

Kazimieras Pukėnas

**SPORTINIŲ TYRIMŲ DUOMENŲ
ANALIZĖ *SPSS* PROGRAMA**

LIETUVOS KŪNO KULTŪROS AKADEMIJA

Kazimieras Pukėnas

**SPORTINIŲ TYRIMŲ DUOMENŲ
ANALIZĖ *SPSS* PROGRAMA**

MOKOMOJI KNYGA

KAUNAS
2005

Mokomąją knygą recenzavo
Doc. dr. Virgilijus SAKALAUŠKAS, doc. dr. Rimantas PLĖŠTYS

2004-12-21 Nr.3
Lietuvos kūno kultūros akademijos Senato rekomenduota

© Lietuvos kūno kultūros akademija, 2005

PUKĖNAS, Kazimieras
Sportinių tyrimų duomenų analizė SPSS programa: mokomoji knyga /
Kazimieras Pukėnas; Lietuvos kūno kultūros akademija. – Kaunas:
LKKA, 2005. – 289 p.

ISBN 9955 – 622 – 18 – 0

TURINYS

PRATARMĖ	8
1. BENDROSIOS ŽINIOS.....	9
1.1. SPSS PROGRAMOS ĮDIEGIMAS	9
1.2. SPSS PROGRAMOS LANGAI.....	9
2. DUOMENŲ REDAKTORIUS	11
2.1. DUOMENŲ PERŽIŪRA.....	11
2.2. KINTAMŲJŲ PERŽIŪRA.....	12
2.2.1. Kintamųjų vardai.....	13
2.2.2. Kintamųjų tipai.....	13
2.2.3. Stulpelio formatas (<i>Width</i>)	14
2.2.4. Dešimtainės skiltys (<i>Decimals</i>).....	14
2.2.5. Kintamojo žymena (<i>Label</i>)	15
2.2.6. Reikšmės žymenos (<i>Values</i>)	15
2.2.7. Praleistos reikšmės (<i>Missing values</i>)	15
2.2.8. Stulpelio plotis (<i>Columns</i>)	16
2.2.9. Įrašo išdėstymas (<i>Alignment</i>)	16
2.2.10. Matavimo skalė (<i>Measure</i>)	16
2.3. DUOMENŲ ĮRAŠYMAS IR REDAGAVIMAS	17
2.3.1. Duomenų kopijavimas ir perkėlimas	18
2.3.2. Naujų kintamųjų reikšmių įterpimas.....	19
2.3.3. Naujų kintamųjų įterpimas	19
2.3.4. Kintamųjų perkėlimas	20
2.3.5. Greitas reikiamo įrašo pasirinkimas.....	20
2.3.6. Duomenų rinkmenos išsaugojimas	20
2.3.7. Pagrindiniai įrankių juostos mygtukai	21
3. DUOMENŲ TRANSFORMAVIMAS	23
3.1. NAUJŲ KINTAMŲJŲ SKAIČIAVIMAS	23
3.1.1. Loginis naujo kintamojo skaičiavimas.....	24
3.1.2. Atsitiktinių skaičių generatoriaus pradinės reikšmės nustatymas.....	25
3.2. PASIKARTOJANČIŲ REIKŠMIŲ SKAIČIAUS NUSTATYMAS.....	26
3.3. REIKŠMIŲ PERKODAVIMAS.....	28
3.3.1. Reikšmių perkodavimas nekeičiant kintamojo	28
3.3.2. Reikšmių perkodavimas įvedant naujus kintamuosius	30
3.3.3. Automatinis perkodavimas	32
3.4. DUOMENŲ RANGAVIMAS	33
3.5. KATEGORINIS TRANSFORMAVIMAS.....	35
3.6. LAIKO EILUČIŲ DUOMENŲ TRANSFORMAVIMAS	35
4. DUOMENŲ RINKMENŲ TVARKYMAS IR TRANSFORMAVIMAS.....	38
4.1. DUOMENŲ RIKIAVIMAS	38
4.2. KINTAMŲJŲ IR STEBĖJIMŲ SUKEITIMAS VIETOMIS	39
4.3. BYLŲ SUJUNGIMAS.....	40
4.3.1. Kaip sujungti bylas su tais pačiais kintamaisiais?	40

4.3.2. Kaip sujungti bylas su skirtingais kintamaisiais?	41
4.4. DUOMENŲ AGREGAVIMAS	43
4.5. STEBĖJIMŲ DUOMENŲ ATRANKA	44
4.5.1. Duomenų atranka pagal loginę sąlygą	46
4.5.2. Atsitiktinė duomenų atranka	47
4.5.3. Diapazoninė duomenų atranka	47
4.6. DUOMENŲ RINKMENOS SKĖLIMAS	48
5. GAUTŲ REZULTATŲ REDAGAVIMAS	50
5.1. VEIKSMAI <i>VIEWER</i> LANGE	51
5.1.1. Kaip parodyti arba paslėpti rezultatus?	51
5.1.2. Kaip kopijuoti, perkelti arba pašalinti gautus rezultatus?	52
5.1.3. Kaip pakeisti rezultatų lygmenį?	53
5.1.4. Kaip pakeisti <i>Outline</i> sąrašo šriftą?	53
5.1.5. Kaip įterpti naują objekto pavadinimą ar papildomą tekstą?	53
5.2. REZULTATŲ IŠSAUGOJIMAS	53
5.3. IŠVESTIES DOKUMENTO SPAUSDINIMAS	54
6. MOBILIOSIOS LENTELĖS	56
6.1. MOBILIŲJŲ LENTELIŲ STRUKTŪROS KEITIMAS	56
6.1.1. Kaip redaguoti lenteles naudojant piktogramas?	56
6.1.2. Kaip sukeisti vietomis eilutes ir stulpelius?	57
6.1.3. Kaip pakeisti lentelės elementų eiliškumą?	57
6.1.4. Kaip perkelti eilutes ir stulpelius?	57
6.1.5. Kaip suskirstyti eilutes ir stulpelius į grupes?	57
6.1.6. Kaip pakeisti eilučių ir stulpelių pavadinimų kryptį?	58
6.1.7. Kaip atkurti originalų lentelės pavidalą?	58
6.1.8. Kaip paslėpti lentelės ląsteles?	58
6.2. MOBILIŲJŲ LENTELIŲ REDAGAVIMAS	58
6.2.1. Kaip pakeisti lentelės išvaizdą?	59
6.2.2. Kaip pakeisti lentelės savybes?	59
6.2.3. Kaip pakeisti atskirų ląstelių savybes?	62
6.2.4. Kaip pakeisti pateiktus rezultatus?	63
7. TRUMPA STATISTIKOS METODŲ APŽVALGA	65
7.1. PIRMINĖS STATISTIKOS TESTO SĄLYGOS	65
7.1.1. Duomenų matavimo skalės	65
7.1.2. Pasiskirstymo pagal normalųjį skirstinį patikrinimas	66
7.1.3. Priklausomos ir nepriklausomos imtys	67
7.2. STATISTINIS KRITERIJUS	67
7.3. STATISTIKOS METODAI	68
7.3.1. Aprašomoji statistika	68
7.3.2. Analitinė statistika	69
8. DAŽNIŲ ANALIZĖ	71
8.1. DAŽNIŲ LENTELĖS	71
8.2. SKAITINĖS DUOMENŲ CHARAKTERISTIKOS	72

8.3. DAŽNIŲ DIAGRAMOS	74
8.3.1. Kaip tvarkyti teksto formatus?	76
8.3.2. Kaip redaguoti diagramos pavadinimą?	77
8.3.3. Kaip tvarkyti diagramos ašių formatus?	77
8.3.4. Kaip tvarkyti duomenų sekų formatus?	79
8.3.5. Kaip tvarkyti kitų diagramos komponentų formatus?	81
8.4. REZULTATŲ FORMATAI	82
9. SKAITINĖS CHARAKTERISTIKOS	83
9.1. APRAŠOMOJI STATISTIKA	83
9.2. STEBĖJIMŲ SUVESTINĖ	85
9.3. DUOMENŲ TYRIMAS (EXPLORE).....	88
10. POŽYMIŲ DAŽNIŲ LENTELĖS	99
10.1. POŽYMIŲ DAŽNIŲ LENTELIŲ SUDARYMAS.....	99
10.2. POŽYMIŲ DAŽNIŲ LENTELIŲ DIAGRAMOS	103
10.3. POŽYMIŲ DAŽNIŲ LENTELIŲ KRITERIJAI	105
10.3.1. Chi-kvadrato (χ^2) testas	106
10.3.2. Koreliacijos koeficientai	108
10.3.3. Nominaliųjų kintamųjų ryšio matai	109
10.3.4. Ranginių kintamųjų ryšio matai	112
10.3.5. Kiti tarpusavio ryšio matai	112
11. HIPOTEZIŲ APIE VIDURKIŲ LYGYBĘ TIKRINIMAS	116
11.1. t-TESTAS NEPRIKLAUSOMOMS IMTIMS	118
11.2. t-TESTAS PRIKLAUSOMOMS IMTIMS	121
12. NEPARAMETRINIAI KRITERIJAI	124
12.1. DVIEJŲ NEPRIKLAUSOMŲ IMČIŲ PALYGINIMAS	124
12.1.1. Mann'o ir Whitney U testas	125
12.1.2. Kolmogorovo ir Smirnovu Z testas.....	125
12.1.3. Wald'o ir Wolfowitz'o testas	125
12.1.4. Mozes'o testas	126
12.2. DVIEJŲ PRIKLAUSOMŲ IMČIŲ PALYGINIMAS	129
12.2.1. Wilcoxon'o testas	129
12.2.2. Ženklu testas.....	131
12.2.3. McNemar'o testas	132
12.2.4. Ribinis homogeniškumo testas.....	134
12.3. KELIŲ NEPRIKLAUSOMŲ IMČIŲ PALYGINIMAS	134
12.3.1. Kruscal'o ir Wallis'o H testas	134
12.4. KELIŲ PRIKLAUSOMŲ IMČIŲ PALYGINIMAS	136
12.4.1. Friedman'o testas	137
12.4.2. Kendall'o W testas	139
12.4.3. Cochran'o Q testas	140
12.5. KOLMOGOROVO IR SMIRNOVO TESTAS	141
12.6. CHI-KVADRATO TESTAS	142
13. KORELIACIJOS RYŠYS.....	146

13.1. PIRSONO KORELIACIJOS KOEFICIENTAS.....	146
13.2. SPIRMENO IR KENDALL'O KORELIACIJOS	148
KOEFICIENTAI	148
13.3. DALINĖ KORELIACIJA.....	148
13.4. SUTAPIMO IR SKIRTINGUMO MATAI	150
14. REGRESINĖ ANALIZĖ	154
14.1. PAPRASTA TIESINĖ REGRESIJA	154
14.1.1. Regresijos lygties skaičiavimas	155
14.1.2. Naujų kintamųjų išsaugojimas	157
14.1.3. Regresijos tiesės braižymas	159
14.1.4. Kitos tiesinės regresijos charakteristikos	163
14.2. DAUGELIO FAKTORIŲ TIESINĖ REGRESIJA	164
14.3. BINARINĖ LOGISTINĖ REGRESIJA.....	166
14.4. DAUGIAREIKŠMĖ LOGISTINĖ REGRESIJA	171
14.5. APROKSIMACIJA KREIVĖMIS	177
15. DISPERSINĖ ANALIZĖ.....	180
15.1. VIENO FAKTORIAUS DISPERSINĖ ANALIZĖ.....	180
15.1.1. Aposterioriniai (<i>Post Hoc</i>) kriterijai	182
15.1.2. Aprioriniai kriterijai	185
15.1.3. Vidurkių trendas.....	186
15.1.4. Pasirinkimai (<i>Options</i>)	187
15.2. BLOKUOTŲJŲ DUOMENŲ VIENO FAKTORIAUS DISPERSINĖ ANALIZĖ.....	191
15.3. DVIEJŲ FAKTORIŲ DISPERSINĖ ANALIZĖ	201
15.4. DAUGIAMATĖ DISPERSINĖ ANALIZĖ	215
15.5. BLOKUOTŲJŲ DUOMENŲ DVIEJŲ FAKTORIŲ DISPERSINĖ ANALIZĖ.....	224
16. KLASTERINĖ ANALIZĖ.....	237
16.1. OBJEKTŲ PANAŠUMO MATAI	238
16.1.1. Intervalinių kintamųjų atstumo matai	238
16.1.2. Stebėjimų dažniais išreikštų kintamųjų atstumo matai.....	239
16.1.3. Binarinių kintamųjų atstumo ir panašumo matai	240
16.2. JUNGIMO METODAI	241
16.3. HIERARCHINĖ KLASTERINĖ ANALIZĖ	242
16.4. <i>k</i> -VIDURKIŲ KLASTERINĖ ANALIZĖ.....	250
17. GRAFIKAI.....	253
17.1. DIALOGINIAI GRAFIKAI.....	254
17.1.1. Stulpelinės diagramos	254
17.1.2. Histogramos	263
17.1.3. Sklaidos diagramos	267
17.2. DIALOGINIŲ GRAFIKŲ REDAGAVIMAS	273
17.2.1. Kaip pažymėti keičiamus diagramos objektus?.....	273
17.2.2. Kaip perkelti diagramos objektus?.....	273

17.2.3. Kaip modifikuoti trimates diagramas?	273
17.2.4. Diagramų vedlys	274
17.2.5. Kaip įterpti į diagramą papildomus elementus?.....	275
17.2.6. Kaip tvarkyti diagramos duomenų sritį?	275
17.2.7. Kaip keisti diagramos objektų planą?	277
17.2.8. Kaip keisti diagramos legendą?	277
17.2.9. Kaip keisti diagramos ašis?.....	280
17.2.10. Dialogo langelis <i>Chart Properties</i>	284
17.2.11. Spalvų ir stilių įrankiai	285
17.2.12. Kaip pakeisti šriftą?.....	286
DALYKINĖ RODYKLĖ	287
LITERATŪRA	289

PRATARMĖ

Šiuolaikinė statistika yra neatsiejama nuo kompiuterinės duomenų analizės, padedančios greitai ir efektyviai spręsti įvairius statistikos uždavinius. Yra daug gerų matematinės statistikos vadovėlių ir mokymo priemonių, tačiau dar labai trūksta lietuviškų knygų, skirtų aprašyti SPSS (angl. — *Statistical Package for the Social Science*) — vienam labiausiai paplitusių statistinės informacijos apdorojimo programinių paketų, tinkamų ir pradedančiajam, ir patyrusiam vartotojui.

Pagrindinis SPSS programinio paketo privalumas — didelė šiuolaikinių statistinių analizės metodų pasirinktis bei duomenų analizės rezultatų vizualizavimo priemonių (duomenų pateikimo lentelių, diagramų, skirstinių kreivių) įvairovė, lengvai įvaldoma dialoginė sąsaja. SPSS programinis paketas taikomas sociologijos, psichologijos, biologijos, medicinos, rinkodaros, kokybės valdymo procese.

Šios knygos tikslas — supažindinti studentus su šiuolaikiniu ir veiksmingu taikomosios statistikos uždavinių sprendimo įrankiu (nagrinėjama SPSS 11.0 versija). Knygoje aprašomi duomenų rinkmenų sudarymo ir redagavimo principai, taip pat tie vienmačiai ir daugiamačiai statistikos metodai ir modeliai, kurie plačiausiai taikomi sportiniams tyrimams: vidurkių palyginimas, koreliacija, regresinė analizė, dispersinė analizė, klasterinė analizė, neparametriniai metodai. Daugiausia dėmesio skiriama šių metodų panaudojimo ypatumams, jų galimybių analizei, gautų rezultatų interpretacijai ir grafinei jų pateikčiai. Visų metodų taikymas iliustruojamas konkrečiais sportiniais pavyzdžiais. Knyga orientuota į skaitytoją, išklausiusį matematinės statistikos kursą. Skaitytojui glaustai primenamos tik tos matematinės statistikos sąvokos ir teiginiai, kurių žinojimas yra būtinas, norint suvokti taikomų metodų esmę ir sėkmingai dirbti su SPSS programiniu paketu. Išsamesnių žinių apie statistinius tyrimo metodus skaitytojas gali taip pat pasisemti iš puikių statistikos vadovėlių (Čekanavičius, Murauskas, 2000), (Čekanavičius, Murauskas, 2002), (Gonestas, Strielčiūnas, 2003), (Sakalauskas, 2003), skirtų skaitytojams, neturintiems ypatingo matematinio parengtumo. Be statistikos pradmenų, skaitytojas turi mokėti dirbti su asmeniniu kompiuteriu *Windows* aplinkoje.

Ši mokomoji knyga skiriama Lietuvos kūno kultūros akademijos ir kitų Lietuvos aukštųjų mokyklų sporto specialybių bakalauro ir magistro studijų studentams, doktorantams.

Autorius nuoširdžiai dėkoja Vilniaus universiteto docentui **V. Sakalauskui** ir Kauno technologijos universiteto docentui **R. Plėščiiui** už nuodugną leidinio rankraščio recenzavimą, kritines pastabas ir naudingus patarimus, leidusius reikšmingai pagerinti pateiktą medžiagą. Autorius taip pat dėkoja LKKA redaktorei **A. Pribušauskaitei** už leidinio kalbinį redagavimą.

1. BENDROSIOŠ ŽINIOS

1.1. SPSS PROGRAMOS ĮDIEGIMAS

Norėdami įdiegti SPSS programą Windows terpėje atlikite tolesnius veiksmus:

- Įdėkite SPSS programos įdiegimo kompaktinį diską į kompiuterio kompaktinių diskų kaupiklį. **My Computer** lange spragtelėkite kompaktinio disko įrenginį.
- Programos įdiegimo dialogo lange spragtelėkite *Install SPSS* ir sekite instrukcijas ekrane.
- Į dialogo langelį **User Information** įveskite vartotojo duomenis, taip pat programos registracijos numerį.
- Spragtelėkite dialogo langelio mygtuką **Next** ir pasirinkite vieną iš trijų programos įdiegimo variantų: standartinį (*Standard*), minimalų (*Minimal*) ar specialų (*Custom*), kuriame galite nurodyti įdiegiamus programos komponentus. Nustatytasis variantas yra *Standard*, kuris rekomenduojamas daugumai vartotojų.
- Spragtelėkite dialogo langelio mygtuką **Next** ir į dialogo langelį **Product license codes** įveskite reikaujamą licencijos kodą. Atkreipkite dėmesį į tai, kad kodo skaitmenų grupės turi būti atskirtos tarpais.
- Spragtelėkite dialogo langelio mygtuką **Next** ir sekite tolesnes instrukcijas ekrane. Spragtelėkite mygtuką **Finish** paskutiniame dialogo langelyje. Pasibaigus licencijos galiojimo laikui, ją galima atnaujinti taip:
 - **Program Files** lange dukart spragtelėkite pele SPSS katalogą, suraskite programos *licrenew.exe* ženklelį, dukart spragtelėkite jį ir atsidariusiame langelyje įrašykite naujos licencijos kodą.

1.2. SPSS PROGRAMOS LANGAI

SPSS programoje yra šie langai:

Duomenų redaktorius (*Data Editor*), kuris atidaromas automatiškai paleidus SPSS programą (nurodžius komandas **Start → Programs → SPSS**). Į jį galima įvesti naujus duomenis arba meniu komandomis **File → Open...** atverti bylas su esamais duomenimis. Savo išvaizda duomenų redaktoriaus langas primena elektroninę lentelę, t. y. darbo lapą, suskirstytą eilutėmis ir stulpeliais, kuris leidžia lengvai ir efektyviai įrašyti naujus duomenis arba modifikuoti esamus. Vienu metu galima atidaryti vieną duomenų redaktoriaus langą.

Rodiny (*Viewer*). Rodinio lange pateikiami visi statistiniai rezultatai, lentelės ir grafikai. Pateiktus duomenis galima redaguoti ir išsaugoti. Rodinio langas

atidaromas automatiškai, kai atliekama komanda, po kurios pateikiami rezultatai.

Eskizinis rodinys (*Draft Viewer*). Eskizinio rodinio lange duomenis galima pateikti kaip paprastą tekstą (vietoje dialoginių dinaminių lentelių).

Dinaminių lentelių redaktorius (*Pivot Table Editor*). Dinaminių lentelių redaktorius suteikia plačias galimybes keisti gautų rezultatų pateikimą — redaguoti tekstą, sukeisti eilučių ir stulpelių duomenis, nuspalvinti, sukurti daugiamates lenteles bei pasirinktinai rodyti duomenis.

Diagramų redaktorius (*Chart Editor*). Leidžia modifikuoti didelės skiriamosios gebos diagramas ir grafikus — keisti spalvas, pasirinkti įvairių tipų ir dydžių šriftą, pakeisti horizontaliąsias ir vertikaląsias ašis, pasukti trimates diagramas, taip pat pakeisti diagramos tipą.

Tekstinio pateikimo redaktorius (*Text Output Editor*). Leidžia redaguoti tekstą, kuris nėra pateikiamas dinaminėje lentelėje.

Sintaksės redaktorius (*Syntax Editor*). Leidžia redaguoti komandų sintaksę, kad būtų galima panaudoti SPSS galimybes, kurios nėra prieinamos per dialogo skydelius.

2. DUOMENŲ REDAKTORIUS

Duomenų redaktorius užtikrina patogų, elektroninės lentelės principu paremtą duomenų rinkmenų sudarymą ir redagavimą. Atskirose elektroninės lentelės eilutėse įrašomi paskirų stebėjimų duomenys (*Cases*), o atskiruose stulpeliuose — atskiri kintamieji (*Variables*). Duomenų redaktorius užtikrina du duomenų rinkmenų pateikimo vaizdus:

- **Duomenų peržiūra** (*Data View*). Pateikia duomenų reikšmes arba duomenų apibūdinimo žymes.
- **Kintamųjų peržiūra** (*Variable View*). Pateikia kintamuosius apibūdinančią informaciją: kintamojo ir kintamojo žymenos pavadinimą, duomenų tipą, matavimo skalę, praleistas reikšmes.

Abiejų roдиниų languose galima papildyti, keisti arba pašalinti duomenų rinkmenoje esančią informaciją.

2.1. DUOMENŲ PERŽIŪRA

	mvj_la	mvj_a5	mvj_a10	mvj_a30	mvj_a60	p1_la	p1_a5	p1
1	,4230	,2800	,2800	,3060	,4120	,0640	,0230	
2	,5650	,3430	,3150	,3200	,3150	,0460	,0110	
3	,5500	,4280	,4250	,4710	,3500	,0350	,0050	
4	,4840	,3960	,3780	,4210	,4620	,0220	,0094	
5	,4900	,4750	,4560	,4780	,4430	,0265	,0100	
6	,5500	,4500	,4560	,5150	,5090	,0190	,0069	
7	,4180	,2280	,2960	,2960	,3370	,0175	,0050	
8	,5930	,4650	,4530	,5120	,4810	,0437	,0090	
9	,3810	,3430	,3460	,3310	,3310	,0306	,0130	
10	,5310	,3310	,3150	,2750	,2900	,0480	,0296	
11	,3460	,2680	,2660	,2870	,3030	,0218	,0200	
12								
13								
14								
15								
16								

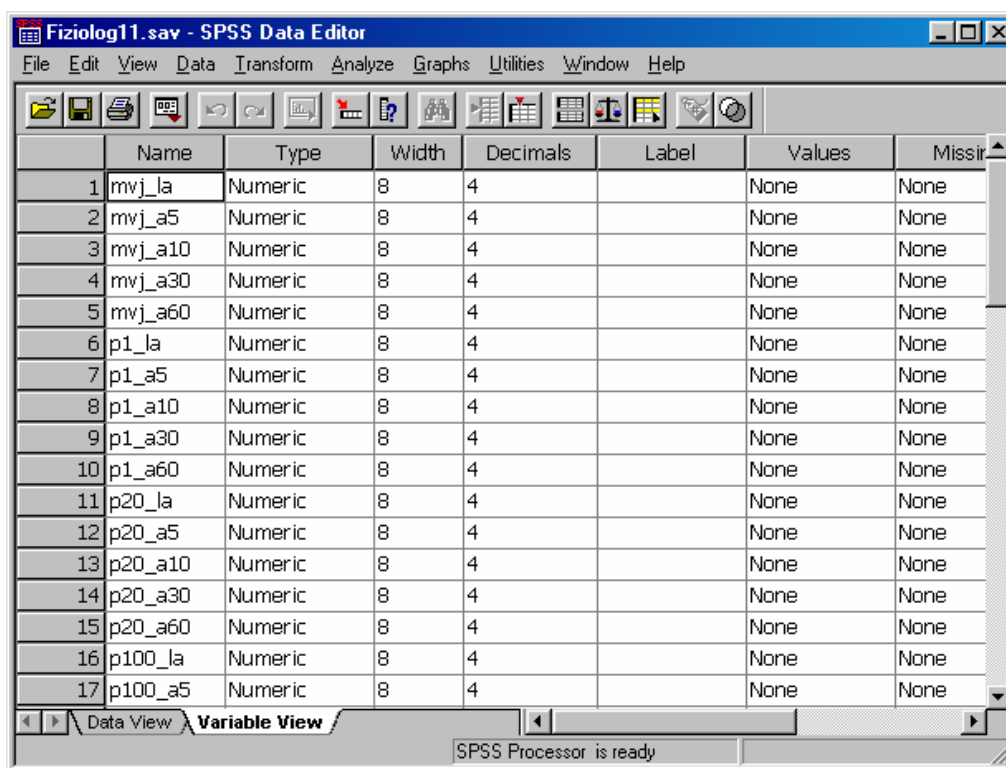
2.1 pav. Duomenų peržiūros langas **Data View**

Dauguma duomenų peržiūros lango savybių yra panašios į kitų programų naudojamų elektroninių lentelių savybes, tačiau yra ir esminių skirtumų:

- Eilutės atitinka atskirus stebėjimus (*Cases*). Pavyzdžiui, atskirų respondentų atsakymai anketoje konkrečiu klausimu.

- Stulpeliai atitinka kintamuosius arba jų charakteristikas (*Variables*). Pavyzdžiui, kiekvienas anketos klausimas yra atskiras kintamasis.
- Eilučių ir stulpelių sankirtoje esančiose ląstelėse įrašomos duomenų reikšmės. Kiekvienoje ląstelėje gali būti įrašyta viena konkretaus kintamojo stebima reikšmė. Skirtingai nuo kitų programų elektroninių lentelių SPSS ląstelėse gali būti tik duomenų reikšmės ir negali būti formulų.
- Duomenų rinkmena yra stačiakampio formos. Stačiakampio dydį lemia kintamųjų ir jų reikšmių skaičius. Negali būti „tuščių“ ląstelių duomenų rinkmeną apibrėžiančiose ribose. Kintamiesiems su skaitmeninėmis reikšmėmis neužpildytos ląstelės yra paverčiamos ląstelėmis su sistemiškai praleistomis reikšmėmis, o raidinės-skaitmeninės sekos (*string*) tipo kintamiesiems neužpildytos ląstelės yra laikomos galiojančiomis.

2.2. KINTAMŲJŲ PERŽIŪRA



2.2 pav. Kintamųjų peržiūros langas *Variable View*

Kintamųjų peržiūros lange pateikiamos kiekvieno duomenų rinkmenos kintamojo savybės. Eilutėse yra nurodomi kintamieji, o stulpeliuose — jų savybės. Į šį langą galima įterpti naujus arba pašalinti esamus kintamuosius, taip pat nurodyti šias kintamųjų savybes ir požymius:

- Kintamojo vardą.
- Duomenų tipą.
- Ženklių skaičių.
- Ženklių po kablelio skaičių.
- Kintamojo bei reikšmių žymenas.
- Vartotojo apibrėžtas praleistas reikšmes.
- Stulpelių plotį.
- Matavimų skalę.

Duomenų ir kintamųjų peržiūros langai duomenų redaktoriaus režime atidaromi spragtelint pele atitinkamo lango (*Data View* arba *Variable View*) auselę.

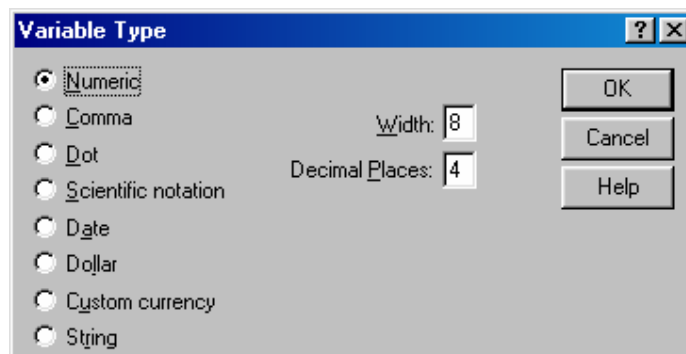
2.2.1. Kintamųjų vardai

Kintamojo vardas yra įrašomas lauko *Name* eilutėje. Kintamojo vardui keliami šie reikalavimai:

- Kintamojo vardas turi prasidėti raide. Kiti simboliai gali būti lotyniškos raidės, skaitmenys, taškas, simboliai @, #, _, \$. Neleidžiami tarpai ir specialūs simboliai, pvz., !, ?, ' , *.
- Paskutinis kintamojo vardo simbolis negali būti taškas arba pabraukimas.
- Kintamojo vardą turi sudaryti ne daugiau kaip aštuoni simboliai.
- Didžiosios ir mažosios raidės traktuojamos vienodai.

2.2.2. Kintamųjų tipai

Pagal nustatymą SPSS programa įvedamus naujus duomenis traktuoja kaip skaitmeninius su didžiausiu aštuonių ženklų skaičiumi ir dviem ženklais po kablelio. Kintamojo tipui pakeisti spragtelėkite pele mygtuką su trimis taškais *Type* ląstelės dešinėje. Atsidarys kintamojo tipo nustatymo dialogo langelis *Define Variable Type*.




2.3 pav. Dialogo langelis *Variable Type* kintamojo tipo nustatymui (pažymėjus **Numeric**)


SPSS yra numatyti šie kintamųjų tipai:

Skaitmeninis (<i>Numeric</i>)	Tai teigiami arba neigiami skaičiai. Prieš neigiamą skaičių rašomas minuso ženklas (prieš teigiamą skaičių pliusas praleidžiamas), o skaičiaus dešimtainė dalis atskiriama dešimtainiu skyrikliu (lietuviškoje vartosenoje — kableliu). Teksto laukelyje <i>Width</i> įrašomas didžiausias ženklų skaičius, įskaitant dešimtainės dalies ženklą. Teksto laukelyje <i>Decimals</i> įrašomas rodomas pozicijų po kablelio skaičius. Įrašant pačius duomenis galima rašyti iki šešiolikos ženklų po kablelio ir jie visi bus naudojami skaičiuojant, nors rodomas bus tik nustatytas ženklų skaičius.
Kablelis (<i>Comma</i>)	Galimos reikšmės — teigiami arba neigiami skaičiai. Naudojamas dešimtainis skyriklis — taškas. Vienas arba keli kableliai naudojami dešimtainių skilčių grupėms (tūkstančiams, milijonams) atskirti.
Taškas (<i>Dot</i>)	Galimos reikšmės — teigiami arba neigiami skaičiai. Naudojamas dešimtainis skyriklis — kablelis. Vienas arba keli taškai naudojami dešimtainių skilčių grupėms atskirti.
EkspONENTINIS formatas (<i>Scientific notation</i>)	Galimos visos skaitinės reikšmės, įskaitant eksponentinį pavaizdavimą, kuris nurodomas raide E arba D ir pliuso ar minuso ženklu.
Data (<i>Date</i>)	Galimos datos ir/arba laiko reikšmės. Pageidaujamą formatą galima pasirinkti išskleidžiamajame sąraše, kuris pasirodo dialogo langelyje <i>Variable Type</i> pažymėjus <i>Date</i> .
Doleris (<i>Dollar</i>)	Naudojamas dolerio ženklas, dešimtainis skyriklis — taškas, dešimtainių skilčių grupių skyriklis — kablelis.
Eilutė (<i>String</i>)	Simbolių eilutė. Galimos reikšmės — raidės, skaitmenys, specialūs simboliai. Paprastai naudojamos ne ilgesnės kaip aštuonių ženklų eilutės. Su tokiais kintamaisiais negalima atlikti skaičiavimo operacijų, bet galima, pvz, suskaičiuoti kartojimąsi.

2.2.3. Stulpelio formatas (*Width*)

Stulpelio formatas *Width* (didžiausias ženklų skaičius) nustatomas dialogo langelyje *Variable Type* arba suktuku , spragtelėjus pele atitinkamo kintamojo *Width* ląstelės dešinėje.

2.2.4. Dešimtainės skiltys (*Decimals*)

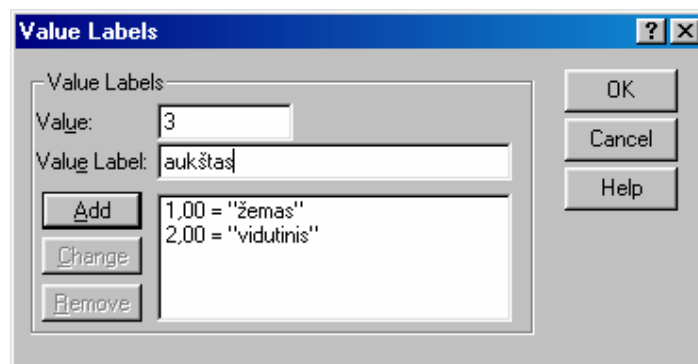
Ženklų skaičius po kablelio (*Decimals*) nustatomas dialogo langelyje *Variable Type* arba suktuku , spragtelėjus pele atitinkamo kintamojo *Decimals* ląstelės dešinėje.

2.2.5. Kintamojo žymena (*Label*)

Kintamojo žymena gali turėti kur kas daugiau simbolių negu kintamojo vardas — iki 256 simbolių ir leidžia detaliau apibūdinti kintamąjį.

2.2.6. Reikšmės žymenos (*Values*)

Reikšmės žymenų paskirtis — smulkiau aprašyti galimas kintamojo reikšmes, ypač tada, kai duomenų rinkmenoje naudojami skaitmeniniai kodai neskaitmeniniam kintamajam. Pavyzdžiui, kintamajam *lytis* galima suteikti žymeną „moteris“ reikšmei „1“ ir žymeną „vyras“ reikšmei „2“. Reikšmės žymenoms įrašyti spragtelėkite pele atitinkamo kintamojo *Values* ląstelės dešinėje mygtuką su trimis taškais. Atsidarys žymenų nustatymo dialogo langelis *Value Labels*.



2.4 pav. Žymenų nustatymo dialogo langelis *Value Labels*

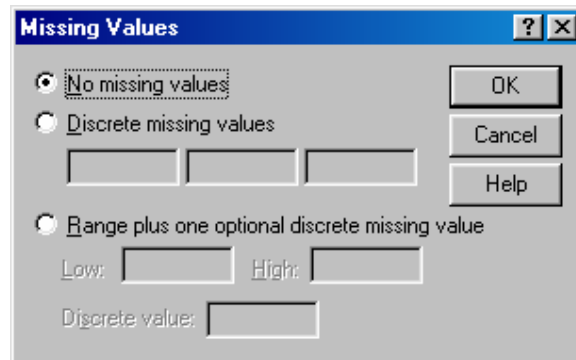
Dialogo langelyje įrašomos kintamojo reikšmės ir jų žymenos. Kiekvienas įrašas įvedamas spragtelint pele *Add* mygtuką.

2.2.7. Praleistos reikšmės (*Missing values*)

SPSS yra galimos dviejų tipų praleistos reikšmės:

- Sisteminės (*System – defined missing values*). Jeigu duomenų matricoje yra neužpildytų skaitmeninių ląstelių, SPSS programa identifikuoja jas kaip praleistas reikšmes. Tokiose ląstelėse įrašomas kablelis.
- Nustatytos vartotojo (*User – defined missing values*). Jeigu dėl kokių nors priežasčių trūksta kintamojo duomenų, pvz., jeigu anketoje nebuvo atsakyta į klausimą, vartotojas gali šiuos duomenis laikyti negaliojančiais ir eliminuoti juos iš tolesnių skaičiavimų;


Praleistoms reikšmėms nustatyti spragtelėkite pele atitinkamo kintamojo *Missing* ląstelės dešinėje mygtuką su trimis taškais. Atsidarys praleistų reikšmių nustatymo dialogo langelis *Missing Values*.



2.5 pav. Praleistų reikšmių nustatymo dialogo langelis *Missing Values*

Galima nustatyti iki trijų atskirų praleistų reikšmių arba reikšmių diapazoną ir vieną atskirą reikšmę. Gali būti nustatytas tik skaitmeninių kintamųjų reikšmių diapazonas. Pagal nustatymą SPSS programa traktuoja duomenis be praleistų reikšmių (*No missing values*).

2.2.8. Stulpelio plotis (*Columns*)

Stulpelio plotis *Columns* (pavaizduojamų ženklų skaičius) gali būti keičiamas suktuku , spragtelėjus pele atitinkamo kintamojo *Columns* ląstelės dešinėje arba duomenų peržiūros lange pele velkant stulpelio kraštinę. Stulpelio plotis turi įtakos tik duomenų pavaizdavimui ir nesvarbus įvestų duomenų formatui.

2.2.9. Įrašo išdėstymas (*Alignment*)

Nustatytas simbolių išdėstymas duomenų peržiūros lango ląstelėse yra toks: „right“ (tekstas pritraukiamas prie dešiniojo narvelio krašto) — skaitmeninio formato kintamųjų ir „left“ (tekstas pritraukiamas prie kairiojo narvelio krašto) — simbolių eilutės formato kintamųjų. Keisti simbolių išdėstymą galima spragtelėjus pele išskleidžiamojo sąrašo mygtuką atitinkamo kintamojo *Alignment* ląstelės dešinėje.

2.2.10. Matavimo skalė (*Measure*)

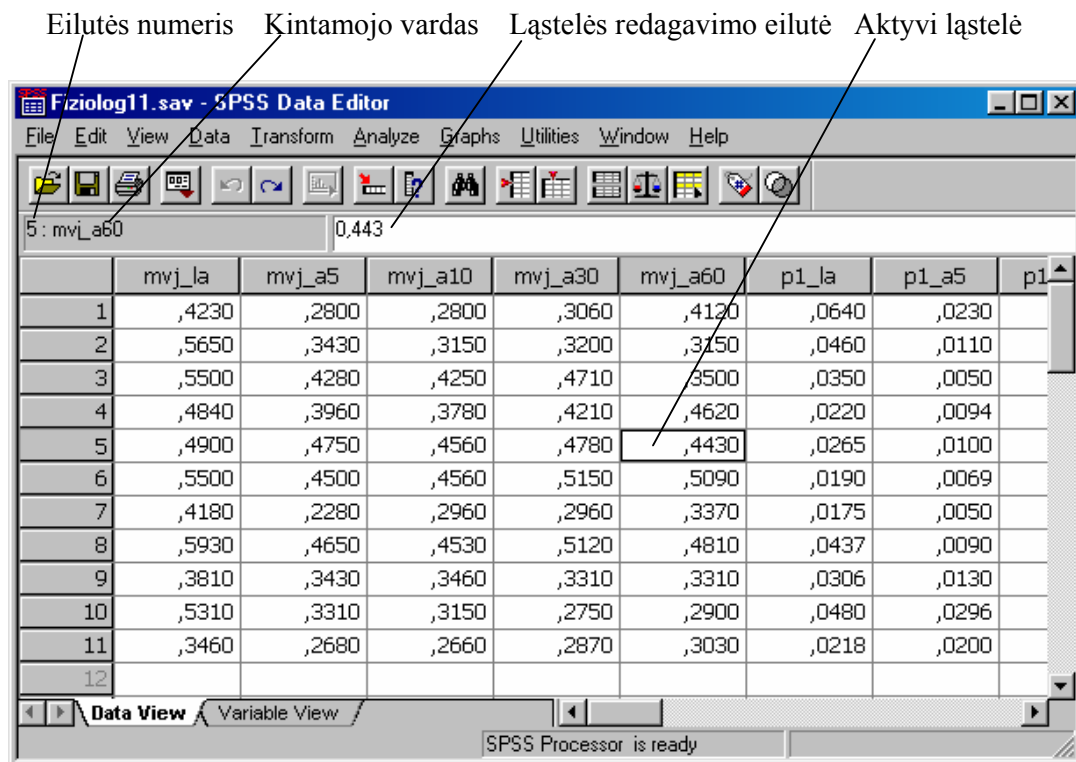
Galimos trys kintamojo skalės: nominalinė (vardinė) — *Nominal*, rangų — *Ordinal* ir metrinė — *Scale*. Nustatyta yra metrinė skalė. Skalių skirtumas turi reikšmę sudarant dialoginius (*interactive*) grafikus, kur vardinė ir rangų

skalės sujungiamos į kategorinę (*Categorical*) tipą. Keisti skalės tipą galima spragtelėjus pele išskleidžiamojo sąrašo mygtuką atitinkamo kintamojo *Measure* ląstelės dešinėje.

2.3. DUOMENŲ ĮRAŠYMAS IR REDAGAVIMAS

Duomenys įrašomi **Data View** lange. Duomenis galima įrašinėti bet kuria tvarka: pagal kintamuosius arba jų įrašus, į pasirinktą sritį arba į atskiras ląsteles. Duomenų įrašymui reikia:

- Spragtelėti pele pasirinktą ląstelę. Pažymėtoji ląstelė tampa aktyvi — tą liudija rėmelis aplink ląstelę. Aktyviosios ląstelės kintamojo vardas ir eilutės numeris rodomi viršutiniame kairiajame duomenų redaktoriaus kampe.
- Įrašyti klaviatūra naują skaitmeninę ar tekstinę informaciją arba redaguoti esamą įrašą. Prieš įrašant kitokius negu skaitmeniniai duomenis būtina nurodyti kintamojo tipą.
- Duomenų įrašymas ląstelėje baigiamas klavišo **Enter** paspaudimu arba klaviatūra ar pele pažymėjus kitą ląstelę.
- Įrašomi duomenis rodomi taip pat ląstelės redagavimo eilutėje duomenų redaktoriaus lango viršuje.



2.6 pav. Duomenų redaktoriaus langas **Data View**

Įrašant duomenis į tuščią stulpelį, duomenų redaktorius automatiškai sukuria naują kintamąjį ir suteikia jam vardą.

Norint įrašyti duomenis pagal reikšmės žymenas, reikia vykdyti komandas **View → Value Labels**, pažymėti reikiamą ląstelę ir pasirinkti išskleidžiamajame sąraše norimą reikšmės žymeną. Suprantama, kintamajam prieš tai turėjo būti suteiktos reikšmių žymenos.

Nurodytas kintamojo tipas ir formatas sąlygoja duomenis, kuriuos galima įrašyti **Data View** lango ląstelėse:

- Negalima įrašyti duomenų, kurių nenumato pasirinktas kintamojo tipas.
- Gali būti įrašomos skaitmeninių kintamųjų sveikų skaičių reikšmės, kurios viršija nustatytą *Width* formatą, tačiau duomenų redaktorius vaizduoja tas reikšmes eksponentiniu formatu arba į ląstelę įrašytomis žvaigždutėmis, tuo parodoma, kad įrašoma reikšmė viršija nustatytą formatą. Kad būtų matoma ląstelėje įrašyta reikšmė, būtina pakeisti kintamojo formatą. (*Pastaba*. Stulpelio pločio keitimas neturi įtakos kintamojo formatui.)

Įrašytus duomenis galima keisti: pakeisti įrašytas duomenų reikšmes, jas kopijuoti arba perkelti, įterpti arba ištrinti įrašus atskirose eilutėse, įterpti arba ištrinti kintamuosius, keisti kintamųjų tvarką.

Įrašytai reikšmei pakeisti:

- **Data View** lange dukart spragtelėkite pele reikiamą ląstelę. Ląstelės reikšmė bus rodoma taip pat ląstelės redagavimo eilutėje.
- Pakeiskite įrašytą reikšmę tiesiog ląstelėje arba ląstelės redagavimo eilutėje.
- Paspauskite **Enter** klavišą arba pažymėkite kitą ląstelę.

2.3.1. Duomenų kopijavimas ir perkėlimas

Kopijavimo arba perkėlimo veiksmams galima:

- Perkelti arba nukopijuoti pavienės ląstelės reikšmę į kitą ląstelę.
- Perkelti arba nukopijuoti pavienės ląstelės reikšmę į ląstelių grupę.
- Perkelti arba nukopijuoti pavienės eilutės reikšmę į kelias eilutes.
- Perkelti arba nukopijuoti pavienio kintamojo (stulpelio) reikšmę į keletą stulpelių.
- Perkelti arba nukopijuoti duomenų bloką (ląstelių grupę).

Duomenims kopijuoti:

- Pažymėkite kopijuojamą ląstelę, eilutę, stulpelį arba ląstelių grupę ir nurodykite komandas **Edit → Copy**.
- Pažymėkite ląstelę, į kurią kopijuosite, eilutę ar stulpelį arba tik pirmą eilutės ar stulpelio arba tik viršutinę kairiąją bloko, į kurią kopijuosite, ląstelę ir nurodykite komandas **Edit → Paste**.

Duomenims perkelti:

- Pažymėkite perkeliama ląstelę, eilutę, stulpelį arba ląstelių grupę ir vykdykite komandas **Edit → Cut**.
- Pažymėkite ląstelę, kurioje turi atsirasti perkeliama duomenys, eilutę ar stulpelį arba tik pirmą eilutės ar stulpelio arba tik viršutinę kairiąją bloko, į kurią perkeliama duomenys, ląstelę ir vykdykite komandas **Edit → Paste**.

Kopijuojant arba perkeliama duomenis, būtina atkreipti dėmesį į tai, kad sutaptų pirminės, t. y. kopijuojamos arba perkeliama ląstelės, ir paskirties ląstelės, t. y., į kurią kopijuojami arba perkeliama duomenys, kintamųjų tipai. Jeigu kintamųjų tipai minėtose ląstelėse nesutampa, duomenų redaktorius bando pakeisti kopijuojamų arba perkeliama duomenų tipą. Jeigu pakeitimas yra negalimas, paskirties ląstelėje įrašoma sisteminė praleista reikšmė.

2.3.2. Naujų kintamųjų reikšmių įterpimas

Duomenų įrašymas į kurio nors kintamojo tuščios eilutės ląstelę automatiškai sukuria naują visų kintamųjų reikšmių įrašą (*new case*). Įrašomos likusių šio įrašo kintamųjų sisteminės praleistos reikšmės. Tuščios eilutės, esančios tarp naujo įrašo ir esamų įrašų, taip pat traktuojamos kaip nauji įrašai su visų kintamųjų sisteminėmis praleistomis reikšmėmis. Jeigu norite naujas kintamųjų reikšmes (naują eilutę) įterpti tarp esančių įrašų:

- Pažymėkite kurią nors ląstelę eilutėje, virš kurios norite įterpti naują įrašą.
- Nurodykite komandas **Data → Insert Case**.

Bus įterpta nauja eilutė su automatiškai įrašytomis visų kintamųjų sisteminėmis praleistomis reikšmėmis.

2.3.3. Naujų kintamųjų įterpimas

Duomenų įrašymas kuriame nors tuščiame stulpelyje duomenų peržiūros lange (**Data View**) arba kurioje nors tuščioje eilutėje kintamųjų peržiūros lange (**Variable View**) automatiškai sukuria naują kintamąjį nustatytu vardu (su priešdėliu *var* ir eilės numeriu) ir nustatyto (skaitmeninio) tipo su sisteminėmis praleistomis kintamojo reikšmėmis. Tušti stulpeliai tarp naujo kintamojo ir esamų kintamųjų duomenų peržiūros lange arba tuščios eilutės tarp naujo kintamojo ir esamų kintamųjų peržiūros lange taip pat traktuojami kaip nauji kintamieji su sisteminėmis praleistomis kintamųjų reikšmėmis. Jeigu norite įterpti naują kintamąjį tarp jau esančių kintamųjų:

- Pažymėkite kurią nors kintamojo, kurio kairėje norite įterpti naują kintamąjį, ląstelę duomenų peržiūros lange arba kurią nors kintamojo, virš kurios norite įterpti naują kintamąjį, ląstelę kintamųjų peržiūros lange.
- Vykdykite komandas **Data → Insert Variable**.

Bus įterptas naujas kintamasis su automatiškai įrašytomis sisteminėmis praleistomis reikšmėmis.

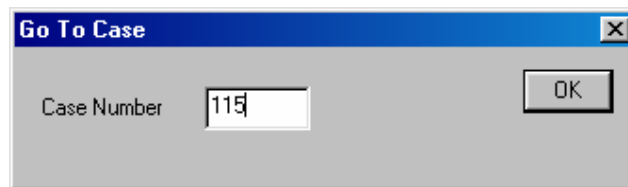
2.3.4. Kintamųjų perkėlimas

Jeigu norite perkelti kintamąjį tarp dviejų esančių kintamųjų, pirmiausia įterpkite naują kintamąjį kaip nurodyta p 2.3.3. Tada:

- Spragtelėkite pele kintamojo, kurį norite perkelti, vardą stulpelio viršuje duomenų peržiūros lange arba eilutės numerį kintamųjų peržiūros lange. Kintamojo stulpelis arba eilutė yra pažymimi.
- Nurodykite komandas **Edit → Cut**.
- Spragtelėkite pele sukurtą kintamojo, į kurio vietą norite perkelti pageidaujamą kintamąjį, vardą stulpelio viršuje duomenų peržiūros lange arba eilutės numerį kintamųjų peržiūros lange. Kintamojo stulpelis arba eilutė yra pažymimi.
- Nurodykite komandas **Edit → Paste**.

2.3.5. Greitas reikiamo įrašo pasirinkimas

Komanda **Go to Case** leidžia greitai pasirinkti reikiamą įrašą (eilutę). Nurodykite komandas **Data → Go to Case...**, pasirodžiusiame langelyje įrašykite norimos eilutės numerį ir spragtelėkite **OK**.



2.7 pav. Dialogo langelis **Go To Case**

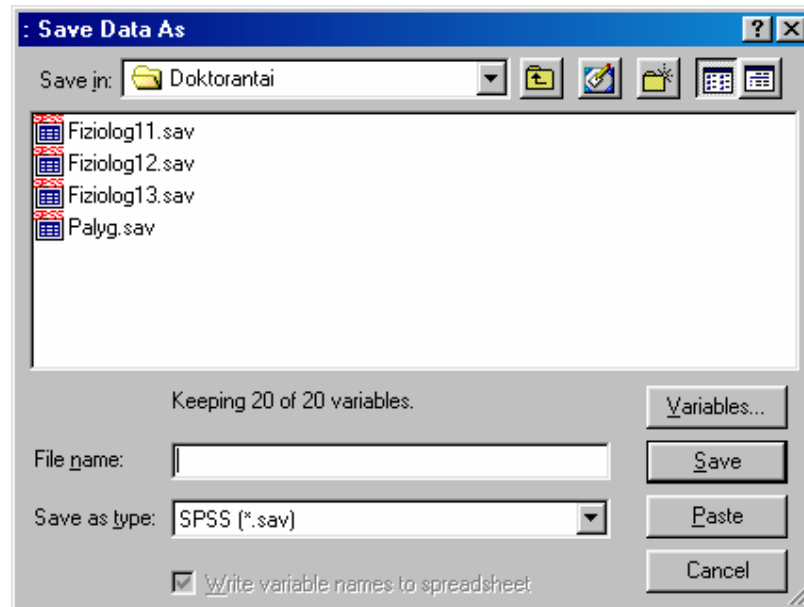
2.3.6. Duomenų rinkmenos išsaugojimas

Norint išsaugoti sukurtą duomenų rinkmeną, būtina atlikti šiuos veiksmus:

- Nurodykite komandas **File → Save As...**
- Pasirodžiusiame dialogo langelyje **Save Data As**, laukelio **Save in** išskleidžiamajame sąrašė dukart spragtelėkite pele reikiamų diskų ir katalogų pavadinimus, kol laukelyje atsiras reikiamo katalogo pavadinimas.
- Laukelyje **File Name** įrašykite bylos vardą.
- Laukelio **Save as type** išskleidžiamajame sąrašė pasirinkite duomenų rinkmenos tipą. Paprastai paliekamas kompiuterio siūlomas variantas **SPSS(*.sav)**.

- Spragtelėkite mygtuką **Save**.

Norint išsaugoti rinkmenoje ne visus kintamuosius, reikia spragtelėti pele dialogo langelio **Save Data As** mygtuką **Variables...** ir pasirodžiusiame dialogo langelyje su kintamųjų sąrašu nurodyti kintamuosius, kurie turi būti išsaugoti.



2.8 pav. Duomenų rinkmenos išsaugojimo dialogo langelis **Save Data As**

2.3.7. Pagrindiniai įrankių juostos mygtukai

Toliau pateikiami dažniausiai vartojami duomenų redaktoriaus įrankių juostos mygtukai, kurie leidžia paspartinti aprašytuosius veiksmus:



Atidaryti bylą (*Open File*).



Išsaugoti bylą (*Save File*). Skirtas bylos, turinčios vardą ir įrašytos į ilgalaikę atmintį, pakartotiniam išsaugojimui po padarytų pakeitimų.



Reikiamos eilutės pasirinkimas (*Go to Case*).



Informacija apie kintamuosius (*Variables*).



Įterpti eilutę (*Insert Cases*). Spragtelėjus pele šį mygtuką, įterpiama nauja eilutė virš pažymėtos ląstelės.



Įterpti kintamąjį (*Insert Variable*). Spragtelėjus pele šį mygtuką, įterpiamas naujas kintamasis kairėje nuo pažymėtojo.



Pasirinkti įrašus (*Select Cases*). Leidžia išrinkti tuos kintamųjų reikšmių įrašus, kurie tenkina tam tikras sąlygas.



Reikšmių žymenos (*Value Labels*). Leidžia pavaizduoti duomenų reikšmes arba reikšmių žymenas.

3. DUOMENŲ TRANSFORMAVIMAS

Kad būtų tenkinamos būtinos sąlygos analizei atlikti ir gauti teisingi rezultatai, pirminius duomenis dažnai tenka modifikuoti.

SPSS turi daug galimybių duomenims transformuoti, nuo duomenų suskirstymo iki naujų kintamųjų sukūrimo, naudojant įvairias matematines išraiškas ir logines sąlygas.

3.1. NAUJŲ KINTAMŲJŲ SKAIČIAVIMAS

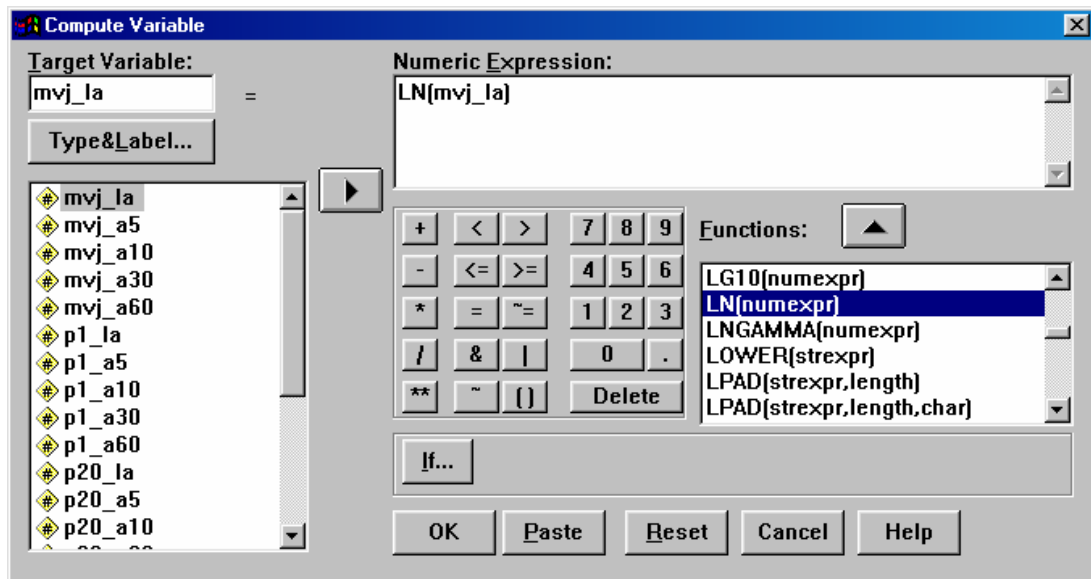
Pagrindinės naujų kintamųjų skaičiavimo galimybės:

- Galima apskaičiuoti tiek skaitmeninių, tiek raidinės-skaitmeninės sekos (*string*) kintamųjų naujas reikšmes.
- Galima sukurti naujus kintamuosius arba pakeisti esamų kintamųjų reikšmes. Nustatomi naujų kintamųjų tipai ir žymenos.
- Remiantis loginėmis sąlygomis galima apskaičiuoti naujas pasirinktų duomenų grupių reikšmes.
- Naujų kintamųjų reikšmėms skaičiuoti galima naudoti daugiau kaip 70 standartinių funkcijų: aritmetinių, statistinių, statistinio pasiskirstymo, atsitiktinių kintamųjų, sekų ir kt.

Norėdami sukurti naują arba transformuoti esamą kintamąjį, atlikite šiuos veiksmus:

- Nurodykite komandas ***Transform → Compute...***
- Atsidariusio dialogo langelio ***Compute Variable*** (3.1pav) laukelyje ***Target Variable*** (tikslinis kintamasis) nurodykite kintamojo, kurio reikšmės skaičiuojamos, vardą. Tai gali būti jau esantis kintamasis arba naujai įtraukiamas į duomenų rinkmeną.
- Laukelyje ***Numeric Expression*** (arba ***String Expression*** priklausomai nuo kintamojo tipo) įrašykite išraišką, pagal kurią bus skaičiuojamas tikslinis kintamasis.
 - Išraiškoje gali būti naudojami esamų kintamųjų vardai, konstantos, aritmetiniai operatoriai ir funkcijos.
 - Išraišką galima įrašyti arba klaviatūra, arba pele, naudojant kintamųjų sąrašą ir dialogo langelio klaviatūrą. Mygtukas su trikampiu leidžia įkelti į formulės laukelį kintamųjų vardus, o dialogo langelio klaviatūra — įterpti skaičius ir ženklus.
 - Pasirinkta funkcija įterpiama spragtelint pele mygtuką su trikampiu. Klaustuku nurodytose vietose įrašomi funkcijos argumentai — tai gali būti kintamieji, konstantos arba kitos funkcijos. Trumpas funkcijos aprašymas pateikiamas spragtelėjus pasirinktą funkciją dešiniuoju klavišu.
 - Sekų (*string*) konstantos rašomos kabutėse arba su apostrofais.

- Skaitmeninės konstantos rašomos amerikietiškuoju formatu, t. y. dešimtainė dalis atskiriama tašku.
- Spragtelėkite pele dialogo langelio mygtuką **Type& Label...** (Tipas ir žymena) ir naujame dialogo langelyje **Compute Variable: Type and Label** įrašykite naujo kintamojo žymeną. Pagal nustatymą naujas kintamasis yra skaitmeninis. Jeigu naujas kintamasis yra raidinės-skaitmeninės sekos tipo, būtina tai nurodyti. Šio tipo kintamieji negali būti naudojami skaičiavimams. Spragtelėkite mygtuką **Continue**.
- Dialogo langelyje **Compute Variable** spragtelėkite mygtuką **OK**.



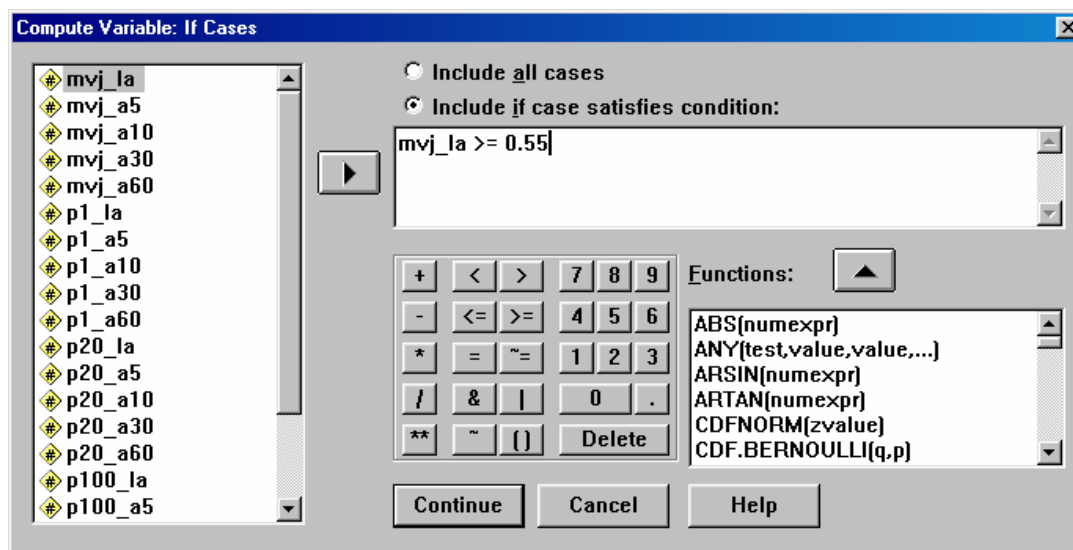
3.1 pav. Kintamojo transformavimo (naujo kintamojo skaičiavimo) dialogo langelis **Compute Variable**

3.1.1. Loginis naujo kintamojo skaičiavimas

SPSS leidžia taikyti ir atskirų stebėjimų reikšmių (*Cases*) transformavimą, naudojant sąlygos formules, pagal kurias kiekvienai kintamojo reikšmei suteikiama loginė reikšmė *true* (sąlyga tenkinama), *false* (sąlyga netenkinama) arba *missing* (praleisti duomenys).

- Jeigu sąlygos formulės sprendinys konkrečiai kintamojo reikšmei yra loginis *true*, ši reikšmė yra transformuojama.
- Jeigu sąlygos formulės sprendinys konkrečiai kintamojo reikšmei yra loginis *false* arba *missing*, reikšmė yra netransformuojama.
- Sąlygos formulei sudaryti gali būti naudojami kintamųjų vardai, konstantos, aritmetiniai operatoriai, skaitmeninės ir kitos funkcijos, santykio: < (mažiau), > (daugiau), <= (mažiau arba lygu), >= (daugiau arba lygu), ~= (nelygu) ir loginiai: & (loginis IR), | (loginis ARBA), ~ (loginis NE) operatoriai iš dialogo langelio klaviatūros.

Sąlygai suformuluoti reikia spragtelėti dialogo langelio *Compute Variable* mygtuką *If...*

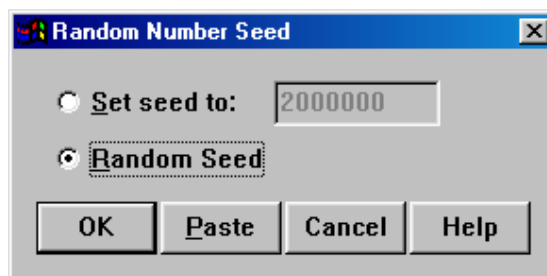


3.2 pav. Sąlygos nustatymo dialogo langelis *Compute Variable: If Cases*

3.1.2. Atsitiktinių skaičių generatoriaus pradinės reikšmės nustatymas

Pseudoatsitiktinių skaičių generatoriaus generuojamos sekos naudojamos duomenims transformuoti (pvz., funkcijų UNIFORM, NORMAL atvejais). Nustatytasis atsitiktinių skaičių sekos pradžios variantas yra *Random Seed*, t. y. kiekvieną kartą naudojama nauja atsitiktinė sekos pradinė reikšmė. Kad atsitiktinių skaičių seka kartotųsi, būtina duoti pradinę atsitiktinių skaičių generatoriaus reikšmę:

- Nurodykite komandas *Transform → Random Number Seed...*
- Dialogo langelio *Random Number Seed* laukelyje *Set seed to* įrašykite bet kokią sveiką skaičių tarp 1 ir 2000000000.



3.3 pav. Pseudoatsitiktinių skaičių generatoriaus pradinės reikšmės nustatymo dialogo langelis *Random Number Seed*

3.2. PASIKARTOJANČIŲ REIKŠMIŲ SKAIČIAUS NUSTATYMAS

Sukuriamas naujas kintamasis, kuris parodo nurodytų kintamųjų tam tikros reikšmės (reikšmių) pasikartojimo skaičių atskiruose stebėjimuose. Pavyzdžiui, anketoje tarp kitų yra ir klausimas:

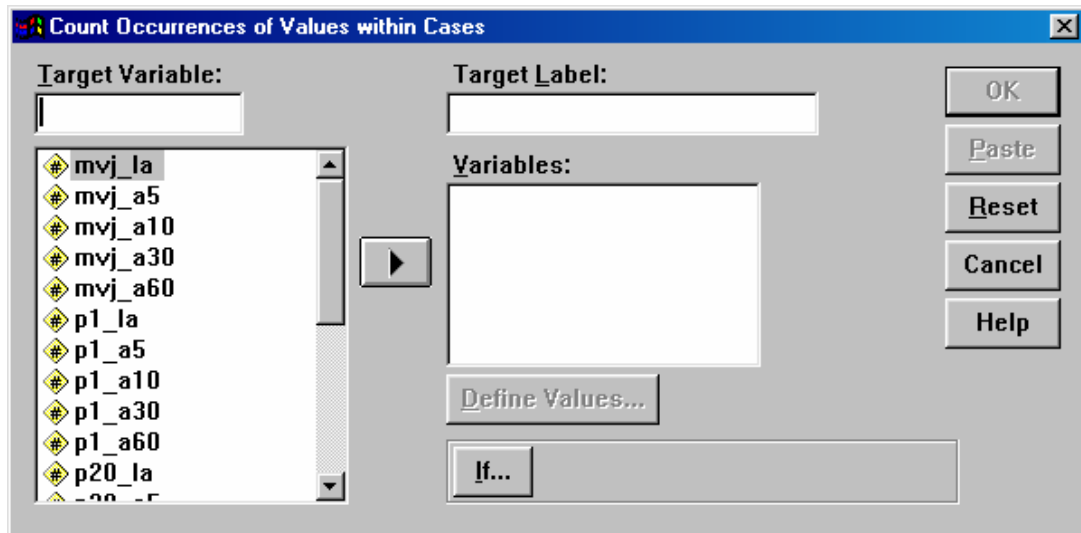
- Kokia sporto šaka iš išvardytų Jūs aktyviai domitės:
 - 1) Plaukimu: taip/ne?
 - 2) Gimnastika: taip/ne?
 - 3) Lengvąja atletika: taip/ne?
 - 4) Tinkliniu: taip/ne?
 - 5) Tenisu: taip/ne?
 - 6) Dviračių sportu: taip/ne?
 - 7) Futbolu: taip/ne?
 - 8) Rankiniu: taip/ne?
 - 9) Krepšiniu: taip/ne?

Suskaičiavus kiekvieno respondento (kiekvieno atskiros eilutės įrašo) reikšmės „taip“ kartojimosi skaičių kintamiesiems 1)–9) gaunamas sporto šakų, kuriomis aktyviai domisi respondentai, skaičius.

Atskiros stebėjimo (*cases*) nustatytos reikšmės kartojimosi skaičiui nustatyti:

- Nurodykite meniu komandas **Transform → Count...**

Atsidarys dialogo langelis **Count Occurences of Values within Cases** (Suskaičiuoti stebėjimo reikšmes).



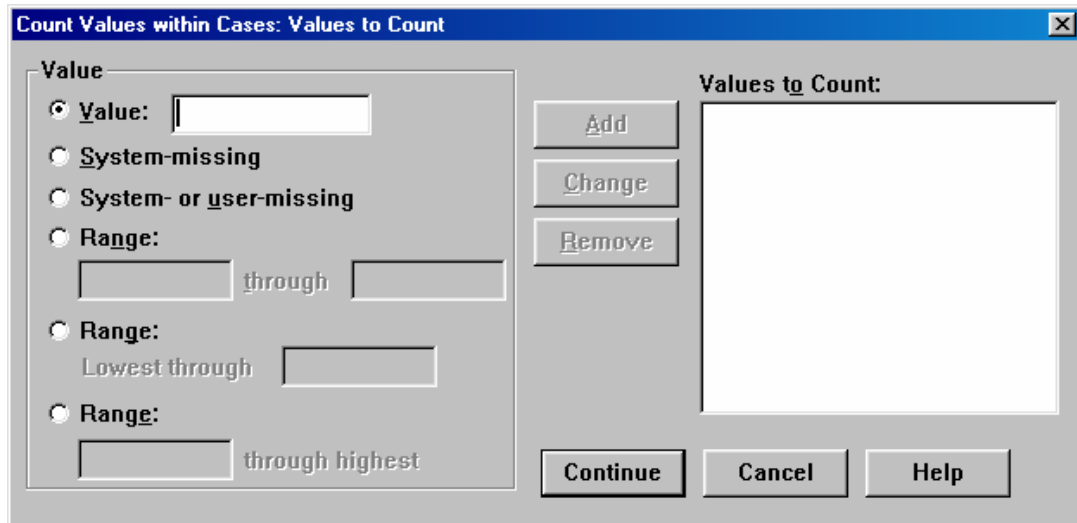
3.4 pav. Dialogo langelis **Count Occurences of Values within Cases**

Dialogo langelis suskirstytas į šias sritis:

- **Target Variable** (Tikslinis kintamasis). Šiame laukelyje nurodomas kintamojo, turinčio suskaičiuotas reikšmes, vardas.

- **Target Label** (Žymena). Laukelyje nurodoma tikslinio kintamojo žymena.
- **Variables** (Kintamieji). Nurodomi tie kintamieji iš bendro kintamųjų, esančių duomenų rinkmenoje, sąrašo, kuriems skaičiuojamas tam tikrų reikšmių buvimas. Sąrašas vienu metu negali turėti skaitmeninių ir raidinių-skaitmeninių sekų tipo kintamųjų.

Spragtelėjus pele dialogo langelio mygtuką **Define Values...** atsidarys naujas dialogo langelis **Count Values within Cases: Values to Count** (Suskaiciuoti stebėjimo reikšmes: skaičiuojamos reikšmės).



3.5 pav. Dialogo langelis **Count Occurences of Values within Cases: Values to Count**

Dialogo langelyje galima nurodyti konkrečią reikšmę, reikšmių diapazoną arba vieno ir kito derinį. Srityje **Value** (Reikšmė) galima pasirinkti vieną iš šių variantų:

- **Value**: Įrašoma konkreti reikšmė, kurios kartojimosi dažnį reikia nustatyti.
- **System missing** (Sistemiškai praleistos): suskaiciuojamas sistemiškai praleistų reikšmių skaičius. Laukelyje **Values to Count** (Skaičiuojamos reikšmės) šis skaičius parodomas kaip SYSMIS. Variantas netaikomas raidinių- skaitmeninių sekų tipo kintamiesiems.
- **System- or user- missing** (Sistemiškai arba vartotojo praleistos): suskaiciuojamas bendras, tiek sisteminų, tiek vartotojo praleistų reikšmių skaičius. Laukelyje **Values to Count** šis skaičius parodomas kaip MISSING.
- **Range through** (Diapazonas): skaičiuojamas patenkančių į apibrėžtą diapazoną reikšmių skaičius. Variantas netaikomas raidinių-skaitmeninių sekų tipo kintamiesiems.

- **Range: Lowest through** (Diapazonas: nuo mažiausios iki): skaičiuojamas patenkančių į diapazoną nuo mažiausios iki nurodytos reikšmių skaičius. Variantas netaikomas raidinių-skaitmeninių sekų tipo kintamiesiems.
- **Range: through highest** (Diapazonas: iki didžiausios): skaičiuojamas patenkančių į diapazoną nuo nurodytos iki didžiausios reikšmių skaičius. Variantas netaikomas raidinių-skaitmeninių sekų tipo kintamiesiems.

Jeigu skaičiuojamas kelių reikšmių kartojimasis, pasirinkus reikiamą variantą, reikia spragtelėti dialogo langelio mygtuką **Add** (Pridėti). Bus suskaičiuotas visų esančių laukelyje **Values to count** reikšmių kartojimasis.

- Patvirtinkite nustatymus spragtelėję mygtuką **Continue**, tada — **OK**.

Jeigu skaičiuojamas tik tam tikras sąlygas tenkinančių reikšmių kartojimasis, sąlygos nurodomos spragtelėjus dialogo langelio mygtuką **If...**, pagal p. 3.1.1 pateiktus reikalavimus.

3.3. REIKŠMIŲ PERKODAVIMAS

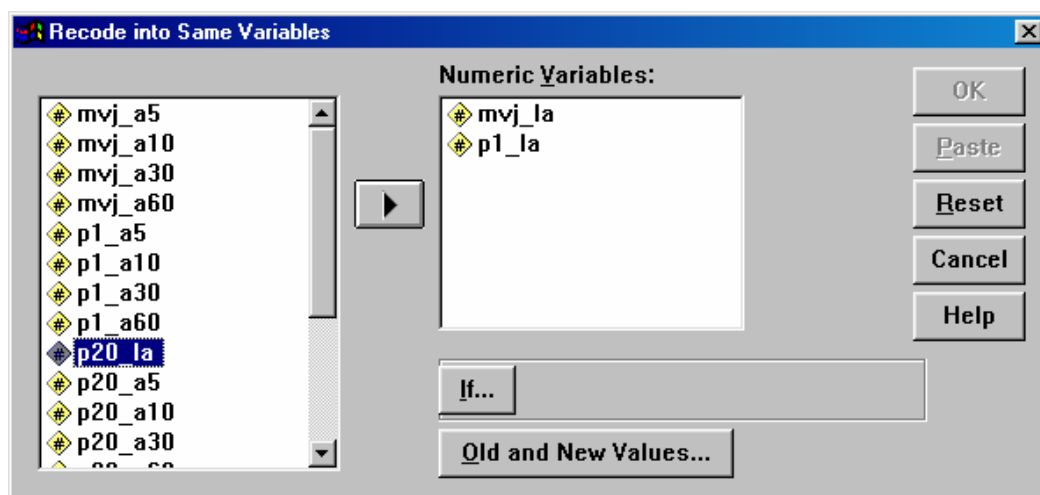
Pirminius duomenis galima perkoduoti. Perkodavimas atliekamas, pvz., pirminių duomenų įvairovei sumažinti, sujungiant šiuos duomenis. Galima perkoduoti tiek esamų kintamųjų reikšmes, tiek sukurti naujus kintamuosius perkoduojamų reikšmių pagrindu.

3.3.1. Reikšmių perkodavimas nekeičiant kintamojo

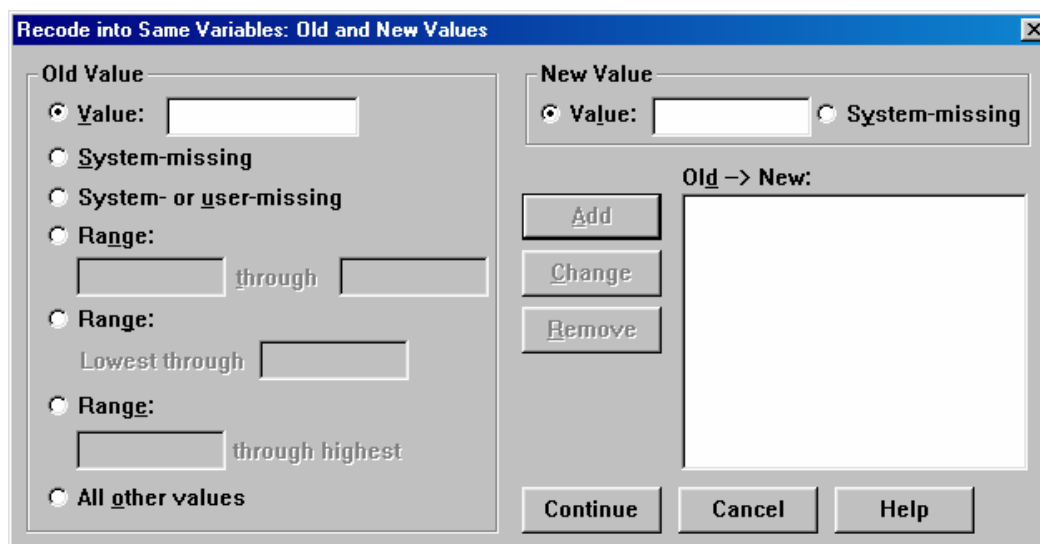
Kintamųjų reikšmių perkodavimas leidžia optimaliau pateikti šias reikšmes arba sujungti jas. Pavyzdžiui, vietoj konkrečių atlyginimų dydžio galima nurodyti atlyginimų diapazonų kategorijas. Galima perkoduoti tiek skaitmeninius, tiek raidinių-skaitmeninių sekų (*string*) kintamuosius. Negalima tik perkoduoti šių skirtingų tipų kintamuosius vienu metu. Kintamųjų reikšmėms perkoduoti:

- Nurodykite komandas **Transform → Recode → Into Same Variables...**
- Pasirodžiusiame dialogo langelyje **Recode into Same Variables** (3.6 pav.) pasirinkite reikiamus kintamuosius (to paties tipo) ir perkelkite į langelį **Numeric Variables** (skaitmeninių kintamųjų atveju). Spragtelėję mygtuką **If...** galite atrinkti perkodavimui tik tam tikras sąlygas tenkinančias reikšmes.
- Spragtelėkite dialogo langelio mygtuką **Old and New Values** (senos ir naujos reikšmės) perkoduojamoms reikšmėms nustatyti. Naujame dialogo langelyje **Recode into Same Variables: Old and New Values** (3.7 pav.), srityje **Old Value** nurodykite kiekvieną perkoduojamą atskirą reikšmę, reikšmių diapazoną arba praleistas reikšmes (sistemiškai praleistos reikšmės ir diapazonai negali būti nurodomi raidinių-skaitmeninių sekų kintamiesiems).

- Dialogo langelio srityje **New Value** (nauja reikšmė) pasirinkite **Value** ir langelyje įrašykite reikšmę, į kurią perkoduojama sena reikšmė ar reikšmių diapazonas, arba pažymėkite **System missing**. Kiekvieną kartą, užpildę **Old Value** ir **New Value** laukelius, spragtelėkite mygtuką **Add**.
- Laukelyje **Old → New** pateikiamas detalus kintamojo (kintamųjų) reikšmių perkodavimo sąrašas. Šį sąrašą galima papildyti, pakeisti ar ištrinti. Sąrašas yra automatiškai rikiuojamas pagal šį eiliškumą: atskiros reikšmės, praleistos reikšmės, diapazonai ir kitos reikšmės.
- Spragtelėkite mygtuką **Continue**, tada — **OK**.



3.6 pav. Reikšmių perkodavimo nekeičiant kintamojo dialogo langelis *Recode into Same Variables*

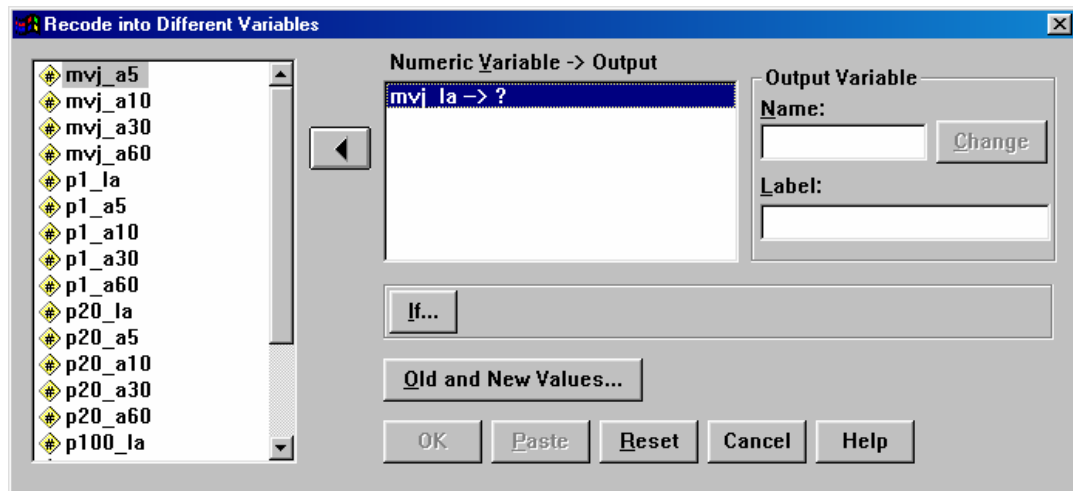


3.7 pav. Dialogo langelis *Recode into Same Variables: Old and New Values*

3.3.2. Reikšmių perkodavimas įvedant naujus kintamuosius

Perkodavimo įvedant naujus kintamuosius metu atskiros esamų kintamųjų reikšmės arba šių reikšmių diapazonai pakeičiami naujų kintamųjų naujomis reikšmėmis. Galima perkoduoti tiek skaitmeninius, tiek raidinių-skaitmeninių sekų kintamuosius. Galima perkoduoti skaitmeninius kintamuosius į raidinių-skaitmeninių sekų kintamuosius, ir atvirkščiai. Vienu metu galima perkoduoti tik vienodo tipo kintamuosius. Kintamųjų reikšmėms perkoduoti:

- Nurodykite komandas **Transform → Recode → Into Different Variables...**
- Pasirodžiusiame dialogo langelyje **Recode into Different Variables** (3.8 pav.) pasirinkite reikiamus kintamuosius (to paties tipo) ir perkeltkite į langelį **Input Variable → Output Variable (Numeric Variable → Output** — pasirinkus skaitmeninį kintamąjį). Pasirodęs klaustuko simbolis primena, kad reikia suteikti vardą naujam kintamajam.

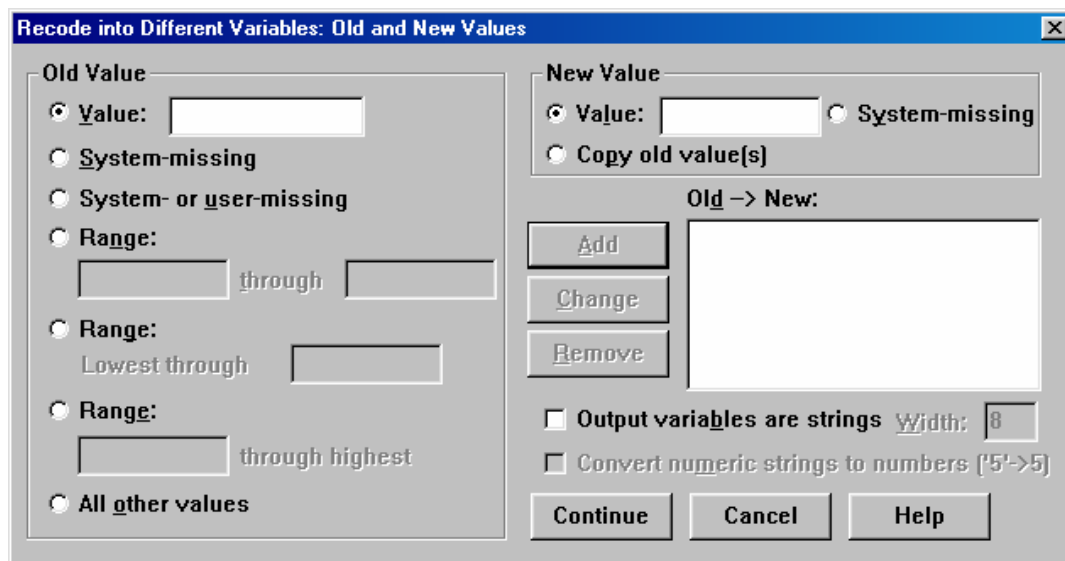


3.8 pav. Reikšmių perkodavimo įvedant naujus kintamuosius dialogo langelis **Recode into Different Variables**

- Įrašykite **Output Name** laukelyje **Name** naujo kintamojo vardą. Spragtelėkite dialogo langelio mygtuką **Change**. Įrašytas vardas pakeis klaustuką laukelyje **Input Variable → Output Variable**. Laukelyje **Label** įrašykite kintamojo žymeną ir spragtelėkite mygtuką **Change**. Tai pakartokite visiems naujiems kintamiesiems. Spragtelėję mygtuką **If...** galite perkodavimui atrinkti tik tam tikras sąlygas tenkinančias reikšmes.
- Spragtelėkite dialogo langelio mygtuką **Old and New Values** (senos ir naujos reikšmės) perkoduojamoms reikšmėms nustatyti. Naujame dialogo langelyje **Recode into Different Variables: Old and New Values** (3.9 pav.), srityje **Old Value** nurodykite kiekvieną perkoduojamą atskirą reikšmę, reikšmių diapazoną arba praleistas reikšmes (sistemiškai praleistos reikšmės ir diapazonai negali būti nurodomi raidinių-

skaitmeninių sekų kintamiesiems). **Old Value** srityje nurodomi duomenys turi atitikti keičiamo kintamojo duomenų tipą.

- Dialogo langelio srityje **New Value** (nauja reikšmė) galite pasirinkti šiuos variantus:
 - **Value** laukelyje įrašyti naują reikšmę.
 - **System missing** — naujo kintamojo reikšmė nurodoma kaip sistemiškai praleista. Reikšmių sąrašo laukelyje **Old→New** ši reikšmė parodoma kaip SYSMIS. Variantas netaikomas raidinių-skaitmeninių sekų tipo kintamiesiems.
 - **Copy old value(s)** — naujo kintamojo reikšmės išsaugomos nepakeistos. Jeigu nauji tiksliniai kintamieji yra raidinių-skaitmeninių sekų tipo, pažymėkite langelį **Output variables are strings**.



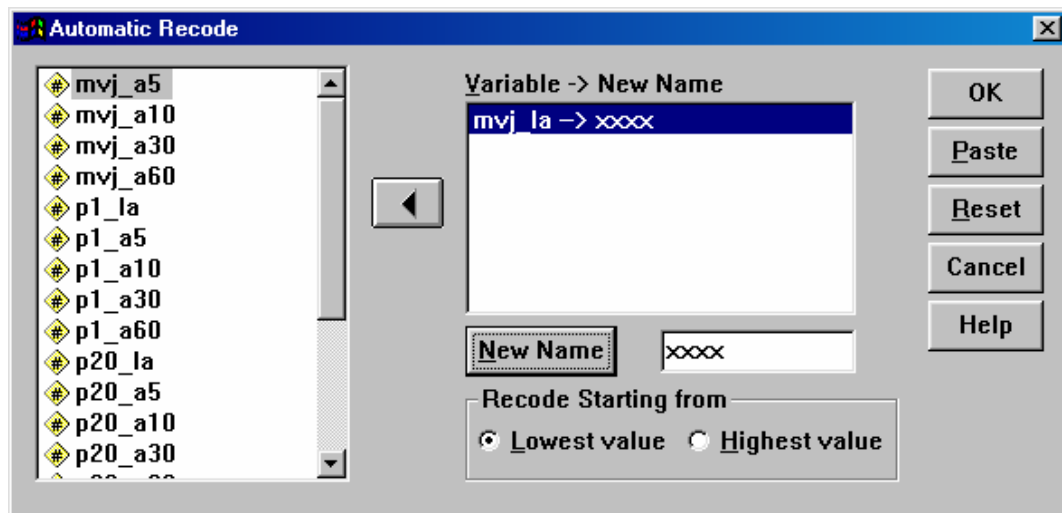
3.9 pav. Dialogo langelis **Recode into Different Variables: Old and New Values**

- Kiekvieną kartą, užpildę **Old Value** ir **New Value** laukelius, spragtelėkite mygtuką **Add**.
- Laukelyje **Old→New** pateikiamas detalus kintamojo (kintamųjų) reikšmių perkodavimo sąrašas. Šį sąrašą galima papildyti, pakeisti ar ištrinti. Sąrašas yra automatiškai rikiuojamas pagal šį eiliškumą: atskiros reikšmės, praleistos reikšmės, diapazonai ir kitos reikšmės.
- Spragtelėkite mygtuką **Continue**, tada — **OK**.

3.3.3. Automatinis perkodavimas

Jeigu reikšmių kategorijos nėra užkoduotos nuoseklia sveikų skaičių seka, pradedant 1, gali būti nevykdomos kai kurios SPSS operacijos. Todėl skaitmeninių ir raidinių-skaitmeninių sekų tipo kintamųjų reikšmėms paversti nuoseklia sveikų skaičių seka SPSS yra numatyta automatinio perkodavimo galimybė. Pavyzdžiu galėtų būti tokių raidinių-skaitmeninių sekų tipo kintamųjų, kaip antai įvairių sporto šakų arba ligų pavadinimų, perkodavimas į nuoseklios sveikų skaičių sekos skaitmeninio tipo kintamąjį. Automatiškai perkoduojant gautas naujas kintamasis išsaugo nustatytas senojo kintamojo ir jo reikšmių žymenas. Jeigu senojo kintamojo reikšmės neturėjo žymenų, tai jos pačios naudojamos kaip naujų perkoduotų reikšmių žymenos. Skaitmeninių ir raidinių-skaitmeninių sekų tipo kintamųjų reikšmėms perkoduoti į nuoseklia sveikų skaičių seką vykdykite šiuos veiksmus:

- Nurodykite komandas **Transform → Automatic Recode...**
- Pasirodžiusiame dialogo langelyje **Automatic Recode** (3.10 pav.) pasirinkite reikiamus kintamuosius ir perkeltkite į langelį **Variable → New Name**. Kiekvienam pasirinktam kintamajam teksto laukelyje (šalia mygtuko **New Name**) įrašykite naujo kintamojo vardą ir spragtelėkite mygtuką **New Name**.
- Spragtelėkite mygtuką **OK**.

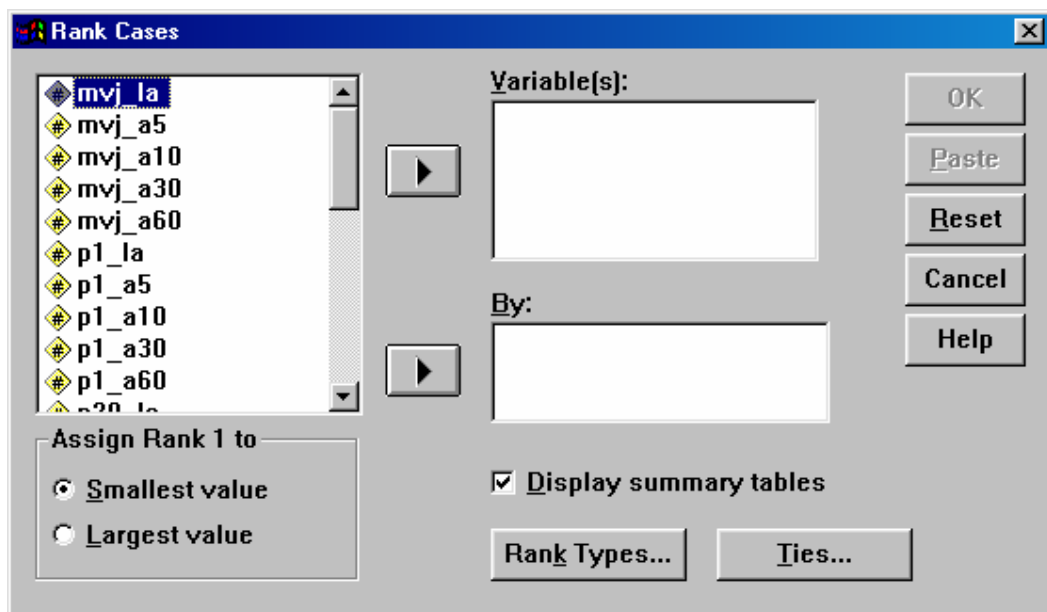


3.10 pav. Automatinio perkodavimo dialogo langelis **Automatic Recode**

3.4. DUOMENŲ RANGAVIMAS

Ranguojant gaunami nauji kintamieji, kurių reikšmės yra rangai, įverčiai ir procentinės skaitmeninių kintamųjų išraiškos. Rangavimas pagal pasirinkto kintamojo atskirų stebėjimų reikšmes atliekamas šia tvarka:

- Nurodykite komandas **Transform → Rank Cases...**
- Atsidariusiame dialogo langelyje **Rank Cases** (3.11 pav.) pasirinkite reikiamus kintamuosius ir perkeltite į langelį **Variable(s)**. Ranguoti galima tik skaitmeninius kintamuosius. Į laukelį **By** galima įrašyti kategorinį kintamąjį — tuo atveju rangai bus priskiriami atskirai pagal šio kintamojo grupes (kategorijas).

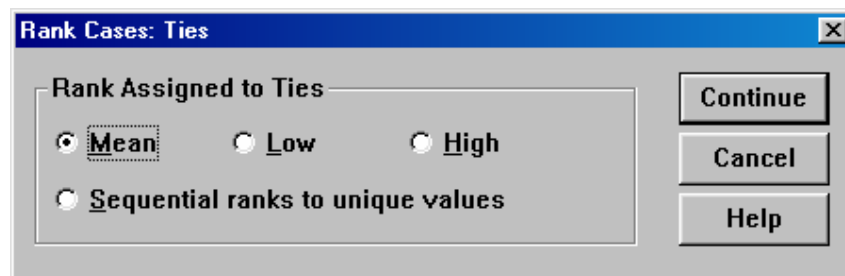


3.11 pav. Rangavimo pagal stebėjimų reikšmes dialogo langelis **Rank Cases**

- Pasirinkite didėjančią arba mažėjančią rangavimo tvarką laukelyje **Assign Rank 1** (suteikti rangą 1) pažymėdami **Smallest value** (mažiausiai reikšmei) arba **Largest value** (didžiausiai reikšmei).
- Spragtelėję mygtuką **Rank Types...** (rangų tipai) atversite dialogo langelį rangų tipui pasirinkti. Standartinis nustatytas variantas yra **Rank**. Trumpai paminėsime kai kuriuos kitus rangų tipus (smulkiau: Бююль, Цефель, 2002):
 - **Fractional Rank** (Santykinis rangas). Tai rango reikšmė, padalyta iš stebėjimų skaičiaus.
 - **Fractional Rank as %** (Santykinis rangas procentais). Tai santykinio rango reikšmė, padauginta iš 100.

- *Niles* (N- procentilės). Vartotojas gali duoti procentilių grupių skaičių, į kurias suskaidomi stebėjimai.
 - *Proportion estimates* (Dalies įverčiai). Suskaičiuojama sukauptoji dalis, esant prielaidai, kad kintamasis pasiskirstęs pagal normalųjį dėsnį.
- Spragtelėję mygtuką **Ties...**(ryšiai) atversite dialogo langelį **Rank Cases: Ties** (3.12 pav), kuriame galite nurodyti, kaip turi būti skaičiuojami rangai, esant vienodiems dydžiams. Standartinis nustatytasis variantas yra **Mean**, pagal kurį vienodiems dydžiams suteikiamo rango reikšmė yra šių dydžių rangų vidurkis. Pasirinkus **Low**, visiems dydžiams nustatomas mažiausias iš rangų, o pasirinkus **High** – didžiausias iš rangų. Pasirinkus **Sequential ranks to unique values** visoms vienodomis stebėjimų reikšmėms suteikiamas vienodas rangas, o kitai pagal dydį reikšmei suteikiamas tolesnis sveikas skaičius. Todėl šiuo atveju didžiausias rangas lygus ne bendram reikšmių skaičiui, o skirtingų reikšmių skaičiui. Išvardyti keturi rangų suteikimo būdai gali būti paaiškinti šiuo paprastu pavyzdžiu:

Reikšmė	Rangai			
	<i>Mean</i>	<i>Low</i>	<i>High</i>	<i>Sequential</i>
10	1	1	1	1
15	3	2	4	2
15	3	2	4	2
15	3	2	4	2
16	5	5	5	3
20	6	6	6	4



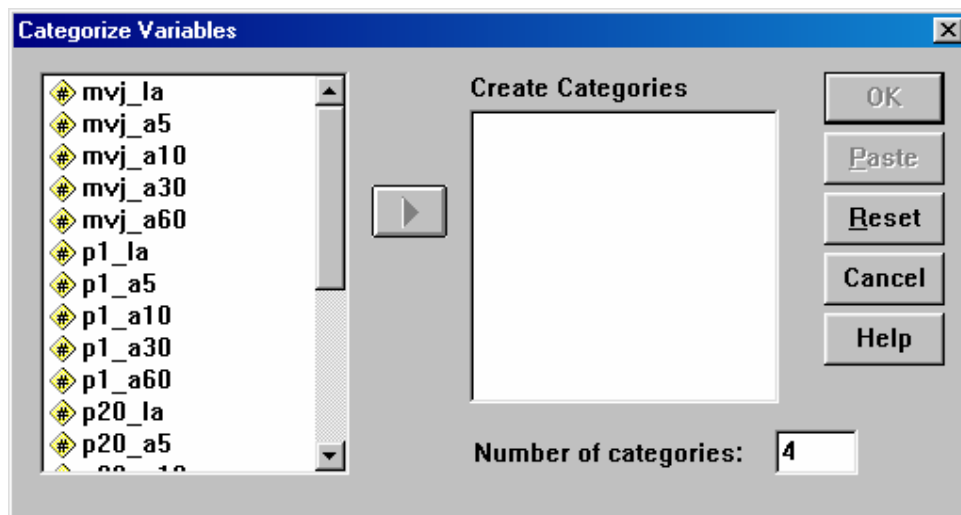
3.12 pav. Rangų skaičiavimo būdo pasirinkimo dialogo langelis **Rank Cases: Ties**

- Rekomenduojamas yra nustatytasis **Mean** variantas. Uždarykite dialogo langelį spragtelėję mygtuką **Continue**. Spragtelėkite mygtuką **OK** dialogo langelyje **Rank Cases**. Duomenų rinkmena bus papildyta naujais kintamaisiais, kurių reikšmės — pasirinktų kintamųjų reikšmėms suteikti rangai.

3.5. KATEGORINIS TRANSFORMAVIMAS

Turimus skaitmeninius duomenis galima transformuoti į diskrečią kategorijų seką. Transformuojant sukuriama nauji kintamieji, kurių duomenys — kategorijos. Duomenys suskirstomi į kategorijas procentiniu santykiu, kiekvienai kategorijai priskiriant apytikriai vienodą stebėjimų skaičių. Duomenų suskirstymas kategorijomis atliekamas šia tvarka:

- Nurodykite komandas **Transform → Categorize Variables...**
- Atsidariusiame dialogo langelyje **Categorize Variables** (3.13 pav.) pasirinkite reikiamus kintamuosius ir perkelkite į langelį **Create Categories for**. Skirstyti į kategorijas galima tik skaitmeninius kintamuosius. Laukelyje **Number of Categories** įrašykite norimą kategorijų skaičių.
- Spragtelėkite mygtuką **OK**.



3.13 pav. Kintamųjų suskirstymo kategorijomis dialogo langelis *Categorize Variables*

3.6. LAIKO EILUČIŲ DUOMENŲ TRANSFORMAVIMAS

Laiko eilutėmis yra vadinami duomenys, gauti matuojant kintamųjų reikšmes reguliariais laiko intervalais. Šios kintamųjų reikšmės sudaro laiko eilučių duomenų rinkmenos atskirus stebėjimus (*cases*). Iš kelių duomenų transformavimų, taikomų laiko eilutėms, trumpai paliesime (plačiau — Sakalauskas, 2003) vieną iš svarbiausių — skaitmeninį filtravimą, leidžiantį labiau išryškinti sistemine komponentę atsitiktinio triukšmo (t. y. atsitiktinio duomenų išsibarstymo) fone. Tam yra naudojami šie metodai: skirtumų (*differences*), slenkamųjų vidurkių (*moving averages*), slenkamųjų medianų (*running medians*), atsilikimo (*lag*) ir paskubos (*lead*).

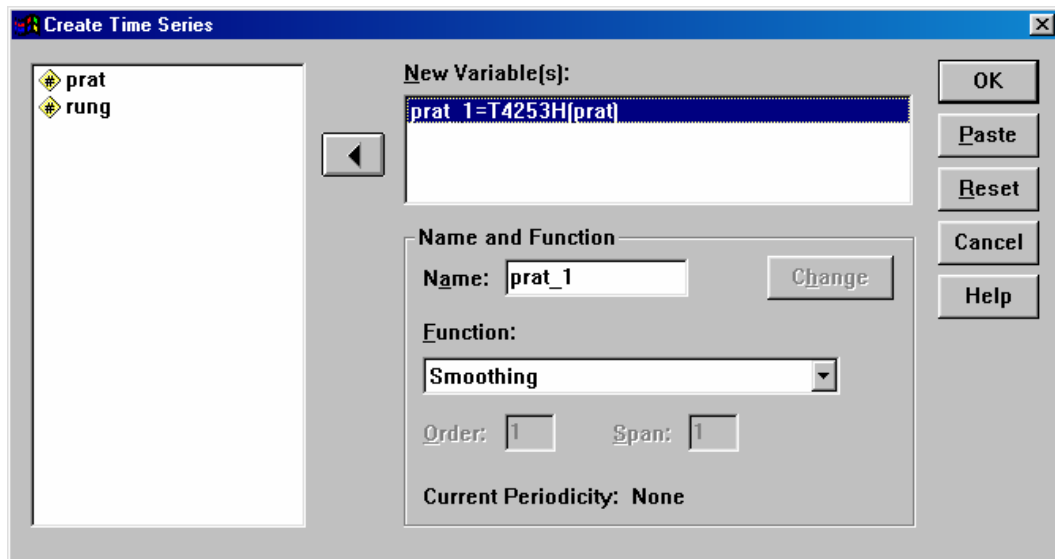
- **Skirtumų metodas.** Pateikia viena paskui kitą einančių laiko eilutės reikšmių skirtumus. Metodo laipsnis nusako prieš tai einančių reikšmių, kurios naudojamos skirtumui skaičiuoti, skaičių. Kadangi prarandama tiek stebėjimų, koks yra laipsnis, laiko eilutės pradžioje pasirodo sistemiškai praleistos reikšmės. Pavyzdžiui, esant skirtumo laipsniui 2, pirmosios dvi naujojo kintamojo reikšmės bus sistemiškai praleistos.
- **Įvertinančių sezoniškumą skirtumų metodas.** Pateikia tam tikru pastoviu laiko intervalu einančių viena paskui kitą eilutės reikšmių skirtumus. Naudojamas, kai pasireiškia sezono komponentė, atspindinti reguliarius nukrypimus nuo bendros tendencijos. Laiko eilutės pradžioje sistemiškai praleistų reikšmių skaičius lygus laipsnio ir laiko intervalų skaičiaus sandaugai. Pavyzdžiui, esant skirtumo laipsniui 2 ir laiko intervalų (mėnesių) skaičiui 12, pirmosios 24 naujojo kintamojo reikšmės bus sistemiškai praleistos.
- **Centruotų slenkamųjų vidurkių metodas.** Šio metodo esmė — kiekvienas laiko eilutės narys yra pakeičiamas tam tikro jį supančių narių skaičiaus ir paties nario vidurkiu. Jeigu vidurkio skaičiavimui naudojamas laiko eilutės narių skaičius yra n , sistemiškai praleistų reikšmių skaičius laiko eilutės pradžioje ir pabaigoje bus lygus $n/2$.
- **Atsiliekančių reikšmių slenkamųjų vidurkių metodas.** Šio metodo esmė — kiekvienas laiko eilutės narys yra pakeičiamas tam tikro skaičiaus prieš jį einančių narių vidurkiu. Sistemiškai praleistų reikšmių skaičius laiko eilutės pradžioje lygus vidurkio skaičiavimui naudojamų laiko eilutės narių skaičiui.
- **Slenkamųjų medianų metodas.** Šio metodo esmė — kiekvienas laiko eilutės narys yra pakeičiamas tam tikro jį supančių narių skaičiaus ir paties nario mediana. Jeigu medianai skaičiuoti naudojamas laiko eilutės narių skaičius yra n , sistemiškai praleistų reikšmių skaičius laiko eilutės pradžioje ir pabaigoje bus lygus $n/2$.
- **Sukauptoji suma.** Sukauptoji laiko eilutės reikšmių, esančių prieš nurodytą narį, ir paties nurodyto nario reikšmės suma.
- **Atsilikimo metodas.** Šio metodo esmė — kiekvienas laiko eilutės narys yra pakeičiamas per tam tikrą narių skaičių prieš jį esančiu nariu. Metodo laipsnis nusako, kurio iš eilės prieš tai esančio nario reikšmė yra paimama. Sistemiškai praleistų reikšmių skaičius laiko eilutės pradžioje lygus laipsnio reikšmei.
- **Paskubos metodas.** Šio metodo esmė — kiekvienas laiko eilutės narys yra pakeičiamas per tam tikrą narių skaičių po jo esančiu nariu. Metodo laipsnis nusako, kurio iš eilės po skaičiuojamo esančio nario reikšmė yra paimama. Sistemiškai praleistų reikšmių skaičius laiko eilutės pabaigoje lygus laipsnio reikšmei.
- **Glodinimas.** Naujos laiko eilutės reikšmės skaičiuojamos pagal sudėtinį glodinimo algoritmą. Pirmiausia pritaikomas slenkamųjų medianų

metodas. Tada, atėmus iš pirminių duomenų suglodintus duomenis, apskaičiuojamos liekanos. Procesas kartojamas atsižvelgiant į gautas liekanas. Metodas dar vadinamas T4253H.

SPSS nustatytas naujo filtruoto kintamojo vardas yra pirmi šeši pirminio kintamojo vardo ženklai, pabraukimo ženklas ir eilės numeris.

Nauji laiko eilučių kintamieji sukuriama šia tvarka:

- Nurodykite komandas **Transform → Create Time Series...**
- Atsidariusiame dialogo langelyje **Create Time Series** (3.14 pav.) pasirinkite kintamąjį (kintamuosius), kurio pagrindu bus sukurtas naujas (filtruotas) kintamasis ir perkeltite į langelį **New Variable(s)**. Pasirenkami gali būti tik skaitmeniniai kintamieji. Laukelyje **Function** pasirinkite funkciją, pagal kurią bus transformuotas pasirinktas kintamasis.
- Laukelyje **Name** galite pakeisti siūlomą naujo kintamojo vardą. Laukelyje **Order** galite nurodyti taikomo metodo laipsnį, o laukelyje **Span** — gretimų laiko eilutės narių, naudojamų transformuojamos eilutės kiekvieno nario reikšmei apskaičiuoti, skaičių.
- Spragtelėkite mygtuką **OK**.



3.14 pav. Naujų (filtruotų) laiko eilučių kintamųjų sukūrimo dialogo langelis **Create Time Series**

4. DUOMENŲ RINKMENŲ TVARKYMAS IR TRANSFORMAVIMAS

SPSS turi plačias galimybes pritaikyti pirminių duomenų rinkmenas keliamiems uždaviniams spręsti. Jūs galite sujungti kelias bylas į vieną, rūšiuoti duomenis skirtinga tvarka, atrinkti dalį stebėjimų, suskirstyti stebėjimus. Svarbiausi duomenų rinkmenų transformavimo būdai yra šie:

- **Duomenų rikiavimas.** Galima rikiuoti stebėjimų duomenis pagal vieno ar kelių kintamųjų reikšmes.
- **Stebėjimų ir kintamųjų sukeitimas vietomis (transponavimas).** SPSS traktuoja elektroninės lentelės eilutes kaip stebėjimus, o stulpelius — kaip kintamuosius. Jeigu duomenų rinkmenoje duomenys įrašyti netinkama tvarka, eilutes ir stulpelius galima sukeisti vietomis.
- **Bylų sujungimas.** Jūs galite sujungti dvi ar daugiau bylų. Gali būti jungiamos bylos su tais pačiais kintamaisiais, bet skirtingais stebėjimais arba su tais pačiais stebėjimais, bet skirtingais kintamaisiais.
- **Dalies stebėjimų atrinkimas.** Galima alikti analizę apsiribojant stebėjimų dalimi arba atlikti vienalaikę analizę su skirtingomis stebėjimų grupėmis.
- **Duomenų agregavimas.** Duomenų agregavimas vieno ar kelių kategorinių kintamųjų pagrindu leidžia praplėsti analizės galimybes.
- **Svorio koeficientų taikymas.** Atskirų stebėjimų duomenims gali būti suteiktas skirtingas svoris.
- **Duomenų pertvarkymas.** Duomenis galima pertvarkyti sukuriant vieną stebėjimą (įrašą) iš daugelio arba sukuriant daug stebėjimų iš vieno.

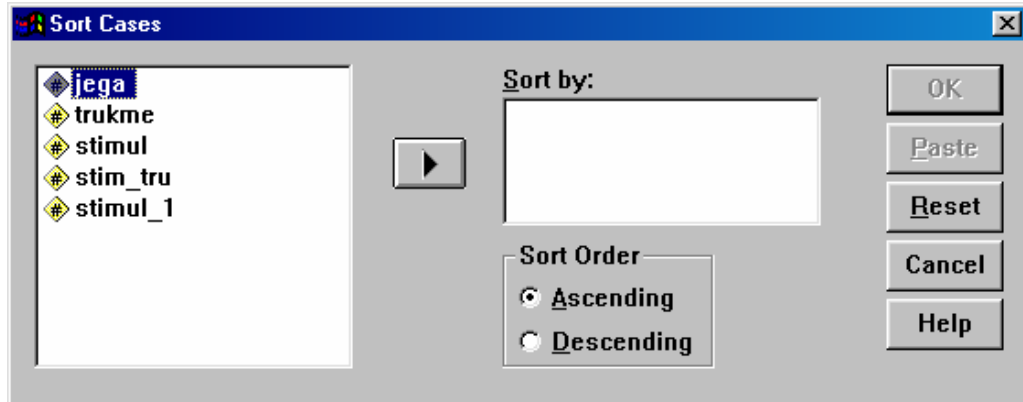
4.1. DUOMENŲ RIKIAVIMAS

Duomenis galima rikiuoti pagal vieno ar kelių kintamųjų reikšmes didėjančia arba mažėjančia tvarka. Pasirinkus rikiavimą pagal kelis kintamuosius, duomenys bus rikiuojami ta tvarka, kuria dialogo langelio **Sort Cases** laukelyje **Sort By** bus įrašyti šie kintamieji. Pavyzdžiui, jeigu pirmuoju rikiavimo kintamuoju bus pasirinktas kintamasis *gender* (lytis), o antruoju — *income* (pajamos), stebėjimai bus surikiuoti pagal lytį ir pagal kiekvienos lyties pajamas. Rikiuojant raidinių- skaitmeninių sekų tipo kintamųjų reikšmes didžiosios raidės turi pirmenybę prieš mažąsias.

Duomenys rikiuojami šia tvarka:

- Nurodykite komandas **Data → Sort Cases...**
- Atsidariusiame dialogo langelyje **Sort Cases** (4.1 pav.) pasirinkite vieną ar kelis kintamuosius, kurių pagrindu bus rikiuojami duomenys ir perkeltkite į langelį **Sort By**.
- Laukelyje **Sort Order** nurodykite rikiavimo tvarką — **Ascending** (didėjimo) arba **Descending** (mažėjimo).

- Spragtelėkite mygtuką **OK**.



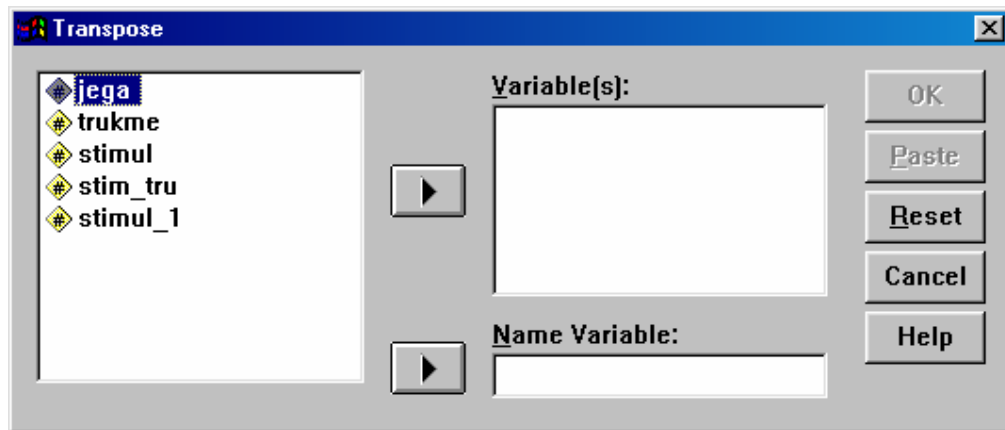
4.1 pav. Duomenų rikiavimo dialogo langelis **Sort Cases**

4.2. KINTAMŲJŲ IR STEBĖJIMŲ SUKEITIMAS VIETOMIS

Vykdamą komandą **Transpose**, sukuriamą naują duomenų rinkmeną, kurioje išeities duomenų eilutės (*cases*) ir stulpeliai sukeičiami vietomis, t. y. eilutėse įrašyti stebėjimai tampa kintamaisiais, o stulpeliuose įrašyti kintamieji — stebėjimais. Sukeičiant duomenis vietomis (transponuojant), automatiškai suteikiami vardai naujiems kintamiesiems (skaitmeniniams kintamiesiems — *var* ir eilės numeris). Taip pat automatiškai sukuriamas naujas skaitmeninės-raidinės sekos tipo kintamasis *case_lbl*, kurio eilutėse įrašyti pradinių duomenų kintamųjų vardai.

Kintamieji ir stebėjimai sukeičiami vietomis šia tvarka:

- Nurodykite komandas **Data → Transpose...**
- Atsidariusiame dialogo langelyje **Transpose** (4.2 pav.) pasirinkite kintamuosius, kuriuos pakeisite į stebėjimus.
- Laukelyje **Name Variable** galite įrašyti vardą kintamojo, kurio reikšmės bus panaudojamos naujų kintamųjų vardams transponuotoje duomenų rinkmenoje. Jei pasirinkote skaitmeninį kintamąjį, naujų kintamųjų vardai prasidės raide *v*, o toliau eis pasirinkto kintamojo skaitmeninė reikšmė.
- Spragtelėkite mygtuką **OK**.



4.2 pav. Duomenų transponavimo dialogo langelis **Transpose**

4.3. BYLŲ SUJUNGIMAS

Dviejų bylų duomenis galima sujungti dviem būdais: sujungti bylas su vienodais kintamaisiais (*variables*), bet skirtingais stebėjimais (*cases*) arba sujungti bylas su tais pačiais stebėjimais, bet skirtingais kintamaisiais.

4.3.1. Kaip sujungti bylas su tais pačiais kintamaisiais?

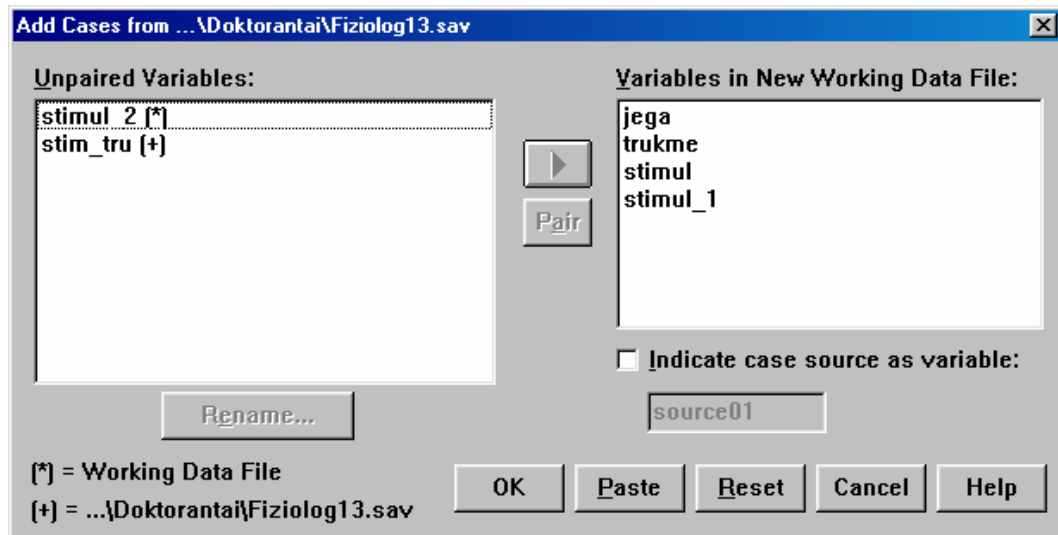
Sujungiant bylas su tais pačiais kintamaisiais, neporiniai kintamieji, t. y. kintamieji, kurie nesikartoja abiejose bylose, į naują duomenų rinkmeną neįtraukiami. Dialogo langelio **Add Cases** (4.3 pav.) laukelyje **Unpaired Variables** (neporiniai kintamieji) tokie kintamieji iš aktyvios bylos pažymimi žvaigždute (*), o kintamieji iš prijungiamos bylos pažymimi pliusu (+). Pagal nustatymą į naują duomenų rinkmeną neįtraukiami:

- Kiekvienos bylos kintamieji, kurių vardai nesikartoja kitoje byloje. Tačiau Jūs galite sukurti poras iš neturinčių poros kintamųjų ir įtraukti jas į naują jungtinę duomenų rinkmeną.
- Kintamieji, kurie vienoje byloje yra apibrėžti kaip skaitmeniniai, o kitoje byloje — kaip raidinės-skaitmeninės sekos (*string*). Skaitmeniniai kintamieji negali būti sujungiami su sekos tipo kintamaisiais.
- Skirtingo formato raidinės-skaitmeninės sekos (*string*) kintamieji. Sujungti gali būti tik vienodą formatą turintys sekos tipo kintamieji.

Laukelyje **Variables in New Working Data File** figūruoja naujos jungtinės duomenų rinkmenos kintamieji. Pagal nustatymą į šį sąrašą įtraukiami kintamieji, kurių vardai ir tipas abiejose bylose sutampa. Jūs galite pašalinti iš šio sąrašo nepageidaujamus kintamuosius.

Bylos su tais pačiais kintamaisiais sujungiamos šia tvarka:

- Atidarykite vieną iš bylų. Šios bylos kintamųjų reikšmės bus pirmosios naujoje jungtinėje byloje.
- Nurodykite komandas **Data → Merge Files → Add Cases...**
- Pasirinkite bylą, kuri bus sujungiama su atidarytąja ir įkelkite duomenis komanda **Open**.
- Esant reikalui, pašalinkite nepageidaujamus kintamuosius iš laukelio **Variables in New Working Data File** ir įkelkite ten bet kurią porą kintamųjų iš sąrašo **Unpaired Variables**, kurie, nepaisant skirtingų vardų sujungiamose bylose, iš esmės yra tapatūs. Dviem neporiniams kintamiesiems įkelti, pirmiausia spragtelėkite laukelyje **Unpaired Variables** vieną pasirinktą kintamąjį, tada, laikydami nuspaustą klavišą **Ctrl** — kitą ir spragtelėkite mygtuką **Pair**. Jungtinėje byloje kintamajam bus suteiktas aktyvios bylos (t. y. atidarytos bylos) kintamojo vardas.



4.3 pav. Dialogo langelis byloms su tais pačiais kintamaisiais sujungti *Add Cases from...*

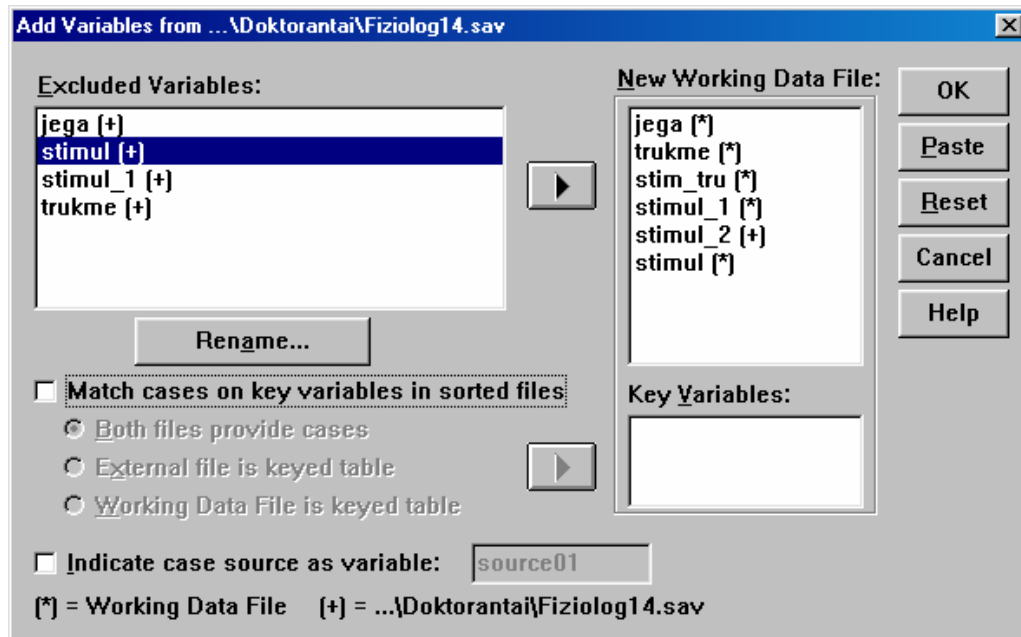
4.3.2. Kaip sujungti bylas su skirtingais kintamaisiais?

Komanda **Add Variables** sujungiamos su skirtingais kintamaisiais bylos turi turėti tuos pačius stebėjimus (*cases*). Pavyzdžiui, vienoje sujungiamų bylų įrašyti vieno testo rezultatai, o kitoje byloje — kito analogiško testo rezultatai. Abiejose bylose stebėjimų duomenys turi būti surikiuoti ta pačia tvarka. Prijungiamos bylos kintamieji, kurių vardai sutampa su atidarytos bylos kintamųjų vardais, į jungtinę bylą neįtraukiami, kadangi traktuojami kaip turintys identišką informaciją.

Bylos su skirtingais kintamaisiais sujungiamos šia tvarka:

- Atidarykite vieną iš bylų. Nurodykite komandas **Data → Merge Files → Add Variables...**

- Pasirinkite bylą, kuri bus sujungiama su atidarytąja ir įkelkite duomenis komanda **Open**.
- Dialogo langelio **Add Variables** (4.4 pav.) laukelyje **Excluded Variables** (neįtraukti kintamieji) pateikiamas sąrašas tų prijungiamos bylos kintamųjų, kurių vardai sutampa su atidarytos bylos kintamųjų vardais.
- Laukelyje **New Working Data File** (nauja darbinė byla) pateikiamas visų kintamųjų sąrašas, į kurį automatiškai įtraukiami visi abiejų bylų kintamieji, turintys nesikartojančius vardus. Kintamieji iš aktyvios bylos pažymimi žvaigždute (*), o kintamieji iš prijungiamos bylos pažymimi pliusu (+).
- Jeigu Jūs norite įtraukti į sujungtą bylą kintamuosius su tapačiais vardais, pirmiausia turite pakeisti jų vardus. Tam tikslui spragtelėkite pasirinkto kintamojo vardą, tada — mygtuką **Rename...**, laukelyje **New Name** įrašykite naują vardą, spragtelėkite mygtuką **Continue** ir įkelkite kintamąjį nauju vardu į laukelį **New Working Data File**.
- Jeigu vienoje iš jungiamų bylų yra mažesnis stebėjimų skaičius (yra praleistų reikšmių), teisingam stebėjimų suderinimui naudokite pagrindinį kintamąjį. Abiejose bylose pagrindinis kintamasis turi būti tuo pačiu vardu, o duomenys turi būti surikiuoti pagal šį kintamąjį didėjančia tvarka. Spragtelėkite laukelyje **Excluded Variables** pasirinkto pagrindinio kintamojo vardą, pažymėkite laukelį **Match cases to the Key Variables list** (suporuoti stebėjimus pagal pagrindinius kintamuosius) ir įkelkite pagrindinį kintamąjį į laukelį **Key Variables**.



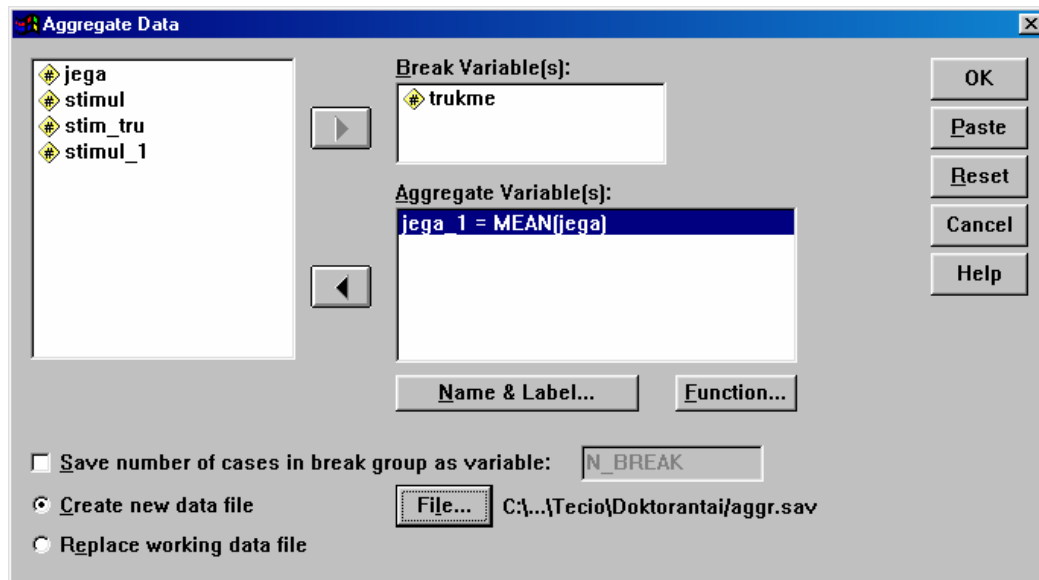
4.4 pav. Dialogo langelis byloms su skirtingais kintamaisiais sujungti **Add Variables from...**

4.4. DUOMENŲ AGREGAVIMAS

Sujungiant vieno ar kelių kategorinių kintamųjų pagrindu atskirų grupių stebėjimų duomenis į suminius stebėjimų duomenis galima sukurti naują jungtinių (agreguotų) duomenų rinkmeną. Pavyzdžiui, atskirų rajonų duomenis Jūs galite sujungti pagal apskritis ir taip sukurti naują duomenų rinkmeną, kurioje analizuojamas vienetas bus apskritis.

Jei norite sujungti duomenis:

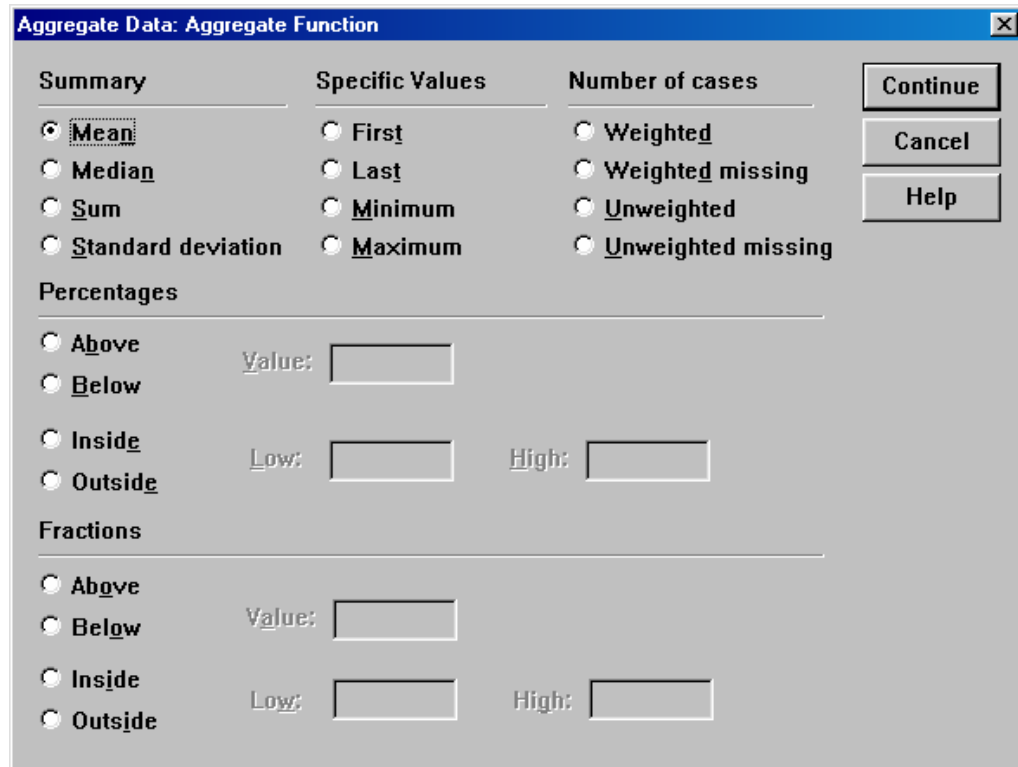
- Nurodykite komandas **Data → Aggregate...**
- Atsidariusiame dialogo langelyje **Aggregate Data** (4.5 pav.) pasirinkite kategorinį kintamąjį (kintamuosius) ir perkeltkite į laukelį **Break Variable(s)** (skaidantysis kintamasis). Šio kintamojo reikšmių pagrindu bus nustatomos agreguojamų duomenų grupės, kurių duomenys bus pakeičiami vienu apskaičiuotu. Kategorinis kintamasis gali būti skaitmeninis arba sekos formato.



4.5 pav. Duomenų agregavimo dialogo langelis **Aggregate Data**

- Pasirinkite kintamuosius, kurių duomenis ketinate sujungti pagal grupes pasirinktos funkcijos pagrindu ir įkelkite į laukelį **Aggregate Variable(s)**. Pagal nustatymą naujam kintamajam yra suteikiamas vardas iš kelių pirminio kintamojo raidžių, pabraukimo ženklo ir eilės numerio. Toliau eina lygybės ženklas, funkcijos, kurios pagrindu vykdomas duomenų agregavimas, pavadinimas ir skliaustuose — pirminio kintamojo vardas. Pirminis kintamasis turi būti skaitmeninis. Spragtelėję mygtuką **Name&Label...** atsidariusiame dialogo langelyje Jūs galite įrašyti kitą kintamojo vardą ir žymeną. Kintamojo vardas neturi viršyti 8 ženklų, o žymena — 120 ženklų.

- Spragtelėję mygtuką **Function...** Jūs galite pasirinkti vietoj nustatytos kitą funkciją, pagal kurią bus agreguojami duomenys. Dialogo langelyje (4.6 pav.) nurodytos šios funkcijos:
 - Suvestinės (*Summary*), įskaitant vidurkį, standartinį nuokrypį, sumą.
 - Procentinės (*Percentages*) arba trupmeninės (*Fractions*) dalies reikšmių, mažesnių arba didesnių už nustatytą.
 - Procentinės arba trupmeninės dalies reikšmių, patenkančių ar nepatenkančių į nustatytą intervalą.



4.6 pav. Dialogo langelis funkcijai duomenų agregavimui parinkti **Aggregate Data: Aggregate Function**.

- Pasirinkę reikiamą funkciją, spragtelėkite mygtuką **Continue**, tada dialogo langelyje **Aggregate Data** — **OK**.

4.5. STEBĖJIMŲ DUOMENŲ ATRANKA

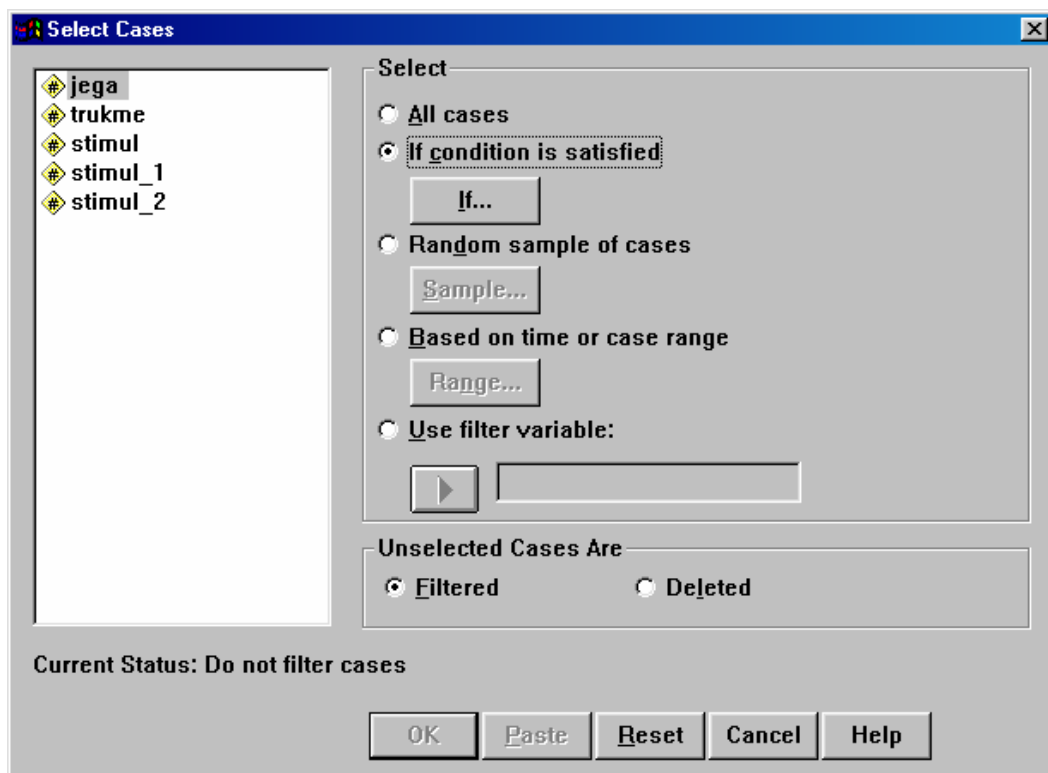
SPSS yra numatyta daug būdų, kaip atrinkti analizei tik tuos stebėjimų duomenis, kurie tenkina nustatytus kriterijus. Visus juos galima suskirstyti į šias tris pagrindines grupes:

- Stebėjimų duomenų atrinkimas pagal tam tikras sąlygas (formules).
- Duomenų atrinkimas atsitiktinės imties būdu.
- Duomenų atrinkimas priklausomai nuo vieno ar kelių kintamųjų reikšmių.

Įrašus (*cases*), kurie netenkina atrankos kriterijų, Jūs galite filtruoti arba pašalinti. Filtruoti įrašai lieka duomenų rinkmenoje, tačiau neanalizuojami. Duomenų statusas nurodomas sukurtu nauju filtravimo kintamuoju *filter_\$*, kurio reikšmė atrinktiems duomenims yra lygi 1, o filtruotiems — 0. Filtruoti įrašai taip pat pažymimi įžambiu brūkšniu **Data Editor** lange. Pašalinti duomenys neatkuriami, tada, kai duomenų rinkmena buvo išsaugota po jų pašalinimo.

Jei norite atrinkti stebėjimų duomenis:

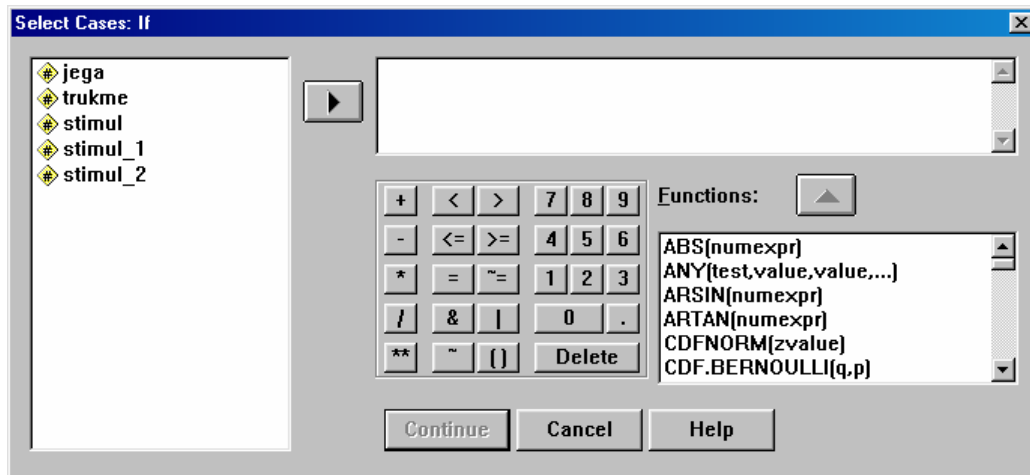
- Nurodykite komandas **Data → Select Cases...**
- Dialogo langelyje **Select Cases** (4.7 pav.) pasirinkite reikiamą duomenų atrankos metodą:
 - **All Cases** (visos reikšmės) — tai variantas pagal nustatymą.
 - **If condition is satisfied** (jeigu sąlyga yra tenkinama).
 - **Random sample of cases** (atsitiktinė duomenų atranka).
 - **Based on time or case range** (pagal diapazoną).
 - **Use filter variable** (pagal filtruojantį kintamąjį).



4.7 pav. Dialogo langelis duomenų atrinkimo kriterijams nustatyti **Select Cases**.

4.5.1. Duomenų atranka pagal loginę sąlygą

- Dialogo langelyje *Select Cases* pažymėkite variantą *If condition is satisfied* ir spragtelėkite mygtuką *If...*
- Atsidarys dialogo langelis *Select Cases: If* (4.8 pav). Šis dialogo langelis sudarytas iš šių pagrindinių dalių:
 - Pirminių kintamųjų sąrašo.
 - Sąlygos redaktoriaus laukelio, kuriame įrašoma loginė sąlyga, pagal kurią bus atrenkami duomenys.
 - Klaviatūros, kuria, spragtelint pele atitinkamą mygtuką, pagal analogiją su skaičiuokliu, galima įrašinėti į sąlygos išraišką skaitmenis, aritmetinius ir loginius operatorius.
 - Funkcijų sąrašo (iš viso apie 140 funkcijų).



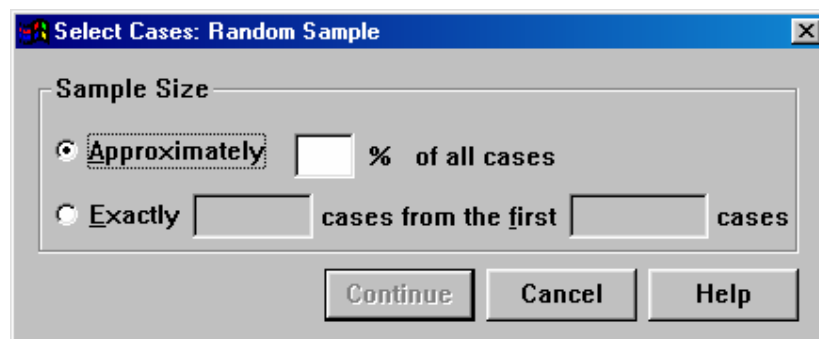
4.8 pav. Dialogo langelis duomenų atrinkimo sąlygai nustatyti *Select Cases: If*

- Redaktoriaus laukelyje parašykite sąlygos duomenims atrinkti išraišką. Pasirinktus kintamuosius, taip pat funkcijas įkelkite į minėtą laukelį spragtelėję atitinkamą mygtuką su trikampiu. Konstantas, aritmetinius, santykio: < (mažiau), > (daugiau), <= (mažiau arba lygu), >= (daugiau arba lygu), ~= (nelygu) bei loginius: & (loginis IR), | (loginis ARBA), ~ (loginis NE) operatorius įrašykite į išraišką naudodamiesi dialogo langelio klaviatūra.
- Spragtelėkite mygtuką *Continue*, tada dialogo langelyje *Select Cases* — *OK*.
- Bus atrinkti tie stebėjimų duomenys, kurių atžvilgiu loginis reiškinys turės sprendinį *true* (sąlyga tenkinama). Duomenys, kurių atžvilgiu loginis reiškinys turės sprendinį *false* (sąlyga netenkinama) arba *missing* (praleisti duomenys), atrinkti nebus.

4.5.2. Atsitiktinė duomenų atranka

Kai stebėjimų skaičius yra didelis, pirminiam hipotezės patikrinimui gali būti tikslinga naudoti nedidelę atsitiktinę imtį.

- Atsitiktinei duomenų atrankai iš stebėjimų visumos dialogo langelyje **Select Cases** pažymėkite variantą **Random sample of cases** ir spragtelėkite mygtuką **Sample...**
- Atsidarys dialogo langelis **Select Cases: Random Sample** (4.9 pav.), kuriame Jūs galite pasirinkti šiuos imties dydžio nurodymo būdus:
 - **Approximately** (apytiksliai). Įrašykite į laukelį procentinę atsitiktinės imties dydžio reikšmę — SPSS programa sugeneruos atsitiktinę imtį, savo dydžiu apytiksliai atitinkančią nurodytą visų stebėjimų dalį.
 - **Exactly** (tiksliai). Šiuo atveju nurodykite tikslų stebėjimų skaičių atsitiktinėje imtyje. Be to, nurodykite skaičių stebėjimų, iš kurių bus paimta atsitiktinė imtis. Pastarasis skaičius neturi viršyti visų stebėjimų skaičiaus duomenų rinkmenoje.

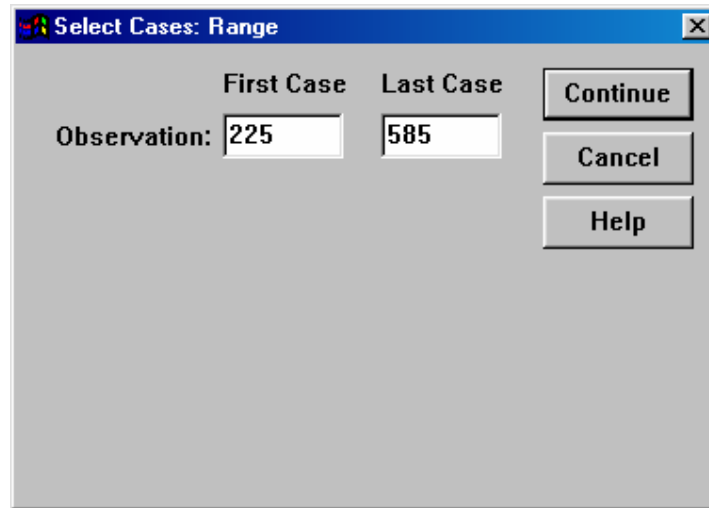


4.9 pav. Dialogo langelis atsitiktinės imties dydžiui nustatyti **Select Cases: Random Sample**

- Spragtelėkite mygtuką **Continue**, tada dialogo langelyje **Select Cases** — **OK**.

4.5.3. Diapazoninė duomenų atranka

- Dialogo langelyje **Select Cases** pažymėkite variantą **Based on time or case range** ir spragtelėkite mygtuką **Range...**
- Atsidariusiame dialogo langelyje **Select Cases: Range** (4.10 pav.) laukelyje **First Case** nurodykite diapazono pradžią (duomenų įrašo eilutės numerį), o laukelyje **Last Case** — diapazono pabaigą. Datos ir laiko diapazonai gali būti nurodyti tik atitinkamo formato duomenims.
- Spragtelėkite mygtuką **Continue**, po to dialogo langelyje **Select Cases** — **OK**.



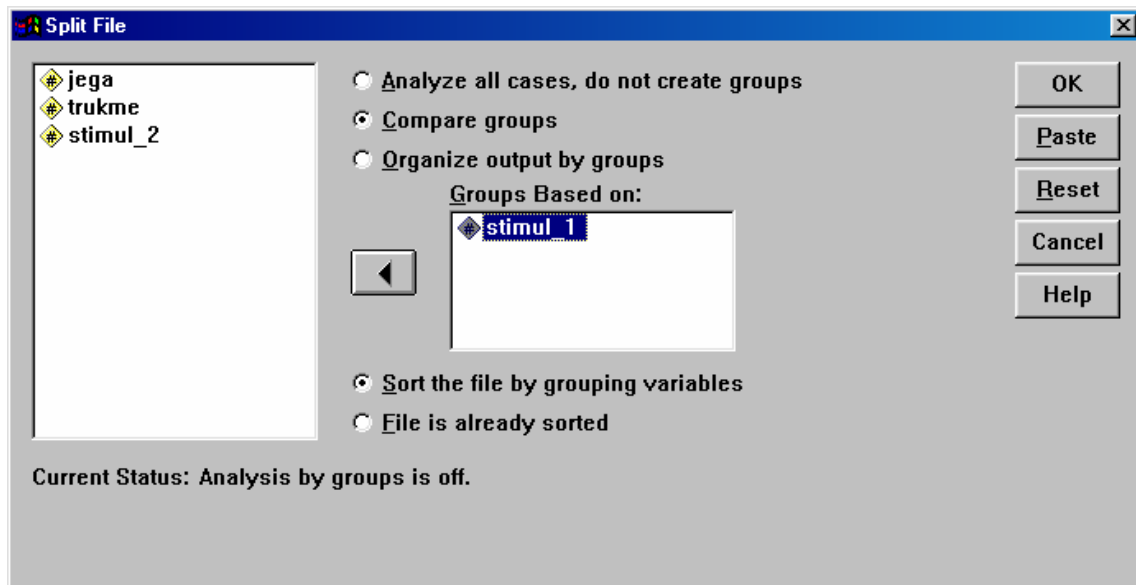
4.10 pav. Dialogo langelis imties diapazonui nustatyti *Select Cases: Range*

4.6. DUOMENŲ RINKMENOS SKĖLIMAS

SPSS galima duomenų rinkmenos atskirų grupių, sudarytų pagal vieno ar daugiau kategorinių kintamųjų reikšmes, analizę. Pasirinkus keletą kategorinių kintamųjų, stebėjimų duomenys bus suskirstyti pagal kategorinių kintamųjų prioritetą. Pavyzdžiui, jeigu pirmuoju kategoriniu kintamuoju bus pasirinktas kintamasis *gender* (lytis), o antruoju — *age* (amžius), stebėjimai bus suskirstyti pagal amžių abiem lyties kategorijoms. Galima nurodyti iki aštuonių kategorinių kintamųjų.

Jei norite suskaldyti duomenų rinkmeną:

- Nurodykite komandas **Data → Split File...**
- Dialogo langelio **Split File** (4.11 pav.) laukelyje **Groups Based on** nurodykite vieną ar kelis kategorinius kintamuosius.
- Pasirinkite norimą išvesties pateikimo metodą:
 - **Compare groups** (palyginti grupes) — atskirų grupių analizės rezultatai pateikiami kartu sudarant galimybę juos lengvai sulyginti;
 - **Organize output by groups** (išvestis pagal grupes) — analizės rezultatai pateikiami atskirai kiekvienai grupei.
- Spragtelėkite dialogo langelio mygtuką **OK**.
Norėdami grąžinti duomenų rinkmeną į pradinę būseną:
- Nurodykite komandas **Data → Split File...**
- Dialogo langelyje **Split File** pažymėkite laukelį **Analyze all cases, do not create groups** (analizuoti visus stebėjimus, nesudarant grupių).
- Spragtelėkite dialogo langelio mygtuką **OK**.

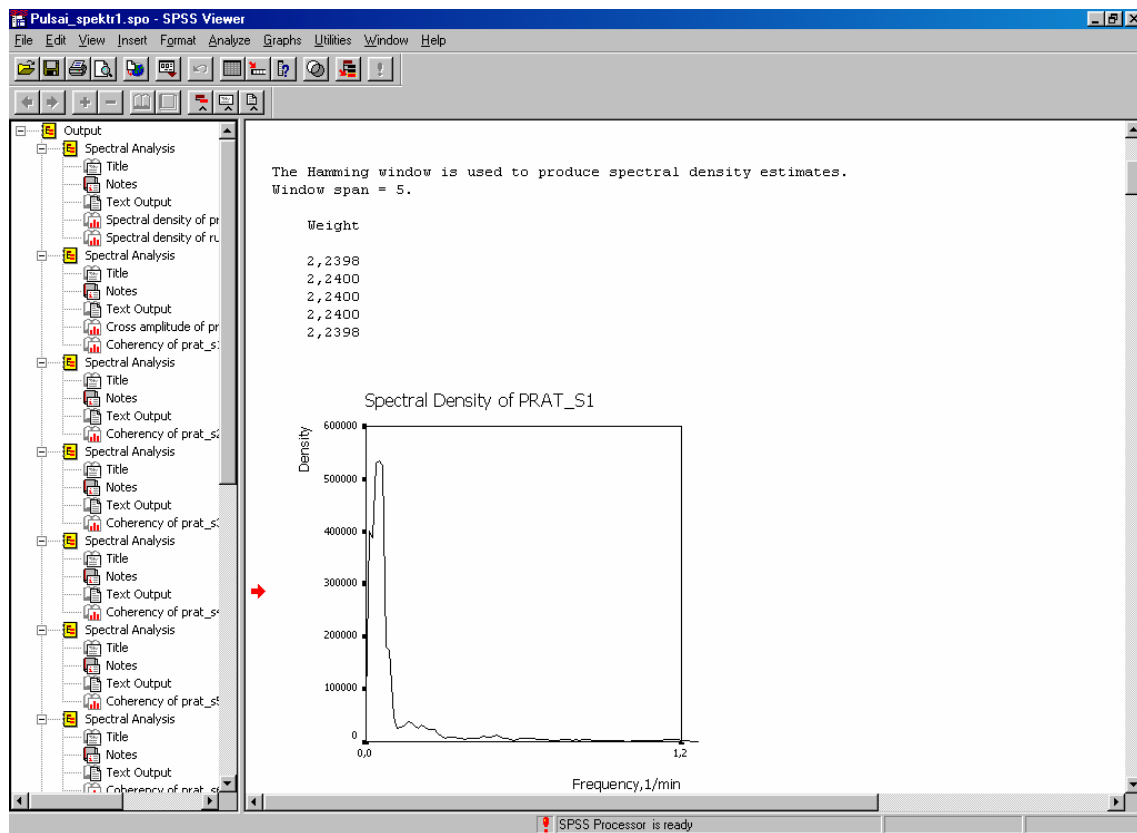


4.11 pav. Dialogo langelis duomenų rinkmenai suskaldyti į grupes **Split File**

5. GAUTŲ REZULTATŲ REDAGAVIMAS

Duomenų analizės rezultatai yra pateikiami rodinio (*Viewer*) lange. Rodinio langas atidaromas automatiškai, kai atliekama komanda, po kurios pateikiami rezultatai. Šiame lange Jūs galite peržiūrėti gautus rezultatus, keisti jų pateikimo pobūdį ir tvarką.

Rodinio langą (5.1 pav.) sudaro dvi dalys. Kairėje (plano) pusėje pateikiama gautų rezultatų apžvalga (*Outline view*), o dešinėje pusėje duodamos lentelės su gautais rezultatais, grafikai ir aiškinamasis tekstas. Lango dalių plotį Jūs galite keisti vilkdami pele šias dalis skiriančią ribą. Rezultatų peržiūrai naudokite slinkties juostas arba spragtelėkite pele *Outline* dalyje pasirinktą poziciją, kad patektumėte į reikiamą peržiūros lango vietą.




5.1 pav. Rodinio langas *Viewer*

Kiekvienos statistinės duomenų apdorojimo procedūros rezultatai ir jų grafinis pavaizdavimas *Viewer* lange pateikiami blokais, kurių kiekvienas yra atskiras objektas.

Outline dalies hierarchiniame (nurodančiame blokų lygmenis ir priklausomybę) sąraše kiekvienas blokas yra įvardijamas pagal vykdomą statistinę procedūrą, o prieš bloko pavadinimą figūruoja atitinkamas bloko

ženklas. Prieš šį ženklą yra išskleidimo ir sutraukties mygtukas (pradžios pozicijoje rodomas minuso ženklas). Kiekvieno bloko viduje pirmiausia yra nurodomas pavadinimas (*Title*) ir pastabos (*Notes*), toliau išvardijami bloko elementai, prieš kuriuos taip pat figūruoja atitinkami simboliai. Tokia objektų hierarchinė struktūra Jums leidžia atlikti reikiamų elementų paiešką, sukeisti juos vietomis, kopijuoti, pašalinti ir t. t.

Jeigu dirbate su keliais **Viewer** langais, gauti rezultatai dedami nurodytame (*designated*) lange. Nurodyto lango nereikia supainioti su aktyvuoju langu, kuris yra atidarytas monitoriaus ekrane. Nurodydami **Viewer** langą, pirmiausia jį suaktyvinkite, tada spragtelėkite pele įrankių juostoje mygtuką su šauktuko ženklu  arba nurodykite komandas **Utilities** → **Designate Window**. Atidarant naują **Viewer** langą, jis automatiškai tampa aktyvus ir nurodytas.

5.1. VEIKSMAI VIEWER LANGE


Šiame lange Jūs galite:

- Peržiūrėti gautus rezultatus.
- Rodyti arba paslėpti pasirinktas lenteles ar grafikus.
- Sukeisti rezultatų rodymo tvarką.
- Kopijuoti ir perkelti duomenis į kitas duomenų rinkmenas.

5.1.1. Kaip parodyti arba paslėpti rezultatus?


Sumažindami vaizduojamos informacijos kiekį, Jūs galite pasirinktinai rodyti arba paslėpti atskirus gautų rezultatų elementus (lenteles, grafikus) ar visus blokus.

Paslėpti dalį gautų rezultatų galima šiais būdais:

- Dvigubai spragtelėkite pele atitinkamą simbolį objektų sąrašė *Outline* dalyje.
- Pažymėkite norimą elementą vieną kartą spragtelėdami pele ir vykdykite komandas **View** → **Hide**.
- Pažymėkite norimą elementą vieną kartą spragtelėdami pele ir spragtelėkite įrankių juostoje uždarytos knygos (*Hide*) mygtuką .

Padaryti paslėptą elementą vėl matomą galima šiais būdais:

- Dukart spragtelėkite pele atitinkamą simbolį objektų sąrašė *Outline* dalyje.
- Pažymėkite norimą elementą vieną kartą spragtelėdami pele ir vykdykite komandas **View** → **Show**.

- Pažymėkite norimą elementą vienąkart spragtelėdami pele ir spragtelėkite įrankių juostoje atverstos knygos (*Show*) mygtuką .

Paslėpti visą atskiros procedūros rezultatų bloką galima šiais būdais:

- Spragtelėkite pele mažą kvadratuką šio bloko ženklo kairėje. Minuso ženklas pavirs pliuso ženklu, ir visas bloko turinys dings.
- Pažymėkite norimą bloką vienąkart spragtelėdami pele ir vykdykite komandas **View → Collapse**.
- Pažymėkite norimą bloką vienąkart spragtelėdami pele ir spragtelėkite įrankių juostoje mygtuką su minuso ženklu (*Collapse*).

Padaryti paslėptą bloką vėl matomą galima šiais būdais:

- Spragtelėkite pele mažą kvadratuką šio bloko ženklo kairėje. Pliuso ženklas pavirs minuso ženklu, ir visas bloko turinys taps matomas.
- Pažymėkite norimą bloką vienąkart spragtelėdami pele ir vykdykite komandas **View → Expand**.
- Pažymėkite norimą bloką vienąkart spragtelėdami pele ir spragtelėkite įrankių juostoje mygtuką su pliuso ženklu (*Expand*).

5.1.2. Kaip kopijuoti, perkelti arba pašalinti gautus rezultatus?

Gautus rezultatus galima pertvarkyti kopijuojant, perkeliant arba pašalinant atskirus elementus arba jų grupes. Norėdami kopijuoti tam tikrus gautų rezultatų duomenis į kitą vietą:

- *Outline* arba pagrindinėje rodinio lango dalyje pažymėkite norimą elementą ar bloką (norėdami pažymėti keletą iš eilės einančių elementų ar blokų, pažymėkite pirmąjį jų, tada, laikydami nuspaudę **Shift** klavišą — paskutinį, o norėdami pažymėti keletą ne iš eilės einančių elementų ar blokų, pažymėkite pirmąjį jų, tada, laikydami nuspaudę **Ctrl** klavišą — visus kitus).
- Nurodykite komandas **Edit → Copy**.
- Spragtelėkite pele elementą, po kurio turi būti įterptas kopijuojamasis ir nurodykite komandas **Edit → Paste After**.

Nukopijuoti elementą (bloką) galite taip pat vilkdami jį pele:

- Spragtelėkite pele kopijuojamą elementą (bloką), neatleisdami pelės klavišo nuspauskite klaviatūros klavišą **Ctrl**, nutempkite kopijuojamą elementą į norimą vietą ir atleiskite pelės klavišą.

Norėdami perkelti rezultatus į kitą vietą:

- Pažymėkite elementą (–us) ar bloką (–us) anksčiau aprašyta tvarka.
- Nurodykite komandas **Edit → Cut**.
- Spragtelėkite pele elementą, po kurio turi būti įterptas perkeliamasis ir nurodykite komandas **Edit → Paste After**.

Perkelti elementą (bloką) galite taip pat vilkdami jį pele:

- Spragtelėkite pele perkeliamą elementą, neatleisdami pelės klavišo nutempkite perkeliamą elementą į norimą vietą ir atleiskite pelės klavišą.

Norėdami pašalinti rezultatus:

- Pažymėkite elementą (–us) ar bloką (–us) anksčiau aprašyta tvarka.
- Nurodykite komandas **Edit → Delete** arba paspauskite klaviatūros klavišą **Delete**.

5.1.3. Kaip pakeisti rezultatų lygmenį?

Perkeliant rezultatus lango *Outline* dalyje, gali būti pakeistas jų lygmuo ir priklausomybė. Atkurti reikiamą hierarchiją galite naudodamiesi įrankių juostos mygtukais **Promote** (paaukštinti lygmenį) — rodyklė į kairę ir **Demote** (pažeminti lygmenį) — rodyklė į dešinę.

5.1.4. Kaip pakeisti *Outline* sąrašo šriftą?

Outline ženklų dydžiui pakeisti nurodykite komandas **View → Outline Size** ir pasirinkite **Small** (mažas) — pradinis nustatymas, **Medium** (vidutinis) arba **Large** (didelis).

Šriftui pakeisti nurodykite komandas **View → Outline Font...** ir pasirinkite norimą šriftą.

5.1.5. Kaip įterpti naują objekto pavadinimą ar papildomą tekstą?

Pavadinimui ar kitam papildomam tekstui įterpti spragtelėkite pele atitinkamą objektą (pavadinimą, lentelę, grafiką ir t. t.), po kurio turėtų eiti Jūsų pageidaujamas tekstas, ir nurodykite komandas **Insert → New Title** (naujas pavadinimas) arba **Insert → New Text** (naujas tekstas). Teksto laukelyje įrašykite reikiamą tekstą. Jeigu reikiamas tekstas yra atskiroje rinkmenoje, nurodykite komandas **Insert → Text File...** ir atsidariusiame dialogo langelyje nurodykite duomenų rinkmenos vardą.

5.2. REZULTATŲ IŠSAUGOJIMAS

Viewer lango turinys išsaugomas kaip atskira duomenų rinkmena su plėtiniu *.spo*. Gautiems rezultatams išsaugoti nurodykite komandas **File → Save As...**, atsidariusiame dialogo langelyje nurodykite bylos saugojimo vietą, vardą ir spragtelėkite mygtuką **Save**.

Viewer rezultatams su slaptažodžiu išsaugoti nurodykite komandas **File → Save with Password...**, dialogo langelyje įrašykite slaptažodį (iki 16 simbolių, didžiosios ir mažosios raidės traktuojamos skirtingai), pakartokite

slaptažodį ir spragtelėkite dialogo langelio mygtuką **OK**. Atsidariusiame dialogo langelyje **Save As...** nurodykite bylos saugojimo vietą, vardą ir spragtelėkite mygtuką **Save**.

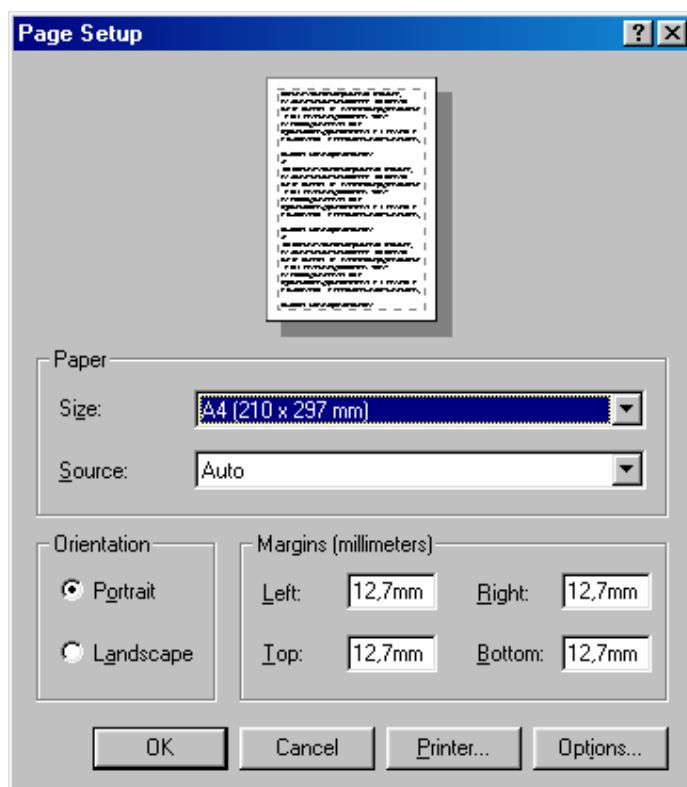
5.3. IŠVESTIES DOKUMENTO SPAUSDINIMAS

Dialogo langelyje, kuris pasirodo nurodžius komandas **File → Print...**, Jūs galite nurodyti šiuos dokumento spausdinimo variantus:

- **All visible output** (matoma išvesties dalis) — spausdinami tik matomi **Viewer** lange objektai, paslėpti objektai nespausdinami.
- **All output** (visa išvestis) — spausdinami visi objektai, įskaitant paslėptus.
- **Selection** (pasirinktinai) — spausdinami nurodyti objektai.

Prieš spausdinimą būtina nustatyti puslapio parametrus:

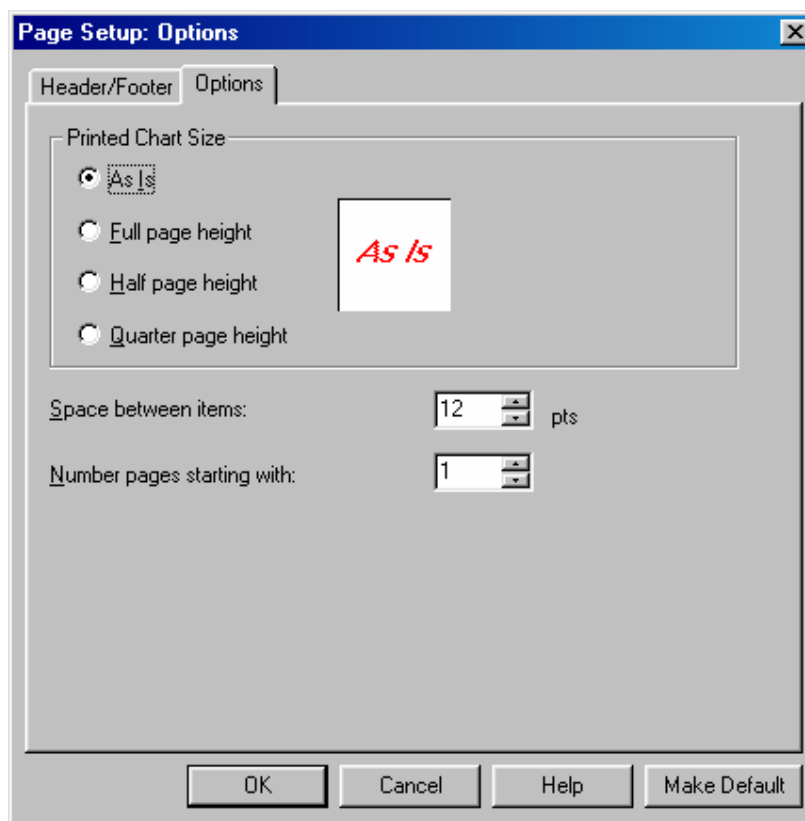
- Nurodykite komandas **File → Page Setup...** ir atsidariusiame dialogo langelyje (5.2 pav.) nurodykite: popieriaus lapo formatą ir orientaciją, dokumento laukelių plotį.



5.2 pav. Lapo parametrų nustatymo dialogo langelis **Page Setup**.

- Spragtelėkite dialogo langelio mygtuką **Options...** ir naujame dialogo langelyje (5.3 pav.), kortelėje **Header/Footer** įrašykite viršutinės ir apatinės puslapių antraščių tekstą.

- Šio dialogo langelio kortelėje **Options** galite pasirinkti spausdinamo grafiko dydį, tarpo tarp spausdinamų objektų dydį ir nustatyti puslapių numeracijos tvarką.



5.3 pav. Lapo parametrų nustatymo dialogo langelis **Page Setup: Options**

6. MOBILIOSIOS LENTELĖS

SPSS daugelis skaičiavimo rezultatų pateikiami vadinamųjų mobiliųjų lentelių (*Pivot Tables*) pavidalu. Tai lentelių forma, leidžianti sukeisti vietomis eilutes, stulpelius ir sluoksnius ir todėl leidžianti įvertinti rezultatus įvairiu aspektu.

6.1. MOBILIŲJŲ LENTELIŲ STRUKTŪROS KEITIMAS

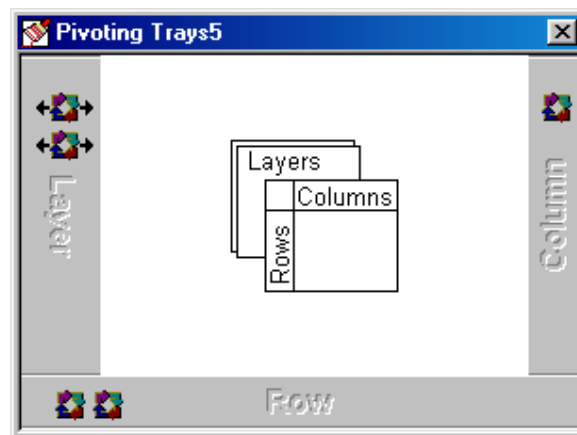
Galimi šie mobiliųjų lentelių struktūros keitimo veiksmai:

- Eilučių ir stulpelių sukeitimas vietomis.
- Eilučių ir stulpelių perkėlimas.
- Daugiasluoksnės struktūros sukūrimas.
- Eilučių ir stulpelių suskirstymas į grupes ir grupių išardymas.
- Lentelės ląstelių paslėpimas.
- Eilučių ir stulpelių pavadinimų krypties pakeitimas.
- Reikiamos išraiškos nustatymas.

Redaguoti mobiliąją lentelę galima dukart spragtelėjus ją pele. Bus aktyvuotas mobiliųjų lentelių redaktorius ir pasikeis meniu juosta.

6.1.1. Kaip redaguoti lenteles naudojant piktogramas?

- Aktyvuokite lentelę dukart spragteldami ją pele bet kurioje vietoje.
- Nurodykite komandas **Pivot → Pivoting Trays**. Atsidarys dialogo langelis **Pivoting Trays** (6.1 pav.) su trimis laukeliais: *Layer* (sluoksnis), *Row* (eilutė) ir *Column* (stulpelis). Kiekviename laukelyje priklausomai nuo lentelės struktūros bus tam tikras ženkliukų (piktogramų) skaičius. Kokį lentelės komponentą reiškia konkretus ženkliukas, sužinosite nukreipę į jį pelės žymeklį — pasirodys atitinkamas užrašas.



6.1 pav. Lentelių redagavimo dialogo langelis **Pivoting Trays**

- Perkeldami pele pasirinktą ženkluką iš vieno laukelio į kitą, galite keisti lentelės struktūrą. Pavyzdžiui, tarkime, kad eilutės (*Row*) laukelyje esantį ženkluką, kuris simbolizuoja kintamąjį su reikšmėmis *Taip* ir *Ne*, Jūs perkeliate į stulpelio (*Column*) laukelį. Prieš perkeliant *Taip* ir *Ne* buvo eilučių žymenos, po perkėlimo taps stulpelių žymenomis.

6.1.2. Kaip sukeisti vietomis eilutes ir stulpelius?

- Aktyvuokite lentelę dukart spragteldami ją pele.
- Nurodykite komandas ***Pivot → Transpose Rows and Columns***. Šis veiksmas tolygus visų piktogramų, esančių eilutės (*Row*) laukelyje, perkėlimui į stulpelio (*Column*) laukelį ir visų piktogramų, esančių stulpelio (*Column*) laukelyje, perkėlimui į eilutės (*Row*) laukelį.

6.1.3. Kaip pakeisti lentelės elementų eiliškumą?

Piktogramų eiliškumas ***Pivoting Trays*** dialogo langelyje lemia lentelės elementų eiliškumą. Norėdami pakeisti lentelės elementų eiliškumą:

- Aktyvuokite lentelę dukart spragteldami pele.
- Nurodykite komandas ***Pivot → Pivoting Trays*** ir vilkdami pele išdėstykite piktogramas kiekviename laukelyje tokia tvarka, kokia pageidaujate.

6.1.4. Kaip perkelti eilutes ir stulpelius?

- Aktyvuokite lentelę dukart spragteldami ją pele.
- Spragtelėkite eilutės ar stulpelio, kurį norite perkelti, pavadinimą.
- Spragtelėkite pavadinimą ir neatleisdami pelės klavišo vilkite į naują poziciją. Atleiskite klavišą ir spragtelėkite iššokusiam kontekstiniam menu ***Insert Before*** (įterpti prieš) arba ***Swap*** (apkeisti). Šiam veiksmui atlikti būtina, kad ***Edit*** menu komanda ***Drag to Copy*** nebūtų pažymėta.

6.1.5. Kaip suskirstyti eilutes ir stulpelius į grupes?

- Aktyvuokite lentelę dukart spragteldami ją pele.
- Pažymėkite (spragtelėkite pele ir vilkite neatleisdami klavišo arba spragtelėkite pele nuspaudę klaviatūros klavišą ***Shift***) surenkamų į grupę eilučių arba stulpelių pavadinimų ląsteles.
- Nurodykite komandas ***Edit → Group***. Bus įterptas surinktų į grupę eilučių arba stulpelių bendras pavadinimas, kurio tekstą galėsite redaguoti jį dukart spragtelėję pele.
- Norėdami išskirstyti sugrupuotas lentelės eilutes arba stulpelius aktyvuokite lentelę, spragtelėkite pele grupės antraštinę ląstelę ir vykdykite komandas ***Edit → Ungroup***.

6.1.6. Kaip pakeisti eilučių ir stulpelių pavadinimų kryptį?

- Aktyvuokite lentelę dukart spragteldami ją pele.
- Vykdykite komandas **Format → Rotate InnerColumn Labels** arba **Rotate OuterRow Labels**.

Pasukti galima tik išorinius eilučių ir vidinius stulpelių pavadinimus.

6.1.7. Kaip atkurti originalų lentelės pavidalą?

Atlikę vieną ar keletą lentelės transformavimo veiksmų, pradinę lentelės pavidalą galite atkurti aktyvavę lentelę ir nurodę komandas **Pivot → Reset Pivots to Defaults**. Komanda **Reset Pivots to Defaults** negrąžina lentelės į pradinę būseną po eilučių ir stulpelių suskirstymo į grupes arba perkėlimo veiksmų.

6.1.8. Kaip paslėpti lentelės ląsteles?

Lentelės struktūrą galima keisti, paslepianč nereikalingas ląsteles. Galima nerodyti šių ląstelių:

- Dydžių pavadinimai.
- Kategoriniai duomenys, įskaitant duomenų pavadinimus ir pačius duomenis.
- Išnašos, pavadinimai ir paaiškinimai.

Atskiroms eilutėms ar stulpeliams paslėpti aktyvuokite lentelę ją dukart spragteldami pele, laikydami nuspaustus klaviatūros klavišus **Ctrl + Alt** spragtelėkite kategorinių duomenų eilutės ar stulpelio pavadinimą ir vykdykite komandas **View → Hide**.

Kad paslėptos eilutės ar stulpeliai vėl būtų matomi, aktyvuokite lentelę ir vykdykite komandas **View → Show All**.

Išnašas, pavadinimus ir paaiškinimus paslėpsite (rodysite) juos pažymėdami ir vykdydami komandas **View → Hide (Show)**.

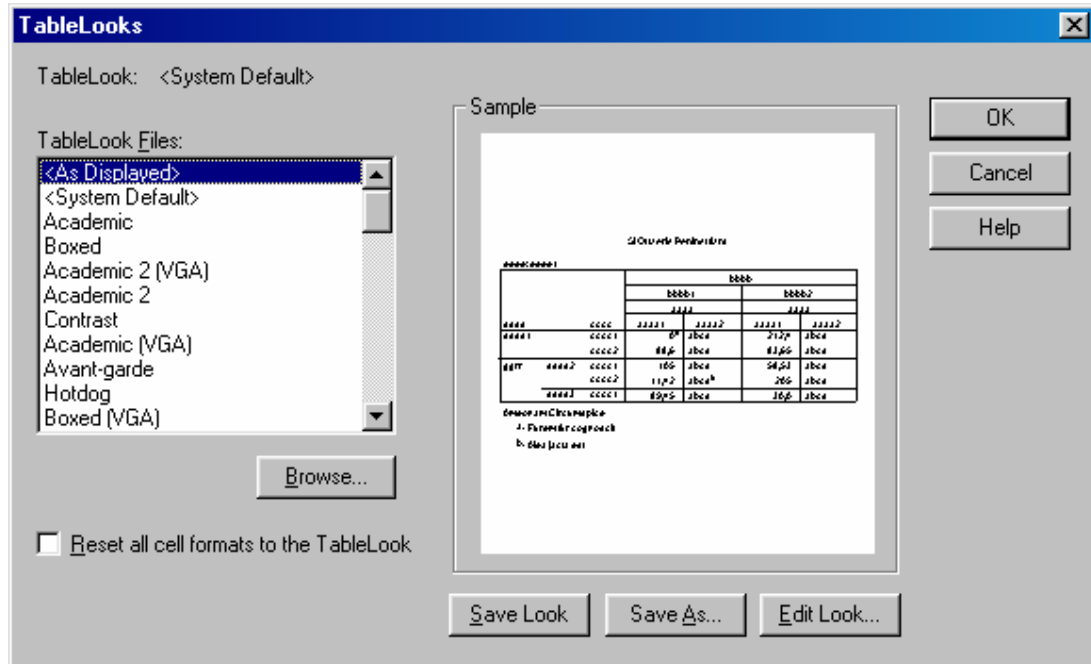
6.2. MOBILIŲJŲ LENTELIŲ REDAGAVIMAS

Be išvardytųjų lentelių struktūros keitimo galimybių yra dar šios papildomos redagavimo galimybės:

- Bendro lentelės vaizdo pasirinkimas.
- Lentelės savybių keitimas.
- Lentelės ląstelių savybių keitimas.
- Lentelės teksto modifikavimas.
- Paaiškinimų ir išnašų įterpimas.

6.2.1. Kaip pakeisti lentelės išvaizdą?

- Aktyvuokite lentelę dukart spragteldami ją pele.
- Nurodykite komandas **Format** → **TableLooks...** (lentelės išvaizda). Atsidarys dialogo langelis **TableLooks** (6.2 pav.), kuriame iš pateikto sąrašo galite pasirinkti pageidaujamą lentelės modelį.
- Spragtelėkite mygtuką **OK**, kad Jūsų aktyvuota lentelė įgautų pasirinktą pavidalą.



6.2 pav. Dialogo langelis **TableLooks** lentelės išvaizdai pasirinkti

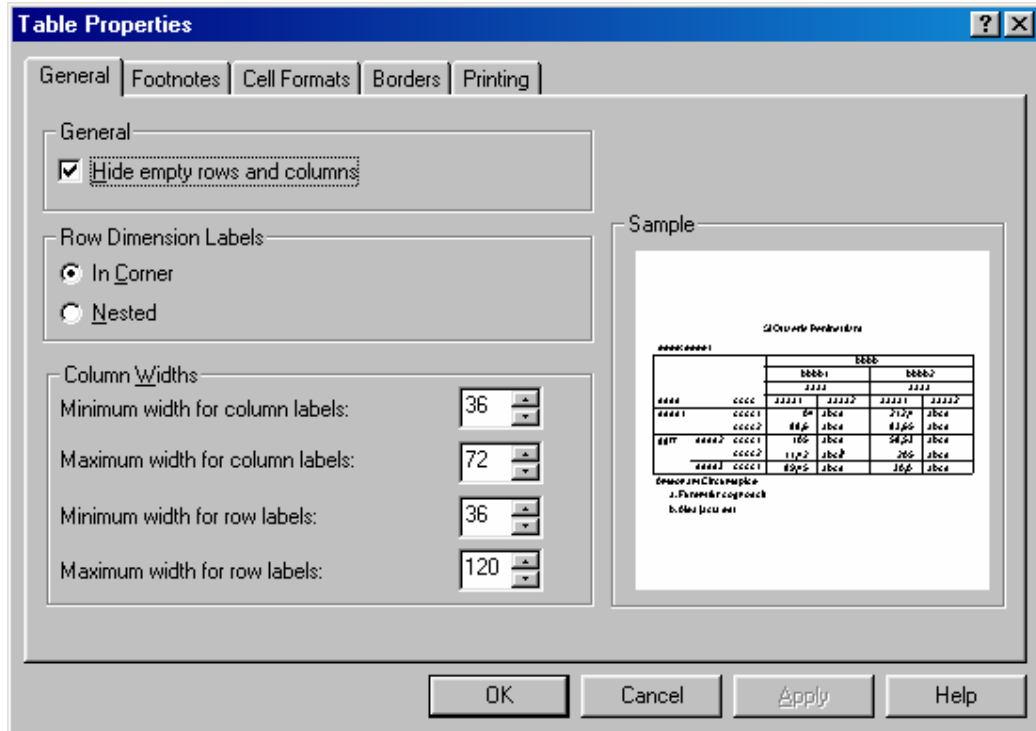
- Spragtelėję dialogo langelio **TableLooks** mygtuką **Edit Look...**, Jūs galite atidaryti pagalbinį dialogo langelį **Table Properties** (lentelės savybės), kuriame galite pakeisti atskirus lentelės komponavimo elementus pageidaujamais. Spragtelėkite mygtuką **Save Look** pakeistam lentelės dizainui išsaugoti arba mygtuką **Save As...** jam išsaugoti nauju vardu. Dialogo langelyje **TableLooks** padaryti pakeitimai galioja tik aktyvuotai lentelei.

6.2.2. Kaip pakeisti lentelės savybes?

Norėdami pakeisti lentelės savybes:

- Aktyvuokite lentelę dukart spragteldami ją pele.
- Nurodykite komandas **Format** → **Table Properties...** (lentelės savybės).

- Dialogo langelyje **Table Properties** (6.3 pav.) pasirinkite reikiamą kortelę (spragtelėkite auselę su kortelės vardu) — **General** (bendri), **Footnotes** (išnašos), **Cell Formats** (ląstelių formatai), **Borders** (rėmeliai), **Printing** (spausdinimas) ir nustatykite pageidaujamus lentelės parametrus.



6.3 pav. Dialogo langelis **Table Properties** lentelės parametrus nustatyti

- Spragtelėkite mygtukus **Apply** ir **OK**.

Galimi pasirinkimai lengvai suvokiami iš dialogo langelio, todėl atskirai panagrinėsime tik sąlygiškai sudėtingesnę – **Cell Formats** (ląstelių formatai) kortelę.

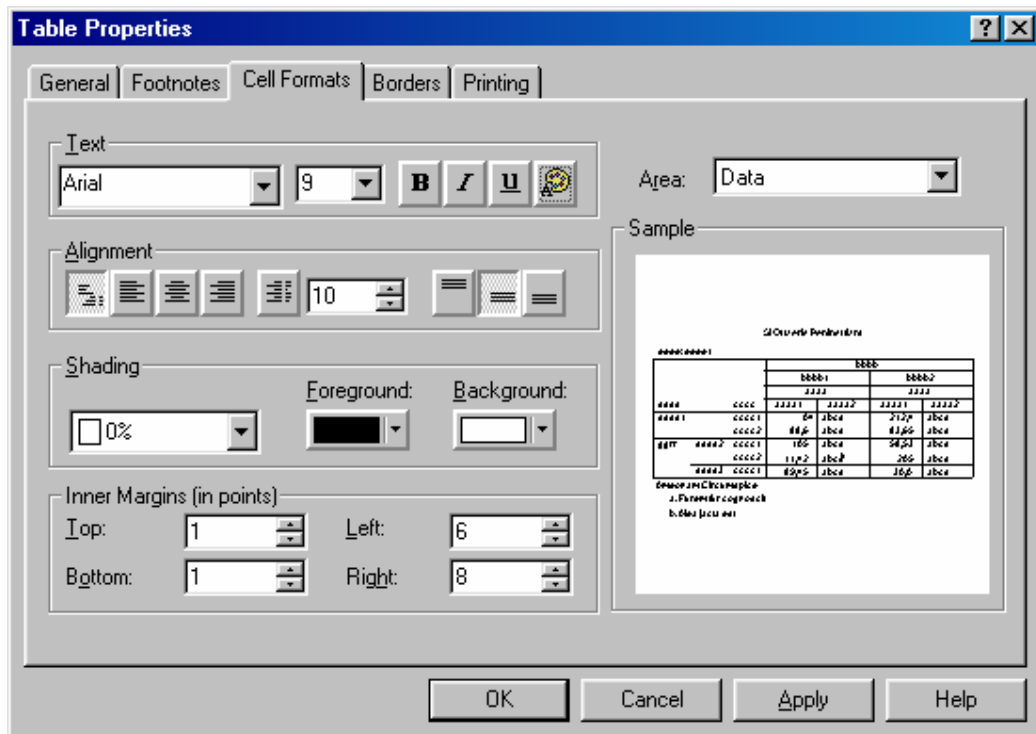
Formato pasirinkimui lentelė yra suskirstyta į šias sritis (6.4 pav.): **Title** (pavadinimas), **Layers** (sluoksniai), **Corner Labels** (kampinės žymenos), **Row Labels** (eilučių pavadinimai), **Column Labels** (stulpelių pavadinimai), **Data** (duomenys), **Caption** (paaiškinimai) ir **Footnotes** (išnašos). Kiekvienai lentelės sričiai Jūs galite nustatyti atitinkamus ląstelių formatus: teksto charakteristikas (šriftą, jo dydį, spalvą, stilių), lygiavimą pagal horizontalią ir vertikalią, ląstelių užtamsinimą, priekinio plano ir fono spalvas ir vidines ląstelių paraštes. **Table Properties** dialogo langelio **Cell Formats** kortelėje nurodyti formatai galioja visai lentelės sričiai. Kaip nustatyti formatus atskiroms ląstelėms, paaiškinama toliau.

Title (pavadinimas)

<i>Layers (sluoksniai)</i>	
<i>Corner Labels (kampinės žymenos)</i>	<i>Column Labels (stulpelių pavadinimai)</i>
<i>Row Labels (eilučių pavadinimai)</i>	<i>Data (duomenys)</i>
<i>Caption (paaiškinimai)</i>	
<i>Footnotes (išnašos)</i>	

6.4 pav. Lentelės formatų nustatymo sritys

- **Cell Formats** kortelėje (6.5 pav.) **Area** (sritis) išskleidžiamajame sąraše pasirinkite lentelės sritį, kurios ląstelių formatus keisite.
- Pasirinkite pageidaujamus šiai sričiai teksto (**Text**), lygiavimo (**Alignment**), užtamsinimo (**Shading**), priekinio plano (**Foreground**), fono (**Background**), vidinių paraščių (**Inner Margins**) parametrus. Jūsų pasirinkimas atsispindės **Sample** (pavyzdys) laukelyje.
- Spragtelėkite mygtukus **Apply** ir **OK**.

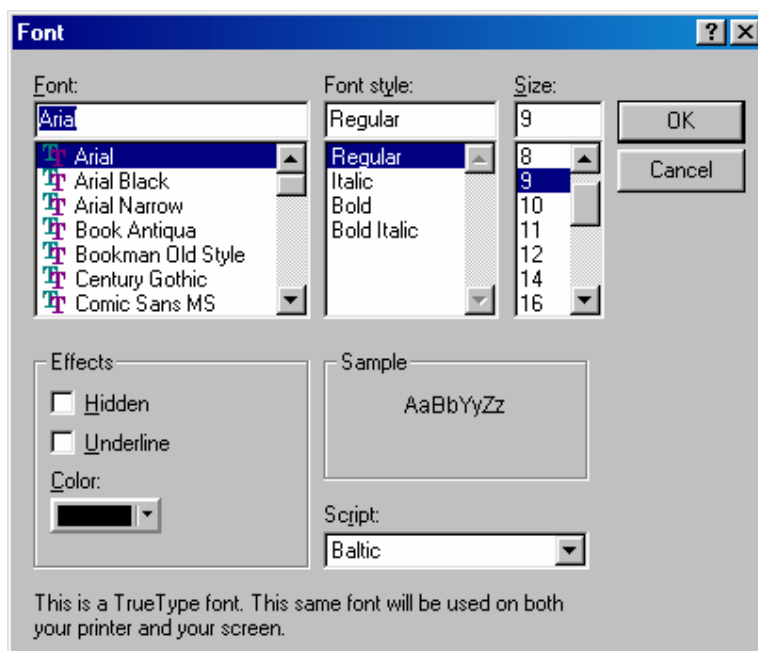
6.5 pav. Dialogo langelio **Table Properties** kortelė **Cell Formats** lentelės parametrus nustatyti

Dialogo langelyje **Table Properties** padaryti pakeitimai galioja tik aktyvuotai lentelei.

6.2.3. Kaip pakeisti atskirų ląstelių savybes?

Atskiros ląstelės šriftui pakeisti:

- Aktyvuokite lentelę dukart spragteldami ją pele. Spragtelėkite pele ląstelę, kurios tekstą norite keisti.
- Nurodykite komandas **Format** → **Font...** ir dialogo langelyje **Font** (6.6 pav.) pasirinkite pageidaujamą šriftą, stilių, šrifto dydį, spalvą ir pabrauktą ar paslėptą tekstą.
- Spragtelėkite mygtuką **OK**.



6.6 pav. Dialogo langelis **Font** atskirų ląstelių šriftui nustatyti

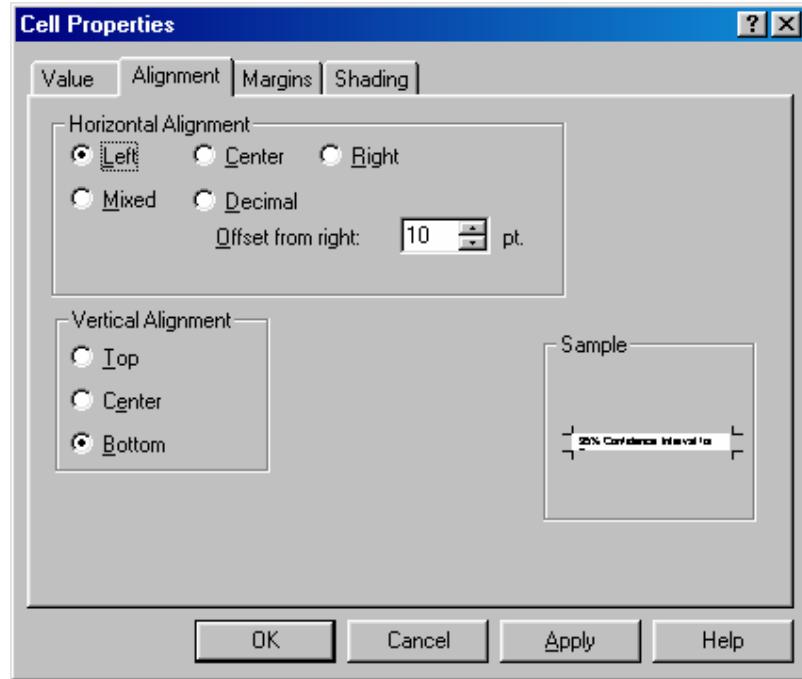
Atskiros stulpelio pločiui pakeisti:

- Aktyvuokite lentelę dukart spragteldami ją pele.
- Nutaikykite pelės žymeklį į lentelės stulpelio antraštės dešinę kraštinę. Pelės žymeklis virsta specialiu stulpelio pločio keitimo ženklu — dvipuse rodykle. Vilkite pele šį ženklą, kol stulpelis pasidarys reikiamo pločio.

Atskirų ląstelių savybėms (skaičių formatams, teksto išdėstymui, vidinėms paraštėms, užtamsinimui) pakeisti:

- Aktyvuokite lentelę dukart spragteldami ją pele.
- Pažymėkite reikiamą ląstelę ir nurodykite komandas **Format** → **Cell Properties...** (ląstelės savybės).
- Dialogo langelio **Cell Properties** (6.7 pav.) kortelėje **Value** (reikšmė) Jūs galite nustatyti skaičių (*Number*), datų (*Date*), laiko (*Time*) ir valiutos (*Currency*) formatus, kortelėje **Alignment** (lygiavimas) — teksto

išdėstymo ląstelėje horizontalia ir vertikalia kryptimi būdus, kortelėje **Margins** (paraštės) — vidinės paraštės ir kortelėje **Shading** (užtamsinimas) — ląstelės užtamsinimą (procentais), priekinio plano ir fono spalvas.



6.7 pav. Dialogo langelis **Cell Properties** atskirų ląstelių savybėms nustatyti

Dialogo langelyje **Cell Properties** nustatyti parametrai turi viršenybę prieš **Table Properties** langelyje nustatytus parametrus. Todėl, keisdami lentelės parametrus, jūs negalite pakeisti atskiroms ląstelėms nustatytų parametrų.

Eilutėms arba stulpeliams pažymėti:

- Aktyvuokite lentelę dukart spragteldami ją pele.
- Pažymėkite eilutę ar stulpelį vienu iš būdų:
 - Spragtelėkite eilutės ar stulpelio antraštę ir nurodykite komandas **Edit → Select → Data and Label Cells**.

Arba

- Spragtelėkite eilutės ar stulpelio antraštę laikydami nuspaustus klavišus **Ctrl+Alt**.

6.2.4. Kaip pakeisti pateiktus rezultatus?

Atskirų ląstelių tekstui pakeisti:

- Aktyvuokite lentelę dukart spragteldami ją pele.

- Dukart spragtelėkite reikiamą ląstelę arba spragtelėkite kartą ir paspauskite klavišą **F2**.
- Redaguokite tekstą. Paspauskite **Enter** klavišą, jei norite pataisytą tekstą įrašyti ląstelėje, arba **Esc** klavišą, jei norite atsisakyti pataisymų.

Paaishkinimui po lentele užrašyti:

- Aktyvuokite lentele dukart spragteldami ją pele.
- Nurodykite komandas **Insert → Caption**. Lentelės apačioje pasirodys užrašas *Table Caption*. Dukart spragtelėkite šį užrašą ir vietoje jo surinkite savo tekstą.

Pastabai (išnašai) po lentele užrašyti:

- Aktyvuokite lentele dukart spragteldami ją pele.
- Spragtelėkite aktyvuotos lentelės antraštę, ląstelę arba paaishkinimą ir nurodykite komandas **Insert → Footnote**. Lentelės apačioje pasirodys užrašas *Footnote*. Dukart spragtelėkite šį užrašą ir jo vietoje surinkite savo tekstą.

7. TRUMPA STATISTIKOS METODŲ APŽVALGA

Darbas su SPSS programa reikalauja pagrindinių statistikos kurso žinių, kurių programos vartotojas gali pasisemti iš knygų ir vadovėlių, orientuotų į skaitytoją, neturintį ypatingo matematinio pasirengimo, o tik susipažinusį su aukštosios matematikos ir tikimybių teorijos pagrindais (Gonestas, Strielčiūnas, 2003), (Sakalauskas, 2003) arba išklausiusį vieno semestro aukštosios matematikos kursą (Čekanavičius, Murauskas, 2000), (Čekanavičius, Murauskas, 2002). Šiame skyriuje trumpai priminsime tiksliai tas sąvokas ir teiginius, kuriuos privalu žinoti vartotojui, kad jis galėtų sėkmingai naudoti SPSS programą.

7.1. PIRMINĖS STATISTIKOS TESTO SĄLYGOS

Prieš taikant apibrėžtą statistikos testą, dažniausiai reikia atsakyti į klausimus:

- Kokiai matavimo skalei priklauso tiriamas kintamasis?
- Ar kintamojo rezultatai intervalų skalėje pasiskirstę pagal normalųjį dėsnį?
- Ar lyginamos imtys yra priklausomos ar nepriklausomos?

7.1.1. Duomenų matavimo skalės

Yra šios tyrimo duomenų matavimo skalės:

- **Vardinė**, arba **nominalinė** skalė. Būdingiausi vardinės skalės pavyzdžiai yra lytis, sportininkų komandos sąrašas, sporto šakų sąrašas ir t. t. Kintamasis, įgyjantis tik dvi reikšmes (kategorijas) dar vadinamas binariniu. Skaičiai, kuriais koduojami atskiri objektai ar jų savybės, neturi jokios empirinės reikšmės, tik rodo, kokia čia ypatybė ar objektas. Kintamųjų, priklausančių vardinei skalei (**nominaliųjų** kintamųjų) apdorojimo galimybės gana ribotos — galima tiksliai įvertinti, kurių objektų (savybių) yra daugiau ar mažiau, koks bendras visų sąraše esančių objektų kiekis. Pagal nominaliuosius kintamuosius dažnai vykdoma kokybinė duomenų klasifikacija, arba grupavimas — imtis suskaldoma pagal šių kintamųjų kategorijas. Gautoms dalinėms imtims taikomi vienodi statistiniai testai, jų rezultatai palyginami tarpusavyje.
- **Rangų** skalė. Joje nustatoma objekto (reiškinio) vieta pagal pasirinktą kiekybinį arba kokybinį požymį vienos rūšies objektų (reiškinų) grupėje. Pavyzdžiui, sportininko užimta vieta varžybose, studentų sportinis aktyvumas (pvz.: 1=nesportuoja, 2=sportuoja retkarčiais, 3=sportuoja intensyviai), automobilio klasė (pvz.: 1=aukščiausios klasės, 2=vidutinės klasės, 3=žemiausios klasės) ir t. t. Šiai skalei priklauso taip pat kintamieji, gauti grupuojant duomenis, pvz. — pagal pinigines įplaukas.

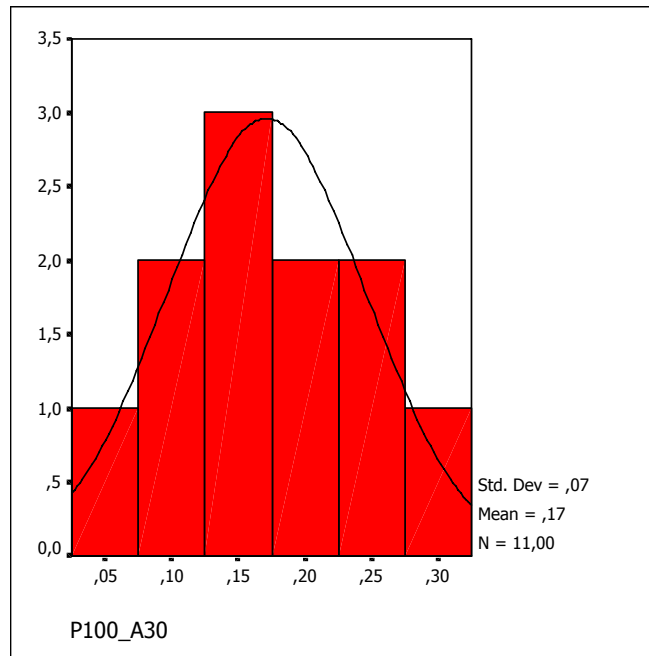
Su matuojamais pagal rangų skalę kintamaisiais (**ranginiais** kintamaisiais) galima atlikti gerokai daugiau statistinių operacijų negu su nominaliaisiais kintamaisiais. Be dažnių įvertinimo, galima apskaičiuoti medianą, rangų koreliacijos koeficientą, palyginti atskiras imtis naudojant neparametrinius testus.

- **Intervalų** skalė. Šioje skalėje nurodomi kiekybiniai kintamųjų reikšmių skirtumai, išreikšti matavimo vienetais (metrais, sekundėmis, laipsniais, taškais ir pan.). Šie skirtumai gali būti tarp atskirų intervalų arba nuo kurio nors pasirinkto atskaitos taško, t. y. nulinė reikšmė dar nereiškia, kad tiriamasis požymis visai nepasireiškia, o tik tai, kad jis nesiskiria nuo sąlyginio atskaitos nulio. Duomenis intervalų skalėje (**intervalinius** kintamuosius) galima apdoroti visais be apribojimų statistikos metodais.
- **Santykių** skalė. Ši skalė skiriasi nuo intervalų skalės tik tuo, kad joje nulinis taškas yra griežtai apibrėžtas ir visiškai atitinka dydžio nebuvimą. SPSS taikymo praktikoje skirtumas tarp kintamųjų, matuojamų pagal intervalų arba santykių skalę, yra neesminis. Todėl toliau nagrinėsime matuojamus pagal intervalų skalę (intervalinius) kintamuosius.

7.1.2. Pasiskirstymo pagal normalųjį skirstinį patikrinimas

Daugelis metodų, kuriais apdorojami pagal intervalų skalę matuojami kintamieji, remiasi hipoteze, kad šių kintamųjų reikšmės pasiskirsčiusios pagal normalųjį dėsnį. Pirminį supratimą apie tiriamo kintamojo reikšmių pasiskirstymą galima gauti nurodžius komandas **Graphs → Histogramm...** ir dialogo langelyje nurodžius **Display Normal Curve** (parodyti normalųjį skirstinį). Gautoje histogramoje bus pateikta taip pat normaliojo skirstinio kreivė (7.1 pav.).

Praktikoje paprastai nepasitaiko imčių, kurių duomenys būtų griežtai pasiskirstę pagal normalųjį skirstinį. Todėl būtina nustatyti, ar galima stebimą skirstinį laikyti normaliuoju ir taikyti analizės metodą, kuris remiasi prielaida, kad duomenys pasiskirstę pagal normalųjį skirstinį, pvz., Stjudento t-testą. Be vizualaus realios histogramos palyginimo su normaliuoju skirstiniu, kuris gali pasirodyti nepakankamai griežtas, galima pritaikyti Kolmogorovo ir Smirnov testą, kuris vykdomas nurodant komandas **Analyze → Nonparametric Tests → 1- Sample K-S...** (žr. 12 skyrių) arba modifikuotą Kolmogorovo ir Smirnov testą, kuris vykdomas nurodant komandas **Analyze → Descriptive Statistics → Explore...** (žr. 9 skyrių). Jeigu imties dydis mažesnis kaip 50 stebėjimų, kartu su pastaruoju vykdomas taip pat Shapiro ir Wilk'o testas.



7.1 pav. Histograma su normaliojo skirstinio kreive

7.1.3. Priklausomos ir nepriklausomos imtys

Dvi imtys priklauso viena nuo kitos, jeigu kiekvienai vienos imties reikšmei galima dėsningai ir vienareikšmiškai nurodyti vieną atitinkamą kitos imties reikšmę. Panašiai apibrėžiama ir kelių imčių priklausomybė.

Dažniausiai pasitaiko priklausomos imtys, kai matavimai atliekami tam tikrais laiko momentais. Tada priklausomos imtys sudaro tiriamo vyksmo parametrų reikšmes skirtingais laiko momentais.

SPSS programoje priklausomos imtys pateikiamos kaip skirtingi kintamieji, kurių atžvilgiu atliekamas atitinkamas testas toje pačioje stebėjimų visumoje. Nepriklausomos imtys, t. y. imtys, kurioms negalima nustatyti dėsningo ir vienareikšmio atitikimo, gali turėti skirtingus stebėjimus, kuriuos paprastai skiria kategorinis vardinės skalės kintamasis.

7.2. STATISTINIS KRITERIJUS

Viena iš statistikos užduočių yra pateikti metodus, kuriais būtų galima objektyviai įvertinti, ar gauti rezultatai nėra atsitiktiniai. Pavyzdžiui, lyginant dviejų imčių vidurkius, galima suformuluoti dvi hipotezes:

- Hipotezė H_0 (nulinė). Nėra skirtumo tarp populiacijos vidurkių. Gautas imčių vidurkių skirtumas atsitiktinis ir neesminis.
- Hipotezė H_1 (alternatyvioji). Tarp vidurkių yra esminis skirtumas. Jo negalima paaiškinti atsitiktiniais nukrypimais.

Hipotezėms tikrinti yra naudojami įvairūs kriterijai. Statistinis kriterijus – tai taisyklė, pagal kurią, remiantis imties duomenimis, hipotezė H_0 priimama arba atmetama. Nesigilindami (plačiau galima rasti: Čekanavičius, Murauskas, 2000 ir Gonestas, Strielčiūnas, 2003) pateiksime statistinio kriterijaus reikšmingumo lygmens, pagal kurį yra atmetama nulinė hipotezė H_0 ir priimama alternatyvioji hipotezė H_1 , sąvoką. Kriterijaus reikšmingumo lygmuo α – tai klaidos atmetus hipotezę H_0 , nors iš tikrųjų ji teisinga, tikimybė: $\alpha = P(H_0 \text{ atmetama} \mid H_0 \text{ teisinga})$.

Dar ši tikimybė vadinama pirmosios rūšies klaida. Tradiciškai reikšmingumo lygmenys yra $\alpha = 0,1$; $\alpha = 0,05$ ir $\alpha = 0,01$.

Tikimybė, kad kriterijaus statistika T (tuo atveju, kai H_0 teisinga) ne mažesnė už stebimą statistikos T realizaciją t^* , vadinama p -reikšme.

$$p = P(T \geq t^*), \text{ kai } H_0 \text{ teisinga.}$$

Bendra taisyklė, tinkanti visų rūšių (su vienu ar dviu alternatyva) nulinėms hipotezėms H_0 ir alternatyvoms H_1 formuluojama taip:

Tegul α yra reikšmingumo lygmuo, o p — p -reikšmė.

Jeigu $p < \alpha$, tai hipotezė H_0 atmetama.

Jeigu $p \geq \alpha$, tai hipotezė H_0 neatmetama.

7.3. STATISTIKOS METODAI

SPSS paketas apima daugelį statistinės duomenų analizės metodų, taikomų įvairiose srityse. Šioje knygoje skirsime dėmesį tik tiems metodams, kurie galėtų būti naudingi sportiniams ir giminingų sričių (biomechanikos, fiziologijos ir kt.) tyrimams.

7.3.1. Aprašomoji statistika

Aprašomoji statistika (angliškai *descriptive statistics*) — tai statistinių duomenų skaitiniai ir grafiniai pateikimo metodai: dažnių lentelės, statistinių charakteristikų skaičiavimas, didelė grafikų įvairovė.

Dažnių lentelės yra sudaromos nominaliesiems ir ranginiams kintamiesiems (žiūr. 8 skyrių).

Ranginių kintamųjų ir pagal intervalų skalę matuojamų kintamųjų, kurie nėra pasiskirstę pagal normalųjį skirstinį, dažniausiai skaičiuojama mediana ir kvartilai.

Matuojamų pagal intervalų skalę kintamųjų, kurių reikšmės pasiskirsčiusios pagal normalųjį skirstinį, dažniausiai skaičiuojamas vidurkis (angl. *mean*) ir sklaidos charakteristikos — standartinis nuokrypis (angl.

standart deviation) arba standartinė paklaida (angl. *standart error mean*, sutrumpintai — *S. E. mean*).

7.3.2. Analitinė statistika

Statistinė analizė neapsiriboja vien tik aprašomojo pobūdžio operacijomis, bet apima vienokius ar kitokius analitinius metodus, kuriais remiantis priimamos arba atmetamos hipotezės pasirinktu reikšmingumo lygmeniu.

Didelė testų dalis yra skirta nustatyti, ar skiriasi dviejų arba daugiau populiacijų požymiai — vidurkiai arba medianos. Priklausomai nuo imčių kiekio, kintamųjų skalės (intervalų ar rangų), nuo to, ar imtys priklausomos ar nepriklausomos, ar kintamieji turi normalųjį skirstinį, taikomi skirtingi testai, pateikti 7.1 lentelėje.

Kiekvienai iš dviejų testų grupių SPSS numatyti atskiri meniu punktai. Pirmos grupės testai vykdomi nurodžius komandas *Analyze → Compare Means* (vidurkių palyginimas), antros grupės *Analyze → Nonparametric Tests* (neparametriniai testai).

Išimčių sudaro blokuotų duomenų (Čekanavičius, Murauskas, 2002) dispersinė analizė. Šis metodas nurodomas komandomis *Analyze → General Linear Model* (bendras tiesinis modelis).

7.1 lentelė. Hipotezių apie vidurkių lygybę tikrinimo metodai

<i>Intervaliniai kintamieji, turintys normalųjį skirstinį</i>	
Problema	Testas
Dviejų nepriklausomų imčių lyginimas	Stjudento t-testas
Dviejų priklausomų imčių lyginimas	Porinis Stjudento t-testas
Trijų ir daugiau nepriklausomų imčių lyginimas	Paprasta dispersinė analizė ANOVA
Trijų ir daugiau priklausomų imčių lyginimas	Blokuotųjų duomenų dispersinė analizė
<i>Ranginiai kintamieji ir intervaliniai kintamieji, neturintys normaliojo skirstinio</i>	
Problema	Testas
Dviejų nepriklausomų imčių lyginimas	Mann'o ir Whitney U-testas
Dviejų priklausomų imčių lyginimas	Wilcoxon'o testas
Trijų ir daugiau nepriklausomų imčių lyginimas	Kruskal'o ir Wallis H-testas
Trijų ir daugiau priklausomų imčių lyginimas	Friedman'o testas

Dažnai tenka lyginti skirtingas nominaliųjų kintamųjų arba ranginių (su nedideliu kategorijų skaičiumi) kintamųjų stebėjimų arba reikšmių grupes. Tokiu atveju sudaromos požymių dažnių lentelės (angl. *crosstabs*) (žr. 10 skyrių).

Koreliaciniai (žr. 13 skyrių) ir regresiniai (žr. 14 skyrių) metodai skirti ryšiui tarp dviejų kintamųjų nustatyti ir aprašyti.

Be išvardintųjų gana paprastų statistikos metodų, vertėtų paminėti sudėtingesnius daugiamatės analizės metodus, kuriais vienu metu analizuojama daugelis kintamųjų. Tai faktorinė analizė, kuri padeda didelio skaičiaus kintamųjų tarpusavio koreliacijas paaiškinti tam tikrų bendrųjų faktorių įtaka, daugiafaktorinė dispersinė analizė (žr. 15 skyrių). Taikant klasterinę analizę, nustatomas objektų panašumas ir jie suskirstomi į klasterius taip, kad skirtumai klasterių viduje būtų kuo mažesni, o tarp klasterių — kuo didesni (žr. 16 skyrių).

8. DAŽNIŲ ANALIZĖ

Dažnių analizės rezultatai ar grafinė medžiaga yra pakankami apibūdinti daugeliui kintamųjų ir yra pirmas duomenų statistinės analizės etapas.

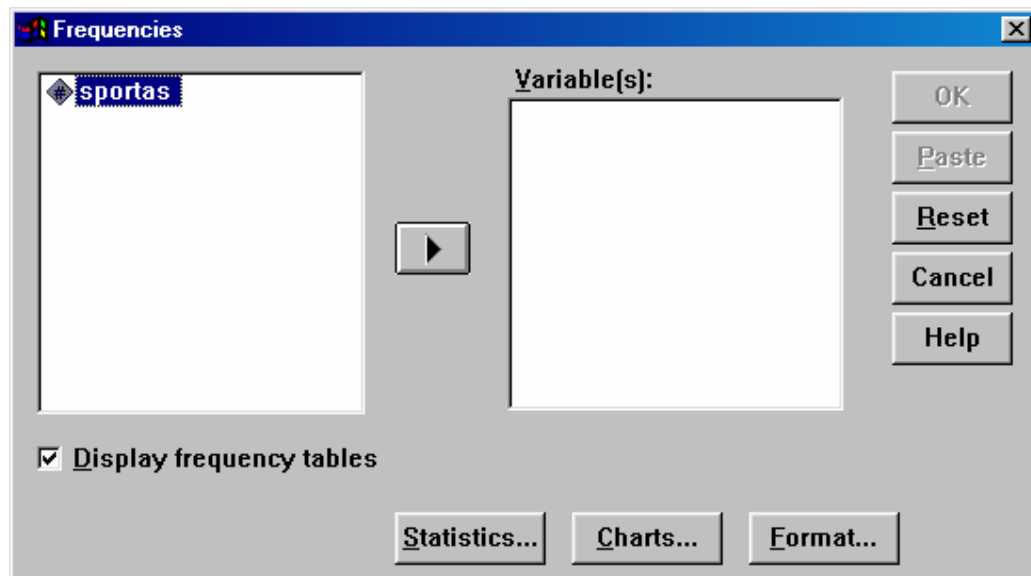
Kategoriniams kintamiesiems (matuojamiems pagal vardinę ar rangų skalę) koduoti naudojami skaitmenys arba trumpos raidinės-skaitmeninės sekos.

Dauguma statistinių duomenų skaitinių charakteristikų, tokių kaip vidurkis arba standartinis nuokrypis, remiasi normaliojo skirstinio prielaida ir tinka kiekybiniam kintamiesiems su simetriniu pasiskirstymu, o kitos (*robust statistics*) — mediana, kvartiliai ir procentiliai gali būti skaičiuojamos kiekybiniam kintamiesiems, nebūtinai turintiems normalųjį skirstinį.

8.1. DAŽNIŲ LENTELĖS

Tiriamo kintamojo (tiriamų kintamųjų) dažnių lentelei gauti:

- Atidarykite bylą su analizuojamais duomenimis.
- Nurodykite komandas *Analyze → Descriptive Statistics → Frequencies...*
- Dialogo langelyje *Frequencies* (8.1 pav.) pažymėkite analizuojamą kintamąjį ir spragtelėję mygtuką su trikampio ženklu įkelkite jį į analizuojamų kintamųjų sąrašą *Variable(s)*.



8.1 pav. Dialogo langelis *Frequencies*

- Įsitikinkite, kad pažymėtas laukelis *Display frequency tables* (rodyti dažnių lenteles) ir spragtelėkite mygtuką *OK*. Rezultatai bus pateikti rodinio (*Viewer*) lange.

8.2 pav. pateiktas elementarus tokios dažnių lentelės pavyzdys. Kintamasis *Sportas* vaizduoja grupės studentų atsakymų į anketos klausimus apie jų sportinį aktyvumą rezultatus. Kintamojo reikšmės ir jų žymenos: 1 — „nesportuoja visai“, 2 — „retkarčiais“, 3 — „dažnai“, 4 — „sportuoja reguliariai“.

SPORTAS					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	nesportuoja visai	4	11,4	11,4	11,4
	retkarčiais	16	45,7	45,7	57,1
	dažnai	10	28,6	28,6	85,7
	sportuoja reguliariai	5	14,3	14,3	100,0
	Total	35	100,0	100,0	

8.2 pav. Dažnių lentelės išvestis

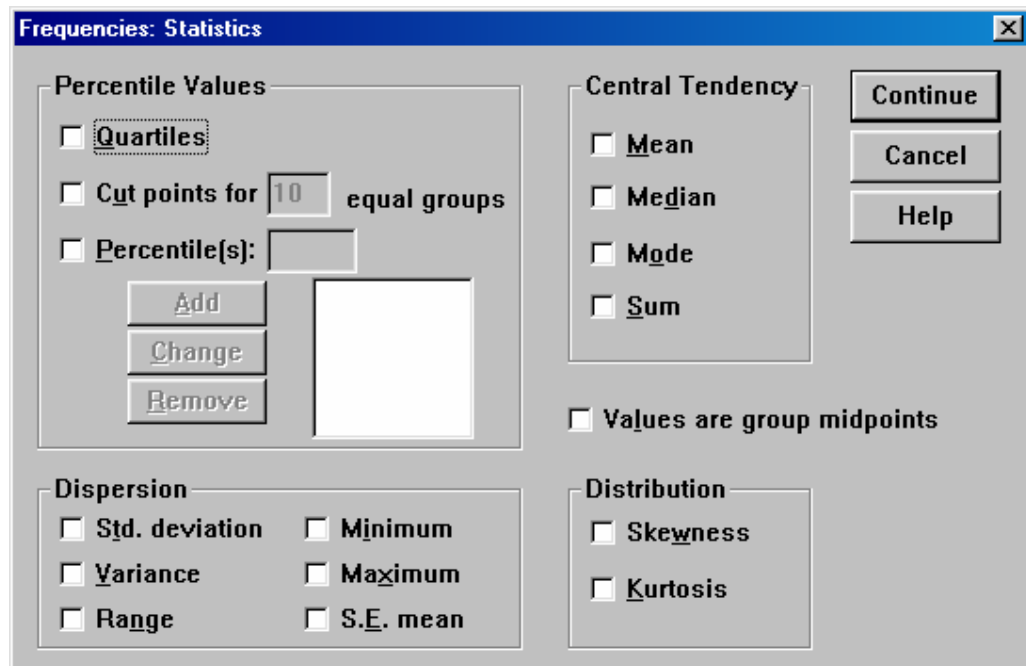
Kiekviena dažnių lentelės eilutė aprašo vieną kintamojo reikšmę. Pirmame stulpelyje pateikiamos visų galiojančių kintamojo reikšmių žymenos. Antrame stulpelyje *Frequency* (dažnis) pateiktas kiekvieno atsakymų į anketos klausimus varianto dažnis, t. y. kiek studentų atsakė, kad „nesportuoja visai“, kiek — „retkarčiais“ ir t. t. Trečiajame stulpelyje *Percent* pateiktas kiekvieno atsakymo procentinis dažnis, kuris gaunamas kiekvieno atsakymo variantų skaičių dalijant iš stebėjimų skaičiaus, t. y. bendro apklausiamų studentų skaičiaus. Ketvirtame stulpelyje *Valid Percent* pateikiamas atsakiusių procentinis dažnis. Kai nėra praleistų duomenų trečio ir ketvirto stulpelių reikšmės sutampa. Penktajame stulpelyje *Cumulative Percent* pateikiamas sukauptas dažnis, kuris gaunamas sumuojant prieš tai buvusias reikšmes. Pavyzdžiui, bendras respondentų, atsakiusių, kad nesportuoja visai ar sportuoja retkarčiais, procentas yra 57,1%.

8.2. SKAITINĖS DUOMENŲ CHARAKTERISTIKOS

Intervalinių kintamųjų skaitinėms charakteristikoms gauti spragtelėkite dialogo langelio *Frequencies* mygtuką *Statistics...* Atsidarys dialogo langelis *Frequencies: Statistics* (8.3 pav.).

Grupėje *Percentile Values* (procentilių reikšmės) galite pasirinkti šiuos variantus:

- **Quartiles** (kvartiliai): Bus pateikti pirmasis, antrasis ir trečiasis kvartiliai. Pirmasis kvartilis (Q_1) — tai taškas surikiuotų didėjimo tvarka duomenų skalėje, žemiau (kairėje) kurio yra ketvirtis (25%) išmatuotų reikšmių. Antrasis kvartilis (Q_2) atskiria du ketvirčius (50%), o trečiasis (Q_3) — tris ketvirčius (75%) išmatuotų reikšmių. Antrasis kvartilis vadinamas mediana.



8.3 pav. Dialogo langelis *Frequencies: Statistics* skaičiuojamoms charakteristikoms pasirinkti

- **Cut points** (dalijimo taškai). Bus pateiktos procentilių, dalijančių surikiuotos didėjimo tvarka imties duomenis į lygias, t. y. turinčias vienodą duomenų skaičių, grupes, reikšmės. Nurodykite grupių skaičių (pirminis nustatytas grupių skaičius — 10).
- **Percentile(s)**: Čia galite nurodyti procentilių reikšmės savo nuožiūra. Laukelyje įrašykite pirmą procentilio reikšmę, ribojamą nuo 0 iki 100, ir spragtelėkite mygtuką **Add**. Pakartokite šiuos veiksmus visoms kitoms procentilių reikšmėms.

Grupėje **Central Tendency** galite pasirinkti šias poslinkio charakteristikas, apibūdinančias centrinės kintamųjų tendencijas: **Mean** (vidurkį), **Median** (mediana), **Mode** (modą), **Sum** (sumą) — visų reikšmių sumą.

Grupėje **Dispersion** galite pasirinkti šias sklaidos charakteristikas, apibūdinančias duomenų išsibarstymą: **Std. Deviation** (standartinis nuokrypis), **Variance** (dispersija), **Range** (plotis), **Minimum** (mažiausia reikšmė), **Maximum** (didžiausia reikšmė), **S.E. mean** (standartinė paklaida).

Paprastai turinčių normalųjų skirstinių intervalinių kintamųjų sklaidos matas yra standartinis nuokrypis arba standartinė paklaida. Standartinio nuokrypio intervale, t. y. intervale, apimančiame standartinio nuokrypio plotį į abi puses nuo vidurkio, yra apie 67% tokio kintamojo reikšmių, dvigubos standartinio nuokrypio reikšmės intervale — apie 95% reikšmių ir trigubos

standartinio nuokrypio reikšmės intervale — apie 99% reikšmių („trijų σ “ taisyklė). Standartinė paklaida leidžia įvertinti populiacijos vidurkio pasikliautinąjį intervalą. Su 95% pasikliovimo lygmeniu galima teigti, kad populiacijos vidurkis yra dvigubos standartinės paklaidos reikšmės intervale į abi puses nuo imties vidurkio, o su 99% pasikliovimo lygmeniu galima teigti, kad populiacijos vidurkis yra trigubos standartinės paklaidos reikšmės intervale į abi puses nuo imties vidurkio.

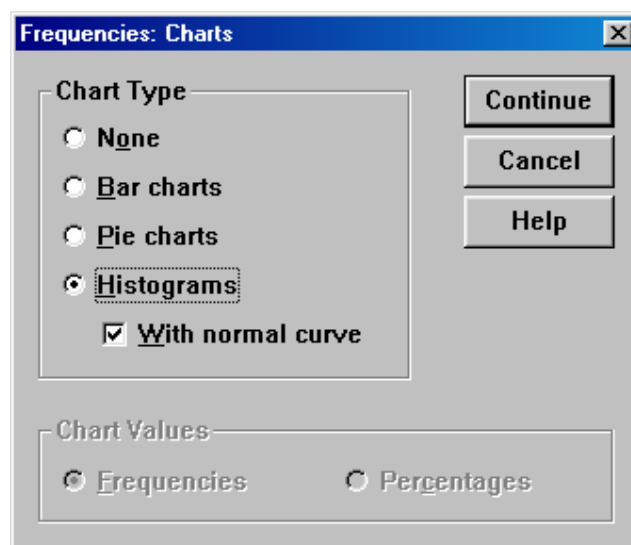
Grupėje **Distribution** (pasiskirstymas) galite pasirinkti šias pasiskirstymo nesimetriškumo charakteristikas: **Skewness** (asimetrijos koeficientas), **Kurtosis** (ekscesas). Normaliojo skirstinio asimetrijos ir eksceso koeficientai yra lygūs nuliui.

Paprastai intervalinių kintamųjų su normaliuoju skirstiniu pagrindinė (poslinkio) charakteristika yra vidurkis, o pagrindinė sklaidos charakteristika — standartinis nuokrypis arba standartinė paklaida. Ranginių kintamųjų arba intervalinių kintamųjų, kurie nėra normaliai pasiskirstę — atitinkamai mediana, taip pat pirmas ir trečias kvartiliai. Nominaliųjų kintamųjų galima tik viena charakteristika — moda.

Pažymėję varnele žymimuosius pasirinktų charakteristikų langelius spragtelėkite mygtuką **Continue** (tęsti).

8.3. DAŽNIŲ DIAGRAMOS

Dažnių analizės rezultatams pavaizduoti grafiškai spragtelėkite dialogo langelio **Frequencies** mygtuką **Charts...** (diagramos). Atsidarys dialogo langelis **Frequencies: Charts** (8.4 pav).



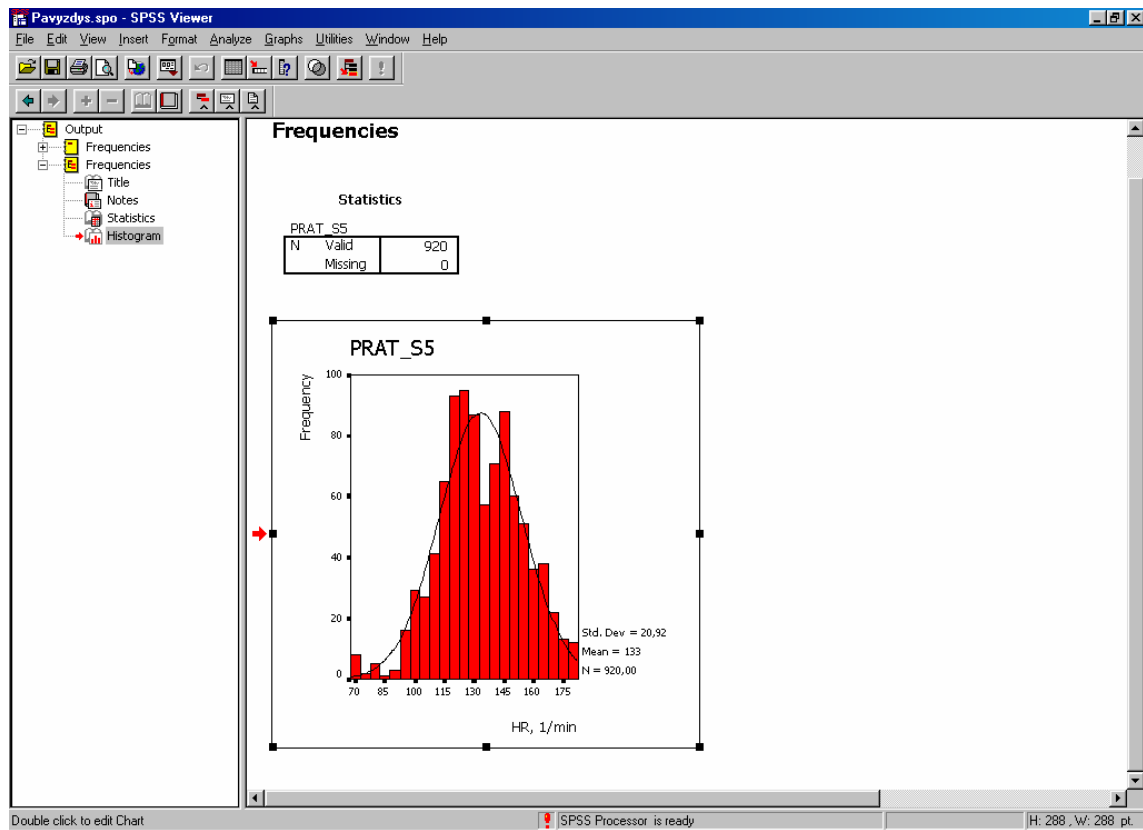
8.4 pav. Dialogo langelis **Frequencies: Charts**

Grupėje **Chart Type** (diagramos tipas) galite pasirinkti diagramos tipą: **Bar Charts** (stulpelinės diagramos), **Pie Charts** (skritulinės diagramos), **Histograms** (histogramos). Pasirinkę stulpelinę diagramą, galite nurodyti, kokie dydžiai turi būti atidėti Y ašyje: **Frequencies** (dažniai) ar **Percentages** (procentai). Pasirinkę histogramą, galite pažymėti laukelį **With normal curve** (su normaliojo skirstinio kreive).

Dialoginių diagramų kūrimui ir redagavimui yra skirtas 17 skyrius. Šiame skyriuje nagrinėjami standartiniai grafikai.

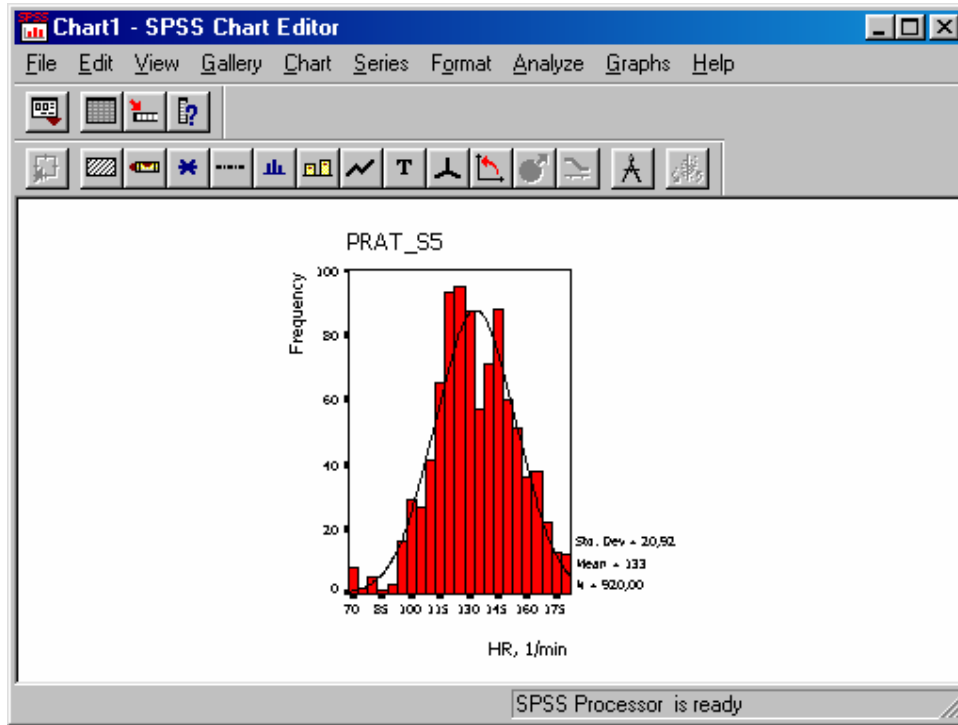
Pasirinkę diagramos tipą, spragtelėkite dialogo langelio **Frequencies: Charts** mygtuką **Continue**, tada — dialogo langelio **Frequencies** mygtuką **OK**. Rodinio (**Viewer**) lange pasirodys Jūsų pasirinkta diagrama.

Iliustracijai 8.5 pav. pavaizduotas **Viewer** langas su krepšininco širdies susitraukimų dažnio (angl. *heart rate* — HR), matuoto reguliariais laiko tarpais per pratybas, histograma, kurioje taip pat pavaizduota normaliojo skirstinio kreivė (histograma pažymėta).



8.5 pav. Histograma **Viewer** lange

Jeigu norite redaguoti diagramą pagal savo poreikius, dukart spragtelėkite pele bet kurioje diagramos vietoje. Diagrama atsidurs diagramų redaktoriaus lange (8.6 pav.).



8.6 pav. Diagramų redaktoriaus langas **Chart Editor**

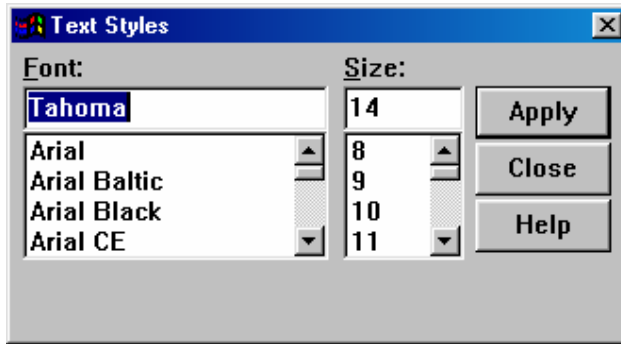
Atkreipkite dėmesį, kad pasikeičia meniu ir įrankių juostos — vyrauja diagramų redagavimo komandos ir įrankiai. Jūs galite keisti kiekvieną diagramos komponentą, taip pat pasirinkti kitą diagramos tipą. Dažniausiai redaguojami:

- Diagramos ir ašių pavadinimai.
- Legenda.
- Diagramos tinklėlis ir ašių formatai.
- Duomenų sekų formatai.
- Pastabų tekstas.
- Vidiniai ir išoriniai rėmeliai.

Padaryti diagramų pakeitimai išsaugomi, kai uždarote diagramų redaktoriaus langą.

8.3.1. Kaip tvarkyti teksto formatus?

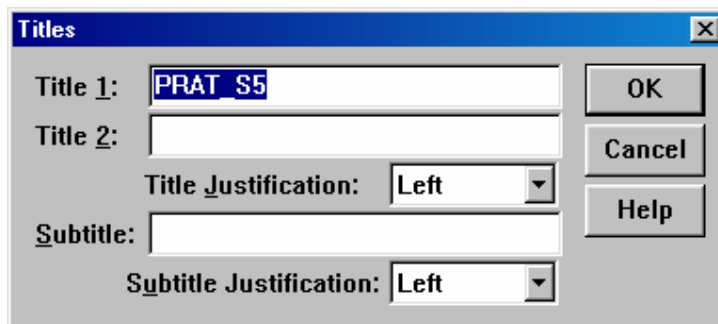
- Spragtelėkite pasirinktą tekstą (diagramos pavadinimą, ašių pavadinimus ir t. t.) — tekstas bus pažymėtas stačiakampiu rėmeliu.
- Nurodykite diagramų redaktoriaus lange komandas **Format → Text...** arba spragtelėkite įrankio **T** mygtuką.
- Dialogo langelyje **Text Styles** (8.7 pav.) galite pasirinkti šriftą ir jo dydį.



8.7 pav. Dialogo langelis **Text Styles** teksto formatui nustatyti

8.3.2. Kaip redaguoti diagramos pavadinimą?

- Nurodykite diagramų redaktoriaus lange komandas **Charts** → **Title...** arba dukart spragtelėkite diagramos pavadinimą.
- Dialogo langelyje **Titles** (8.8 pav.), laukelyje **Title 1 (2)** galite pakeisti diagramos pavadinimą, laukelyje **Subtitle** įrašyti paantraštę ir pakeisti užrašų vietą vaizdo atžvilgiu.

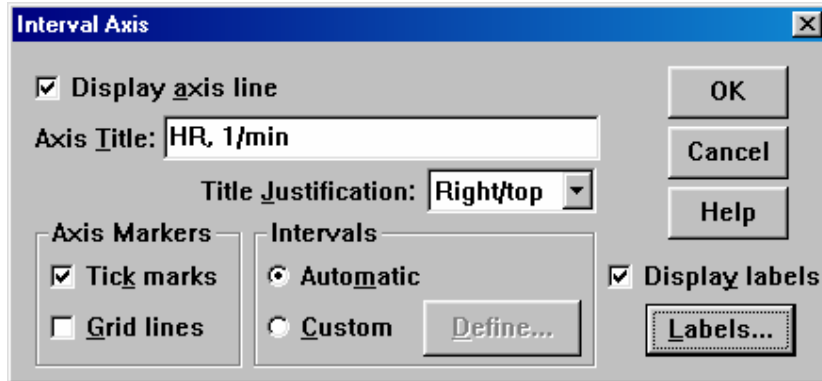


8.8 pav. Dialogo langelis **Titles** diagramos pavadinimui redaguoti

8.3.3. Kaip tvarkyti diagramos ašių formatus?

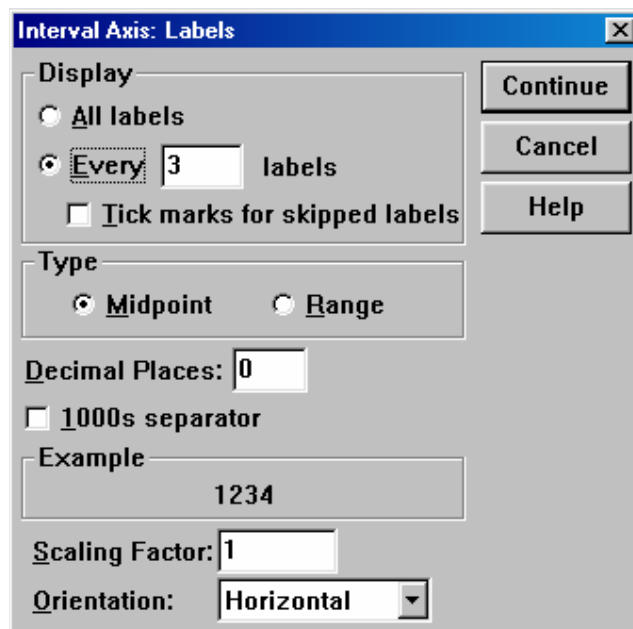
- Diagramų redaktoriaus lange pažymėkite (spragtelėkite) ašį, kurios formatus norite tvarkyti ir nurodykite komandas **Charts** → **Axis...** arba dukart spragtelėkite reikiamą ašį.
- Jei pasirinkote X ašį, dialogo langelyje **Interval Axis** (8.9 pav., jei tvarkoma histograma) galite:
 - Redaguoti ašies pavadinimą (**Axis Title**).
 - Nustatyti pavadinimo teksto vietą (**Title Justification**).
 - Nustatyti papildomą tinklę **Grid lines** — linijas, brėžiamas nuo kiekvienos ašies padalų (**Tick marks**) per visą braižymo sritį.

- Pažymėti grupėje *Intervals* vietoje *Automatic* – *Custom* (individualus formatas), spragtelėti mygtuką *Define...* (nustatyti) ir atidaryti naują dialogo langelį *Interval Axis: Define Custom Intervals* nustatyti pageidaujamą histogramos intervalų kiekį arba plotį.



8.9 pav. Dialogo langelis *Interval Axis* X ašies formatams tvarkyti

- Spragtelėti mygtuką *Labels...* ir dialogo langelyje *Interval Axis: Labels* (8.10 pav.) nustatyti, kas kiek padalų rodyti žymę, žymės tipą, kiek skaitmenų rodyti po kablelio (*Decimal Places*), mastelio koeficientą (*Scaling Factor*), žymės rašymo kryptį (*Orientation*).




8.10 pav. Dialogo langelis *Interval Axis: Labels* padalų žymių formatams tvarkyti

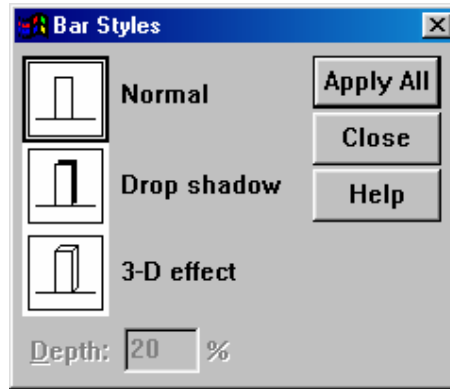
- Jei pasirinkote Y ašį, dialogo langelyje *Scale Axis* (8.11 pav., jei tvarkoma histograma) galite:
 - Redaguoti ašies pavadinimą (*Axis Title*).
 - Nustatyti ašies pavadinimo teksto vietą (*Title Justification*).
 - Nustatyti linijinę (*Linear*) arba logaritminę (*Log*) skalę.
 - Nustatyti mažiausią (*Minimum*) ir didžiausią (*Maximum*) reikšmių skalės rodomas reikšmes.
 - Nustatyti stambių padalų (*Major Divisions*) ir smulkių padalų (*Minor Divisions*) reikšmes, taip pat, kokias padalas ir tinklę rodyti.
 - Spragtelėti mygtuką *Labels...* ir dialogo langelyje *Scale Axis: Labels* nustatyti, kiek skaitmenų rodyti po kablelio (*Decimal Places*), mastelio koeficientą (*Scaling Factor*).

8.11 pav. Dialogo langelis *Scale Axis* Y ašies formatams tvarkyti

8.3.4. Kaip tvarkyti duomenų sekų formatus?

Stulpelinės diagramos stulpelių tipui pakeisti:

- Nurodykite diagramų redaktoriaus lange komandas *Format → Bar Style...* arba spragtelėkite įrankio  mygtuką.
- Dialogo langelyje (8.12 pav.) pasirinkite stulpelių tipą: *Normal* (normalų) — tai numatytasis variantas, *Drop shadow* (su šešėliu), *3D effect* (erdvinį). Pasirinkę erdvinį variantą, laukelyje *Depth* (gylis) nurodykite skaičių erdvinės diagramos gylio išpildžiui išreikšti.
- Spragtelėkite mygtuką *Apply All* (pritaikyti visiems), tada — *Close*.



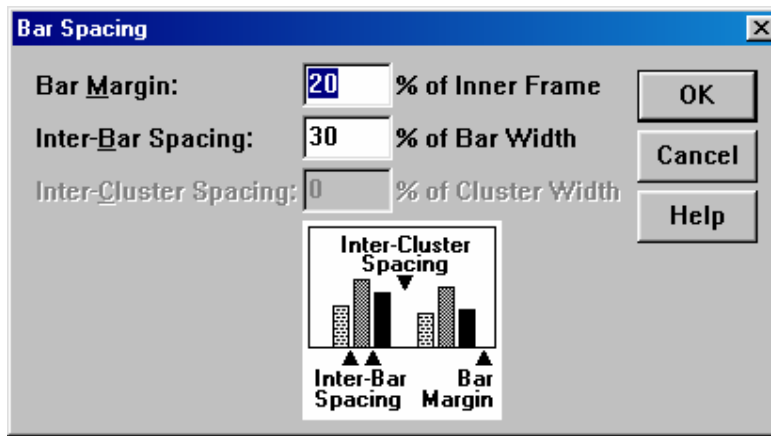
8.12 pav. Dialogo langelis **Bar Styles** stulpelinės diagramos stulpelių tipui parinkti

Diagramos stulpelių išdėstymui pakeisti:

- Nurodykite diagramų redaktoriaus lange komandas **Charts → Bar Spacing...**


Stulpelinės diagramos dialogo langelio **Bar Spacing** (8.13 pav.) laukelyje **Bar Margin** galite pakeisti nustatytą kraštinių stulpelių padėtį (procentais nuo vidinio rėmelio pločio) ir **Inter-Bar Spacing** laukelyje — tarpą tarp stulpelių (procentais nuo stulpelio pločio).

Histogramos dialogo langelis **Bar Spacing** turi tik laukelį **Bar Margin**.



8.13 pav. Dialogo langelis **Bar Spacing** stulpelių išdėstymui parinkti

Duomenų sekos spalvoms pakeisti:


- Spragtelėkite seką diagramų redaktoriaus lange ir nurodykite komandas **Format → Color...** arba spragtelėkite įrankio  mygtuką.
- Dialogo langelyje (8.14 pav.) spragtelėkite pele reikiamą pasirinkties mygtuką: **Fill** (užpildyti) arba **Border** (rėmeliai) ir pasirinkite spalvą. Daugiau spalvų galite rasti spragtelėję mygtuką **Edit...**

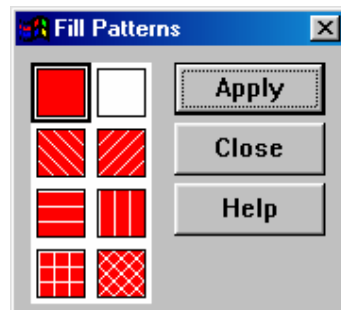
- Spragtelėkite mygtuką *Apply*.



8.14 pav. Dialogo langelis *Colors* duomenų sekos spalvai parinkti

Duomenų sekos raštui nustatyti:

- Spragtelėkite seką diagramų redaktoriaus lange ir nurodykite komandas *Format* → *Fill Pattern...* arba spragtelėkite įrankio  mygtuką.
- Dialogo langelyje (8.15 pav.) spragtelėkite pele reikiamą rašto pavyzdį ir mygtuką *Apply*.



8.15 pav. Dialogo langelis *Fill Patterns* duomenų sekos raštui nustatyti

8.3.5. Kaip tvarkyti kitų diagramos komponentų formatus?

Išoriniam diagramos rėmeliui pavaizduoti:

- Diagramų redaktoriaus lange meniu *Chart* pažymėkite varnele (spragtelėkite pele) komandą *Outer Frame* (išorinis rėmelis).

Kitų diagramos komponentų (išnašų, pastabų, legendų) formatų tvarkymui meniu *Chart* komandomis: *Footnote...* (išnašos), *Legend...* (legenda), *Annotation...* (pastabos) atidarykite atitinkamą paprastą dialogo langelį, kuris nurodys tolesnius veiksmus.

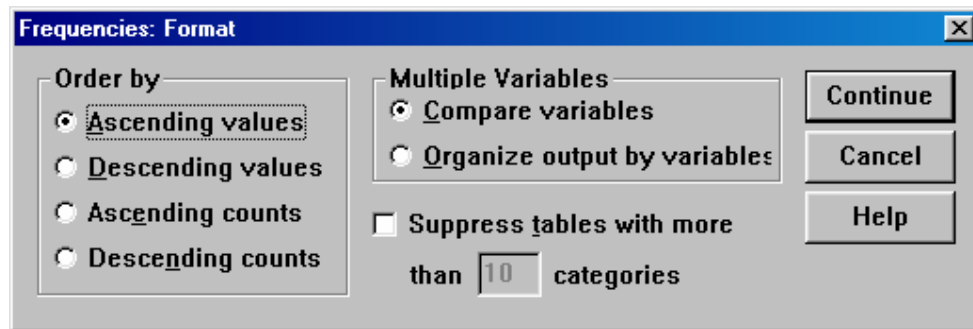
8.4. REZULTATŲ FORMATAI

Nustatytiems rezultatų pateikimo formatams pakeisti spragtelėkite dialogo langelio **Frequencies** mygtuką **Format...** Atsidarys dialogo langelis **Frequencies: Format** (8.16 pav.).

Komandų grupėje **Order by** (rikiavimas pagal) galite surikiuoti dažnių lentelėse pateikiamus rezultatus pagal duomenų dydį didėjančia (**Ascending values**) arba mažėjančia (**Descending values**) tvarka ar pagal duomenų atskirų reikšmių kiekį (t. y. tų reikšmių pasikartojimo dažnį) didėjančia (**Ascending counts**) arba mažėjančia tvarka (**Descending counts**).

Jeigu analizuojate kelis kintamuosius, galite visų kintamųjų rezultatus rodyti vienoje bendroje lentelėje — tam pažymėkite variantą **Compare variables** (palyginti kintamuosius) arba kiekvieną kintamąjį atskiroje lentelėje — tam pažymėkite variantą **Organize output by variables** (išvestis pagal kintamuosius).

Pažymėję komandą **Suppress tables with more than n categories** (uždrausti lenteles, didesnes kaip n kategorijų), apribosite lentelių dydį iki nurodyto kintamųjų skaičiaus.



8.16 pav. Dialogo langelis **Frequencies: Format** rezultatų formatui nustatyti

9. SKAITINĖS CHARAKTERISTIKOS

Skaitinės statistinių duomenų charakteristikos paprastai skaičiuojamos kintamųjų, matuojamų pagal intervalų skalę. SPSS tam tikslui naudojamos keturios meniu komandos:

- **Analyze → Descriptive Statistics** (aprašomoji statistika) → **Frequencies...** (dažniai), žr. 8 skyrių.
- **Analyze → Descriptive Statistics → Descriptives...**
- **Analyze → Descriptive Statistics → Explore...** (tirti).
- **Analyze → Reports** (ataskaitos) → **Case summaries...** (stebėjimų suvestinė).

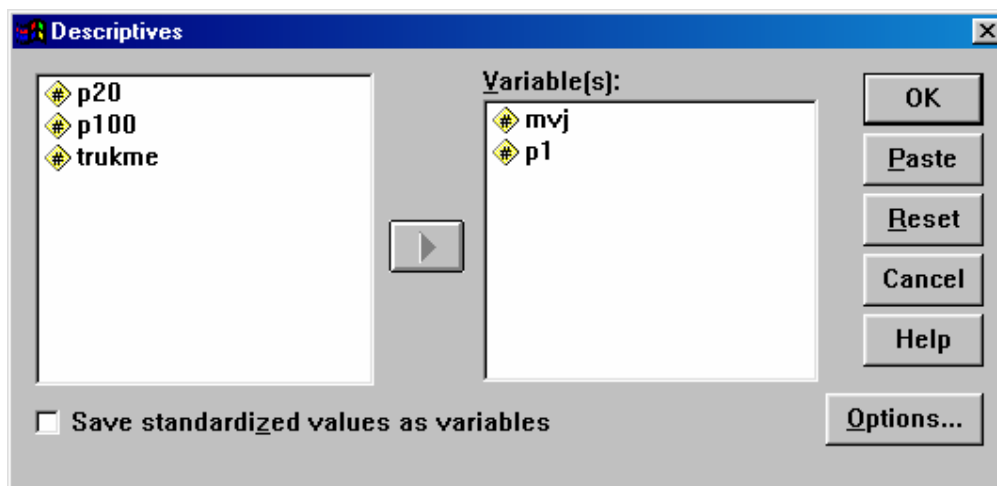
Kiekviena šių komandų turi tam tikrų savitumų.

9.1. APRAŠOMOJI STATISTIKA

Vykdamas komandas **Analyze → Descriptive Statistics → Descriptives...** pateikiamos analizuojamų kintamųjų skaitinės charakteristikos (vienoje lentelėje), taip pat standartizuotos kintamųjų reikšmės (z reikšmės). Dauguma skaitinių charakteristikų remiasi kintamųjų normaliojo skirstinio prielaida.

Tiriamo kintamojo (tiriamų kintamųjų) skaitinėms charakteristikoms apskaičiuoti:

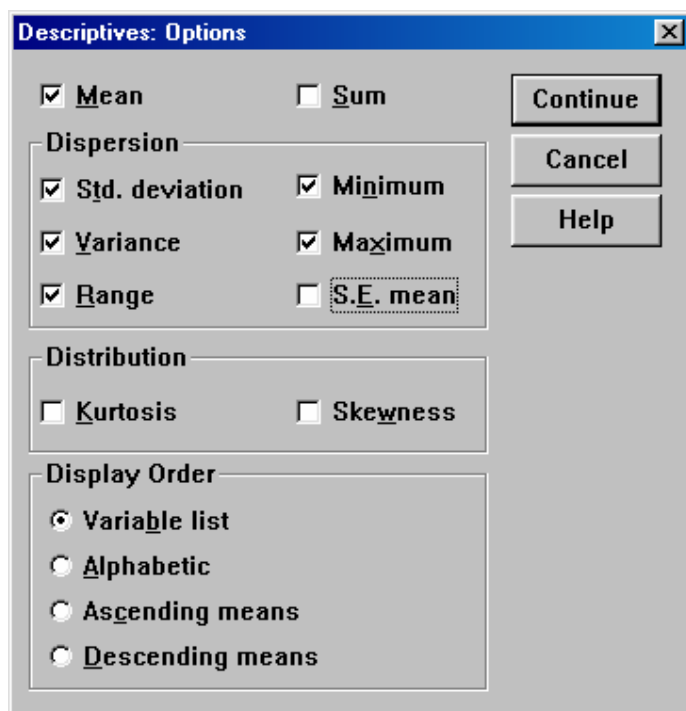
- Atidarykite bylą su analizuojamais duomenimis.
- Nurodykite komandas **Analyze → Descriptive Statistics → Descriptives...**
- Dialogo langelyje **Descriptives** (9.1 pav.) pažymėkite analizuojamą kintamąjį ir spragtelėję mygtuką su trikampio ženklu įkelkite jį į analizuojamų kintamųjų sąrašą **Variable(s)**.



9.1 pav. Dialogo langelis **Descriptives**

- Spragtelėkite mygtuką **Options...**

- Dialogo langelyje **Descriptives: Options** (9.2 pav.) galite pasirinkti šias skaitines kintamųjų charakteristikas:
 - **Mean** (vidurkis) – tai numatytasis variantas.
 - **Std. Deviation** (standartinis nuokrypis).
 - **Variance** (dispersija).
 - **Range** (plotis).
 - **Sum** (suma).
 - **Minimum** (mažiausia reikšmė).
 - **Maximum** (didžiausia reikšmė).
 - **S.E. mean** (standartinė paklaida).
 - **Kurtosis** (ekscesas).
 - **Skewness** (asimetrijos koeficientas).



9.2 pav. Dialogo langelis **Descriptives: Options**

- Jeigu atliekate kelių kintamųjų analizę, **Display Order** (vaizdavimo tvarka) dialogo langelio srityje galite pasirinkti duomenų išvesties eiliškumą:
 - **Variable list** (pagal analizuojamų kintamųjų sąrašą).
 - **Alphabetic** (pagal kintamųjų vardus).
 - **Ascending means** (pagal didėjančius vidurkius).
 - **Descending means** (pagal mažėjančius vidurkius).
- Spragtelėkite mygtuką **Continue**.

- Pažymėkite dialogo langelyje **Descriptives** komandą **Save standardized values as variables**, jeigu norite išsaugoti standartizuotas reikšmes duomenų rinkmenoje kaip atskirus kintamuosius.
- Spragtelėkite mygtuką **OK**.

9.3 pav. pateiktas **Descriptive Statistics** išvesties pavyzdys su dviejų kintamųjų *MVJ* ir *P1* apskaičiuotomis pasirinktų skaitinių charakteristikų reikšmėmis.

Descriptive Statistics							
	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance
MVJ	55	,3650	,2280	,5930	,395800	,0925244	,0086
P1	55	,0630	,0010	,0640	,013444	,0134953	,0002
Valid N (listwise)	55						

9.3 pav. **Descriptive Statistics** išvesties pavyzdys

Kintamųjų standartizacija (z-transformacija) naudojama tais atvejais, kai reikia atlikti veiksmus su kintamaisiais, kurių reikšmės gali skirtis keliomis eilėmis. Z-transformacija transformuoja normalųjį kintamąjį su vidurkiu \bar{X} ir dispersija s^2 į standartizuotą normalųjį kintamąjį su vidurkiu 0 ir dispersija 1, t. y.

$$X \sim N(\bar{X}, s^2) \xrightarrow{Z} N(0,1) \quad (9.1)$$

9.2. STEBĖJIMŲ SUVESTINĖ

Stebėjimų suvestinėje pateikiamos kintamųjų pogrupių (stebėjimų grupių) skaitinės charakteristikos pagal vieno ar kelių kategorinių kintamųjų kategorijas. Tokie kategoriniai kintamieji dažniausiai būna lytis, amžius, sporto šaka, poveikis ir t. t. Pavyzdžiu galėtų būti atskirų Lietuvos rajonų moksleivių sportinių rodiklių statistinių duomenų suvestinė.

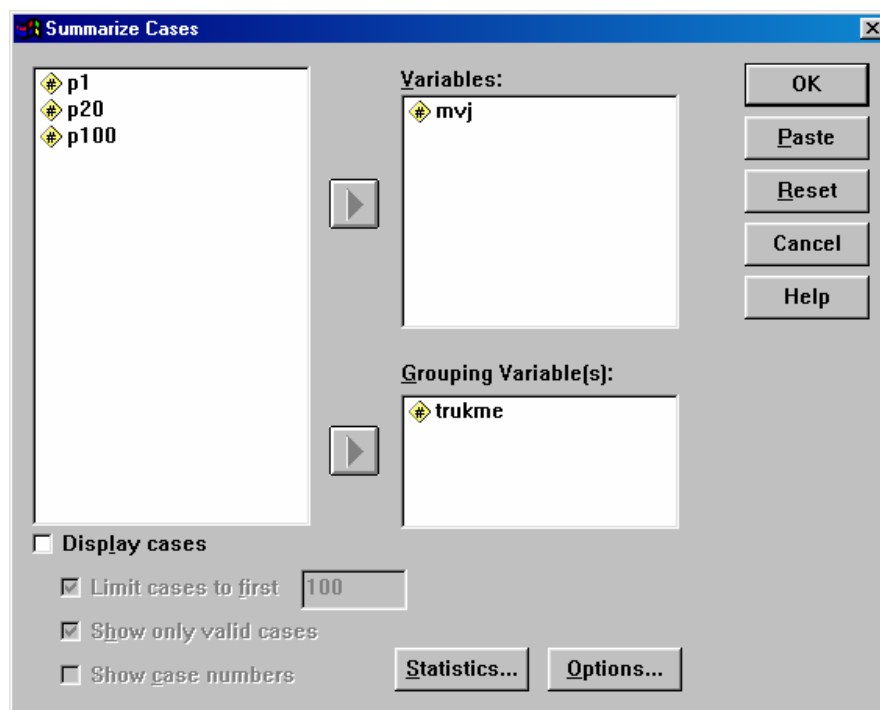
Stebėjimų suvestinėje taip pat pateikiamos suminės (apimančios visas stebėjimų grupes) kiekvieno kintamojo skaitinės charakteristikos, o atskirų kategorijų (stebėjimų grupių) rezultatai gali būti nerodomi.

Kategoriniai kintamieji paprastai yra nominalieji arba ranginiai, jų reikšmės gali būti skaitmenys arba trumpos raidinės-skaitmeninės sekos.

Kai kurios skaitinės charakteristikos, tokios kaip vidurkis arba standartinis nuokrypis, remiasi normaliojo skirstinio prielaida ir tinka kiekybiniais kintamiesiems su simetriniu pasiskirstymu, o kitos (*robust statistics*) — mediana, plotis gali būti skaičiuojamos ir kiekybinių kintamųjų, nebūtinai turinčių normalųjį skirstinį.

Tiriamų kintamųjų skaitinių charakteristikų suvestinei gauti:

- Atidarykite bylą su analizuojamais duomenimis.
- Nurodykite komandas **Analyze → Reports → Case Summaries...**
- Dialogo langelyje **Summarize Cases** (9.4 pav.) įkelkite pasirinktus kintamuosius į analizuojamų kintamųjų sąrašą **Variables**.
- Pasirinkite vieną ar kelis kategorinius kintamuosius ir įkelkite į sąrašą **Grouping Variable(s)**, jeigu norite gauti rezultatus pagal kategorijas.
- Nuimkite **Display cases** varnelę, jeigu nenorite, kad būtų rodomos pasirinktos stebėjimo reikšmės (šias reikšmes galite matyti analizuojamų duomenų byloje).

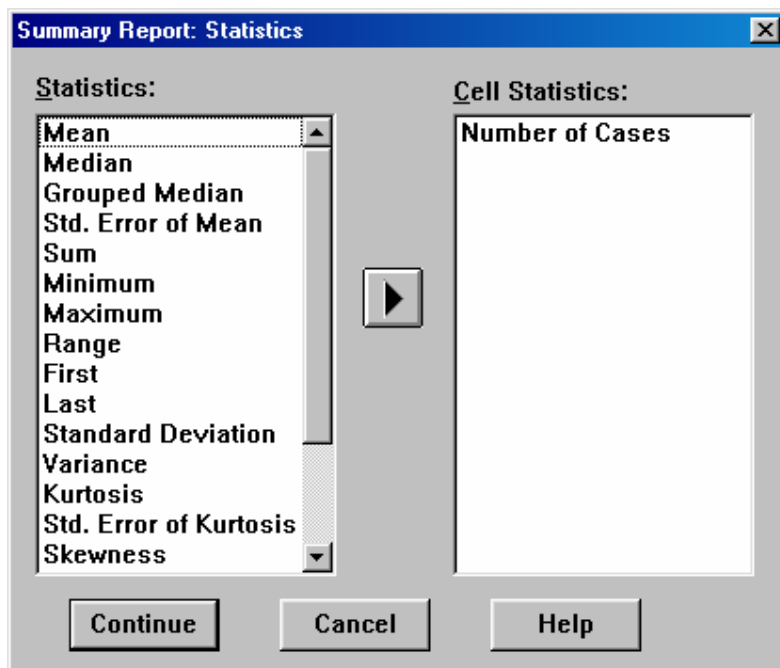


9.4 pav. Dialogo langelis **Summarize Cases**

- Spragtelėkite dialogo langelio mygtuką **Statistics...** Atsidarys naujas dialogo langelis **Summary Report: Statistics** (9.5 pav.), kuriame Jūs galite pasirinkti šias skaitines kintamųjų charakteristikas: **Number of Cases** (stebėjimų skaičių), **Mean** (vidurkį), **Median** (mediana), **Grouped Median** (paskirstytų duomenų mediana), **Std. Error of Mean** (standartinę vidurkio paklaidą), **Sum** (sumą), **Minimum** (mažiausią reikšmę), **Maximum** (didžiausią reikšmę), **Range** (plotį), **Variance** (dispersiją), **Standart Deviation** (standartinę nuokrypį), **Kurtosis** (ekscesa), **Standart Error of Kurtosis** (standartinę eksceso paklaidą), **Skewness** (asimetrijos koeficientą), **Standart Error of Skewness** (standartinę asimetrijos koeficiento paklaidą), **First** (pirmą kintamojo reikšmę grupėje), **Last**

(paskutinę kintamojo reikšmę grupėje), *Geometric Mean* (geometrinį vidurkį), *Harmonic Mean* (harmoninį vidurkį), *Percentage of total Sum* (procentą nuo bendros sumos), *Percentage of total N* (procentą nuo bendro kiekio).

- Perkelkite pasirinktus parametrus į dialogo langelio *Cell Statistics* sąrašą. Kokia tvarka bus išrikiuoti parametrai *Cell Statistics* sąraše, tokia bus pateikti išvesties lentelėje.
- Spragtelėkite mygtuką *Continue*.



9.5 pav. Dialogo langelis *Summary Report: Statistics* skaičiuojamoms charakteristikoms pasirinkti

- Spragtelėkite dialogo langelio *Summarize Cases* mygtuką *Options...* Naujame dialogo langelyje *Options* galite pakeisti išvesties pavadinimą, įterpti paaiškinimą (po išvesties lentelės).
- Spragtelėkite mygtuką *Continue*, tada — *OK* dialogo langelyje *Summarize Cases*.

9.6 pav. parodytas *Case Summaries* išvesties pavyzdys, kuriame pateiktos trijų sporto mokyklų 15 jaunųjų krepšininkų 20 m nuotolio bėgimo testo rezultatų (kintamasis *greit_20*) pasirinktų skaitinių charakteristikų reikšmės, suskirstytos pagal tris kategorinio kintamojo *mokykl* reikšmes (mokyklos sąlygiškai sunumeruotos skaičiais nuo 1 iki 3), taip pat bendros reikšmės (*Total*).

Case Processing Summary

	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
GREIT_20 * MOKYKL	45	100,0%	0	,0%	45	100,0%

Case Summaries

GREIT_20

MOKYKL	N	Minimum	Maximum	Range	Mean	Std. Deviation
1	15	3,35	3,75	,40	3,5420	,15077
2	15	3,34	3,75	,41	3,5640	,14505
3	15	3,38	3,80	,42	3,5764	,15287
Total	45	3,34	3,80	,46	3,5608	,14687

9.6 pav. *Case Summaries* išvesties pavyzdys

9.3. DUOMENŲ TYRIMAS (EXPLORE)

Tiriant duomenis *Explore* metodika yra gaunamos suminės kintamųjų arba tų kintamųjų stebėjimų grupių skaitinės charakteristikos ir įvairios diagramos. *Explore* naudojama duomenims patikrinti, jiems aprašyti, įvesties klaidoms aptikti, prielaidoms patikrinti, skirtumams tarp stebėjimų grupių apibūdinti. Toks duomenų patikrinimas (peržvalga) leidžia pastebėti neįprastas arba ekstremalias reikšmes (išskirtis), duomenų spragas ir kitokius ypatumus. Duomenų tyrimas leis nuspręsti, kuris metodas yra tinkamiausias šiems duomenims analizuoti, sprendimui transformuoti duomenis priimti, jeigu jie neatitinka normaliojo skirstinio, arba taikyti neparametrinius testus.

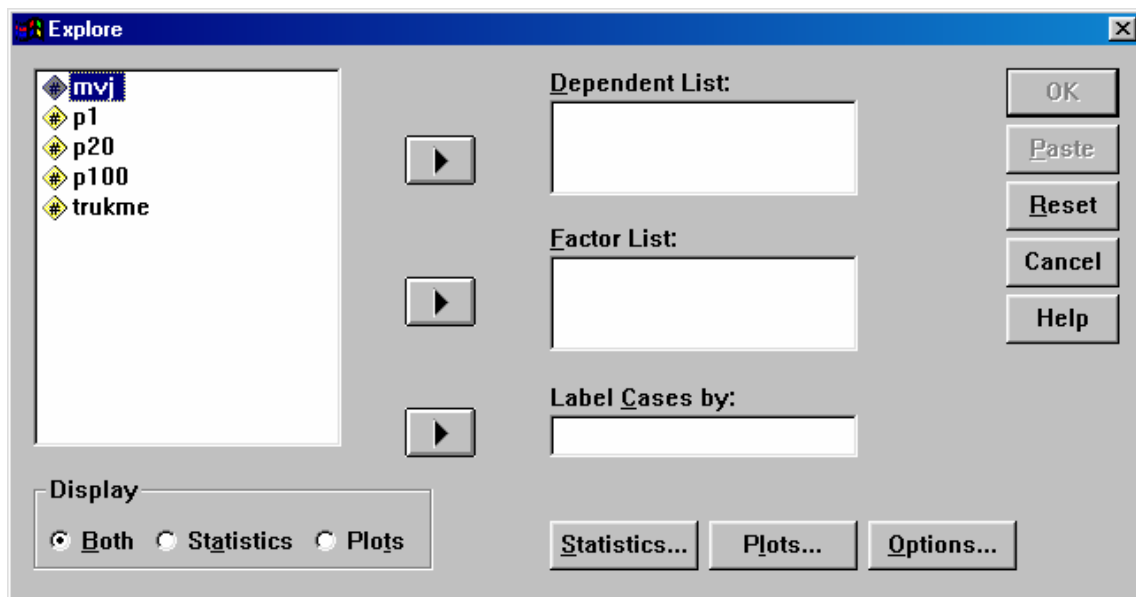
Explore yra pateikiamos šios nustatytosios skaitinės charakteristikos:

Mean (vidurkis), **Median** (mediana), **5% Trimmed Mean** (nupjautasis vidurkis — apskaičiuotas be 5% mažiausių reikšmių ir 5% didžiausių reikšmių), **Standart Error** (standartinė paklaida), **Variance** (dispersija), **Standart Deviation** (standartinis nuokrypis), **Minimum** (mažiausia reikšmė), **Maximum** (didžiausia reikšmė), **Range** (plotis), **Interquartile Range** (tarpkvartilinis plotis — plotis tarp pirmojo ir trečiojo kvartilų), **Kurtosis** (ekscesas), **Standart Error of Kurtosis** (standartinė eksceso paklaida), **Skewness** (asimetrijos koeficientas), **Standart Error of Skewness** (standartinė asimetrijos koeficiento paklaida), **Confidence Interval for the Mean** (pasikliautinasis vidurkio intervalas), **M-estimators** (M-įverčiai): **Huber's M-estimator**, **Andrew's wave estimator**, **Hampel's redescending M-estimator**, **Tukey's biweight estimator**, **Outliers** (išskirtys — penkios mažiausios ir penkios didžiausios reikšmės), **Percentiles** (procentiliai). Taip pat pateikiami **Kolmogorovo ir Smirnov** testo (**Lilliefors** modifikacija) ir **Shapiro ir Wilk** testo skirstinio normalumui patikrinti rezultatai.

Explore metodika naudojama intervalinių kintamųjų analizei. Kategorinis kintamasis, kuris leidžia kintamųjų duomenys suskirstyti į stebėjimų grupes, turi turėti pagrįstą (kiek galima mažesni) reikšmių (kategorijų) skaičių. Šios reikšmės gali būti skaitmenys arba trumpos raidinės-skaitmeninės sekos. Analizuojami duomenys nebūtinai turi turėti normalųjį skirstinį.

Duomenis su **Explore** tirkite taip:

- Atidarykite bylą su analizuojamais duomenimis.
- Nurodykite komandas **Analyze → Descriptive Statistics → Explore...**
- Dialogo langelyje **Explore** (9.7 pav.) įkelkite pasirinktus kintamuosius į analizuojamų kintamųjų sąrašą **Dependent List**.
- Jeigu ketinate atlikti kintamųjų analizę pagal atskiras stebėjimų grupes, į sąrašą **Factor List** įkelkite vieną ar kelis kategorinius kintamuosius, pagal kuriuos bus nustatytos stebėjimų grupės. Jeigu tokios analizės atlikti neketinate, **Factor List** sąrašo nenaudokite.

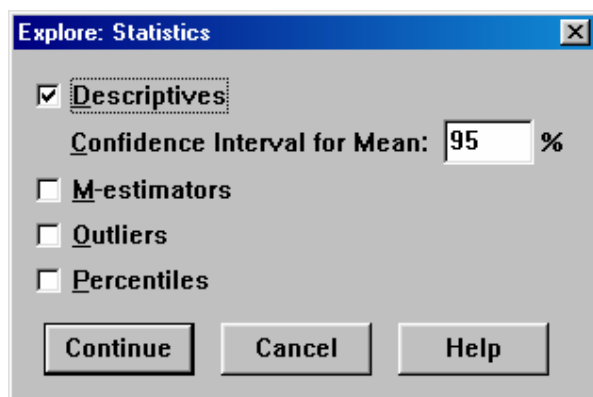


9.7 pav. Dialogo langelis **Explore**

- Komandų grupėje **Display** pasirinkite, kas turi būti pateikta išvestyje: **Both** (abu, t. y. skaitiniai parametrai ir grafikai), **Statistics** (skaitiniai parametrai), **Plots** (grafikai).
- Spragtelėkite mygtuką **Statistics...** Dialogo langelyje **Explore: Statistics** (9.8 pav.) nustatytasis variantas yra **Descriptives** (aprašomoji statistika), pagal kurią skaičiuojamos anksčiau pateiktos skaitinės charakteristikos. Papildomai galite pasirinkti:
 - **M-estimators** (M-įverčiai). M-įverčių esmė yra ta, kad vidurkis skaičiuojamas suteikiant atskiriems stebėjimams skirtingą svorį, kuris mažėja tostant nuo pasiskirstymo centro. Taigi M-įverčiai sumažina

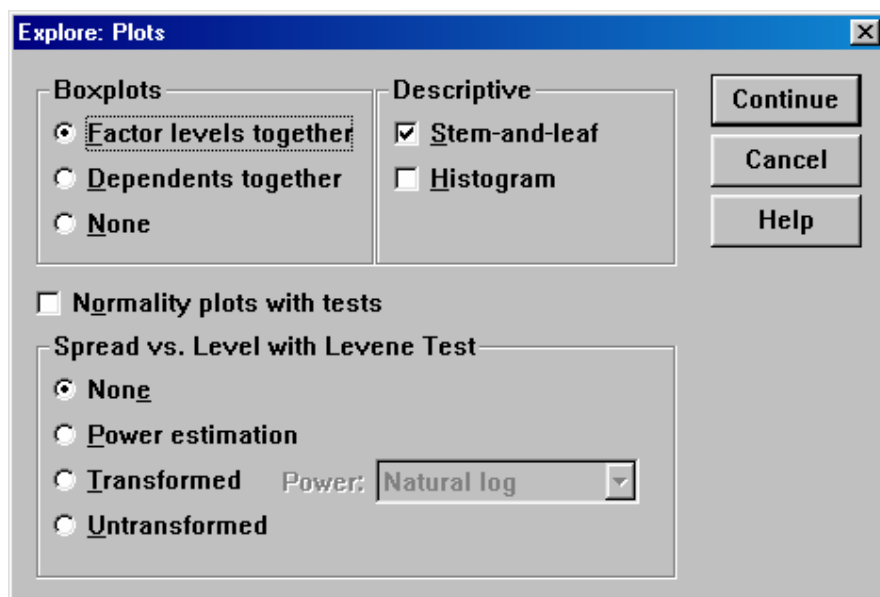
ekstremalių reikšmių įtaką vidurkiui. Galima teigti, kad paprastas vidurkis skaičiuojamas su vienetiniu visų stebėjimų svoriu.

- **Outliers** (išskirties reikšmės). Pateikiamos penkios mažiausios ir penkios didžiausios reikšmės.
 - **Percentiles** (procentiliai). Yra skaičiuojamos septynių procentilių — 5, 10, 25, 50, 75, 90 ir 95 procentų — reikšmės.
- Spragtelėkite mygtuką **Continue**.



9.8 pav. Dialogo langelis **Explore: Statistics** skaičiuojamoms charakteristikoms pasirinkti

- Spragtelėkite **Explore** dialogo langelio mygtuką **Plots...** Dialogo langelyje **Explore: Plots** (9.9 pav.) galite pasirinkti diagramas ir grafikus, kurie bus pateikti išvestyje:
 - **Boxplots** (stulpelinės sklaidos diagramos). Stulpelinę sklaidos diagramą sudaro stačiakampis, užimantis erdvę nuo pirmojo iki trečiojo kvartilio, t. y. nuo 25 iki 75 procentilių. Linija stačiakampio viduje atitinka medianą. Be to, šioje diagramoje pavaizduojama mažiausia ir didžiausia reikšmės (ne išskirtys — išskirčių reikšmės pažymimos specialiais simboliais). Yra du alternatyvūs stulpelinės sklaidos diagramos sudarymo variantai, jeigu turite daugiau kaip vieną priklausomą kintamąjį (*dependent variable*). Pasirinkę **Factor levels together**, gausite kiekvieno priklausomo kintamojo atskiras stulpelines sklaidos diagramas. Kiekvienoje šių diagramų (bendroje koordinačių sistemoje) bus pavaizduotos visos kintamojo kategorijos, kurias lemia faktorinis kintamasis. Pasirinkę **Dependents together** gausite atskiras stulpelines sklaidos diagramas kiekvienos kategorijos, kurią lemia faktorinis kintamasis. Kiekvienoje šių diagramų (bendroje koordinačių sistemoje) bus pavaizduoti visi priklausomi kintamieji. Šis variantas yra patogus, kai yra matuojami tie patys atskirų kintamųjų parametrai ir juos norima vizualiai palyginti.
- Pasirinkus **None**, ši diagrama nebus pavaizduota.



9.9 pav. Dialogo langelis **Explore: Plots** diagramoms ir grafikams pasirinkti

- **Descriptive** (aprašomoji). Jūs galite pasirinkti **Stem-and-leaf** (diagramą-medį) ir **Histogram** (histogramą). Trumpai paaikšinsime pirmąją diagramą. Diagrama-medis (9.14 pav.) yra histogramos ir lentelės sąrašo kombinacija. Kaip ir diagramoje, kiekvienos eilutės ilgis atitinka stebėjimų, patenkančių į tam tikrą intervalą, skaičių. Tačiau šioje diagramoje pateikiama ir kiekvieno stebėjimo reikšmė. Tam skaitmeniniai stebėjimų duomenys suskaldomi į dvi dedamąsias: kamieną, kuris rodo pirmą skaitmenį (arba skaitmenų grupę) ir lapą, kuris rodo likusius. Kamienas atitinka tas skaitmeninių reikšmių skiltis, kurios nesikeičia tam tikrame intervale, o lapai atitinka tas skaitmeninių reikšmių skiltis, kurios keičiasi tame intervale.
- **Normality plots with tests** (skirstinio normalumo grafikas ir testas). Pažymėjus šį laukelį, bus pateiktas **Kolmogorovo ir Smirnov**o testo rezultatas su reikšmingumo korekcija remiantis Lilliefors'u skirstinio normalumui patikrinti (9.11 pav.). Jei gauta p -reikšmė (*Sign*) mažesnė už reikšmingumo lygmenį 0,05, skirstinys esmiškai skiriasi nuo normaliojo. Kai imtį sudaro mažiau kaip 50 stebėjimų, atliekamas taip pat **Shapiro ir Wilk** testas. Pateikiami taip pat du grafikai, leidžiantys vizualiai įvertinti, kiek arti normaliojo yra tiriamo kintamojo skirstinys: skirstinio normalumo grafikas ($Q-Q$ grafikas) ir skirstinio normalumo grafikas be tendencijos komponentės. $Q-Q$ grafike (9.15 pav.) kiekvieno stebėjimo reikšmė lyginama su ta reikšme, kuri turėtų būti pagal normalųjį skirstinį. Stebėjimų duomenys atidedami X ašyje, laukiami pagal normalųjį skirstinį — Y ašyje. Visi duomenys

standartizuojami (z-transformacija). Tuo atveju, kai stebėjimų duomenys tiksliai atitinka normalųjį skirstinį, visi taškai grafike yra tiesėje. *Q-Q* grafike be tendencijos dedamosios (9.16 pav.) pateikiami stebėjimų duomenų skirtumai nuo laukiamų pagal normalųjį skirstinį (Y ašyje) priklausomai nuo stebėjimų duomenų (X ašyje). Tuo atveju, kai stebėjimų duomenys tiksliai atitinka normalųjį skirstinį, visi taškai grafike yra horizontalioje tiesėje, nubrėžtoje per Y ašies nulinę reikšmę. Visi duomenys taip pat standartizuojami (z-transformacija).

- ***Spread vs. Level with Levene Test*** (sklaida — lygmuo su Levene testu). Tikrinamas priklausomų kintamųjų atskirų stebėjimų grupių (kurias lemia kategorinis kintamasis) dispersijų homogeniškumas. Pasirinkę variantą ***Untransformed***, gausite pradinių (netransformuotų) duomenų diagramą. Pasirinkę variantą ***Power estimation*** (eksponentinis įvertis), galėsite įvertinti, kiek skiriasi atskirų stebėjimų grupių dispersijos, t. y. patikrinti jų homogeniškumą (9.12 pav.). Jei gauta kriterijaus *p*-reikšmė (*Sign*) mažesnė už reikšmingumo lygmenį 0,05, tai dispersijos skiriasi esmiškai. Diagramoje (9.18 pav.) parodoma kiekvienos grupės duomenų reikšmių sklaidos priklausomybė nuo centrinės reikšmės, t. y. X ašyje atidedama medianos logaritmo reikšmė, o Y ašyje — tarpkvartilinio pločio logaritmo reikšmė. ***Power estimation*** įvertina transformacijos, dėl kurios dispersijos būtų galimai lygios, laipsnį. Duomenų transformacijai pasirinkite variantą ***Transformed*** ir atitinkamą transformacijos laipsnį išskleidžiamajame sąraše ***Power***. Reikia pažymėti, kad duomenų transformaciją gali taikyti tik patyrę vartotojai, nes netiesinė transformacija pakeičia proporcijas tarp duomenų grupių, o išvados daromos transformuotų duomenų pagrindu. Pasirinkus ***None*** (numatytasis variantas), testo rezultatai ir diagrama nebus pateikti.

- Spragtelėkite mygtuką ***Continue***.
- Spragtelėję ***Explore*** dialogo langelio mygtuką ***Options...***, naujame dialogo langelyje ***Explore: Options*** galite nurodyti, kaip turi būti traktuojamos praleistos reikšmės. Nustatytasis variantas — ***Exclude cases listwise*** (visų kintamųjų praleistos reikšmės į skaičiavimus neįtraukiamos).
- Spragtelėkite dialogo langelio ***Explore*** mygtuką ***OK***.

Explore pateikiamus rezultatus pailiustruosime pavyzdžiu.

Buvo analizuojami dviejų irklutojų grupių (po 30 sportininkų kiekvienoje grupėje) išvystomo didžiausio galingumo (W) matavimo rezultatai. Kintamasis pavadintas *p_max*, kintamojo žymena — *didžiausias galingumas*. Faktorinis kintamasis pavadintas *grup*, faktorinio kintamojo žymena — *grupė*. Dialogo langelyje ***Explore*** kintamąjį *p_max* įkeliamė į laukelį ***Dependent List***, o kintamąjį *grup* — į laukelį ***Factor List***. Dialogo langelyje ***Explore: Statistics*** paliekame nustatytąjį variantą ***Descriptives***, o dialogo langelyje ***Explore: Plots***

pasirenkame: **Boxplots: Factor levels together**, **Descriptive: Stem-and-leaf** ir **Histogram**, **Normality plots with tests**, **Spread vs. Level with Levene Test**: **Power estimation**.

Gauti rezultatai pateikti 9.10 – 9.18 pav.:

- 9.10 pav. pateikta skaitinių charakteristikų (*Descriptives*) lentelė.

Case Processing Summary						
grupė		Cases				
		Valid		Missing		Total
		N	Percent	N	Percent	N
didžiausias galingumas	1	30	100,0%	0	,0%	30
	2	30	100,0%	0	,0%	30

a)

Descriptives					
grupė			Statistic	Std. Error	
didžiausias galingumas	1	Mean	586,2333	16,20208	
		95% Confidence Interval for Mean	Lower Bound Upper Bound		553,0964 619,3703
		5% Trimmed Mean	588,1481		
		Median	597,0000		
		Variance	7875,220		
		Std. Deviation	88,74243		
		Minimum	369,00		
		Maximum	764,00		
		Range	395,00		
		Interquartile Range	130,7500		
		Skewness	-,347		
		Kurtosis	,106		
		2	Mean		590,8333
	95% Confidence Interval for Mean		Lower Bound Upper Bound	570,2066 611,4600	
	5% Trimmed Mean		591,2407		
	Median		593,0000		
	Variance		3051,385		
	Std. Deviation		55,23934		
	Minimum		477,00		
	Maximum		699,00		
	Range		222,00		
	Interquartile Range		74,2500		
	Skewness		-,125		
	Kurtosis		-,172		

b)

9.10 pav. Skaitinių charakteristikų lentelė **Descriptives**

- 9.11 pav. pateikti normalaus skirstinio testo rezultatai. Kadangi vienai imties grupei tenka 30 stebėjimų, t. y. mažiau negu 50, reikėtų vadovautis **Shapiro ir Wilk** testo rezultatais: pirmos grupės p -reikšmė (stebimasis reikšmingumo lygmuo — *Sig.*) yra 0,861, antros grupės — 0,911, t. y. viršija reikšmingumo lygmenį 0,05, todėl darome išvadą, kad abiejų grupių duomenys turi normalųjį skirstinį.

Tests of Normality						
grupė	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
didžiausias galingumas 1	,126	30	,200*	,981	30	,861
2	,089	30	,200*	,984	30	,911

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

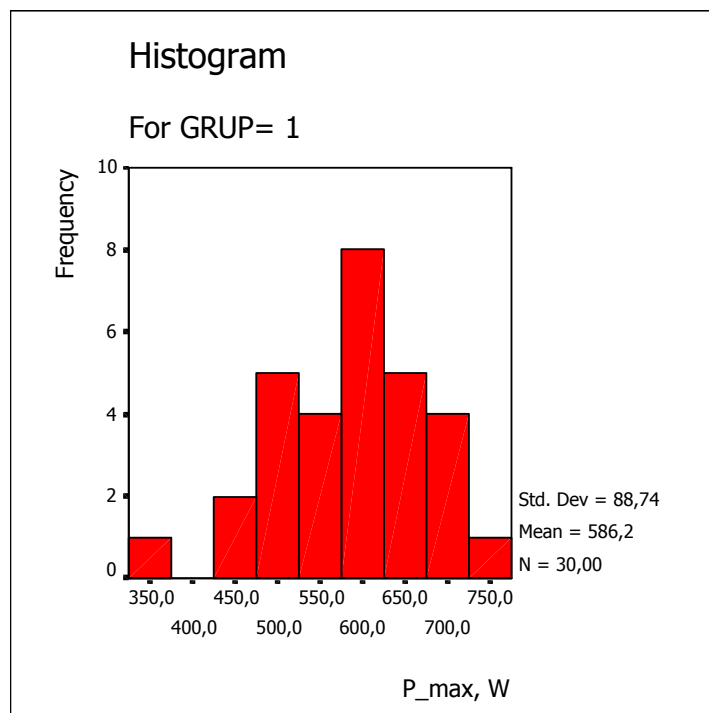
9.11 pav. Normaliojo skirstinio testo rezultatai

- 9.12 pav. pateikti **Levene** testo rezultatai dispersijų homogeniškumui patikrinti. Visais atvejais p -reikšmės (*Sig.*) yra mažesnės už reikšmingumo lygmenį, todėl darome išvadą, kad abiejų grupių dispersijos skiriasi esmiškai.

Test of Homogeneity of Variance					
		Levene Statistic	df1	df2	Sig.
didžiausias galingumas	Based on Mean	5,295	1	58	,025
	Based on Median	4,338	1	58	,042
	Based on Median and with adjusted df	4,338	1	47,794	,043
	Based on trimmed mean	5,076	1	58	,028

9.12 pav. Dispersijų homogeniškumo testo rezultatai

- 9.13 pav. parodyta pirmos sportininkų grupės didžiausio išvystomo galingumo histograma.
- 9.14 pav. parodyta pirmos sportininkų grupės didžiausio išvystomo galingumo *Stem-and-Leaf* diagrama. Jos kamieną (*Stem*) sudaro šimtai, atskirus lapus (*Leaf*) — dešimtys. *Frequency* stulpelyje parodytas stebėjimų reikšmių, patenkančių į tam tikrą diapazoną, kiekis.



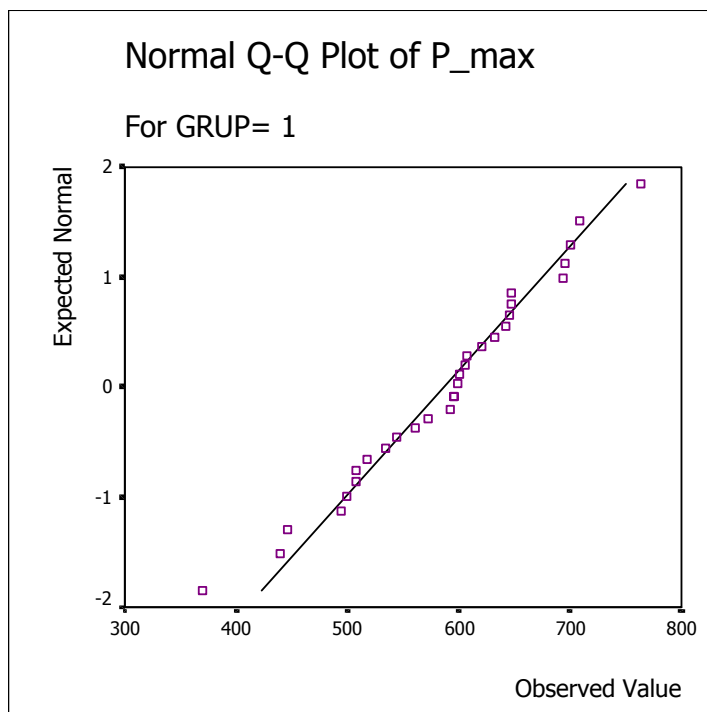
9.13 pav. Didžiausio išvystomo galingumo histograma

didžiausias galingumas Stem-and-Leaf Plot for
GRUP= 1

Frequency	Stem &	Leaf
1,00	3 .	6
2,00	4 .	44
1,00	4 .	9
6,00	5 .	000134
6,00	5 .	679999
9,00	6 .	000234444
2,00	6 .	99
2,00	7 .	00
1,00	7 .	6
Stem width: 100,00		
Each leaf: 1 case(s)		

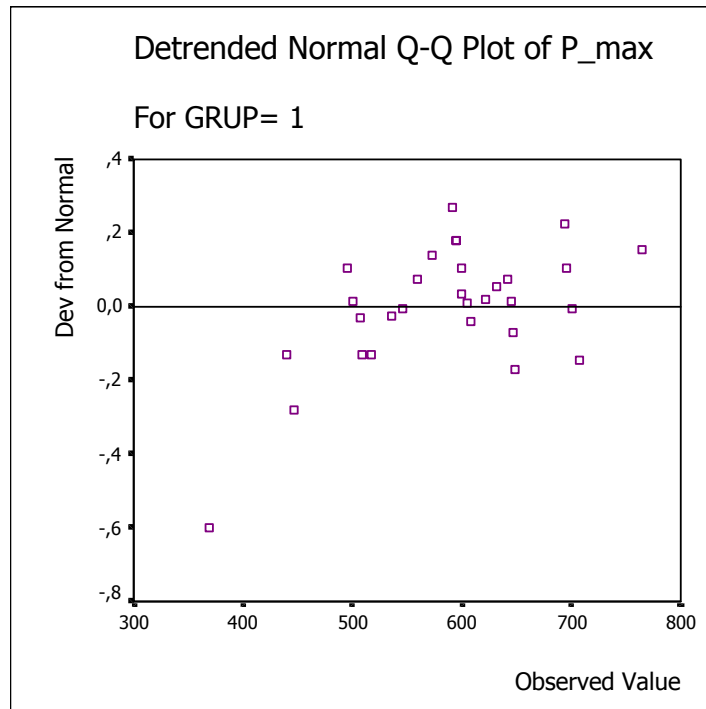
9.14 pav. Didžiausio išvystomo galingumo Stem-and- Leaf diagrama.

- 9.15 pav. parodytas pirmos sportininkų grupės didžiausio išvystomo galingumo skirstinio normalumo grafikas ($Q-Q$ grafikas), kuriame stebėjimų duomenų reikšmės pakankamai arti tiesės, t. y. skirstinys esmiškai nesiskiria nuo normaliojo.

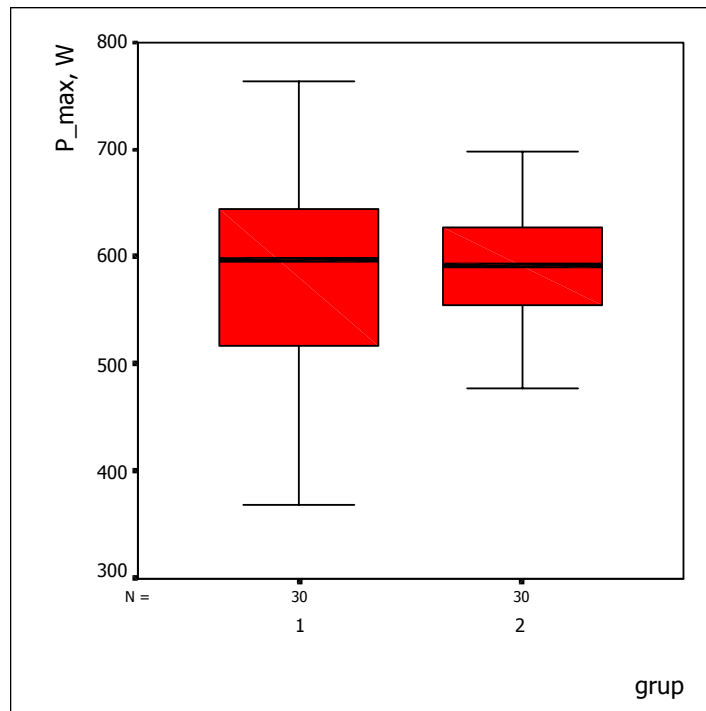


9.15 pav. Didžiausio išvystomo galingumo pasiskirstymo normalumo ($Q-Q$) grafikas.

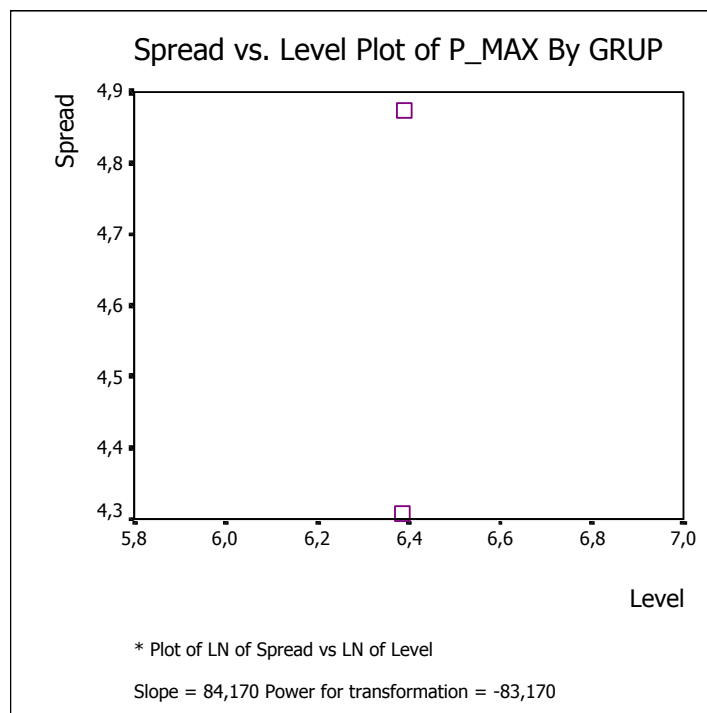
- 9.16 pav. parodytas pirmos sportininkų grupės didžiausio išvystomo galingumo skirstinio normalumo grafikas ($Q-Q$ grafikas) be tendencijos dedamosios.
- 9.17 pav. parodyta pirmos ir antros sportininkų grupės didžiausio išvystomo galingumo stulpelinė sklaidos diagrama.
- 9.18 pav. parodyta pirmos ir antros sportininkų grupės didžiausio išvystomo galingumo *Spread vs. Level* (sklaida-lygmuo) diagrama. *Power for transformation* rodiklis rodo, kad duomenis reikėtų transformuoti aukštu laipsniu, norint, kad dispersijos būtų kiek galima lygios.



9.16 pav. Didžiausio išvystomo galingumo pasiskirstymo normalumo (Q-Q) grafikas be tendencijos komponentės



9.17 pav. Pirmos ir antros sportininkų grupės didžiausio išvystomo galingumo stulpelinė sklaidos diagrama



9.18 pav. Pirmos ir antros sportininkų grupės didžiausio išvystomo galingumo Spread vs. Level (sklaida-lygmuo) diagrama

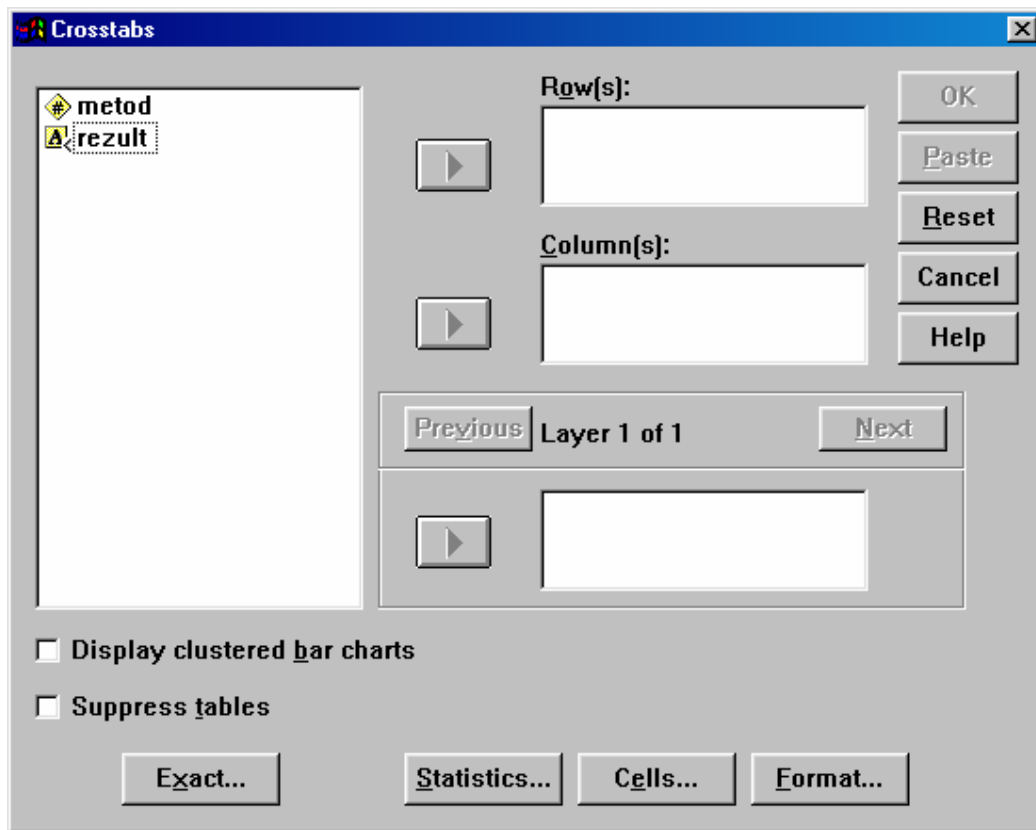
10. POŽYMIŲ DAŽNIŲ LENTELĖS

Šiame skyriuje nagrinėsime priklausomybę tarp nominaliųjų ir ranginių kintamųjų. SPSS tam tikslui naudojamos vadinamosios požymių dažnių lentelės (*Crosstabs*) ir didelė testų įvairovė priklausomybės laipsniui tarp nagrinėjamų kintamųjų įvertinti. Ranginiai ir nominalieji kintamieji vadinami kategoriniais, o kategoriniai kintamieji savo ruožtu dažnai — požymiais (Čekanavičius, Murauskas, 2000). Iš čia šių lentelių pavadinimas. Analizei priklausomybės tarp matuojamų pagal intervalų skalę kintamųjų yra skiriamas 13 skyrius.

10.1. POŽYMIŲ DAŽNIŲ LENTELIŲ SUDARYMAS

Požymių dažnių lentelės sudaromos taip:

- Atidarykite bylą su analizuojamais duomenimis.
- Nurodykite komandas *Analyze* → *Descriptive Statistics* → *Crosstabs...* Atsidarys dialogo langelis *Crosstabs* (10.1 pav.).



10.1 pav. Dialogo langelis *Crosstabs*

- Iš sąrašo įkelkite į laukelį **Row(s)** kintamuosius, kuriuos norite dėti į dažnių lentelės eilutes, o į laukelį **Colum(s)** – kintamuosius, kuriuos norite dėti į dažnių lentelės stulpelius. Kiekvienam dviejų kintamųjų deriniui bus sukurta atskira dažnių lentelė. Pavyzdžiui, jeigu **Row(s)** sąrašė yra trys kintamieji, o **Colum(s)** sąrašė – du kintamieji, gausime $3 \times 2 = 6$ dažnių lenteles.

Pateiksime pavyzdį. Dvi vienodo meistriškumo vaikų grupės buvo treniruojamos dviem skirtingomis metodikomis. Po tam tikro laiko abi grupės buvo testuotos ir rezultatai įvertinti įverčiais „gerai“, „vidutiniškai“, „blogai“. Mūsų tikslas yra nustatyti, kuri metodika geresnė. Iš pradžių sudarysime dažnių lentelę. Pirmą kintamąjį (metodiką) pavadinsime *metod* ir suteiksime skaitmeninį formatą (kintamojo reikšmės 1 ir 2), antrą kintamąjį pavadinsime *rezult* ir suteiksime raidinės sekos formatą (kintamojo reikšmės *g*, *v* ir *b*, o reikšmių žymenos – atitinkamai *geri*, *vidutin.* ir *blogi*). Dialogo langelyje **Crosstabs** kintamąjį *metod* įkeliamo į eilučių sąrašą **Row(s)**, o kintamąjį *rezult* – į stulpelių sąrašą **Colum(s)**. Spragtelėję **OK**, **Viewer** lange gauname 10.2 pav. parodytą lentelę.

METOD * RESULT Crosstabulation

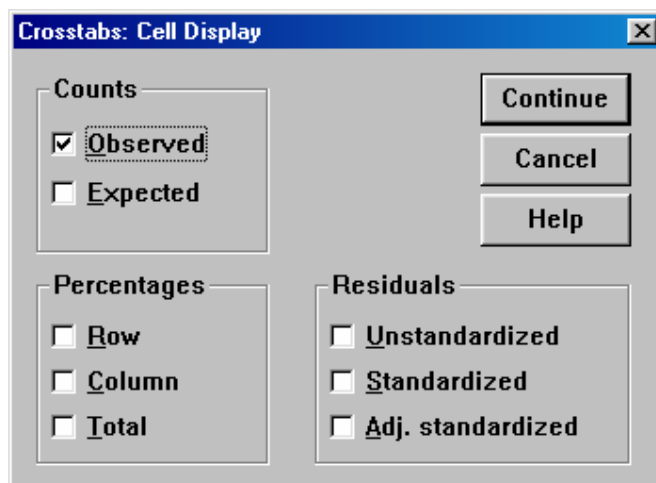
Count		RESULT			Total
		blogi	geri	vidutin	
METOD	1	4	6	17	27
	2	6	9	9	24
Total		10	15	26	51

10.2 pav. Standartinio formato požymių dažnių lentelė

Dažnių lentelėje kintamojo *metod* reikšmės įrašytos į atskiras eilutes, o kintamojo *rezult* reikšmės (tiksliau, jų žymenos) — į atskirus stulpelius. Kiekvienoje lentelės ląstelėje yra įrašytas stebėjimų skaičius (pasirodymo dažnis) — *Count*, o *Total* eilutėje ir stulpelyje — atitinkamai eilučių ir stulpelių reikšmių sumos. Šios standartinio formato lentelės, kuri gaunama pagal SPSS nustatytus pradinis parametrus, ląstelėse rodoma tik absoliuti stebėjimų skaičiaus reikšmė. Kad sužinotumėte procentinę šių reikšmių dalį bendro kiekio atžvilgiu:

- Spragtelėkite dialogo langelio **Crosstabs** mygtuką **Cells...**
- Dialogo langelio **Crosstabs: Cell Display** (10.3 pav.) komandų grupėje **Percentages** pasirinkite vieną ar kelis šių variantų:
 - **Row** (pagal eilutes): procentinės reikšmės skaičiuojamos pagal eilutes, t. y. kiekvienos ląstelės reikšmė eilutės sumos atžvilgiu.
 - **Column** (pagal stulpelius): procentinės reikšmės skaičiuojamos pagal stulpelius, t. y. kiekvienos ląstelės reikšmė stulpelio sumos atžvilgiu.

- **Total** (iš viso): kiekvienos ląstelės reikšmė bendro stebėjimų skaičiaus atžvilgiu.



10.3 pav. Dialogo langelis **Crosstabs: Cell Display**

Pažymėję visus tris **Percentages** laukelius, mūsų atveju gausime 10.4 pav. parodytą lentelę.

METOD * RESULT Crosstabulation

			RESULT			Total
			blogi	geri	vidutin	
METOD 1	Count		4	6	17	27
	% within METHOD		14,8%	22,2%	63,0%	100,0%
	% within RESULT		40,0%	40,0%	65,4%	52,9%
	% of Total		7,8%	11,8%	33,3%	52,9%
2	Count		6	9	9	24
	% within METHOD		25,0%	37,5%	37,5%	100,0%
	% within RESULT		60,0%	60,0%	34,6%	47,1%
	% of Total		11,8%	17,6%	17,6%	47,1%
Total	Count		10	15	26	51
	% within METHOD		19,6%	29,4%	51,0%	100,0%
	% within RESULT		100,0%	100,0%	100,0%	100,0%
	% of Total		19,6%	29,4%	51,0%	100,0%

10.4 pav. Požymių dažnių lentelė su procentinėmis stebėjimų skaičiaus reikšmėmis eilučių, stulpelių ir bendros sumos atžvilgiu

Geriau suprasti priklausomybę tarp kintamųjų padeda teoriškai tikėtinas (laukiamas) stebėjimų skaičius (pasirodymo dažnis), kuris skaičiuojamas kaip atitinkamos eilutės ir atitinkamo stulpelio sumų sandauga, padalyta iš visų stebėjimų sumos. Lyginant tikėtiną stebėjimų skaičių su eksperimentiniu būdu nustatytu ir atsižvelgiant į nukrypimo pobūdį (daugiau ar mažiau), galima

padaryti tam tikras išvadas apie vieno kintamojo reikšmių priklausomybę nuo kito kintamojo.

Tikėtinam stebėjimų skaičiui surasti:

- Spragtelėkite dialogo langelio **Crosstabs** mygtuką **Cells...**
- Dialogo langelio **Crosstabs: Cell Display** komandų grupėje **Counts**, kurioje numatytasis variantas yra **Observed** (stebimas), pažymėkite taip pat laukelį **Expected** (tikėtinas, laukiamas).
- Komandų grupėje **Residuals** (liekanos) galite pasirinkti šiuos liekanų, parodančių, kiek stipriai skiriasi tikėtini stebėjimai nuo esamų, variantus:
 - **Unstandardized** (nenormuotos): pateikiamas skirtumas tarp tikėtinų ir nustatytų reikšmių.
 - **Standardized** (normuotos): tai liekanos, kurių vidurkis lygus 0, o standartinis nuokrypis lygus 1.
 - **Adj. Standardized** (patikslintos normuotos): normuotos liekanos, išreikštos standartiniu nuokrypiu nuo vidurkio.

Mūsų atveju gausime 10.5 pav. parodytą lentelę. Tikėtinės (teorinės) reikšmės yra artimos gautoms (nustatytoms) — tai rodo, kad požymiai yra nepriklausomi.

METOD * REZULT Crosstabulation

			REZULT			Total
			blogi	geri	vidutin	
METOD 1	Count		4	6	17	27
	Expected Count		5,3	7,9	13,8	27,0
2	Count		6	9	9	24
	Expected Count		4,7	7,1	12,2	24,0
Total	Count		10	15	26	51
	Expected Count		10,0	15,0	26,0	51,0

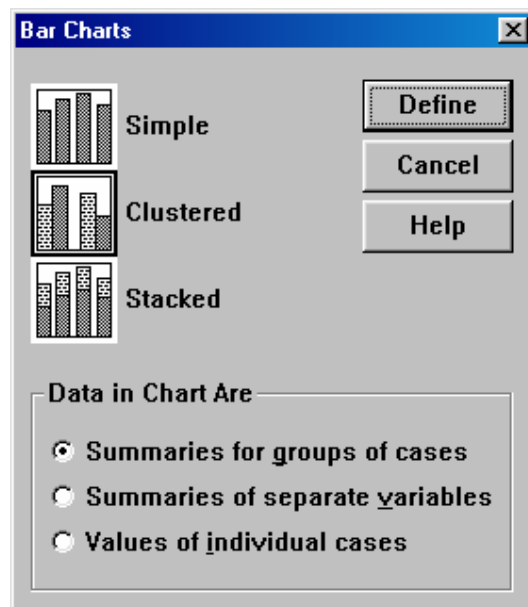
10.5 pav. Požymių dažnių lentelė su tikėtino stebėjimų skaičiaus reikšmėmis

Kintamųjų, nurodytų **Row(s)** ir **Column(s)** laukeliuose, požymių dažnių lentelės galima sudaryti pagal kiekvieną kintamojo, nurodyto **Layer** (sluoksnis) laukelyje, kategoriją. Pateiktame pavyzdyje toks kintamasis, turintis kelias reikšmių kategorijas, galėtų būti vaikų amžius. Norint sudaryti kitą sluoksnį, reikia spragtelėti mygtuką **Next**. Sudaromos kiekvienos pirmojo ir antrojo sluoksnio kintamųjų kategorijų kombinacijos ir t. t. lentelės, jeigu yra daugiau negu du sluoksnio kintamieji.

10.2. POŽYMIŲ DAŽNIŲ LENTELIŲ DIAGRAMOS

Kad geriau būtų galima suvokti duomenis, pateikiamus požymių dažnių lentelėse, juos galima parodyti grafiškai. Tam gerai tinka klasterinės (grupinės) stulpelinės diagramos. Norėdami pavaizduoti duomenis klasterine stulpeline diagrama, darykite taip:

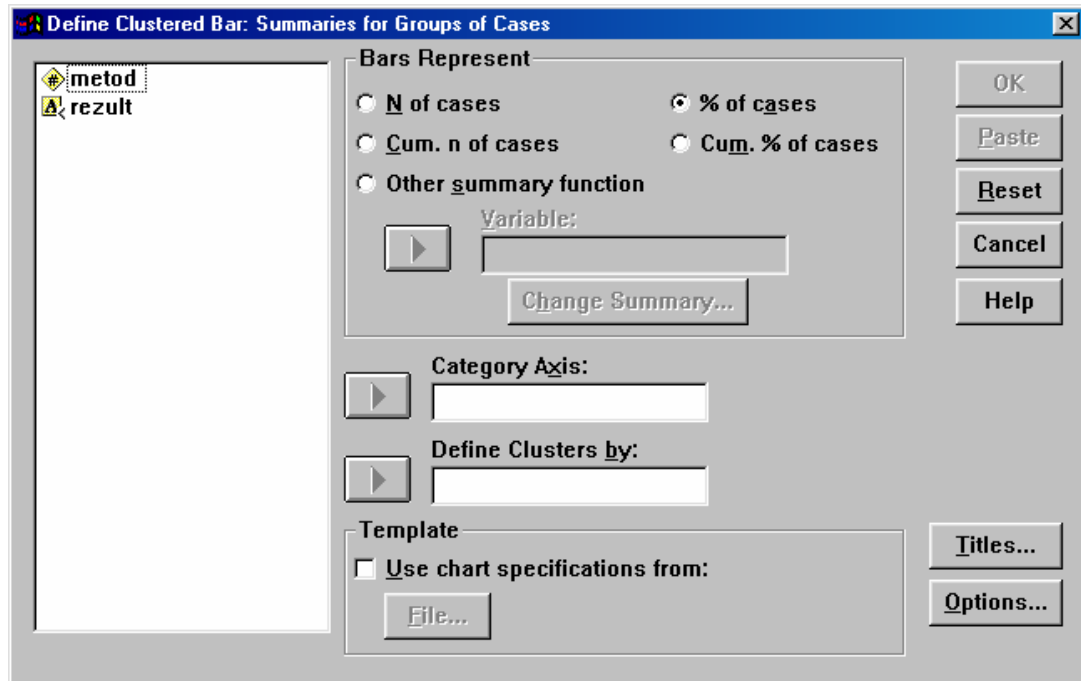
- Atidarykite bylą su analizuojamais duomenimis.
- Nurodykite komandas **Graphs → Bar...**
- Dialogo langelyje **Bar Charts** (stulpelinės diagramos) (10.6 pav.) pasirinkite **Clustered** (klasterinės), laukelyje **Data in Chart Are** – duomenų pateikimo būdą diagramoje: **Summaries for groups of cases** (kintamojo kategorijų suvestinė) – tai numatytasis variantas, **Summaries of separate variables** (atskirų kintamųjų suvestinė), **Values of individual cases** (atskirų stebėjimų reikšmės).



10.6 pav. Dialogo langelis **Bar Charts**

- Spragtelėkite dialogo langelio **Bar Charts** mygtuką **Define** (nustatyti).
- Atsidariusiame naujame dialogo langelyje **Define Clustered Bar: Summaries for groups of cases** (jei palikote numatytąjį variantą **Summaries for groups of cases** dialogo langelyje **Bar Charts**) (10.7 pav.) galite pasirinkti stulpelių reikšmes (**Bars Represent**): **N of cases** (stebėjimų skaičius), **Cum. n of cases** (sukauptas stebėjimų skaičius), **Other summary function** (kitos sumavimo funkcijos), **% of cases** (stebėjimų procentas), **Cum. % of cases** (sukauptas stebėjimų procentas).
- Įkelkite iš kintamųjų sąrašo į laukelį **Category Axis** (X ašis) kintamąjį, kuris bus atidėtas X ašyje.

- Į laukelį **Define Clusters by** (nustatyti grupes pagal) įkelkite kintamąjį, pagal kurį bus sudarytos stulpelių sekos.



10.7 pav. Dialogo langelis **Define Clustered Bar: Summaries for Groups of Cases** stulpelinės diagramos parametrus nustatyti

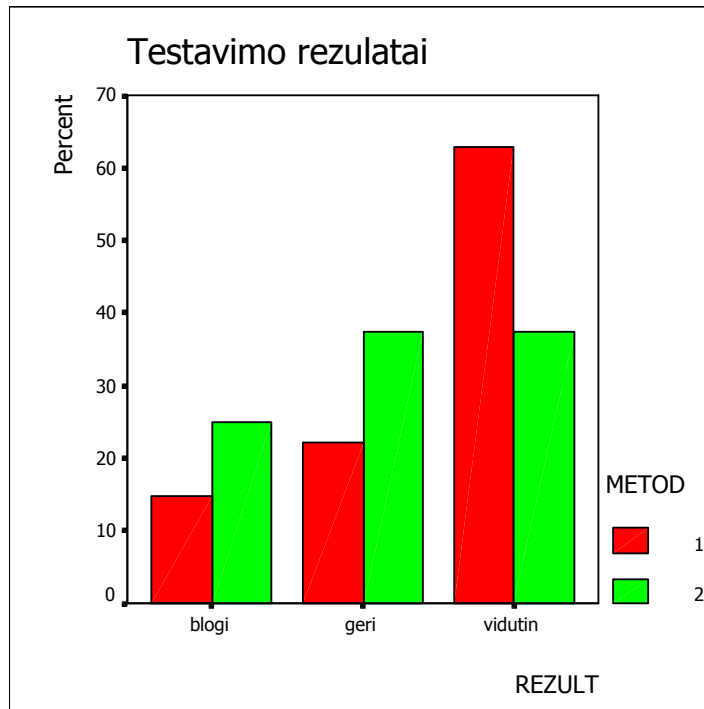
- Spragtelėkite mygtuką **Titles...** ir atsidariusiame dialogo langelyje **Titles** (10.8 pav.) įrašykite diagramos pavadinimą, paantraštę ir reikalingas pastabas.



10.8 pav. Dialogo langelis **Titles** diagramos pavadinimui ir pastaboms įrašyti

- Spragtelėkite mygtuką **Continue**, tada — **OK** dialogo langelyje **Define Clustered Bar: Summaries for groups of cases**. **Viewer** lange pasirodys stulpelinė diagrama, kurią galite redaguoti remiantis 8.3 skyriuje pateiktais paaiškinimais.

10.9 pav. parodyta stulpelinė klasterinė diagrama, sudaryta remiantis 10.2 pav. požymių dažnių lentelės duomenimis. Kintamasis *rezult* buvo įrašytas **Category Axis** laukelyje, o kintamasis *metod* — **Define Clusters by** laukelyje.



10.9 pav. Stulpelinės klasterinės diagramos pavyzdys

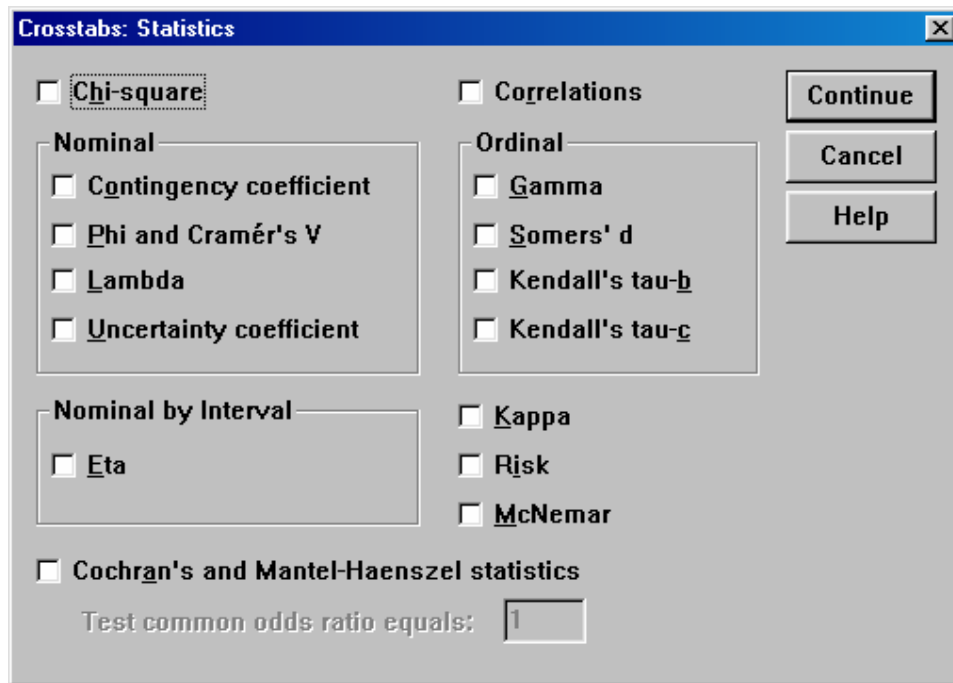
Požymių dažnių lentelių stulpelinės diagramos nustatytąjį variantą galite gauti ir pažymėję dialogo langelio **Crosstabs** laukelį **Display clustered bar charts** (rodyti klasterinę stulpelinę diagramą).

10.3. POŽYMIŲ DAŽNIŲ LENTELIŲ KRITERIJAI

Statistinėms išvadoms apie užrašytus dažnių lentelėmis kategorinius kintamuosius (požymius) gauti:

- Spragtelėkite dialogo langelio **Crosstabs** mygtuką **Statistics...**
- Atsidariusiame naujame dialogo langelyje **Crosstabs: Statistics** (10.10 pav.) galite pasirinkti šiuos kriterijus:
 - **Chi-square** (Chi-kvadrato (χ^2) testas).
 - **Correlation** (koreliacijos koeficientai).
 - **Nominal** — koeficientai nominaliųjų kintamųjų tarpusavio ryšiui įvertinti.

- **Ordinal** — koeficientai ranginių kintamųjų tarpusavio ryšiui įvertinti.
- **Nominal by Interval** — koeficientai tarpusavio ryšiui tarp nominaliųjų ir intervalinių kintamųjų įvertinti.
- **Kappa** — kapa koeficientas.
- **Risk** — rizikos matas.
- **McNemar** testas.
- **Cochran's and Mantel-Haenszel statistics** testas.



10.10 pav. Dialogo langelis **Crosstabs: Statistics** požymių kriterijams pasirinkti

10.3.1. Chi-kvadrato (χ^2) testas

Chi-kvadrato testas užima išskirtinę vietą statistiniuose skaičiavimuose. Vienas iš populiariausių neparametrinių kriterijų — Chi-kvadrato testas naudojamas hipotezėms apie kintamojo skirstinį populiacijoje tikrinti (t. y., ar empirinio ir teorinio skirstinių skirtumas yra reikšmingas), dviejų kintamųjų nepriklausomumui ar vieno kintamojo homogeniškumui tikrinti.

SPSS Chi-kvadrato kriterijus skaičiuojamas trejopai: pagal Pirsono (*Pearson*) formulę, pagal tikėtinumo santykio (*Likelihood Ratio*) formulę ir pagal Mantelio ir Haenzelio (*Linear-by-Linear*) formulę. Kai duomenys aprašomi keturlauke (2×2) dažnių lentele ir kai tikėtinas stebėjimų skaičius mažiau negu penki, papildomai skaičiuojamas tikslus Fišerio (*Fisher's*) testas. Tikslus χ^2 testas mažoms imtims taip pat pasirenkamas spragtelėjus **Crosstabs** dialogo langelio mygtuką **Exact...**

Pirsono formulė Chi-kvadrato kriterijaus reikšmei apskaičiuoti yra ši:

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}, \quad (10.1)$$

čia O_i — nustatyti dažniai, E_i — tikėtini dažniai, k — kintamųjų kategorijų grupių skaičius (dažnių lentelės ląstelių skaičius).

Asimptotinio Chi-kvadrato testo rezultatų patikimumą sąlygoja šie reikalavimai: bent 80% dažnių lentelės ląstelių tikėtini dažniai turi būti ne mažesni kaip 5 arba tikėtinas dažnis bet kurioje lentelės ląstelėje turi būti ne mažesnis kaip 1, stebėjimų skaičius turi būti ne mažesnis kaip 30.

Tikėtinumo santykio formulė Chi-kvadrato kriterijaus reikšmei apskaičiuoti yra ši:

$$\chi^2 = -2 \cdot \sum_{i=1}^k O_i \cdot \ln \frac{E_i}{O_i}, \quad (10.2)$$

Esant didelėms imtims pagal Pirsono formulę ir tikėtinumo santykio formulę gaunami artimi rezultatai.

Mantelio-Haenzelio formulė Chi-kvadrato kriterijaus reikšmei apskaičiuoti yra ši:

$$\chi^2 = r^2 \cdot (n - 1), \quad (10.3)$$

čia r — Pirsono koreliacijos koeficientas.

Išvesties lentelėje šis testas nurodomas **Linear-by-Linear** pavadinimu. Netaikomas nominaliems kintamiesiems.

Chi-kvadrato testo rezultatams gauti:

- Spragtelėkite dialogo langelio **Crosstabs** mygtuką **Statistics...**
- Atsidariusiame naujame dialogo langelyje **Crosstabs: Statistics** pasirinkite **Chi-square**.
- Spragtelėkite dialogo langelio **Crosstabs: Statistics** mygtuką **Continue**, tada — dialogo langelio **Crosstabs** mygtuką **OK**.

Išvestyje bus pateiktos Chi-kvadrato kriterijaus reikšmės, taip pat kriterijaus p -reikšmė (*Asymp. Sig.(2-sided)*), kuria remiantis priimamas sprendimas:

- Kai tikrinamas požymių nepriklausomumas (kintamųjų pora pasitaiko vienoje populiacijoje) — požymiai statistiškai priklausomi, kai $p < \alpha$;

požymiai statistiškai nepriklausomi, kai $p \geq \alpha$, čia α — nustatytas reikšmingumo lygmuo.

- Kai tikrinamas požymių homogeniškumas (keliose populiacijose pasitaiko tas pats kintamasis) — požymio skirstiniai skirtingose populiacijose skiriasi reikšmingai, kai $p < \alpha$; požymio skirstiniai skirtingose populiacijose skiriasi nereikšmingai, kai $p \geq \alpha$.

10.11 pav. yra parodyti Chi-kvadrato testo rezultatai 10.1 skyriuje pateikto pavyzdžio metodikoms įvertinti.

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	3,296 ^a	2	,192
Likelihood Ratio	3,332	2	,189
N of Valid Cases	51		

a. 1 cells (16,7%) have expected count less than 5. The minimum expected count is 4,71.

10.11 pav. Chi-kvadrato testo rezultatų lentelės pavyzdys

Matome, kad Chi-kvadrato testo kriterijaus p -reikšmė (*Asymp. Sig. (2-sided)*) tiek pagal Pirsono formulę (*Pearson Chi-Square*), tiek pagal tikėtimumo santykio formulę (*Likelihood Ratio*) yra maždaug vienoda ir viršija reikšmingumo lygmenį 0,05. Vadinasi, nėra esminio skirtumo tarp dviejų analizuojamų metodikų.

10.3.2. Koreliacijos koeficientai

Koreliacijos koeficientas yra tiesinės priklausomybės tarp kintamųjų kiekybinio įvertinimo kriterijus arba ryšio stiprumo matas. Matuojamų pagal rangų skalę kintamųjų (tik skaitmeninių reikšmių) yra skaičiuojamas Spirmeno (*Spearman*) koreliacijos koeficientas, o matuojamų pagal intervalų skalę kintamųjų — Pirsono (*Pearson*) koreliacijos koeficientas. Turinčius dvi kategorijas (binarinius) nominaliuosius kintamuosius galima laikyti ranginiais kintamaisiais.

Koreliacijos koeficientui apskaičiuoti:

- Spragtelėkite dialogo langelio **Crosstabs** mygtuką **Statistics...**
- Atsidariusiame naujame dialogo langelyje **Crosstabs: Statistics** pasirinkite **Correlations**.
- Spragtelėkite dialogo langelio **Crosstabs: Statistics** mygtuką **Continue**, tada — dialogo langelio **Crosstabs** mygtuką **OK**.

Išvestyje bus pateiktos Pirsono (*Pearson's R*) ir Spirmeno (*Spearman Correlation*) koreliacijos koeficientų reikšmės, taip pat kriterijaus p -reikšmė (*Approx. Sig.*), kuria remiantis sprendžiama, ar koreliacija statistiškai

reikšminga: populiacijos kintamieji koreliuoja, kai $p < \alpha$, kintamieji nekoreliuoja, kai $p \geq \alpha$, čia α — nustatytas reikšmingumo lygmuo.

Paskaičiuosime koreliacijos koeficientą tarp mūsų nagrinėjamų kintamųjų: treniruočių metodikos *metod* ir testavimo rezultatų *rezult*. Pirmiausia raidinės sekos kintamajam *rezult* suteiksime skaitmeninį formatą su šiomis kategorijų reikšmėmis: 1 — *geri*, 2 — *vidutin*, 3 — *blogi*. Taigi, abu kintamieji bus ranginiai.

Skaičiavimo rezultatai parodyti 10.12 pav.

Symmetric Measures					
		Value	Asymp. Std. Error ^a	Approx. T ^b	Approx. Sig.
Interval by Interval	Pearson's R	-,037	,142	-,257	,798 ^c
Ordinal by Ordinal	Spearman Correlation	-,048	,145	-,338	,737 ^c
N of Valid Cases		51			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

c. Based on normal approximation.

10.12 pav. Koreliacijos koeficiento skaičiavimo rezultatų pavyzdys

Kadangi kintamieji nepriklauso intervalų skalei, Pirsono koreliacijos koeficiento reikšmės nenagrinėjame. Spirmeno koreliacijos koeficiento reikšmė labai maža (– 0,048), t. y. praktiškai nėra ryšio tarp testavimo rezultatų ir taikytos treniruočių metodikos. Pats gautas rezultatas nėra statistiškai reikšmingas, nes kriterijaus p -reikšmė (*Approx. Sig.*) daug didesnė už nustatytą reikšmingumo lygmenį 0,05 ($0,737 > 0,05$), todėl nulinė hipotezė, teigianti, kad nėra koreliacijos tarp dviejų kintamųjų (koreliacijos koeficientas lygus nuliui) negali būti atmesta. Tokio rezultato, aišku, buvo galima tikėtis, žinant Chi-kvadrato testo rezultatus.

10.3.3. Nominaliųjų kintamųjų ryšio matai

Koreliacijos koeficiento negalima taikyti tarpusavio ryšiui tarp nominaliųjų kintamųjų, turinčių daugiau kaip dvi kategorijas, apibūdinti, nes jie yra tarpusavyje nelyginami, jų negalima surikiuoti. Todėl priklausomybei tarp tokių kintamųjų nustatyti paprastai naudojamas Chi-kvadrato testas. Tačiau SPSS taip pat yra pateikiami kriterijai kiekybiniam ryšio tarp dviejų nominaliųjų kintamųjų įvertinimui. Šie kriterijai parodo dviejų priklausančių nominalinei skalei kintamųjų priklausomumo ar nepriklausomumo laipsnį. Kriterijaus reikšmė, lygi nuliui, atitinka visišką kintamųjų nepriklausomumą, o reikšmė, lygi vienetui — didžiausią priklausomumą. Šie tarpusavio ryšio matai negali turėti neigiamų reikšmių, kadangi negalint išrikiuoti kintamųjų, negalima nustatyti priklausomybės krypties.

Trumpai apibūdinsime šiuos kriterijus, kuriuos galite pasirinkti **Crosstabs: Statistics** dialogo langelio kriterijų grupėje **Nominal**.

Contingency Coefficient (kontingencijos koeficientas)

Šis koeficientas skaičiuojamas pagal formulę

$$c = \sqrt{\frac{\chi^2}{\chi^2 + N}}, \quad (10.4)$$

čia N — bendras stebėjimų skaičius.

Kadangi N visada daugiau už nulį, koeficiento reikšmė mažesnė už vienetą ir priklauso nuo stebėjimų skaičiaus. Todėl pagal šį koeficientą negalima lyginti kelių atvejų su skirtingais N .

Phi (coefficient) – ϕ koeficientas

Koeficientas skaičiuojamas pagal formulę

$$\phi = \sqrt{\frac{\chi^2}{N}}, \quad (10.5)$$

Naudojamas tik tada, kai duomenys aprašomi 2×2 tipo lentelėmis.

Cramer's V (Kramerio V koeficientas)

Lentelėms 2×2 Kramerio V koeficientas sutampa su ϕ koeficientu. Koeficientas skaičiuojamas pagal formulę

$$V = \sqrt{\frac{\chi^2}{N \cdot (k - 1)}}, \quad (10.6)$$

čia k — mažiausias iš eilučių ir stulpelių skaičiaus.

Visi išvardytieji kriterijai yra skaičiuojami remiantis Chi-kvadrato kriterijumi. Kiti matuojamų pagal nominalinę skalę kintamųjų tarpusavio ryšio matai: **Lambda** (lambda, λ), **Goodman and Kruskal's tau** (lambda modifikacija — Gudmano ir Kruskalio tau, τ) ir **Uncertainly coefficient** (neapibrėžtumo koeficientas) koeficientai skaičiuojami remiantis vadinamąja proporcingo klaidos mažinimo koncepcija. Jie įvertina vieno požymio kategorijos

nuspėjamumo santykinį klaidos sumažėjimą, kai žinoma kito požymio kategorija.

Pateiksime pavyzdį. Buvo daroma trijų fakultetų studentų apklausa apie jų domėjimąsi sportu. Anketoje buvo tik vienas klausimas („ar Jūs aktyviai domitės sportu?“), į kurį reikėjo atsakyti „taip“ arba „ne“. Taigi turime du vardinės skalės kintamuosius: *sportas*, kurio reikšmės 1 (*taip*) ir 2 (*ne*) ir *fak_tas*, kurio reikšmėms 1, 2, 3 suteiksime sąlygines žymes — atitinkamai *pirmas*, *antras*, *trečias*. Dialogo langelyje **Crosstabs: Statistics** pasirinksiame visus **Nominal** variantus. Rezultatai parodyti 10.13 pav.

SPORTAS * FAK_TAS Crosstabulation

Count		FAK_TAS			Total
		pirmas	antras	trečias	
SPORTAS	taip	12	15	7	34
	ne	9	2	15	26
Total		21	17	22	60

Directional Measures

			Value	Asymp. Std. Error ^a	Approx. T ^b	Approx. Sig.
Nominal by Nominal	Lambda	Symmetric	,250	,109	2,183	,029
		SPORTAS Dependent	,308	,150	1,749	,080
		FAK_TAS Dependent	,211	,110	1,749	,080
	Goodman and Kruskal tau	SPORTAS Dependent	,207	,092		,002 ^c
		FAK_TAS Dependent	,100	,049		,003 ^c
	Uncertainty Coefficient	Symmetric	,127	,062	2,050	,001 ^d
		SPORTAS Dependent	,166	,080	2,050	,001 ^d
		FAK_TAS Dependent	,104	,050	2,050	,001 ^d

- a. Not assuming the null hypothesis.
b. Using the asymptotic standard error assuming the null hypothesis.
c. Based on chi-square approximation
d. Likelihood ratio chi-square probability.

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	,455	,002
	Cramer's V	,455	,002
	Contingency Coefficient	,414	,002
N of Valid Cases		60	

- a. Not assuming the null hypothesis.
b. Using the asymptotic standard error assuming the null hypothesis.

10.13 pav. Ryšio matų tarp dviejų vardinės skalės kintamųjų skaičiavimo rezultatai

Pirmoje lentelėje (*Crosstabulation*) parodyti gauti dažniai (iš viso buvo apklausta 60 studentų). Antroje (*Directional Measures*) ir trečioje (*Symmetric Measures*) lentelėje pateiktos aksčiau minėtų koeficientų reikšmės.

Atsižvelgiant į tai, kad šių koeficientų reikšmės retai kada būna artimos vienetui, ryši tarp kintamųjų (t. y. tarp studijų tam tikrame fakultete ir aktyviu domėjimusi sportu) galime vertinti kaip vidutinį ir statistiškai reikšmingą — visų koeficientų kriterijaus *p*-reikšmė (*Approx. Sig.*) mažesnė už nustatytą reikšmingumo lygmenį (0,05). Ryšys tarp kintamųjų pagal kriterijus Chi-kvadrato pagrindu kiek stipresnis negu pagal „kryptingus“ kriterijus. Tuo atveju, kai neaišku, kuris kintamasis gali būti priskirtas priklausomam, skaičiuojami abu variantai ir pateikiamas vidurkis (*Symmetric* — antroje lentelėje).

10.3.4. Ranginių kintamųjų ryšio matai

Be dažniausiai taikomo Spirmeno ranginės koreliacijos koeficiento, yra keli koeficientai, kurių skaičiavimai remiasi nukrypimų (vadinamųjų inversijų) nuo nustatytos tvarkos skaičiumi. Inversijų skaičius nustatomas surikiavus didėjimo tvarka vieno iš dviejų kintamųjų, tarp kurių norima nustatyti tarpusavio ryšio laipsnį, reikšmės ir šalia jų užrašant atitinkamas kito kintamojo reikšmes. Antro kintamojo nukrypimų nuo nustatytos tvarkos skaičius ir bus inversijų skaičius. Inversijų ir sutapimų skaičiaus pagrindu (arba kitaip — suderintų ir nesuderintų porų skaičiaus skirtumo pagrindu) skaičiuojamas kintamųjų tarpusavio ryšio koeficientas, kurio reikšmės yra diapazone nuo -1 iki $+1$. Dialogo langelyje **Crosstabs: Statistics, Ordinal** grupėje galite pasirinkti šiuos koeficientus: **Gamma**, **Somers'd**, **Kendall's tau-b** (taikomas kvadratinių dažnių lentelių atveju, t. y., kai eilučių ir stulpelių skaičius vienodas), **Kendall's tau-c** (taikomas bet kokio tipo lentelėms).

10.3.5. Kiti tarpusavio ryšio matai

Eta (Nominal by Interval) koeficientas yra taikomas, kai nepriklausomas kintamasis yra nominalusis arba ranginis (kategorinis kintamasis), o priklausomas kintamasis — intervalinis. Kategorinis kintamasis turi turėti skaitmeninį kodavimą (*Numeric* tipo).

Kappa koeficientas taikomas dviejų ekspertų, vertinančių tą patį objektą ar reiškinį, išvadų suderinamumui nustatyti. Koeficiento reikšmė 1 rodo visišką ekspertų vertinimų sutapimą. **Kappa** koeficientas taikomas tik tada, kai abudu kintamieji turi tas pačias kategorijų reikšmes ir vienodą kategorijų skaičių.

Risk (rizikos laipsnis) yra skaičiuojamas keturlaukėms 2×2 dažnių lentelėms, sudarytoms laikantis tam tikrų toliau išdėstytų reikalavimų. Skaičiuojant rizikos matą, analizuojamas vadinamasis rizikos kintamasis, kuris turi dvi kategorijas ir nurodo, įvyko tam tikras įvykis ar ne. Rizikos kintamasis analizuojamas atžvilgiu priežastinio (nepriklausomo) kintamojo, kuris irgi turi

būti binarinis (turintis dvi kategorijas). Šį teiginį pailiustruosime paprastu pavyzdžiu. Buvo atlikta 90 respondentų apklausa apie polinkį į depresiją. Apklausos rezultatai pateikti 10.1 lentelėje.

10.1 lentelė. Apklausos rezultatai

Depresija	Taip	Ne
Moterys	a = 12	b = 44
Vyrai	c = 3	d = 31

Depresija yra rizikos kintamasis, o lytis — nepriklausomas (priežastinis) kintamasis. SPSS rizikos matas yra apskaičiuojamas pagal šias formules:

$$R_0 = \frac{a \cdot d}{b \cdot c}, \text{ galimybių santykis (Odds Ratio),} \quad (10.7)$$

$$R_1 = \frac{a \cdot (c + d)}{c \cdot (a + b)}, \text{ santykinės rizikos koeficientas,} \quad (10.8)$$

$$R_2 = \frac{b \cdot (c + d)}{d \cdot (a + b)}, \text{ santykinės rizikos koeficientas,} \quad (10.9)$$

čia a, b, c, d — lentelėje pateikti dažniai.

Dialogo langelyje **Crosstabs: Statistics** pasirinkę variantą **Risk** gausime 10.14 pav. parodytus rezultatus

LYTIS * depresija Crosstabulation

Count		depresija		Total
		taip	ne	
LYTIS	moterys	12	44	56
	vyrai	3	31	34
Total		15	75	90

Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for LYTIS (moterys / vyrai)	2,818	,733	10,828
For cohort depresija = taip	2,429	,738	7,993
For cohort depresija = ne	,862	,725	1,024
N of Valid Cases	90		

10.14 pav. Rizikos laipsnio įvertinimo pavyzdys

Pirmasis *Risk Estimate* (rizikos įvertinimas) lentelėje pateiktas dydis — galimybių santykis (*Odds Ratio*) — rodo, kad rizika susirgti depresija moterims yra 2,818 karto didesnė, negu vyrams. Kiti du koeficientai vadinami santykinės rizikos koeficientais.

Kad apskaičiuoti santykinės rizikos koeficientai būtų teisingi, reikia laikytis šių taisyklių:

- Nepriklausomą (priežastinį) kintamąjį dėkite į lentelės eilučių (**Row(s)**) sąrašą, o rizikos kintamąjį — į stulpelių (**Column(s)**) sąrašą.
- Pirmoje lentelės eilutėje turi būti didžiausios rizikos grupė.
- Pirmame lentelės stulpelyje turi būti duomenys, jeigu įvykis įvyks.

McNemar'o testas yra taikomas tada, kai tiriamas dvireikšmis (binarinis) kintamasis (nuostata, gebėjimai, sveikata ir t. t.) matuojamas du kartus — iki poveikio ir po jo. Plačiau apie šį testą — 12 skyriuje.

Cochran's and Mantel-Haenszel testas. Taikant šį testą, apskaičiuojamas galimybių santykis kintamiesiems, kurie aprašomi 2×2 lentelėmis, pagal kiekvieną kintamojo, nurodyto **Layer** (sluoksnis) laukelyje, kategoriją ir patikrinama, ar šios kategorijos esmiškai skiriasi savo galimybių santykiu (kiek jis skiriasi nuo 1 ar kito nustatyto dydžio).

Šio testo rezultatams gauti:

- Įdėkite kintamuosius į dialogo langelio **Crosstabs** laukelius **Row(s)** ir **Column(s)** anksčiau aprašyta tvarka.
- Įdėkite kintamąjį, pagal kurio kategorijas bus skaičiuojamas galimybių santykis, į dialogo langelio **Crosstabs** laukelį **Layer**.
- Spragtelėkite dialogo langelio **Crosstabs** mygtuką **Statistics...**
- Atsidariusiame naujame dialogo langelyje **Crosstabs: Statistics** pasirinkite **Cochran's and Mantel-Haenszel statistics**.
- Laukelyje **Test common odds ratio equals** (bendras galimybių santykis) palikite nustatytąją reikšmę 1.
- Spragtelėkite mygtuką **Continue**, tada — **OK** dialogo langelyje **Crosstabs**.

Pavyzdžiui, suskirstę apklausos duomenis (duomenys hipotetiniai!) į dvi amžiaus kategorijas (iki 20 m. ir daugiau kaip 20 m.) ir įkėlę kintamąjį *amzius* į dialogo langelio **Crosstabs** laukelį **Layer**, gautume 10.15 pav. parodytus rezultatus. Lentelėje *Mantel-Haenszel Common Odds Ratio Estimate* (Mantelio ir Haenzelio bendro galimybių santykio įvertis) pateikta kriterijaus *p*-reikšmė (*Asymp. Sig. (2-sided)*) rodo, kad abiejų amžiaus kategorijų galimybių santykis reikšmingai skiriasi nuo 1. Panašus rezultatas pateiktas ir lentelėje *Test of Homogeneity of the Odds Ratio*.

LYTIS * DEPRES * AMZIUS Crosstabulation

Count

AMZIUS			DEPRES		Total
			taip	ne	
<20	LYTIS	moterys	12	44	56
		vyrų	3	31	34
	Total		15	75	90
>20	LYTIS	moterys	8	40	48
		vyrų	2	30	32
	Total		10	70	80

Tests for Homogeneity of the Odds Ratio

Statistics		Chi-Squared	df	Asymp. Sig. (2-sided)
Conditional	Cochran's	4,322	1	,038
Independence	Mantel-Haenszel	3,406	1	,065
Homogeneity	Breslow-Day	,003	1	,954
	Tarone's	,003	1	,954

Under the conditional independence assumption, Cochran's statistic is asymptotically distributed as a 1 df chi-squared distribution, only if the number of strata is fixed, while the Mantel-Haenszel statistic is always asymptotically distributed as a 1 df chi-squared distribution. Note that the continuity correction is removed from the Mantel-Haenszel statistic when the sum of the differences between the observed and the expected is 0.

Mantel-Haenszel Common Odds Ratio Estimate

Estimate			2,892
ln(Estimate)			1,062
Std. Error of ln(Estimate)			,528
Asymp. Sig. (2-sided)			,044
Asymp. 95% Confidence Interval	Common Odds Ratio	Lower Bound	1,027
		Upper Bound	8,143
	ln(Common Odds Ratio)	Lower Bound	,027
		Upper Bound	2,097

The Mantel-Haenszel common odds ratio estimate is asymptotically normally distributed under the common odds ratio of 1,000 assumption. So is the natural log of the estimate.

10.15 pav. Galimybių santykio pagal atskiras kategorijas skaičiavimo pavyzdys

11. HIPOTEZIŲ APIE VIDURKIŲ LYGYBĘ TIKRINIMAS

Hipotezių apie populiacijų vidurkių lygybę tikrinimas yra vienas iš dažniausiai taikomų statistinės analizės metodų. Remiantis gautais rezultatais nulinė hipotezė H_0 , kad nėra esminio skirtumo tarp populiacijų, iš kurių paimtos imtys, vidurkių yra priimama arba atmetama ir priimama jai alternatyvioji hipotezė H_1 , kad skirtumas tarp populiacijos vidurkių yra reikšmingas.

Šiame skyriuje nagrinėsime populiacijų vidurkių palyginimo metodus, kurie taikomi, kai kintamieji turi normalųjį skirstinį. Jeigu ši sąlyga netenkinama, taikomi neparametriniai imčių palyginimo metodai.

Lyginant populiacijų vidurkius, galimos šios situacijos:

1. Nagrinėjame dvi nepriklausomas imtis.
2. Nagrinėjame dvi priklausomas imtis.
3. Nagrinėjame daugiau kaip dvi nepriklausomas imtis.
4. Nagrinėjame daugiau kaip dvi priklausomas imtis.

Atitinkamai taikomi šie statistiniai testai:

1. **t**-testas nepriklausomoms imtimis (Stjudento testas).
2. Porinių imčių Stjudento **t**-testas.
3. Vieno faktoriaus dispersinė analizė.
4. Blokuotųjų duomenų vienfaktorinė dispersinė analizė.

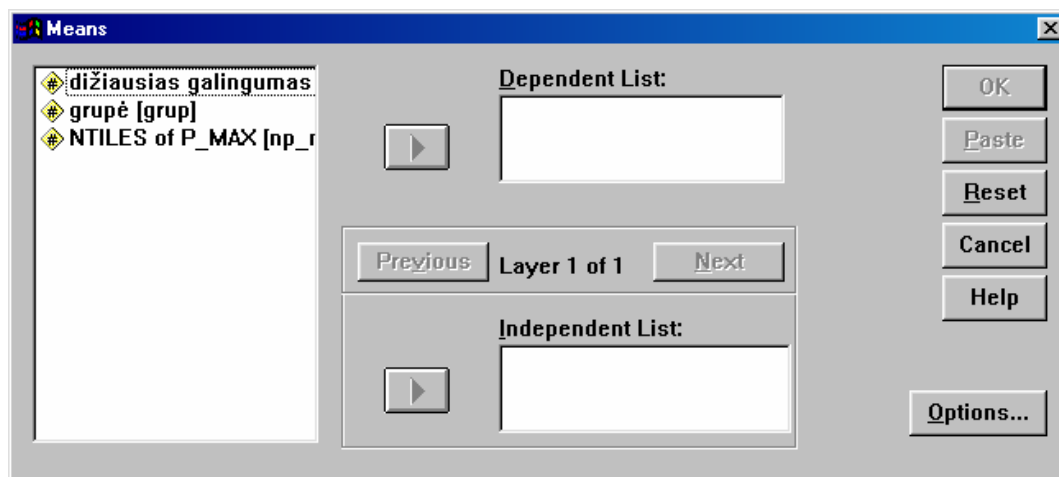
Šiame skyriuje nagrinėjami **t**-testai, kurie pasirenkami komandomis **Analyze → Compare Means** (vidurkių palyginimas). 3 ir 4 p. nurodyti testai aprašomi 15 skyriuje, skirtame dispersinei analizei.

Meniu **Analyze → Compare Means**, be nurodytų testų, dar yra komanda **Means...** (vidurkiai), kurią nurodžius galima apskaičiuoti nagrinėjamų kintamųjų vidurkius ir kitas skaitines charakteristikas atskirai pagal kategorinio kintamojo kategorijas. Nagrinėjami kintamieji turi būti skaitmeniniai, o kategorinis kintamasis — skaitmeninis arba trumpos raidinės-skaitmeninės sekos tipo.

Kintamųjų vidurkiams ir kitoms skaitinėms charakteristikoms atskirai pagal kategorinio kintamojo kategorijas apskaičiuoti:

- Atidarykite bylą su analizuojamais duomenimis.
- Nurodykite komandas **Analyze → Compare Means → Means...**
- Dialogo langelyje **Means** (11.1 pav.) įkelkite pasirinktus kintamuosius į analizuojamų kintamųjų sąrašą **Dependent List** (priklausomi kintamieji).
- Vienu iš dviejų būdų nurodykite kategorinius kintamuosius:
 - Pasirinkite vieną ar kelis kategorinius kintamuosius ir įkelkite į sąrašą **Independent List** (nepriklausomi kintamieji). Rezultatą gausite atskirai kiekvienam kategoriniam kintamajam.
 - Kategorinius kintamuosius priskirkite skirtingiems sluoksniams (**Layer**). Tam, įkėlę pirmą kategorinį kintamąjį į sąrašą **Independent List**, spragtelėkite **Layer** mygtuką **Next**, tada į sąrašą **Independent List**

įkelkite antrą kategorinį kintamąjį ir t. t. Rezultatą gausite vienoje lentelėje kiekvienam kategorinių kintamųjų deriniui.



11.1 pav. Dialogo langelis *Means*

- Spragtelėkite dialogo langelio mygtuką **Options...** Atsidarys naujas dialogo langelis **Means: Options**, kuriame galite pasirinkti šias skaitines tiriamų kintamųjų charakteristikas atskirai kiekvienai kategorinio kintamojo kategorijai: **Number of Cases** (stebėjimų skaičių), **Mean** (vidurki), **Median** (mediana), **Grouped Median** (suskirstytų duomenų mediana), **Std. Error of Mean** (standartinę vidurkio paklaidą), **Sum** (sumą), **Minimum** (mažiausią reikšmę), **Maximum** (didžiausią reikšmę), **Range** (plotį), **Variance** (dispersiją), **Standart Deviation** (standartinį nuokrypį), **Kurtosis** (ekscesą), **Standart Error of Kurtosis** (standartinę eksceso paklaidą), **Skewness** (asimetrijos koeficientą), **Standart Error of Skewness** (standartinę asimetrijos koeficiento paklaidą), **First** (pirmą kintamojo reikšmę grupėje), **Last** (paskutinę kintamojo reikšmę grupėje), **Geometric Mean** (geometrinį vidurki), **Harmonic Mean** (harmoninį vidurki), **Percentage of total Sum** (procentą nuo bendros sumos), **Percentage of total N** (procentą nuo bendro kiekio).
- Perkelkite pasirinktus parametrus į dialogo langelio **Cell Statistics** sąrašą. Kokia tvarka bus išrikiuoti parametrai **Cell Statistics** sąraše, tokia jie bus pateikti išvesties lentelėje.
- Pažymėkite laukelį **Anova table and eta**, jeigu norite atlikti vienfaktorinę dispersinę analizę. Tačiau, patartina naudoti daug platesnių galimybių testą, aprašytą 11.3 skyriuje.
- Spragtelėkite mygtuką **Continue**, tada — **OK** dialogo langelyje **Means**.

11.1. t-TESTAS NEPRIKLAUSOMOMS IMTIMS

t-testas nepriklausomoms imtims (*Independent-samples T test* arba *Two-samples T test*) taikomas kintamųjų, stebimų dviejose populiacijose, vidurkių palyginimui šiomis sąlygomis:

- Matuojami kintamieji laikomi normaliai pasiskirsčiusiais. Tačiau, kai imties tūris pakankamai didelis, t-testas nepriklausomoms imtims nėra kritiškas nukrypimams nuo šio reikalavimo. Tikrinant pasiskirstymą (histogramą) vizualiai, reikia įsitikinti, kad jis yra simetriškas ir kad nėra pavienių išskirčių.
- Imtys iš populiacijos paimitos atsitiktinės atrankos būdu.

Pateikiami du t-testo variantai priklausomai nuo populiacijų dispersijų:

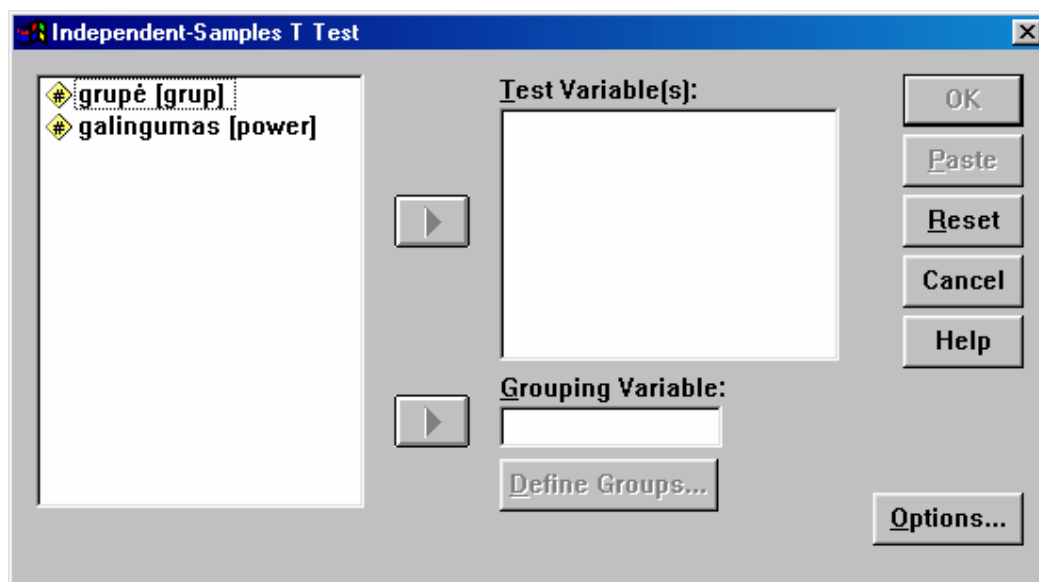
- Kai laikoma, kad dispersijos skiriasi statistiškai nereikšmingai.
- Kai laikoma, kad dispersijos skiriasi statistiškai reikšmingai.

Ar dispersijas galima laikyti lygiomis, ar jos reikšmingai skiriasi sprendžiama iš *Levene* dispersijų homogeniškumo testo rezultatų.

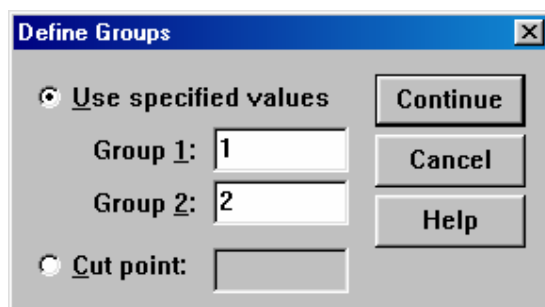
Skaitmeninės tiriamo kintamojo reikšmės (stebėjimai) duomenų rinkmenoje turi būti įrašytos viename stulpelyje. Kategorinis kintamasis, kuris skiria šias reikšmes į dvi grupes, gali būti skaitmeninio formato (pvz., reikšmės 1 ir 2) arba trumpos raidinės-skaitmeninės sekos formato (pvz., *taip* ir *ne*).

t-testo nepriklausomoms imtims rezultatams gauti:

- Atidarykite bylą su analizuojamais duomenimis.
- Nurodykite komandas *Analyze → Compare Means → Independent-Samples T Test...*
- Dialogo langelyje *Independent-Samples T Test* (11.2 pav.) įkelkite pasirinktą kintamąjį į testuojamų kintamųjų sąrašą *Test Variable(s)*. Jeigu pasirinkote kelis kintamuosius, bus atskirai apskaičiuoti kiekvieno kintamojo t-testo rezultatai.
- Kategorinį kintamąjį įkelkite į laukelį *Grouping Variable*.
- Spragtelėkite dialogo langelio mygtuką *Define Groups* (nustatyti grupes). Naujame dialogo langelyje *Define Groups* (11.3 pav.) galite nurodyti dvi kategorinio kintamojo reikšmes. Jeigu kintamasis turi daugiau negu dvi reikšmes, nenurodytos reikšmės nebus analizuojamos. Jeigu turite skaitmeninį kategorinį kintamąjį, pažymėję laukelį *Cut point* (pjūvio taškas) galite nurodyti reikšmę, kuri padalys kategorinį kintamąjį į dvi kategorijas: visos reikšmės, mažesnės už nurodytą, sudarys vieną grupę, reikšmės, didesnės už nurodytą — kitą grupę. Pavyzdžiui, taip galima sudaryti dvi amžiaus grupes iš detalaus amžiaus sąrašo.
- Spragtelėkite mygtuką *Continue*, tada — *OK* dialogo langelyje *Independent-Samples T Test*.



11.2 pav. Dialogo langelis *Independent-Samples T Test*



11.3 pav. Dialogo langelis *Define Groups* grupėms nustatyti

Nepriklausomų imčių *t*-testo rezultatų analizei pasitelksime pavyzdį. Veloergometru buvo testuojamos dvi skirtingo treniruotumo sportininkų grupės, po 30 sportininkų kiekvienoje. Tyrimo tikslas — nustatyti ar gauti galingumo rodikliai skiriasi reikšmingai. Kintamąjį, kurio reikšmės yra abiejų grupių testavimo rezultatai, pavadinsime *power* (žymena — *galingumas*), o kategorinį kintamąjį, kuris parodo, kuriai grupei priklauso konkretus testavimo rezultatas, pavadinsime *group* (žymena — *grupė*). Duomenų rinkmenos fragmentas parodytas 11.4 pav.

Kintamąjį *power* įkeliamo į dialogo langelio *Independent-Samples T Test* laukelį *Test Variable(s)*, o kintamąjį *group* — į laukelį *Grouping Variable*. Dialogo langelyje *Define Groups*, laukelyje *Group 1* įrašome 1, o laukelyje *Group 2* įrašome 2.

	group	power	var	var	var	var
20	1	761,00				
21	1	768,00				
22	1	771,00				
23	1	753,00				
24	1	756,00				
25	1	772,00				
26	1	761,00				
27	1	794,00				
28	1	779,00				
29	1	782,00				
30	1	785,00				
31	2	619,00				
32	2	643,00				
33	2	644,00				
34	2	629,00				
35	2	640,00				

11.4 pav. Dviejų sportininkų grupių veloergometro testo rezultatai (fragmentas)

Gauti rezultatai parodyti 11.5 pav. Pimoje a) lentelėje — *Group Statistics* (grupių statistika) pateikiamos kiekvienos sportininkų grupės charakteristikos: stebėjimų skaičius (*N*), vidurkis (*Mean*), standartinis nuokrypis (*Std. Deviation*), standartinė paklaida (*Std. Error Mean*).

Group Statistics

	grupė	N	Mean	Std. Deviation	Std. Error Mean
galingumas	1	30	768,8000	15,56831	2,84237
	2	30	638,1000	16,41982	2,99783

a)

Independent Samples Test

		Levene's Test for Equality of Variances	t-test for Equality of Means						
		Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
galingumas	Equal variances assumed	,705	31,638	58	,000	130,7000	4,13111	122,43	138,97
	Equal variances not assumed		31,638	57,836	,000	130,7000	4,13111	122,43	138,97

b)

11.5 pav. Nepriklausomų imčių *t*-testo rezultatai

Antroje b) lentelėje (*Independent Samples Test*) pateikiama:

- **Levene** testo dispersijų lygybei nustatyti rezultatas. Nulinė hipotezė, kad dispersijos statistiškai reikšmingai nesiskiria (*Equal variances assumed*), atmetama ir priimama alternatyvioji hipotezė, kad dispersijos skiriasi statistiškai reikšmingai (*Equal variances not assumed*), kai Levene testo *p*-reikšmė (*Sig.*) mažesnė už nustatytą reikšmingumo lygmenį (0,05). Šiame pavyzdyje dispersijų skirtumas yra nereikšmingas ($p = 0,705$), todėl turi būti pasirenkamas variantas *Equal variances assumed*.
- *t*-kriterijaus reikšmė *t*, laisvės laipsnių skaičius *df*, *p*-reikšmė — *Sig. (2-tailed)*, vidurkių skirtumas (*Mean Difference*), vidurkių skirtumo standartinė paklaida (*Std. Error Difference*) ir pasikliautinis intervalas (*95% Confidence Interval of the Difference*).

Šiame pavyzdyje gautas reikšmingas skirtumas tarp dviejų sportininkų grupių galingumo rodiklių, nes gauta *t*-testo *p*-reikšmė (*Sig. (2-tailed)*) mažesnė už nustatytą reikšmingumo lygmenį (lygi nuliui trijų ženklų po kablelio tikslumu).

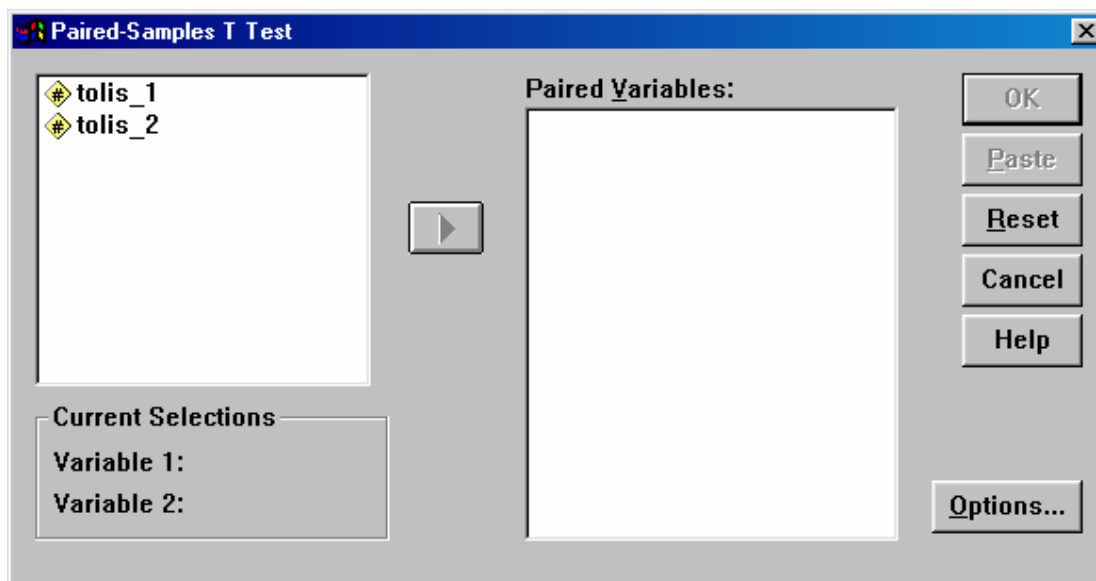
Spragtelėję dialogo langelio ***Independent-Samples T Test*** mygtuką ***Options...*** atsidariusiame dialogo langelyje ***Independent-Samples T Test: Options*** galite pakeisti nustatytą skirtumo pasikliautinojo intervalo patikimumo lygmenį.

11.2. *t*-TESTAS PRIKLAUSOMOMS IMTIMS

Sporte dažnai tenka pakartotinai matuoti tų pačių sportininkų tuos pačius rodiklius po tam tikro laiko (pvz., sportininkai testuojami prieš treniruočių ciklą ir po jo). Tokių testavimų tikslas — nustatyti, ar reikšmingai pagerėjo sportininkų rodikliai po treniruočių ciklo. Tam yra taikomas porinis *t*-testas (***Paired-samples T test*** arba ***Dependent T test***). Prielaidos yra tos pačios kaip ir *t*-testo nepriklausomoms imtims. Bet nuo *t*-testo nepriklausomoms imtims šis visų pirma skiriasi tuo, kad duomenys yra susiję — turime tos pačios grupės matavimų poras (kiekviena matavimų pora — tai to pačio sportininko testavimo rezultatai eksperimento pradžioje ir pabaigoje). Kiekviena matavimų pora turi būti tas pats stebėjimas duomenų rinkmenoje.

t-testo priklausomoms imtims rezultatams gauti:

- Atidarykite bylą su analizuojamais duomenimis.
- Nurodykite komandas ***Analyze → Compare Means → Paired-Samples T Test...***
- Dialogo langelio ***Paired-Samples T Test*** (11.6 pav.) pradiniam kintamųjų sąrašę pažymėkite du kintamuosius ir perkeltkite į testuojamų kintamųjų sąrašę ***Paired Variables***. Jeigu suklydote ir perkėlėte ne tuos kintamuosius, pažymėkite juos ir spragtelėję mygtuką su pasikeitusios krypties trikampę rodykle grąžinkite į bendrą kintamųjų sąrašę.
- Spragtelėkite mygtuką ***OK***.



11.6 pav. Dialogo langelis *Paired-Samples T Test*

Porinio *t*-testo pateiktų rezultatų analizei pasitelksime pavyzdį. Grupės berniukų šolio į tolį rezultatai buvo užfiksuoti kontrolinių varžybų metu ir po dviejų mėnesių treniruočių. Norima įsikinti, kad sportininkų rezultatų pagerėjimas yra patikimas. Kontrolinių varžybų rezultatus pavadinsime kintamuoju *tolis_1* ir įrašysime į vieną duomenų rinkmenos stulpelį. Po dviejų mėnesių treniruočių pasiektus rezultatus pavadinsime *tolis_2* ir įrašysime į kitą duomenų rinkmenos stulpelį. Abu kintamuosius įkeliamo į dialogo langelio *Paired-Samples T Test* laukelį *Paired Variables*. Spragtelėjus mygtuką **OK**, rodinio (*Viewer*) lange bus pateikti porinio *t*-testo rezultatai (11.7 pav.).

Pimoje *a*) lentelėje – *Paired Samples Statistics* (porinių imčių statistika) pateiktos šios kiekvieno testo skaitinės charakteristikos: vidurkis (*Mean*), stebėjimų skaičius (*N*), standartinis nuokrypis (*Std. Deviation*), standartinė paklaida (*Std. Error Mean*).

Antroje *b*) lentelėje – *Paired Samples Correlations* (porinių imčių koreliacija) pateikta Pirsono koreliacijos koeficiento reikšmė ir *p*-reikšmė (*Sig.*), iš kurios sprendžiama, ar koreliacija statistiškai reikšminga. Šiame pavyzdyje koreliacijos koeficientas lygus 0,963 ir koreliacija statistiškai reikšminga (*Sig.* lygus nuliui trijų ženklų po kablelio tikslumu). Taigi tarp kintamųjų yra labai stipri teigiama koreliacija.

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	TOLIS_1	5,5530	30	,26364	,04813
	TOLIS_2	5,6257	30	,30382	,05547

a)

Paired Samples Correlations

	N	Correlation	Sig.
Pair 1 TOLIS_1 & TOLIS_2	30	,963	,000

b)

Paired Samples Test

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
Pair 1 TOLIS_1 - TOLIS_2	-,0727	,08686	,01586	-,1051	-,0402	-4,582	29	,000

c)

11.7 pav. Porinio *t*-testo rezultatai

Trečioje c) lentelėje – *Paired Samples Test* (porinių imčių testas) pateikiama:

- Skirtumo vidurkis (*Mean*), standartinis nuokrypis (*Std. Deviation*), standartinė paklaida (*Std. Error Mean*) ir pasikliautinis intervalas (*95% Confidence Interval of the Difference*).
- t*-kriterijaus reikšmė *t*, laisvės laipsnių skaičius *df*, *p*-reikšmė (stebimasis reikšmingumo lygmuo) – *Sig. (2-tailed)*.

Šiame pavyzdyje gauta *t*-testo *p*-reikšmė (*Sig. (2-tailed)*) mažesnė už nustatytą reikšmingumo lygmenį (lygi nuliui trijų ženklų po kablelio tikslumu), tai rodo statistiškai reikšmingą vidurkių skirtumą (sportininkų rezultatai pagerėjo statistiškai reikšmingai).

Spragtelėję dialogo langelio *Paired-Samples T Test* mygtuką *Options...* atsidariusiame dialogo langelyje *Paired-Samples T Test: Options* galite pakeisti nustatytą skirtumo pasikliautinio intervalo patikimumo lygmenį.

12. NEPARAMETRINIAI KRITERIJAI

Neparametriniai kriterijai taikomi tais atvejais, kai duomenys nėra pasiskirstę pagal normalųjį dėsnį arba priklauso rangų, o ne intervalų skalei. Šiame skyriuje nagrinėjami kriterijai, kuriems pakanka kintamųjų tolydumo. Dar viena teigiama neparametrinių kriterijų savybė — jie gali būti taikomi mažoms imtims. Tačiau, jeigu stebimi normalieji kintamieji, o imtys didelės, tada geriau taikyti parametrinius kriterijus, nes jie yra galingesni. SPSS yra pateikiama vartotojui nemaža neparametrinių testų įvairovė. Populiariausi yra dviejų ir daugiau nepriklausomų arba priklausomų imčių palyginimo kriterijai, Chi-kvadrato (χ^2) kriterijus:

- **Chi-Square Test (Chi-kvadrato testas).** Šiuo testu galima patikrinti, kiek reikšmingai skiriasi stebimi vardinės skalės arba pateikti dažnių lentelėmis tyrimo duomenys nuo planuotų prieš tyrimą.
- **1-Sample Kolmogorov-Smirnov Test** (vienos imties Kolmogorovo ir Smirnov testas). Šiuo testu galima patikrinti, ar realus skirstinys atitinka vieną iš šių skirstinių: normalųjį, tolygųjį, eksponentinį ar Puasono.
- **2 Independent Samples Tests** (testai dviem nepriklausomoms imtims):
 - **Mann-Whitney-U-test** (Mann'o ir Whitney U testas).
 - **Moses extreme reaction test** (Mozes'o ekstreminės reakcijos testas).
 - **Z Kolmogorov-Smirnov test** (Kolmogorovo ir Smirnov Z testas).
 - **Wald-Wolfowitz runs test** (Wald'o ir Wolfowitz'o serijų testas).
- **K Independent Samples Tests** (k nepriklausomų imčių testai):
 - **H Kruskal-Wallis test** (Kruskal'o ir Wallis'o testas).
 - **Median test** (medianos testas).
- **2 Related Samples Tests** (testai dviem priklausomoms imtims):
 - **Wilcoxon test** (Wilcoxon'o testas).
 - **Sign test** (ženklų testas).
 - **McNemar test** (McNemar'o testas).
- **K Related Samples Test** (k priklausomų imčių testai):
 - **Friedman test** (Friedman'o testas).
 - **W Kendall test** (W Kendall'o testas).
 - **Q Cochran test** (Q Cochran'o testas).

Neparametrinių hipotezių atveju dažniausiai lyginami skirstiniai. Beveik visi šio skyriaus kriterijai galingiausi, kai alternatyva „skirstiniai skiriasi“ suprantama, kad jie skiriasi tik postūmio parametru (vidurkiu, mediana).

12.1. DVIEJŲ NEPRIKLAUSOMŲ IMČIŲ PALYGINIMAS

SPSS Jūs galite pasirinkti keturis skirtumo reikšmingumo tarp dviejų nepriklausomų imčių (grupių) nustatymo testus.

12.1.1. Mann'o ir Whitney U testas

Tai žinomiausias ir plačiausiai taikomas dviejų nepriklausomų imčių neparametrinio palyginimo testas. Mann'o ir Whitney U testas yra Stjudento t -testo dviem nepriklausomoms imtims neparametrinis analogas. Jis yra grindžiamas jungtinės imties analize. Mann'o ir Whitney U testo esmę trumpai galima paaiškinti taip: jungtiniai abiejų imčių duomenys, gauti testuojant tuo pačiu testu, išdėstomi pagal rangus. Tada kiekvienos imties rangai (gauti ranguojant jungtinius duomenis) sumuojami atskirai. Jeigu teisinga nulinė hipotezė, t. y. kintamųjų skirstiniai vienodi, rangai bus pasiskirstę tarp grupių atsitiktiniu būdu. Kiekvienos imties rangų sumos pagrindu skaičiuojama Mann'o ir Whitney kriterijaus U statistika, kuria remiantis priimamas sprendimas dėl statistinės hipotezės:

$$\begin{cases} H_0: \text{kintamųjų skirstiniai yra vienodi,} \\ H_1: \text{kintamųjų skirstiniai nėra vienodi.} \end{cases}$$

Daroma išvada, kad skirstiniai skiriasi, jeigu p -reikšmė mažesnė už α ir skirstiniai nesiskiria, jeigu p -reikšmė didesnė už α arba jai lygi, čia α — reikšmingumo lygmuo.

12.1.2. Kolmogorovo ir Smirnov Z testas

Kolmogorovo ir Smirnov testą tikslinga naudoti tada, kai testuojamų kintamųjų kategorijų skaičius yra nedidelis. Tokiu atveju, naudojant Mann'o ir Whitney testą vienodi rangai būtų suteikti keliems vienodiems stebėjimams, t. y. atsirastų rangų sekos neapibrėžtumas. Kolmogorovo ir Smirnov testas paremtas didžiausio skirtumo tarp abiejų imčių sukaupytųjų dažnių skirstinių apskaičiavimu. Iš to sprendžiama, ar skirtumas tarp imčių yra reikšmingas.

12.1.3. Wald'o ir Wolfowitz'o testas

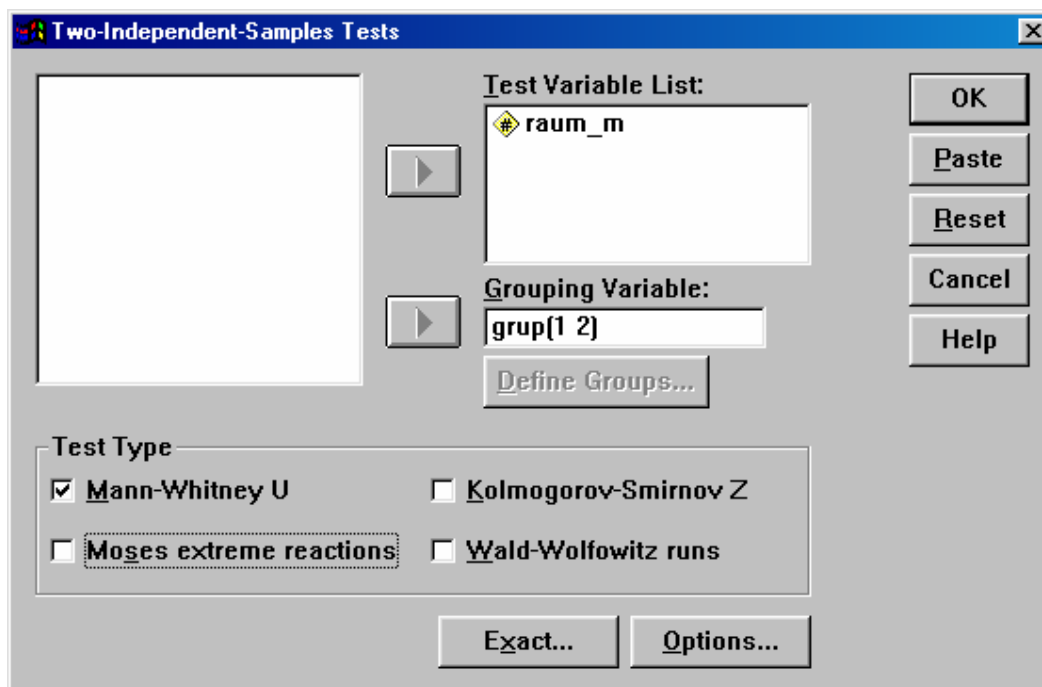
Testuojant Wald'o ir Wolfowitz'o testu taip pat yra sudaroma jungtinė variacinė eilutė. Kriterijaus statistika yra serijų skaičius r . Šiuo atveju serija yra tos pačios imties narių seka, prieš kurią ir po kurios yra kitos imties nariai. Remiantis kriterijaus statistika, nulinė hipotezė H_0 atmetama arba neatmetama. Wald'o ir Wolfowitz'o testas yra tinkamesnis, kai skiriasi lyginamųjų skirstinių sklaidos ir formos charakteristikos. Jeigu skirstinių sklaida ir forma maždaug vienoda, o skiriasi jų medianos, tai labiau taikytinas Mann'o ir Whitney U testas.

12.1.4. Mozes'o testas

Mozes'o testu yra tikrinamas dviejų suranguotų (ranguojant jungtinius duomenis) imčių pločio skirtumas. Viena imtis yra apibrėžiama kaip kontrolinė, kita – kaip eksperimentinė. Kad būtų sumažinta išskirties reikšmių įtaka, iš abiejų kontrolinės grupės skirstinio pusių pašalinama 5% reikšmių.

Dviejų nepriklausomų imčių palyginimo testo rezultatams gauti:

- Atidarykite bylą su analizuojamais duomenimis.
- Nurodykite komandas *Analyze → Nonparametric Tests → 2 Independent Samples...*
- Dialogo langelyje *Two-Independent-Samples Tests* (12.1 pav.) įkelkite vieną ar kelis kintamuosius į testuojamų kintamųjų sąrašą *Test Variable List*. Jeigu pasirinkote kelis kintamuosius, testo rezultatai bus apskaičiuoti atskirai kiekvienam kintamajam.
- Kategorinį kintamąjį įkelkite į laukelį *Grouping Variable*.
- Spragtelėkite dialogo langelio mygtuką *Define Groups* (nustatyti grupes). Naujame dialogo langelyje *Define Groups* nurodykite dvi kategorinio kintamojo reikšmes, pagal kurias identifikuojamos dvi lyginamos testuojamo kintamojo grupės.



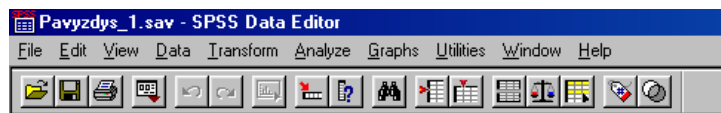
12.1 pav. Dialogo langelis *Two-Independent-Samples Tests*

- Nustatytasis testo variantas dialogo langelyje *Two-Independent-Samples Tests* yra *Mann'o-Whitney U* testas. Remdamiesi anksčiau pateikta

trumpa testų charakteristika, galite pasirinkti tinkamiausią konkrečiam atvejui testą.

Pavyzdys. Lyginamos dvi po 20 sportininkų grupės pagal raumenų masę. Kadangi histograma ir Shapiro ir Wilk testas mažoms imtims (žr. 9.3 skyrių) rodo akivaizdų duomenų skirstinio skirtumą nuo normaliojo, vietoj Stjudento *t*-testo naudosime neparametrinius testus. Nesant bent kiek aiškesnio duomenų pasikartojimo, paliekame nustatytąjį Mann'o ir Whitney testą.

- Kintamąjį, kurio reikšmės yra abiejų grupių testavimo rezultatai, pavadinsime *raum_m*, o kategorinį kintamąjį, kuris parodo, kuriai grupei priklauso konkretus testavimo rezultatas, pavadinsime *grup*. Duomenų rinkmenos fragmentas parodytas 12.2 pav.
- Kintamąjį *raum_m* įkeliamo į dialogo langelio **Two-Independent-Samples Tests** laukelį **Test Variable List**, o kintamąjį *grup* – į laukelį **Grouping Variable**. Spragtelėję mygtuką **Define Groups...** atsidariusio dialogo langelio **Define Groups** laukelyje **Group 1** įrašome 1, o laukelyje **Group 2** įrašome 2.



	raum_m	grup	var	var	var	var
1	49,60	1				
2	58,30	1				
3	51,10	1				
4	41,70	1				
5	37,90	1				
6	42,70	1				
7	44,40	1				
8	40,50	1				
9	40,90	1				
10	38,10	1				
11	47,30	1				
12	40,30	1				
13	49,20	1				
14	58,10	1				
15	51,50	1				
16	41,30	1				
17	38,00	1				
18	42,50	1				
19	44,10	1				
20	43,00	1				
21	42,80	2				
22	46,70	2				
23	52,30	2				

12.2 pav. Dviejų sportininkų grupių raumenų masės rodikliai (duomenų rinkmenos fragmentas)

Pasirinkto Mann'o ir Whitney testo rezultatai parodyti 12.3 pav. Lentelėje *a) Ranks* (rangai) pateikiamas abiejų grupių stebėjimų skaičius, rangų vidurkiai (*Mean Rank*), rangų suma (*Sum of Ranks*). Čia didesnėms reikšmėms yra suteikiami mažesni rangai. Lentelėje *b) Test Statistics* (Testo statistika) pateikiama parametro *U* reikšmė (*Mann-Whitney U*), mažesnioji rangų sumos reikšmė (*Wilcoxon W*), parametro *Z* reikšmė, *p*-reikšmė (*Asymp. Sig. (2-tailed)*), kuria reikėtų remtis, kai stebėjimų skaičius daugiau kaip 30, tiksli *p*-reikšmė (*Exact Sig. [2*(1-tailed Sig.)]*), kuria reikėtų remtis, kai stebėjimų skaičius mažiau kaip 30. Kadangi *p*-reikšmė didesnė už reikšmingumo lygmenį, nulinės hipotezės atmesti nėra pagrindo, t. y. abiejų sportininkų grupių duomenys reikšmingai nesiskiria.

Mann-Whitney Test

Ranks

	GRUP	N	Mean Rank	Sum of Ranks
RAUM_M	1	20	19,60	392,00
	2	20	21,40	428,00
	Total	40		

a)

Test Statistics^b

	RAUM_M
Mann-Whitney U	182,000
Wilcoxon W	392,000
Z	-,487
Asymp. Sig. (2-tailed)	,626
Exact Sig. [2*(1-tailed Sig.)]	,640 ^a

a. Not corrected for ties.

b. Grouping Variable: GRUP

b)

12.3 pav. Mann'o ir Whitney testo rezultatai

Spragtelėję mygtuką **Options...**, galite:

- Pasirinkti kai kurias papildomas **Descriptive** charakteristikas (vidurki, mažiausią reikšmę, didžiausią reikšmę, standartinę nuokrypį, stebėjimų skaičių) bei kvartilius (**Quartiles**) – 25, 50 ir 75 procentilius. Tačiau reikia turėti omeny, kad šios charakteristikos turi prasmę tik testuojamam kintamajam, bet ne kategoriniam.
- Nurodyti, kaip traktuoti praleistas reikšmes: **Exclude cases test-by-test** — įvertinamos kiekvieno nustatyto testo praleistos reikšmės (nustatytasis variantas), **Exclude cases listwise** — stebėjimai, kuriuose pasitaiko kurio nors kintamojo praleistų reikšmių, pašalinami iš visų skaičiavimų.

Spragtelėję mygtuką **Exact...**, galite pasirinkti tikslų testą mažoms imtims (vietoje asimptotinio).

12.2. DVIEJŲ PRIKLAUSOMŲ IMČIŲ PALYGINIMAS

Priklausomos imtys dažniausiai lyginamos norint patikrinti kokio nors poveikio ar laiko įtaką tiriamam kintamajam. Tada testas atliekamas du kartus: pirmą kartą grupė testuojama prieš poveikį ir antrą kartą — po daryto poveikio (treniruočių, gydymo, auklėjimo priemonių ir t. t.) ar praėjus tam tikram laikui. Šiam tikslui SPSS galite pasirinkti tris testus, iš kurių nustatytasis būna Wilcoxon'o testas. Vertas dėmesio taip pat ženklų testas, o dviejų binarinių kintamųjų palyginimui naudojamas Chi-kvadrato kriterijumi paremtas McNemar'o testas.

Testuojami duomenys turi būti skaitmeniniai, ranguojami. Nors reikalavimai kintamųjų pasiskirstymo pobūdžiui nekeliami, daroma prielaida, kad porinių testavimo reikšmių skirtumų skirstinys populiacijoje yra simetrinis.

12.2.1. Wilcoxon'o testas

Tai pats populiariausias dviejų priklausomų imčių lyginimo testas, taikomas tolydiems duomenims. Jis yra Stjudento porinio t -testo neparametrinis analogas. Testo esmę trumpai galima apibūdinti taip: randamas skirtumas tarp abiejų testavimo reikšmių (iš pirmojo testavimo reikšmių atimamos atitinkamos antrojo testavimo reikšmės) ir absoliučiosios gautų skirtumų reikšmės yra išdėstomos pagal rangus, paliekant rangams turėtus skirtumų ženklus. Kriterijaus statistika Z gaunama įvertinant rangų su minuso ir pliuso ženklais sumas bei nulinių skirtumų skaičių. Remiantis šia statistika priimamas sprendimas dėl statistinės hipotezės:

$$\begin{cases} H_0: \text{kintamųjų skirstiniai yra vienodi,} \\ H_1: \text{kintamųjų skirstiniai nėra vienodi.} \end{cases}$$

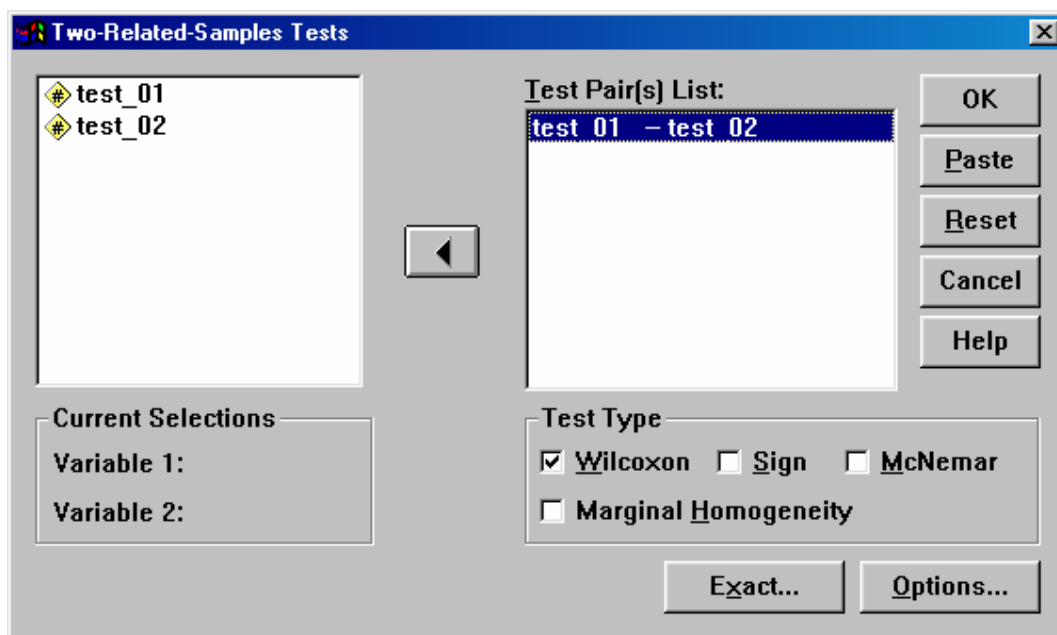
SPSS paketu tikrinama tik dvipusė alternatyva (H_1 : skirstiniai skiriasi).

Wilcoxon'o testas yra galingiausias, kai kintamųjų skirstiniai skiriasi tik postūmio parametru, t. y. X skirstinys sutampa su $Y + a$ skirstiniu. Būtent tokiems duomenims jį ir rekomenduojama taikyti.

Pavyzdys. Lyginami plaukikų grupės (12 sportininkų) aerobinio pajėgumo duomenys — deguonies suvartojimas prieš treniruočių ciklą ir po jo. Naudosime Wilcoxon'o testą, nes pagal Shapiro ir Wilk testą duomenys nėra pasiskirstę pagal normalųjį dėsnį.

- Nurodome komandas **Analyze → Nonparametric Tests → 2 Related Samples...**

- Dialogo langelyje **Two-Related-Samples Tests** (12.4 pav.) pažymime abu kintamuosius — *test_01* (pradiniai duomenys) ir *test_02* (duomenys po treniruočių ciklo) ir spragtelėję mygtuką su rodykle įkeliamė juos į laukelį **Test Pair(s) List**. Galima įkelti ir daugiau kintamųjų porų. Norint grąžinti duomenų porą atgal į bendrą duomenų rinkmenos sąrašą, reikia ją pažymėti ir spragtelėti mygtuką su pasikeitusios krypties rodykle.
- Paliekame nustatytąjį variantą — **Wilcoxon** ir spragtelime mygtuką **OK**.



12.4 pav. Dialogo langelis **Two-Related-Samples Tests**

Wilcoxon'o testo rezultatai parodyti 12.5 pav.

Lentelėje *a)* **Ranks** (rangai) pateikiamas rangų skaičius N , rangų vidurkis (*Mean Rank*) ir rangų suma (*Sum of Ranks*) neigiamoms ir teigiamoms skirtumų tarp abiejų grupių reikšmėms (šiuo atveju, didesnėms reikšmėms yra suteikiami aukštesni rangai), taip pat nulinių skirtumų skaičius. Lentelėje *b)* **Test Statistics** (testo statistika) pateikiama parametro Z reikšmė ir p -reikšmė (*Asymp. Sig. (2-tailed)*).

Kadangi gauta p -reikšmė viršija nustatytą reikšmingumo lygmenį (0,05), nulinės hipotezės atmesti negalima ir darome išvadą, kad po treniruočių ciklo plaukikų aerobinio pajėgumo duomenys reikšmingai nepasikeitė.

Dialogo langelio **Two-Related-Samples Tests** mygtukų **Exact...** ir **Options...** paskirtis ir funkcijos analogiškos, kaip ir dialogo langelio **Two-Independent-Samples Tests**.

Wilcoxon Signed Ranks Test

Ranks

		N	Mean Rank	Sum of Ranks
TEST_02 - TEST_01	Negative Ranks	3 ^a	5,67	17,00
	Positive Ranks	9 ^b	6,78	61,00
	Ties	0 ^c		
	Total	12		

a. TEST_02 < TEST_01

b. TEST_02 > TEST_01

c. TEST_01 = TEST_02

a)

Test Statistics^b

	TEST_02 - TEST_01
Z	-1,726 ^a
Asymp. Sig. (2-tailed)	,084

a. Based on negative ranks.

b. Wilcoxon Signed Ranks Test

b)

12.5 pav. **Wilcoxon'o** testo rezultatai

12.2.2. Ženklių testas

Ženklių testas yra pagrįstas teigiamų ir neigiamų skirtumų tarp testavimo reikšmių skaičiaus nustatymu ir gali būti taikomas, kai skirtumas tarp dviejų imčių nėra žymus. Tačiau, jis nusileidžia Wilcoxon'o testui efektyvumu, nes pastarasis atsižvelgia ne tik į skirtumo tarp dviejų imčių ženklą, bet ir į to skirtumo dydį.

Pavyzdys. 60 pacientų prieš pradedant reabilitacijos kursą ir po jo buvo paprašyti įvertinti savo savijautą. Atsakymų variantai: „gera“, „patenkinama“, „bloga“. Po reabilitacijos kurso dalis pacientų pakeitė savo vertinimus (apklausos rezultatų suvestinė parodyta 12.1 lentelėje). Reikia nustatyti, ar reabilitacija reikšmingai paveikė pacientų savijautą.

Duomenų rinkmenoje kintamieji pavadinti *test_01* (apklausos duomenys prieš poveikį) ir *test_02* (apklausos duomenys po poveikio). Kintamųjų reikšmės: 1–žymena „gera“, 2–žymena „patenkinama“, 3–žymena „bloga“.

12.1 lentelė. Apklausos rezultatų suvestinė

Prieš poveikį		Po poveikio		
Vertinimas	Vertinimų kiekis	1–gera	2–patenkinama	3–bloga
1–gera	5	3	2	
2–patenkinama	16	8	6	2
3–bloga	39	7	19	13

- Nurodome komandas *Analyze* → *Nonparametric Tests* → *2 Related Samples...*
- Dialogo langelyje *Two-Related-Samples Tests* (12.4 pav.) pažymime abu kintamuosius — *test_01* (pradiniai duomenys) ir *test_02* (duomenys po poveikio) ir spragtelėję mygtuką su rodykle įkeliamė juos į laukelį *Test Pair(s) List*.
- Pasirenkame testą *Sign* ir spragtelime mygtuką *OK*.

Ženklių testo rezultatai parodyti 12.6 pav. Gauti 34 neigiami skirtumai ($\text{test_02} < \text{test_01}$), kurie parodo savijautos pagerėjimą ir 4 teigiami skirtumai, liudijantys savijautos pablogėjimą. 22 atvejais savijautos pakitimų nebuvo. Kriterijaus *p*-reikšmė (*Asymp. Sig. (2-tailed)*) mažesnė už reikšmingumo lygmenį, todėl nulinę hipotezę atmetame — reabilitacija turėjo reikšmingą įtaką pacientų savijautos pagerėjimui.

Frequencies

		N
TEST_02 - TEST_01	Negative Differences ^a	34
	Positive Differences ^b	4
	Ties ^c	22
	Total	60

a. TEST_02 < TEST_01

b. TEST_02 > TEST_01

c. TEST_01 = TEST_02

Test Statistics^a

	TEST_02 - TEST_01
Z	-4,704
Asymp. Sig. (2-tailed)	,000

a. Sign Test

12.6 pav. Ženklių (*Sign*) testo rezultatai

12.2.3. McNemar'o testas

McNemar'o testas taikomas dvireikšmiams (binariniais) kintamiesiems. Paprastai testas naudojamas kartotinių bandymų, kada kiekvienas tiriamas

objektas testuojamas prieš ir po tam tikro poveikio, rezultatams įvertinti. Iš testo rezultatų sprendžiama apie taikyto poveikio reikšmingumą. Tokio tipo uždaviniuose binarinio kintamojo reikšmės koduojamos +, – arba *taip*, *ne*. Nulinė hipotezė H_0 teigia, kad populiacijos dalis, kuriai matuojamo kintamojo reikšmė pasikeitė iš + į –, lygi daliai, kuriai kintamojo reikšmė pasikeitė iš – į +, t. y. tikrinama proporcijų lygybė.

Pavyzdys. Atsitiktinai parinktų 40 asmenų grupė buvo apklausta du kartus skirtingu laiku. Apklausus pirmą kartą ir užfiksavus jų atsakymus tos pačios grupės asmenims buvo padarytas poveikis — perskaityta paskaita testo tematika. Tada buvo atlikta antra apklausa, po kurios pastebėta, kad dalis asmenų pakeitė savo anksčiau išsakytą nuomonę. Poveikio (šiuo atveju — paskaitos) reikšmingumui nustatyti panaudosime McNemar'o testą.

Duomenų rinkmenoje kintamieji pavadinti *test_1* (apklausos duomenys prieš poveikį) ir *test_2* (apklausos duomenys po poveikio). Kintamųjų reikšmės: 1 — žymena „*palankiai*“, 2 — žymena „*nepalankiai*“.

- Nurodome komandas **Analyze → Nonparametric Tests → 2 Related Samples...**
- Dialogo langelyje **Two-Related-Samples Tests** (12.4 pav.) pažymime abu kintamuosius – *test_1* ir *test_2* ir spragtelėję mygtuką su rodykle įkeliamo juos į laukelį **Test Pair(s) List**.
- Pasirenkame **McNemar** testą ir spragtelime mygtuką **OK**.

Testo rezultatai parodyti 12.7 pav.

McNemar Test

Crosstabs

TEST_1 & TEST_2		
TEST_1	TEST_2	
	1	2
1	6	3
2	15	16

a)

Test Statistics ^b	
	TEST_1 & TEST_2
N	40
Exact Sig. (2-tailed)	,008 ^a

a. Binomial distribution used.

b. McNemar Test

b)

12.7 pav. **McNemar'o** testo rezultatai

Pirmoje a) lentelėje parodyta, kad pirmos apklausos metu *palankiai* atsakė 9 asmenys ir *nepalankiai* – 31 asmuo. Antros apklausos metu 6 iš šių 9 patvirtino vertinimą *palankiai*, o 3 pakeitė nuomonę į *nepalankiai*, o 15 iš 31 pakeitė nuomonę į *palankiai* ir 16 patvirtino nuomonę *nepalankiai*. Remiantis antroje lentelėje pateikta *p*-reikšme (*Exact Sig. (2-tailed)*) galima teigti, kad paskaita reikšmingai paveikė apklausiamųjų nuomonę.

12.2.4. Ribinis homogeniškumo testas

Ribinis homogeniškumo testas (*Marginal Homogeneity Test*) taikomas kategoriniams duomenims. Savo esme jis yra tolygus McNemar'o testui, tik skirtingai nuo pastarojo, kuris taikomas binarinių kintamųjų analizei, ribinis homogeniškumo testas taikomas daugiareikšmių kintamųjų analizei.

12.3. KELIŲ NEPRIKLAUSOMŲ IMČIŲ PALYGINIMAS

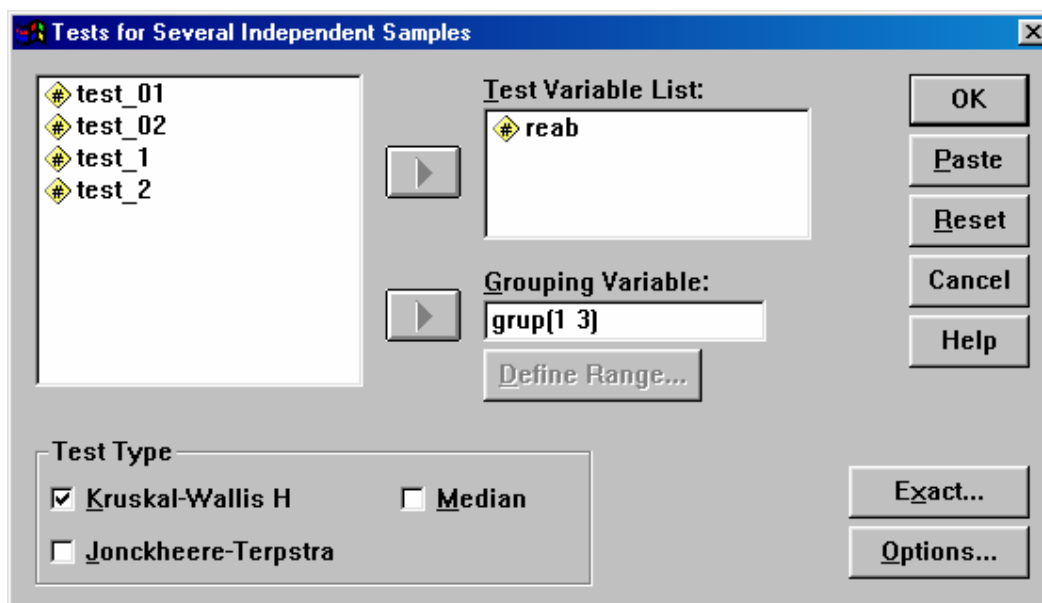
Kelių nepriklausomų imčių palyginimo neparametriniai kriterijai skirti hipotezei apie dviejų ir daugiau populiacijų skirstinių lygybę tikrinti. SPSS paketu lyginamos kelios vieno kintamojo stebėjimų grupės. Testai turi būti atlikti naudojant vienodą metodiką, o testavimo duomenys turi būti skaitmeniniai, ranguojami. Iš kelių SPSS pateikiamų tos paskirties testų panagrinsime plačiausiai naudojamą Kruscal'o ir Wallis'o H testą.

12.3.1. Kruscal'o ir Wallis'o H testas

Kruscal'o ir Wallis'o testas yra Mann'o ir Whitney testo modifikacija, taikoma daugiau negu dviem nepriklausomoms imtims palyginti. Jis taip pat paremtas jungtine visų imčių reikšmių rangine seka.

Kelių nepriklausomų imčių palyginimo testo rezultatams gauti:

- Atidarykite bylą su analizuojamais duomenimis.
- Nurodykite komandas **Analyze → Nonparametric Tests → K Independent Samples...**
- Dialogo langelyje **Tests for Several Independent Samples** (12.8 pav.) įkelkite vieną ar kelis kintamuosius į testuojamų kintamųjų sąrašą **Test Variable List**. Jeigu pasirinkote kelis kintamuosius, testo rezultatai bus apskaičiuoti kiekvieno kintamojo atskirai.
- Kategorinį kintamąjį įkelkite į laukelį **Grouping Variable**.
- Spragtelėkite dialogo langelio mygtuką **Define Range** (nustatyti rangą). Naujame dialogo langelyje **Several Independent Samples: Define** nurodykite mažiausią ir didžiausią kategorinio kintamojo, pagal kurį identifikuojamos testuojamo kintamojo grupės, reikšmes.
- Palikite nustatytąjį testą **Kruskal-Wallis H** ir spragtelėkite mygtuką **OK**.



12.8 pav. Dialogo langelis *Tests for Several Independent Samples*

Pavyzdys. Norima nustatyti, kuris reabilitacijos metodas padeda sportininkams greičiausiai atsigauti po kelių raiščių sužeidimo. Buvo tiriamos trys sportininkų grupės po aštuonis tiriamuosius, kuriems buvo taikytos trijų rūšių reabilitacijos programos.

Po nustatyto reabilitacijos laikotarpio gydytojų ekspertų grupė patikrino tiriamuosius ir įvertino jų būklę dešimties balų sistema. Reikia patikrinti hipotezę:

$$\begin{cases} H_0: \text{visi reabilitacijos metodai yra vienodi,} \\ H_1: \text{reabilitacijos metodai skiriasi.} \end{cases}$$

- Kintamąjį, kurio reikšmės yra trijų grupių testavimo rezultatai, pavadinsime *reab*, o kategorinį kintamąjį, kuris parodo, kuriai grupei priklauso konkretus testavimo rezultatas, pavadinsime *grup*.
- Kintamąjį *reab* įkeliamė į dialogo langelio *Tests for Several Independent Samples* laukelį *Test Variable List*, o kintamąjį *grup* – į laukelį *Grouping Variable*.
- Spragtelėję mygtuką *Define Range* naujo dialogo langelio *Several Independent Samples: Define* laukelyje *Minimum* įrašome 1, o laukelyje *Maximum* įrašome 2.
- Spragtelėję dialogo langelio *Several Independent Samples: Define* mygtuką *Continue*, tada — pagrindinio dialogo langelio *Tests for Several Independent Samples* mygtuką *OK*, gauname 12.9 pav. parodytus Kruskal'o ir Wallis'o testo rezultatus.

Pirmoje *a)* lentelėje *Ranks* (rangai) pateikiami kiekvienos grupės vidutiniai rangai (*Mean Rank*), gauti ranguojant jungtinius visų grupių duomenis. Antroje *b)* lentelėje *Test Statistics* (testo statistika) pateikiama Chi-kvadrato (*Chi-Square*) kriterijaus reikšmė, laisvės laipsnių (*df*) skaičius ir *p*-reikšmė — *Asymp. Sig.* Šiame pavyzdyje *p*-reikšmė yra mažesnė už nustatytą reikšmingumo lygmenį (0,05), tai rodo esant reikšmingą skirtumą tarp taikytų reabilitacijos metodų. Informaciją apie pačius skirstinius, t. y. kurie skirstiniai „linkę“ būti didesni už kitus, galima sužinoti iš lentelės *Ranks* vidutinių rangų stulpelio *Mean Rank* — kuo vidutinis rangas didesnis, tuo labiau atitinkamas skirstinys „linkęs“ įgyti didesnes, palyginti su kitais skirstiniais, reikšmes (ranguojant mažiausiai reikšmei skiriama pirma vieta ir t. t.).

Norint nustatyti, kurios būtent grupės skiriasi reikšmingai, galima palyginti visas grupes poromis pagal Mann'o ir Whitney testą.

Kruskal-Wallis Test

Ranks

	GRUP	N	Mean Rank
REAB	1	8	17,56
	2	8	11,63
	3	8	8,31
	Total	24	

a)

Test Statistics^{a,b}

	REAB
Chi-Square	7,262
df	2
Asymp. Sig.	,026

a. Kruskal Wallis Test

b. Grouping Variable: GRUP

b)

12.9 pav. **Kruskal'o ir Wallis'o** testo rezultatai

12.4. KELIŲ PRIKLAUSOMŲ IMČIŲ PALYGINIMAS

Kelių priklausomų imčių testais tikrinamas skirtumo tarp kelių priklausomų kintamųjų skirstinių reikšmingumas. Testai turi būti atlikti taikant vienodą metodiką, o testavimo duomenys turi būti skaitmeniniai, ranguojami. Kaip ir kiti neparametriniai testai, jie nereikalauja prielaidos apie konkretų duomenų pasiskirstymo dėsnį. Iš pateikiamų SPSS trijų šio tipo testų dažniausiai naudojamas Friedman'o testas. Kendall'o W testas ir Cochran'o Q testas taikomi atskirai numatytais atvejais.

12.4.1. Friedman'o testas

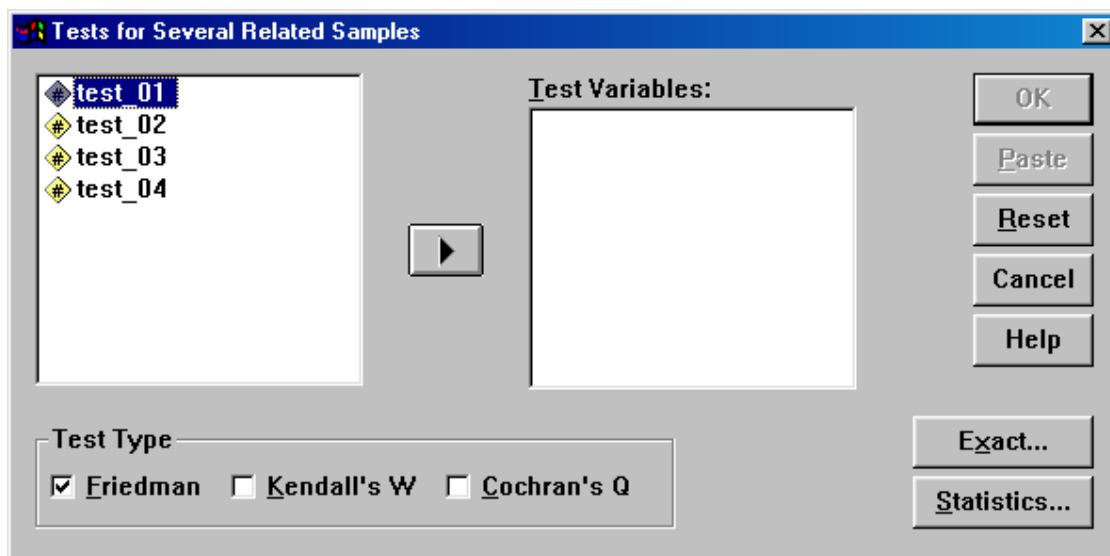
Friedman'o testas praplečia Wilcoxon'o testo galimybes daugiau kaip dviem priklausomoms imtims palyginti. Kriterijaus statistikai apskaičiuoti kiekvienas kintamųjų rinkinio stebėjimas (duomenų rinkmenos eilutė) ranguojamas atskirai, tada sudedami kiekvienam kintamajam tekę rangai. Remiantis apskaičiuota statistika priimamas sprendimas dėl statistinės hipotezės:

$$\begin{cases} H_0: \text{kintamųjų skirstiniai vienodi,} \\ H_1: \text{kintamųjų skirstiniai nėra vienodi.} \end{cases}$$

Daroma išvada, kad skirstiniai skiriasi, jeigu p -reikšmė mažesnė už α ir skirstiniai nesiskiria, jeigu p -reikšmė didesnė už α arba jai lygi, čia α — reikšmingumo lygmuo.

Kelių priklausomų imčių palyginimo Friedman'o testu rezultatams gauti:

- Atidarykite bylą su analizuojamais duomenimis.
- Nurodykite komandas *Analyze* → *Nonparametric Tests* → *K Related Samples...*
- Dialogo langelyje *Tests for Several Related Samples* (12.10 pav.) įkelkite kintamuosius į testuojamų kintamųjų sąrašą *Test Variables*.
- Įsitikinkite, kad nustatytas yra *Friedman'o* testas ir spragtelėkite mygtuką *OK*.



12.10 pav. Dialogo langelis *Tests for Several Related Samples*

Pirmas pavyzdys. Pakartosime 12.2.1 skyrelio pavyzdį, kiek pakeisdami testo sąlygas. Treniruočių ciklo metu, reguliariais intervalais, keturis kartus buvo tikrinami plaukikų grupės (12 sportininkų) aerobinio pajėgumo duomenys —

deguonies suvartojimas. Kadangi pagal histogramos pobūdį, taip pat Shapiro ir Wilk testą duomenų pasiskirstymas skiriasi nuo normaliojo skirstinio, naudosime neparametrinį kelių priklausomų imčių Friedman'o testą.

- Dialogo langelyje **Tests for Several Related Samples** (12.10 pav.) pažymime kintamuosius – *test_01*, *test_02*, *test_03*, *test_04* ir spragtelėję mygtuką su rodykle įkeliamė juos į laukelį **Test Variables**.

Friedman'o testo rezultatai parodyti 12.11 pav. Lentelėje *Ranks* (rangai) pateikiami analizuojamų kintamųjų vidutiniai rangai (*Mean Rank*). Lentelėje *Test Statistics* (testo statistika) pateikiamas stebėjimų skaičius *N*, Chi- kvadrato kriterijaus reikšmė (*Chi-Square*), laisvės laipsnių skaičius *df* ir *p*-reikšmė (*Asymp. Sig.*).

Mūsų nagrinėjamu atveju *p*-reikšmė $0,005 < 0,05$, todėl darome išvadą, kad treniruočių ciklo metu plaukikų aerobinio pajėgumo duomenys keitėsi statistiškai reikšmingai. Nustačius šį faktą, Wilcoxon'o porinio palyginimo testu galima nustatyti, kokiais būtent laiko momentais duomenys skiriasi reikšmingai.

Antras pavyzdys. Norima nustatyti, ar studentai vienodai vertina tris pasirenkamus dalykus, kuriuos bendrumo dėlei pavadinsime *a*, *b* ir *c*. Patikrinimui atrinkta 10 studentų ir kiekvieno iš jų paprašyta suteikti šiems dalykams rangus: 1 — labiausiai mėgstamam dalykui, 2 — vidutiniškai mėgstamam, 3 — mažiausiai mėgstamam. Kintamuosius duomenų rinkmenoje pavadinsime atitinkamai *dalyk_a*, *dalyk_b*, *dalyk_c*.

Friedman'o testo rezultatai parodyti 12.12 pav. Galima teigti, kad studentai statistiškai reikšmingai skirtingai vertina pasirenkamus dalykus.

Friedman Test

Ranks	
	Mean Rank
TEST_01	1,42
TEST_02	2,50
TEST_03	3,08
TEST_04	3,00

a)

Test Statistics ^a	
N	12
Chi-Square	12,700
df	3
Asymp. Sig.	,005

a. Friedman Test

b)

12.11 pav. Pirmam pavyzdžio **Friedman'o** testo rezultatai

Friedman Test

Ranks

	Mean Rank
DALYK_A	1,40
DALYK_B	2,60
DALYK_C	2,00

a)

Test Statistics^a

N	10
Chi-Square	7,200
df	2
Asymp. Sig.	,027

a. Friedman Test

b)

12.12 pav. Antrojo pavyzdžio **Friedman'o** testo rezultatai

12.4.2. Kendall'o W testas

Kendall'o konkordacijos koeficientas W naudojamas daugiau negu dviejų ranginių kintamųjų priklausomybei įvertinti. Jis dar vadinamas ekspertų vertinimo testu, nes juo yra nustatomas grupės ekspertų sutarimas. Koeficientas W gali turėti reikšmės intervale tarp nulio (nėra sutarimo) ir vieneto (visiškas sutarimas). Galima apskaičiuoti ne tik patį konkordacijos koeficientą, bet ir patikrinti, ar jis statistiškai reikšmingai skiriasi nuo nulio. Konkordacija statistiškai reikšminga, jeigu p -reikšmė yra mažesnė už pasirinktą reikšmingumo lygmenį α . SPSS duomenų rinkmenoje recenzuojamieji (vertinami) subjektai sudaro atskirus kintamuosius (stulpelius), o ekspertai — atskirus stebėjimus (eilutes). Kiekvienas ekspertas pagal nustatytus kriterijus suteikia rangą kiekvienam vertinamam subjektui.

Pavyzdys. Trijų sporto laikraščių apžvalgininkai vertino populiarios futbolo komandos žaidėjų pasirodymą per atsakingas rungtynes. 14 žaidėjų žaidimas buvo vertinamas balais nuo 1 (aukščiausias įvertis) iki 6 (žemiausias įvertis). Kintamuosius duomenų rinkmenoje pavadinsime f_{01}, \dots, f_{14} .

- Dialogo langelyje **Tests for Several Related Samples** (12.10 pav.) pažymime kintamuosius f_{01}, \dots, f_{14} ir spragtelėję mygtuką su rodykle įkeliamo juos į laukelį **Test Variables**. Pasirenkame **Kendall's W** testą ir spragtelime mygtuką **OK**.

Kendall'o testo rezultatai parodyti 12.13 pav. Lentelėje a) *Ranks* (rangai) pateikiamas visų žaidėjų vidutinis rangas (*Mean Rank*). Lentelėje b) *Test Statistics* (testo statistika) pateikiamas stebėjimų skaičius N , Kendall'o

konkordacijos koeficiento W reikšmė (*Kendall's W*), Chi- kvadrato kriterijaus reikšmė (*Chi-Square*), laisvės laipsnių skaičius df ir p -reikšmė (*Asymp. Sig.*).

Gauta konkordacijos koeficiento reikšmė ($W=0,763$) liudija, kad trijų laikraščių apžvalgininkai gana sutartinai vertino 14 futbolininkų pasirodymą per atsakingas rungtynes. Konkordacija yra statistiškai reikšminga, nes $p=0,005 < 0,05$.

Kendall's W Test

Ranks	
	Mean Rank
F_01	7,83
F_02	5,00
F_03	8,83
F_04	9,50
F_05	2,50
F_06	2,67
F_07	4,67
F_08	10,17
F_09	12,50
F_10	6,17
F_11	13,33
F_12	3,33
F_13	9,17
F_14	9,33

a)

Test Statistics	
N	3
Kendall's W ^a	,763
Chi-Square	29,775
df	13
Asymp. Sig.	,005

a. Kendall's Coefficient of Concordance

b)

12.13 pav. **Kendall'o** testo rezultatai

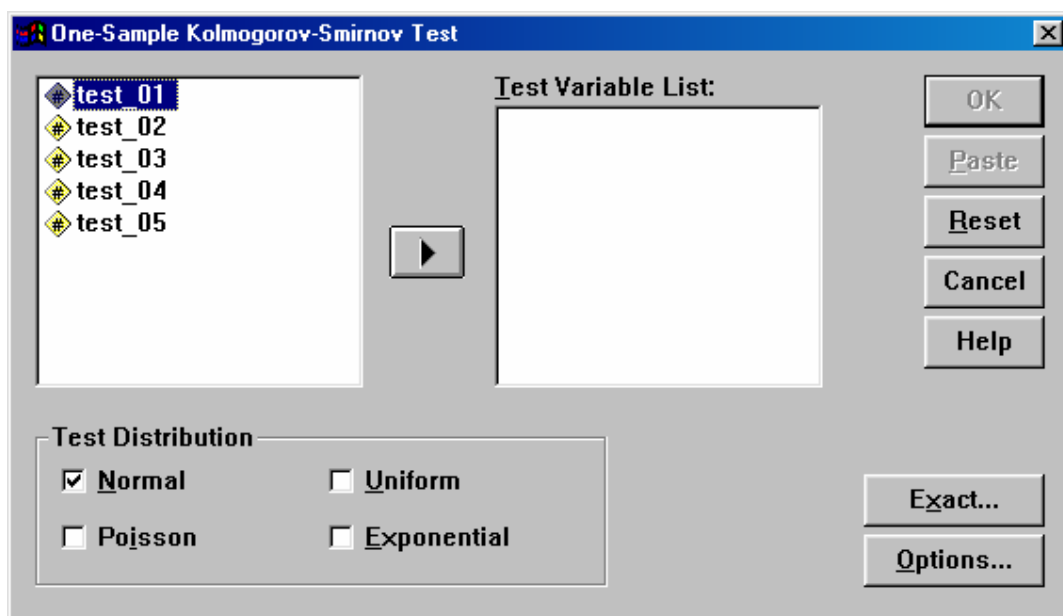
12.4.3. Cochran'o Q testas

Cochran'o testas yra tapatus Friedman'o testui, tiktai pritaikytas binariniais duomenimis. Jis išplečia McNemar'o testo galimybes ir leidžia atlikti k priklausomų imčių su binariniais duomenimis analizę. Cochran'o testo taikymo pavyzdys galėtų būti hipotezės, kad respondentų atsakymai į atskirus anketos klausimus skiriasi nereikšmingai, patikrinimas.

12.5. KOLMOGOROVO IR SMIRNOVO TESTAS

Šiuo testu galima patikrinti, ar realus skirstinys atitinka vieną iš šių skirstinių: normalųjį, tolygųjį, eksponentinį ar Puasono. Dažniausiai yra tikrinamas skirstinio normalumas, kadangi nuo gautų rezultatų priklauso, kokie analizės metodai — parametriniai ar neparametriniai bus taikomi. Norėdami patikrinti, ar pradiniai duomenys pasiskirstę pagal normalųjį dėsnį:

- Atidarykite bylą su analizuojamais duomenimis.
- Nurodykite komandas *Analyze* → *Nonparametric Tests* → *1-Sample K-S...*
- Dialogo langelyje *One Sample Kolmogorov-Smirnov Test* (12.14 pav.), kuriame nustatytasis variantas yra skirstinio normalumo patikrinimas, įkelkite pasirinktą kintamąjį į testuojamų kintamųjų sąrašą *Test Variable List*. Jeigu pasirinkote kelis kintamuosius, kiekvienas kintamasis bus testuojamas atskiru testu.
- Spragtelėkite mygtuką *OK*.



12.14 pav. Dialogo langelis *One-Sample Kolmogorov-Smirnov Test*

Skirstinys reikšmingai skiriasi nuo normaliojo, jeigu gauta p -reikšmė (*Asymp. Sig. (2-tailed)*) mažesnė už nustatytą reikšmingumo lygmenį (paprastai — 0,05).

Spragtelėję mygtuką *Options...*, Jūs galite:

- Pasirinkti kai kurias papildomas *Descriptive* charakteristikas (vidurkį, mažiausią reikšmę, didžiausią reikšmę, standartinį nuokrypį, stebėjimų skaičių) bei kvartilius (*Quartiles*) — 25, 50 ir 75 procentilius.

- Nurodyti, kaip traktuoti praleistas reikšmes: ***Exclude cases test-by-test*** — praleistos reikšmės įvertinamos kiekvienam nustatytam testui (nustatytasis variantas), ***Exclude cases listwise*** — stebėjimai, kuriuose pasitaiko kurio nors kintamojo praleistų reikšmių, pašalinami iš visų skaičiavimų.

Spragtelėję mygtuką ***Exact...***, galite pasirinkti mažoms imtims tikslų testą (vietoj asimptotinio).

12.6. CHI-KVADRATO TESTAS

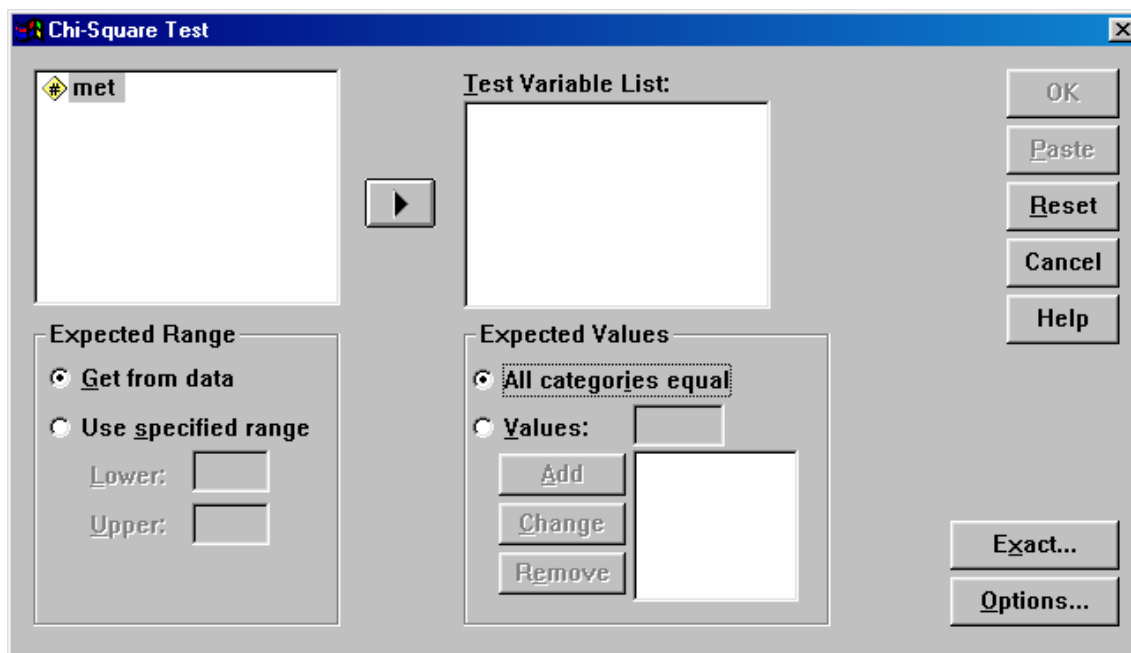
Chi-kvadrato testu (žr. taip pat 10.3.1 skyrelį) tikrinama, kiek gauti tyrimo duomenys sutampa su planuotais prieš tyrimą. Testas taikomas skaitmeniniams kategoriniams (nominaliesiems arba ranginiams) kintamiesiems: lyginami kiekvienos kintamojo kategorijos eksperimentiniu būdu gauti dažniai (*observed frequency*) su teoriškai numatytais (tikėtinais) dažniais (*expected frequency*). Teoriškai numatyti dažniai dažniausiai yra tolygiai pasiskirstę, t. y. visos kintamojo kategorijos turi vienodas proporcijas. SPSS taip pat yra galimybė vartotojui pačiam nustatyti šias proporcijas.

Kaip ir kiti neparametriniai testai, Chi-kvadrato testas nereikalauja prielaidos apie konkretų duomenų pasiskirstymo dėsnį. Kiekvienos kintamojo kategorijos tikėtini dažniai turi būti ne mažiau kaip 1 ir ne daugiau kaip 20% visų kategorijų tikėtini dažniai gali būti mažesni už 5.

Chi-kvadrato testo rezultatams gauti:

- Atidarykite bylą su analizuojamais duomenimis.
- Nurodykite komandas ***Analyze → Nonparametric Tests → Chi-Square...***
- Dialogo langelyje ***Chi-Square Test*** (12.15 pav.) įkelkite pasirinktą kintamąjį į testuojamų kintamųjų sąrašą ***Test Variable List***. Jeigu pasirinkote kelis kintamuosius, kiekvienas kintamasis bus testuojamas atskiru testu.
- Jeigu norite, kad būtų analizuojamos visos kintamojo kategorijos (grupės), dialogo langelio skyriuje ***Expected Range*** (tikėtinas diapazonas) palikite nustatytąjį variantą ***Get from data*** (iš pradinių duomenų), o jeigu norite apriboti analizuojamų kategorijų skaičių, pasirinkite variantą ***Use specified range*** (nurodytas diapazonas) ir nurodykite apatinę (***Lower***) ir viršutinę (***Upper***) ribas.
- Pagal nustatytąjį variantą visų kintamojo kategorijų tikėtini dažniai vienodi. Jeigu norite nurodyti kitas tikėtinų dažnių proporcijas, pažymėkite langelį ***Values*** dialogo langelio skyriuje ***Expected Values***, nurodykite pirmos kategorijos proporcijos dydį ir spragtelėkite mygtuką ***Add***. Pakartokite šį veiksmą visoms kitoms kintamojo kategorijoms. Svarbi yra šių reikšmių įvedimo tvarka — ji turi atitikti didėjančias kintamojo kategorijų reikšmes, t. y. pirmoji proporcijos reikšmė turi atitikti mažiausią kategorijos reikšmę, o paskutinė proporcijos reikšmė —

didžiausią kategorijos reikšmę. Kiekvienos kategorijos tikėtino dažnio proporcija skaičiuojama dalijant nurodytą reikšmę iš visų nurodytų reikšmių sumos. Pavyzdžiui, įrašyti skaičiai 3, 4, 5, 4 reikš šias proporcijas: $3/16$, $4/16$, $5/16$, $4/16$.

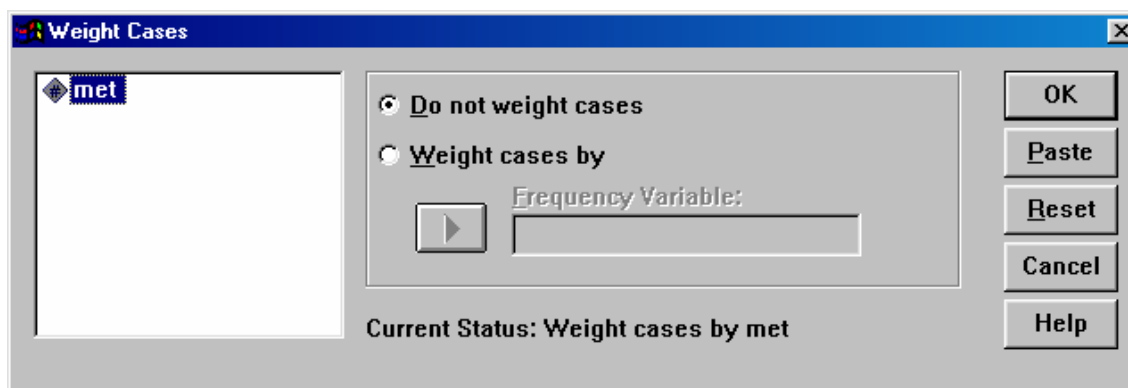


12.15 pav. Dialogo langelis *Chi-Square Test*

Pavyzdys. Krepšininkas B per 7 rungtynes metė kamuolį į krepšį iš vidutinių nuotolių 127 kartus. Buvo manoma, kad žaidėjo taiklumas yra 60%. Iš 127 metimų 64 buvo taiklūs. Ar galima daryti išvadą, kad krepšininko taiklumas pablogėjo?

Šiame pavyzdyje duomenis jau turime dažnių lentelės pavidalu — taiklūs metimai — 64, netaiklūs — 63. Kad uždavinį būtų galima išspręsti SPSS programa, pirmiausia reikia sudaryti duomenų rinkmeną. Kintamasis, kurį pavadinsime *met*, turi tik du stebėjimus — 63 (žymena *netaikl*) ir 64 (žymena *taikl*). Tokiems koncentruotiems duomenims yra taikomas svorių suteikimas (*Weight Cases*):

- Nurodome komandas **Data → Weight Cases...**
- Dialogo langelyje **Weight Cases** (12.16 pav.) pažymime variantą **Weight cases by** ir perkeliame kintamąjį *met* į laukelį **Frequency Variable** (dažnių kintamasis).
- Spragtelime mygtuką **OK**.



12.16 pav. Dialogo langelis *Weight Cases*

- Nurodome komandas *Analyze → Nonparametric Tests → Chi-Square...*
 - Dialogo langelyje *Chi-Square Test* (12.15 pav.) įkeliame kintamąjį *met* į testuojamų kintamųjų sąrašą *Test Variable List*.
 - Pažymime langelį *Values* dialogo langelio skyriuje *Expected Values*, nurodome nepataikytų metimų tikėtinų dažnių proporcijos dydį — 0,4 ir spragtelime mygtuką *Add*. Tada nurodome pataikytų metimų tikėtinų dažnių proporcijos dydį — 0,6 ir vėl spragtelime mygtuką *Add*.
 - Spragtelime dialogo langelio *Chi-Square Test* mygtuką *OK*.
- Chi-kvadrato testo rezultatai parodyti 12.17 pav.

Chi-Square Test

MET			
	Observed N	Expected N	Residual
netaikl	63	50,8	12,2
taikl	64	76,2	-12,2
Total	127		

a)

Test Statistics	
	MET
Chi-Square ^a	4,883
df	1
Asymp. Sig.	,027

a. 0 cells (,0%) have expected frequencies less than 5. The minimum expected cell frequency is 50,8.

b)

12.17 pav. *Chi-kvadrato* testo rezultatai

Pirmoje a) lentelėje pateikiami analizuojamo kintamojo kiekvienos kategorijos eksperimentiniu būdu gauti dažniai (*ObservedN*), tikėtini dažniai

(*ExpectedN*) ir liekanos (*Residual*). Antroje *b*) lentelėje (*Test Statistics*) pateikiama Chi-kvadrato kriterijaus reikšmė (*Chi-Square*), laisvės laipsnių skaičius (*df*) ir *p*-reikšmė (*Asymp. Sig.*). Kadangi $p < 0,05$, nulinę hipotezę, kad krepšininko taiklumas lygus tikėtinam, atmetame. Stebimi rezultatai reikšmingai skiriasi nuo tikėtinų. Kad jie blogesni už tikėtinus matyti iš *a*) lentelės.

13. KORELIACIJOS RYŠYS

Šis skyrius skirtas koreliacijos ryšio tarp dviejų kintamųjų analizei. Tiesinės dviejų kintamųjų priklausomybės mato — koreliacijos koeficiento r skaičiavimo metodika priklauso nuo kintamųjų skalės:

- Pasiskirsčiusiems pagal normalųjį dėsnį intervaliniams kintamiesiems yra skaičiuojamas Pirsono (*Pearson*) koreliacijos koeficientas.
- Intervaliniams kintamiesiems, kuriems normalumo prielaida nėra tenkinama, ir ranginiams kintamiesiems yra skaičiuojamas Spirmeno (*Spearman*) arba Kendall'o τ -b koreliacijos koeficientas.
- Dvireikšmiams (binariniais) kintamiesiems yra skaičiuojami panašumo ir skirtingumo rodikliai.

Koreliacijos koeficientas nematuoja netiesinės priklausomybės.

13.1. PIRSONO KORELIACIJOS KOEFICIENTAS

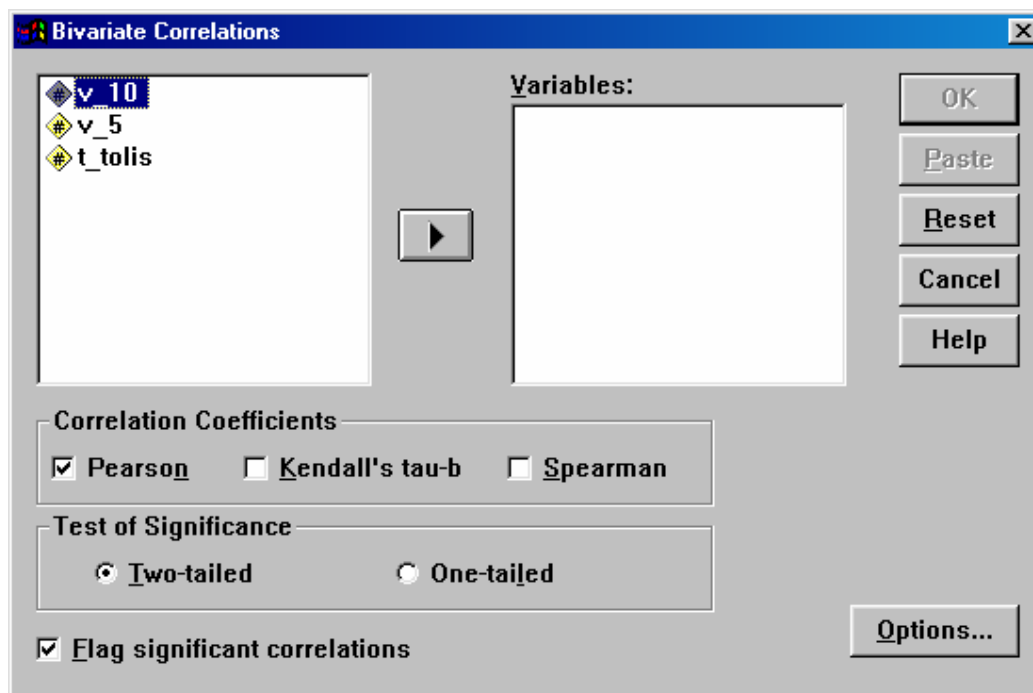
Pirsono koreliacijos koeficientas skaičiuojamas pagal formulę

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{(n-1) \cdot s_x \cdot s_y}, \quad (13.1)$$

čia x_i, y_i — abiejų kintamųjų imties reikšmės, \bar{x}, \bar{y} — šių kintamųjų imties vidurkiai, s_x, s_y — kintamųjų standartiniai nuokrypiai, n — imties variantų (kintamųjų reikšmių porų) skaičius.

Pirsono koreliacijos koeficiento reikšmei tarp kintamųjų porų apskaičiuoti, t. y. visų analizuojamų kintamųjų koreliacinei matricai gauti:

- Atidarykite bylą su analizuojamais duomenimis.
- Nurodykite komandas **Analyze → Correlate... → Bivariate...**
- Dialogo langelyje **Bivariate Correlations** (13.1 pav.) perkeltkite reikiamus kintamuosius į testuojamų kintamųjų sąrašą **Variables**.
- Nustatytasis **Correlation Coefficients** variantas yra **Pearson**, o **Test of Significance** (reikšmingumo testas) — **Two-tailed** (dvipusis). Jeigu ryšio kryptis yra žinoma, pasirinkite **One-tailed** (vienpusį) reikšmingumo testą.
- Spragtelėję mygtuką **Options...**, naujame dialogo langelyje **Bivariate Correlations: Options** galite nurodyti skaičiuoti šias papildomas charakteristikas:
 - **Means and standart deviations** (kiekvieno kintamojo vidurkį ir standartinį nuokrypį).
 - **Cross-product deviations and covariances** (momentų sandaugos sumą — 13.1 formulės skaitiklio reikšmę ir kovariaciją — 13.1 formulės skaitiklio reikšmę, padalytą iš $n-1$).



13.1 pav. Dialogo langelis *Bivariate Correlations*

Pavyzdys. Tiriamas ryšys tarp trišuolininko išibėgėjimo greičio per paskutinius 10 m ir 5 m tarpsnius ir trišuolio rezultato. Kintamieji duomenų rinkmenoje pavadinti v_{10} , v_5 ir t_{tolis} .

Skaiciavimo rezultatai parodyti 13.2 pav. Lentelėje *Correlations* pateikta koreliacinė matrica, t. y. visų galimų kintamųjų porų Pirsono koreliacijos koeficiento (*Pearson Correlation*) reikšmės ir p -reikšmė — *Sig. (2-tailed)*, pagal kurią vertinama, ar koreliacija statistiškai reikšmingai skiriasi nuo nulio, taip pat analizuojamų kintamųjų reikšmių porų skaičius N .

Nulinė hipotezė, reiškianti, kad koreliacijos koeficientas statistiškai reikšmingai nesisikiria nuo nulio (koreliacija nėra statistiškai reikšminga), atmetama (t. y. koreliacija nenulinė, kintamieji priklausomi), jeigu $p < \alpha$;

Nulinė hipotezė neatmetama, jeigu $p \geq \alpha$, čia α — nustatytas reikšmingumo lygmuo.

Šiame pavyzdyje stipriausias koreliacijos ryšys ($r = 0,9$) yra tarp trišuolio rezultato ir išibėgėjimo greičio per paskutinius 10 m. Koreliacija yra statistiškai reikšminga, nes visais atvejais $p < 0,05$.

Correlations

		V_10	V_5	T_TOLIS
V_10	Pearson Correlation	1	,775*	,900*
	Sig. (2-tailed)	,	,005	,000
	N	11	11	11
V_5	Pearson Correlation	,775*	1	,645*
	Sig. (2-tailed)	,005	,	,032
	N	11	11	11
T_TOLIS	Pearson Correlation	,900*	,645*	1
	Sig. (2-tailed)	,000	,032	,
	N	11	11	11

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

13.2 pav. Pirsono koreliacijos koeficiento skaičiavimo rezultatai

13.2. SPIRMENO IR KENDALL'O KORELIACIJOS KOEFICIENTAI

Spirmeno koreliacijos koeficientas skaičiuojamas tada, kai kintamųjų duomenys pateikti rangais arba kai nėra tenkinama normalumo prielaida. Pavyzdžiui, reikia patikrinti, ar yra koreliacija tarp sportininkų užimtos vietos varžybose ir taikyto testo rodiklių. Šiuo atveju matuojamiems pagal intervalų skalę kintamiesiems taip pat suteikiami rangai ir tolesnis skaičiavimas atliekamas tik su rangais. Taip pat gali būti skaičiuojamas Kendall'o rangų koreliacijos koeficientas, kurio skaičiavimas pagrįstas suderintų ir nesuderintų duomenų porų skaičiaus skirtumu (plačiau – Čekanavičius, Murauskas, 2002). Minėti koeficientai skaičiuojami, pasirinkus dialogo langelyje **Bivariate Correlations Kendall's tau-b** arba **Spearman** variantus. Kaip ir Pirsono koreliacijos koeficiento atveju, pateikiami empiriniai koreliacijos koeficientai ir kriterijaus p -reikšmė. Jeigu tikrinant hipotezę apie Spirmeno ar Kendall'o koreliacijos koeficientą p -reikšmė lygi p , o reikšmingumo lygmuo lygus α , darome išvadą, kad:

kintamieji koreliuoja, kai $p < \alpha$,

kintamieji nekoreliuoja, kai $p \geq \alpha$.

13.3. DALINĖ KORELIACIJA

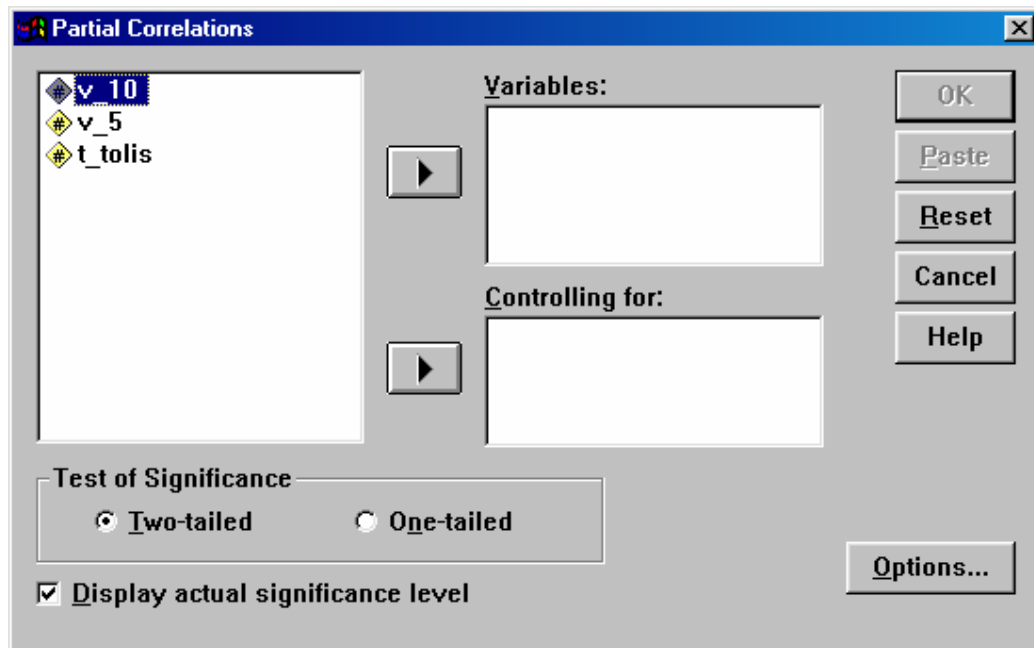
Dažnai tenka įvertinti dviejų kintamųjų tarpusavio ryšį, kurį veikia kiti kintamieji, galintys jį iškreipti. Tada yra skaičiuojami dalinės koreliacijos koeficientai, kurie leidžia įvertinti dviejų tiriamų kintamųjų tarpusavio ryšį, kai kitų kintamųjų įtaka yra eliminuojama. Dalinės koreliacijos koeficientas apskaičiuojamas pagal formulę

$$r_{xy,z} = \frac{r_{xy} - r_{xz} \cdot r_{yz}}{\sqrt{(1 - r_{xz}^2) \cdot (1 - r_{yz}^2)}}, \quad (13.2)$$

čia r_{xy} , r_{xz} , r_{yz} — poriniai koreliacijos koeficientai, nustatyti pagal Pirsono arba Spirmeno formules tarp kintamųjų x ir y , x ir z bei y ir z .

SPSS paketu dalinės koreliacijos koeficiento reikšmei apskaičiuoti:

- Atidarykite bylą su analizuojamais duomenimis.
- Nurodykite komandas **Analyze → Correlate... → Partial...**
- Dialogo langelyje **Partial Correlations** (13.3 pav.) perkeltite reikiamus kintamuosius į testuojamų kintamųjų sąrašą **Variables**, o vieną ar kelis kontroliuojamuosius kintamuosius (kintamuosius, į kurių įtaką skaičiuojant dalinę koreliaciją atsižvelgta) — į laukelį **Controlling for**.
- Spragtelėkite mygtuką **OK**.



13.3 pav. Dialogo langelis **Bivariate Correlations**

Išvada, ar gautoji dalinė koreliacija statistiškai reikšminga, daroma taip pat, kaip ir tikrinant hipotezę apie Pirsono ar Spirmeno koreliaciją — koreliacija statistiškai reikšminga, jeigu p -reikšmė mažesnė už pasirinktąjį reikšmingumo lygmenį.

13.4. SUTAPIMO IR SKIRTINGUMO MATAI

Be anksčiau nagrinėtų koreliacijos koeficientų, SPSS pateikia didelę sutapimo ir skirtingumo matų įvairovę, iš kurių reikėtų išskirti skaitlingą matų grupę ryšiui tarp binarinių kintamųjų įvertinti. Šiais matais taip pat remiasi faktorinė analizė, klasterinė analizė ir kiti analizės metodai. Sutapimo ir skirtingumo rodiklius galima skaičiuoti tiek tarp atskirų stebėjimų porų, tiek tarp kintamųjų porų.

Matuojamų pagal intervalų skalę kintamųjų sutapimo ir skirtingumo matų skaičiavimui skirtas 16 skyrius. Šiame skyriuje pagrindinį dėmesį skirsime binarinių kintamųjų ryšio matams. Toliau pateikiamos formulės kai kurių binarinių kintamųjų ryšio matams apskaičiuoti. Plačiau galite rasti (Буюль, Цефелъ, 2002 ir Čekanavičius, Murauskas, 2000).

Formulėse simboliais a , b , c ir d pažymėti 2×2 dažnių lentelės (keturlaukės matricos) dažniai. Tokios lentelės pavyzdys galėtų būti atsakymai (*taip* arba *ne*) į du anketos klausimus, kuriuos bendrumo dėlei pavadinsime *1 klausimas* ir *2 klausimas* (13.1 lentelė)

13.1 lentelė. 2×2 dažnių lentelės pavyzdys

	1 klausimas	
2 klausimas	<i>taip</i>	<i>ne</i>
<i>taip</i>	$a = 9$	$b = 7$
<i>ne</i>	$c = 3$	$d = 11$

Lentelėje a reiškia į *1 klausimą* ir *2 klausimą* atsakiusių *taip* skaičių, b — į *1 klausimą* atsakiusių *ne* ir į *2 klausimą* atsakiusių *taip* skaičių, c — į *1 klausimą* atsakiusių *taip* ir į *2 klausimą* atsakiusių *ne* skaičių, d — į *1 klausimą* atsakiusių *ne* ir į *2 klausimą* atsakiusių *ne* skaičių.

Russel and Rao

$$RR = \frac{a}{a + b + c + d},$$

Simple matching

(paprastas atitikimas)

$$SM = \frac{a + d}{a + b + c + d},$$

Rogers and Tanimoto

$$RT = \frac{a + d}{a + d + 2(b + c)},$$

Sokal and Sneath 1

$$SS1 = \frac{2(a + d)}{2(a + d) + b + c},$$

Hamann
$$HAMANN = \frac{(a + d) - (b - c)}{a + b + c + d},$$

Yule Q

(keturlaukės koreliacijos koeficientas)

$$Q = \frac{ad - bc}{ad + bc},$$

Phi 4-point correlation

(tarpusavio sutapimo rodiklis)

$$PHI = \frac{ad - bc}{\sqrt{(a + b)(a + c)(b + d)(c + d)}},$$

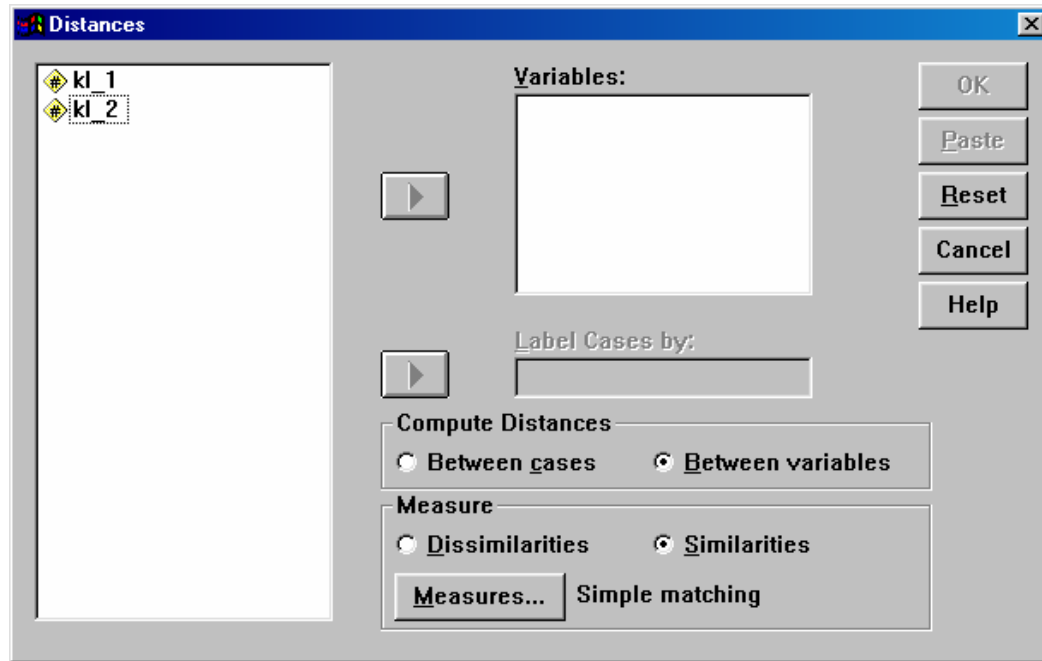
Dispersion

$$DISPER = \frac{ad - bc}{(a + b + c + d)^2}.$$

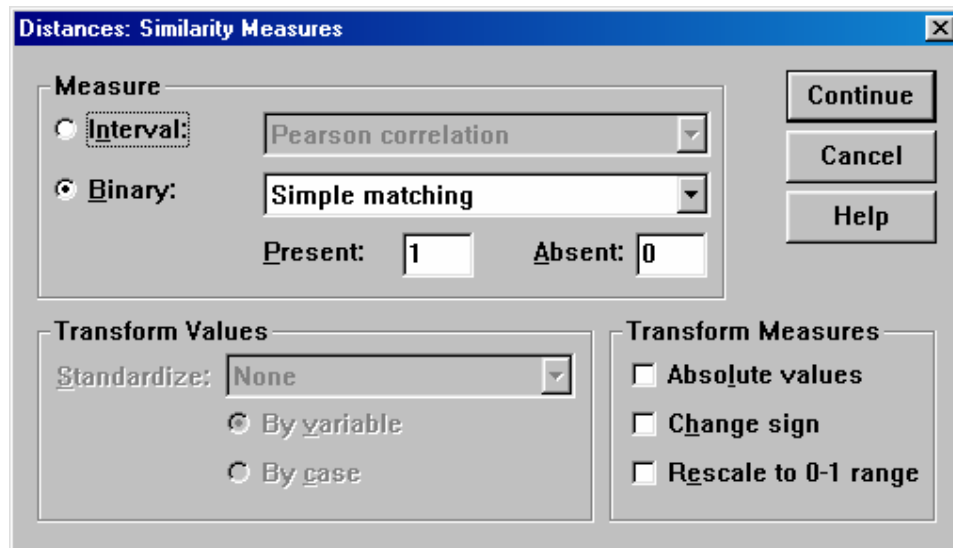
Atstumo (*Distance*) matricai tarp visų analizuojamų kintamųjų arba stebėjimų porų gauti:

- Atidarykite bylą su analizuojamais duomenimis.
- Nurodykite komandas **Analyze → Correlate... → Distances...**
- Dialogo langelyje **Distances** (13.4 pav.) pasirinkite mažiausiai vieną kintamąjį, jeigu norite apskaičiuoti sutapimo (skirtingumo) rodiklius tarp stebėjimų ir mažiausiai du skaitmeninius kintamuosius, jeigu norite apskaičiuoti šiuos rodiklius tarp kintamųjų. Įkelkite pasirinktus kintamuosius į testuojamų kintamųjų sąrašą **Variables**.
- Pasirinkite vieną iš atstumo matų (**Compute Distances**) skaičiavimo variantų — **Between cases** (tarp stebėjimų) arba **Between variables** (tarp kintamųjų).
- Pažymėkite laukelį **Similarities**, jeigu norite apskaičiuoti sutapimo matus, ir spragtelėkite mygtuką **Measures...** (matai). Atsidariusiame dialogo langelyje **Distances: Similarity Measures** (13.5 pav.) nurodykite duomenų tipą — **Interval** ar **Binary** ir atitinkamame išskleidžiamajame sąraše pasirinkite vieną iš tam duomenų tipui tinkančių sutapimo matų.
- Nurodžius binarinių duomenų tipą (**Binary**), nustatytosios binarinių duomenų reikšmės yra šios: **Present** (esantis) — 1, **Absent** (nesantis) — 0. Jeigu Jūsų duomenų rinkmenoje yra nurodytos kitos binarinių duomenų reikšmės, Jūs turite pakoreguoti laukelių **Present** ir **Absent** įrašus.
- Komandų grupės **Transform Values** laukelyje **Standardize** galite nurodyti pageidaujamą pradinių duomenų standartizavimo būdą. Ši transformacija netaikoma binariniais kintamiesiems.

- Komandų grupėje **Transform Measures** galite nurodyti transformuoti apskaičiuotų matų reikšmes į absoliutines (*Absolute values*), pakeisti ženklą (*Change sign*), normuoti 0—1 diapazone (*Rescale to 0-1 range*).
- Spragtelėkite mygtuką **Continue**.



13.4 pav. Dialogo langelis **Distances**



13.5 pav. Dialogo langelis **Distances: Similarity Measures**

- Pažymėkite dialogo langelio **Distances** laukelį **Dissimilarities**, jeigu norite apskaičiuoti skirtingumo matus ir spragtelėkite mygtuką **Measures...** (matai). Atsidariusiame dialogo langelyje **Distances: Dissimilarity Measures** (13.6 pav.) nurodykite duomenų tipą — **Interval**, **Counts** ar **Binary** ir atitinkamame išskleidžiamajame sąrašė pasirinkite vieną iš tam duomenų tipui tinkančių skirtingumo matų (plačiau — 16 skyrius).

13.6 pav. Dialogo langelis **Distances: Similarity Measures**

- Spragtelėkite mygtuką **Continue**.
- Spragtelėkite mygtuką **OK** pagrindiniame dialogo langelyje **Distances**.

Pavyzdys. 13.7 pav. parodyti ryšio mato — **Simple matching** (paprastas atitikimas) skaičiavimo rezultatai, lyginant dviejų respondentų atsakymus į du klausimus (remiantis 13.1 lentelės duomenimis).

Proximity Matrix

	Simple matching Measure	
	KL_1	KL_2
KL_1	1,000	,667
KL_2	,667	1,000

This is a similarity matrix

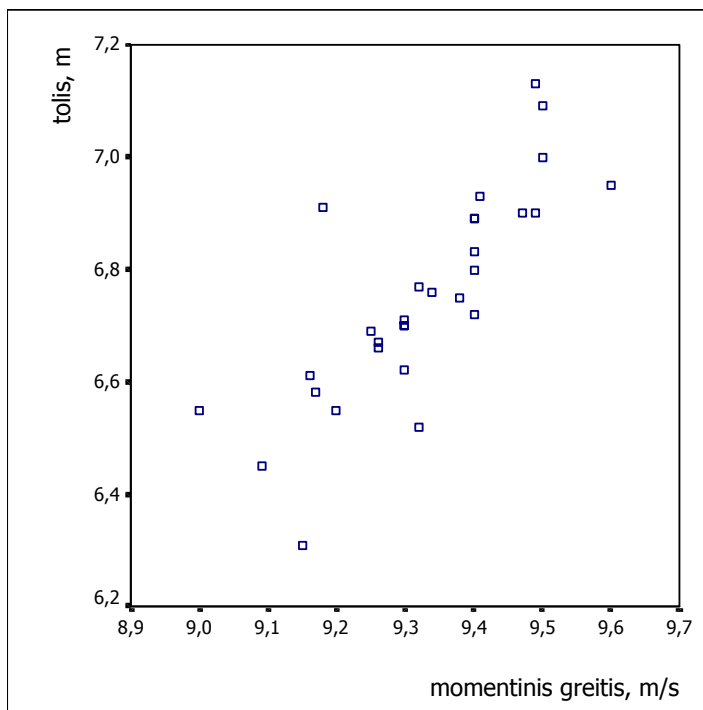
13.7 pav. Ryšio mato **Simple matching** skaičiavimo rezultatai

14. REGRESINĖ ANALIZĖ

Koreliacija apibūdina ryšio tarp kintamųjų stiprumą, o regresinė analizė įgalina nustatyti šio ryšio pobūdį ir aprašyti priklausomojo (pasekmės) kintamojo vidutinių reikšmių priklausomybę nuo vieno ar kelių nepriklausomųjų (priežasties) kintamųjų reikšmių matematine formule ir kartu — prognozuoti šio kintamojo reikšmes. SPSS paketas turi dideles regresinės analizės galimybes. Šioje knygoje aptarsime tik plačiausiai naudojamus ir sportiniams tyrimams taikomus regresinės analizės metodus: paprastą tiesinę regresiją (vienas nepriklausomas kintamasis) ir daugelio faktorių (daugialypę) tiesinę regresiją (keletas nepriklausomų kintamųjų), binarinę logistinę regresiją, daugiareikšmę logistinę regresiją, netiesinio (kreivių) priartėjimo metodus.

14.1. PAPRASTA TIESINĖ REGRESIJA

Akivaizdus tiesinės priklausomybės pavyzdys galėtų būti ryšys tarp momentinio bėgimo greičio ir šuolio į tolį rezultato, kurių koreliacinis laukas (sklaidos diagrama) parodyta 14.1 pav. Sklaidos diagramoje galima išvelgti tiesioginę kintamųjų priklausomybę, o taškų, atitinkančių stebimus rezultatus, visuma koncentruojasi palei tiesią.



14.1 pav. Koreliacinis laukas tarp momentinio greičio ir šuolio į tolį rezultatų

Bendriausias tiesinis tikimybinis modelis, nusakantis priklausomo intervalinio kintamojo Y ir nepriklausomo intervalinio kintamojo X sąryšį, užrašomas šia stochastine lygtimi

$$Y = a + bX + e, \quad (14.1)$$

čia a ir b yra nežinomos konstantos, e — atsitiktinė paklaida. Regresijos tiesės lygtis yra

$$\hat{y}(x) = \hat{a} + \hat{b}x, \quad (14.2)$$

čia \hat{a} ir \hat{b} yra parametrų a ir b įverčiai. Koefficientų \hat{a} ir \hat{b} reikšmės nustatomos vadinamu mažiausiųjų kvadratų metodu — minimizuojant sumą

$$SSE = \sum_{i=1}^n (y_i - a - bx_i)^2 \quad (14.3)$$

Tiesinės regresijos prielaidos: kiekvienai fiksuotai nepriklausomo kintamojo reikšmei x_i priklausomo kintamojo Y_i reikšmės turi būti pasiskirsčiusios pagal normalųjį dėsnį, priklausomo kintamojo Y_i dispersija turi būti lygi esant visoms nepriklausomo kintamojo reikšmėms x_i (homoskedastiškumo reikalavimas), ryšys tarp priklausomo kintamojo ir nepriklausomo kintamojo turi būti tiesinis.

14.1.1. Regresijos lygties skaičiavimas

Regresijos lygties koeficientams apskaičiuoti:

- Atidarykite bylą su analizuojamais duomenimis.
- Nurodykite komandas **Analyze → Regression... → Linear...**
- Dialogo langelyje **Linear Regression** (14.2 pav.) priklausomą (pasekmės) kintamąjį įkelkite į laukelį **Dependent**, o nepriklausomą (priežasties) kintamąjį — į laukelį **Independent(s)**.
- Nekeisdami kitų nustatymų spragtelėkite mygtuką **OK**.

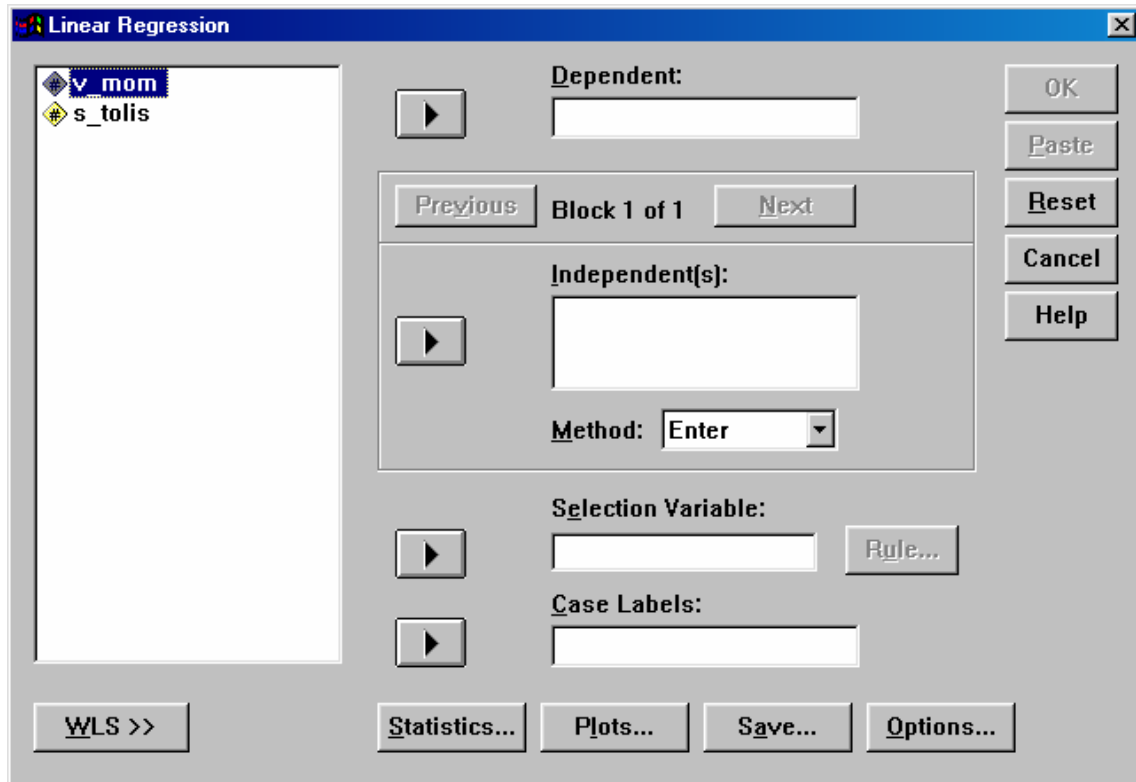
Pavyzdys. 14.3 pav. parodyti regresijos lygties tarp momentinio greičio — nepriklausomas kintamasis v_mom ir šuolio į tolį rezultato — priklausomas kintamasis s_tolis pagrindiniai skaičiavimo rezultatai.

Lentelės *Coefficients* stulpelio *Unstandardized Coefficients B* dalyje pateikiamas poslinkis pagal ordinačių ašį \hat{a} (*Constant*) ir regresijos koeficientas \hat{b} . Taigi tiesioginė regresijos lygtis bus tokia

$$s_tolis = 1,094 \cdot v_mom - 3,452.$$

Stulpelyje *Beta* pateikiamas standartizuotų duomenų ($y_i \sim N(0, \sigma^2)$) regresijos lygties koeficientas. Paprastosios tiesinės regresijos atveju šis koeficientas sutampa su pradinių duomenų koreliacijos koeficientu.

Stulpeliuose *t* ir *Sig.* Pateikiamos hipotezių $H_0: a = 0$ ir $H_0: b = 0$ statistikos ir *p*-reikšmės.



14.2 pav. Dialogo langelis *Linear Regression*

Nulinė hipotezė H_0 atmetama (koeficientai statistiškai reikšmingai skiriasi nuo 0), jeigu *p*-reikšmė $p < \alpha$, čia α yra pasirinktasis reikšmingumo lygmuo.

Šiuo atveju pirmoji statistinė hipotezė $H_0: a = 0$ yra neaktuali, o antrąją statistinę hipotezę $H_0: b = 0$ atmetame, nes *p*-reikšmė lygi $0,000 < 0,05$. Taigi, prognozuojant šuolio į tolį rezultatą, greitis yra svarbus.

Paskutiniuose dviejuose lentelės *Coefficients* stulpeliuose pateikti koeficientų *a* ir *b* 95% pasikliautinųjų intervalų režiai.

Lentelėje *ANOVA* pateikta kvadratų suma (*Sum of Squares*): regresijos kvadratų suma (*Regression*), liekamųjų paklaidų kvadratų suma (*Residual*) ir visa kvadratų suma (*Total*).

Lentelėje *Model Summary* pateiktas determinacijos koeficientas r^2 (*R Square*) ir pataisytasis determinacijos koeficientas (*Adjusted R Square*) r_{adj}^2 .

Determinacijos koeficientą galima interpretuoti kaip santykį dispersijos dalies, kurią paaiškina regresijos modelis, su visa dispersija. Paprastosios

tiesinės regresijos atveju determinacijos koeficientas sutampa su empirinio Pirsono koreliacijos koeficiento kvadratu. Jeigu duomenys idealiai atitinka regresijos tiesės lygtį, t. y. visos y_i patenka ant regresijos tiesės, determinacijos koeficientas yra lygus 1. Jeigu regresijos tiesės lygtis visiškai netinka prognozei, jis lygus 0.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,801 ^a	,641	,628	,11442

a. Predictors: (Constant), V_MOM

a)

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	,655	1	,655	50,022	,000 ^a
	Residual	,367	28	,013		
	Total	1,022	29			

a. Predictors: (Constant), V_MOM

b. Dependent Variable: S_TOLIS

b)

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-3,452	1,443		-2,393	,024
	V_MOM	1,094	,155	,801	7,073	,000

a. Dependent Variable: S_TOLIS

c)

14.3 pav. Pagrindinių regresinės analizės rezultatų išvesties pavyzdys

14.1.2. Naujų kintamųjų išsaugojimas

Daugelį pagalbinių reikšmių, gaunamų skaičiuojant regresijos lygtį, galima išsaugoti kaip naujus kintamuosius ir panaudoti tolesniems skaičiavimams.

- Spragtelėkite dialogo langelio **Linear Regression** mygtuką **Save...**
- Atsidarys naujas dialogo langelis **Linear Regression: Save** (14.4 pav.), kuriame galite nurodyti išsaugoti prognozuojamas reikšmes (**Predicted Values**), liekanas (**Residuals**) ir kitas charakteristikas.

Linear Regression: Save

Predicted Values

- ☒ Unstandardized
- ☐ Standardized
- ☐ Adjusted
- ☐ S.E. of mean predictions

Distances

- ☐ Mahalanobis
- ☐ Cook's
- ☐ Leverage values

Prediction Intervals

- ☐ Mean ☐ Individual
- Confidence Interval: 95 %

Residuals

- ☐ Unstandardized
- ☐ Standardized
- ☐ Studentized
- ☐ Deleted
- ☐ Studentized deleted

Influence Statistics

- ☐ DfBeta(s)
- ☐ Standardized DfBeta(s)
- ☐ DfFit
- ☐ Standardized DfFit
- ☐ Covariance ratio

Save to New File

- ☐ Coefficient statistics **File...**

Export model information to XML file

Browse

Continue **Cancel** **Help**

14.4 pav. Dialogo langelis **Linear Regression: Save**

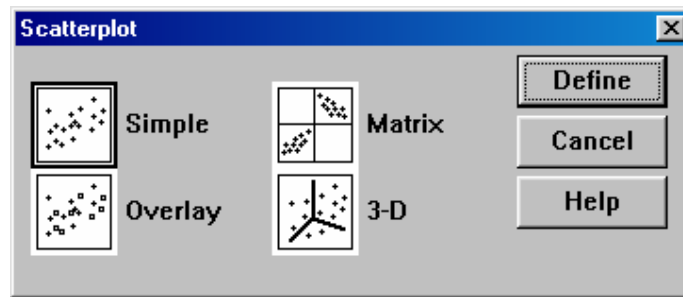
Dažniausiai tikslinga išsaugoti prognozuojamas reikšmes — **Unstandardized** (nestandardizuotos reikšmės) ir **Standardized** (standartizuotos reikšmės). Pasirinkus **Unstandardized**, bus skaičiuojamos regresijos lygties reikšmės kiekvienai stebėjimo (nepriklausomo kintamojo) reikšmei. Pasirinkus **Standardized** variantą regresijos lygties reikšmės normalizuojamos. Kiekvienam apskaičiuotam kintamajam nepriklausomai nuo to, ar skaičiuojamos prognozuojamos reikšmės, liekanos ar kitos charakteristikos, SPSS pakete automatiškai suteikiamas naujas vardas. Nestandardizuotoms reikšmėms suteikiami vardai *pre_1* (*predicted value*), *pre_2* ir t. t., o standartizuotoms *zpr_1* ir t. t. Nauji kintamieji bus pateikti duomenų rinkmenoje pradinių kintamųjų sąrašo gale. Norėdami pasinaudoti regresijos galimybėmis prognozavimo tikslais, įrašykite naujas nepriklausomo kintamojo reikšmes pradinių duomenų sąrašo gale ir pakartokite regresijos lygties skaičiavimo

procedūrą (išsaugojant prognozuojamas reikšmes) — naujas kintamasis apims reikšmes, atitinkančias Jūsų įrašytus naujus duomenis.

14.1.3. Regresijos tiesės braižymas

Skaidos diagramoje regresijos tiesei pavaizduoti:

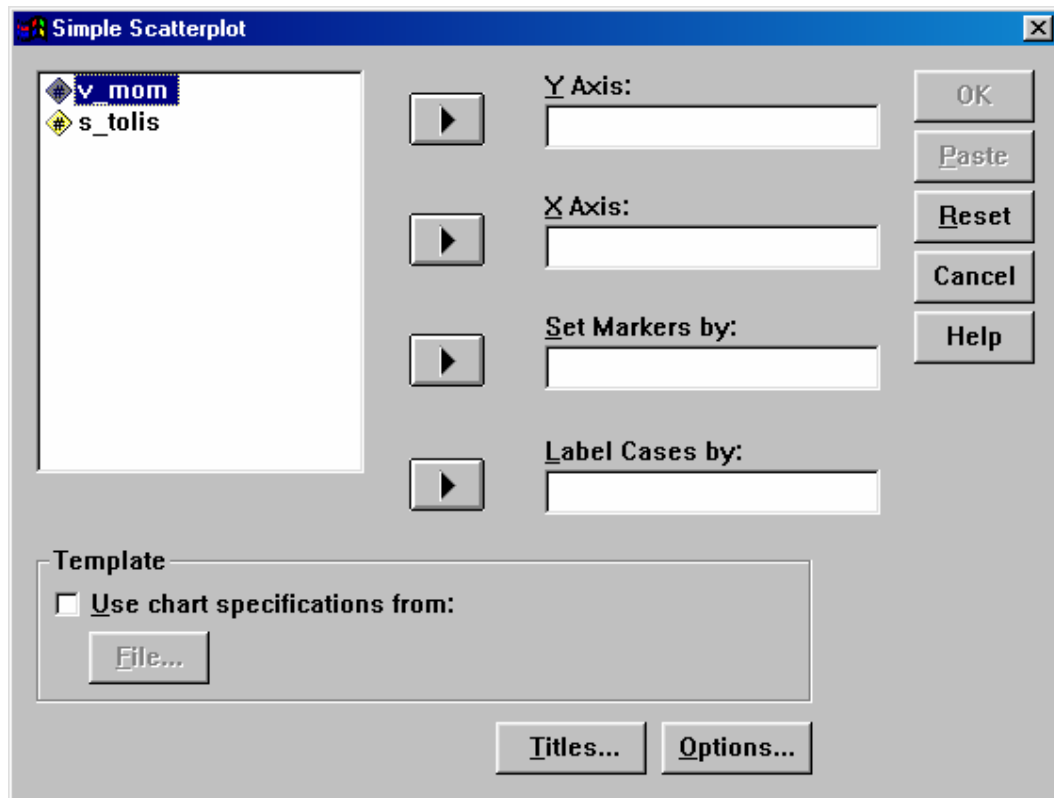
- Nurodykite meniu komandas **Graphs... → Scatter...**
- Atsidariusiame dialogo langelyje **Scatterplot** (14.5 pav.) palikite pirminį nustatymą — **Simple** (paprasta) ir spragtelėkite mygtuką **Define** (nustatyti).



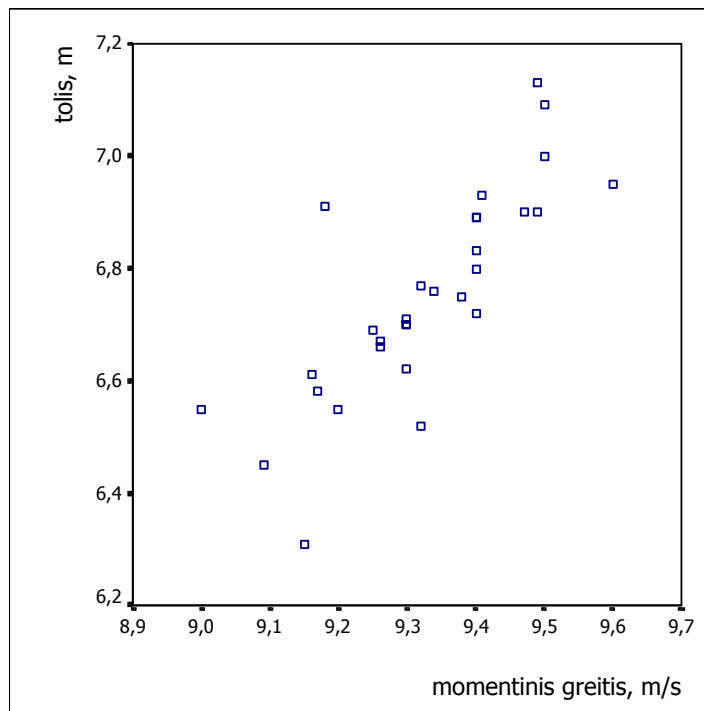
14.5 pav. Dialogo langelis **Scatterplot** skaidos diagramos tipui pasirinkti

- Atsidarys naujas dialogo langelis **Simple Scatterplot** (paprasta skaidos diagrama, 14.6 pav.), kuriame į laukelį **Y Axis** (Y ašis) perkeltite priklausomą kintamąjį, o į laukelį **X Axis** (X ašis) — nepriklausomą kintamąjį.
- Spragtelėję dialogo langelio **Simple Scatterplot** mygtuką **Titles...** naujame dialogo langelyje **Titles** įrašykite diagramos pavadinimą ir spragtelėkite mygtuką **Continue**.
- Spragtelėkite pagrindinio dialogo langelio **Simple Scatterplot** mygtuką **OK**.

Tolesnius veiksmus paaiškinsime remdamiesi mūsų anksčiau pateiktu pavyzdžiu. Įkėlę kintamąjį *s_tolis* į dialogo langelio **Simple Scatterplot** laukelį **Y Axis**, o kintamąjį *v_mom* — į laukelį **X Axis** ir spragtelėję mygtuką **OK**, **Viewer** lange gausime 14.7 pav. parodytą skaidos diagramą. Norėdami koreguoti pirminę diagramą — pakeisti ašių pavadinimų šriftą ir vietą, ašių diapazonus, pasirinkti pageidaujama spalvą ir t. t., dukart ją spragtelėkite. Diagrama bus įkelta į diagramų redaktoriaus langą (žr. 8.3 skyrių). Dukart spragtelėkite pele tą diagramos objektą (diagramos pavadinimą, X ašį, Y ašį, duomenų seką ir t. t.), kurio formatus norite keisti, atsidariusiame dialogo langelyje padarykite reikiamus pakeitimus ir spragtelėkite mygtuką **OK**. Kai diagrama bus sutvarkyta pagal Jūsų poreikius, uždarykite diagramų redaktoriaus langą — visi padaryti pakeitimai atsispindės **Viewer** lange.



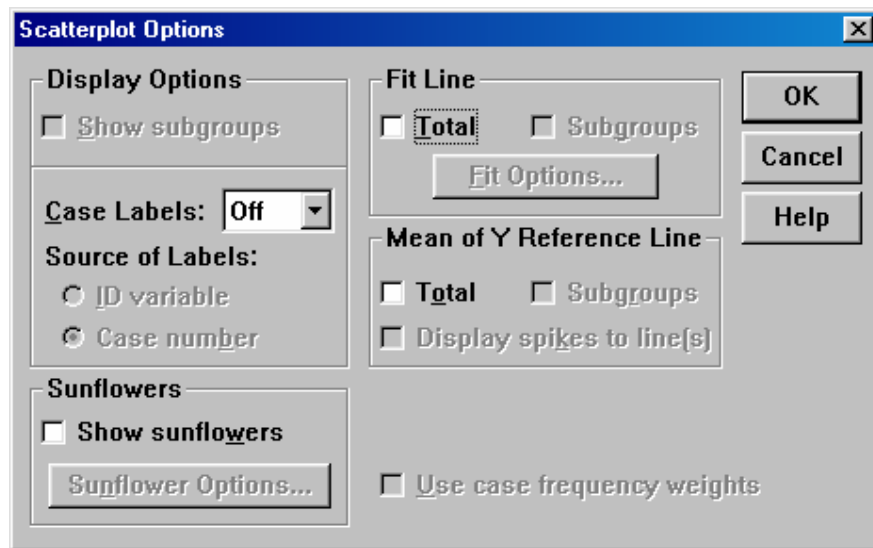
14.6 pav. Dialogo langelis **Simple Scatterplot** sklaidos diagramai sudaryti



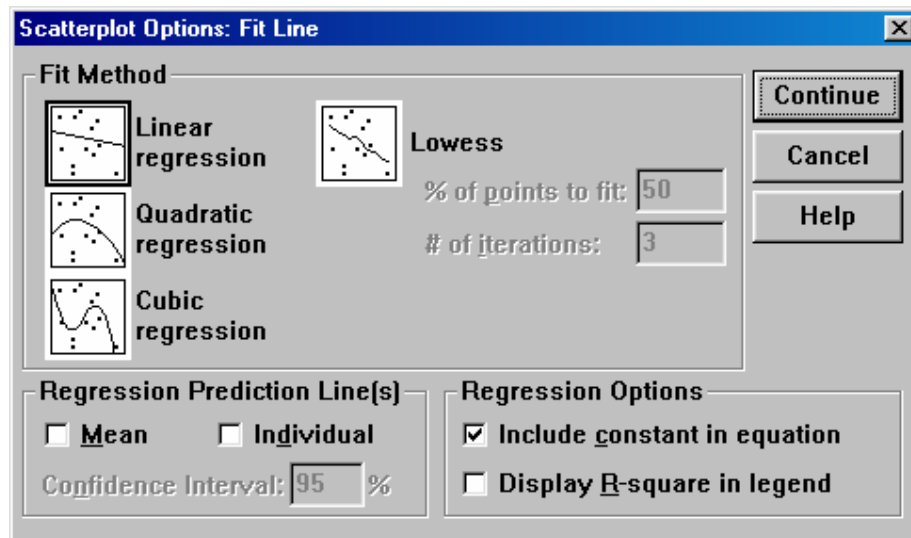
14.7 pav. Momentinio greičio ir šuolio į tolį rezultatų sklaidos diagrama

- Dukart spragtelime diagramą ir įkeliame ją į diagramų redaktoriaus langą.
- Diagramų redaktoriaus lange nurodome komandas **Chart... → Options...**. Atsidarys dialogo langelis **Scatterplot Options** (14.8 pav.).
- Dialogo langelio skyriuje **Fit Line** (pritaikyta kreivė) pažymėkime žymimąjį langelį **Total** (visai duomenų rinkmenai) ir spragtelime mygtuką **Fit Options** (pritaikymo variantai). Atsidarys naujas dialogo langelis **Scatterplot Options: Fit Line** (14.9 pav.).
- Įsitikiname, kad nustatytasis variantas yra **Linear Regression** (tiesinė regresija) ir spragtelime mygtuką **Continue**.
- Spragtelime mygtuką **OK** dialogo langelyje **Scatterplot Options**, uždarome diagramų redaktoriaus langą ir spragtelime šalia diagramos **Viewer** lange. Sklaidos diagramoje bus pavaizduota regresijos tiesė (14.10 pav.).

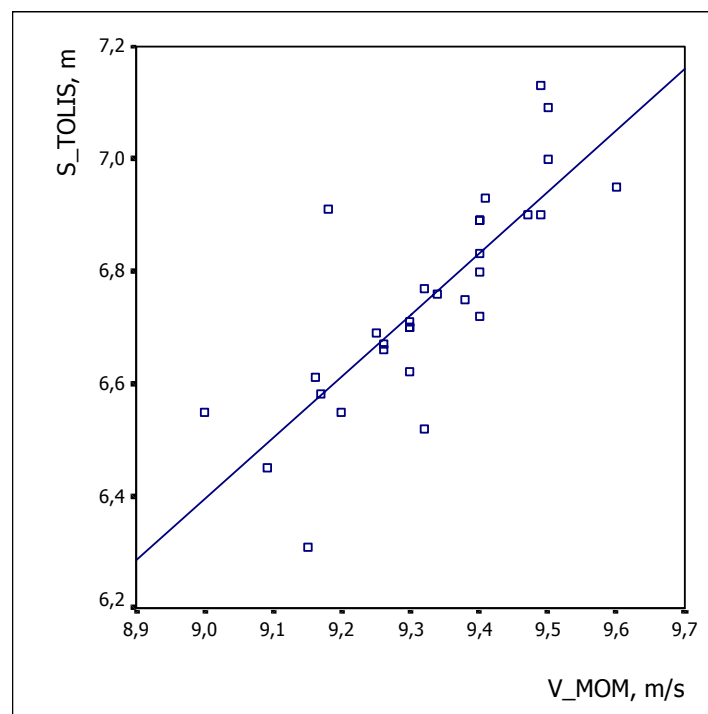
Pažymėję dialogo langelio **Scatterplot Options: Fit Line** skyriuje **Regression Prediction Line(s)** žymimąjį langelį **Mean** ar **Individual** gausite dvi priklausomo kintamojo vidutinės prognozuojamos reikšmės pasikliautinojo intervalo (**Confidence Interval**) kreives arba dvi priklausomo kintamojo prognozuojamų reikšmių pasikliautinojo intervalo (**Confidence Interval**) kreives (14.11 pav.). Akivaizdu, kad vidurkio pasikliautinis intervalas yra siauresnis už atskirų reikšmių pasikliautinąjį intervalą. Nustatytoji pasikliautino intervalo reikšmė yra 95%.



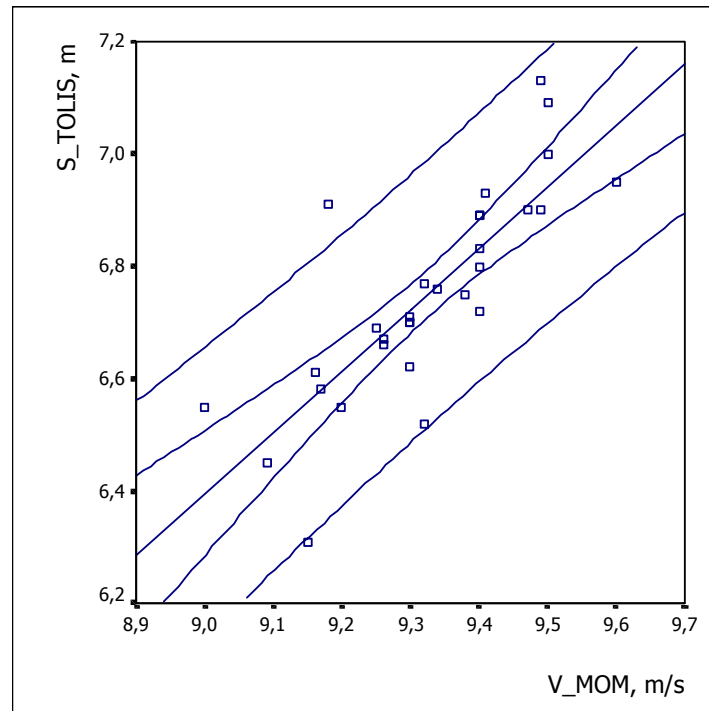
14.8 pav. Dialogo langelis **Scatterplot Options**



14.9 pav. Dialogo langelis *Scatterplot Options: Fit Line*



14.10 pav. Momentinio greičio ir šuolio į tolį rezultatų sklaidos diagrama su pavaizduota regresijos tiese



14.11 Sklaidos diagrama su regresijos tiese ir pavaizduotomis priklausomo kintamojo vidurkio ir atskirų reikšmių 95% pasikliautinąjį intervalą žyminčiomis kreivėmis

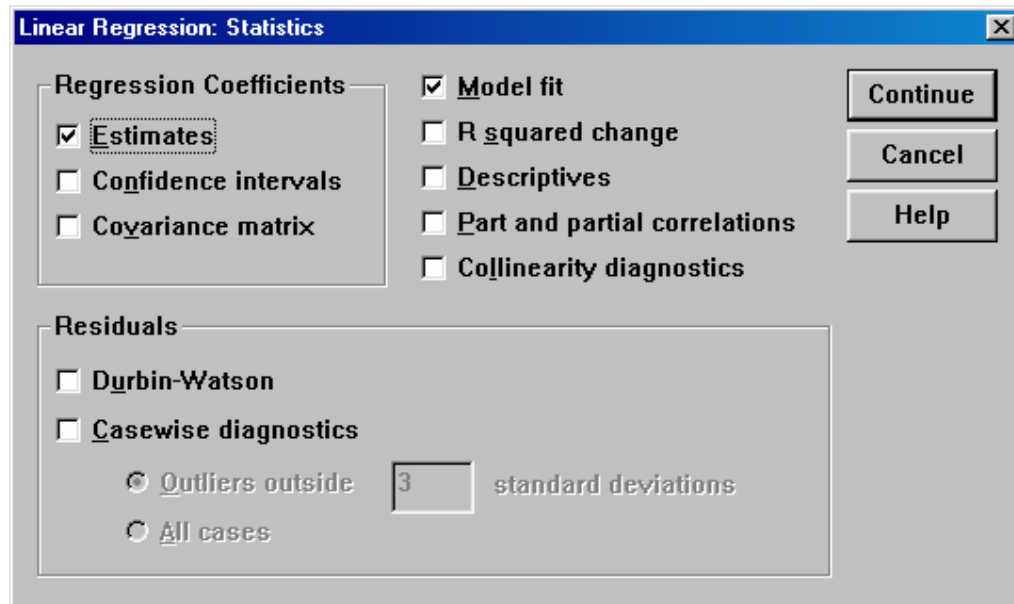
14.1.4. Kitos tiesinės regresijos charakteristikos

Spragtelėję dialogo langelio **Linear Regression** mygtuką **Plots...** galite pasirinkti sklaidos diagramas tarp bet kurių dviejų dialogo langelyje **Linear Regression: Plots** nurodytų kintamųjų (*ZPRED* — standartizuotos prognozuojamos reikšmės, *ZRESID* — standartizuotos liekanos ir t. t.), kurios padeda aptikti žymiai besiskiriančias nuo kitų savo dydžiu pavienes kintamųjų reikšmes — išskirtis (*outliers*), klaidingas reikšmes ir t. t. Pažymėję šio dialogo langelio skyriuje **Standardized Residual Plots** žymimąjį langelį **Histogram** ir **Normal probability plot**, gausite standartizuotų liekanų histogramą ir normaliojo skirstinio kreivę, taip pat prielaidą patikrinti, kad priklausomo kintamojo reikšmės kiekvienai nepriklausomo kintamojo reikšmei pasiskirsčiusios pagal normalųjį dėsnį.

Spragtelėję dialogo langelio **Linear Regression** mygtuką **Statistics...** dialogo langelyje **Linear Regression: Statistics** (14.12 pav.) galite pasirinkti šias charakteristikas:

- **Regression Coefficients** (regresijos koeficientai): **Estimates** (įverčiai) — nustatytasis variantas — pateikiamos regresijos koeficientų, jų standartinės paklaidos reikšmės; **Confidence intervals** (pasikliautiniai intervalai); **Covariance matrix** (kovariacinė matrica).

- **Model fit** (modelio tinkamumas) — nustatytasis variantas — pateikiamos determinacijos koeficiento, pataisyto determinacijos koeficiento, standartinės paklaidos reikšmės, taip pat dispersinės analizės (regresijos ANOVA) rezultatai.
- **Descriptives** (aprašomosios statistikos charakteristikos) — pateikiamas kiekvieno kintamojo stebėjimų skaičius, vidurkis, standartinis nuokrypis.
- **Durbin-Watson** — Durbino ir Watsono kriterijus, kuris taikomas prielaidai apie liekamųjų paklaidų nepriklausomumą (taip pat, kad visi Y_i yra nepriklausomi, t. y. nėra autokoreliacijos) patikrinti. Durbino ir Watsono statistika d kinta nuo 0 iki 4. Kuo d arčiau 2, tuo mažiau tikėtina, kad tarp liekamųjų paklaidų (ir tarp kintamojo Y reikšmių) yra autokoreliacija.
- Kitas šio dialogo langelio galimybes panagrinėsime tolesniame skyrelyje, skirtame daugialypei regresijai.



14.12 pav. Dialogo langelis **Linear Regression: Statistics**

14.2. DAUGELIO FAKTORIŲ TIESINĖ REGRESIJA

Daugelio faktorių (daugialypėje) tiesinėje regresijoje yra nagrinėjami ryšiai tarp priklausomo kintamojo ir kelių nepriklausomų kintamųjų. Šiuo atveju, regresijos funkcija užrašoma formule

$$\hat{y}(\vec{x}) = \hat{a} + \hat{b}_1 x_1 + \hat{b}_2 x_2 + \dots + \hat{b}_k x_k, \quad (14.4)$$

čia x_1, \dots, x_n — nepriklausomi kintamieji, \hat{a} ir $\hat{b}_1, \dots, \hat{b}_k$ yra parametrų a ir b_1, \dots, b_k įverčiai. Koefficientų \hat{a} ir $\hat{b}_1, \dots, \hat{b}_k$ reikšmės nustatomos mažiausiųjų kvadratų metodu. Daugialypės tiesinės regresijos prielaidos yra tos pačios kaip ir paprastosios regresijos. Papildomai reikalaujama, kad vienas nepriklausomas kintamasis nebūtų likusiųjų tiesinė daugdara. Be to, reikia atsižvelgti į tai, kad nepriklausomais laikomi kintamieji gali koreliuoti tarpusavyje.

Daugialypės regresijos lygties koeficientams apskaičiuoti, jų nelygybės nuliui reikšmingumui patikrinti, taip pat daugialypės regresijos prielaidų galiojimui patikrinti:

- Atidarykite bylą su analizuojamais duomenimis.
- Nurodykite komandas **Analyze → Regression... → Linear...**
- Dialogo langelyje **Linear Regression** (14.2 pav.) priklausomą (pasekmės) kintamąjį įkelkite į laukelį **Dependent**, o nepriklausomus (priežasties) kintamuosius — į laukelį **Independent(s)**.

Išskleidžiamajame sąraše **Method** galima pasirinkti, kaip nepriklausomi kintamieji bus įtraukiami į skaičiavimą. Nustatytasis **Enter** variantas, pagal kurį visi atrinkti nepriklausomi kintamieji įtraukiami į skaičiavimą vienu metu, rekomenduojamas taikyti tik tada, kai yra vienas nepriklausomas kintamasis. Daugelio faktorių analizei paprastai pasirenkamas vienas iš laipsniškų metodų. Tiesioginis (**Forward**) metodas užtikrina laipsnišką nepriklausomų kintamųjų, turinčių didžiausius dalinės koreliacijos su priklausomu kintamuoju koeficientus, įtraukimą į regresijos lygties skaičiavimą. Pasirinkus atvirkštinį (**Backward**) metodą, skaičiavimas pradedamas su visais nepriklausomais kintamaisiais, pašalinant tolimesnių skaičiavimų metu kintamuosius su mažiausiais dalinės koreliacijos koeficientais. Plačiausiai taikomas **Stepwise** (po žingsnį) metodas. Pasirinkus šį metodą, nepriklausomi kintamieji gali būti pateikiami skaičiavimui blokais, kurie kiekviename žingsnyje apdorojami kartu.

- Spragtelėję dialogo langelio **Linear Regression** mygtuką **Statistics...** dialogo langelyje **Linear Regression: Statistics** (14.12 pav.) pasirinkite šias daugialypės regresijos modelį apibūdinančias charakteristikas:
 - **Durbin-Watson** — Durbino ir Vatsono kriterijų prielaidai apie liekamųjų paklaidų nepriklausomumą patikrinti. Šio kriterijaus taikymas nesikiria nuo aprašytojo 14.1.4 skyrelyje.
 - **Collinearity diagnostics**. Patikrinama, ar nepriklausomi kintamieji nėra multikolinearūs, t. y. ar tarp jų nėra stiprai koreliuojančių — daugialypės regresijos modelis tinkamiausias prognozuoti tada, kai visi nepriklausomi kintamieji X_1, \dots, X_k tarpusavyje nekoreliuoja, o priklausomybė sieja tik juos ir priklausomą kintamąjį Y . Tam yra skaičiuojamas dispersijos mažėjimo rodiklis **VIF** (**Variance Inflation Factor**). Paprastai laikoma, kad kintamasis yra „per daug multikolinearus“, jeigu $VIF > 4$.
 - **Part and partial correlations** — dalinės koreliacijos koeficientai.

- Spragtelėkite mygtuką **OK**.

Išvestyje yra pateikiamos regresijos funkcijos koeficientų reikšmės ir statistinės hipotezės apie jų nelygybę nuliui patikrinimo rezultatai. Lentelėje *ANOVA* gauname vieną iš dviejų išvadų: 1) visi koeficientai $b_j = 0$, t. y. regresijos modelis su duomenimis nesuderintas; 2) bent vienas koeficientas $b_j \neq 0$, t. y. regresijos modelis bent jau iš dalies prognozėms tinka. Hipotezė H_0 (visi koeficientai lygūs nuliui) atmetama, kai p -reikšmė $p < \alpha$, čia α — nustatytas reikšmingumo lygmuo. Lentelėje *Coefficients* pagal Stjudento kriterijų sprendžiama, ar konkretus $b_j \neq 0$. Nulinė hipotezė H_0 (koeficientas lygus nuliui) atmetama, kai p -reikšmė $p < \alpha$. Pagal abu anksčiau paminėtus kriterijus nustatome, ar koeficientai b_j statistiškai reikšmingai skiriasi nuo nulio, ir pagal tai sprendžiame, ar prognozuojamos Y priklauso nuo X_j .

Papildomą informaciją apie priklausomybės stiprumą ir apie tai, kurie kintamieji prognozei svarbesni, pateikia standartizuotieji koeficientai (*Standardized Coefficients, Beta*). Absoliučiu didumu didesnė šio koeficiento reikšmė rodo didesnę Y priklausomybę nuo X_j .

Kaip ir vieno kintamojo atveju, nepriklausomų kintamųjų įtakai Y įgyjamoms reikšmėms vertinti lentelėje *Model Summary* pateikiamas determinacijos koeficientas (*R Square*), koreguotasis determinacijos koeficientas (*Adjusted R Square*), taip pat daugialypės koreliacijos koeficientas R (kvadratinė šaknis iš determinacijos koeficiento).

Pateikiamos kitos dialogo langelyje **Linear Regression: Statistics** pasirinktos charakteristikos.

14.3. BINARINĖ LOGISTINĖ REGRESIJA

Binarinė logistinė regresija yra naudojama tyrinėti binarinių kintamųjų priklausomybei nuo bet kurios skalės nepriklausomų kintamųjų (kategoriniai nepriklausomi kintamieji dirbtinai perkoduojami į keletą binarinių kintamųjų). Kadangi binarinis kintamasis paprastai reiškia įvykį, kuris gali įvykti arba neįvykti, naudojant binarinės logistinės regresiją, yra apskaičiuojama šio įvykio tikimybė, sąlygojama nepriklausomų kintamųjų. Tikimybė p_i , kad atsitiktinis dydis Y_i įgys reikšmę 1, apskaičiuojama pagal formulę

$$p_i = \frac{\exp\{z(\vec{x}_i)\}}{1 + \exp\{z(\vec{x}_i)\}}, \quad z(\vec{x}_i) = a + b_1 x_{1i} + b_2 x_{2i} + \dots + b_k x_{ki}, \quad (14.5)$$

čia x_{1i}, \dots, x_{ki} — nepriklausomų kintamųjų reikšmės.

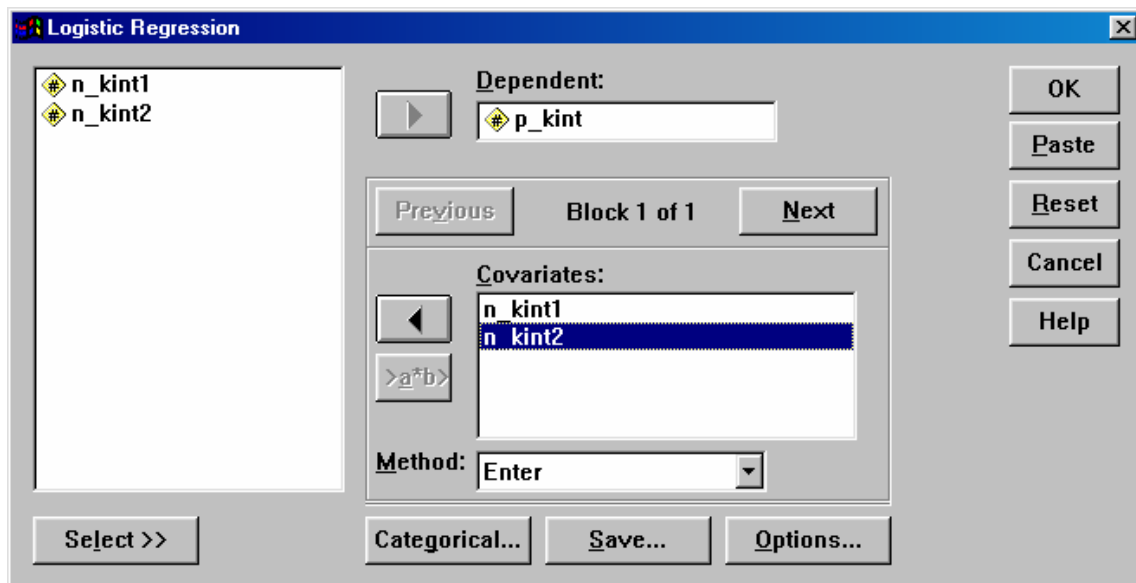
Parametrų a, b_1, \dots, b_k įverčiai $\hat{a}, \hat{b}_1, \dots, \hat{b}_k$ parenkami taip, kad tikėtinumo funkcija (žr. Čekanavičius, Murauskas, 2002) būtų didžiausia.

Jeigu $P(Y=1|\vec{x})$ įvertis yra daugiau negu 0,5, prognozuojama, kad Y reikšmė yra 1; jeigu $P(Y=1|\vec{x})$ įvertis yra mažiau negu 0,5, prognozuojama, kad Y reikšmė yra 0.

Binarinės logistinės regresijos koeficientai ir jų reikšmingumo rodikliai skaičiuojami taip:

- Atidarykite bylą su analizuojamais duomenimis.
- Nurodykite komandas **Analyze → Regression... → Binary Logistic...**
- Dialogo langelyje **Logistic Regression** (14.13 pav.) priklausomą (pasekmės) kintamąjį įkelkite į laukelį **Dependent**, o nepriklausomus (priežasties) kintamuosius — į laukelį **Covariates**.

Išskleidžiamajame sąraše **Method** galima pasirinkti, kaip nepriklausomi kintamieji bus įtraukiami į skaičiavimą. Nustatytas yra **Enter** variantas, kai visi atrinkti nepriklausomi kintamieji įtraukiami į skaičiavimą vienu metu. Alternatyva šiam metodui yra progresinė ir atgalinė atranka. Taikant tiesioginės atrankos (**Forward**) metodus, iš pradžių nustatomos konstantos, o toliau laipsniškai įtraukiami nepriklausomi kintamieji, turintys stiprų koreliacinį ryšį su priklausomu kintamuoju. Tada, remiantis atitinkamais kriterijais tikrinama, kurie kintamieji turi būti pašalinti iš regresijos lygties. Taikant atgalinės atrankos (**Backward**) metodus, skaičiavimas pradedamas su visais nepriklausomais kintamaisiais, pašalinant tolesnio darbo metu priklausomam kintamajam mažai įtakos turinčius nepriklausomus kintamuosius. Jeigu yra tik vienas nepriklausomas kintamasis, būtina pasirinkti **Enter** variantą. Daugelio kintamųjų atveju reikėtų pasirinkti vieną iš laipsniškųjų metodų.



14.13 pav. Dialogo langelis **Logistic Regression**

- Spragtelėkite mygtuką **Categorical...**, jeigu tarp nepriklausomų kintamųjų yra kategorinių (t. y. nominaliųjų kintamųjų) ir naujame dialogo langelyje **Logistic Regression: Define Categorical Variables** įkelkite kategorinius kintamuosius į sąrašą **Categorical Covariates**. Kontrastų išskleidžiamajame sąraše **Contrast** palikite nustatytąjį variantą **Indicator** arba pasirinkite taip pat dažnai naudojamą **Deviation**. Kategorinis kintamasis bus suskaldytas į fiktyvius binarinius kintamuosius (pseudokintamuosius), kurių skaičius bus lygus kategorijų skaičiui. Spragtelėkite dialogo langelio **Logistic Regression: Define Categorical Variables** mygtuką **Continue** ir grįžkite į pagrindinį dialogo langelį.
- Spragtelėję mygtuką **Save...** galite išsaugoti duomenų rinkmenoje papildomus kintamuosius: **Predicted Values** (prognozuojamas reikšmės), **Residuals** (liekanas), **Influence** (stebėjimų įtaką prognozuojamoms reikšmėms).
- Spragtelėkite mygtuką **OK**.

Išvestyje pateikiamus binarinės logistinės regresijos rezultatus ir statistines išvadas paaiškinsime remdamiesi pavyzdžiu.

Pavyzdys. Studentai įvaldė naują kompiuterinę programą ir turėjo atlikti testą. Priklausomas kintamasis *test* lygus 1, jeigu testo rezultatai teigiami, ir lygus 0, jeigu testo rezultatai neigiami. Kiek valandų pratybų dirbo kiekvienas studentas, rodo intervalinis kintamasis *val*. Ar studentas turėjo darbo su panašiomis programomis patirtį, rodo binarinis kintamasis *patirtis* (1 — *taip*, 0 — *ne*). Duomenų rinkmena parodyta 14.14 pav. Įvertinsime tikimybę, kad studento, kuris dirbo 30 val. ir turėjo darbo su panašiomis programomis patirtį, testo rezultatai bus teigiami. Pagrindiniai skaičiavimo rezultatai pateikti 14.15 pav.

Lentelėje *Variables in the Equation* pateikiamos apskaičiuotų koeficientų $\hat{a}, \hat{b}_1, \hat{b}_2$ reikšmės (*B* stulpelyje) ir jų nelygybės nuliui reikšmingumas. Wald'o kriterijus yra Stjudento kriterijaus tiesinėje regresijoje analogas, t. y. jis atsako į klausimą, ar konkretus koeficientas $b_j \neq 0$. Hipotezė H_0 atmetama (taigi $b_j \neq 0$), jeigu p -reikšmė $p < \alpha$. Hipotezė H_0 neatmetama, jeigu $p \geq \alpha$, čia α — reikšmingumo lygmuo. Visi šio pavyzdžio regresijos lygties koeficientai reikšmingai nelygūs nuliui, nes p -reikšmė (*Sig.*) yra mažesnė už 0,05. Šios lentelės *Exp(B)* stulpelyje pateikiama galimybių santykio reikšmė. Galimybių santykis rodo, kaip kinta Y galimybė įgyti reikšmę 1. Šiuo pavyzdžiu galime teigti, kad studentas, turintis darbo su analogiškais programomis patirtį, padidino savo galimybę išlaikyti testą 8,444 karto. Analogiškai viena papildoma pratybų valanda padidina galimybę išlaikyti testą 1,305 karto.

Tikrinant, ar logistinės regresijos modelis tinka prognozėms, būtina patikrinti, ar dažnai žinomos reikšmės sutampa su reikšmėmis, kurias remdamiesi logistine regresija prognozuotume. Jeigu duomenų, iš kurių gauti

parametrų įverčiai, prognozės netikslios, tada logistinės regresijos taikyti negalima, net jei kiti rodikliai ir statistinės išvados rodytų, kad ji tinkama.

Lentelėje *Classification Table* pateikiamos priklausomo kintamojo stebimos (*Observed*) ir apskaičiuoto regresijos modelio pagrindu tikėtinos (*Predicted*) reikšmės. Iš 13 kintamojo *test* reikšmių 0 (*ne*) testu buvo patvirtintos 10, o kitos 3 priskirtos prie 1 (*taip*). Iš 17 kintamojo *test* reikšmių 1 (*taip*) buvo patvirtintos 14, o kitos 3 priskirtos prie 0 (*ne*). Bendras teisingo atpažinimo procentas yra 80%. Nėra griežtų taisyklių, nusakančių, koks teisingo klasifikavimo procentas laikytinas patenkinamas. Paprastai reikalaujama, kad kiekvienos kategorijos jis būtų ne mažesnis kaip 50%, o logistinė regresija apskritai taikoma tik tuo atveju, kai $y_i = 0$ sudaro ne mažiau kaip 20% ir ne daugiau kaip 80% visų stebėjimų.

	test	patirtis	val	var	var
1	1	0	27		
2	0	0	28		
3	1	1	44		
4	1	1	25		
5	1	0	40		
6	0	1	32		
7	1	0	36		
8	1	1	38		
9	0	0	30		
10	0	1	26		
11	1	0	35		
12	1	1	38		
13	0	0	24		
14	1	1	45		
15	0	0	31		
16	1	1	39		
17	0	0	23		
18	0	0	27		
19	1	1	22		
20	1	1	30		
21	1	1	39		
22	0	1	27		
23	0	0	33		
24	1	1	39		
25	1	1	27		
26	0	0	27		
27	0	0	20		
28	1	0	35		
29	1	1	39		
30	0	0	30		
31					

14.14 pav. Duomenų rinkmena binarinei logistinei regresijai apskaičiuoti

Lentelėje *Omnibus Tests of Model Coefficients* pateiktas Chi-kvadrato (*Chi-square*) modelio suderinamumo kriterijus, kuris yra regresijos ANOVA analogas. Hipotezė H_0 atmetama (taigi bent vienas $b_j \neq 0$), jeigu p -reikšmė $p < \alpha$. Hipotezė H_0 neatmetama (taigi visi $b_j = 0$), jeigu $p \geq \alpha$, čia α —

reikšmingumo lygmuo. Šio pavyzdžio p -reikšmė (*Sig.*) $p = 0,000 < 0,05$, todėl nulinę hipotezę atmetame.

Logistinė regresija turi net keletą tiesinės regresijos determinacijos koeficientų analogų. Visiems jiems galioja ta pati taisyklė: kuo didesnis koeficientas, tuo geriau logistinė regresija suderinta su duomenimis, nors patys koeficientai yra nelyginami dėl skirtingo jų apibrėžimo. Šie koeficientai pateikiami lentelėje *Model Summary*. Tai *Cox & Snell R Square* pseudodeterminacijos koeficientas ir normuotas šio koeficiento variantas — *Nagelkerke R Square*. Koeficientai rodo tą dispersijos dalį, kurią paaiškina logistinės regresijos modelis. Šioje lentelėje taip pat pateikiama dviguba tikėtinumo funkcijos logaritmo reikšmė $-2 \text{ Log likelihood}$ (plačiau — Čekanavičius, Murauskas, 2002).

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	16,134	2	,000
	Block	16,134	2	,000
	Model	16,134	2	,000

a)

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	24,920	,416	,558

b)

Classification Table^a

Observed			Predicted		Percentage Correct
			TEST		
			0	1	
Step 1	TEST	0	10	3	76,9
		1	3	14	82,4
Overall Percentage					80.0

a. The cut value is ,500

c)

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1	PATIRTIS	2,133	1,084	3,872	1	,049	8,444
	VAL	,266	,113	5,513	1	,019	1,305
	Constant	-8,866	3,613	6,020	1	,014	,000

a. Variable(s) entered on step 1: PATIRTIS, VAL.

d)

14.15 pav. Pagrindiniai binarinės logistinės regresijos skaičiavimo rezultatai

Turėdami koeficientų reikšmes, galime apskaičiuoti dydžio z reikšmę ir norimo įvykio tikimybę p pagal 14.5 formulę. Įrašę koeficientų reikšmes į z išraišką, gauname

$$z = -8,866 + 2,133 \cdot 1 + 0,266 \cdot 30 = 1,247;$$

$$\hat{P}(\text{testas teigiamas} | 30, 1) = \exp\{1,247\} / (1 + \exp\{1,247\}) = 0,78.$$

Gautas rezultatas visada reiškia įvykio tikimybę, kad priklausomas kintamasis iš dviejų galimų kodavimo reikšmių turės didesnę reikšmę.

14.4. DAUGIAREIKŠMĖ LOGISTINĖ REGRESIJA

Šis metodas yra logistinės regresijos variantas, kai kategorinis priklausomas kintamasis nėra binarinis, o turi daugiau kaip dvi kategorijas. Nepriklausomi kintamieji turi būti kategoriniai (nominalieji arba ranginiai), tačiau numatyta galimybė į regresijos modelį įtraukti ir intervalų skalės kintamuosius.

Tikimybė, kad kategorinis priklausomas kintamasis turės i -os kategorijos reikšmę, apskaičiuojama pagal formulę

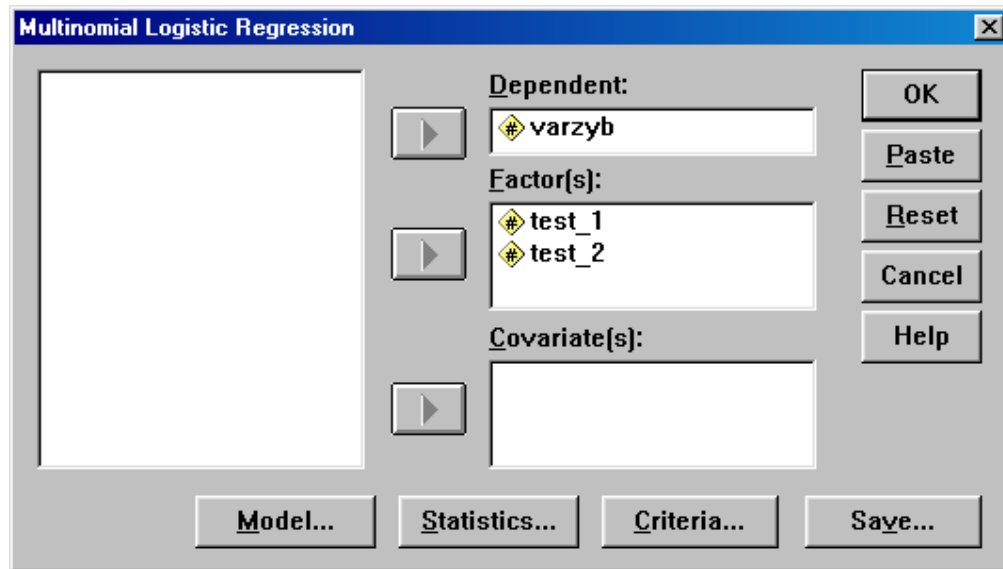
$$p_i = \frac{e^{g_i}}{\sum_{k=1}^n e^{g_k}}, \quad (14.6)$$

čia $g_i = b_{i0} + b_{i1} + \dots + b_{im}$, b_i — apskaičiuoti parametriniai i -tos priklausomo kintamojo kategorijos koeficientai (paskutinės n -tos kategorijos koeficientai prilyginami nuliui), m — nepriklausomų kintamųjų (faktorių) skaičius.

Daugiareikšmės logistinės regresijos koeficientai bei jų reikšmingumo rodikliai skaičiuojami taip:

- Atidarykite bylą su analizuojamais duomenimis.
- Nurodykite komandas **Analyze → Regression... → Multinomial Logistic...**
- Dialogo langelyje **Multinomial Logistic Regression** (14.16 pav.) priklausomą kintamąjį įkelkite į laukelį **Dependent**, o nepriklausomus kategorinius kintamuosius — į laukelį **Factor(s)**. Jeigu ketinate į regresijos modelį įtraukti intervalų skalės kintamuosius, įkelkite juos į laukelį **Covariate(s)**.
- Spragtelėję mygtuką **Statistics...** naujame dialogo langelyje **Multinomial Logistic Regression: Statistics** (14.17 pav.) galite pasirinkti papildomų charakteristikų skaičiavimą. Pažymėję **Cell probabilities** žymimąjį langelį rezultatų išvestyje gausite procentines priklausomo kintamojo kategorijų

tikimybių reikšmės visoms kategorinių nepriklausomų kintamųjų kombinacijoms ir išvengsite būtinybės skaičiuoti šias reikšmes pagal formules. Spragtelėkite dialogo langelio **Multinomial Logistic Regression: Statistics** mygtuką **Continue** ir grįžkite į pagrindinį dialogo langelį.

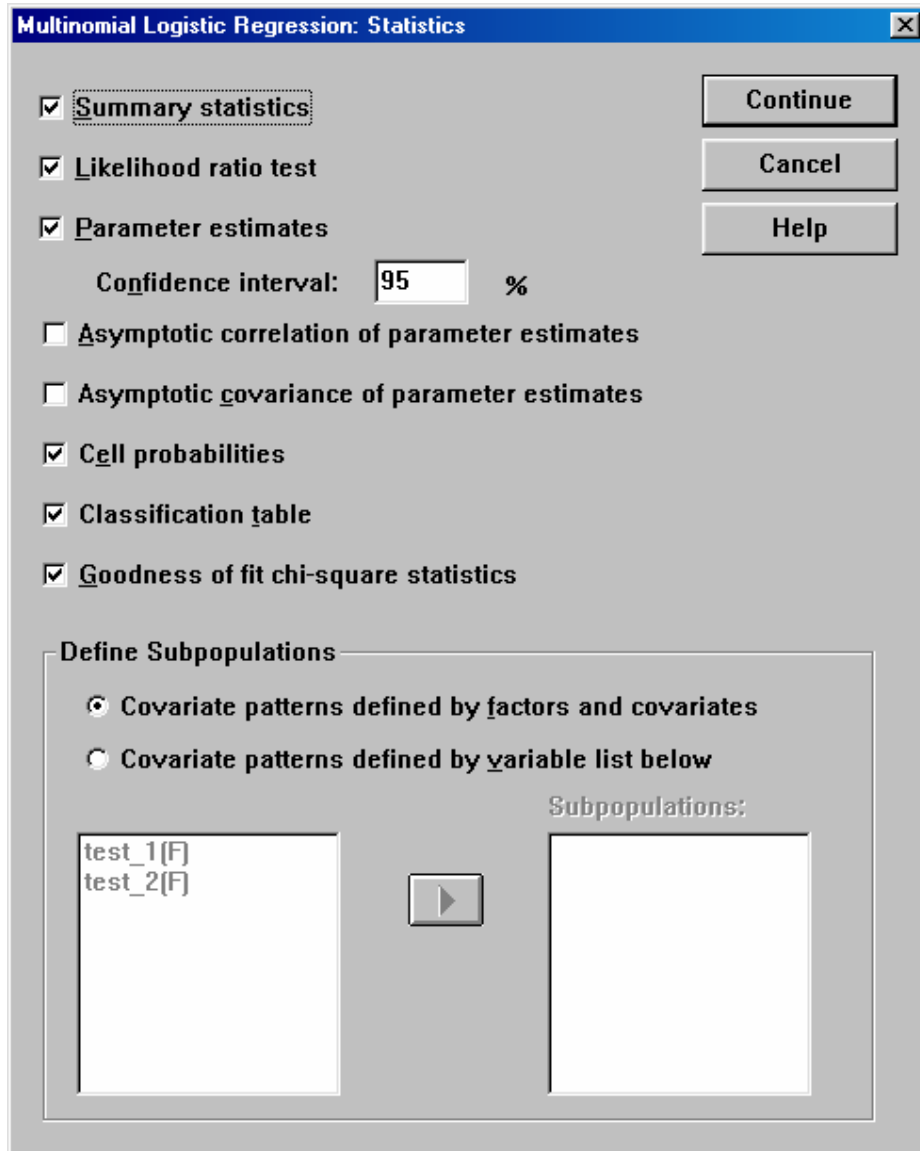


14.16 pav. Dialogo langelis **Multinomial Logistic Regression**

- Spragtelėję mygtuką **Save...**, dialogo langelyje **Multinomial Logistic Regression: Save** galite išsaugoti duomenų rinkmenoje papildomus kintamuosius: **Estimated response probabilities** (tikimybių, kad priklausomas kintamasis kiekvieno stebėjimo metu įgys konkrečią reikšmę, įverčiai), **Predicted category** (numatomos priklausomo kintamojo reikšmės, t. y. tos, kurių tikimybė kiekvieno stebėjimo metu yra didžiausia), **Predicted category probabilities** (numatomų priklausomo kintamojo reikšmių tikimybės) ir kt.
- Spragtelėkite pagrindinio dialogo langelio **Multinomial Logistic Regression** mygtuką **OK**.

Išvestyje pateikiamus daugiareikšmės logistinės regresijos rezultatus ir statistines išvadas paaiškinsime remdamiesi pavyzdžiu.

Pavyzdys (duomenys hipotetiniai). Jaunieji sportininkai prieš varžybas buvo testuojami dviem skirtingais testais: pirmo testo įverčiai — 0 (*nepatenkinamai*) ir 1 (*patenkinamai*), antro testo įverčiai — 1 (*puikiai*), 2 (*gerai*), 3 (*patenkinamai*). Varžybose pasiekti rezultatai buvo vertinami šiais įverčiais: 1 (*puikiai*), 2 (*gerai*), 3 (*žemiau galimybių*). Mus domina tikimybė, kad sportininkas, kurio pirmo testo įvertis yra 1 (*patenkinamai*), o antro testo įvertis yra 1 (*puikiai*), analogiškose varžybose pasieks *puikiai* vertinamą rezultatą.



14.17 pav. Dialogo langelis *Multinomial Logistic Regression: Statistics*

- Dialogo langelyje *Multinomial Logistic Regression* (14.16 pav.) priklausomą kintamąjį *varzyb* įkeliamo į laukelį *Dependent*, o nepriklausomus — *test_1* ir *test_2* — į laukelį *Factor(s)*.
- Dialogo langelyje *Multinomial Logistic Regression: Statistics* (14.17 pav.) papildomai pažymime šias opcijas: *Cell probabilities*, *Classification table* ir *Goodness of fit chi-square statistic*.

Pagrindiniai skaičiavimo rezultatai pateikti 14.18.1 — 14.18.3 pav.

14.18.1 pav. išvesties lentelėje a) *Case Processing Summary* pateikiama pradinių duomenų suvestinė: priklausomo kintamojo ir nepriklausomų kintamųjų stebėjimų skaičius pagal šių kintamųjų kategorijas, kiekvienos

kategorijos stebėjimų procentinė dalis tarp visų šio kintamojo stebėjimų, bendras stebėjimų skaičius.

Case Processing Summary

		N	Marginal Percentage
VARZYB	puikiai	16	38,1%
	gerai	17	40,5%
	žemiau galimybių	9	21,4%
TEST_1	nepatenkinamai	21	50,0%
	patenkinamai	21	50,0%
TEST_2	puikiai	15	35,7%
	gerai	17	40,5%
	patenkinamai	10	23,8%
Valid		42	100,0%
Missing		0	
Total		42	
Subpopulation		6 ^a	

a. The dependent variable has only one value observed in 1 (16,7%) subpopulations.

a)

Parameter Estimates

VARZYB		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp(B)	
								Lower Bound	Upper Bound
puikiai	Intercept	-17,574	1,665	111,462	1	,000			
	[TEST_1=0]	-4,673	1,724	7,348	1	,007	9,340E-03	3,183E-04	,274
	[TEST_1=1]	0 ^a			0				
	[TEST_2=1]	41,205	1,530	725,185	1	,000	7,9E+17	3,914E+16	1,576E+19
	[TEST_2=2]	21,350	,000		1		1,9E+09	1871752153	1871752153
	[TEST_2=3]	0 ^a			0				
gerai	Intercept	,879	1,181	,554	1	,457			
	[TEST_1=0]	-2,818	1,488	3,589	1	,058	5,971E-02	3,235E-03	1,102
	[TEST_1=1]	0 ^a			0				
	[TEST_2=1]	21,050	,000		1		1,4E+09	1386178998	1386178998
	[TEST_2=2]	2,779	1,306	4,524	1	,033	16,098	1,244	208,370
	[TEST_2=3]	0 ^a			0				

a. This parameter is set to zero because it is redundant.

b)

Classification

Observed	Predicted			
	puikiai	gerai	žemiau galimybių	Percent Correct
puikiai	12	4	0	75,0%
gerai	6	10	1	58,8%
žemiau galimybių	0	3	6	66,7%
Overall Percentage	42,9%	40,5%	16,7%	66,7%

c)

14.18.1 pav. Pagrindiniai daugiareikšmės logistinės regresijos skaičiavimo rezultatai

Lentelėje *b) Parameter Estimates* pateikiamos apskaičiuotų koeficientų reikšmės (*B* stulpelyje) ir jų nelygybės nuliui reikšmingumas. Wald'o kriterijus, kaip ir binarinės logistinės regresijos atveju, atsako į klausimą, ar konkretus koeficientas $b_{ij} \neq 0$. Hipotezė H_0 atmetama (taigi $b_{ij} \neq 0$), jeigu p -reikšmė $p < \alpha$. Hipotezė H_0 neatmetama, jeigu $p \geq \alpha$, čia α — reikšmingumo lygmuo. Šio pavyzdžio regresijos lygties koeficientai reikšmingai nelygūs nuliui, vienas yra ties reikšmingumo riba (p -reikšmė $p = 0,058$). Lentelės *Exp(B)* stulpelyje pateikiama galimybių santykio reikšmė.

Lentelėje *c) Classification* pateikiamos priklausomo kintamojo stebimos (*Observed*) ir apskaičiuoto regresijos modelio pagrindu tikėtinos (*Predicted*) reikšmės. Bendras teisingo atpažinimo procentas yra 66,7%, o kiekvienos kategorijos jis yra ne mažesnis kaip 50%.

14.18.2 pav. lentelėje *a) Model Fitting Information* pateiktas Chi-kvadrato (*Chi-square*) modelio suderinamumo kriterijus. Hipotezė H_0 atmetama (taigi bent vienas $b_{ij} \neq 0$), jeigu p -reikšmė $p < \alpha$. Hipotezė H_0 neatmetama (taigi visi $b_{ij} = 0$), jeigu $p \geq \alpha$, čia α — reikšmingumo lygmuo. Šiame pavyzdyje p -reikšmė (*Sig.*) $p = 0,000 < 0,05$, todėl nulinę hipotezę atmetame.

Lentelėje *b) Likelihood Ratio Test* pateikti pateiktas Chi-kvadrato (*Chi-square*) modelio suderinamumo kriterijus atskirai pagal veikiančius faktorius. Šiame pavyzdyje gautas reikšmingumo lygmuo rodo, kad abu nepriklausomi kintamieji *test_1* ir *test_2* turi reikšmingą įtaką priklausomam kintamajam.

Lentelėje *c) Pseudo R Square* pateikiamos *Cox & Snell*, *Nagelkerke* ir *McFadden* pseudodeterminacijos koeficientų reikšmės.

Įrašius į 14.5 formulę *Parameter Estimates* (parametrų įverčiai) lentelėje pateiktas koeficientų reikšmės galima apskaičiuoti norimo įvykio tikimybę, t. y. tikimybę, kad priklausomas kintamasis esant tam tikroms nepriklausomų kintamųjų reikšmėms įgys konkrečią reikšmę.

Tačiau paprasčiau šių tikimybių reikšmės procentais gauti iš lentelės *Observed and Predicted Frequencies* (14.18.3 pav.), kuri gaunama pažymėjus dialogo langelyje **Multinomial Logistic Regression: Statistics** žymimąjį langelį **Cell probabilities**. Šiame pavyzdyje mus dominanti tikimybė yra 0,846.

Model Fitting Information

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	55,796			
Final	19,331	36,465	6	,000

a)

Likelihood Ratio Tests

Effect	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	19,331 ^a	,000	0	,
TEST_1	30,252 ^b	10,921	2	,004
TEST_2	45,353	26,022	4	,000

The chi-square statistic is the difference in -2 log-likelihoods between the final model and a reduced model. The reduced model is formed by omitting an effect from the final model. The null hypothesis is that all parameters of that effect are 0.

- This reduced model is equivalent to the final model because omitting the effect does not increase the degrees of freedom.
- There is possibly a quasi-complete separation in the data. Either the maximum likelihood estimates do not exist or some parameter estimates are infinite.

b)

Pseudo R-Square

Cox and Snell	,580
Nagelkerke	,659
McFadden	,408

c)

14.18.2 pav. Pagrindiniai daugiareikšmės logistinės regresijos skaičiavimo rezultatai

Observed and Predicted Frequencies

TEST_2	TEST_1	VARZYB	Frequency			Percentage	
			Observed	Predicted	Pearson Residual	Observed	Predicted
puikiai	nepatenkinamai	puikiai	2	3,233	-,935	28,6%	46,2%
		gerai	5	3,767	,935	71,4%	53,8%
		žemiau galimybių	0	,000	,000	,0%	,0%
	patenkinamai	puikiai	8	6,767	1,207	100,0%	84,6%
		gerai	0	1,233	-1,207	,0%	15,4%
		žemiau galimybių	0	,000	,000	,0%	,0%
gerai	nepatenkinamai	puikiai	2	,767	1,492	28,6%	11,0%
		gerai	3	4,353	-1,055	42,9%	62,2%
		žemiau galimybių	2	1,880	,102	28,6%	26,9%
	patenkinamai	puikiai	4	5,233	-,781	40,0%	52,3%
		gerai	6	4,647	,858	60,0%	46,5%
		žemiau galimybių	0	,120	-,348	,0%	1,2%
patenkinamai	nepatenkinamai	puikiai	0	,000	,000	,0%	,0%
		gerai	1	,880	,137	14,3%	12,6%
		žemiau galimybių	6	6,120	-,137	85,7%	87,4%
	patenkinamai	puikiai	0	,000	,000	,0%	,0%
		gerai	2	2,120	-,152	66,7%	70,7%
		žemiau galimybių	1	,880	,152	33,3%	29,3%

The percentages are based on total observed frequencies in each subpopulation.

14.18.3 pav. Daugiareikšmės logistinės regresijos stebimų ir numatomų dažnių lentelės Observed and Predicted Frequencies pavyzdys

14.5. APROKSIMACIJA KREIVĖMIS

SPSS pakete yra pateikiama 11 įvairių kreivių tipų, kuriomis galima aproksimuoti realius duomenis (dažniausiai tai laiko eilutės) ir prognozuoti naujas priklausomo kintamojo reikšmes. Aproksimacijos modeliai išreiškiami šiomis formulėmis:

Tiesinis $\hat{y} = b_0 + b_1 \cdot x,$

Logaritminis $\hat{y} = b_0 + b_1 \cdot \ln(x),$

Atvirkštinis $\hat{y} = b_0 + \frac{b_1}{x},$

Kvadratinis $\hat{y} = b_0 + b_1 \cdot x + b_2 \cdot x^2,$

Kubinis $\hat{y} = b_0 + b_1 \cdot x + b_2 \cdot x^2 + b_3 \cdot x^3,$

Laipsnio $\hat{y} = b_0 \cdot x^{b_1},$

Rodiklinis $\hat{y} = b_0 \cdot b_1^x,$

S $\hat{y} = e^{b_0 + b_1 \cdot x},$

Logistinis
$$\hat{y} = \frac{1}{\frac{1}{u} + b_0 \cdot b_1^x},$$

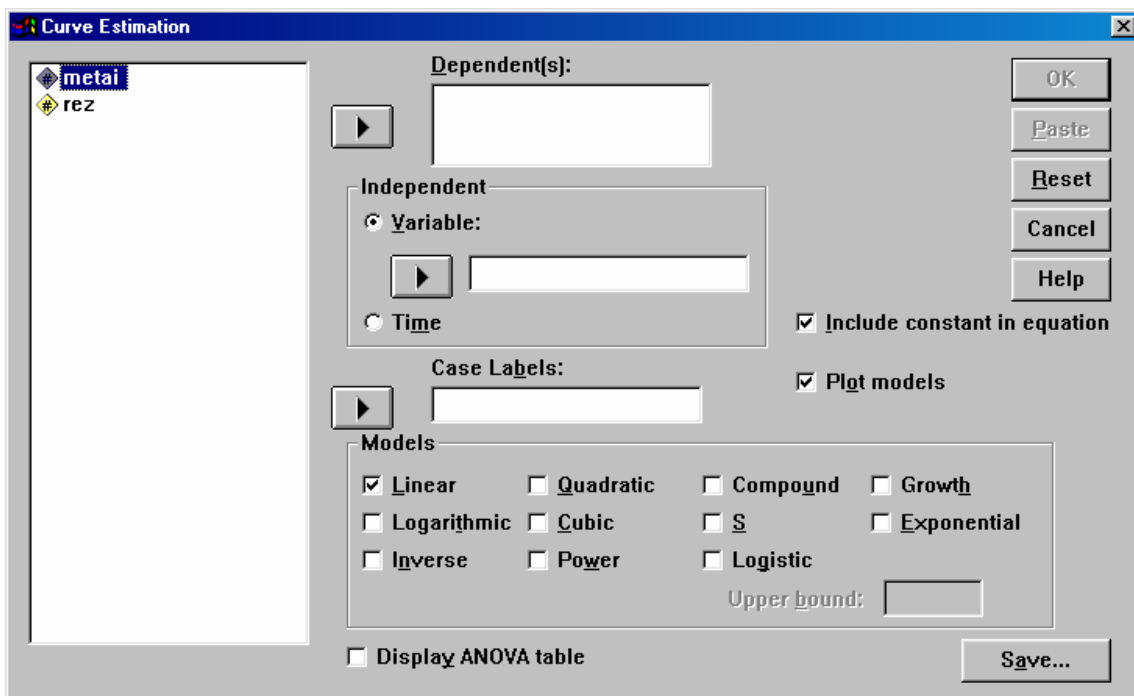
Augimo
$$\hat{y} = e^{b_0 + b_1 \cdot x},$$

Eksponentinis
$$\hat{y} = b_0 \cdot e^{b_1 \cdot x},$$

čia b_0, b_1, b_2, b_3 — koeficientai, u — parametras, nurodantis viršutinę priklausomo kintamojo ribą.

Turimų duomenų kaitai atitinkama kreivė aproksimuoti ir naujoms priklausomo kintamojo reikšmėms prognozuoti:

- Atidarykite bylą su analizuojamais duomenimis.
- Nurodykite komandas **Analyze → Regression... → Curve Estimation...**
- Dialogo langelyje **Curve Estimation** (14.19 pav.) pasirinkite vieną ar kelis priklausomus kintamuosius ir įkelkite į laukelį **Dependent(s)**. Kiekvienam priklausomam kintamajam bus pateiktas atskiras aproksimacijos rezultatas.



14.19 pav. Dialogo langelis **Curve Estimation**

- Įkelkite į **Independent Variable** laukelį nepriklausomą kintamąjį arba pažymėkite žymimąjį laukelį **Time**, jeigu analizuojate laiko eilučių duomenis su vienodais intervalais tarp stebėjimų.

- Pasirinkite tinkamą aproksimacijos modelį.

Norint nuspręsti, kuris būtent modelis geriausiai tinka konkrečioje situacijoje, pirmiausia reikia nubrėžti duomenų grafiką. Kai duomenų pasiskirstymas artimas tiesei, pasirinkite paprastą tiesinės regresijos modelį. Jeigu duomenų pasiskirstymas gerokai skiriasi nuo tiesinio, pabandykite transformuoti savo duomenis. Jeigu tai neduoda reikiamų rezultatų, panaudokite sudėtingesnę modelį, kurio funkciją labiausiai primena Jūsų duomenų pasiskirstymo grafikas.

- Spragtelėję mygtuką **Save...**, dialogo langelyje **Curve Estimation: Save** galite nurodyti išsaugoti duomenų rinkmenoje kaip naujus kintamuosius apskaičiuotas numatomas reikšmes, jų nuokrypius.
- Spragtelėkite mygtuką **OK** pagrindiniame dialogo langelyje **Curve Estimation**.

Išvesties lange **Viewer** yra pateikiami pasirinkto regresijos modelio koeficientai, jų nelygybės nuliui reikšmingumas, esamų ir prognozuojamų duomenų kitimo grafikas.

15. DISPERSINĖ ANALIZĖ

SPSS paketas turi dideles dispersinės analizės galimybes. Galima atlikti vienmatę dispersinę analizę (*Univariate*), daugiamatę dispersinę analizę (*Multivariate*), blokuotųjų duomenų dispersinę analizę (*Repeated Measures*), dispersijos komponentų skaičiavimą (*Variance Components*), kovariacinę analizę (kai nepriklausomas kintamasis yra intervalinis). Vienmatės dispersinės analizės metodu yra tiriama vieno nepriklausomo kintamojo (faktoriaus) įtaka (vieno faktoriaus dispersinė analizė) ar kelių nepriklausomų kintamųjų (faktorių) įtaka (daugelio faktorių dispersinė analizė) vienam priklausomam kintamajam, daugiamatės dispersijos analizės metodu — keliems priklausomiems kintamiesiems.

Apžvelgsime šiuos plačiausiai taikomus metodus:

- Vienmatę (vieno faktoriaus ir dviejų faktorių) dispersinę analizę.
- Daugiamatę dispersinę analizę.
- Blokuotųjų duomenų dispersinę analizę.

15.1. VIENO FAKTORIAUS DISPERSINĖ ANALIZĖ

Vieno faktoriaus dispersinė analizė — tai Stjudento *t*-kriterijaus apibendrinimas keletui nepriklausomų imčių. Vieno faktoriaus dispersinė analizė (***One-Way ANOVA***) taikoma tada, kai populiacijos viena nuo kitos skiriamos tik pagal vieną požymį. Kategorinis kintamasis (populiacijos požymis), pagal kurį populiacijos skiriamos viena nuo kitos, vadinamas nepriklausomu kintamuoju arba faktoriumi. Pavyzdžiui, tiriant sportininko parengtumą varžyboms, toks nepriklausomas kintamasis gali būti taikoma treniruočių metodika. Priklausomo kintamojo sąvoka vartojama įvardyti matuojamam parametrai, kuris priklauso nuo veikiančio faktoriaus (nepriklausomo kintamojo). Dispersinės analizės tikslas — nuspręsti, ar priklausomo kintamojo, išmatuoto skirtingose populiacijose, vidurkiai statistiškai reikšmingai skiriasi. Metodas leidžia atsakyti į klausimą, ar iš kelių nepriklausomų imčių bent dviejų vidurkiai statistiškai reikšmingai skiriasi.

Vieno faktoriaus dispersinė analizė, kuriai žymėti vartojama santrumpa ANOVA (angl. – *ANalysis Of VAriance*) grindžiama dviejų dispersijos įverčių palyginimu (Čekanavičius, Murauskas, 2002). Visa kvadratų suma *SST* (*sum of squares total*), kuri apima visų duomenų ir bendrojo vidurkio skirtumus, yra suskaidoma į dviejų komponentų — vidinės kvadratų sumos *SSW* (*sum of squares within*), įvertinančios kiekvienos imties duomenų sklaidą (apie imties vidurkį) ir grupių kvadratų sumos *SSB* (*sum of squares between*), įvertinančios imčių vidurkių ir bendrojo vidurkio skirtumus, sumą

$$SST = SSW + SSB, \quad (15.1)$$

čia

$$SST = \sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \bar{X})^2, \quad SSW = \sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2,$$

$$SSB = \sum_{i=1}^k n_i (\bar{X}_i - \bar{X})^2, \quad (15.2)$$

X_{ij} — i -osios imties j -asis stebėjimas, \bar{X} — bendrasis visų imčių vidurkis,

\bar{X}_i — i -osios imties vidurkis, k — nepriklausomų populiacijų skaičius,

n_i — i -osios imties didumas;

ANOVA hipotezei tikrinti yra sudaroma statistika tarpgruvinės ir grupių vidinės dispersijos pagrindu.

Tikrinama statistinė hipotezė:

$$\begin{cases} H_0: \mu_1 = \mu_2 = \dots = \mu_k, \\ H_1: \text{ bent du vidurkiai skiriasi,} \end{cases}$$

čia μ_j — nepriklausomų kintamųjų vidurkiai.

Sprendimo priėmimo taisyklė:

H_0 atmetama (ne visi vidurkiai lygūs), jeigu $p < \alpha$;

H_0 neatmetama (vidurkių skirtumo nerasta), jeigu $p \geq \alpha$;

čia α — nustatytasis reikšmingumo lygmuo.

Nustačius, kad yra statistiškai reikšmingai besiskiriančių imčių vidurkių, galima nustatyti, kurių imčių vidurkiai statistiškai reikšmingai skiriasi. Tam yra naudojami vadinamieji daugkartinių lyginimų kriterijai.

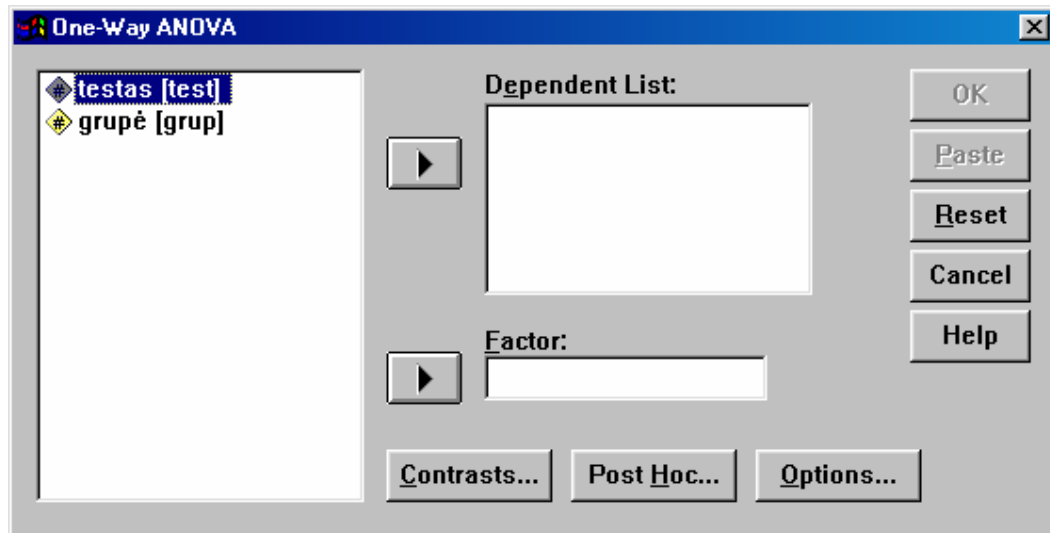
Kad būtų galima patikimai patvirtinti arba atmesti nulinę hipotezę apie priklausomo kintamojo, stebimo skirtingose populiacijose, vidurkių lygybę taikant vieno faktoriaus dispersinės analizės metodą, turi būti įvykdytos šios sąlygos:

- Kintamieji $X_1 \dots X_k$ populiacijose turi būti pasiskirstę pagal normalųjį dėsnį. Tačiau, jeigu imtys yra pakankamai didelės arba jų didumai mažai skiriasi, nedideli skirstinių nuokrypiai nuo normaliojo nėra kritiški.
- Kintamieji $X_1 \dots X_k$ turi būti nepriklausomi.
- Faktorinis (nepriklausomas) kintamasis turi būti sveikas skaičius, o priklausomas kintamasis — intervalinis.
- Kintamųjų $X_1 \dots X_k$ dispersijos turi būti lygios — turi būti tenkinama dispersijų homogeniškumo sąlyga. Kuo mažiau skiriasi imčių didumas, tuo mažiau ANOVA jautri populiacijų dispersijų skirtumams. Paprastai didžiausioji iš imčių dispersijų neturi būti didesnė už mažiausiąją daugiau kaip tris kartus.

Vieno faktoriaus dispersinės analizės esmę paaiškinsime pasitelkę paprastą pavyzdį. Tarkime, kad treneris penkioms vienodo amžiaus ir pajėgumo grupėms (po 10 sportininkų kiekvienoje) taiko skirtingą treniruotės metodiką norėdamas rasti efektyviausią. Bendrumo dėlei sportininkų testavimo rezultatus vertinsime dešimties balų sistemoje. Kintamąjį, kurio reikšmės yra sportininkų testavimo rezultatai, pavadinsime *test* (žymena — *testas*), o faktorinį kintamąjį, kuris parodo kuriai grupei priklauso konkretus testavimo rezultatas, pavadinsime *group* (žymena — *grupė*). Faktorinio kintamojo reikšmės yra 1, 2, 3, 4 ir 5. Patogumo dėlei faktorinio kintamojo *group* reikšmes duomenų redaktoriaus stulpelyje suskirstome po dešimt didėjančia tvarka, o šalia, kitame stulpelyje, įrašome atitinkamus testavimo rezultatus — kintamojo *test* reikšmes.

- Nurodome komandas **Analyze → Compare Means → One-Way ANOVA...**
- Dialogo langelyje **One-Way ANOVA** (15.1 pav.) kintamąjį *test* perkeliame į priklausomų kintamųjų sąrašą **Dependent List**, o kintamąjį *group* — į laukelį **Factor**.

Panagrinėkime, kokius parametrus ir papildomus testus galima nurodyti dialogo langeliuose, kurie pasirodo spragtelėjus mygtukus **Contrasts...**, **Post Hoc...**, **Options...**



15.1 pav. Vieno faktoriaus dispersinės analizės dialogo langas **One-Way ANOVA**

15.1.1. Aposterioriniai (*Post Hoc*) kriterijai

ANOVA atsako į klausimą, ar yra statistiškai reikšmingai besiskiriančių imčių vidurkių, bet nenurodo, kurių būtent imčių vidurkiai statistiškai reikšmingai skiriasi. Kriterijai, leidžiantys nustatyti, kurių imčių vidurkiai statistiškai reikšmingai skiriasi, yra vadinami daugkartinių lyginimų kriterijais.

Šie kriterijai yra skirstomi į planuotus iš anksto (apriorinius) kriterijus ir *Post Hoc* (aposteriorinius) kriterijus. Jeigu sprendžiame prieš vieno faktoriaus analizę, turime apriorinius, iš anksto planuotus lyginimus, jeigu paskui, kai jau žinomi analizės rezultatai, turime aposteriorinius (*Post Hoc*) lyginimus. Tie patys kriterijai, tik su skirtingais reikšmingumo lygmenimis, gali būti taikomi ir kaip aprioriniai, ir kaip *Post Hoc* lyginimai.

Yra daug skirtingų *Post Hoc* kriterijų. Dialogo langelyje **One-Way ANOVA: Post Hoc Multiple Comparisons** (15.2 pav.), kuris pasirodo spragtelėjus pagrindinio dialogo langelio **One-Way ANOVA** mygtuką **Post Hoc...** pateikiama 18 tokių aposteriorinių daugkartinio lyginimo kriterijų.

Kriterijai, taikomi kai kintamųjų dispersijos yra lygios (*Equal Variances Assumed*):

LSD, Bonferroni, Sidak'o, Scheffe, R-E-G-W F (Ryan'o-Einot'o-Gabriel'o-Welsch'o F testas), R-E-G-W Q (Ryan'o-Einot'o-Gabriel'o-Welsch'o Q testas), S-N-K (Student'o-Newman'o-Keuls'o), Tukey'o, Tukey'o b, Duncan'o, Hochberg'o GT2, Gabriel'o, Waller'o-Duncan'o, Dunnett'o.

Kriterijai, taikomi kai kintamųjų dispersijos nėra lygios (*Equal Variances Not Assumed*):

Tamhane T2, Dunnett'o T3, Games-Howell'o, Dunnett'o C.

One-Way ANOVA: Post Hoc Multiple Comparisons

Equal Variances Assumed

☐ LSD ☐ S-N-K ☐ Waller-Duncan

☐ Bonferroni ☐ Tukey ☐ Type I/Type II Error Ratio: 100

☐ Sidak ☐ Tukey's-b ☐ Dunnett

☐ Scheffe ☐ Duncan ☐ Control Category: Last

☐ R-E-G-W F ☐ Hochberg's GT2 ☐ Test

☐ R-E-G-W Q ☐ Gabriel ☐ 2-sided ☐ < Control ☐ > Control

Equal Variances Not Assumed

☐ Tamhane's T2 ☐ Dunnett's T3 ☐ Games-Howell ☐ Dunnett's C

Significance level: .05

Continue Cancel Help

15.2 pav. Dialogo langelis **One-Way ANOVA: Post Hoc Multiple Comparisons** aposteriorinams testams pasirinkti

Paprastai tyrimuose apsiribojama keliais pagrindiniais *Post Hoc* kriterijais, kuriuos čia trumpai apibūdinsime (plačiau — Čekanavičius, Murauskas, 2002).

LSD kriterijus

Pagal Fišerio mažiausiai reikšmingo skirtumo LSD (*Least Significant Difference*) kriterijų visos imčių poros lyginamos naudojant *t*-kriterijų (tada, kai ANOVA parodo, kad vidurkiai statistiškai reikšmingai skiriasi).

LSD yra liberaliausias *Post Hoc* kriterijus, t. y. jis dažniausiai randa statistiškai reikšmingus vidurkių skirtumus.

Bonferroni kriterijus

Pagal Bonferroni kriterijų visos imčių poros lyginamos taikant Stjudento kriterijų, esant reikšmingumo lygmeniui $\alpha = \alpha_E / C$; čia α_E — eksperimento reikšmingumo lygmuo (t. y. tikimybė lyginant visas įmanomas poras nors kartą neteisingai nustatyti statistiškai reikšmingą dviejų imčių vidurkių skirtumą), α — reikšmingumo lygmuo lyginant vieną imčių porą, $C = k(k-1)/2$, k — imčių skaičius. Bonferroni kriterijus netaikytinas, kai imčių yra daug, nes labai sumažėja α ir nefiksuojamas statistiškai reikšmingas vidurkių skirtumas, nors tikrieji populiacijų vidurkiai ir skiriasi (t. y. labai išauga antrosios rūšies klaidos tikimybė).

Tukey'o kriterijus

Tukey'o HSD (*Honestly Significant Difference* — ganėtinai statistiškai reikšmingo skirtumo) kriterijus — vienas dažniausiai taikomų *Post Hoc* kriterijų. Jis yra gera alternatyva Bonferroni kriterijui, kai imčių daug. Tukey'o kriterijus yra labai konservatyvus, t. y. labiausiai nelinkęs atmesti nulinės hipotezės — vidurkių skirtumus pripažinti statistiškai reikšmingais.

Tukey'o kriterijaus rezultatų išvestyje išskiriamos taip pat homogeniškų imčių grupės.

S-N-K (Student'o-Newman'o-Keuls'o) kriterijus

S-N-K kriterijus taikomas kaip homogeniškų imčių grupių išskyrimo priemonė. S-N-K kriterijus yra kiek konservatyvesnis už Bonferroni kriterijų, bet ne toks konservatyvus kaip Tukey'o kriterijus. Taikant S-N-K, kaip ir LSD kriterijų, statistiškai reikšmingi imčių vidurkių skirtumai dažnai randami net tada, kai tikrieji populiacijų vidurkiai nesiskiria. Todėl S-N-K kriterijus taikomas ribotai.

Daugelis kitų *Post Hoc* kriterijų grindžiami ką tik aprašytaisiais.

15.1.2. Aprioriniai kriterijai

Tradiciškai aprioriniams priskiriami kriterijai, kurie taikomi hipotezėms apie tiesines populiacijų vidurkių kombinacijas (kontrastus) tikrinti. Taigi, yra tikrinamos ne visos imtys poromis, o pasirinktos kombinacijos.

Tiesiniu populiacijų vidurkių kontrastu vadinama suma

$$C_1\mu_1 + C_2\mu_2 + \dots + C_k\mu_k, \quad (15.3)$$

čia koeficientai C_i tenkina lygybę

$$C_1 + C_2 + \dots + C_k = 0. \quad (15.4)$$

Kontrasto koeficientus, t. y. konkrečius skaičius pasirenkame patys. Koeficientai gali būti teigiami, neigiami, sveikieji ar trupmeniniai skaičiai. Imdami tik du nenulinius koeficientus 1 ir -1 , gauname hipotezę apie dviejų vidurkių lygybę. Norėdami atskirti vieną nuo kitos kelias populiacijų grupes, pvz., palyginti pirmųjų dviejų populiacijų kombinaciją su penktąja, galime pasirinkti tokius koeficientus: $-1 \quad -1 \quad 0 \quad 0 \quad 2$. Populiacijos, kurios atitinka nulinius koeficientus, iš analizės eliminuojamos.

Tarkime, kad turime k imčių $(x_{11}, x_{12}, \dots, x_{1n}), (x_{21}, x_{22}, \dots, x_{2n}), \dots, (x_{k1}, x_{k2}, \dots, x_{kn})$, kurios gautos matuojant nepriklausomus normaliuosius atsitiktinius dydžius $X_1 \sim N(\mu_1, \sigma^2), X_2 \sim N(\mu_2, \sigma^2), \dots, X_k \sim N(\mu_k, \sigma^2)$ pagal intervalų skalę.

Statistinei hipotezei

$$\begin{cases} H_0: C_1\mu_1 + C_2\mu_2 + \dots + C_k\mu_k = 0, \\ H_1: C_1\mu_1 + C_2\mu_2 + \dots + C_k\mu_k \neq 0 \end{cases}$$

(čia C_i — yra iš anksto pasirinkti skaičiai, tenkinantys (15.4) lygybę) patikrinti:

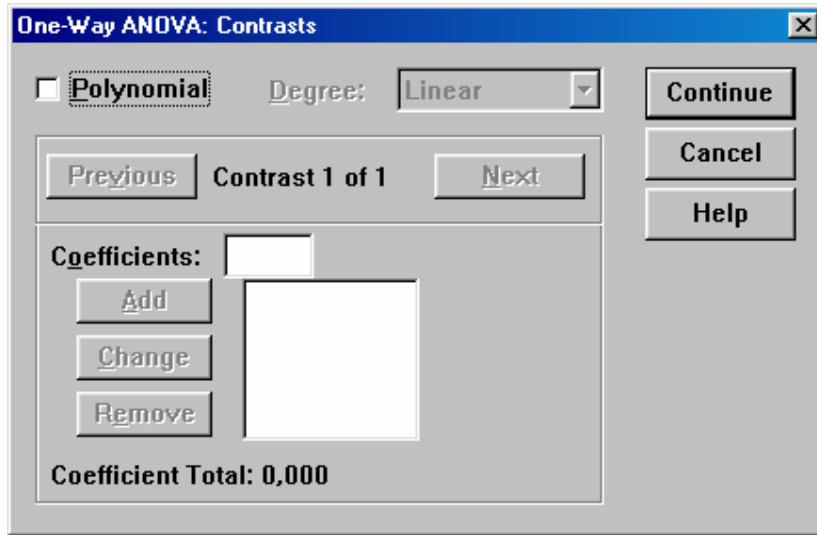
- Spragtelėkite dialogo langelio **One-Way ANOVA** mygtuką **Contrasts...**
- Dialogo langelio **One-Way ANOVA: Contrasts** (15.3 pav.) laukelyje **Coefficients** įrašykite pirmojo koeficiento reikšmę ir spragtelėkite mygtuką **Add**. Tokiu pat būdu įrašykite likusius koeficientus. Įrašę visus vieno kontrasto koeficientus, spragtelėję mygtuką **Next** galite pradėti įrašinėti kito kontrasto koeficientus.
- Nurodę koeficientų reikšmes visiems kontrastams, spragtelėkite mygtuką **Continue**.

Sprendimo priėmimo taisyklė:

H_0 atmetama (kontrastas statistiškai reikšmingai skiriasi nuo nulio), jeigu $p < \alpha$;

H_0 neatmetama, jeigu $p \geq \alpha$;

čia α — nustatytasis reikšmingumo lygmuo.



15.3 pav. Dialogo langelis **One-Way ANOVA: Contrasts**

15.1.3. Vidurkių trendas

Kai nepriklausomo faktoriaus reikšmės rodo skirtingą treniruotumo laipsnį, streso lygį ir pan., t. y. nepriklausomas kintamasis (faktorių), pagal kurį yra skiriamos populiacijos, yra ne šiaip kategorinis, bet gautas iš intervalinio kintamojo, gali dominti populiacijose stebimo kintamojo vidurkių kitimo tendencija (vidurkių trendas). Iškart galima tikrinti kelias hipotezes: ir apie tiesinį vidurkių trendą, ir apie kvadratinį, ir t. t. iki penktojo laipsnio imtinai. Tikrai imčių turi būti bent viena daugiau nei tendą atitinkančio daugianario didžiausias laipsnis.

Pavyzdžiui, statistinė hipotezė apie tiesinį vidurkių trendą formuluojama taip:

$$\begin{cases} H_0: \text{vidurkliai } \mu_1, \mu_2, \dots, \mu_k \text{ nesudaro tiesinio trendo,} \\ H_1: \text{vidurkliai } \mu_1, \mu_2, \dots, \mu_k \text{ sudaro tiesinį tendą.} \end{cases}$$

Sprendimo priėmimo taisyklė:

H_0 atmetama (trendas tinka), jeigu $p < \alpha$;

H_0 neatmetama, jeigu $p \geq \alpha$;

čia α — nustatytasis reikšmingumo lygmuo.

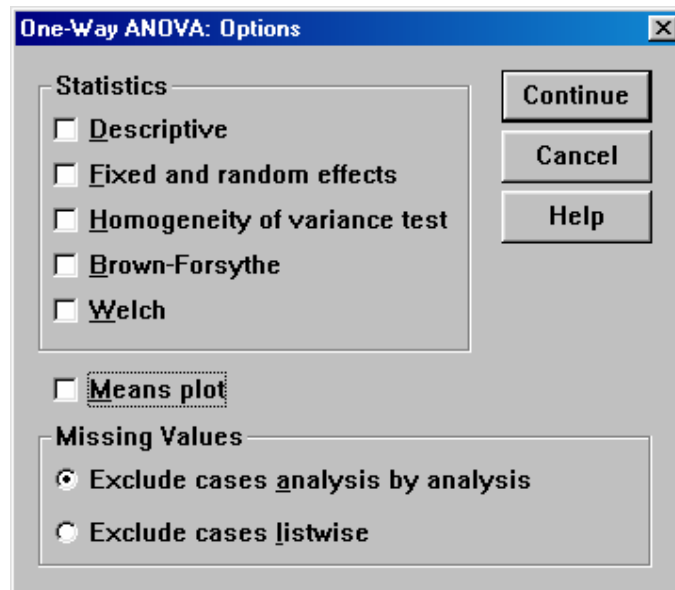
Trendų analizei atlikti, t. y. patikrinti hipotezei apie tiesinį ar polinominį populiacijų vidurkių tendą:

- Dialogo langelio **One-Way ANOVA: Contrasts** (15.3 pav.) pažymėkite laukelį **Polynomial**, o išskleidžiamajame sąrašė **Degree** pasirinkite polinomo laipsnį.

15.1.4. Pasirinkimai (*Options*)

Dialogo langelyje **One-Way ANOVA: Options** (15.4 pav.), kuris pasirodo spragtelėjus pagrindinio dialogo langelio **One-Way ANOVA** mygtuką **Options...** galite pasirinkti:

- **Descriptive** (skaitines kiekvienos imties charakteristikas): stebėjimų skaičių, vidurkį, standartinį nuokrypį, standartinę paklaidą, didžiausią ir mažiausią reikšmę, 95% pasikliautinąjį intervalą).
- **Homogeneity of variance test** (dispersijų homogeniškumo testą).
- **Means plot** (linijinį vidurkių grafiką), kitus parametrus.



15.4 pav. Dialogo langelis **One-Way ANOVA: Options**

Pateiktame pavyzdyje pasirinksiame šiuos ANOVA variantus:

- Dialogo langelyje **One-Way ANOVA: Contrasts** pažymime laukelį **Polynomial**, o išskleidžiamajame sąrašė **Degree** nurodome trendo kubinį polinomo laipsnį (**Cubic**). Laukelyje **Coefficients** įrašome šiuos kontrasto koeficientus: 2 0 0 –1 –1, t. y. patikrinsime hipotezę apie dvigubo pirmos grupės vidurkio lygybę ketvirtos ir penktos grupių vidurkių sumai.
- Dialogo langelyje **One-Way ANOVA: Post Hoc Multiple Comparisons** pasirenkame Tukey'o *Post Hoc* kriterijų.
- Dialogo langelyje **One-Way ANOVA: Options** pažymime **Descriptive** ir **Homogeneity of variance test** laukelius.

Vieno faktoriaus dispersinės analizės rezultatai pateikti 15.5.1 — 15.5.2 pav.

15.5.1 pav. pimoje a) lentelėje — *Descriptives* (aprašomoji statistika) pateiktos šios kiekvienos grupės skaitinės charakteristikos: stebėjimų skaičius

(*N*), vidurkis (*Mean*), standartinis nuokrypis (*Std. Deviation*), standartinė paklaida (*Std. Error Mean*), vidurkio 95% pasikliautinasis intervalas (95% *Confidence Interval for Mean*), mažiausia reikšmė (*minimum*), didžiausia reikšmė (*maximum*).

Antroje *b*) lentelėje — *Test Homogeneity of Variances* (dispersijų homogeniškumo testas) pateiktas Levene testo rezultatas. Kadangi *p*-reikšmė (*Sig.*) gerokai viršija nustatytą reikšmingumo lygmenį (0,05), priimame hipotezę apie dispersijų lygybę.

Trečios *c*) — *ANOVA* lentelės pirmoje eilutėje *Between Groups (Combined)* pateikta *p*-reikšmė (*Sig.*) lygi nuliui trijų skaičių po kablelio tikslumu, t. y. mažesnė už nustatytą reikšmingumo lygmenį (0,05), todėl hipotezę apie vidurkių lygybę atmetame. Taikomos treniruočių metodikos statistiškai skiriasi reikšmingai.

Iškart po eilutės, skirtos grupių skirtumams, išdėstyta informacija apie trendus. Visi trendai tinka, nes atitinkamų kontrastų *p*-reikšmės (*Sig.*) mažesnės už pasirinktą reikšmingumo lygmenį. Jeigu vidurkiams tinka keli trendai, dažniausiai pasirenkamas mažesnės eilės trendas. Šiuo atveju tai būtų tiesinis trendas.

Ketvirtoje *d*) lentelėje — *Contrast Coefficients* (kontrastų koeficientai) pateikti mūsų pasirinkti kontrasto koeficientai, o penktoje *e*) lentelėje — *Contrast Tests* (kontrastų testas) *p*-reikšmė (*Sig.(2-tailed)*), mažesnė už pasirinktą reikšmingumo lygmenį, leidžia atmesti nulinę hipotezę apie kontrasto reikšmę, t. y. dviguba pirmos grupės vidurkio reikšmė statistiškai reikšmingai skiriasi nuo ketvirtos ir penktos grupių vidurkių sumos.

15.5.2 pav. *a*) lentelėje *Post Hoc Test Multiple Comparisons* (daugkartiniai lyginimai) pateikti Tukey'o HSD kriterijaus rezultatai. Statistiškai reikšmingi vidurkių skirtumai, t. y. tie, kurių *p*-reikšmė (*Sig.*) mažesnė už nustatytą reikšmingumo lygmenį, pažymėti žvaigždutėmis.

15.5.2 pav. *b*) lentelėje *Homogenous Subsets* (homogeniškos grupės) remiantis Tukey'o HSD testu išskirti du homogeniški pogrupiai (pirma, antra ir trečia bei pirmą, ketvirtą ir penktą treniruočių metodikas), tarp kurių skirtumas yra statistiškai reikšmingas (reikšmingumo lygmeniu 0,05), tačiau viduje kurių vidurkiai skiriasi nereikšmingai — *p*-reikšmė (*Sig.*) atitinkamai lygi 0,333 ir 0,089.

Descriptives

testas								
	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1	10	6,60	1,430	,452	5,58	7,62	4	9
2	10	5,70	1,494	,473	4,63	6,77	3	8
3	10	5,50	,972	,307	4,80	6,20	4	7
4	10	7,80	1,317	,416	6,86	8,74	6	10
5	10	8,10	1,197	,379	7,24	8,96	6	10
Total	50	6,74	1,639	,232	6,27	7,21	3	10

a)

Test of Homogeneity of Variances

testas			
Levene Statistic	df1	df2	Sig.
,364	4	45	,833

b)

ANOVA

testas			Sum of Squares	df	Mean Square	F	Sig.
Between Groups	(Combined)		56,120	4	14,030	8,362	,000
	Linear Term	Contrast	26,010	1	26,010	15,503	,000
		Deviation	30,110	3	10,037	5,982	,002
	Quadratic Term	Contrast	17,150	1	17,150	10,222	,003
		Deviation	12,960	2	6,480	3,862	,028
	Cubic Term	Contrast	7,290	1	7,290	4,345	,043
		Deviation	5,670	1	5,670	3,379	,073
	Within Groups		75,500	45	1,678		
Total		131.620	49				

c)

Contrast Coefficients

Contrast	grupė				
	1	2	3	4	5
1	2	0	0	-1	-1

d)

Contrast Tests

			Value of Contrast	Std. Error	t	df	Sig. (2-tailed)
testas	Assume equal variances	1	-2,70	1,003	-2,691	45	,010
	Does not assume equal	1	-2,70	1,065	-2,535	16,102	,022

e)

15.5.1 pav. Vieno faktoriaus dispersinės analizės rezultatai

Post Hoc Tests

Multiple Comparisons

Dependent Variable: testas

Tukey HSD

(I) grupē	(J) grupē	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	,90	,579	,534	-,75	2,55
	3	1,10	,579	,333	-,55	2,75
	4	-1,20	,579	,250	-2,85	,45
	5	-1,50	,579	,089	-3,15	,15
2	1	-,90	,579	,534	-2,55	,75
	3	,20	,579	,997	-1,45	1,85
	4	-2,10*	,579	,006	-3,75	-,45
	5	-2,40*	,579	,001	-4,05	-,75
3	1	-1,10	,579	,333	-2,75	,55
	2	-,20	,579	,997	-1,85	1,45
	4	-2,30*	,579	,002	-3,95	-,65
	5	-2,60*	,579	,000	-4,25	-,95
4	1	1,20	,579	,250	-,45	2,85
	2	2,10*	,579	,006	,45	3,75
	3	2,30*	,579	,002	,65	3,95
	5	-,30	,579	,985	-1,95	1,35
5	1	1,50	,579	,089	-,15	3,15
	2	2,40*	,579	,001	,75	4,05
	3	2,60*	,579	,000	,95	4,25
	4	,30	,579	,985	-1,35	1,95

*. The mean difference is significant at the .05 level.

a)

Homogeneous Subsets

testas

Tukey HSD^a

grupē	N	Subset for alpha = .05	
		1	2
3	10	5,50	
2	10	5,70	
1	10	6,60	6,60
4	10		7,80
5	10		8,10
Sig.		,333	,089

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 10,000.

b)

15.5.2 pav. Vieno faktoriaus dispersinēs analizēs rezultāti

15.2. BLOKUOTŲJŲ DUOMENŲ VIENO FAKTORIAUS DISPERSINĖ ANALIZĖ

Biologijoje, medicinoje, farmacijoje bei kitose mokslo srityse tų pačių objektų tiriamą požymį tenka matuoti keletą kartų. Taip atsiranda vadinamieji duomenų blokai. Blokuotųjų duomenų dispersinė analizė yra Stjudento t -kriterijaus priklausomoms imtims apibendrinimas didesniai imčių skaičiui. Porinio Stjudento kriterijaus atveju duomenis sudaro stebėjimų poros $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, o bendru atveju laikysime, kad duomenis sudaro b -mačiai stebėjimų vektoriai — duomenų blokai: $(x_{11}, x_{12}, \dots, x_{1b}), (x_{21}, x_{22}, \dots, x_{2b}), \dots, (x_{n1}, x_{n2}, \dots, x_{nb})$. Duomenų bloką sudaro visi vieno tiriamojo požymio matavimai. Blokų gaunama tiek, kiek ir tiriamųjų. Pavyzdžiui, tiriant treniruočių proceso efektyvumą arba tikrinant naudojamų testų patikimumą, tos pačios sportininkų grupės tyrimo rezultatai registruojami tam tikrais laiko intervalais. Šioje situacijoje duomenų bloką sudarys visi konkretaus individo parametrų matavimai, o blokų skaičius bus lygus tiriamų sportininkų skaičiui. Blokuotųjų duomenų dispersinė analizė tinka ne vien pakartotiniams matavimams. Svarbu tik, kad bloką sudarančius duomenis vienytų bendras faktorius. Pavyzdžiui, k ekspertų vertina n metodikų. Blokuotųjų duomenų imtys surašomos duomenų rinkmenos stulpeliuose. Eilutėse esantys blokai nepriklauso vienas nuo kito.

Blokuotųjų duomenų dispersinės analizės prielaidos daug kuo panašios į vieno faktoriaus dispersinės analizės prielaidas, aprašytas 15.1 skyriuje. Tarkime, tą patį intervalinį kintamąjį matuojame skirtingomis sąlygomis. Kintamąjį, pamatuotą pirmomis sąlygomis, pažymime X_1 , antrosiomis — X_2 ir t. t. iki X_b . Kiekvienas iš b kintamųjų buvo pamatuotas n kartų. Blokuotųjų duomenų modelio prielaidos:

- Kintamieji X_j pasiskirstę pagal normalųjį dėsnį.
- Kintamųjų X_j dispersijos lygios.
- Kintamųjų skirtumų $X_j - X_i$ ir $X_k - X_l$ dispersijos lygios ($i \neq j, k \neq l$).

Antroji ir trečioji sąlygos dar vadinamos sferiškumo prielaidomis. Šias prielaidas dar galima soformuluoti taip: antroji prielaida reikalauja, kad visų įmanomų X_j ir X_k kovariacijos matricos visi pagrindinės įstrižainės elementai būtų lygūs (šią įstrižainę sudaro X_j dispersijos), o trečioji prielaida reikalauja, kad visi likusieji elementai (visos X_j ir X_k , $j \neq k$, koreliacijos) būtų lygūs, bet nebūtinai sutaptų su įstrižainės elementais. Yra keletas kriterijų, kuriais tikrinamas kovariacijų matricos sferiškumas. Jeigu trečioji prielaida netenkinama, naudojamos specialios pataisos.

Kaip ir nepriklausomų imčių atveju, kriterijaus imčių vidurkiams palyginti statistika yra grindžiama dispersijų įverčių santykiu (Čekanavičius, Murauskas, 2002). Visa kvadratų suma SST (*sum of squares total*), kuri apima visų duomenų ir bendrojo vidurkio skirtumus, yra suskaidoma į šių komponentų — blokų kvadratų sumos $SSBL$, įvertinančios blokų (tiriamųjų)

įtaką, eksperimento sąlygų kvadratų sumos $SSTR$ ir paklaidos kvadratų sumos SSE , sumą

$$SST = SSBL + SSTR + SSE, \quad (15.5)$$

čia

$$SST = \sum_{i=1}^n \sum_{j=1}^b (X_{ij} - \bar{X})^2, \quad SSBL = b \sum_{i=1}^n (\bar{X}_{i.} - \bar{X})^2, \\ SSTR = n \sum_{j=1}^b (\bar{X}_{.j} - \bar{X})^2, \quad SSE = SST - SSTR - SSBL, \quad (15.6)$$

X_{ij} — j -osios imties i -asis elementas, \bar{X} — bendrasis visų imčių vidurkis, $\bar{X}_{i.}$ — i -osios eilutės (bloko) vidurkis, $\bar{X}_{.j}$ — j -tojo stulpelio (j -osios imties) vidurkis, n — eilučių (blokų) skaičius, b — stulpelių (imčių) skaičius.

Blokuotųjų duomenų dispersinė analizė yra grindžiama dviejų dispersijos įverčių (sunormavus SSE ir $SSTR$) palyginimu.

Tarkime, kad turime b imčių, gautų matuojant pagal intervalų skalę priklausomus normaliuosius, tenkinančius sferiškumo sąlygą, kintamuosius $X_1 \sim N(\mu_1, \sigma^2)$, $X_2 \sim N(\mu_2, \sigma^2)$, ..., $X_b \sim N(\mu_b, \sigma^2)$, kurių vidurkliai μ_1, \dots, μ_b , dispersijos σ ir kovariacijos $cov(X_i, X_j)$ yra nežinomos.

Tikrinama statistinė hipotezė:

$$\begin{cases} H_0: \mu_1 = \mu_2 = \dots = \mu_b, \\ H_1: \text{ bent du vidurkliai skiriasi;} \end{cases}$$

Sprendimo priėmimo taisyklė:

H_0 atmetama, jeigu $p < \alpha$;

H_0 neatmetama, jeigu $p \geq \alpha$;

čia α — nustatytasis reikšmingumo lygmuo.

Kaip jau minėjome anksčiau, SPSS paketu tikrinama sferiškumo hipotezė. Jeigu ši hipotezė atmetama (atitinkama p -reikšmė gaunama mažesnė už reikšmingumo lygmenį), tada atsakymo apie vidurkių lygybę reikia ieškoti žiūrint į pataisytą kriterijų atitinkančią p -reikšmę.

SPSS pakete kiekvieno bandymo rezultatai atitinka tiriamą požymį veikiančio faktoriaus (**Within-Subject Factor**) atitinkamą reikšmę (kategoriją) ir duomenų rinkmenoje sudaro atskirą priklausomą (šiuo atveju, nuo veikiančio faktoriaus) kintamąjį. Veikiančio faktoriaus reikšmių (kategorijų) skaičius lygus atliktų bandymų skaičiui.

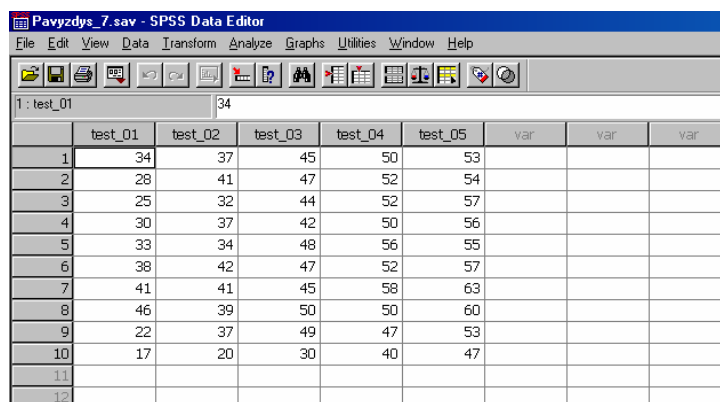
Pavyzdys. Panagrinėsime jaunųjų krepšininkų komandos (10 sportininkų) baudų metimo rezultatų kitimo priklausomybę nuo treniruočių ciklo

trukmės (duomenys hipotetiniai). Per tam tikrą laikotarpį, vienodais laiko tarpais, krepšininkai buvo testuoti penkis kartus. Kiekvieno testavimo metu buvo atliekamos dvi serijos po 50 metimų, abiejų serijų rezultatai sumuojami. Duomenų rinkmenoje gauti rezultatai sudaro penkių kintamųjų *test_01*, ..., *test_05* reikšmės (15.6 pav.).

- Nurodome komandas *Analyze* → *General Linear Model* → *Repeated Measures...*
- Dialogo langelio *Repeated Measures Define Factor(s)* (15.7 pav.) laukelyje *Number of Levels* nurodome veikiančio faktoriaus (*Within-Subject Factor*) lygmenų skaičių — 5. Nustatytasis faktoriaus pavadinimas yra *Factor 1*. Galima suteikti kitą vardą, įrašant jį laukelyje *Within-Subject Factor Name*.
- Spragtelime mygtuką *Add*. Kadangi šiame pavyzdyje turime tik vieną veikiančią faktorių spragtelime mygtuką *Define* (nustatyti).
- Naujame dialogo langelyje *Repeated Measures* (kartotiniai matavimai, 15.8 pav.) įkeliamo kintamuosius *test_01*, ..., *test_05* į sąrašą *Within-Subject Variables*.

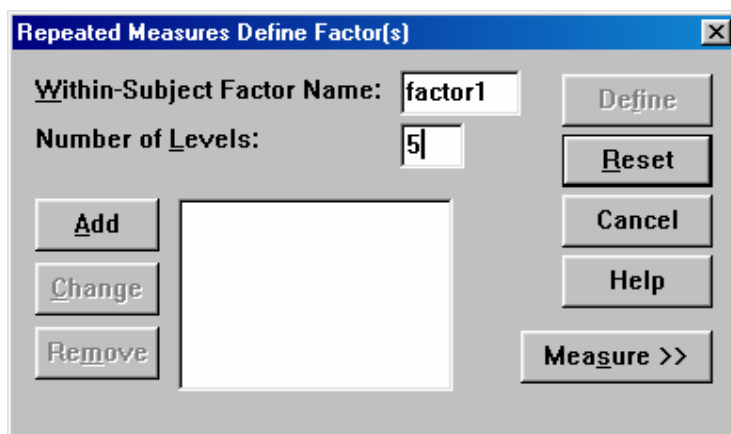
Mygtukais dialogo langelio apačioje galima nurodyti papildomų parametrų skaičiavimą arba pakeisti nustatytuosius. Panagrinėsime svarbiausius iš jų (kai kurie testai yra numatyti tik kelių faktorių dispersinei analizei ir plačiau nagrinėjami 15.4 skyrelyje).

Spragtelėję mygtuką *Plots...* galite nurodyti nubraižyti atskirų bandymų (priklausomų kintamųjų) vidurkių priklausomybės (trendų) nuo veikiančio faktoriaus reikšmių grafiką. Blokuotųjų duomenų vieno faktoriaus analizės atveju tereikia įkelti veikiančią faktorių iš sąrašo *Factors* į laukelį *Horizontal Axis* ir spragtelėti mygtuką *Add*.

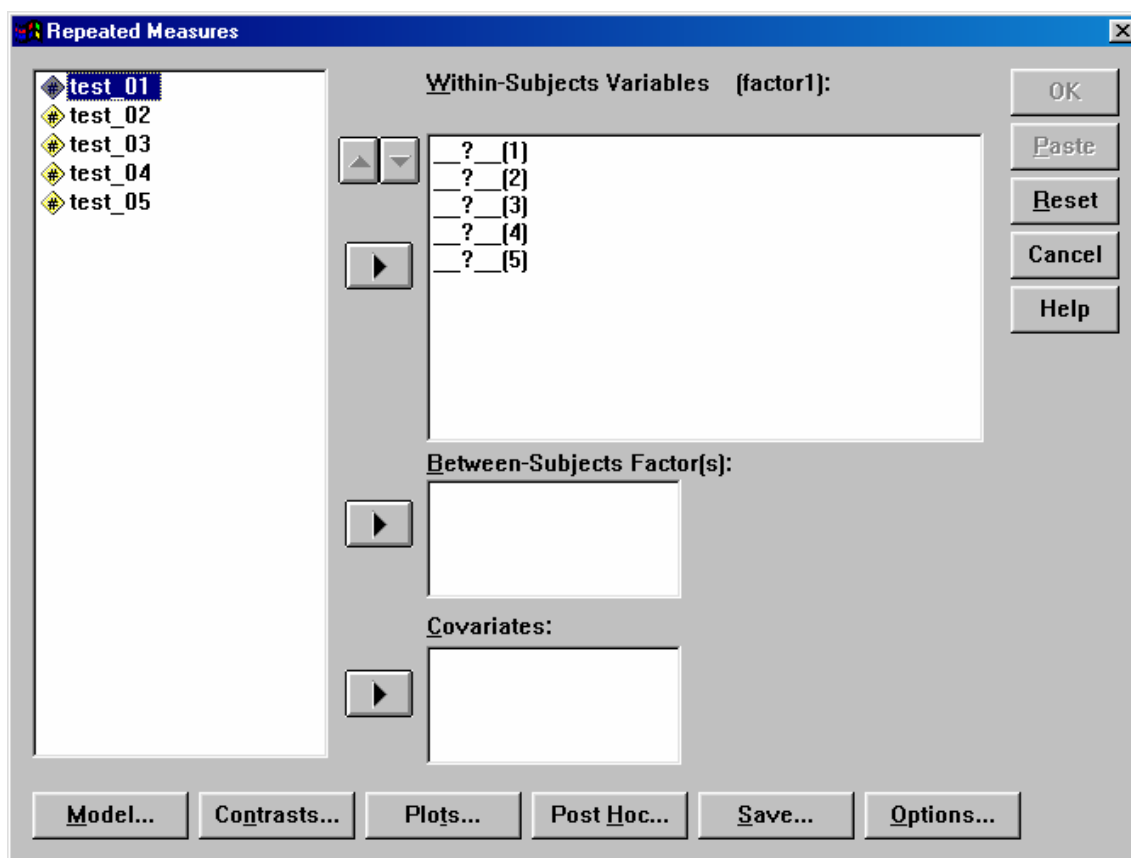


	test_01	test_02	test_03	test_04	test_05	var	var	var
1	34	37	45	50	53			
2	28	41	47	52	54			
3	25	32	44	52	57			
4	30	37	42	50	56			
5	33	34	48	56	55			
6	38	42	47	52	57			
7	41	41	45	58	63			
8	46	39	50	50	60			
9	22	37	49	47	53			
10	17	20	30	40	47			
11								
12								

15.6 pav. Krepšininkų komandos baudos metimo kartotinių testų rezultatai



15.7 pav. Dialogo langelis *Repeated Measures Define Factor(s)*



15.8 pav. Dialogo langelis *Repeated Measures*

Post Hoc kriterijus dialogo langelyje ***Post Hoc...*** galima pasirinkti tik blokuotųjų duomenų kelių faktorių dispersinės analizės atveju, kai grupės faktorius — *Between-Subjects Factors* turi ne mažiau kaip tris reikšmes (kategorijas). Blokuotųjų duomenų vieno faktoriaus dispersinės analizės atveju

Post Hoc... dialogo langelis yra neaktyvus, o kai kuriuos *Post Hoc* kriterijus galima nurodyti **Options...** dialogo langelyje.

Spragtelėję mygtuką **Options...** dialogo langelyje **Repeated Measures: Options** (15.9 pav.), galite:

- Pasirinkti iš sąrašo **Factor(s) and Factor Interactions** (faktoriai ir faktorių sąveika) faktorius ir faktorių sąveikas, kurių priklausomų kintamųjų populiacijų vidurkių įverčius norite gauti. Vieno faktoriaus dispersinės analizės atveju (kaip pateiktame pavyzdyje) galite pasirinkti bendro visų kintamųjų vidurkio įvertį (**OVERALL**) ir atskirų priklausomų kintamųjų populiacijų vidurkių įverčius pagal veikiančio faktoriaus reikšmes. Pasirinktą faktorių įkelkite į sąrašą **Display Means for**. Įkėlę į sąrašą **Display Means for** veikiančią faktorių (šiuo atveju — *factor1*), papildomai galite pažymėti laukelį **Compare main effects**, kad gautumėte visų priklausomų kintamųjų vidurkių įverčių porinius palyginimus, t. y. pritaikyti *Post Hoc* kriterijų. Vieną iš trijų kriterijų (LSD, Bonferroni, Sidak'o) pasirinkite išskleidžiamajame sąrašo **Confidence interval adjustment**.

Repeated Measures: Options

Estimated Marginal Means

Factor(s) and Factor Interactions:

[OVERALL]
factor1

Display Means for:

☐ Compare main effects

Confidence interval adjustment:

LSD (none)

Display

☐ Descriptive statistics
☐ Estimates of effect size
☐ Observed power
☐ Parameter estimates
☐ SSCP matrices
☐ Residual SSCP matrix

☐ Transformation matrix
☐ Homogeneity tests
☐ Spread vs. level plots
☐ Residual plots
☐ Lack of fit test
☐ General estimable function

Significance level: .05 **Confidence intervals are 95%**

Continue Cancel Help

15.9 pav. Dialogo langelis **Repeated Measures: Options**

- Pasirinkti komandų grupėje **Display** įvairias charakteristikas (paminėsime tik dažniausiai naudojamas):
 - **Descriptive statistics** (aprašomoji statistika): imčių vidurkius, standartinį nuokrypį, kiekvieno priklausomo kintamojo stebėjimų skaičių.
 - **Estimates of effect size** (poveikio įvertis). Dalinis Eta-kvadrato (η^2) koeficientas (*Partial Eta Squared*) padeda atsakyti į klausimą, kaip matuojamo kintamojo reikšmė priklauso nuo to, kokiomis sąlygomis buvo matuojama (t. y. kokią įtaką faktorius daro priklausomiems kintamiesiems palyginus su atsitiktine paklaida). Blokuotųjų duomenų atveju koeficientas η^2 skaičiuojamas pagal formulę

$$\eta^2 = \frac{SSTR}{SSTR + SSE} \quad (15.7)$$

$0 \leq \eta^2 \leq 1$; Kuo koeficientas η^2 didesnis, tuo didesnė sąlygų (veikiančio faktoriaus) įtaka matavimo rezultatams lyginant su atsitiktine paklaida.

- **Observed power** (stebimoji galia). Pateikiama kriterijaus galia. Kriterijaus galia β — tai tikimybė atmesti hipotezę H_0 , kai ji klaidinga: $\beta = P(H_0 \text{ atmetama} / H_0 \text{ klaidinga})$, t. y. $\beta = 1 - P$ (antrosios rūšies klaida).
- **Parameter estimates** (parametriniai įverčiai). Pateikiami kintamųjų priklausomybės nuo veikiančių faktorių parametrinio (regresinio) modelio rezultatai: koeficientų reikšmės, standartinė paklaida, *t*-kriterijaus reikšmė, *p*-reikšmė, pasikliautiniai intervalai, kriterijaus galia.
- Šiame pavyzdyje paliksime nustatytuosius dialogo langelio **Repeated Measures** ir papildomų dialogo langelių, išskviečiamų mygtukais **Model...**, **Contrast...**, **Post Hoc...**, **Save...**, parametrus. Dialogo langelyje **Options...** į sąrašą **Display Means for** įkelsime *factor 1*, pažymėsime laukelį **Compare main effects**, o išskleidžiamajame sąraše **Confidence interval adjustment** pasirinksime Bonferroni variantą, komandų grupėje **Display** pažymėsime **Descriptive statistics** ir **Estimates of effect size** laukelius. Dialogo langelyje **Plots...** pasirinksime vidurkių trendo grafiką. Krepšininkų komandos baudos metimo kartotinių bandymų dispersinės analizės pagrindiniai rezultatai parodyti 15.10.1 — 15.10.3 pav.

15.10.1 pav. pirmoje a) lentelėje *Descriptive Statistics* pateikti penkių testų vidurkiai ir standartiniai nuokrypiai. Matome, kad imčių vidurkiai skiriasi —treniruojantis didėja. Tačiau ar šis skirtumas statistiškai reikšmingas?

Antroje *b*) lentelėje *Mauchly's Test of Sphericity* Mauchly kriterijumi yra tikrinama sferiškumo prielaida (blokuotųjų duomenų dispersinės analizės atveju Mauchly kriterijaus rezultatai pateikiami automatiškai). Kadangi *p*-reikšmė (*Sig.*) lygi 0,424, t. y. didesnė už reikšmingumo lygmenį 0,05, galima teigti, kad sferiškumo prielaida tenkinama.

Trečioje *c*) lentelėje *Multivariate Tests* (daugiamačiai testai; vieno faktoriaus atveju šioje lentelėje pateikiami šio faktoriaus poveikio rezultatai — mūsų atveju tai *Factor 1*) pateikta keturių kriterijų *Pillai's Trace* (Pillai pėdsakas), *Wilks' Lambda* (Wilks'o lambda), *Hotelling's Trace* (Hotelling'o pėdsakas) ir *Roy's Largest Root* (didžiausia būdingoji šaknis pagal Roy'o metodą) *p*-reikšmė (*Sig.*) yra lygi nuliui trijų ženklų po kablelio tikslumu. Kadangi *p*-reikšmė mažesnė už reikšmingumo lygmenį $\alpha = 0,05$, nulinė hipotezė H_0 apie vidurkių lygybę atmetama — atskirų testavimų rezultatai (jaunųjų sportininkų baudos pataikymo vidurkis) skiriasi statistiškai reikšmingai. Konservatyviausiu yra laikomas *Pillai's Trace* kriterijus.

Kad pakartotiniai testai turi didelę įtaką rezultatams rodo ir didelė (0,970) koeficiento η^2 (*Partial Eta Squared*) reikšmė.

15.10.2 pav. pirmoje *a*) lentelėje *Tests of Within-Subject Effects* (poveikio tiriamajam požymiui testas) taip pat pateikti keturi sąlygų eilutės *Factor* variantai. Viršutinės eilutės *Sphericity Assumed* atitinka mūsų nagrinėjamą atvejį, kai sferiškumo prielaida yra tenkinama. Atitinkama *p*-reikšmė (*Sig.*) yra lygi nuliui trijų ženklų po kablelio tikslumu. Kadangi *p*-reikšmė mažesnė už reikšmingumo lygmenį $\alpha = 0,05$, nulinė hipotezė H_0 apie vidurkių lygybę atmetama – atskirų testavimų rezultatai (jaunųjų sportininkų baudos pataikymo vidurkis) skiriasi statistiškai reikšmingai. Taigi 15.10.1 pav. *c*) ir 15.10.2 pav. *a*) lentelėse pateikiami tie patys rezultatai, gauti skirtingais metodais.

Jeigu sferiškumo hipotezė nepasitvirtina, reikia naudotis viena iš trijų siūlomų pataisų: *Greenhouse-Geisser*, *Huynh-Feld* arba *Lower-bound* (apatinio rėžio).

15.10.2 pav. antroje *b*) lentelėje *Pairwise Comparisons* (poriniai palyginimai) pateikti Bonferroni *Post Hoc* testo rezultatai. Statistiškai reikšmingai skiriasi visų galimų testų porų rezultatai, išskyrus pirmo ir antro testo porą.

15.10.3 pav. pateikiamas jaunųjų sportininkų baudos pataikymo vidurkių trendo grafikas *a*).

Descriptive Statistics

	Mean	Std. Deviation	N
TEST_01	31,40	8,872	10
TEST_02	36,00	6,446	10
TEST_03	44,70	5,697	10
TEST_04	50,70	4,900	10
TEST_05	55,50	4,327	10

a)

Mauchly's Test of Sphericity^b

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon ^a		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
FACTOR1	,287	9,267	9	,424	,635	,906	,250

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

- a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.
- b. Design: Intercept
Within Subjects Design: FACTOR1

b)

Multivariate Tests^b

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
FACTOR1	Pillai's Trace	,970	47,966 ^a	4,000	6,000	,000	,970
	Wilks' Lambda	,030	47,966 ^a	4,000	6,000	,000	,970
	Hotelling's Trace	31,977	47,966 ^a	4,000	6,000	,000	,970
	Roy's Largest Root	31,977	47,966 ^a	4,000	6,000	,000	,970

- a. Exact statistic
- b.
Design: Intercept
Within Subjects Design: FACTOR1

c)

15.10.1 pav. Blokuotųjų duomenų dispersinės analizės rezultatai

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
FACTOR1	Sphericity Assumed	3998,120	4	999,530	72,157	,000	,889
	Greenhouse-Geisser	3998,120	2,542	1573,061	72,157	,000	,889
	Huynh-Feldt	3998,120	3,626	1102,716	72,157	,000	,889
	Lower-bound	3998,120	1,000	3998,120	72,157	,000	,889
Error(FACTOR1)	Sphericity Assumed	498,680	36	13,852			
	Greenhouse-Geisser	498,680	22,875	21,801			
	Huynh-Feldt	498,680	32,631	15,282			
	Lower-bound	498,680	9,000	55,409			

a)

Pairwise Comparisons

Measure: MEASURE_1

(I) FACTOR1	(J) FACTOR1	Mean Difference (I-J)	Std. Error	Sig. ^a	95% Confidence Interval for Difference ^a	
					Lower Bound	Upper Bound
1	2	-4,600	2,012	,481	-12,024	2,824
	3	-13,300*	2,246	,002	-21,588	-5,012
	4	-19,300*	2,161	,000	-27,272	-11,328
	5	-24,100*	1,871	,000	-31,002	-17,198
2	1	4,600	2,012	,481	-2,824	12,024
	3	-8,700*	1,126	,000	-12,854	-4,546
	4	-14,700*	1,461	,000	-20,091	-9,309
	5	-19,500*	1,432	,000	-24,783	-14,217
3	1	13,300*	2,246	,002	5,012	21,588
	2	8,700*	1,126	,000	4,546	12,854
	4	-6,000*	1,414	,022	-11,218	-,782
	5	-10,800*	1,451	,000	-16,155	-5,445
4	1	19,300*	2,161	,000	11,328	27,272
	2	14,700*	1,461	,000	9,309	20,091
	3	6,000*	1,414	,022	,782	11,218
	5	-4,800*	,940	,006	-8,270	-1,330
5	1	24,100*	1,871	,000	17,198	31,002
	2	19,500*	1,432	,000	14,217	24,783
	3	10,800*	1,451	,000	5,445	16,155
	4	4,800*	,940	,006	1,330	8,270

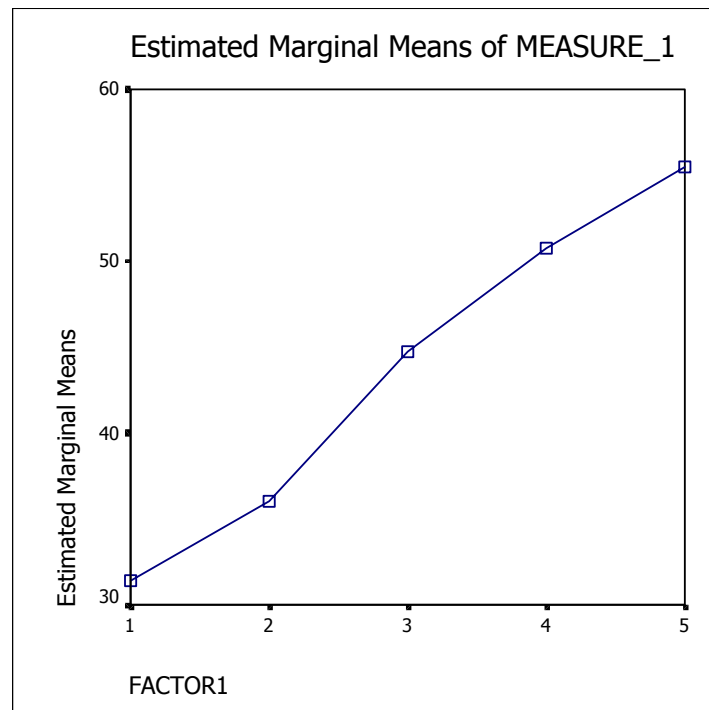
Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

a. Adjustment for multiple comparisons: Bonferroni.

b)

15.10.2 pav. Blokuotųjų duomenų dispersinės analizės rezultatai



a)

15.10.3 pav. Blokuotųjų duomenų dispersinės analizės rezultatai

15.3. DVIEJŲ FAKTORIŲ DISPERSINĖ ANALIZĖ

SPSS pakete realizuota daugelio faktorių dispersinė analizė. Apsiribosime dažniausiai naudojama dviejų faktorių dispersine analize — kelių faktorių dispersinės analizės atveju atsiranda daugiau tarpusavio sąveikų ir jas sunkiau interpretuoti. Dviejų ir vieno faktoriaus dispersinės analizės duomenų struktūros labai panašios, tik dviejų faktorių dispersinėje analizėje populiacijos skiriamos atsižvelgiant į du faktorius (nepriklausomus kintamuosius). Pagrindinis dviejų faktorių dispersijos analizės tikslas — nustatyti, ar priklausomo intervalinio kintamojo vidurkiai skirtingose populiacijose skiriasi, taip pat — nustatyti nepriklausomų faktorių tarpusavio sąveiką. Trumpai pažymėsime, kad faktorių tarpusavio sąveika yra tada, kai skirtingoms faktorių reikšmėms nustatomi nevienodi populiacijų vidurkių skirtumai.

Priėmę, kad pagal du nepriklausomus faktorius A ir B yra skiriama $a \times b$ nepriklausomų populiacijų ir pažymėję (i, j) populiaciją, kurioje faktorius A įgyja reikšmę i , o faktorius B — reikšmę j , o X_{ijk} — k -ąją (i, j) imties elementą, galime užrašyti struktūrinį dvifaktoriinės dispersinės analizės modelį, pagal kurį visa kvadratų suma SST , apimanti visų duomenų skirtumus nuo bendrojo vidurkio, skyla į keturias komponentes — faktoriaus A kvadratų sumą $SSFA$, faktoriaus B kvadratų sumą $SSFB$, kvadratų sumą $SSAB$ faktorių A ir B tarpusavio sąveikai ir paklaidos kvadratų sumą SSE , t. y.

$$SST = SSFA + SSFB + SSAB + SSE, \quad (15.8)$$

čia

$$SST = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (X_{ijk} - \bar{X})^2, \quad (15.9)$$

$$SSFA = nb \sum_{i=1}^a (\bar{X}_{i.} - \bar{X})^2, \quad SSFB = na \sum_{j=1}^b (\bar{X}_{.j} - \bar{X})^2, \quad (15.10)$$

$$SSAB = n \sum_{i=1}^a \sum_{j=1}^b (\bar{X}_{ij} - \bar{X}_{i.} - \bar{X}_{.j} + \bar{X})^2, \quad (15.11)$$

$$SSE = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (X_{ijk} - \bar{X}_{ij})^2, \quad (15.12)$$

$\bar{X}_{i.}$ — duomenų lentelės i -osios eilutės vidurkis, $\bar{X}_{.j}$ — duomenų lentelės j -tojo stulpelio vidurkis, \bar{X}_{ij} — (i, j) populiacijos kintamojo vidurkis, n — imčių didumai, $N = nab$ — bendras visų stebėjimų skaičius.

Dviejų faktorių dispersinės analizės prielaidos:

- Kintamųjų X_{ij} normalumas.
- Kintamųjų X_{ij} dispersijų lygybė.
- Kintamųjų X_{ij} nepriklausomumas.

Dispersinės analizės metodas nėra kritiškas nukrypimams nuo normaliojo pasiskirstymo, tačiau duomenų pasiskirstymas turi būti simetrinis.

Dviejų faktorių dispersinė analizė leidžia formuluoti tris statistines hipotezes: apie faktoriaus A įtaką, apie faktoriaus B įtaką ir apie faktorių A ir B tarpusavio sąveiką. Intervalų skalėje matuojamiems nepriklausomiems vienodai pasiskirsčiusiems kintamiesiems $X_{ij} \sim N(\mu_{ij}, \sigma^2)$, $i = 1, \dots, a$, $j = 1, \dots, b$, kurių vidurkiai μ_{ij} ir dispersijos σ^2 nežinomi, galima tikrinti šias hipotezes

$$\begin{cases} H_0: \mu_{1.} = \mu_{2.} = \dots = \mu_{a.}, \\ H_1: \text{ bent du vidurkiai skiriasi;} \end{cases} \quad (15.13)$$

čia nulinė hipotezė H_0 reiškia, kad faktoriaus A įtakos nėra;

$$\begin{cases} H_0: \mu_{.1} = \mu_{.2} = \dots = \mu_{.b}, \\ H_1: \text{ bent du vidurkiai skiriasi;} \end{cases} \quad (15.14)$$

čia nulinė hipotezė H_0 reiškia, kad faktoriaus B įtakos nėra;

$$\begin{cases} H_0: \mu_{ij} - \mu_{i.} - \mu_{.j} + \mu = 0, \\ H_1: \mu_{ij} - \mu_{i.} - \mu_{.j} + \mu \neq 0; \end{cases} \quad (15.15)$$

čia nulinė hipotezė H_0 reiškia, kad faktorių A ir B tarpusavio sąveikos nėra;

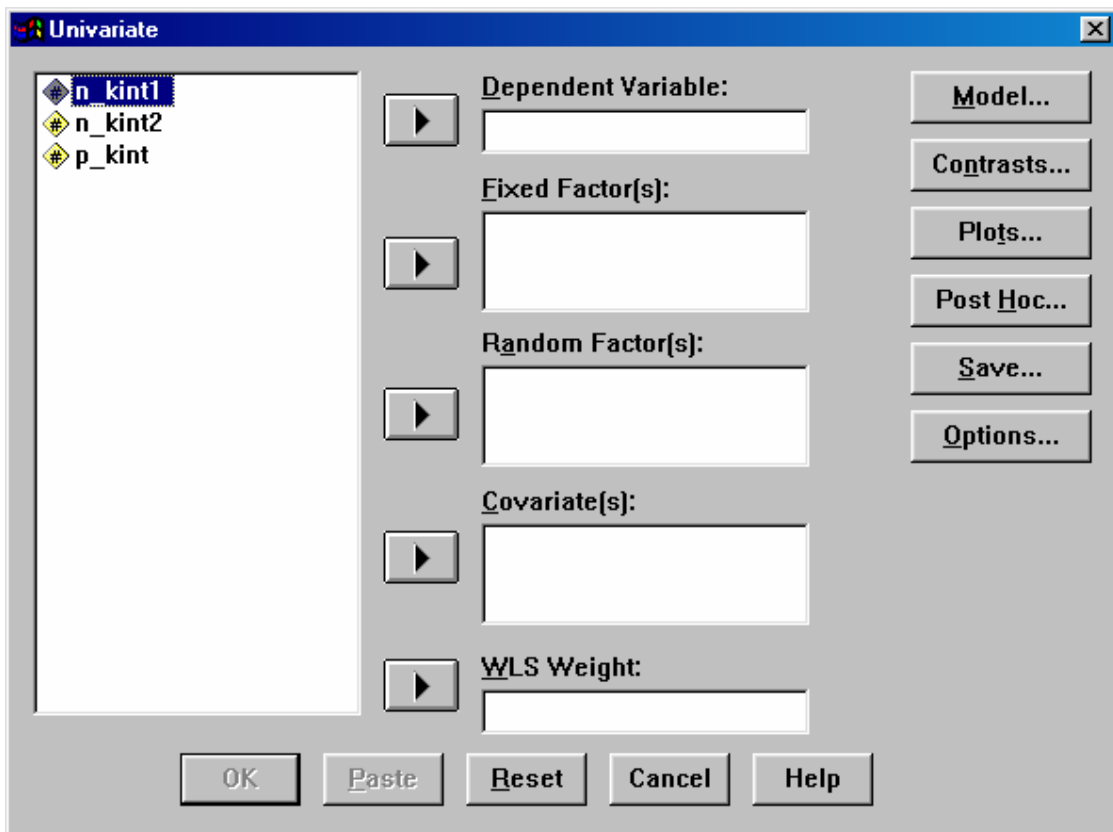
$$\mu_{i.} = \frac{\sum_{j=1}^b \mu_{ij}}{b}, \quad \mu_{.j} = \frac{\sum_{i=1}^a \mu_{ij}}{a}, \quad \mu — \text{ bendrasis vidurkis.}$$

Kiekvienai statistinei hipotezei galioja ta pati sprendimo taisyklė. Tegul reikšmingumo lygmuo lygus α , o p -reikšmė yra p . Tuomet nulinę hipotezę H_0 atmetame, jeigu $p < \alpha$; nulinę hipotezę H_0 priimame, jeigu $p \geq \alpha$. Svarbu tik žinoti, ką kokia nulinė hipotezė reiškia, t. y. p -reikšmę priskirti atitinkamam faktoriui.

Vienmatei kelių faktorių dispersinei analizei atlikti:

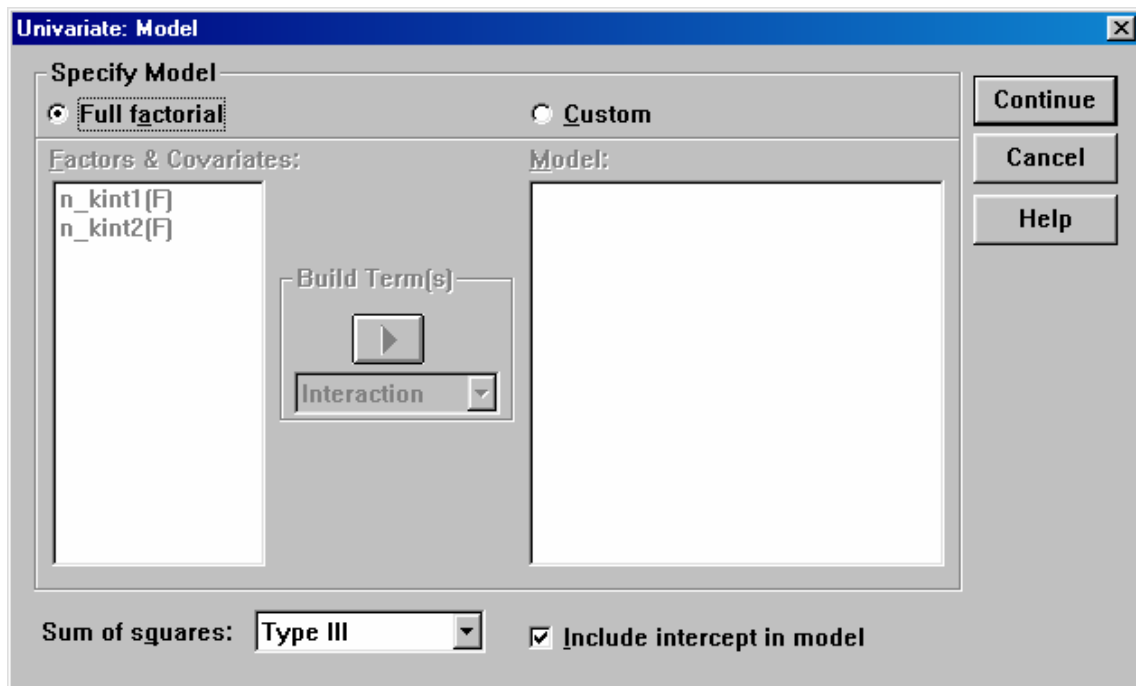
- Atidarykite bylą su analizuojamais duomenimis.
- Nurodykite komandas *Analyze* → *General Linear Model* → *Univariate...*
- Atsidariusiame dialogo langelyje *Univariate* (15.11 pav.) kintamųjų sąraše pažymėkite priklausomą kintamąjį ir įkelkite jį į laukelį *Dependent Variable*.

Dialogo langelyje yra trys laukeliai nepriklausomiems kintamiesiems nurodyti. Į laukelį *Fixed Factor(s)* (nustatyti faktoriai) įkelkite tuos kintamuosius (faktorius), kuriems nurodomos visos klasifikacinės reikšmės (pvz., lytis: vyras — moteris, išsilavinimas: vidurinis — aukštasis neuniversitetinis — aukštasis universitetinis ir t. t.). Į laukelį *Random Factor(s)* (atsitiktiniai faktoriai) įkelkite kintamuosius, kurių reikšmės atrenkamos atsitiktinai iš begalinės generalinės galimų reikšmių populiacijos (tuo atveju, kai nagrinėjami atsitiktiniai faktoriai, paprastai yra atliekama dispersijos komponentų analizė — *Variance Components...*). Į laukelį *Covariate(s)* įkelkite intervalinius nepriklausomus kintamuosius.



15.11 pav. Dialogo langelis *Univariate*

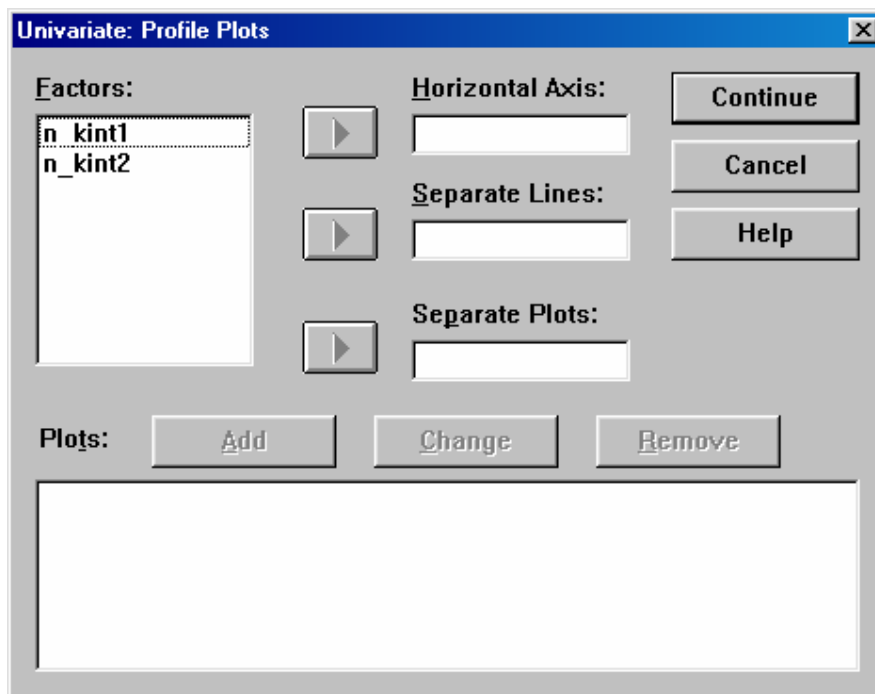
- Spragtelėję mygtuką **Model...**, naujame dialogo langelyje **Univariate: Model** (15.12 pav.) galite pasirinkti dispersinės analizės modelį. Nustatytas yra **Full factorial** (visas faktorinis) variantas, pagal kurį yra įvertinama visų faktorių įtaka bei šių faktorių sąveika (išskyrus sąveiką tarp nepriklausomų intervalinių kintamųjų — kovariančių (*Covariates*)). Pasirinkę alternatyvųjį variantą **Custom** galite įvertinti sąveiką tarp nurodytų faktorių ir tarp faktorių ir kovariančių. Sąveikos lygis yra nurodomas išskleidžiamajame sąrašė **Build Term(s)**. Nustatytasis variantas — **Interaction**, pagal kurį įvertinami visi sąveikos lygiai. Sąrašė **Factors & Covariates** nustatyti (fiksiuoti) faktoriai yra pažymimi raide *F*, o kovariantės — *C*. Atsitiktiniai faktoriai pažymimi raide *R*. Išskleidžiamajame sąrašė **Sum of squares** galite pasirinkti kvadratų sumų skaičiavimo metodą. Dažniausiai naudojamas **Type III** kvadratų sumų skaičiavimo metodas, kuris būna nustatytas iš anksto. Spragtelėkite dialogo langelio **Univariate: Model** mygtuką **Continue**.



15.12 pav. Dialogo langelis **Univariate: Model**

- Spragtelėję dialogo langelio **Univariate** mygtuką **Contrasts...** naujame dialogo langelyje **Univariate: Contrasts** galite pasirinkti kontrastų tipą. Kontrastai (žr. 15.1 skyrių) taikomi hipotezėms apie tiesines populiacijų vidurkių kombinacijas tikrinti. Galima pasirinkti šiuos kontrastų tipus:
 - **Deviation** (nuokrypis). Palygina pagal veikiančio faktoriaus reikšmes suskirstytų populiacijų kintamųjų vidurkius (išskyrus kategoriją, kuri laikoma pagrindine) su bendru visų populiacijų vidurkiu.

- **Simple** (paprastas). Palygina pagal veikiančio faktoriaus reikšmes suskirstytų populiacijų kintamųjų vidurkius su pasirinktą faktoriaus reikšmę atitinkančios populiacijos vidurkiu.
 - **Difference** (skirtumo). Palygina pagal veikiančio faktoriaus reikšmes suskirstytų populiacijų (išskyrus pirmąją) kintamųjų vidurkius su prieš tai esančių populiacijų kintamųjų vidurkiu.
 - **Helmert'o**. Palygina pagal veikiančio faktoriaus reikšmes suskirstytų populiacijų (išskyrus paskutiniąją) kintamųjų vidurkius su paskesniųjų populiacijų kintamųjų vidurkiu.
 - **Repeated** (pakartotinis). Palygina pagal veikiančio faktoriaus reikšmes suskirstytų populiacijų (išskyrus paskutiniąją) kintamųjų vidurkius su paskesniosios populiacijos kintamojo vidurkiu.
 - **Polynomial** (polinominis). Leidžia patikrinti hipotezę apie populacijose stebimų kintamųjų vidurkių kitimo tendenciją (vidurkių tendą): tiesinį vidurkių tendą, kvadratinį ir t. t.
- Spragtelėkite dialogo langelio **Univariate** mygtuką **Plots...**(diagramos). Atsidarys naujas dialogo langelis **Univariate: Profile Plots** (15.13 pav.) vidurkių tendų linijinei diagramai sudaryti.



15.13 pav. Dialogo langelis **Univariate: Profile Plots**

- Įkelkite nepriklausomą kintamąjį, nuo kurio priklausomo kintamojo vidurkio priklausomybę norite pavaizduoti, į laukelį **Horizontal Axis** (horizontalioji ašis). Antrą nepriklausomą kintamąjį galite įkelti į laukelį **Separate Lines** (atskiros linijos) — bus nubrėžtos atskiros kreivės

kiekvienai šio kintamojo reikšmei. Kiekvienai nepriklausomo kintamojo, įkelto į laukelį **Separate Plots** (atskiros diagramos), reikšmei bus sudarytos atskiros diagramos. Praktiškai galima teigti, kad faktorių tarpusavio sąveikos nėra, jeigu atitinkami trendų grafikai sutampa arba yra identiškos formos ir skiriasi tik postūmiu. Visais kitais atvejais tarpusavio sąveika yra.

- Įkėlę į laukelius **Horizontal Axis**, **Separate Lines**, **Separate Plots** reikiamus kintamuosius, kiekvieną kartą spragtelėkite mygtuką **Add**. Sudarę norimų diagramų sąrašą, spragtelėkite mygtuką **Continue** ir grįžkite į pagrindinį dialogo langelį.
- Spragtelėkite dialogo langelio **Univariate** mygtuką **Post Hoc...** Atsidarys naujas dialogo langelis **Univariate: Post Hoc Multiple Comparisons for Observed Means** aposterioriniams (*Post Hoc*) kriterijams pasirinkti (15.14 pav.).

Nustačius faktą, kad ne visų imčių vidurkiai lygūs, *Post Hoc* kriterijai padeda nustatyti, kurių imčių vidurkiai statistiškai reikšmingai skiriasi.

Kaip ir vieno faktoriaus dispersinės analizės atveju, yra pateikiama 18 aposteriorinių (*Post Hoc*) kriterijų, iš kurių dažniausiai naudojami apibūdinti 15.1 skyriuje. Pridursime, kad be **Tukey'o** kriterijaus, kitas labiausiai konservatyvus, t. y. labiausiai nelinkęs atmesti nulinės hipotezės — vidurkių skirtumus pripažinti statistiškai reikšmingais, yra **Scheffe** kriterijus.

Univariate: Post Hoc Multiple Comparisons for Observed Means

Factor(s):
n_kint1
n_kint2

Post Hoc Tests for:
n_kint1

Equal Variances Assumed

<input type="checkbox"/> LSD	<input type="checkbox"/> S-N-K	<input type="checkbox"/> Waller-Duncan
<input type="checkbox"/> Bonferroni	<input type="checkbox"/> Tukey	Type I/Type II Error Ratio: 100
<input type="checkbox"/> Sidak	<input type="checkbox"/> Tukey's-b	<input type="checkbox"/> Dunnett
<input type="checkbox"/> Scheffe	<input type="checkbox"/> Duncan	Control Category: Last
<input type="checkbox"/> R-E-G-W F	<input type="checkbox"/> Hochberg's GT2	Test:
<input type="checkbox"/> R-E-G-W Q	<input type="checkbox"/> Gabriel	<input checked="" type="radio"/> 2-sided <input type="radio"/> < Control <input type="radio"/> > Control

Equal Variances Not Assumed

<input type="checkbox"/> Tamhane's T2	<input type="checkbox"/> Dunnett's T3	<input type="checkbox"/> Games-Howell	<input type="checkbox"/> Dunnett's C
---------------------------------------	---------------------------------------	---------------------------------------	--------------------------------------

15.14 pav. Dialogo langelis **Univariate: Post Hoc Multiple Comparisons for Observed Means** aposterioriniams testams pasirinkti

- Norėdami nustatyti, kurios iš skiriamų pagal nepriklausomus kintamuosius (faktorius) populiacijų kintamieji turi statistiškai reikšmingai besiskiriančius vidurkius, įkelkite tuos nepriklausomus kintamuosius (faktorius) į laukelį **Post Hoc Tests for**.

Pastaba. *Post Hoc* kriterijai taikomi tik tuomet, kai imčių yra ne mažiau kaip trys.

- Pasirinkę reikiamą *Post Hoc* kriterijų, spragtelėkite mygtuką **Continue** ir grįžkite į pagrindinį dialogo langelį.
- Spragtelėkite dialogo langelio **Univariate** mygtuką **Options...** Atsidarys naujas dialogo langelis **Univariate: Options** (15.15 pav.).

Univariate: Options

Estimated Marginal Means

Factor(s) and Factor Interactions:

(OVERALL)
n_kint1
n_kint2
n_kint1*n_kint2

Display Means for:

☐ Compare main effects

Confidence interval adjustment:
LSD (none)

Display

☐ Descriptive statistics
☐ Estimates of effect size
☐ Observed power
☐ Parameter estimates
☐ Contrast coefficient matrix

☐ Homogeneity tests
☐ Spread vs. level plot
☐ Residual plot
☐ Lack of fit
☐ General estimable function

Significance level: .05 Confidence intervals are 95%

Continue Cancel Help

15.15 pav. Dialogo langelis **Univariate: Options**

- Pasirinkite iš sąrašo **Factor(s) and Factor Interactions** (faktoriai ir faktorių sąveika) faktorius arba faktorių sąveikas, kuriems norite gauti populiacijų kintamųjų vidurkių įverčius. Galite pasirinkti populiacijų kintamųjų vidurkių įverčius pagal visas šių faktorių reikšmes arba bendrojo vidurkio įvertį (**OVERALL**), sąlygojamą visų veikiančių faktorių. Pasirinktus faktorius įkelkite į sąrašą **Display Means for**.

- Pasirinkite komandų grupėje **Display** norimas charakteristikas (kai kurios šių charakteristikų analogiškos blokuotųjų duomenų atvejui ir buvo apibūdintos 15.2 skyriuje, todėl čia jų nekartosime):
 - **Estimates of effect size** (įtakos įvertis). Dalinis Eta-kvadrato (η^2) koeficientas (*Partial Eta Squared*) lygina duomenų skirtumus, atsiradusius dėl faktoriaus (ar faktorių sąveikos), su skirtumais, kuriuos paaiškina atsitiktinė paklaida. $0 \leq \eta^2 \leq 1$. Kuo η^2 didesnis, tuo faktoriaus įtaka rezultatams didesnė, palyginti su skirtumais, atsirandančiais dėl imties atsitiktinumo. Dviejų faktorių dispersinėje analizėje koeficientas η^2 skaičiuojamas taip:

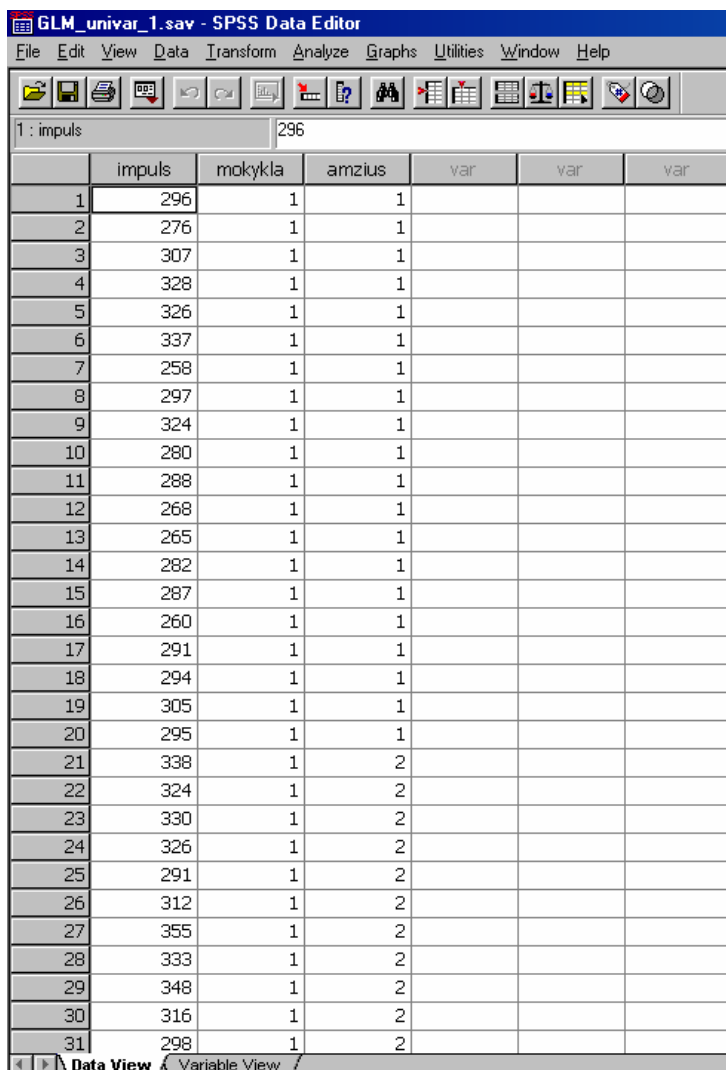
$$\eta_A^2 = \frac{SSFA}{SSFA + SSE}, \eta_B^2 = \frac{SSFB}{SSFB + SSE}, \eta_{AB}^2 = \frac{SSAB}{SSAB + SSE}; \quad (15.16)$$

- **Contrast coefficient matrix** — pateikiami pasirinkto kontrasto koeficientai.
- **Homogeneity tests** (homogeniškumo testas). Pateikiamas priklausomo kintamojo dispersijų homogeniškumo Levene testo rezultatas pagal visas veikiančių faktorių reikšmių kombinacijas.
- **General estimable function** (bendroji aproksimacinė funkcija). Pateikiama aproksimacinė funkcija, pagrįsta tiesine kontrastų matricos eilutėse esančių koeficientų kombinacija.
- Spragtelėkite mygtuką **Continue**, o paskui mygtuką **OK** pagrindiniame dialogo langelyje **Univariate**.

Pavyzdys. Buvo tiriamas dviejų krepšinio mokyklų jaunųjų krepšininkų fizinis parengtumas — matuojamas atsispyrimo jėgos impulsas, N_s (duomenys hipotetiniai). Intervalinis priklausomas kintamasis duomenų rinkmenoje pavadintas *impuls*. Vienas faktorius (nepriklausomas kintamasis) yra *mokykla*, turintis reikšmes 1 ir 2. Antras faktorius (nepriklausomas kintamasis) yra *amžius* (duomenų rinkmenoje — kintamasis *amzius*), turintis tris reikšmes: 1 (*14 metų*), 2 (*15 metų*) ir 3 (*16 metų*). Duomenų rinkmenos fragmentas (iš viso yra 120 stebėjimų) parodytas 15.16 pav.

- Dialogo langelyje **Univariate** (15.11 pav.) priklausomą kintamąjį *impuls* įkeliamo į laukelį **Dependent Variable**, o nepriklausomus kintamuosius *mokykla* ir *amzius* — į laukelį **Fixed Factor(s)**.
- Paliekame dialogo langelyje **Univariate: Model** (15.12 pav.) nustatytąjį modelį **Full factorial**, kontrastų nenaudosime.
- Dialogo langelyje **Univariate: Profile Plots** (15.13 pav.) nepriklausomą kintamąjį *amzius* įkeliamo į laukelį **Horizontal Axis**, o nepriklausomą kintamąjį *mokykla* — į laukelį **Separate Lines**.
- Dialogo langelyje **Univariate: Post Hoc Multiple Comparisons for Observed Means** (15.14 pav.) įkeliamo nepriklausomą kintamąjį *amzius* į

laukelį *Post Hoc Test for* (atlikti aposteriorinį testą binariniam kintamajam *mokykla* nėra prasmės) ir pasirenkame *Tukey'o* ir *Scheffe* testus.



	impuls	mokykla	amzius	var	var	var
1	296	1	1			
2	276	1	1			
3	307	1	1			
4	328	1	1			
5	326	1	1			
6	337	1	1			
7	258	1	1			
8	297	1	1			
9	324	1	1			
10	280	1	1			
11	288	1	1			
12	268	1	1			
13	265	1	1			
14	282	1	1			
15	287	1	1			
16	260	1	1			
17	291	1	1			
18	294	1	1			
19	305	1	1			
20	295	1	1			
21	338	1	2			
22	324	1	2			
23	330	1	2			
24	326	1	2			
25	291	1	2			
26	312	1	2			
27	355	1	2			
28	333	1	2			
29	348	1	2			
30	316	1	2			
31	298	1	2			

15.16 pav. Duomenų rinkmenos fragmentas

- Dialogo langelyje *Univariate: Options* (15.15 pav.) įkeliamo *mokykla*amzius* į laukelį *Display Means for* ir pažymime laukelius *Descriptive statistics*, *Estimates of effect size*, *Parameter estimates* ir *Homogeneity tests*.
- Spragtelime dialogo langelio *Univariate* mygtuką *OK*. Pagrindiniai skaičiavimo rezultatai parodyti 15.17.1 — 15.17.4 pav.

15.17.1 pav. lentelėje a) *Between-Subjects Factors* (tarpsubjektiniai faktoriai) pateikiama nepriklausomų kintamųjų suvestinė — kintamųjų reikšmės, jų stebėjimų skaičius.

Lentelėje *b) Descriptive Statistics* (aprašomoji statistika) pateikiami priklausomo kintamojo vidurkiai ir standartiniai nuokrypiai visoms nepriklausomų kintamųjų reikšmėms.

Lentelėje *c) Levene's Test of Equality of Error Variances* pateikiami Levene testo dispersijų homogeniškumui įvertinti rezultatai. Gauta *p*-reikšmė 0,362 rodo, kad dispersijos skiriasi nereikšmingai, t. y. dispersijų homogeniškumo sąlyga yra įvykdoma.

Between-Subjects Factors

	Value Label	N
MOKYKLA 1		60
2		60
AMŽIUS 1	14 metų	40
2	15 metų	40
3	16 metų	40

a)

Descriptive Statistics

Dependent Variable: IMPULS

MOKYKLA	AMŽIUS	Mean	Std. Deviation	N
1	14 metų	293,20	22,881	20
	15 metų	330,15	30,020	20
	16 metų	352,60	23,761	20
	Total	325,32	35,363	60
2	14 metų	306,70	15,928	20
	15 metų	335,35	26,162	20
	16 metų	356,60	23,623	20
	Total	332,88	30,116	60
Total	14 metų	299,95	20,625	40
	15 metų	332,75	27,918	40
	16 metų	354,60	23,474	40
	Total	329,10	32,926	120

b)

Levene's Test of Equality of Error Variances^a

Dependent Variable: IMPULS

F	df1	df2	Sig.
1,104	5	114	,362

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept+MOKYKLA+AMŽIUS+MOKYKLA * AMŽIUS

c)

15.17.1 pav. Pagrindiniai dviejų faktorių dispersinės analizės rezultatai

15.17.2 pav. lentelės a) *Test of Between-Subject Effects* eilutėje *Corrected Model* nurodyta visa kvadratų suma be paklaidos. *Intercept* atitinka statistinę hipotezę apie bendrojo vidurkio lygybę nuliui. Mūsų atveju ši hipotezė neaktuali. Eilutės *MOKYKLA*, *AMŽIUS* IR *MOKYKLA*AMŽIUS* duoda atsakymą apie mokyklos įtaką, amžiaus įtaką ir mokyklos ir amžiaus sąveikos įtaką jaunųjų krepšininkų fizinio parengtumo rodikliams. Šias hipotezes atitinkančios *p*-reikšmės pateiktos stulpelyje *Sig.* Matome, kad statistiškai reikšmingai skiriasi tik skirtingo amžiaus sportininkų rezultatai. Tuo tarpu pagal mokyklas šis skirtumas statistiškai nereikšmingas. Nėra ir tarpusavio sąveikos tarp mokyklos ir sportininkų amžiaus. Eilutėje *Error* pateikiama paklaidos reikšmė, eilutėje *Corrected Total* pateikiama visa kvadratų suma — *SST* pagal (15.9) formulę, eilutėje *Total* — *SST* su pridėta *Intercept* suma.

Lentelėje b) *Estimated Marginal Means MOKYKLA*AMŽIUS* yra pateikiamas priklausomo kintamojo vidurkis visoms nepriklausomų kintamųjų kombinacijoms, standartinė paklaida ir pasikliautinis intervalas.

Tests of Between-Subjects Effects

Dependent Variable: IMPULS

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	62784,700 ^a	5	12556,940	21,616	,000	,487
Intercept	12996817,200	1	12996817,2	22373,08	,000	,995
MOKYKLA	1717,633	1	1717,633	2,957	,088	,025
AMŽIUS	60531,800	2	30265,900	52,101	,000	,478
MOKYKLA * AMŽIUS	535,267	2	267,633	,461	,632	,008
Error	66224,100	114	580,913			
Total	13125826,000	120				
Corrected Total	129008,800	119				

a. R Squared = ,487 (Adjusted R Squared = ,464)

a)

Estimated Marginal Means

MOKYKLA * AMŽIUS

Dependent Variable: IMPULS

MOKYKLA	AMŽIUS	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
1	14 metų	293,200	5,389	282,524	303,876
	15 metų	330,150	5,389	319,474	340,826
	16 metų	352,600	5,389	341,924	363,276
2	14 metų	306,700	5,389	296,024	317,376
	15 metų	335,350	5,389	324,674	346,026
	16 metų	356,600	5,389	345,924	367,276

b)

15.17.2 pav. Pagrindiniai dviejų faktorių dispersinės analizės rezultatai

15.17.3 pav. lentelėje a) *Multiple Comparison* pateikti aposteriorinių (*Post Hoc*) **Tukey'o** ir **Scheffe** testų rezultatai. Iš *p*-reikšmės (*Sig.*) matyti, kad visų amžiaus grupių vidurkiai statistiškai reikšmingai skiriasi. Tai atspindi ir b) lentelėje *Homogenous Subsets* (homogeniniai pogrupiai) pateikti rezultatai — nėra vienašalių pogrupių, kuriuose vidurkiai nesiskirtų.

Post Hoc Tests

Multiple Comparisons

Dependent Variable: IMPULS

	(I) AMŽIUS	(J) AMŽIUS	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Tukey HSD	14 metų	15 metų	-32,80*	5,389	,000	-45,60	-20,00
		16 metų	-54,65*	5,389	,000	-67,45	-41,85
	15 metų	14 metų	32,80*	5,389	,000	20,00	45,60
		16 metų	-21,85*	5,389	,000	-34,65	-9,05
	16 metų	14 metų	54,65*	5,389	,000	41,85	67,45
		15 metų	21,85*	5,389	,000	9,05	34,65
Scheffe	14 metų	15 metų	-32,80*	5,389	,000	-46,17	-19,43
		16 metų	-54,65*	5,389	,000	-68,02	-41,28
	15 metų	14 metų	32,80*	5,389	,000	19,43	46,17
		16 metų	-21,85*	5,389	,000	-35,22	-8,48
	16 metų	14 metų	54,65*	5,389	,000	41,28	68,02
		15 metų	21,85*	5,389	,000	8,48	35,22

Based on observed means.

*. The mean difference is significant at the ,05 level.

a)

Homogeneous Subsets

IMPULS

AMŽIUS	N	Subset		
		1	2	3
Tukey HSD ^{a,t}	14 metų	299,95		
	15 metų		332,75	
	16 metų			354,60
	Sig.	1,000	1,000	1,000
Scheffe ^{a,b}	14 metų	299,95		
	15 metų		332,75	
	16 metų			354,60
	Sig.	1,000	1,000	1,000

Means for groups in homogeneous subsets are displayed.

Based on Type III Sum of Squares

The error term is Mean Square(Error) = 580,913.

a. Uses Harmonic Mean Sample Size = 40,000.

b. Alpha = ,05.

b)

15.7.3 pav. Pagrindiniai dviejų faktorių dispersinės analizės rezultatai

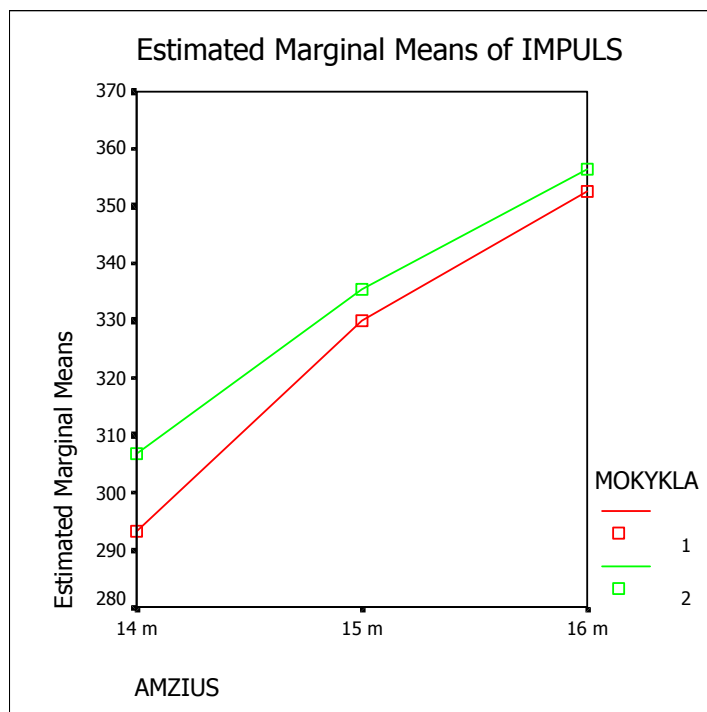
Parameter Estimates

Dependent Variable: IMPULS

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	356,600	5,389	66,167	,000	345,924	367,276
[MOKYKLA=1]	-4,000	7,622	-,525	,601	-19,099	11,099
[MOKYKLA=2]	0 ^a	,	,	,	,	,
[AMŽIUS=1]	-49,900	7,622	-6,547	,000	-64,999	-34,801
[AMŽIUS=2]	-21,250	7,622	-2,788	,006	-36,349	-6,151
[AMŽIUS=3]	0 ^a	,	,	,	,	,
[MOKYKLA=1] *	-9,500	10,779	-,881	,380	-30,853	11,853
[AMŽIUS=1]						
[MOKYKLA=1] *	-1,200	10,779	-,111	,912	-22,553	20,153
[AMŽIUS=2]						
[MOKYKLA=1] *	0 ^a	,	,	,	,	,
[AMŽIUS=3]						
[MOKYKLA=2] *	0 ^a	,	,	,	,	,
[AMŽIUS=1]						
[MOKYKLA=2] *	0 ^a	,	,	,	,	,
[AMŽIUS=2]						
[MOKYKLA=2] *	0 ^a	,	,	,	,	,
[AMŽIUS=3]						

a. This parameter is set to zero because it is redundant.

a)



b)

15.17.4 pav. Pagrindiniai dviejų faktorių dispersinės analizės rezultatai

15.17.4 pav. a) lentelėje *Parameter Estimates* pateiktos regresinio modelio, nusakančio priklausomo kintamojo *IMPULS* priklausomybę nuo veikiančių faktorių ir šių faktorių sąveikos, koeficientų reikšmės bei šių koeficientų reikšmingumas. Modelio rezultatai sutampa su 15.17.2 pav. lentelės a) *Test of Between-Subject Effects* rezultatais — statistiškai reikšmingi yra koeficientai, apibrėžiantys priklausomo kintamojo priklausomybę nuo sportininkų amžiaus. Tuo tarpu koeficientai, apibrėžiantys priklausomybę nuo mokyklos, mokyklos ir sportininko amžiaus sąveikos, nėra statistiškai reikšmingi.

15.17.4 pav. b) diagramoje *Estimated Marginal Means of IMPULS* parodytas priklausomo kintamojo vidurkio įverčio pokytis priklausomai nuo nepriklausomo kintamojo *amzius* reikšmių. Dvi kreivės atitinka dvi binarinio nepriklausomo kintamojo *mokykla* reikšmes. Iš grafiko akivaizdžiai matyti, kad nėra veikiančių faktorių sąveikos — kreivės praktiškai yra identiškos formos ir skiriasi tik postūmiu.

15.4. DAUGIAMATĖ DISPERSINĖ ANALIZĖ

Daugiamatė dispersinė analizė yra taikoma tada, kai reikia ištirti vieno ar kelių faktorių įtaką keliems priklausomiems kintamiesiems, tarp kurių yra koreliacijos ryšys. Kai koreliacijos ryšio tarp priklausomų kintamųjų nėra, tikslinga taikyti vienmatę dispersinę analizę kiekvienam priklausomam kintamajam.

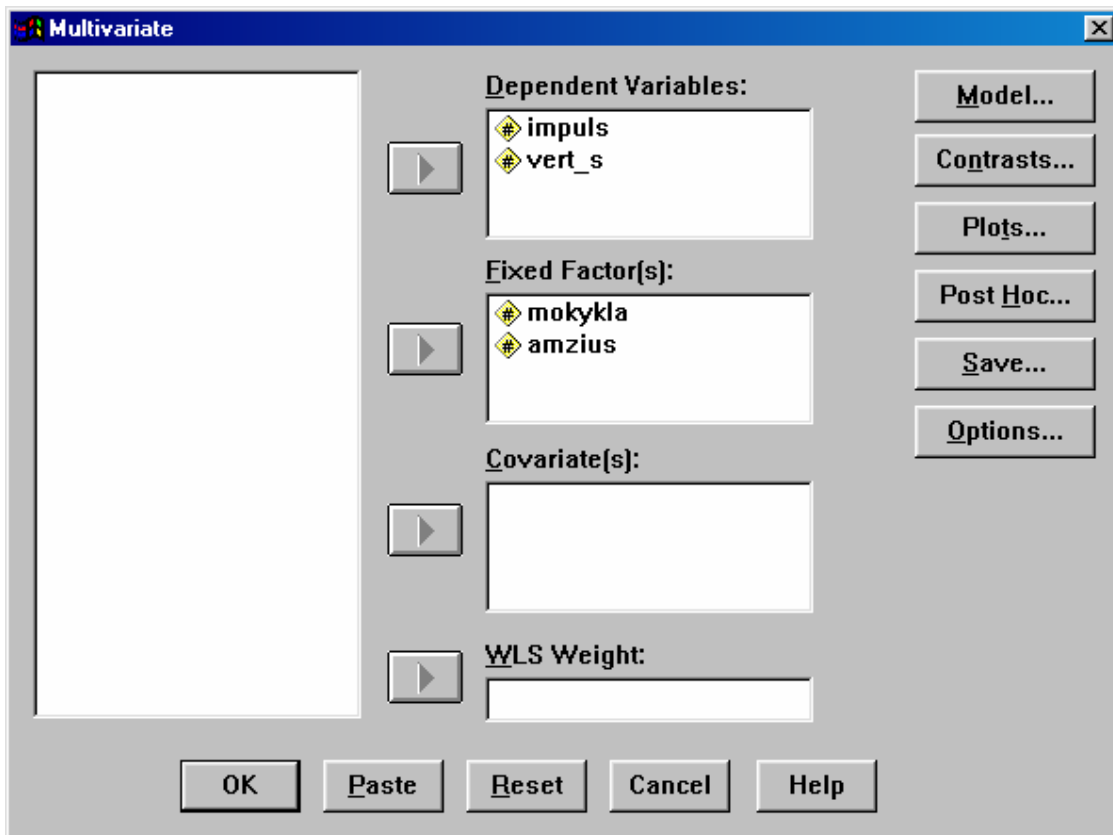
Daugiamatė dispersinė analizė gali būti taikoma, kai tenkinamos šios sąlygos:

- Priklausomų kintamųjų reikšmės turi būti pasiskirsčiusios pagal normalųjį dėsnį. Dispersinės analizės metodas nėra kritiškas nukrypimams nuo normaliojo pasiskirstymo, tačiau duomenų pasiskirstymas turi būti simetrinis.
 - Priklausomų kintamųjų vektorių dispersijos turi būti homogeniškos, taip pat tenkinama kovariacijos matricos sferiškumo prielaida. Šioms prielaidoms patikrinti naudojami Levene ir Box'o M testai.
- Daugiamati kelių faktorių dispersinei analizei atlikti:
- Atidarykite bylą su analizuojamais duomenimis.
 - Nurodykite komandas *Analyze → General Linear Model → Multivariate...*
 - Atsidariusiame dialogo langelyje *Multivariate* (15.18 pav.) kintamųjų sąraše pažymėkite priklausomus kintamuosius ir įkelkite juos į laukelį *Dependent Variables*. Į laukelį *Fixed Factor(s)* (nustatyti faktoriai) įkelkite kategorinius nepriklausomus kintamuosius (faktorius), o į laukelį *Covariate(s)* įkelkite intervalinius nepriklausomus kintamuosius.

Mygtukais *Model...*, *Contrasts...*, *Plots...*, *Post Hoc...*, *Save...*, *Options...* atidarius atitinkamus dialogo langelius galima nurodyti įvairius skaičiavimo variantus ir pasirinkti norimas charakteristikas. Šios galimybės yra analogiškos toms, kurios buvo aptartos nagrinėjant vienmatę dispersinę analizę, todėl plačiau ties jomis neapsistosime.

Pavyzdys. Papildysime 15.3 skyriaus vienmatės dispersinės analizės pavyzdį antru priklausomu kintamuoju — vertikalaus šuolio aukščiu, cm (*vert_s*). Taigi, turėsime du priklausomus intervalinius kintamuosius *impuls* ir *vert_s* ir du nepriklausomus kintamuosius — binarinę *mokykla*, turintį reikšmes 1 ir 2 ir nominalinę *amzius*, turintį tris reikšmes (kategorijas) — 1, 2 ir 3.

- Dialogo langelyje *Multivariate* (15.18 pav.) priklausomus kintamuosius *impuls* ir *vert_s* įkeliamė į laukelį *Dependent Variables*, o nepriklausomus kintamuosius *mokykla* ir *amzius* — į laukelį *Fixed Factor(s)*.
- Spragtelėję mygtuką *Plots...*, dialogo langelyje *Multivariate: Profile Plots* nepriklausomą kintamąjį *amzius* įkeliamė į laukelį *Horizontal Axis*, o nepriklausomą kintamąjį *mokykla* — į laukelį *Separate Lines*.



15.18 pav. Dialogo langelis *Multivariate*

- Spragtelėję mygtuką *Post Hoc...* dialogo langelyje *Multivariate: Post Hoc Multiple Comparisons for Observed Means* įkeliame nepriklausomą kintamąjį *amzius* į laukelį *Post Hoc Test for* ir pasirenkame *Tukey'o* testą.
- Dialogo langelyje *Multivariate: Options* įkeliame *mokykla*amzius* į laukelį *Display Means for* ir pažymime laukelius *Descriptive statistics*, *Estimates of effect size*, *Parameter Estimates* ir *Homogeneity tests*.
- Nekeisdami kitų nustatymų spragtelime dialogo langelio *Multivariate* mygtuką *OK*.

Pagrindiniai skaičiavimo rezultatai parodyti 15.19.1 — 15.19.6 pav.

15.19.1 pav. lentelėje a) *Descriptive Statistics* (aprašomoji statistika) pateikiami priklausomų kintamųjų *impuls* ir *vert_s* vidurkiai ir standartiniai nuokrypiai visoms nepriklausomų kintamųjų reikšmėms.

Lentelėje b) *Box's Test of Equality of Covariance Matrices* pateikiami priklausomų kintamųjų kovariacijos matricų lygybės Box'o M testo rezultatai. Box'o M testas yra labai jautrus kovariacijos matricų lygybės pažeidimui, o taikomas dispersinės analizės metodas, priešingai, tam yra ganėtinai atsparus, todėl reikšmingumo lygmuo α dažnai pasirenkamas $\alpha = 0,001$. Jeigu kovariacijos matricų lygybės sąlyga yra nevykdoma, t. y. *p*-reikšmė (*Sig.*) yra

mažesnė už reikšmingumo lygmenį, pirmenybę reikia teikti 15.19.2 pav. a) lentelėje *Multivariate Tests* (daugiamačiai testai) *Pillai's Trace* (Pillai pėdsakas) testui, kuris yra galingiausias iš išvardytų lentelėje arba bandyti transformuoti duomenis.

Lentelėje c) *Levene's Test of Equality of Error Variances* (priklausomų kintamųjų nuokrypių dispersijų lygybės Levene testas) pateikiama *p*-reikšmė rodo, kad abiejų priklausomų kintamųjų dispersijos reikšmingai nesiskiria pagal nepriklausomų kintamųjų reikšmes.

Pagal 15.19.2 pav. a) lentelėje *Multivariate Tests* (daugiamačiai testai) pateiktų testų *Pillai's Trace* (Pillai pėdsakas), *Wilks' Lambda* (Wilks'o lambda), *Hotelling's Trace* (Hotelling'o pėdsakas) ir *Roy's Largest Root* (didžiausia būdingoji šaknis pagal Roy'o metodą) *p*-reikšmę (*Sig.*) galime padaryti bendrą išvadą, kad imtyse, sudarytose pagal faktoriaus *mokykla* ir pagal faktoriaus *amžius* reikšmes, kintamųjų vidurkiai statistiškai reikšmingai skiriasi. Tuo tarpu, šių faktorių tarpusavio sąveikos nėra. Iš visų anksčiau paminėtų testų *Pillai's Trace* testas yra galingiausias (Бююль, Цефелъ, 2002).

Konkretesnį atsakymą apie kiekvieno faktoriaus ir šių faktorių sąveikos poveikį kiekvienam priklausomam kintamajam duoda 15.19.2 pav. b) lentelėje *Tests of Between-Subject Effects* pateikti rezultatai. Analogiškai vienmatės dispersinės analizės atveju eilutėje *Corrected Model* nurodyta visa kvadratų suma be paklaidos. *Intercept* atitinka statistinę hipotezę apie bendrojo vidurkio lygybę nuliui. Mūsų atveju ši hipotezė neaktuali. Eilutės *MOKYKLA*, *AMŽIUS* IR *MOKYKLA*AMŽIUS* duoda atsakymą apie mokyklos įtaką, amžiaus įtaką ir mokyklos bei amžiaus sąveikos įtaką jaunųjų krepšininkų fizinio parengtumo rodikliams – atsispyrimo jėgos impulsui (*IMPULS*) ir vertikalaus šuolio aukščiui (*VERT_S*). Šias hipotezes atitinkančios *p*-reikšmės pateiktos stulpelyje *Sig.* Matyti, kad statistiškai reikšmingai skiriasi skirtingo amžiaus sportininkų abiejų rodiklių vidurkiai. Tuo tarpu pagal mokyklas šis skirtumas statistiškai reikšmingas tik vertikalaus šuolio aukščiui, o atsispyrimo jėgos impulso vidurkiai skiriasi nereikšmingai. Nėra tarpusavio sąveikos tarp mokyklos ir sportininkų amžiaus abiejų rodiklių. Taigi patikslinama bendra lentelėje *Multivariate Tests* pateikta išvada, kad kad imtyse, sudarytose pagal faktoriaus *mokykla* reikšmes, statistiškai reikšmingai skiriasi tik vertikalaus šuolio aukščio vidurkiai, o atsispyrimo jėgos impulso vidurkiai reikšmingai nesiskiria.

15.19.3 pav. lentelėje a) *MOKYKLA*AMŽIUS* yra pateikiami priklausomų kintamųjų vidurkiai visoms nepriklausomų kintamųjų kombinacijoms, standartinė paklaida ir pasikliautinis intervalas.

Descriptive Statistics

	MOKYKLA	AMZIUS	Mean	Std. Deviation	N
IMPULS	1	14 metų	293,20	22,881	20
		15 metų	330,15	30,020	20
		16 metų	352,60	23,761	20
		Total	325,32	35,363	60
	2	14 metų	306,70	15,928	20
		15 metų	335,35	26,162	20
		16 metų	356,60	23,623	20
		Total	332,88	30,116	60
	Total	14 metų	299,95	20,625	40
		15 metų	332,75	27,918	40
		16 metų	354,60	23,474	40
		Total	329,10	32,926	120
VERT_S	1	14 metų	381,25	19,114	20
		15 metų	391,90	27,543	20
		16 metų	397,80	22,185	20
		Total	390,32	23,836	60
	2	14 metų	392,50	20,669	20
		15 metų	401,75	21,136	20
		16 metų	413,65	22,869	20
		Total	402,63	22,936	60
	Total	14 metų	386,88	20,459	40
		15 metų	396,83	24,741	40
		16 metų	405,73	23,643	40
		Total	396,48	24,099	120

a)

Box's Test of Equality of Covariance Matrices^a

Box's M	101,881
F	6,489
df1	15
df2	71084,38
Sig.	,000

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept+MOKYKLA+AMZIUS+MOKYKLA * AMZIUS

b)

Levene's Test of Equality of Error Variances^a

	F	df1	df2	Sig.
IMPULS	1,104	5	114	,362
VERT_S	,622	5	114	,683

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept+MOKYKLA+AMZIUS+MOKYKLA * AMZIUS

c)

15.19.1 pav. Pagrindiniai daugiamečių kelių faktorių dispersinės analizės rezultatai

Multivariate Tests^c

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	,997	19408,62 ^a	2,000	113,000	,000	,997
	Wilks' Lambda	,003	19408,62 ^a	2,000	113,000	,000	,997
	Hotelling's Trace	343,515	19408,62 ^a	2,000	113,000	,000	,997
	Roy's Largest Root	343,515	19408,62 ^a	2,000	113,000	,000	,997
MOKYKLA	Pillai's Trace	,074	4,501 ^a	2,000	113,000	,013	,074
	Wilks' Lambda	,926	4,501 ^a	2,000	113,000	,013	,074
	Hotelling's Trace	,080	4,501 ^a	2,000	113,000	,013	,074
	Roy's Largest Root	,080	4,501 ^a	2,000	113,000	,013	,074
AMŽIUS	Pillai's Trace	,501	19,038	4,000	228,000	,000	,250
	Wilks' Lambda	,500	23,414 ^a	4,000	226,000	,000	,293
	Hotelling's Trace	,999	27,981	4,000	224,000	,000	,333
	Roy's Largest Root	,998	56,891 ^b	2,000	114,000	,000	,500
MOKYKLA * AMŽIUS	Pillai's Trace	,022	,633	4,000	228,000	,639	,011
	Wilks' Lambda	,978	,630 ^a	4,000	226,000	,642	,011
	Hotelling's Trace	,022	,627	4,000	224,000	,644	,011
	Roy's Largest Root	,021	1,174 ^b	2,000	114,000	,313	,020

a. Exact statistic

b. The statistic is an upper bound on F that yields a lower bound on the significance level.

c. Design: Intercept+MOKYKLA+AMŽIUS+MOKYKLA * AMŽIUS

a)

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	IMPULS	62784,700 ^a	5	12556,940	21,616	,000	,487
	VERT_S	11861,875 ^b	5	2372,375	4,724	,001	,172
Intercept	IMPULS	12996817,200	1	12996817,2	22373,08	,000	,995
	VERT_S	18863091,075	1	18863091,1	37562,72	,000	,997
MOKYKLA	IMPULS	1717,633	1	1717,633	2,957	,088	,025
	VERT_S	4551,008	1	4551,008	9,063	,003	,074
AMŽIUS	IMPULS	60531,800	2	30265,900	52,101	,000	,478
	VERT_S	7113,800	2	3556,900	7,083	,001	,111
MOKYKLA * AMŽIUS	IMPULS	535,267	2	267,633	,461	,632	,008
	VERT_S	197,067	2	98,533	,196	,822	,003
Error	IMPULS	66224,100	114	580,913			
	VERT_S	57248,050	114	502,176			
Total	IMPULS	13125826,000	120				
	VERT_S	18932201,000	120				
Corrected Total	IMPULS	129008,800	119				
	VERT_S	69109,925	119				

a. R Squared = ,487 (Adjusted R Squared = ,464)

b. R Squared = ,172 (Adjusted R Squared = ,135)

b)

15.19.2 pav. Pagrindiniai daugiamečių kelių faktorių dispersinės analizės rezultatai

15.19.3 pav. lentelėje b) *Multiple Comparison* pateikti aposteriorinių (Post Hoc) **Tukey'o** testų rezultatai. Iš *p*-reikšmės (Sig.) matyti, kad visų amžiaus grupių kintamojo *IMPULS* vidurkiai statistiškai reikšmingai skiriasi. Tuo tarpu kintamojo *VERT_S* reikšmingai skiriasi tik 14 ir 16 m. amžiaus grupių

vidurkiai. Tai atspindi ir 15.19.4 pav. a) ir b) lentelėse *Homogenous Subsets* (homogeniniai pogrupiai) pateikti rezultatai — kintamajam *IMPULS* nėra vienašalių pogrupių, kuriuose vidurkiai nesiskirtų, o kintamajam *VERT_S* yra du vienašaliai pogrupiai.

MOKYKLA * AMŽIUS

Dependent Variable	MOKYKLA	AMŽIUS	Mean	Std. Error	95% Confidence Interval	
					Lower Bound	Upper Bound
IMPULS	1	14 metų	293,200	5,389	282,524	303,876
		15 metų	330,150	5,389	319,474	340,826
		16 metų	352,600	5,389	341,924	363,276
	2	14 metų	306,700	5,389	296,024	317,376
		15 metų	335,350	5,389	324,674	346,026
		16 metų	356,600	5,389	345,924	367,276
VERT_S	1	14 metų	381,250	5,011	371,324	391,176
		15 metų	391,900	5,011	381,974	401,826
		16 metų	397,800	5,011	387,874	407,726
	2	14 metų	392,500	5,011	382,574	402,426
		15 metų	401,750	5,011	391,824	411,676
		16 metų	413,650	5,011	403,724	423,576

a)

Post Hoc Tests

Multiple Comparisons

Tukey HSD

Dependent Variable	(I) AMŽIUS	(J) AMŽIUS	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
IMPULS	14 metų	15 metų	-32,80*	5,389	,000	-45,60	-20,00
		16 metų	-54,65*	5,389	,000	-67,45	-41,85
	15 metų	14 metų	32,80*	5,389	,000	20,00	45,60
		16 metų	-21,85*	5,389	,000	-34,65	-9,05
	16 metų	14 metų	54,65*	5,389	,000	41,85	67,45
		15 metų	21,85*	5,389	,000	9,05	34,65
VERT_S	14 metų	15 metų	-9,95	5,011	,120	-21,85	1,95
		16 metų	-18,85*	5,011	,001	-30,75	-6,95
	15 metų	14 metų	9,95	5,011	,120	-1,95	21,85
		16 metų	-8,90	5,011	,182	-20,80	3,00
	16 metų	14 metų	18,85*	5,011	,001	6,95	30,75
		15 metų	8,90	5,011	,182	-3,00	20,80

Based on observed means.

*. The mean difference is significant at the ,05 level.

b)

15.19.3 pav. Pagrindiniai daugiamečių kelių faktorių dispersinės analizės rezultatai

Homogeneous Subsets

IMPULS

Tukey HSD^{a,b,c}

AMŽIUS	N	Subset		
		1	2	3
14 metų	40	299,95		
15 metų	40		332,75	
16 metų	40			354,60
Sig.		1,000	1,000	1,000

Means for groups in homogeneous subsets are displayed.

Based on Type III Sum of Squares

The error term is Mean Square(Error) = 580,913.

- Uses Harmonic Mean Sample Size = 40,000.
- The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.
- Alpha = ,05.

a)

VERT_S

Tukey HSD^{a,b,c}

AMŽIUS	N	Subset	
		1	2
14 metų	40	386,88	
15 metų	40	396,83	396,83
16 metų	40		405,73
Sig.		,120	,182

Means for groups in homogeneous subsets are displayed.

Based on Type III Sum of Squares

The error term is Mean Square(Error) = 502,176.

- Uses Harmonic Mean Sample Size = 40,000.
- The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.
- Alpha = ,05.

b)

15.19.4 pav. Pagrindiniai daugiameitės kelių faktorių dispersinės analizės rezultatai

15.19.5 pav. a) lentelėje *Parameter Estimates* pateiktos regresinio modelio, nusakančio priklausomų kintamųjų *IMPULS* ir *VERT_S* priklausomybę nuo veikiančių faktorių ir šių faktorių sąveikos, koeficientų reikšmės bei jų reikšmingumas. Statistiškai reikšmingi yra koeficientai, apibrėžiantys priklausomo kintamojo *IMPULS* priklausomybę nuo sportininkų amžiaus ir priklausomo kintamojo *VERT_S* priklausomybę nuo sportininkų amžiaus ir

lankomos mokyklos. Koeficientai, apibrėžiantys abiejų priklausomų kintamųjų priklausomybę nuo mokyklos ir sportininko amžiaus sąveikos, nėra statistiškai reikšmingi.

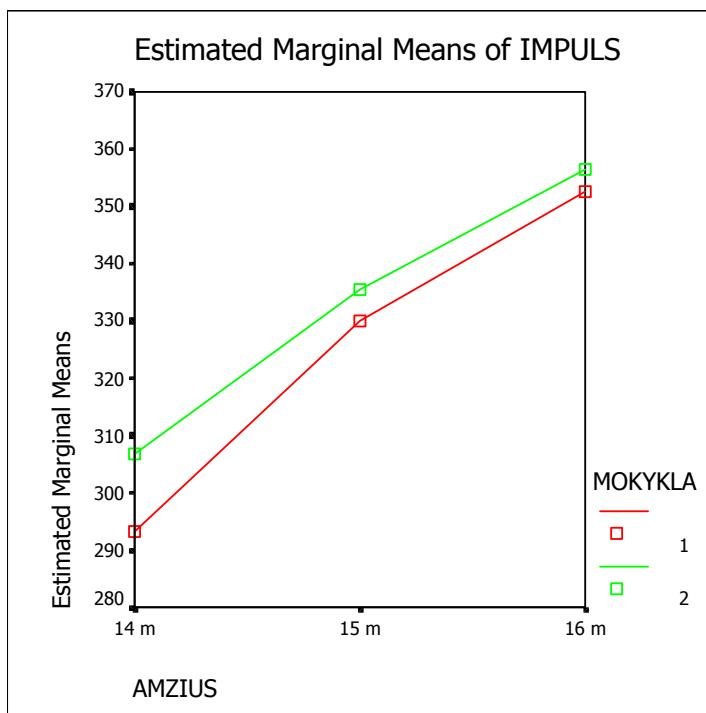
15.19.6 pav. diagramose *Estimated Marginal Means of IMPULS* ir *Estimated Marginal Means of VERT_S* parodyti priklausomų kintamųjų vidurkių trendai — vidurkių priklausomybė nuo faktoriaus *amžius* reikšmių. Dvi kreivės kiekviename grafike atitinka dvi binarinio faktoriaus *mokykla* reikšmes. Iš grafiko galima padaryti išvadą, kad nėra veikiančių faktorių sąveikos — kreivės praktiškai yra identiškos formos ir skiriasi tik postūmiu.

Parameter Estimates

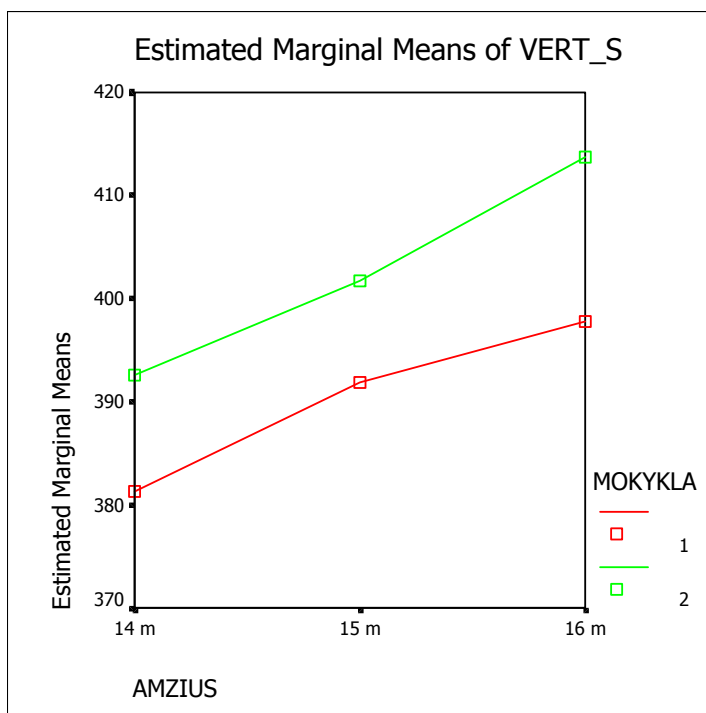
Dependent Variable	Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
IMPULS	Intercept	356,600	5,389	66,167	,000	345,924	367,276
	[MOKYKLA=1]	-4,000	7,622	-,525	,601	-19,099	11,099
	[MOKYKLA=2]	0 ^a	,	,	,	,	,
	[AMŽIUS=1]	-49,900	7,622	-6,547	,000	-64,999	-34,801
	[AMŽIUS=2]	-21,250	7,622	-2,788	,006	-36,349	-6,151
	[AMŽIUS=3]	0 ^a	,	,	,	,	,
	[MOKYKLA=1] *	-9,500	10,779	-,881	,380	-30,853	11,853
	[AMŽIUS=1]						
	[MOKYKLA=1] *	-1,200	10,779	-,111	,912	-22,553	20,153
	[AMŽIUS=2]						
	[MOKYKLA=1] *	0 ^a	,	,	,	,	,
	[AMŽIUS=3]						
	[MOKYKLA=2] *	0 ^a	,	,	,	,	,
	[AMŽIUS=1]						
	[MOKYKLA=2] *	0 ^a	,	,	,	,	,
	[AMŽIUS=2]						
	[MOKYKLA=2] *	0 ^a	,	,	,	,	,
	[AMŽIUS=3]						
VERT_S	Intercept	413,650	5,011	82,551	,000	403,724	423,576
	[MOKYKLA=1]	-15,850	7,086	-2,237	,027	-29,888	-1,812
	[MOKYKLA=2]	0 ^a	,	,	,	,	,
	[AMŽIUS=1]	-21,150	7,086	-2,985	,003	-35,188	-7,112
	[AMŽIUS=2]	-11,900	7,086	-1,679	,096	-25,938	2,138
	[AMŽIUS=3]	0 ^a	,	,	,	,	,
	[MOKYKLA=1] *	4,600	10,022	,459	,647	-15,253	24,453
	[AMŽIUS=1]						
	[MOKYKLA=1] *	6,000	10,022	,599	,551	-13,853	25,853
	[AMŽIUS=2]						
	[MOKYKLA=1] *	0 ^a	,	,	,	,	,
	[AMŽIUS=3]						
	[MOKYKLA=2] *	0 ^a	,	,	,	,	,
	[AMŽIUS=1]						
	[MOKYKLA=2] *	0 ^a	,	,	,	,	,
	[AMŽIUS=2]						
	[MOKYKLA=2] *	0 ^a	,	,	,	,	,
	[AMŽIUS=3]						

a. This parameter is set to zero because it is redundant.

15.19.5 pav. Pagrindiniai daugiamečių kelių faktorių dispersinės analizės rezultatai



a)



b)

15.19.6 pav. Pagrindiniai daugiamačės kelių faktorių dispersinės analizės rezultatai

15.5. BLOKUOTŲJŲ DUOMENŲ DVIEJŲ FAKTORIŲ DISPERSINĖ ANALIZĖ

Blokuotųjų duomenų struktūra gali būti daug sudėtingesnė nei aptartoji 15.2 skyriuje. Naudodamiesi SPSS paketo blokuotųjų duomenų analizės metodu (*GLM Repeated Measures*) galite spręsti ir šiuos sudėtingesnius uždavinius — ištirti priklausymo tam tikrai grupei (*between-subjects*) ir matavimo sąlygų (*within-subjects*) faktorių įtaką.

Blokuotųjų duomenų dispersinės analizės metodu taip pat gali būti tiriami keli priklausomi kintamieji, sąlygojami grupės ir sąlygų faktorių bei jų sąveikos. Pavyzdžiui, n kartų tam tikrais laiko intervalais matuojamas kelių grupių individų širdies susitraukimų ir kvėpavimo dažnis. Tokiais atvejais faktiškai turime daugiamatę blokuotųjų duomenų dispersinę analizę.

Blokuotųjų duomenų struktūrą, statistinių hipotezių formulavimą išsamiau paaiškinsime remdamiesi dažnai naudojamu dviejų faktorių dispersinės analizės modeliu. Tarkime, turime a nepriklausomų grupių po n tiriamųjų, kurių kiekvienas buvo pakartotinai matuotas (testuotas) po b kartų vis kitomis sąlygomis. Situacija primena dviejų faktorių dispersinę analizę, tačiau ne visi kintamieji yra nepriklausomi. Kiekvieno tiriamojo duomenys sudaro duomenų bloką, o iš viso blokų yra tiek, kiek ir tiriamųjų — $a \times n$. Blokai nepriklausomi, tačiau duomenys bloko viduje gauti matuojant priklausomus kintamuosius. Skirtingai nuo vieno faktoriaus blokuotųjų duomenų analizės, čia yra ne viena duomenų blokų grupė, o a grupių, suskirstytų pagal požymį A .

Blokuotųjų duomenų dviejų faktorių dispersinės analizės prielaidos:

- Be aptartos duomenų priklausomybės, visų kintamųjų skirstinys turi būti normalus, t. y. tenkinama normalumo prielaida.
- Visų a nepriklausomų duomenų grupių dispersijos turi būti lygios. Kiekvienai iš šių grupių turi galioti sferiškumo prielaida.
- Visų grupių apskaičiuotos stulpelių (t. y. atskirų matavimų) kovariacijos matricos turi sutapti.

Kaip ir kitais dispersinės analizės atvejais, galimi nedideli reikalavimų nukrypimai.

Kaip ir nepriklausomų duomenų dviejų faktorių dispersinės analizės atveju formuluojamos trys statistinės hipotezės: apie grupės (faktoriaus A) įtaką rezultatams, apie sąlygų (faktoriaus B) įtaką ir apie faktorių A ir B tarpusavio sąveikos poveikį. Intervalų skalėje matuojamiems nepriklausomiems vienodai pasiskirsčiusiems kintamiesiems $X_{ij} \sim N(\mu_{ij}, \sigma^2)$, $i = 1, \dots, a, j = 1, \dots, b$, kurių vidurkliai μ_{ij} ir dispersijos σ^2 nežinomi, galima tikrinti šias hipotezes

$$\begin{cases} H_0: \mu_1. = \mu_2. = \dots = \mu_{a.}, \\ H_1: \text{ bent du vidurkliai skiriasi;} \end{cases} \quad (15.17)$$

čia nulinė hipotezė H_0 reiškia, kad visų grupių vidutiniai matavimų rezultatai vienodi;

$$\begin{cases} H_0: \mu_{.1} = \mu_{.2} = \dots = \mu_{.b}, \\ H_1: \text{ bent du vidurkiai skiriasi}; \end{cases} \quad (15.18)$$

čia nulinė hipotezė H_0 reiškia, kad visomis sąlygomis pakartotinai matuojant vidutiniai matavimų rezultatai nesiskiria;

$$\begin{cases} H_0: A \text{ ir } B \text{ tarpusavio sąveikos nėra}, \\ H_1: A \text{ ir } B \text{ tarpusavio sąveika yra}; \end{cases} \quad (15.19)$$

$$\text{čia } \mu_{i.} = \frac{\sum_{j=1}^b \mu_{ij}}{b}, \quad \mu_{.j} = \frac{\sum_{i=1}^a \mu_{ij}}{a};$$

Kiekvienai statistinei hipotezei galioja ta pati sprendimo taisyklė. Tegul reikšmingumo lygmuo lygus α , o p -reikšmė yra p . Tuomet nulinę hipotezę H_0 atmetame, jeigu $p < \alpha$; nulinę hipotezę H_0 priimame, jeigu $p \geq \alpha$. Svarbu tik neužmiršti, ką kokia nulinė hipotezė reiškia: jeigu grupės faktorių A atitinkanti p -reikšmė $p < \alpha$, tada grupių vidurkiai statistiškai reikšmingai skiriasi; jeigu sąlygų faktorių B atitinkanti p -reikšmė $p < \alpha$, tada skirtingomis sąlygomis gautų rezultatų vidurkiai statistiškai reikšmingai skiriasi; jeigu p -reikšmė, atitinkanti faktorių A ir B tarpusavio sąveiką, $p < \alpha$, tada nusprendžiame, kad faktorių tarpusavio sąveika yra, priešingu atveju — nėra.

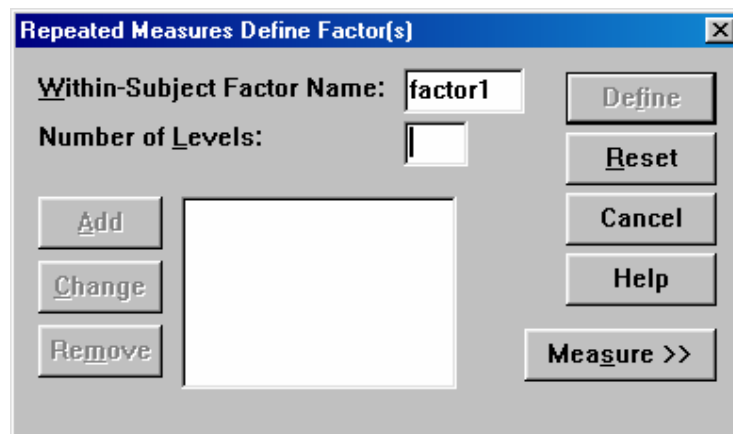
Yra realizuoti ir *Post Hoc* kriterijai, kurie interpretuojami kaip ir kitais dispersinės analizės atvejais, t. y. dviejų grupių rezultatų vidurkiai statistiškai reikšmingai skiriasi, jeigu atitinkama p -reikšmė yra mažesnė už pasirinktą reikšmingumo lygmenį α . Skaičiuojami koeficientai η^2 ir tikrinama sferiškumo prielaida. Kaip ir blokuotųjų duomenų vieno faktoriaus dispersinės analizės atveju, jeigu sferiškumo prielaida netenkinama, reikia naudotis viena iš trijų SPSS pakete siūlomų pataisų.

Grupės faktoriai (*between-subjects factor*) gali būti skaitmeniniai arba raidinės-skaitmeninės sekos formato. Matavimo sąlygų (tarbandyminiai) faktoriai nurodomi dialogo langelyje **Repeated Measures Define Factor(s)**. Intervaliniai nepriklausomi kintamieji (*Covariates*) turi būti pastovūs kiekvienai matavimo sąlygų (tarbandyminių) faktorių reikšmei. Duomenų rinkmenoje kiekvieno parametro (priklausomo kintamojo) kartotinių matavimų duomenys sudaro atskirus kintamuosius. Taigi, matavimo sąlygų (tarbandyminio) faktoriaus (*within-subject factor*) reikšmių skaičius yra lygus kartotinių bandymų skaičiui. Esant keletui tarbandyminių faktorių, kartotinių bandymų skaičius bus lygus tarbandyminių faktorių reikšmių sandaugai. Pavyzdžiui, matavimai atliekami penkias dienas tris kartus per dieną. Bendras matavimų

skaičius bus 12, o tarpbandyminiai faktoriai bus *diena* (keturios reikšmės) ir *kartas* (3 reikšmės).

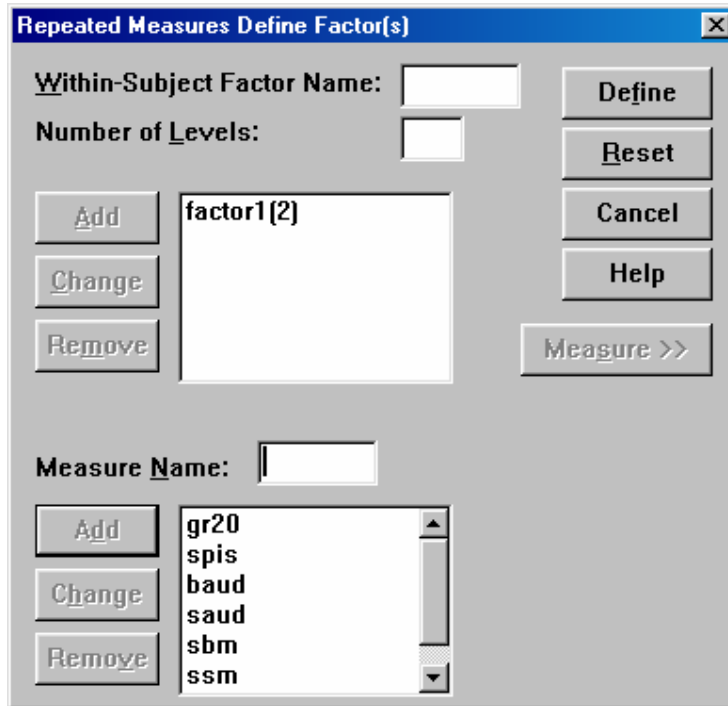
Daugiamatei blokuotųjų duomenų dispersinei analizei atlikti:

- Atidarykite bylą su analizuojamais duomenimis.
- Nurodykite komandas **Analyze → General Linear Model → Repeated Measures...**
- Atsidariusiame dialogo langelyje **Repeated Measures Define Factor(s)** (15.20 pav.) galite nurodyti vieną ar daugiau tarpbandyminių (*within-subjects*) faktorių. Svarbus yra šių faktorių apibrėžimo eiliškumas — pirmiausia nurodomas stambiausio laiko mastelio faktorius (pvz., *diena*), toliau — tie faktoriai, kurie sudaro prieš tai esančio faktoriaus reikšmės smulkesnę dalį (pvz., *kartas per dieną* ar *valanda* ir t. t.). Reikia taip pat pažymėti, kad tarpbandyminiai faktoriai nėra duomenų rinkmenos kintamieji, o Jūsų nustatyti šiame dialogo langelyje. Tarpbandyminio faktoriaus pavadinimą įrašykite laukelyje **Within- Subject Factor Name**, o laukelyje **Number of Levels** nurodykite faktoriaus reikšmių skaičių. Spragtelėkite mygtuką **Add**. Šiuos veiksmus kartokite, kol nustatysite visus faktorius.



15.20 pav. Dialogo langelis **Repeated Measures Define Factor(s)**

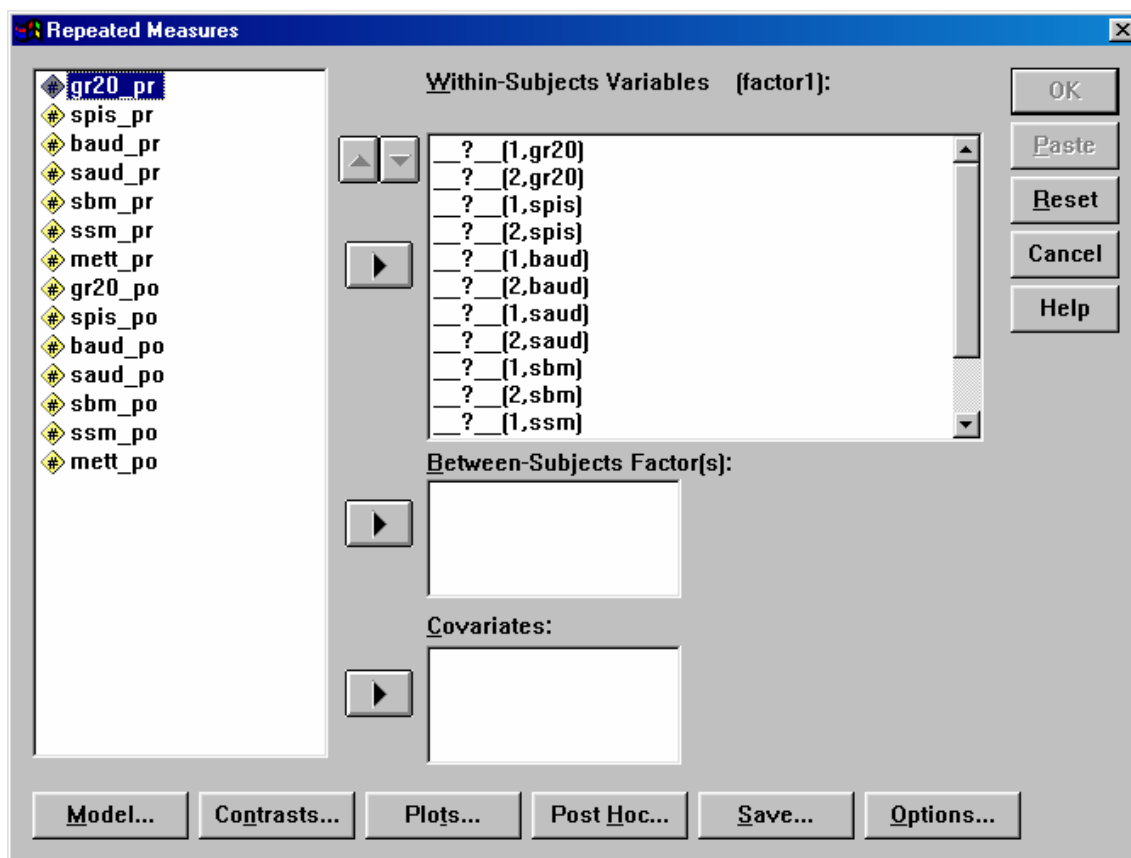
- Jeigu vienu metu buvo matuojami keli parametrai (kaip mūsų anksčiau pateiktame pavyzdyje — tam tikrais laiko intervalais matuojamas tų pačių individų širdies susitraukimų ir kvėpavimo dažnis), spragtelėkite mygtuką **Measure** (matavimai). Išplėstinio dialogo langelio **Repeated Measures Define Factor(s)** (15.21 pav.) laukelyje **Measure Name** nurodykite matuojamo parametro (priklausomo kintamojo) pavadinimą ir spragtelėkite mygtuką **Add**. Šiuos veiksmus kartokite tol, kol nurodysite visus matuojamus parametrus.



15.21 pav. Išplėstinis dialogo langelis *Repeated Measures Define Factor(s)*

- Dialogo langelyje *Repeated Measures* (15.22 pav.) pažymėkite priklausomą kintamąjį, kuris atitinka pirmąją sąrašo *Within – Subjects Variables* matuojamo parametro ir tarpbandyminio faktoriaus kombinaciją ir įkelkite šį kintamąjį į pirmąją sąrašo *Within – Subjects Variables* poziciją. Šį veiksmą kartokite tol, kol į reikiamas sąrašo *Within – Subjects Variables* pozicijas įkelsite visus priklausomus kintamuosius.
- Norėdami pakeisti sąrašo *Within – Subjects Variables* kintamųjų eiliškumą, naudokitės mygtukais „žemyn“, „aukštyn“.
- Įkelkite (jeigu tai numatyta tyrimo metodikoje) grupės faktorius į laukelį *Between – Subjects Factor(s)*, o intervalinius faktorius — į laukelį *Covariates*.

Mygtukais *Model...*, *Contrasts...*, *Plots...*, *Post Hoc...*, *Save...*, *Options...* atidarius atitinkamus dialogo langelius galima nurodyti įvairius skaičiavimo variantus ir pasirinkti norimas charakteristikas. Šios galimybės yra analogiškos toms, kurios buvo aptartos nagrinėjant vieno faktoriaus blokuotųjų duomenų dispersinę analizę, todėl pateiktame pavyzdyje atkreipsime dėmesį tik į tas charakteristikas, kurios būdingos kelių faktorių blokuotųjų duomenų dispersinei analizei.



15.22 pav. Dialogo langelis **Repeated Measures**

Pavyzdys. Jaunieji krepšininkai buvo testuojami du kartus — prieš treniruočių ciklą ir po jo. Testavimo metodika paremta 7 testais (toliau nurodomi taip pat priklausomų kintamųjų — matuojamų parametų pavadinimai duomenų rinkmenoje):

- 20 m bėgimas — *gr20*.
- Specialioji ištvermė — *spis*.
- Baudos metimai — *baud*.
- Bėgimas šaudykle — *saud*.
- Šuolis nemojant — *sbm*.
- Šuolis mojant — *ssm*.
- Metimo tikslumas — *mett*.

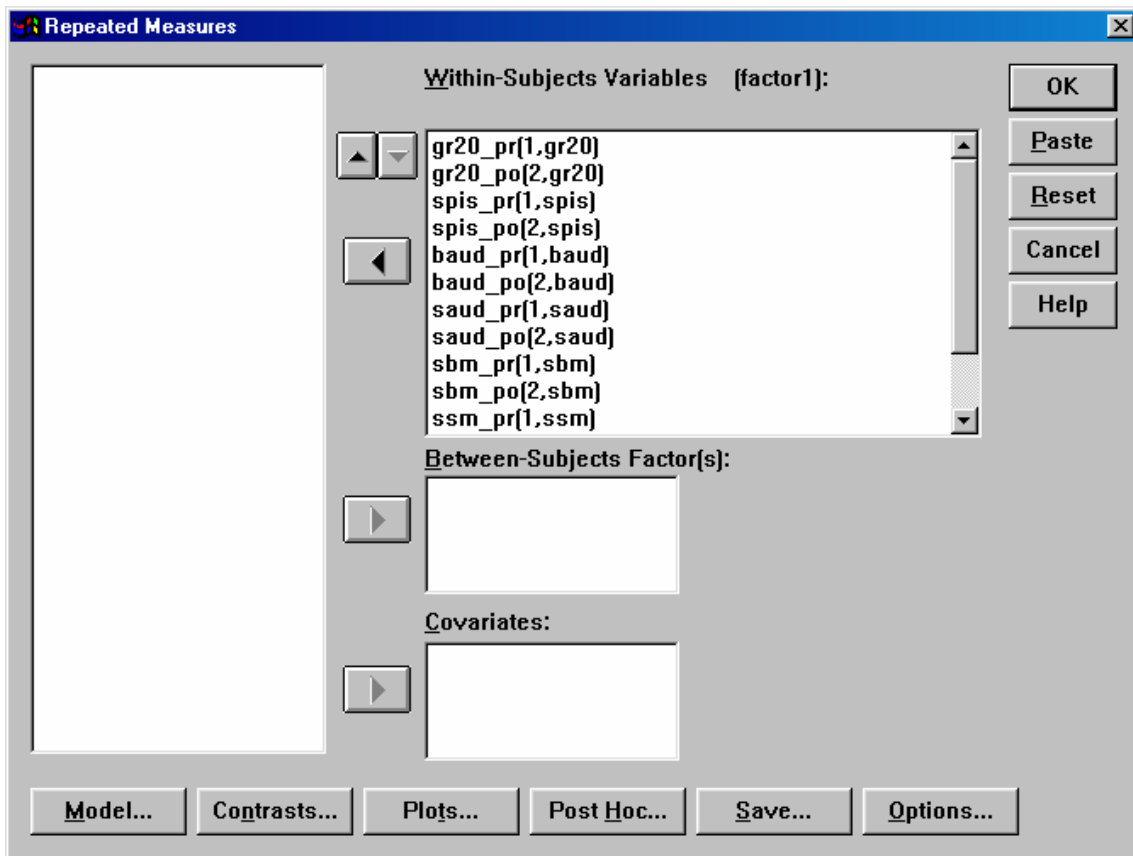
Duomenų rinkmenoje (15.23 pav.) kintamieji, kurių duomenys yra testavimo rezultatai prieš treniruočių ciklą, turi priesagą *_pr*, o kintamieji, kurių duomenys yra testavimo rezultatai po treniruočių ciklo, turi priesagą *_po*.

	gr20_pr	spis_pr	baud_pr	saud_pr	sbrn_pr	ssm_pr	mett_pr	gr20_po	spis_po	baud_po	saud_po	sbrn_po	ssm_po
1	3,35	16,00	28,00	44,30	43,80	52,70	7,00	3,53	18,00	25,00	43,70	44,70	52,00
2	3,63	16,00	20,00	48,60	40,50	47,10	9,00	3,75	17,00	24,00	47,15	45,80	52,00
3	3,75	15,00	20,00	46,40	41,60	47,50	7,00	3,58	15,00	19,00	46,30	44,20	50,10
4	3,41	17,00	19,00	46,00	36,20	40,40	9,00	3,34	15,00	22,00	45,60	39,20	49,00
5	3,57	17,00	18,00	49,00	34,90	39,70	13,00	3,53	17,00	26,00	44,50	32,20	38,30
6	3,35	16,00	23,00	45,00	41,10	47,50	5,00	3,75	16,00	24,00	44,80	40,00	50,60
7	3,63	16,00	25,00	48,20	42,80	51,30	8,00	3,58	17,00	24,00	47,40	41,20	51,40
8	3,75	16,00	23,00	47,10	41,10	49,70	6,00	3,34	16,00	24,00	46,90	42,50	53,40
9	3,41	16,00	20,00	44,60	43,80	52,70	9,00	3,53	18,00	24,00	45,30	44,10	51,70
10	3,57	16,00	14,00	47,00	40,50	47,10	5,00	3,75	17,00	26,00	44,78	37,20	46,30
11	3,35	15,00	21,00	48,40	41,60	47,50	7,00	3,58	17,00	19,00	45,90	44,70	52,00
12	3,63	17,00	18,00	47,20	36,20	40,40	7,00	3,34	18,00	25,00	45,10	45,80	52,00
13	3,75	17,00	28,00	44,30	34,90	39,70	9,00	3,53	17,00	24,00	43,70	44,20	50,10
14	3,41	16,00	20,00	48,60	41,10	47,50	7,00	3,75	15,00	19,00	47,15	39,20	49,00
15	3,57	16,00	20,00	46,40	42,80	51,30	9,00	3,58	15,00	22,00	46,30	32,20	38,30
16	,	16,00	19,00	46,00	41,10	49,70	13,00	,	17,00	26,00	45,60	40,00	50,60
17	,	16,00	18,00	49,00	,	,	5,00	,	16,00	24,00	44,50	41,20	51,40
18	,	16,00	23,00	45,00	,	,	8,00	,	17,00	24,00	44,80	42,50	53,40
19	,	15,00	25,00	48,20	,	,	6,00	,	16,00	24,00	47,40	44,10	51,70
20	,	17,00	23,00	47,10	,	,	9,00	,	18,00	24,00	46,90	37,20	46,30
21	,	17,00	20,00	44,60	,	,	,	,	17,00	26,00	45,30	,	,
22	,	16,00	14,00	47,00	,	,	,	,	17,00	19,00	44,78	,	,
23	,	,	,	,	,	,	,	,	,	,	,	,	,
24													

15.23 pav. Testavimo duomenų rinkmena

Kadangi matuojamieji parametrai yra daugiau ar mažiau priklausomi, treniruočių ciklo įtakos reikšmingumą atskiriems krepšinininkų rodikliams įvertinsime naudodamiesi daugiamatės blokuotųjų duomenų dispersinės analizės metodu.

- Dialogo langelyje **Repeated Measures Define Factor(s)** (15.20 pav.) paliekame siūlomą tarpbandyminio faktoriaus pavadinimą *factor 1*, o laukelyje **Number of Levels** nurodome šio faktoriaus reikšmių (kategorijų) skaičių — 2. Taigi, faktoriaus reikšmė 1 reikš prieš treniruočių ciklą gautus testavimo rezultatus, o faktoriaus reikšmė 2 — po treniruočių ciklo gautus testavimo rezultatus. Spragtelime mygtuką **Add**.
- Spragtelime mygtuką **Measure**, išplėstinio dialogo langelio **Repeated Measures Define Factor(s)** (15.21 pav.) laukelyje **Measure Name** įrašome pirmo testuojamo parametro pavadinimą *gr20* ir spragtelime šalia esantį mygtuką **Add**. Tokiu būdu nurodome visus testuojamus parametrus (15.21 pav. parodytas užpildytas dialogo langelis).
- Spragtelime mygtuką **Define** ir dialogo langelyje **Repeated Measures** (15.22 pav.) pagal sąrašą **Within – Subjects Variables (factor 1)** nurodytas testuojamo parametro ir testo eilės numerio kombinacijas įkeliam į šio sąrašo klaustukais nurodytas vietas atitinkamus priklausomus kintamuosius (15.24 pav.).



15.24 pav. Užpildytas dialogo langelio *Repeated Measures* sąrašas *Within – Subjects Variables (factor 1)*

- Kadangi šiame pavyzdyje nėra grupės faktorių (toks faktorius galėtų būti, pvz., skirtingos jaunųjų krepšininkų mokyklos) ir intervalinių nepriklausomų kintamųjų, dialogo langelio *Repeated Measures* laukelius *Between- Subjects Factor(s)* ir *Covariates* paliekame neužpildytus.
- Aposteriorinius (*Post Hoc*) testus blokuotųjų duomenų dispersinės analizės atveju dialogo langelyje *Post Hoc...* galima pasirinkti tik tada, kai yra nurodytas ne mažiau kaip tris kategorijas turintis vienas ar keli grupės faktoriai (*Between-Subject Factors*). Jei šių faktorių nėra (kaip šiame pavyzdyje) *Post Hoc...* dialogo langelis *Multivariate: Post Hoc Multiple Comparisons for Observed Means* būna neaktyvus, o kai kuriuos *Post Hoc* kriterijus poriniams palyginimams pagal sąlygų faktoriaus (*Within-Subject Factor*) reikšmes galima nurodyti *Options...* dialogo langelyje (sąlygų faktorius turi turėti ne mažiau kaip tris reikšmes, jeigu turės tik dvi — gausime tuos pačius rezultatus, kurie pateikiami lentelėje *Univariate Tests*).
- Dialogo langelyje *Repeated Measures: Options* įkeliamo *factor 1* iš sąrašo *Factor(s) and Factor Interactions* į laukelį *Display Means for*. Komandų grupėje *Display* pažymime *Descriptive statistics, Estimates of*

effect size. Dispersijų homogeniškumo testą **Homogeneity tests**, kuris leidžia pateikti Levene testo rezultatus ir Box'o M kovariacijos matricių lygybės testo rezultatus visiems priklausomiems kintamiesiems pagal visas grupės faktorių reikšmių kombinacijas, galima pasirinkti tik tada, kai yra nurodyti grupės faktoriai, jei šių faktorių nėra (kaip šiame pavyzdyje) testo laukelis būna neaktyvus, o rezultatų išvestyje automatiškai bus pateikti Mauchly sferiškumo testo rezultatai.

- Dialogo langelyje **Repeated Measures: Contrasts** paliekame nustatytą **Polynomial** variantą. Nekeisdami kitų nustatymų spragtelime dialogo langelio **Repeated Measures** mygtuką **OK**.

Pagrindiniai skaičiavimo rezultatai parodyti 15.25.1 — 15.25.4 pav.

15.25.1 pav. a) lentelėje *Within-Subjects Factors* (tarpbandyminiai faktoriai) pateikiama matuojamų parametrų (*Measures*), tarpbandyminio faktoriaus reikšmių ir visų priklausomų kintamųjų (*Dependent Variables*) suvestinė.

Lentelėje b) *Descriptive Statistics* (aprašomoji statistika) pateikiami visų priklausomų kintamųjų vidurkiai, standartiniai nuokrypiai ir stebėjimų skaičius.

Iš 15.25.2 pav. a) lentelėje *Multivariate Tests* (daugiamačiai testai) pateiktų testų *Pillai's Trace* (Pillai pėdsakas), *Wilks' Lambda* (Wilks'o lambda), *Hotelling's Trace* (Hotelling'o pėdsakas) ir *Roy's Largest Root* (didžiausia būdingoji šaknis pagal Roy'o metodą) *p*-reikšmės (*Sig.*) galime padaryti bendrą išvadą, kad imtyse, sudarytose pagal tarpbandyminio (*Within-Subjects*) faktoriaus *Factor 1* reikšmes, kintamųjų vidurkiai statistiškai reikšmingai nesiskiria (*Sig.*=0,224), t. y. tarpbandyminio faktoriaus bendra įtaka nėra reikšminga. Kaip minėjome anksčiau, konservatyviausiu yra laikomas *Pillai's Trace* kriterijus.

Sferiškumo sąlyga pagal Mauchly testą (b) lentelė *Mauchly's test of Sphericity*) visada tenkinama, kai yra atliekami tik du bandymai.

Lentelėje c) *Multivariate (Test of Within Subjects Effects)* pakartojami daugiamačio testo rezultatai.

15.25.3 pav. lentelėje *Univariate Tests* (vienmačiai testai) pateikiami testų tarpbandyminio faktoriaus įtakai atskiriems parametrams įvertinti rezultatai. Kadangi sferiškumo sąlyga yra tenkinama galima vadovautis testo *Sphericity Assumed* (sferiškumo prielaida tenkinama) rezultatais. Matyti, kad tarpbandyminio faktoriaus įtaka yra reikšminga parametrai *saud* (bėgimas šaudykle) — *p*-reikšmė $p = 0,006$ ir ties reikšmingumo riba — parametrai *baud* (baudos mėtymas) — *p*-reikšmė $p = 0,103$, o kitiems parametrams yra nereikšminga.

15.25.4 pav. a) lentelėje *Test of Within-Subject Contrasts* pateikti hipotezės, kad tarpbandyminio faktoriaus įtaką atskiriems parametrams galima išreikšti atitinkamo laipsnio daugianariu, tikrinimo rezultatai. Kadangi šiame pavyzdyje yra tik du bandymai, tikrinama tik pirmojo laipsnio (*linear*)

daugianario hipotezė, o šios lentelės duomenys atitinka lentelės *Univariate Tests* duomenis, t. y. gauname hipotezę apie dviejų vidurkių lygybę.

Within-Subjects Factors

Measure	FACTOR1	Dependent Variable
GR20	1	GR20_PR
	2	GR20_PO
SPIS	1	SPIS_PR
	2	SPIS_PO
BAUD	1	BAUD_PR
	2	BAUD_PO
SAUD	1	SAUD_PR
	2	SAUD_PO
SBM	1	SBM_PR
	2	SBM_PO
SSM	1	SSM_PR
	2	SSM_PO
METT	1	METT_PR
	2	METT_PO

a)

Descriptive Statistics

	Mean	Std. Deviation	N
GR20_PR	3,5420	,15077	15
GR20_PO	3,5640	,14505	15
SPIS_PR	16,1333	,63994	15
SPIS_PO	16,5333	1,12546	15
BAUD_PR	21,1333	3,75817	15
BAUD_PO	23,1333	2,41622	15
SAUD_PR	46,7400	1,63916	15
SAUD_PO	45,6387	1,22080	15
SBM_PR	40,1933	3,09550	15
SBM_PO	41,1467	4,49060	15
SSM_PR	46,8067	4,65533	15
SSM_PO	49,0800	4,70231	15
METT_PR	7,8000	2,00713	15
METT_PO	8,0667	2,15362	15

b)

15.25.1 pav. Pagrindiniai daugiamačių blokuotųjų duomenų dispersinės analizės rezultatai

Multivariate Tests^b

Effect			Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Between Subjects	Intercept	Pillai's Trace	1,000	5846,723 ^a	7,000	8,000	,000	1,000
		Wilks' Lambda	,000	5846,723 ^a	7,000	8,000	,000	1,000
		Hotelling's Trace	5115,882	5846,723 ^a	7,000	8,000	,000	1,000
		Roy's Largest Root	5115,882	5846,723 ^a	7,000	8,000	,000	1,000
Within Subjects	FACTOR1	Pillai's Trace	,605	1,752 ^a	7,000	8,000	,224	,605
		Wilks' Lambda	,395	1,752 ^a	7,000	8,000	,224	,605
		Hotelling's Trace	1,533	1,752 ^a	7,000	8,000	,224	,605
		Roy's Largest Root	1,533	1,752 ^a	7,000	8,000	,224	,605

a. Exact statistic

b.

Design: Intercept

Within Subjects Design: FACTOR1

a)

Mauchly's Test of Sphericity^b

Within Subjects Effect	Measure	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon ^a		
						Greenhouse-Geisser	Huynh-Feldt	Lower-bound
FACTOR1	GR20	1,000	,000	0	,	1,000	1,000	1,000
	SPIS	1,000	,000	0	,	1,000	1,000	1,000
	BAUD	1,000	,000	0	,	1,000	1,000	1,000
	SAUD	1,000	,000	0	,	1,000	1,000	1,000
	SBM	1,000	,000	0	,	1,000	1,000	1,000
	SSM	1,000	,000	0	,	1,000	1,000	1,000
	METT	1,000	,000	0	,	1,000	1,000	1,000

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

b.

Design: Intercept

Within Subjects Design: FACTOR1

b)

Tests of Within-Subjects Effects

Multivariate^{b,c}

Within Subjects Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
FACTOR1	Pillai's Trace	,605	1,752 ^a	7,000	8,000	,224	,605
	Wilks' Lambda	,395	1,752 ^a	7,000	8,000	,224	,605
	Hotelling's Trace	1,533	1,752 ^a	7,000	8,000	,224	,605
	Roy's Largest Root	1,533	1,752 ^a	7,000	8,000	,224	,605

a. Exact statistic

b.

Design: Intercept

Within Subjects Design: FACTOR1

c. Tests are based on averaged variables.

c)

15.25.2 pav. Pagrindiniai daugiamečių blokuotųjų duomenų dispersinės analizės rezultatai

Univariate Tests

Source	Measure		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
FACTOR1	GR	Sphericity Assumed	3,630E-03	1	3,630E-03	,135	,719	,010
		Greenhouse-Geisser	3,630E-03	1,000	3,630E-03	,135	,719	,010
		Huynh-Feldt	3,630E-03	1,000	3,630E-03	,135	,719	,010
		Lower-bound	3,630E-03	1,000	3,630E-03	,135	,719	,010
	SPIS	Sphericity Assumed	1,200	1	1,200	1,714	,212	,109
		Greenhouse-Geisser	1,200	1,000	1,200	1,714	,212	,109
		Huynh-Feldt	1,200	1,000	1,200	1,714	,212	,109
		Lower-bound	1,200	1,000	1,200	1,714	,212	,109
	BAUD	Sphericity Assumed	30,000	1	30,000	3,043	,103	,179
		Greenhouse-Geisser	30,000	1,000	30,000	3,043	,103	,179
		Huynh-Feldt	30,000	1,000	30,000	3,043	,103	,179
		Lower-bound	30,000	1,000	30,000	3,043	,103	,179
	SAUD	Sphericity Assumed	9,097	1	9,097	10,631	,006	,432
		Greenhouse-Geisser	9,097	1,000	9,097	10,631	,006	,432
		Huynh-Feldt	9,097	1,000	9,097	10,631	,006	,432
		Lower-bound	9,097	1,000	9,097	10,631	,006	,432
	SBM	Sphericity Assumed	6,816	1	6,816	,531	,478	,037
		Greenhouse-Geisser	6,816	1,000	6,816	,531	,478	,037
		Huynh-Feldt	6,816	1,000	6,816	,531	,478	,037
		Lower-bound	6,816	1,000	6,816	,531	,478	,037
	SSM	Sphericity Assumed	38,760	1	38,760	2,220	,158	,137
		Greenhouse-Geisser	38,760	1,000	38,760	2,220	,158	,137
		Huynh-Feldt	38,760	1,000	38,760	2,220	,158	,137
		Lower-bound	38,760	1,000	38,760	2,220	,158	,137
	METT	Sphericity Assumed	,533	1	,533	,118	,737	,008
		Greenhouse-Geisser	,533	1,000	,533	,118	,737	,008
		Huynh-Feldt	,533	1,000	,533	,118	,737	,008
		Lower-bound	,533	1,000	,533	,118	,737	,008
Error(FACTOR1)	GR	Sphericity Assumed	,377	14	2,691E-02			
		Greenhouse-Geisser	,377	14,000	2,691E-02			
		Huynh-Feldt	,377	14,000	2,691E-02			
		Lower-bound	,377	14,000	2,691E-02			
	SPIS	Sphericity Assumed	9,800	14	,700			
		Greenhouse-Geisser	9,800	14,000	,700			
		Huynh-Feldt	9,800	14,000	,700			
		Lower-bound	9,800	14,000	,700			
	BAUD	Sphericity Assumed	138,000	14	9,857			
		Greenhouse-Geisser	138,000	14,000	9,857			
		Huynh-Feldt	138,000	14,000	9,857			
		Lower-bound	138,000	14,000	9,857			
	SAUD	Sphericity Assumed	11,980	14	,856			
		Greenhouse-Geisser	11,980	14,000	,856			
		Huynh-Feldt	11,980	14,000	,856			
		Lower-bound	11,980	14,000	,856			
	SBM	Sphericity Assumed	179,629	14	12,831			
		Greenhouse-Geisser	179,629	14,000	12,831			
		Huynh-Feldt	179,629	14,000	12,831			
		Lower-bound	179,629	14,000	12,831			
	SSM	Sphericity Assumed	244,415	14	17,458			
		Greenhouse-Geisser	244,415	14,000	17,458			
		Huynh-Feldt	244,415	14,000	17,458			
		Lower-bound	244,415	14,000	17,458			
	METT	Sphericity Assumed	63,467	14	4,533			
		Greenhouse-Geisser	63,467	14,000	4,533			
		Huynh-Feldt	63,467	14,000	4,533			
		Lower-bound	63,467	14,000	4,533			

15.25.3 pav. Pagrindiniai daugiamatės blokuotųjų duomenų dispersinės analizės rezultatai

Tests of Within-Subjects Contrasts

Source	Measure	FACTOR1	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
FACTOR1	GR	Linear	3,630E-03	1	3,630E-03	,135	,719	,010
	SPIS	Linear	1,200	1	1,200	1,714	,212	,109
	BAUD	Linear	30,000	1	30,000	3,043	,103	,179
	SAUD	Linear	9,097	1	9,097	10,631	,006	,432
	SBM	Linear	6,816	1	6,816	,531	,478	,037
	SSM	Linear	38,760	1	38,760	2,220	,158	,137
	METT	Linear	,533	1	,533	,118	,737	,008
Error(FACTOR1)	GR	Linear	,377	14	2,691E-02			
	SPIS	Linear	9,800	14	,700			
	BAUD	Linear	138,000	14	9,857			
	SAUD	Linear	11,980	14	,856			
	SBM	Linear	179,629	14	12,831			
	SSM	Linear	244,415	14	17,458			
	METT	Linear	63,467	14	4,533			

a)

Tests of Between-Subjects Effects

Transformed Variable: Average

Source	Measure	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	GR	378,714	1	378,714	22458,49	,000	,999
	SPIS	8003,333	1	8003,333	8198,537	,000	,998
	BAUD	14696,533	1	14696,533	1454,417	,000	,990
	SAUD	64003,635	1	64003,635	19269,38	,000	,999
	SBM	49621,467	1	49621,467	2933,231	,000	,995
	SSM	68956,896	1	68956,896	2619,384	,000	,995
	METT	1888,133	1	1888,133	456,806	,000	,970
Error	GR	,236	14	1,686E-02			
	SPIS	13,667	14	,976			
	BAUD	141,467	14	10,105			
	SAUD	46,501	14	3,322			
	SBM	236,838	14	16,917			
	SSM	368,559	14	26,326			
	METT	57,867	14	4,133			

b)

15.25.4 pav. Pagrindiniai daugiamačių blokuotųjų duomenų dispersinės analizės rezultatai

Estimated Marginal Means

FACTOR1

Estimates

Measure	FACTOR1	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
GR	1	3,542	,039	3,459	3,625
	2	3,564	,037	3,484	3,644
SPIS	1	16,133	,165	15,779	16,488
	2	16,533	,291	15,910	17,157
BAUD	1	21,133	,970	19,052	23,215
	2	23,133	,624	21,795	24,471
SAUD	1	46,740	,423	45,832	47,648
	2	45,639	,315	44,963	46,315
SBM	1	40,193	,799	38,479	41,908
	2	41,147	1,159	38,660	43,633
SSM	1	46,807	1,202	44,229	49,385
	2	49,080	1,214	46,476	51,684
METT	1	7,800	,518	6,688	8,912
	2	8,067	,556	6,874	9,259

15.25.5 pav. Pagrindiniai daugiamačės blokuotųjų duomenų dispersinės analizės rezultatai

15.25.4 pav. b) lentelėje *Test of Between-Subjects Effects* pateikti testo, kuris įvertina grupės faktorių įtaką, rezultatai. Šiame pavyzdyje tokių faktorių nėra, todėl pateikiami formalūs hipotezės, teigiančios, kad priklausomų kintamųjų (matuojamų parametru) vidurkis nesiskiria reikšmingai nuo nulio, tikrinimo rezultatai. Kadangi iš tikrųjų visų matuojamų parametru vidurkis yra daug daugiau už nulį, nulinę hipotezę atmetame, nes p -reikšmė (*Sig.*) $p=0,000$.

15.25.5 pav. lentelėje *Estimates* yra pateikiamas priklausomų kintamųjų (matuojamų parametru) vidurkis, standartinė paklaida ir pasikliautinis intervalas kiekvienai tarpbandyminio faktoriaus reikšmei.

16. KLAŠTERINĖ ANALIZĖ

Klašterinė analizė — tai metodas identifikuoti homogenines objektų arba stebėjimų grupes (klašterius) — objektai suskirstomi taip, kad skirtumai klašterių viduje būtų kuo mažesni, o tarp klašterių — kuo didesni. Pagrindinės klašterinės analizės sąvokos yra panašumas ir skirtingumas (atstumas): atstumas nurodo, kiek objektai yra nutolę vienas nuo kito (skirtingi), o panašumas rodo objektų artumą. Panašūs objektai priklauso tam pačiam klašteriui, nutolę objektai — skirtingiems klašteriams. Pavyzdžiui, valstybės gali būti suskirstytos į homogenines grupes pagal biudžeto ir ekonomikos rodiklius, turizmo rinka — pagal įvairius paklausos rodiklius ir t. t.

Pats paprasčiausias būdas suvokti dviejų kintamųjų, kurių duomenys sudaro keletą homogeninių grupių, skirstymo į klašterius reiškini, yra nubraižyti šių dviejų kintamųjų sklaidos (*Scatter*) diagramą. Realiai tokia galimybė pasitaiko labai retai — paprastai klašterių struktūros nėra aiškiai atskirtos ir persidengia, o kintamųjų skaičius žymiai viršija du, t. y. analizė atliekama n -matėje erdvėje.

Skiriamos dvi pagrindinės klašterinės analizės metodų klasės — hierarchiniai ir nehierarchiniai metodai (Čekanavičius, Murauskas, 2002). Hierarchiniais metodais nustatoma bendra visų klašterių tarpusavio priklausomybių struktūra ir tik tada sprendžiama, koks klašterių skaičius optimalus. Savo ruožtu hierarchiniai metodai skirstomi į jungimo ir skaidymo metodus. Taikant jungimo metodus, iš pradžių visi stebėjimai traktuojami kaip atskiri klašteriai. Pirmuoju žingsniu du stebėjimai yra sujungiami į klašterį, kiekvienu kitu žingsniu naujas stebėjimas yra jungiamas prie esamo klašterio arba du klašteriai sujungiami. Suformuotas klašteris vėliau jau negali būti skaldomas — jis gali būti tik jungiamas su kitais klašteriais. Vyksmas kartojamas tol, kol lieka vienas klašteris. Tyrėjas pats sprendžia, kuriuo etapu objektų paskirstymas į klašterius yra optimalus. Skaidymo metodai yra loginė jungimo metodų priešingybė — vienintelis klašteris nuosekliai skaidomas į dalis. Nehierarchiniai metodai paprastai taikomi tada, kai iš anksto žinomas klašterių skaičius ir norima tiriamus objektus klašterizuoti.

Šiame skyriuje aprašysime dažniausiai naudojamus hierarchinį jungimo metodą (taikomą nedidelėms imtims — stebėjimų skaičius iki 300) bei didelėms imtims taikomą nehierarchinį — k -vidurkių metodą.

Taikant klašterinės analizės metodus reikia turėti omenyje, kad:

- klašterinėje analizėje yra daug euristinių, neturinčių teorinio pagrindimo, metodų;
- klašterinės analizės metodai turi nemažai specifiškumo;
- tiems patiems duomenims taikant skirtingus klašterinės analizės metodus, galima gauti skirtingus rezultatus.

16.1. OBJEKTŲ PANAŠUMO MATAI

Stebėjimų klasterinių grupių nustatymo pagrindas yra atstumų arba panašumo matrica. Priklausomai nuo to, kokiai matavimų skalei priklauso analizuojami kintamieji, SPSS pakete yra naudojami skirtingi atstumo ir panašumo matai.

16.1.1. Intervalinių kintamųjų atstumo matai

Kai objektus apibūdinantys požymiai matuojami pagal intervalų arba santykių skalę, naudojami metriniai atstumo matai. Šiuos matus galima būtų vadinti skirtingumo matais — kuo didesnė reikšmė, tuo objektai mažiau panašūs. Nevienodai skirtingai matuojamų požymių įtakai išvengti naudojamos šių kintamųjų standartizuotosios reikšmės (z -reikšmės).

Intervalinių kintamųjų panašumo matas gali būti taip pat koreliacijos koeficientas.

Euklido atstumas

Euklido atstumas tarp taško x ir taško y — tai trumpiausias atstumas tarp šių taškų. Plokštumoje arba trimatėje erdvėje — tai taškus jungianti tiesė. Bendru n -matės erdvės atveju šis atstumas apskaičiuojamas pagal formulę

$$Dist(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (16.1)$$

Euklido atstumo kvadratas

Tai nustatytasis variantas SPSS programoje. Šiuo matu tiksliau įvertinami dideli skirtumai. Euklido atstumo kvadratas turi būti naudojamas pasirinkus centrų, medianos arba Ward'o klasterizacijos metodą.

$$Dist(x, y) = \sum_{i=1}^n (x_i - y_i)^2 \quad (16.2)$$

Kosinuso panašumo matas

$$Sim(x, y) = \frac{\sum_{i=1}^n (x_i \cdot y_i)}{\sqrt{\sum_{i=1}^n (x_i^2 \cdot y_i^2)}} \quad (16.3)$$

Šio panašumo (*Similarity*) mato reikšmių diapazonas yra tarp -1 ir $+1$.

Čebyšovo atstumas

Atstumu tarp dviejų stebėjimų yra laikoma didžiausia absoliuti skirtumo tarp visų šių stebėjimų porų reikšmė

$$Dist(x, y) = \max_i |x_i - y_i| \quad (16.4)$$

Bloko atstumo matas

Tai absoliučių skirtumų tarp visų stebėjimų porų reikšmių suma

$$Dist(x, y) = \sum_{i=1}^n |x_i - y_i| \quad (16.5)$$

Pirsono koreliacijos panašumo matas

Netaikomas tuo atveju, kai yra tik du kintamieji.

16.1.2. Stebėjimų dažniais išreikštų kintamųjų atstumo matai

Atstumas tarp dviejų stebėjimų nustatomas lyginant priklausančių šiems stebėjimams kintamųjų pasirodymo dažnius.

Chi-kvadrato (χ^2) matas

Tai nustatytasis variantas SPSS pakete.

$$Dist(x, y) = \sqrt{\sum_{i=1}^n \frac{(O_{X_i} - E_{X_i})^2}{E_{X_i}} + \sum_{i=1}^n \frac{(O_{Y_i} - E_{Y_i})^2}{E_{Y_i}}}, \quad (16.6)$$

čia O_{X_i}, O_{Y_i} — eksperimentiniu būdu gauti dažniai, E_{X_i}, E_{Y_i} — tikėtini dažniai.

Fi-kvadrato (ϕ^2) matas

$$Dist(x, y) = \sqrt{\frac{\sum_{i=1}^n \frac{(O_{X_i} - E_{X_i})^2}{E_{X_i}} + \sum_{i=1}^n \frac{(O_{Y_i} - E_{Y_i})^2}{E_{Y_i}}}{N}}, \quad (16.7)$$

čia N — bendra dažnių suma.

16.1.3. Binarinių kintamųjų atstumo ir panašumo matai

SPSS pakete yra nustatytas šis dvireikšmių (binarinių) duomenų kodavimas: 1 — įvykis įvyko, 0 — įvykis neįvyko. 16.1 lentelėje raidėmis a , b , c ir d pažymėti dviejų binarinių kintamųjų visų galimų stebėjimų kombinacijų dažniai. Remiantis šiais dažniais yra apskaičiuojama daug įvairių atstumo ir panašumo matų. Dalis jų pateikta 13 skyriuje, dalis pateikiama toliau.

16.1 lentelė. Galimos dviejų dvireikšmių kintamųjų kombinacijos

1 kintamasis	2 kintamasis		Σ
	įvyko	neįvyko	
įvyko	a	b	$a+b$
neįvyko	c	d	$c+d$
Σ	$a+c$	$b+d$	$m = a+b+c+d$

Euklido atstumas

$$Dist = \sqrt{b+c} \quad (16.8)$$

Euklido atstumo kvadratas

$$Dist = b+c \quad (16.9)$$

Didumų skirtumas

$$Dist = \frac{(b-c)^2}{(a+b+c+d)^2} \quad (16.10)$$

Pavyzdinis skirtumas

$$Dist = \frac{b \cdot c}{(a+b+c+d)^2} \quad (16.11)$$

Sklaida

$$Dist = \frac{b+c}{4 \cdot (a+b+c+d)} \quad (16.12)$$

Formos koeficientas

$$Dist = \frac{(a + b + c + d) \cdot (b + c) - (b - c)^2}{(a + b + c + d)^2} \quad (16.13)$$

Lance ir Williams'o atstumas

$$Dist = \frac{b + c}{2a + b + c} \quad (16.14)$$

16.2. JUNGIMO METODAI

SPSS pakete yra pateikiami septyni duomenų jungimo (klasterizavimo) metodai. Nustatytasis variantas yra *Between-groups linkage* (ryšys tarp grupių).

Ryšys tarp grupių

Atstumas tarp klasterių yra lygus atstumo tarp visų galimų stebėjimų porų, kai vienas stebėjimas yra imamas iš vieno klasterio, o kitas — iš antro, vidurkiui.

Ryšys grupių viduje

Atstumas tarp klasterių yra apskaičiuojamas remiantis visomis abiejų klasterių stebėjimų poromis, tarp jų ir esančiomis klasterio viduje.

Artimiausio kaimyno

Atstumas tarp klasterių yra apibrėžiamas kaip atstumas tarp priklausančių skirtingiems klasteriams dviejų artimiausių stebėjimų reikšmių.

Tolimiausio kaimyno

Atstumas tarp klasterių yra apibrėžiamas kaip atstumas tarp priklausančių skirtingiems klasteriams dviejų tolimiausių stebėjimų reikšmių.

Centrų klasterizacija

Abiejuose klasteriuose yra apskaičiuojami stebėjimų reikšmių vidurkiai, o atstumas tarp klasterių yra apskaičiuojamas kaip atstumas tarp šių vidurkių.

Medianos klasterizacija

Metodas panašus į centrų klasterizacijos metodą, tačiau abudu klasteriai turi vienodą svorį (nepriklausantį nuo klasterio stebėjimų skaičiaus).

Ward'o metodas

Taikant Ward'o metodą, imama klasterių pora ir apskaičiuojama požymių vektorių nuokrypių nuo jungtinio klasterio centro kvadratų suma. Analogiškai daroma su kita klasterių pora ir t. t. Jungiami tie klasteriai, kuriuos sujungus nuokrypių suma mažiausia.

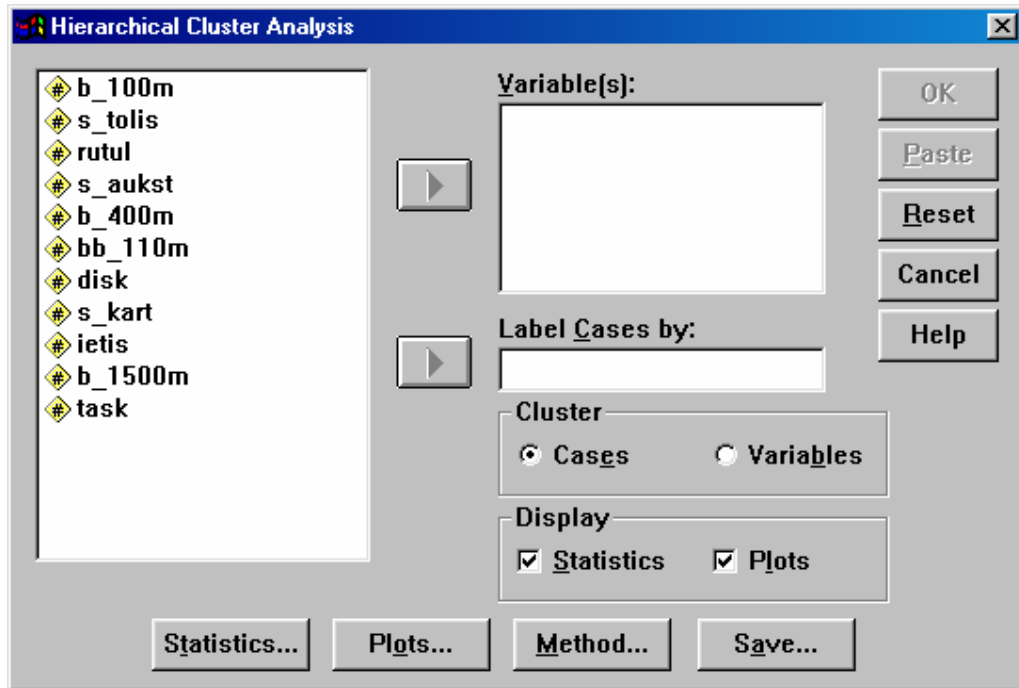
Iš visų aprašytųjų metodų dažniausiai yra naudojamas nustatytasis variantas — ***Between-groups linkage*** (ryšys tarp grupių). Dažnai naudojamas ir Ward'o metodas (***Ward's method***), tačiau taikant šį metodą gaunama daugiau (mažesnių) klasterių, metodo rezultatams didelę įtaką daro objektai su požymių išskirtimis. Apskritai išskirtys iškraipo visų klasterizavimo metodų rezultatus, todėl prieš klasterizuojant pageidautina jas pašalinti. Kintamieji taip pat neturi būti koreliuoti, visuose grupėse tikėtinas tas pats elementų skaičius. Objektams klasterizuoti rekomenduojama taikyti keletą klasterizavimo metodų — tada galima tikėtis patikimesnių rezultatų.

16.3. HIERARCHINĖ KLASTERINĖ ANALIZĖ

Hierarchinei klasterinei analizei atlikti:

- Atidarykite bylą su analizuojamais duomenimis.
- Nurodykite komandas ***Analyze → Classify → Hierarchical Cluster...***
- Atsidariusiame dialogo langelyje ***Hierarchical Cluster Analysis*** (16.1 pav.) įkelkite analizuojamus kintamuosius į laukelį ***Variable(s)***. Klasifikuojant stebėjimus (pažymėtas ***Cluster*** laukelis ***Cases***), šiame laukelyje būtina įrašyti mažiausiai vieną kintamąjį, o klasifikuojant kintamuosius (pažymėtas ***Cluster*** laukelis ***Variables***) — mažiausiai tris kintamuosius.
- Spragtelėję mygtuką ***Method...***, naujame dialogo langelio ***Hierarchical Cluster Analysis: Method*** (16.2 pav.) išskleidžiamajame sąrašė ***Cluster Method*** galite pasirinkti klasterių suformavimo metodą. Kaip alternatyva nustatytajam ***Between-groups linkage*** (ryšys tarp grupių) metodui — jis yra rekomenduojamas — gali būti pasirinktas vienas iš 16.2 skyriuje paminėtų metodų. Atstumo ir panašumo matų pasirinkimas priklauso nuo to, kokiai skalei priklauso analizuojami kintamieji: intervalų, išreikšti dažniais ar binariniais. Pažymėję atitinkamą ***Measure*** grupės laukelį — ***Interval***, ***Counts*** ar ***Binary*** galite pasirinkti vieną iš 16.1 skyriuje paminėtų metodų. Nustatytieji intervaliniams kintamiesiems, dažnių kintamiesiems ir binariniais kintamiesiems matai atitinkamai yra

Squared Euclidean distance (Euklido atstumo kvadratas), **Chi-square measure** (Chi-kvadrato matas) ir **Squared Euclidean distance** (Euklido atstumo kvadratas). **Transform Values** išskleidžiamajame sąrašė **Standardize** galite pasirinkti gerokai besiskiriančių duomenų standartizacijos metodą — paprastai pasirenkamas z-standartizacijos metodas (**Z scores**). Spragtelėkite mygtuką **Continue**.



16.1 pav. Dialogo langelis *Hierarchical Cluster Analysis*

- Spragtelėkite dialogo langelio *Hierarchical Cluster Analysis* mygtuką **Statistics...** Naujame dialogo langelyje *Hierarchical Cluster Analysis: Statistics* (16.3 pav.), be duomenų jungimo į klasterius eiliškumo išvesties **Agglomeration schedule**, galite taip pat nurodyti atstumo ir panašumo matų išvestį — **Proximity matrix**, duomenų priklausomybės konkrečiam klasteriui išvestį **Cluster Membership: Single solution** (vienas sprendimas — nurodomas konkretus klasterių skaičius) ar **Range of solutions** (sprendimų sritis — nurodomas klasterių skaičiaus intervalas).
- Spragtelėję dialogo langelio *Hierarchical Cluster Analysis* mygtuką **Plots...**, naujame dialogo langelyje *Hierarchical Cluster Analysis: Plots* (16.4 pav.) galite pasirinkti klasterizacijos vyksmą vaizduojantį grafiką—dendogramą (medžio pavidalo diagramą) **Dendrogram** ir jungimo schemą (nenurodant atstumų) **Icicle** (angl. varveklis) visam klasteriui ar keliems jo sudarymo žingsniams. Spragtelėkite mygtuką **Continue**.

Hierarchical Cluster Analysis: Method

Cluster Method: **Between-groups linkage**

Measure

☒ Interval: **Squared Euclidean distance**
 Power: **2** Root: **2**

☐ Counts: **Chi-square measure**

☐ Binary: **Squared Euclidean distance**
 Present: **1** Absent: **0**

Transform Values

Standardize: **None**
☒ By variable
☐ By case

Transform Measures

☐ Absolute values
☐ Change sign
☐ Rescale to 0-1 range

Continue
 Cancel
 Help

16.2 pav. Dialogo langelis *Hierarchical Cluster Analysis: Method*

Hierarchical Cluster Analysis: Statistics

☒ Agglomeration schedule

☐ Proximity matrix

Cluster Membership

☒ None

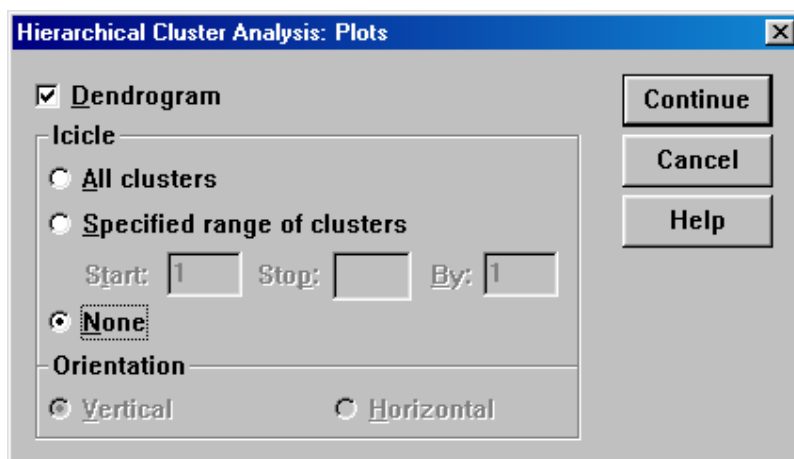
☐ Single solution: clusters

☐ Range of solutions:
 From through clusters

Continue
 Cancel
 Help

16.3 pav. Dialogo langelis *Hierarchical Cluster Analysis: Statistics*

- Spragtelęję dialogo langelio *Hierarchical Cluster Analysis* mygtuką **Save...**, naujame dialogo langelyje *Hierarchical Cluster Analysis: Save* galite nurodyti išsaugoti duomenų rinkmenoje atskirų stebėjimų priklausomybės konkrečiam klasteriui duomenis (*Cluster Membership*).
- Spragtelėkite pagrindinio dialogo langelio *Hierarchical Cluster Analysis* mygtuką **OK**.



16.4 pav. Dialogo langelis *Hierarchical Cluster Analysis: Plots*

Pavyzdys. Pabandykime suskirstyti į klasterius didelio meistriškumo dešimtkovinėms pagal jų atskirų rungčių rezultatus. Pasinaudosime pirmųjų 23 sportininkų per Sidnėjaus olimpinės žaidynės rezultatais. Duomenų rinkmenoje (16.5 pav.) atskiras dešimtkovės rungtis atitinkantys kintamieji pavadinti šiais vardais:

- 1) 100 m bėgimas — *b_100m*;
 - 2) šuolis į tolį — *s_tolis*;
 - 3) rutulio stūmimas — *rutul*;
 - 4) šuolis į aukštį — *s_aukst*;
 - 5) 400 m bėgimas — *b_400m*;
 - 6) 110 m barjerinis bėgimas — *bb_110m*;
 - 7) disko metimas — *disk*;
 - 8) šuolis su kartimi — *s_kart*;
 - 9) ieties metimas — *ietis*;
 - 10) 1500 m bėgimas — *b_1500m*;
 - 11) bendra taškų suma — *task*;
- Dialogo langelyje *Hierarchical Cluster Analysis* (16.1 pav.) įkeliamo visus kintamuosius (išskyrus kintamąjį *task*) į laukelį *Variable(s)*. Paliekame nustatytąjį *Clusters* variantą *Cases* ir nustatytuosius *Display* variantus *Statistics* ir *Plots*.
 - Dialogo langelyje *Hierarchical Cluster Analysis: Statistics* (16.3 pav.) paliekame pažymėtą laukelį *Agglomeration schedule*, taip pat pažymime *Cluster Membership* laukelį *Range of solutions* ir nurodome tikėtiną optimalių klasterių kiekį — nuo (*from*) 2 iki (*through*) 3.
 - Dialogo langelyje *Hierarchical Cluster Analysis: Plots* (16.4 pav.) pažymime laukelį *Dendrogram*, taip pat *Icicle* grupės laukelį *All clusters* ir *Orientation* grupės laukelį *Vertical*.

decaathlon.sav - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

1 : b_100m 11,25

	b_100m	s_tolis	rutul	s_aukst	b_400m	bb_110m	disk	s_kart	ietis	b_1500m	task
1	11,25	7,43	15,48	2,27	48,90	15,13	49,28	4,70	61,32	268,95	8488,00
2	10,87	7,45	14,97	1,97	47,71	14,46	44,36	5,10	61,76	273,02	8399,00
3	11,18	7,44	14,20	1,97	48,29	14,81	43,66	5,20	64,16	263,20	8328,00
4	10,62	7,38	15,02	2,03	49,06	14,72	44,80	4,90	64,04	285,11	8306,00
5	11,02	7,43	12,92	1,97	47,44	14,40	41,20	5,20	57,46	256,64	8286,00
6	10,83	7,72	13,58	2,12	48,34	14,18	43,06	4,90	52,18	274,07	8272,00
7	11,18	7,05	14,12	2,06	49,34	14,39	41,68	5,70	61,60	291,20	8216,00
8	11,05	6,95	15,34	2,00	48,21	14,36	41,32	4,80	63,00	265,86	8189,00
9	11,15	7,12	14,52	2,03	49,15	14,66	42,36	4,90	66,46	269,62	8180,00
10	11,23	7,28	15,25	1,97	48,60	14,76	48,02	5,20	59,48	292,24	8167,00
11	10,94	7,45	15,34	1,97	49,94	14,25	41,86	4,80	66,64	295,89	8143,00
12	11,18	7,34	14,48	1,94	49,02	15,11	42,76	4,70	65,84	256,74	8114,00
13	11,02	7,29	12,92	2,06	48,23	14,94	39,54	5,00	56,80	257,85	8093,00
14	10,99	7,37	13,61	1,97	47,83	14,70	43,88	4,30	66,54	268,97	8083,00
15	11,03	7,45	14,20	1,97	48,94	15,44	41,66	4,70	64,00	267,48	8036,00
16	11,09	7,08	14,51	2,03	49,89	14,78	43,20	4,90	57,18	268,54	8021,00
17	11,57	7,00	16,60	1,94	49,84	15,00	46,66	4,90	60,20	286,04	7869,00
18	11,07	7,04	13,41	1,94	47,97	14,96	40,38	4,50	51,50	262,41	7860,00
19	10,89	7,07	15,84	1,79	49,68	15,38	45,32	4,90	60,48	277,84	7859,00
20	11,52	7,36	13,93	1,94	49,99	15,64	38,82	4,60	67,04	266,42	7781,00
21	11,49	7,02	13,80	2,03	50,60	15,22	39,08	4,70	60,92	262,93	7753,00
22	11,38	7,08	14,31	2,00	50,24	14,97	46,34	4,40	55,68	272,68	7745,00
23	11,30	6,97	13,23	2,15	49,98	15,38	38,72	4,60	54,34	277,84	7743,00

16.5 pav. Dešimtkovininų rezultatų duomenų rinkmena

- Dialogo langelyje **Hierarchical Cluster Analysis: Method** (16.2 pav.) paliekame nustatytą **Between-groups linkage** klasterizacijos metodą ir **Squared Euclidean distance** matuojamų pagal intervalų skalę kintamųjų atstumo matą. Kadangi atskirų rungčių rezultatai matuojami skirtingai, **Transform Values Standardize** laukelyje nurodome **Z scores** ir pažymime laukelį **By variable** (transformacija pagal kintamuosius).
- Spragtelime pagrindinio dialogo langelio **Hierarchical Cluster Analysis** mygtuką **OK**.

Pagrindiniai skaičiavimo rezultatai parodyti 16.6.1 — 16.6.2 pav.

16.6.1 pav. a) lentelėje **Agglomeration Schedule**, stulpelyje **Cluster Combined** pateikiamas duomenų jungimo į klasterius eiliškumas: pirmame žingsnyje buvo sujungti 12 ir 15 stebėjimai, antrame — 8 ir 9 stebėjimai ir t. t. Optimalaus klasterių skaičiaus problema nėra sprendžiama analitiškai — „teisingas“ skaičius labai priklauso nuo interpretacijos. Koks klasterių skaičius yra optimalus, paprastai sprendžiama vadovaujantis lentelės **Agglomeration Schedule** rodikliu **Coefficients**, kuris reiškia atstumą tarp dviejų klasterių pagal pasirinktą atstumo matą. Pasiekus etapą, kur šis atstumas šuoliškai padidėja, sukaupimo į klasterius vyksmą reikėtų laikyti užbaigtą, nes, priešingu atveju, gali būti sujungti toli vienas nuo kito esantys klasteriai. Šiame pavyzdyje nėra šuoliško atstumo mato pokyčio, tai rodo, kad suskirstymas į klasterius bus labai

sąlygiškas. Tarkime, kad optimalus klasterių skaičius yra 3. Stulpeliuose *Stage Cluster First Appears* (žingsnis, kuriame klasteris pasirodo pirmą kartą arba žingsnis, kuriame šis klasteris figūravo paskutinį kartą) ir *Next Stage* (kitas žingsnis, kuriame figūruoja klasteris) pateikiami papildomi klasterių sudarymo eigos duomenys.

Lentelėje *b) Cluster Membership* parodoma, kuriam klasteriui yra priskiriamas kiekvienas stebėjimas. Kadangi dialogo langelyje ***Hierarchical Cluster Analysis: Statistics*** pažymėjome atitinkamą laukelį ***Range of solutions*** ir nurodėme tikėtiną optimalų klasterių kiekį — nuo 2 iki 3, lentelėje yra pateikiami du stebėjimų priskyrimo atskiriems klasteriams variantai — kai yra trys klasteriai ir kai yra du klasteriai.

16.6.2 pav. dendrogramoje (*Dendrogram*) ir *Vertical Icicle* schemeje pagal *Agglomeration Shedule* lentelės duomenis vizualiai parodytas duomenų jungimo į klasterius vyksmas. Koeficientų reikšmės dendrogramoje yra transformuotos (*Rescaled Distance Cluster Combine*) į skalės nuo 0 iki 25 reikšmes.

Matome, kad į trečią klasterį pateko 7, 10, 17 ir 19 vietas užėmusių sportininkų rezultatai, į antrą — visų likusiųjų, išskyrus 1 vietą užėmusį sportininką, kurio rezultatai sudaro pirmą klasterį. Gautą klasterių kiekį galima logiškai paaiškinti — į pirmą klasterį patenka nugalėtojas, vienodai gerai pasirodęs visose rungtyse, į antrą klasterį patenka aukštus rezultatus bėgimo ir šuolių rungtyse pasiekę sportininkai, į trečią — aukštus rezultatus metimo rungtyse, bet kuklesnius rezultatus bėgimo ir šuolių rungtyse pasiekę sportininkai. Kaip minėjome anksčiau, šis suskirstymas yra sąlygiškas, nes didelio meistriškumo dešimtkovinė parodo aukštus rezultatus visų dešimties rungčių.

Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	12	15	2,764	0	0	8
2	8	9	4,231	0	0	9
3	5	13	4,765	0	0	12
4	2	3	4,801	0	0	9
5	16	22	6,226	0	0	13
6	20	21	7,113	0	0	14
7	4	11	7,247	0	0	16
8	12	14	7,774	1	0	11
9	2	8	8,896	4	2	11
10	10	17	9,809	0	0	15
11	2	12	9,849	9	8	16
12	5	18	10,598	3	0	17
13	16	23	12,236	5	0	14
14	16	20	13,978	13	6	19
15	10	19	14,225	10	0	20
16	2	4	14,411	11	7	18
17	5	6	16,358	12	0	18
18	2	5	18,608	16	17	19
19	2	16	21,424	18	14	21
20	7	10	22,225	0	15	21
21	2	7	23,771	19	20	22
22	1	2	27,010	0	21	0

a)

Cluster Membership

Case	3 Clusters	2 Clusters
1:Case 1	1	1
2:Case 2	2	2
3:Case 3	2	2
4:Case 4	2	2
5:Case 5	2	2
6:Case 6	2	2
7:Case 7	3	2
8:Case 8	2	2
9:Case 9	2	2
10:Case 10	3	2
11:Case 11	2	2
12:Case 12	2	2
13:Case 13	2	2
14:Case 14	2	2
15:Case 15	2	2
16:Case 16	2	2
17:Case 17	3	2
18:Case 18	2	2
19:Case 19	3	2
20:Case 20	2	2
21:Case 21	2	2
22:Case 22	2	2
23:Case 23	2	2

b)

16.6.1 pav. Pagrindiniai hierarcinės klasterinės analizės rezultatai

16.4. *k*-VIDURKIŲ KLASTERINĖ ANALIZĖ

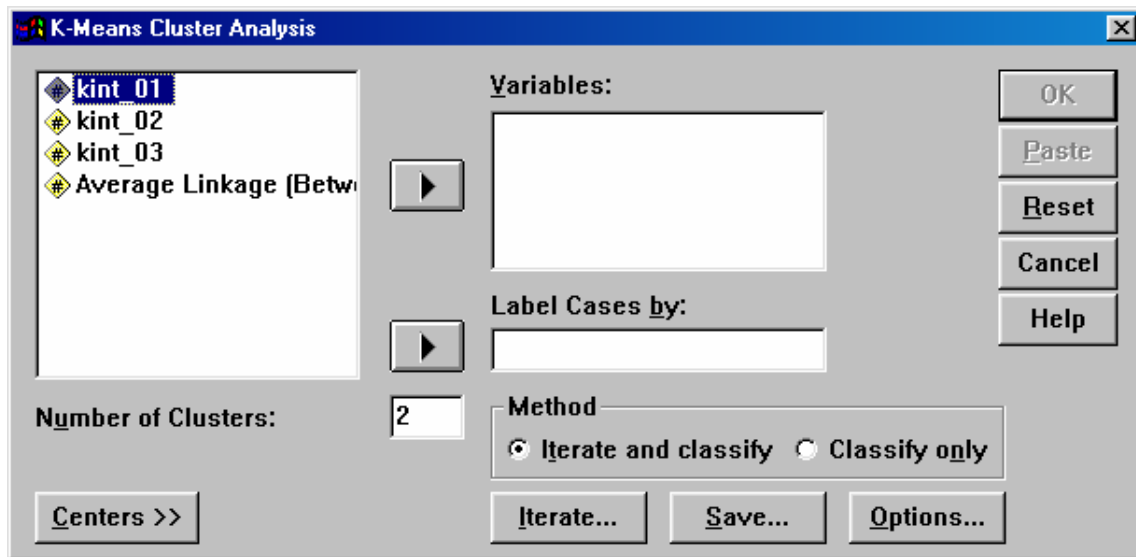
Hierarchiniai analizės metodai yra tikslūs, tačiau didėjant stebėjimų skaičiui, reikia didelių skaičiavimo pajėgumų atstumų matricai apskaičiuoti ir jos elementams išsaugoti — skaičiavimo laikas ilgėja proporcingai stebėjimų skaičiui, pakeltam trečiuoju laipsniu. Todėl dideliame stebėjimų kiekiui apdoroti yra naudojami kiti metodai. Šių algoritmų trūkumas yra tas, kad būtina iš anksto nurodyti klasterių skaičių. Apriorinį klasterių skaičių galima nustatyti atlikus stebėjimų atsitiktinės imties hierarchinę klasterinę analizę. Kita problema yra klasterių centrų pradinių reikšmių nustatymas, kuri SPSS pakete yra sprendžiama iteracijos būdu.

k-vidurkių klasterinė analizė yra atliekama kiekybiniais (intervalų ar santykio skalės) kintamiesiems. Jeigu kintamieji yra binariniai ar išreikšti dažniais, naudokite hierarchinę klasterinę analizę.

Kaip atstumo matas yra naudojamas tik Euklido atstumas.

k-vidurkių klasterinės analizės rezultatams gauti:

- Atidarykite bylą su analizuojamais duomenimis.
- Nurodykite komandas **Analyze → Classify → K-Means Cluster...**
- Atsidariusiame dialogo langelyje **K-Means Cluster Analysis** (16.7 pav.) įkelkite analizuojamus kintamuosius į laukelį **Variables**.
- Laukelyje **Number of Clusters** įrašykite apriorinį klasterių skaičių. Mažiausias klasterių skaičius yra 2.
- **Method** grupėje palikite nustatytąjį **Iterate and classify** variantą.
- Spragtelėkite mygtuką **Iterate...** ir atsidariusiame dialogo langelyje **K-Means Cluster Analysis: Iterate** nurodykite didžiausią iteracijų skaičių (**Maximum Iterations**), kuris gali būti nuo 1 iki 999, ir konvergavimo kriterijų (**Convergence Criterion**), iš kurio sprendžiama, kada turi būti nutrauktos iteracijos. Iteracijos yra nutraukiamos, kada nė vieno klasterio centras nepasikeičia didesniu atstumu, negu nustatyta reikšmė kriterijaus, kuris reiškia mažiausio atstumo tarp klasterių pradinių centrų dalį. Konvergavimo kriterijus turi būti didesnis už 0, bet mažesnis už 1. Spragtelėkite mygtuką **Continue**.
- Spragtelėkite mygtuką **Save...** Atsidarys dialogo langelis **K-Means Cluster: Save New Variables**, kuriame būna pažymėtas laukelis **Cluster membership**, reiškiantis, kad atskirų stebėjimų priklausomybės konkrečiam klasteriui duomenys bus išsaugoti duomenų rinkmenoje kaip atskiras kintamasis. Galite taip pat pažymėti laukelį **Distance from cluster center** ir tuo nurodyti išsaugoti duomenų rinkmenoje kiekvieno stebėjimo Euklido nuotolio nuo klasterio centro reikšmę. Spragtelėkite mygtuką **Continue**.

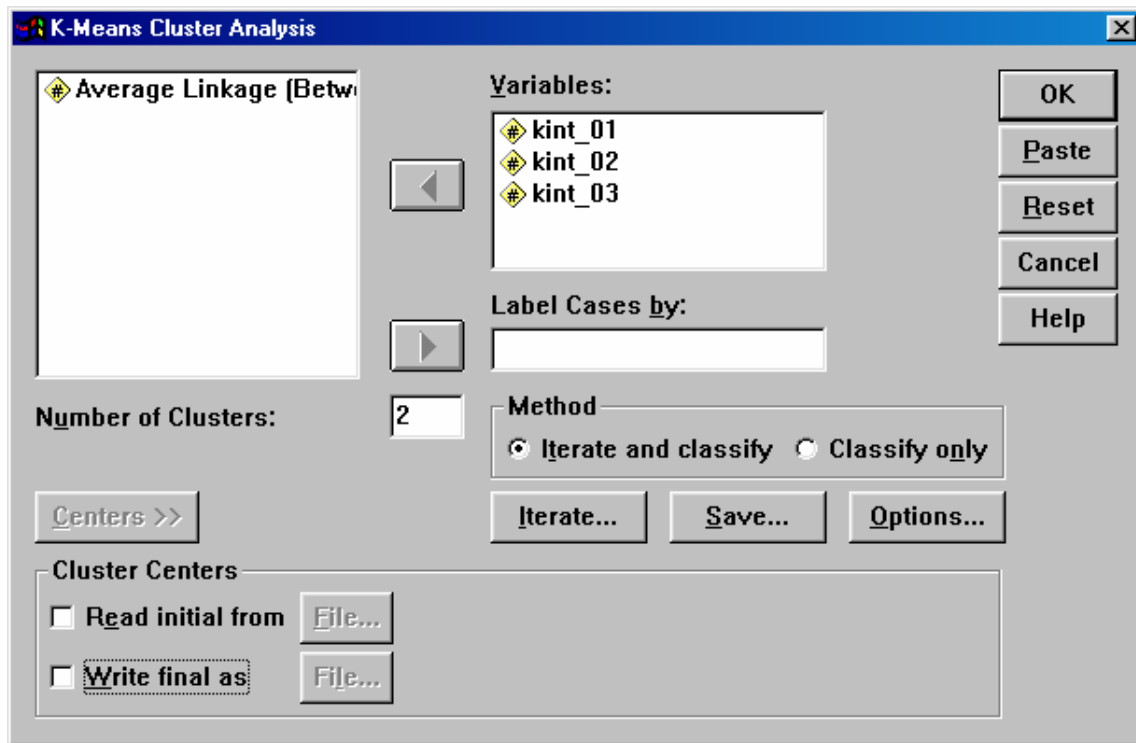


16.7 pav. Dialogo langelis *K-Means Cluster Analysis*

- Spragtelęję mygtuką **Options...**, atsidariusiame dialogo langelyje *Means Cluster Analysis: Options* grupėje *Statistics* galite pasirinkti šias charakteristikas: *Initial cluster centers* (pradiniai klasterių centrai), *ANOVA table* (ANOVA lentelė), *Cluster information for each case* (kiekvieno stebėjimo priklausomybės klasteriui duomenys). Spragtelėkite mygtuką **Continue**.
- Spragtelėkite pagrindinio dialogo langelio *K-Means Cluster Analysis* mygtuką **OK**.

Be išdėstytosios galima taip pat skaičiavimo eiga nustatant pradinis klasterių centrus. Tada:

- Iš daugelio stebėjimų duomenų rinkmenos sudarykite atsitiktinę imtį.
- Dialogo langelyje *K-Means Cluster Analysis* turi būti nurodytas metodas *Iterate and classify*.
- Spragtelėkite mygtuką **Centers>>**, išplėstiniame dialogo langelyje (16.8 pav.) pažymėkite laukelį *Write final as*, spragtelėkite šalia esantį mygtuką **File...** ir dialogo langelyje *K-Means Cluster Analysis: Write to File* įrašykite bylos, kurioje bus išsaugoti duomenys apie klasterių centrų koordinates, vardą ir spragtelėkite šio langelio mygtuką **Save**.
- Spragtelėkite dialogo langelio *K-Means Cluster Analysis* mygtuką **OK**.
- Atlikę pirminę klasterinę analizę su imties duomenimis, atkurkite visą duomenų rinkmeną, dialogo langelyje *K-Means Cluster Analysis* pasirinkite metodą *Classify only*.



16.8 pav. Išplėstinis dialogo langelis *K-Means Cluster Analysis*

- Pažymėkite laukelį **Read initial from** (nuskaityti pirminius duomenis iš), spragtelėkite šalia esantį mygtuką **File...** ir pažymėję dialogo langelyje *K-Means Cluster Analysis: Read from File* bylą, kurioje saugomi duomenys apie klasterių centrų koordinatas, spragtelėkite šio langelio mygtuką **Open**.
- Mygtukais **Save...** ir **Options...** atidarykite atitinkamus dialogo langelius ir pasirinkite reikiamus parametrus.
- Spragtelėkite pagrindinio dialogo langelio *K-Means Cluster Analysis* mygtuką **OK**.

17. GRAFIKAI

Vienas iš SPSS paketo privalumų yra didelė grafikų ir diagramų įvairovė. Grafikai gali būti sudaromi tiek naudojant grafikų meniu komandas, tiek statistinių metodų meniu komandas. Vartotojas gali pasirinkti standartinius grafikus ir vadinamuosius dialoginius (*interactive*) grafikus, kurių grafinės informacijos pateikimo galimybės didesnės.

SPSS paketo standartinių grafikų apžvalga ir pagrindinės charakteristikos yra pateikiamos dialogo langelyje **Main Chart Gallery** (pagrindinių grafikų galerija), kurį galite išsikviesti nurodę meniu komandas **Graphs → Gallery**.

Reikiamam grafiko tipui pasirinkti, nurodykite meniu komandą **Graphs** ir pageidaujamą grafiko tipą:

- **Bar...** (stulpelinė diagrama).
- **Line...** (linijinė diagrama).
- **Area...** (plotinė diagrama).
- **Pie...** (skritulinė diagrama).
- **High-Low...** (didžiausių ir mažiausių reikšmių diagrama).
- **Pareto...** (Pareto diagrama — stulpelinė diagrama, kurioje stulpeliai išdėstomi mažėjančia tvarka, o papildoma kreivė rodo sukauptus dažnius).
- **Control...** (kontrolinė diagrama, kurioje rodomas kintamojo pokytis laike ir atsitiktinių faktorių sąlygojama sklaidos sritis).
- **Boxplot...** (stulpelinė sklaidos diagrama).
- **Error Bar...** (paklaidos stulpelinė diagrama).
- **Scatter...** (sklaidos, arba taškinė diagrama).
- **Histogram...** (histograma).
- **P-P...** (skirstinio normalumo patikrinimo diagrama, kurioje vaizduojama teorinių sukaupųjų dažnių priklausomybė nuo faktinių sukaupųjų dažnių).
- **Q-Q...** (skirstinio normalumo patikrinimo diagrama, kurioje vaizduojama teorinių dažnių priklausomybė nuo faktinių dažnių).
- **ROC Curve...** (specializuotas grafikas, skirtas priklausomybei tarp dviejų testo charakteristikų pavaizduoti).
- **Sequence...** (sekos), **Time Series...** (laiko eilutės).

Pasirinkus grafiko tipą, svarbiausias grafiko sudarymo momentas yra nurodyti atsidariusiame dialogo langelyje reikiamus kintamuosius. Kai kurių standartinių tipų grafikų redagavimo principai buvo apžvelgti 8.3 skyriuje. Kadangi šie principai yra bendri visiems standartinių grafikų tipams, plačiau ties jais neapsistosisime. Juolab kad grafikų sudarymo ir redagavimo dialogo langeliai SPSS yra informatyvūs ir lengvai suvokiami. Todėl pereisime prie dialoginių grafikų sudarymo ir redagavimo principų išdėstymo.

17.1. DIALOGINIAI GRAFIKAI

Pradedant 8 versija į SPSS programų paketą įtraukti vadinamieji dialoginiai (angl. *interactive*) grafikai, kurių galimybės didesnės negu ankstesniųjų, dabar vadinamų standartiniais, grafikų. Dialoginiai grafikai geriausiai tinka nedidelėms duomenų rinkmenoms arba didelių rinkmenų dalims. Kai kurių tipų grafikai (kontrolės diagramos, laiko eilučių diagramos ir kt.) galimi tik standartinio formato. Galima sudaryti šių tipų dialoginius grafikus:

- **Bar...** (stulpelinė diagrama).
- **Dot...** (taškinė diagrama).
- **Line...** (linijinė diagrama).
- **Ribbon...** (juostinė diagrama).
- **Drop-Line...** (juostinė su ryšio linijomis).
- **Area...** (plotinė diagrama).
- **Pie...** (skritulinė diagrama).
- **Boxplot...** (stulpelinė sklaidos diagrama).
- **Error Bar...** (paklaidos stulpelinė diagrama).
- **Histogram...** (histograma).
- **Scatterplot...** (sklaidos, arba taškinė, diagrama).

Remdamiesi pavyzdžiais, panagrinėsime šias sportiniams tyrimams dažnai naudojamas diagramas: stulpelinę, histogramą ir sklaidos diagramą. Kitų diagramų sudarymo ir redagavimo principai yra analogiški.

Dažniausiai diagramomis yra vaizduojami duomenų rinkmenos duomenys, tačiau dialoginės diagramos gali būti sudarytos tiesiogiai iš mobiliųjų lentelių duomenų. Tam reikia dukart spragtelėti pele lentelę ir ją aktyvuoti, pažymėti eilutes, stulpelius ar atskiras ląsteles, kurių duomenis norite pavaizduoti, spragtelėti dešiniu pelės klavišu pažymėtą sritį, kontekstiniame meniu pasirinkti komandą **Create Graph** (sudaryti grafiką) ir pasirinkti diagramos tipą.

17.1.1. Stulpelinės diagramos

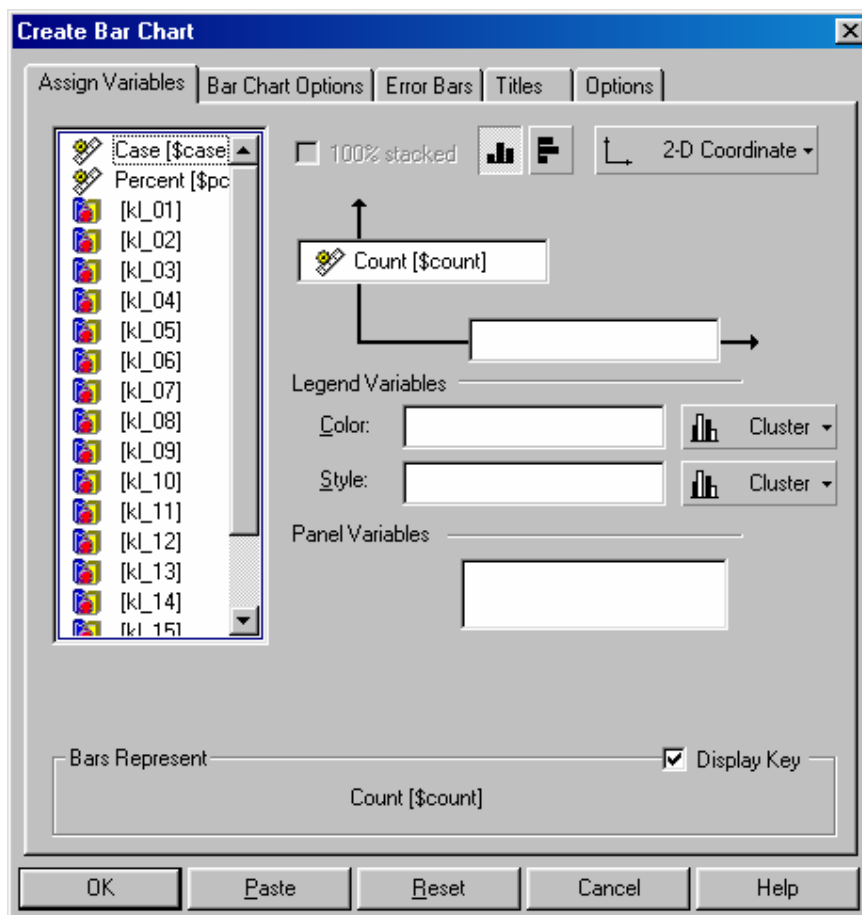
- Atidarykite bylą su analizuojamais duomenimis.
- Nurodykite komandas **Graphs → Interactive → Bar...**

Atsidarys dialogo langelis **Create Bar Chart** (17.1 pav.), kurio struktūra yra bendra sudarant visų tipų diagramas. Dialogo langelyje yra penkios kortelės, kurių pirmoji **Assign Variables** (priskirti kintamuosius) atsidaro iš karto, kai yra atidaromas dialogo langelis. Ją sudaro: kintamųjų sąrašas, laukeliai kintamiesiems nurodyti, diagramos variantų (vertikali ar horizontali) mygtukai, keli išskleidžiamieji sąrašai.

Kintamieji sąrašė gali būti išrikiuoti pagal alfabetą arba pagal kintamųjų tipą. Norėdami pakeisti rikiavimo tvarką, dešiniu pelės klavišu spragtelėkite

vieno iš kintamųjų pavadinimą ir pasirodžiusiame kontekstiniame meniu pasirinkite reikiamą kintamųjų rikiavimo eiliškumą. Tame pačiame kontekstiniame meniu taip pat galite nurodyti, kaip turi būti pateikti kintamieji — pagal savo vardus ar pagal žyvenes. Kadangi žyvenos paprastai būna ilgos ir negali būti visos surašytos į kintamųjų sąrašą, patariama palikti kintamųjų pateikimą pagal vardus.

Sąrašo kintamuosius galima skirstyti į du tipus: kategorinius ir metrinės skalės. Šie du kintamųjų tipai yra identifikuojami skirtingais simboliais, rašomais prieš kintamojo pavadinimą. Kategoriniams kintamiesiems yra priskiriami vardinės arba rangų skalės kintamieji. Kintamojo tipą galima pakeisti spragtelėjus kintamojo pavadinimą dešiniu pelės klavišu.



17.1 pav. Dialogo langelis **Create Bar Chart**

Prie metrinės skalės kintamųjų priklauso taip pat sisteminiai kintamieji *\$count* (dažnis), *\$pct* (procentai) ir *\$case* (stebėjimai), kurie yra naudojami sudarant stulpelines diagramas atitinkamai pagal dažnius, procentinius rodiklius ir atskirus stebėjimus.

Priklausomai nuo to, kokia diagrama bus pasirinkta — dvimatė (**2-D Coordinate**), su erdviniu efektu (**3-D Effect**) ar trimatė (**3-D Coordinate**) — išskleidžiamajame sąraše dialogo langelio viršutiniame dešiniajame kampe bus pateikta $x - y$ ar $x - y - z$ koordinačių ašių schema. Pagal nustatymą y ašies laukelyje figūruoja kintamasis *\$count* (dažnis). Palikę šį pirminį nustatymą, gausite stulpelinę diagramą, vaizduojančią dažnių priklausomybę. Analizuojamas kintamasis turi būti įrašytas į x ašies laukelį.

Visuose mūsų anksčiau nagrinėtuose dialogo langeliuose, norint perkelti kintamąjį iš kintamųjų sąrašo į kurį nors testuojamų kintamųjų laukelį, reikia jį pažymėti spragtelint pele ir tada spragtelėti mygtuką su rodykle, nurodančia perkėlimo kryptį. O dialoginių diagramų dialogo langeliuose kintamieji perkeliama velkant. Nukreipkite pelės žymeklį į kintamąjį ir kai jis igauna rankos simbolį, perkeltkite kintamąjį į reikiamą laukelį. Taigi, į y ašies laukelį galite įkelti kitą sisteminį kintamąjį arba metrinės skalės kintamąjį, o į x ašies laukelį — analizuojamą kintamąjį. Įkeliant į y ašies laukelį metrinės skalės kintamąjį, dialogo langelio apatiniame kairiajame kampe, po užrašu **Bars Represent** (stulpeliai reiškia) atsiranda išskleidžiamasis sąrašas, kuriame galite pasirinkti y ašyje atidedamo kintamojo skaitinę charakteristiką (vidurkį ir t. t.). Nuspausdami klavišą **Ctrl** galite priskirti koordinačių ašiai kelis kintamuosius.

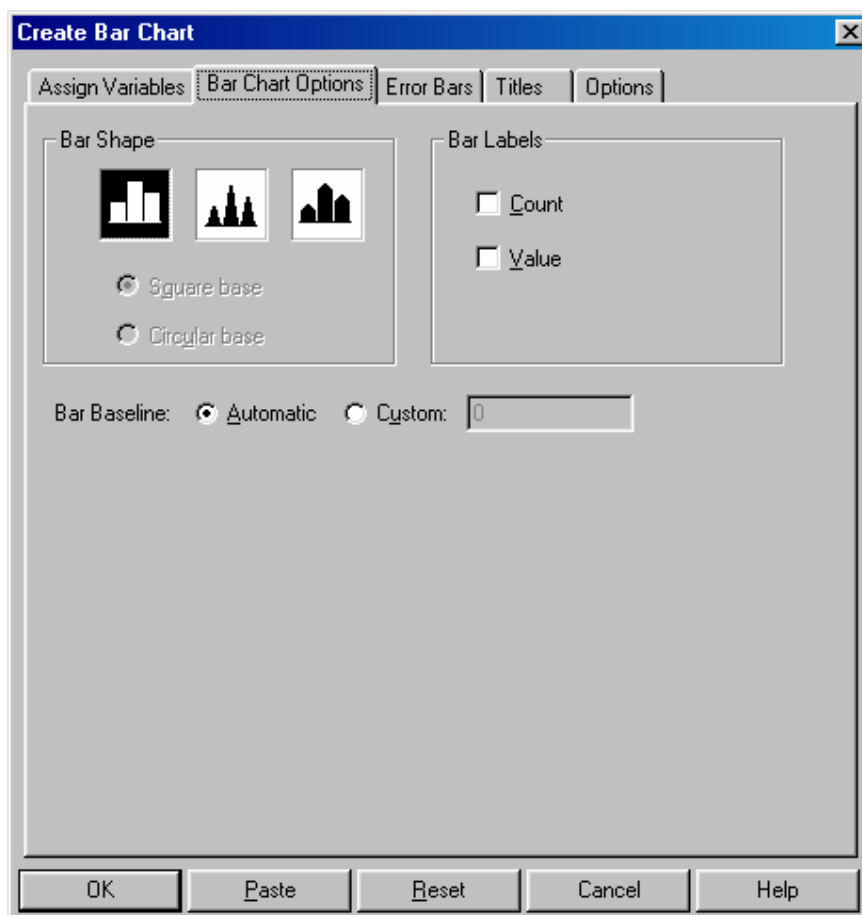
Kai stulpelinėje diagramoje reikia pavaizduoti kelias sekas, yra naudojama legenda, kuri spalviniu ar rašto kodavimu nurodo stulpelių priklausomybę konkrečiai sekai. Kintamąjį, kuris nusako duomenų priklausomybę apibrėžtai sekai, įkelkite į **Legend Variables** laukelį **Color**, jeigu norite sekas skirti spalviniu žymėjimu, arba įkelkite į laukelį **Style**, jeigu norite sekas skirti skirtingais raštais. Abiem atvejais šalia esančiame išskleidžiamajame sąraše turi būti užrašas **Cluster** (grupė). Pasirinkę vietoj **Cluster** variantą **Stack**, gausite sudėtinę diagramą, kurioje kelios sekos bus pavaizduotos viename stulpelyje. Kita galimybė skirti sekas — kintamąjį, kuris nusako duomenų priklausomybę apibrėžtai sekai, reikia įkelti į laukelį **Panel Variables** — bus nubraižytos kiekvienos į šį laukelį įkelto kintamojo reikšmės (kategorijos) atskiros diagramos. Įkėlus į laukelį **Panel Variables** kelis kategorinius kintamuosius, bus nubraižytos kiekvienos šių kintamųjų reikšmių kombinacijos atskiros diagramos.

Dialogo langelio apačioje aktyvus laukelis **Display Key** reiškia, kad diagramoje bus papildomas užrašas, primenantis, ką reiškia stulpeliai. Paprastai šis laukelis daromas neaktyvus.

Dialogo langelio kortelėje **Bar Chart Options** (17.2 pav.) mygtukais **Bar Shape** galite pasirinkti stulpelių formą, o pažymėję laukelį **Square base** (kvadratinis pagrindas) arba **Circular base** (apvalus pagrindas) pasirinkti diagramos su erdviniu efektu (**3-D Effect**) stulpelių pagrindo formą. Aktyvavę **Bar Labels** (stulpelių žymenos) žymimąjį laukelį **Count** (kiekis) arba **Value** (reikšmė), nurodysite įrašyti šias reikšmes stulpeliuose. Aktyvavę **Bar Baseline**

(stulpelių pagrindo linija) žymimąjį laukelį **Custom**, galite nurodyti apatinę stulpelių reikšmę.

Pažymėję dialogo langelio kortelėje **Error Bars** (17.3 pav.) žymimąjį langelį **Display Error Bars**, galite pavaizduoti kiekvieno stulpelio vidurkio pasikliautinąjį intervalą, standartinę nuokrypį ar standartinę paklaidą (kai y ašyje yra pavaizduotas metrinės skalės kintamasis). Mygtukais **Shape** galite pasirinkti paklaidos vaizdavimo formą, o mygtukais **Direction** — vaizdavimo kryptį.

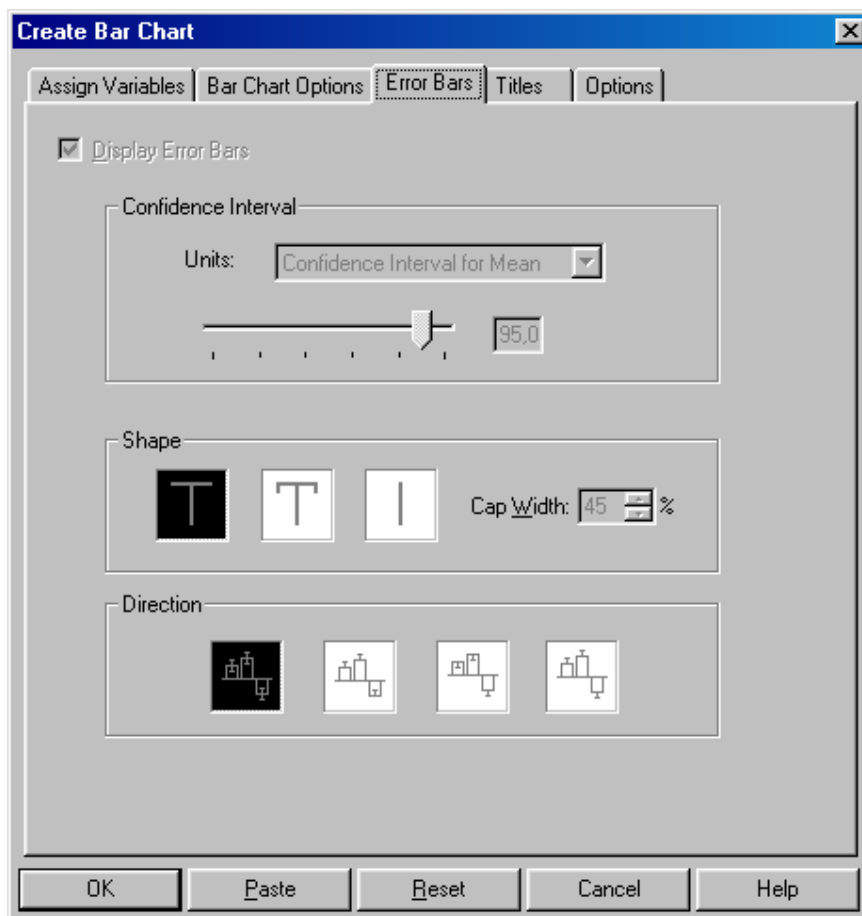


17.2 pav. Dialogo langelio **Create Bar Chart** kortelė **Bar Chart Options**

Kortelės **Titles** laukelyje **Chart Title** galite nurodyti diagramos pavadinimą, laukelyje **Chart Subtitle** — paantraštę, o laukelyje **Caption** — įrašyti komentarus.

Kortelėje **Options** (17.4 pav.) galima nurodyti pasirinktų kategorinių kintamųjų (**Categorical Order** išskleidžiamajame sąrašė **Variable**) rikiavimo kriterijų (**Categorical Order** išskleidžiamajame sąrašė **Order by**) ir rikiavimo tvarką (**Ascending** — didėjančia tvarka, **Descending** — mažėjančia), metrinės skalės kintamųjų (**Scale Range** išskleidžiamajame sąrašė **Variable**) — duomenų

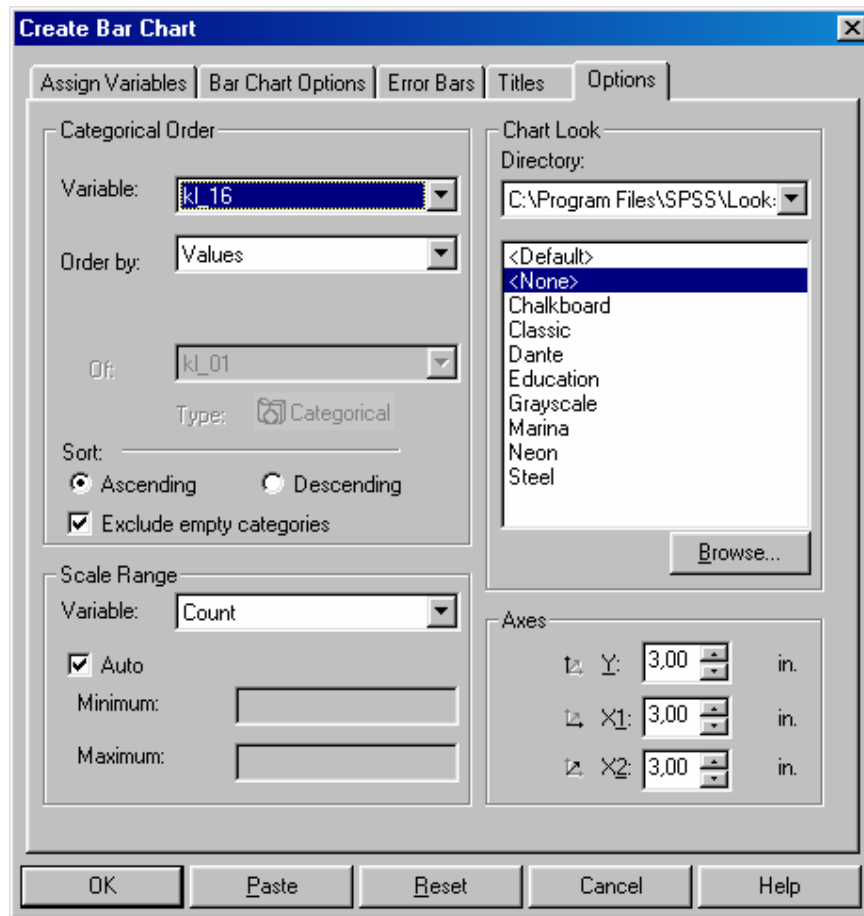
diapazoną (*Minimum*, *Maximum*), pasirinkti nestandartinius diagramos sudarymo formatus, pakeisti koordinačių ašių matmenis.



17.3 pav. Dialogo langelio *Create Bar Chart* kortelė *Error Bars*

Pavyzdžiui, 17.5 pav. pateikiame stulpelinę diagramą, kurioje pavaizduoti grupės respondentų atsakymų į klausimą „Kaip dažnai Jūs sportuojate?“ rezultatai. *X* ašyje atidėtas kategorinis kintamasis, kurio kategorijos yra atsakymų į klausimą variantai, o *y* ašyje — atsakymų dažniai. Dažnių reikšmės nurodomos stulpeliuose (pažymėjus *Bar Chart Options* kortelės *Bar Labels* grupės laukelį *Count*).

Nubraižytą diagramą galima redaguoti, įterpiant naujus arba pašalinant esamus elementus, keičiant ašių užrašus, šrifto tipą ir dydį, linijų tipą, storį ir spalvą, sekų spalvas ir raštus, ašių skales ir k. t., kad būtų pasiektas norimas efektas, t. y. duomenys būtų pavaizduoti kuo aiškiau. Šiame skyriuje panagrinėsime, kaip modifikuoti stulpelinei diagramai būdingus formatus. Plačiau dialoginių diagramų bendri redagavimo principai nagrinėjami 17.2 skyriuje.



17.4 pav. Dialogo langelio **Create Bar Chart** kortelė **Options**

Jeigu norite redaguoti dialoginę diagramą, ją dukart spragtelėkite pele. Diagrama bus aktyvuota, atsiras dialoginių diagramų redagavimo įrankių juostos (17.6 pav.).

Stulpelinės diagramos stulpelių formatui pakeisti:

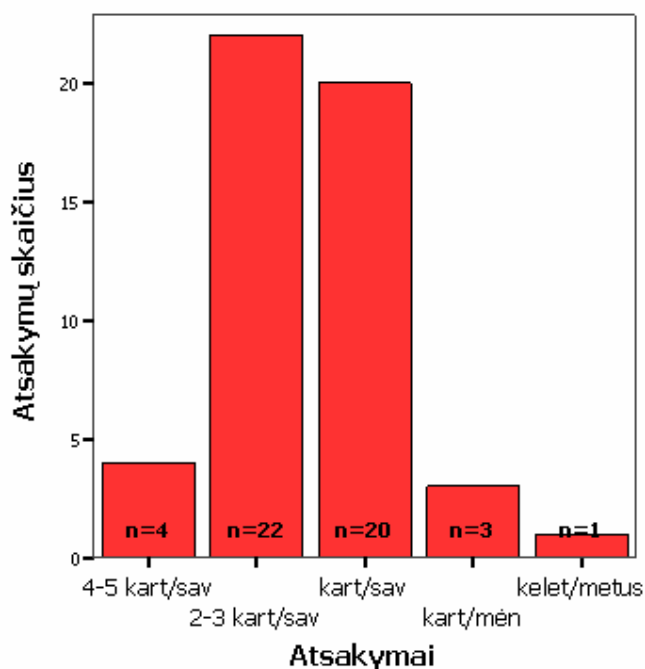
- Nurodykite komandas **Format → Graph Elements → Bar** arba spragtelėkite dešiniu pelės klavišu aktyvuotos diagramos stulpelį ir kontekstiniame meniu pasirinkite **Bars...**
- Atsidariusio dialogo langelio **Bars** (17.7 pav.) kortelėje **Bar Options** galite pasirinkti:
 - Stulpelių formą (**Bar Shape**), taip pat trimatės diagramos (**3-D**) stulpelių pagrindo formą **Square base** (kvadratinis pagrindas) arba **Circular base** (apvalus pagrindas).
 - Stulpelių aukštį atitinkančių dažnių (**Count**) arba stebėjimų reikšmių (**Values**) įrašymo vietą (komandų grupėje **Bar Labels**).
 - Stulpelių ploto spalvą (**Color**) bei raštą (**Style**) — komandų grupėje **Bar Fill**.

- Stulpelių rėmelių spalvą (*Color*), linijų stilių (*Style*) ir storį (*Weight*) — komandų grupėje **Bar Border**.
- Bazinės linijos (**Bar Baseline**) padėtį.
- Dialogo langelio **Bars** kortelėje **Summary Function** galite pasirinkti y ašyje atidėto metrinės skalės kintamojo skaitinę charakteristiką (vidurkį ir k. t.) ir šios skaitinės charakteristikos užrašo padėtį diagramoje.
- Dialogo langelio **Bars** kortelėje **Bar Width** (17.8 pav.) galite pasirinkti stulpelių plotį (**Bar Width**) arba sudėtinei diagramai — stulpelių grupių plotį (**Cluster Width**).

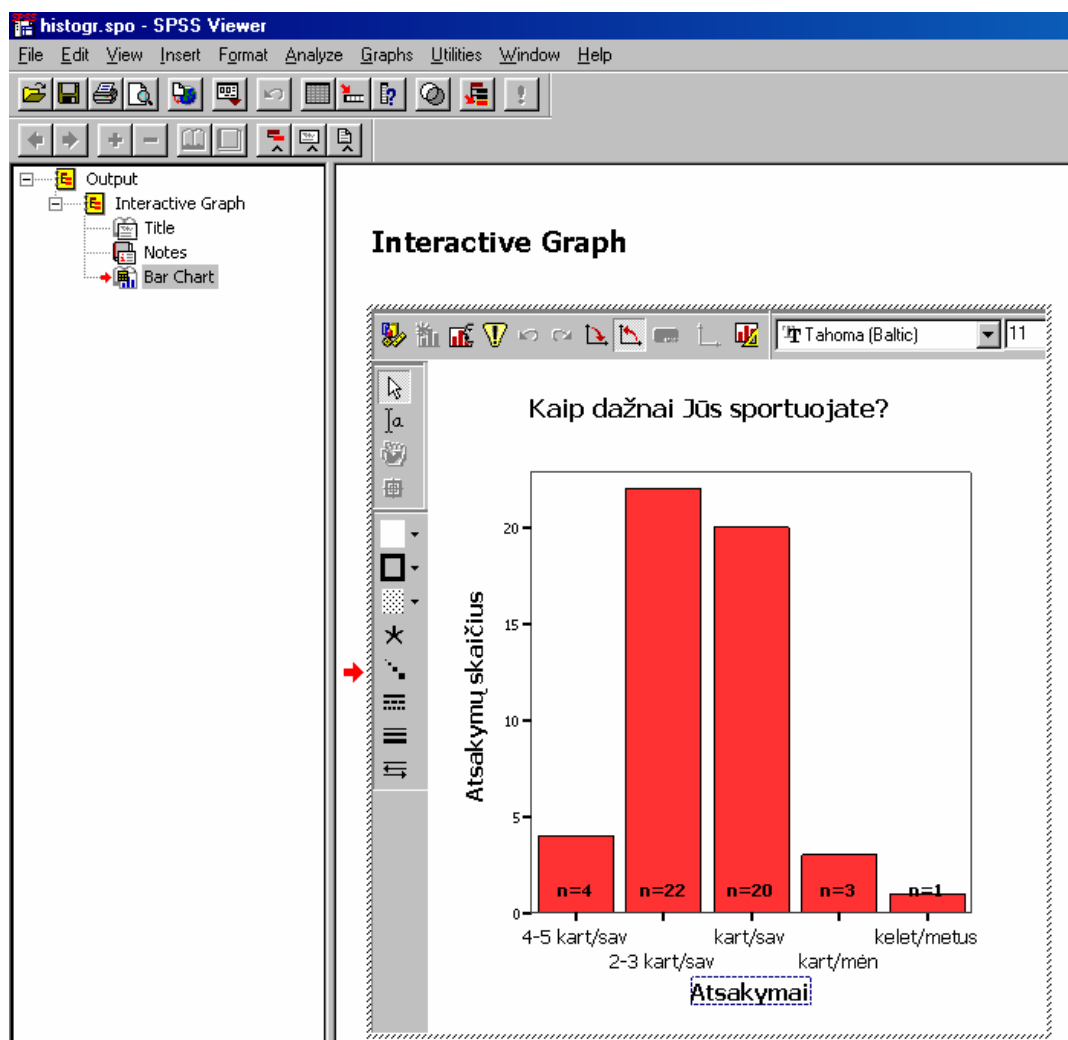
Jeigu norite pakeisti stulpelinės diagramos atskirų stulpelių formatus, kurių nesąlygoja vaizduojami duomenys, t. y. spalvas, rėmelių linijų tipus ir spalvas, stulpelių žymenas:

- Aktyvuokite diagramą dukart spragtelėdami ją pele, pažymėkite vieną ar kelis stulpelius ir nurodykite komandas **Edit → Bar Properties...** arba spragtelėkite dešiniu klavišu pasirinktą stulpelį ar kelis stulpelius ir kontekstiniame meniu pasirinkite **Properties**.
- Atsidariusiame dialogo langelyje **Bar Properties** pasirinkite pageidaujamą stulpelio(-ių) formatą.

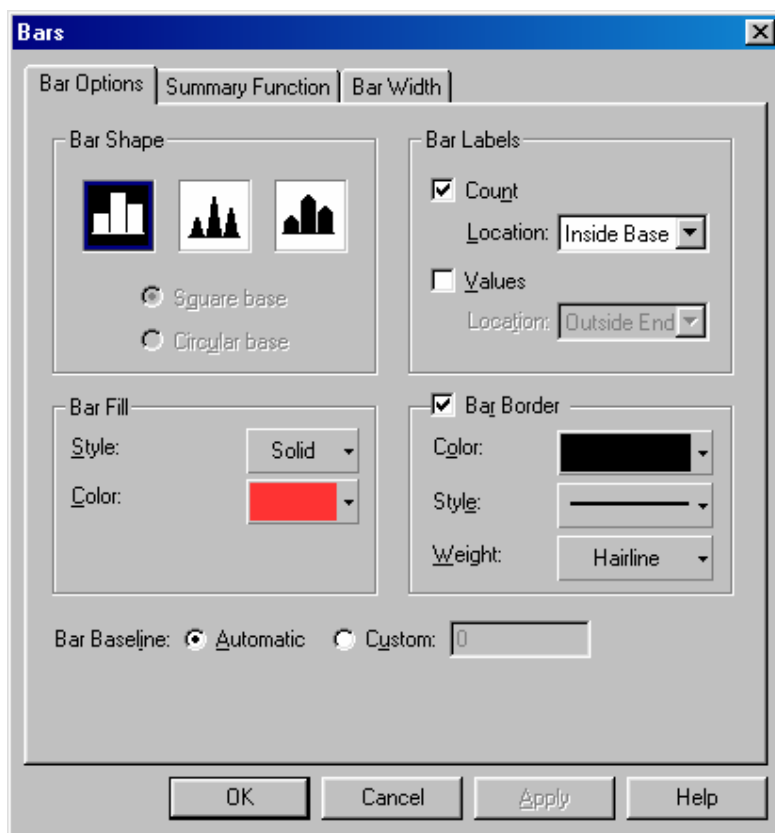
Kaip dažnai Jūs sportuojate?



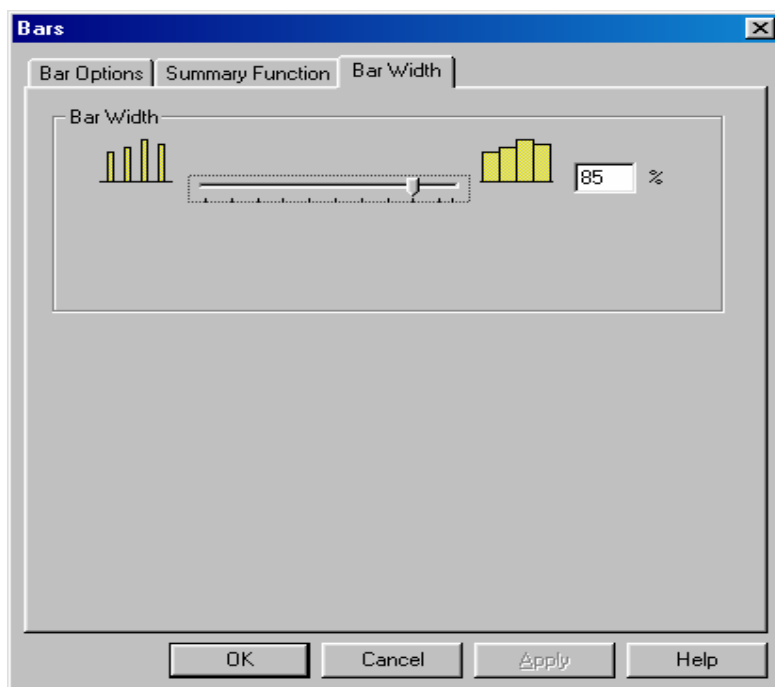
17.5 pav. Stulpelinės diagramos pavyzdys



17.6 pav. Aktyvuota dialoginė stulpelinė diagrama



17.7 pav. Dialogo langelis stulpelinei diagramai redaguoti **Bars**



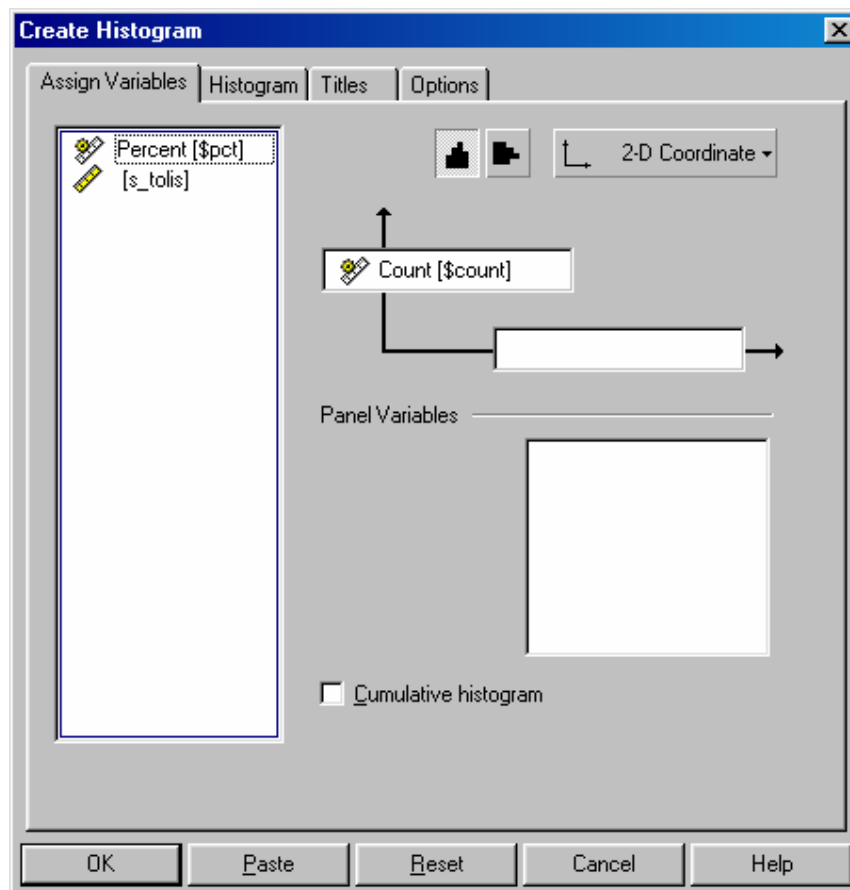
17.8 pav. Dialogo langelio **Bars** kortelė **Bar Width**

17.1.2. Histogramos

Histogramomis vaizduojamas intervalų skalės kintamųjų pasiskirstymas. Kintamojo reikšmės yra suskirstomos į intervalus, nustatomas į šiuos intervalus patenkančių reikšmių skaičius (dažnis) ir rodikliai pavaizduojami stačiakampių stulpelių, kurių kiekvieno aukštis yra proporcingas atitinkamo intervalo duomenų dažniui, seka. Pagal pradinius nustatymus intervalų skaičių ir plotį programa parenka automatiškai, tačiau šiuos parametrus gali nustatyti ir vartotojas.

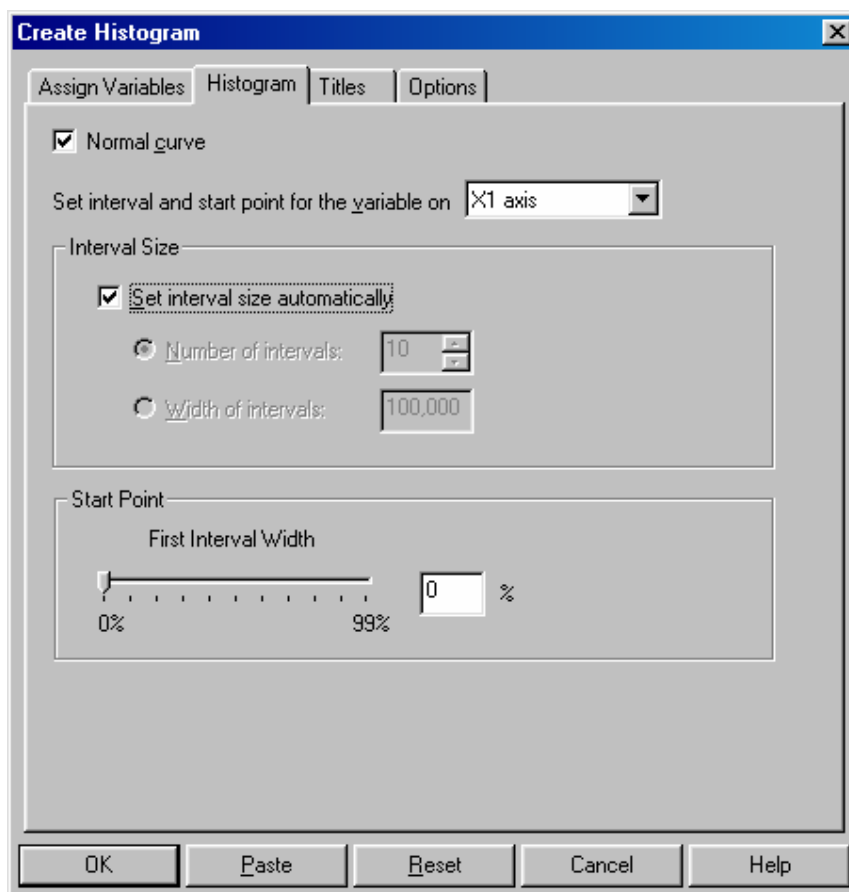
Histogramai nubraižyti:

- Atidarykite bylą su analizuojamais duomenimis.
- Nurodykite komandas **Graphs → Interactive → Histogram...**
- Dialogo langelyje **Create Histogram** (17.9 pav.) įkelkite analizuojamą kintamąjį į x ašies laukelį, o y ašies laukelyje palikite programos nustatytą sisteminių kintamųjų *\$count*. Į laukelį **Panel Variables** galima įkelti kategorinį kintamąjį, pagal kurio reikšmes bus sudarytos atskiros histogramos. Pažymėję laukelį **Cumulative histogram**, gausite sukaupųjų dažnių histogramą.



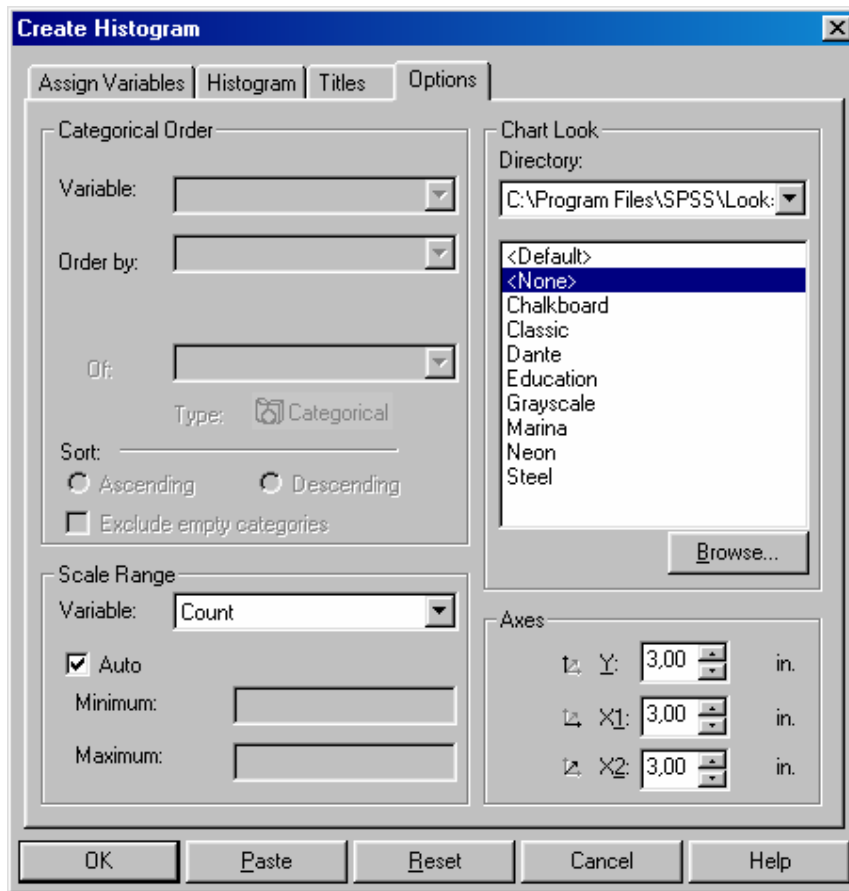
17.9 pav. Dialogo langelis **Create Histogram**

- Atidarykite dialogo langelio **Create Histogram** kortelę **Histogram** (17.10 pav.). Pažymėję laukelį **Normal curve**, galite nurodyti nubraižyti papildomą teorinio normaliojo skirstinio kreivę (pasirinkus trimatę histogramą, normaliojo skirstinio kreivė nevaizduojama), o panaikinę laukelio **Set interval size automatically** (automatinis intervalo dydžio nustatymas) pažymėjimą — patys nurodyti intervalų skaičių laukelyje **Number of intervals** arba intervalų plotį laukelyje **Width of intervals**.



17.10 pav. Dialogo langelio **Create Histogram** kortelė **Histogram**

- Kortelėje **Titles** nurodykite histogramos pavadinimą ir reikiamus komentarus. Kortelėje **Options** (17.11 pav.) galima nurodyti pasirinktų kategorinių kintamųjų (**Categorical Order** išskleidžiamajame sąraše **Variable**) rikiavimo kriterijų (**Categorical Order** išskleidžiamajame sąraše **Order by**) ir rikiavimo tvarką (**Ascending** — didėjančia tvarka, **Descending** — mažėjančia), metrinės skalės kintamųjų (**Scale Range** išskleidžiamajame sąraše **Variable**) — duomenų diapazoną (**Minimum**, **Maximum**), pasirinkti nestandartinius histogramos sudarymo formatus, pakeisti koordinatinių ašių matmenis.



17.11 pav. Dialogo langelio **Create Histogram** kortelė **Options**

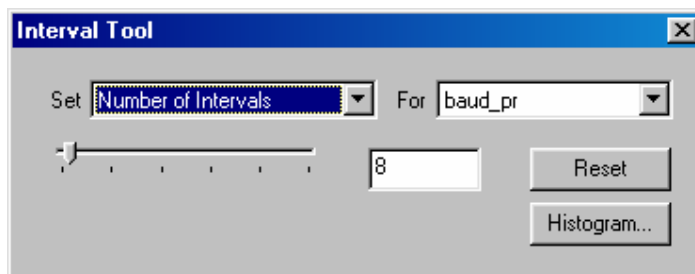
Dialoginei histogramai redaguoti:

- Dukart spragtelėkite ją pele ir nurodykite komandas **Format → Graph Elements → Histogram** arba spragtelėkite histogramą dešiniu pelės klavišu ir kontekstiniame meniu pasirinkite **Histogram...**
- Atsidariusiame dialogo langelyje **Histogram** (17.12 pav.) galite pasirinkti:
 - Komandų grupėje **Fill** — stulpelių ploto spalvą (**Color**) bei raštą (**Style**).
 - Komandų grupėje **Border** — stulpelių rėmelių spalvą (**Color**), linijų stilių (**Style**) bei storį (**Weight**).
 - Sukaupųjų dažnių histogramą (pažymėję laukelį **Cumulative Histogram**).
 - Teorinę normaliojo skirstinio kreivę (**Normal Curve**), kuri bus nubraižyta pagal vaizduojamų duomenų vidurkį ir dispersiją (pasirinkus trimatę histogramą, normaliojo skirstinio kreivė nevaizduojama). Galima keisti skirstinio kreivės spalvą (**Color**), stilių (**Style**) ir storį (**Weight**).
- Spragtelėję dialogo langelio **Histogram** mygtuką **Interval Tool...**, dialogo langelio **Interval Tool** (17.13 pav.) išskleidžiamajame sąrašė **Set** galite

pasirinkti nustatomą parametą: **Number of Intervals** (intervalų skaičius), **Interval Size** (intervalų dydis), **Start Point** (starto taškas) ir nustatyti šio parametro reikšmę. Nustatę reikiamą reikšmę, paspauskite klaviatūros klavišą **Enter**.



17.12 pav. Dialogo langelis **Histogram**



17.13 pav. Dialogo langelis **Interval Tool**

Jeigu norite pakeisti sudėtinės histogramos pasirinktos dalies formatus:

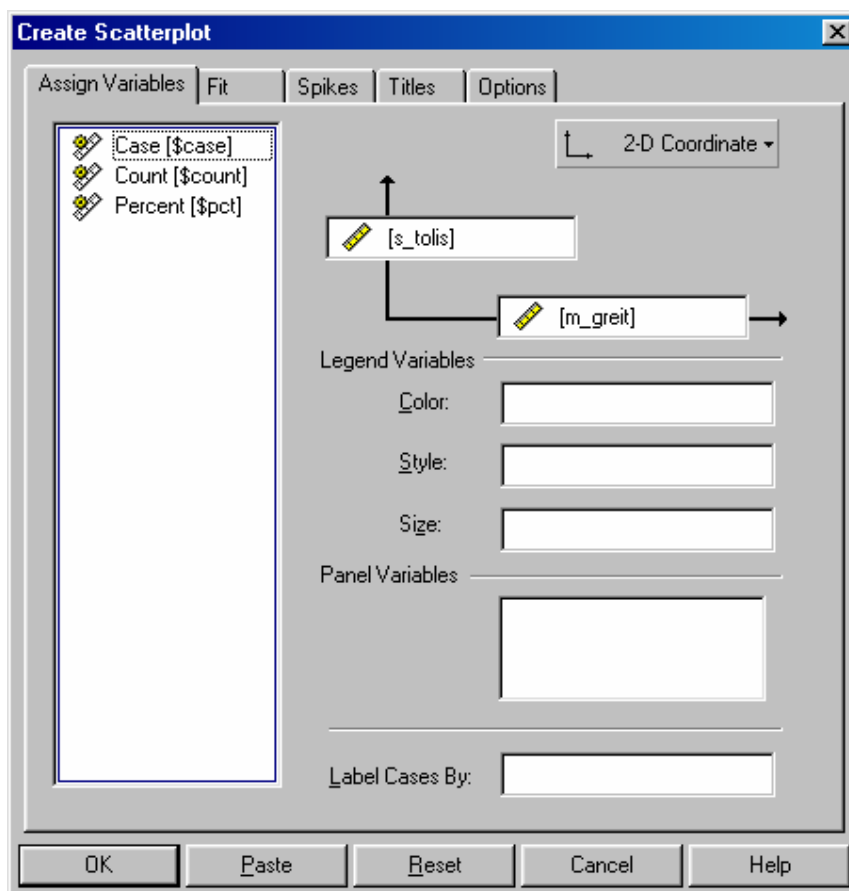
- Pažymėkite histogramą ir nurodykite komandas **Format → Graph Elements → Histogram** arba spragtelėkite dešiniu klavišu pasirinktą histogramą ir kontekstiniame meniu pasirinkite **Properties**.
- Atsidariusiame dialogo langelyje **Histogram Properties** galite pasirinkti histogramos spalvą, rėmelius ir normaliojo skirstinio kreivės parametrus.

17.1.3. Sklaidos diagramos

Sklaidos diagramomis yra vaizduojama priklausomybė tarp dviejų intervalų skalės kintamųjų. Kiekviena duomenų pora stačiakampėje koordinatinių sistemoje yra pažymima tašku, o duomenų visuma — taškų sanakaupa. Galimos ir trimatės sklaidos diagramos, nors jas sunkiau interpretuoti. Sklaidos diagramų analizė padeda pritaikyti adekvačią matematinę išraišką duomenų priklausomybei apibūdinti, aptikti smarkiai besiskiriančias nuo kitų ir gal būt klaidingas reikšmes.

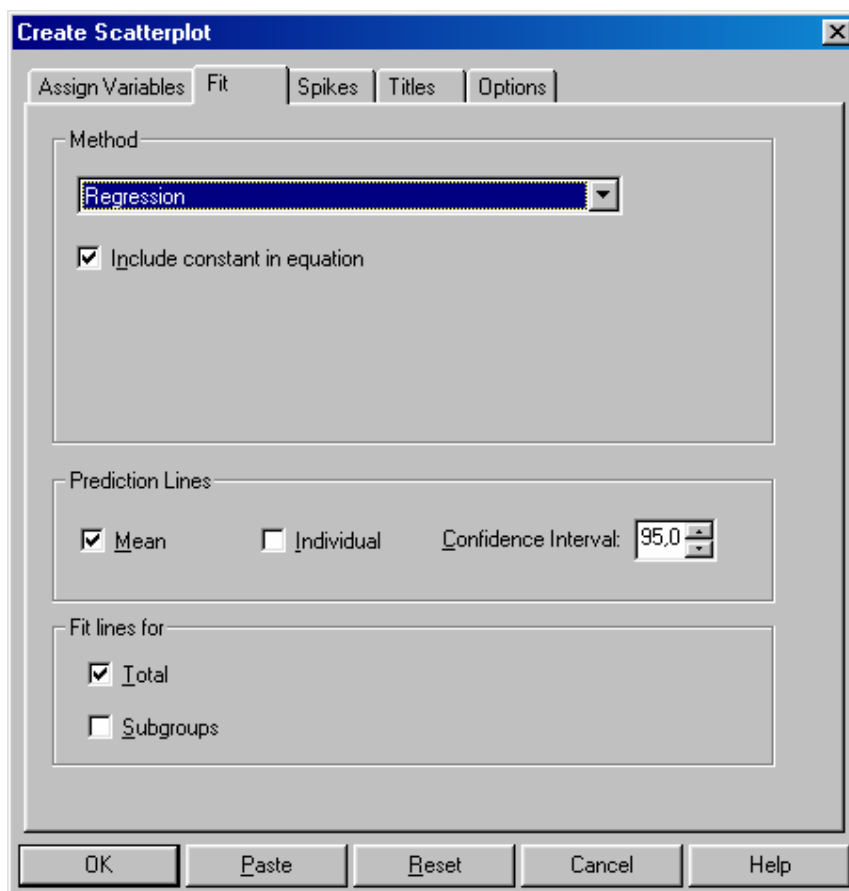
Sklaidos diagramai nubraižyti:

- Atidarykite bylą su analizuojamais duomenimis.
- Nurodykite komandas **Graphs → Interactive → Scatterplot...**
- Dialogo langelyje **Create Scatterplot** (17.14 pav.) vieną kintamąjį (dažniausiai tai nepriklausomas kintamasis) įkelkite į x ašies laukelį, o kitą kintamąjį (dažniausiai tai priklausomas kintamasis) — į y ašies laukelį. Trimatėje sklaidos diagramoje (pasirinkę **3-D Coordinate** variantą) trečią kintamąjį įkelkite į z ašies laukelį. Jeigu sklaidos diagramoje norite išskirti atskiras duomenų grupes, turite tris galimybes:
 - Duomenų priklausomybę apibrėžtai grupei nusakantį kategorinį kintamąjį įkelkite į **Legend Variables** laukelį **Color** — duomenų grupės bus išskirtos spalviniu žymėjimu.
 - Duomenų priklausomybę apibrėžtai grupei nusakantį kategorinį kintamąjį įkelkite į **Legend Variables** laukelį **Style** — duomenų grupės bus pavaizduotos skirtingais simboliais.
 - Duomenų priklausomybę apibrėžtai grupei nusakantį kategorinį kintamąjį įkelkite į **Legend Variables** laukelį **Size** — duomenų grupės bus pavaizduotos skirtingo dydžio simboliais.
- Duomenų priklausomybę apibrėžtai grupei nusakantį kategorinį kintamąjį įkėlę į laukelį **Panel Variables**, gausite atskiras diagramas kiekvienai šio kintamojo kategorijai. Įkėlus į laukelį **Panel Variables** kelis kategorinius kintamuosius, atskiros diagramos bus nubraižytos kiekvienai šių kintamųjų reikšmių kombinacijai.
- Jeigu norite nubraižyti sklaidos diagramą, kurioje būtų vaizduojamos vieno kintamojo stebėjimų reikšmės, įkelkite šį kintamąjį į y ašies laukelį, o sisteminį kintamąjį **Case** — į x ašies laukelį.



17.14 pav. Dialogo langelis **Create Scatterplot**

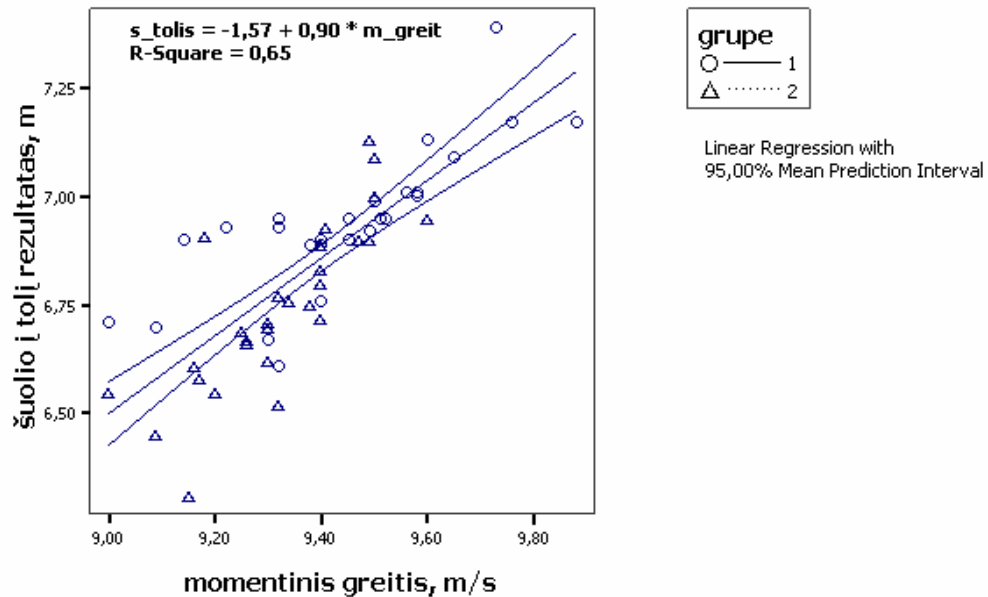
- Jeigu norite, kad stebėjimai (diagramos taškai) turėtų pavadinimus, kintamąjį, kuris nusako atskirų stebėjimų pavadinimus, įkelkite į laukelį **Label Cases By**. Tačiau tai rekomenduojama tik esant nedideliame stebėjimų skaičiui. Priešingu atveju, stebėjimų žymenos uždengs viena kitą ir bus neįskaitomos.
- Jeigu norite pavaizduoti sklaidos diagramoje regresijos tiesę, atidarykite dialogo langelio **Create Scatterplot** kortelę **Fit** (17.15 pav.) ir išskleidžiamajame sąraše **Method** pažymėkite **Regression**. Grupėje **Prediction Lines** (prognozės kreivės) nurodykite variantą **Mean** (vidurkis), o grupėje **Fit Lines for** (pritaikyti kreives) nurodykite, kam brėžiama regresijos tiesė — visoms duomenų grupėms bendrai (**Total**) ar atskirai kiekvienai grupei (**Subtotal**).



17.15 pav. Dialogo langelio **Create Scatterplots** kortelė **Fit**

- Kortelėje **Spikes** (smaigai) galite pasirinkti įvairius taškų sujungimo linijomis su koordinacių pradžia, ašimis ir t. t. variantus. Kortelėje **Titles** įrašykite reikiamus diagramos pavadinimus ir komentarus. Kortelėje **Options** galite nurodyti pasirinktų kategorinių kintamųjų (**Categorical Order** išskleidžiamajame sąrašė **Variable**) rikiavimo kriterijų (**Categorical Order** išskleidžiamajame sąrašė **Order by**) ir rikiavimo tvarką (**Ascending** — didėjančia tvarka, **Descending** — mažėjančia), metrinės skalės kintamųjų (**Scale Range** išskleidžiamajame sąrašė **Variable**) duomenų diapazoną (**Minimum**, **Maximum**), pasirinkti nestandartinius histogramos sudarymo formatus, pakeisti koordinacių ašių matmenis.

Pavyzdžiai. 17.16 pav. parodyta sklaidos diagrama, kurioje pavaizduota dviejų sportininkų grupių šuolio į tolį priklausomybė nuo momentinio įsibėgėjimo greičio. Abiejų grupių duomenys pavaizduoti skirtingais simboliais. Diagramoje pavaizduota bendra abiem grupėms regresijos tiesė ir linijos, žyminčios pasikliautinojo intervalo ribas.



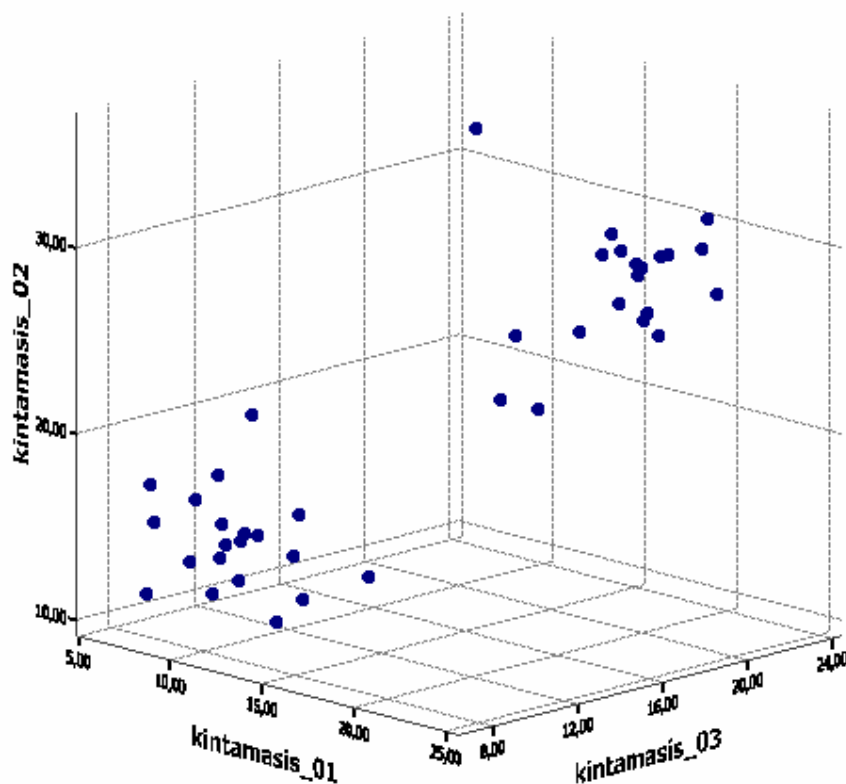
17.16 pav. Dviejų duomenų grupių sklaidos diagramos su regresijos tiesės pavyzdys

17.17 pav. yra pateiktas trimatės sklaidos diagramos pavyzdys.

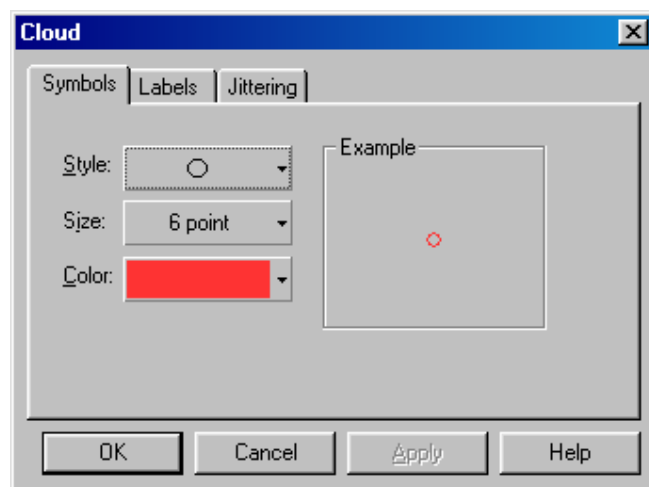
Sklaidos diagramoje duomenis žymintiems simboliams pakeisti:

- Dukart spragtelėkite diagramą pele ir nurodykite komandas **Format** → **Graph Elements** → **Cloud** arba spragtelėkite diagramą dešiniu pelės klavišu ir kontekstiniame meniu pasirinkite **Cloud...**
- Atsidariusiame dialogo langelyje **Cloud** (17.18 pav.) galite pasirinkti:
 - Kortelėje **Symbols** — simbolių tipą (išskleidžiamajame sąraše **Style**), simbolių dydį (išskleidžiamajame sąraše **Size**) ir simbolių spalvą (išskleidžiamajame sąraše **Color**); negalima keisti tų simbolių parametrų, kurie legendoje nusako duomenų priklausomybę vienai ar kitai kategorijai;
 - Kortelėje **Labels** (17.19 pav.) — rodyti ar nerodyti (priklausomai ar pažymėsite laukelį **Display labels**) stebėjimų žymenas, kurias nusako dialogo langelio **Create Scatterplot** laukelyje **Label Cases By** nurodytas kintamasis; jei kintamasis nėra nurodytas, stebėjimai bus pažymėti eilės numeriu; išskleidžiamajame sąraše **Location** galite pasirinkti žymenų padėtį, išskleidžiamajame sąraše **Connector** — jungiamųjų linijų tarp žymenos ir simbolio tipą, o išskleidžiamajame sąraše **Color** — šių linijų spalvą;
 - Kortelėje **Jittering** (17.20 pav.) — duomenis vaizduojančių simbolių išsibarstymo laipsnį. Šis išsibarstymas dirbtinai sukūriamas tam, kad

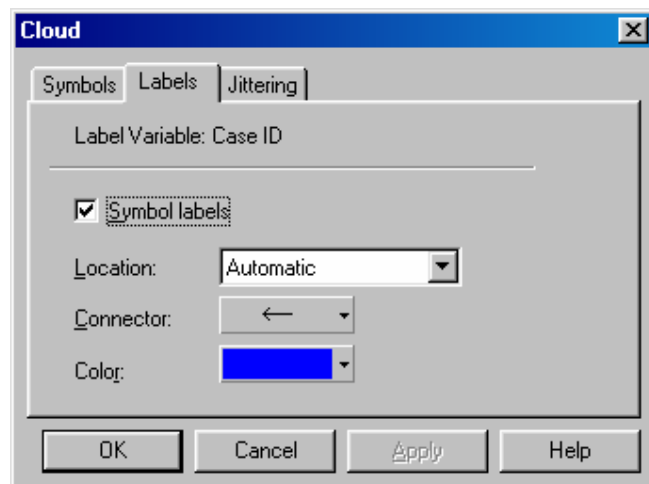
vienodus duomenis vaizduojantys simboliai visiškai neuždengtų vienas kito.



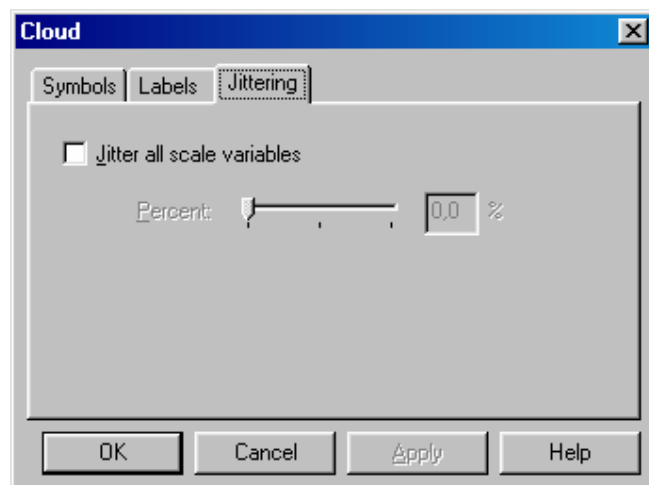
17.17 pav. Trimatės sklaidos diagramos pavyzdys



17.18 pav. Dialogo langelio **Cloud** kortelė **Symbols**



17.19 pav. Dialogo langelio **Cloud** kortelė **Labels**



17.20 pav. Dialogo langelio **Cloud** kortelė **Jittering**

17.2. DIALOGINIŲ GRAFIKŲ REDAGAVIMAS

Sukurtą diagramą galima redaguoti, suteikiant jos elementams norimą stilių, dydį ir spalvą, parenkant šriftą, fono spalvas ir t. t. Toliau pateikiamos dialoginių diagramų redagavimo galimybės įgalina:

- Keisti pavadinimus.
- Koreguoti ašių formatus.
- Keisti legendas.
- Keisti duomenų sekų dydį ir spalvas.
- Naudotis diagramų vedlio (*Chart Manager*) galimybėmis redaguojant ar šalinant diagramos objektus.
- Pertvarkyti diagramos objektus.
- Keisti diagramos objektų stilių ir spalvą.
- Įterpti arba pašalinti naujus kintamuosius ir kitus elementus.

Redaguojama diagrama pirmiausia turi būti aktyvuota, dukart spragtelint ją pele. Diagramą aktyvavus, pasikeičia kai kurios meniu komandos, o diagramos kairėje ir viršuje atsiranda diagramos redagavimo įrankių juostos.

17.2.1. Kaip pažymėti keičiamus diagramos objektus?

Diagramos elementai gali būti pažymimi atskirai arba grupėmis.

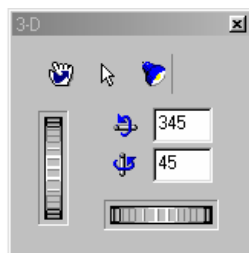
- Spragtelėkite dešiniu pelės klavišu pasirinktą aktyvuotos diagramos elementą (stulpelį, sklaidos diagramos tašką ir pan.). Kontekstinio meniu apačioje bus užrašytas objekto pavadinimas (*Bars*, *Cloud* ir t. t.) ir užrašas *Properties* (savybės). Jeigu norite keisti visus pasirinkto objekto elementus (pavyzdžiui, visus sekos stulpelius), pasirinkite objekto pavadinimą, o jeigu — tik pasirinktą elementą (arba kelis elementus), pasirinkite *Properties*. Jeigu norite keisti apibrėžtos kategorijos objektus, pažymėkite tą kategoriją legendoje.
- Arba
- Dukart spragtelėkite pele pasirinktą aktyvuotos diagramos objektą. Atsidarys dialogo langelis pasirinkto objekto elementams redaguoti.

17.2.2. Kaip perkelti diagramos objektus?

- Aktyvuokite diagramą dukart spragtelėdami ją pele.
- Nuvilkite pele pasirinktą objektą (legendą, diagramos pavadinimą, ašių pavadinimus, pastabas) į naują padėtį.

17.2.3. Kaip modifikuoti trimates diagramas?

Trimatei diagramai modifikuoti yra skirta specializuotų įrankių grupė (*3-D Palette*), kuri pasirodo aktyvavus trimatę diagramą (17.21 pav.).



17.21 pav. Įrankiai trimatei diagramai modifikuoti



Šviesos šaltinis (*Light source*). Spragtelėję šį įrankį, galite pasirinkti menamo šviesos šaltinio padėtį ir kartu — tinkamą diagramos išvaizdą;



Sukimo ranka (*Rotation Hand*). Spragtelėję šį įrankį ir nutaikę rankos simbolį į apibrėžtą diagramos vietą, galite sukinėti diagramą apie vertikalią ir horizontalią ašį ir taip keisti stebėtojo žiūrėjimo kampą diagramos pagrindo atžvilgiu.



Sukimo skalė (*Rotation Dial*). Sukdami pele stačią arba gulsčią skritulį, galite sukinėti diagramą apie vertikalią ir horizontalią ašį ir taip keisti stebėtojo žiūrėjimo kampą diagramos pagrindo atžvilgiu.



Įrankis, kuriuo galite iškviesti įrankių grupę **3-D Palette**, jeigu pastaroji nepasirodo aktyvavus trimatę diagramą.

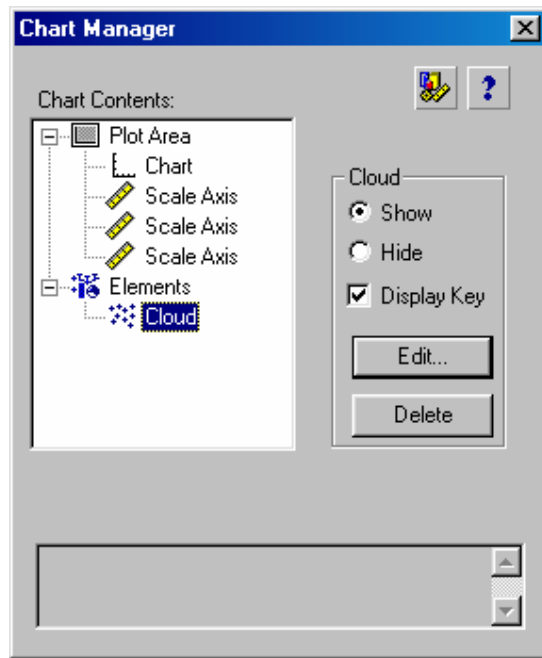
17.2.4. Diagramų vedlys

Diagramų vedlys (*Chart Manager*) užtikrina patogų ir paprastą priėjimą prie diagramos elementų, kurie gali būti atrinkti, o paskui — keičiami, paslėpti ar pašalinami.

- Diagramų vedliui išsikviesti, nurodykite komandas **Edit → Chart Manager** arba spragtelėkite pele diagramų vedlio įrankį



- Dialogo langelio **Chart Manager** (17.22 pav.) sąrašė pasirinkite reikiamą grafiko komponentą, kurį galite paslėpti (**Hide**), pašalinti (**Delete**) arba kurio parametrus galite keisti vadovaudamiesi naujo dialogo langelio, kuris pasirodo spragtelėjus mygtuką **Edit...** (redaguoti), nuorodomis.



17.22 pav. Dialogo langelis **Chart Manager**

17.2.5. Kaip įterpti į diagramą papildomus elementus?

Į diagramą galima įterpti naujus elementus, suteikiančius papildomos informacijos. Pavyzdžiui, stulpelinėje diagramoje galima papildomai įterpti vidurkio pasikliautinąjį intervalą vaizduojančius elementus, linijinėje diagramoje — vidurkį ar medianą vaizduojančias linijas, sklaidos diagramoje — regresijos liniją ir t.t.

Norėdami įterpti į diagramą papildomus elementus:

- Aktyvuokite diagramą, dukart ją spragtelėdami pele.
- Nurodykite komandas **Insert** → (nurodykite įterpiamą elementą) arba spragtelėkite pele įrankį **Insert Element** (įterpti elementą)



ir iš sąrašo pasirinkite reikiamą elementą.

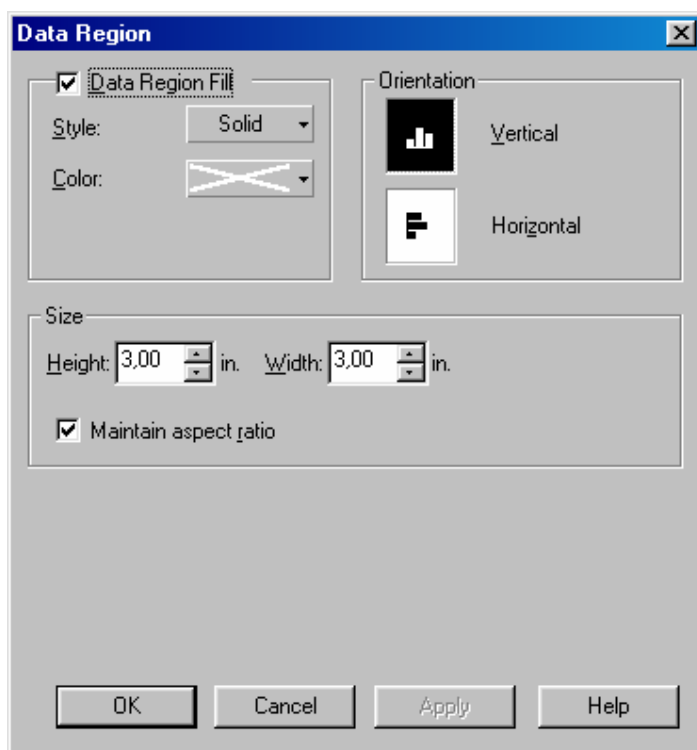
Į diagramą turi būti įterpiami tik tie elementai, kurie yra suderinami su diagramoje pateikiama informacija. Pavyzdžiui, negalima pavaizduoti pasikliautinąjo intervalo stulpelinėje diagramoje, kurioje stulpeliai reiškia ne vidurkius, o paprastus stebėjimus ar dažnius.

17.2.6. Kaip tvarkyti diagramos duomenų sritį?

Dvimatės diagramos duomenų sritis yra plotas, o trimatės diagramos — erdvė, kurioje vaizduojami duomenys. Kiti diagramos objektai yra išdėstomi duomenų srities atžvilgiu.

Dvimatės (2-D) diagramos duomenų srities išvaizdai pakeisti:

- Aktivuoskite diagramą, dukart ją spragtelėdami pele.
- Nurodykite komandas **Format → Data region...** arba spragtelėkite diagramų vedlio **Chart Manager** įrankį, dialogo kotelėje **Chart Manager** pažymėkite **Chart** ir spragtelėkite mygtuką **Edit...**
- Dialogo kortelėje **Data region** (17.23 pav.) galite pasirinkti:
 - Duomenų srities foną (**Data Region Fill**) — mygtukais **Style** (Stilius) ir **Color** (Spalva).
 - Duomenų srities dydį (**Size**) — suktukais **Height** (aukštis) ir **Width** (plotis). Pažymėjus laukelį **Maintain aspect ratio** (išlaikyti krypties proporcijas), vieno matmens pakeitimas automatiškai sukelia kito matmens pasikeitimą.



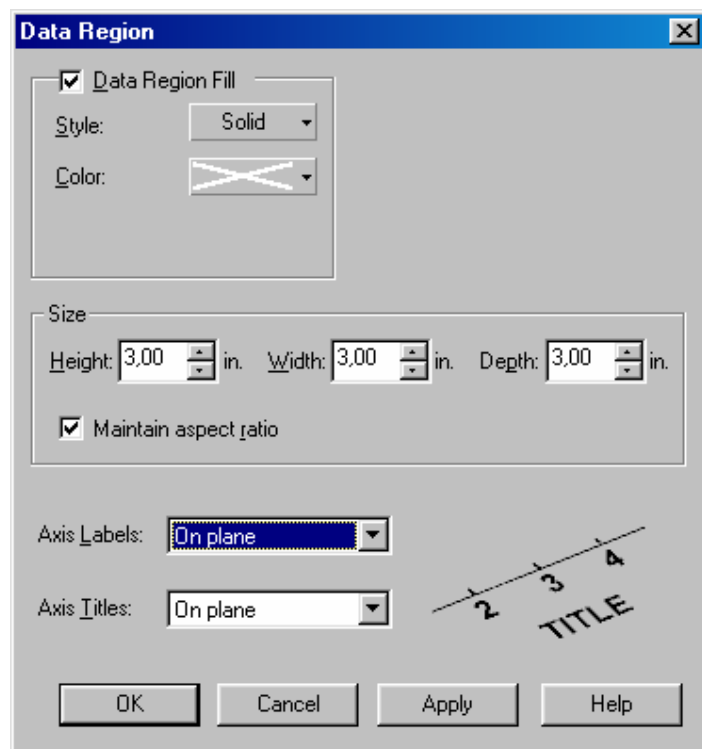
17.23 pav. Dialogo langelis **2-D Data Region**

Trimatės (3-D) diagramos duomenų srities išvaizdai pakeisti:

- Nurodykite komandas **Format → Data Region...** arba spragtelėkite diagramų vedlio **Chart Manager** įrankį, dialogo kotelėje **Chart Manager** pažymėkite **Chart** ir spragtelėkite mygtuką **Edit...**
- Dialogo kortelėje **Data Region** (17.24 pav.) galite pasirinkti:
 - Duomenų srities foną (**Data Region Fill**) — mygtukais **Style** (Stilius) ir **Color** (Spalva);
 - Duomenų srities dydį (**Size**) — suktukais **Height** (aukštis), **Width** (plotis) ir **Depth** (gylis). Pažymėjus laukelį **Maintain aspect ratio**

(išlaikyti krypčių proporcijas), vieno matmens pakeitimas automatiškai sukelia kitų matmenų pasikeitimą.

- Ašių padalų žymenų (*Axis Labels*) ir ašių pavadinimų (*Axis Titles*) orientaciją — *On Plane* (palei ašis) ar *Facing Front* (priekyje).



17.24 pav. Dialogo langelis *3-D Data Region*

17.2.7. Kaip keisti diagramos objektų planą?

Diagramos objektai (pavadinimai, paantraštės, pastabos, legendos ir esmėžodžiai) gali būti įvairiai išdėstyti aplink duomenų sritį. Kai diagrama yra sukurta, galite tuos objektus nutempti pele į naujas pozicijas. Norėdami grąžinti juos į nustatytą pradinę padėtį, nurodykite komandas **Format → Arrange All**.

17.2.8. Kaip keisti diagramos legendą?

Duomenų priklausomybė vienai ar kitai grupei (kategorijai) nurodoma naudojant skirtingų spalvų, dydžių ir stilių simboliai. Spalvos ir dydžiai gali būti naudojami tiek diskrečioms kategorijoms, tiek ištisinės skalės dydžiams skirti. Spalvų, dydžių ar stilių visuma yra iššifruojama vadinamoje diagramos legendoje. Atitinkami kintamieji legendai yra priskiriami pagrindiniame diagramos sudarymo dialogo langelyje, tačiau tai galima padaryti ir modifikuojant diagramą.

Jeigu legenda yra kuriama remiantis kategoriniu kintamuoju, duomenys diagramoje yra suskirstomi į atskiras skirtingų spalvų, dydžių ar stilių simbolių grupes, kurias sąlygoja tas kategorinis kintamasis. Priklausomai nuo diagramos tipo vienai diagramai gali būti sukurta iki trijų legendų pagal tris kintamuosius.

Jeigu legenda yra kuriama remiantis metrinės skalės kintamuoju, duomenys diagramoje yra vaizduojami tolygiai besikeičiančios spalvos ar dydžio simboliais.

Jeigu norite perkelti legendą į kitą vietą:

- Aktyvuokite diagramą, ją dukart spragtelėdami pele.
- Pele nuvilkite legendą į reikiamą vietą.

Jeigu norite pakeisti legendos savybes:

- Aktyvuokite diagramą, ją dukart spragtelėdami pele.
- Nurodykite komandas **Format → Data Legend** ir iš sąrašo pasirinkite **Color Legend**, **Style Legend** ar **Size Legend** — sąrašė nurodomų legendos tipų skaičius priklauso nuo to, pagal kelis kintamuosius buvo sukurta legenda.

Arba

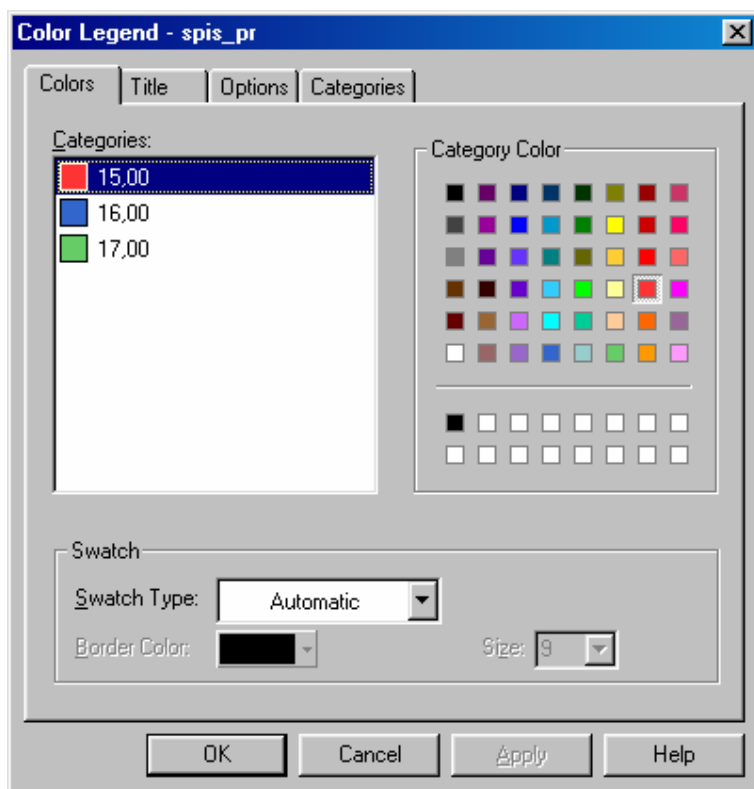
- Dukart spragtelėkite legendos plote pele.

Arba

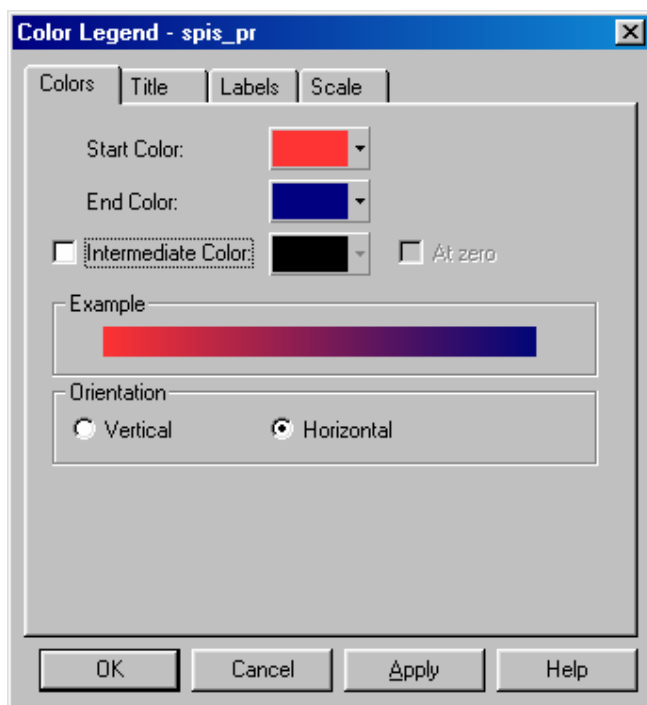
- Atidarykite **Chart Manager** dialogo langelį, pažymėkite legendą ir spragtelėkite šio dialogo langelio mygtuką **Edit**.

Panagrinėsime legendų, skirstančių duomenis pagal spalvas (**Color Legend**), modifikavimo principus. Legendų, skirstančių duomenis kitais pagrindais (**Style Legend**, **Size Legend**), modifikavimo principai analogiški.

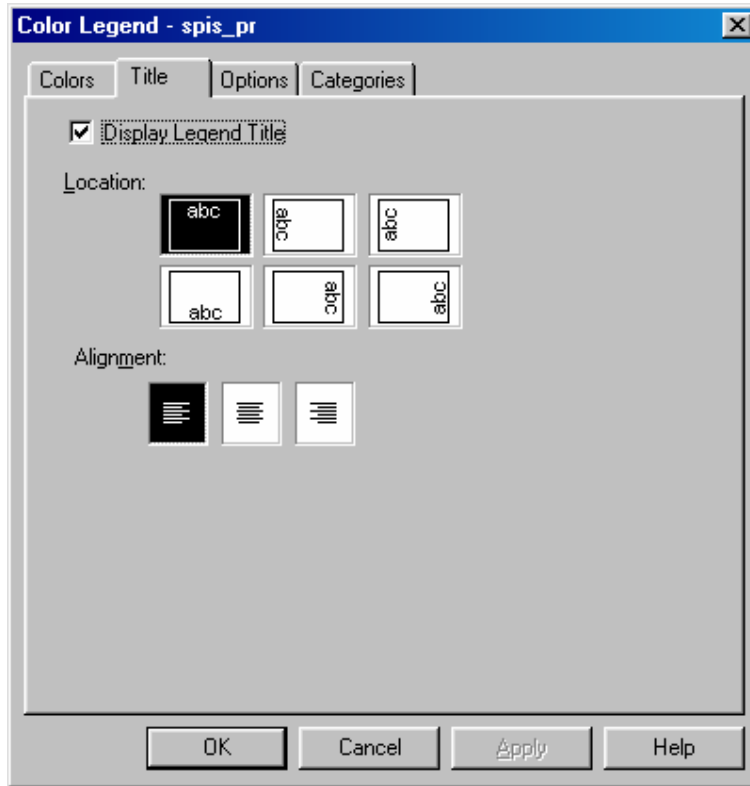
- Atidarykite vienu iš anksčiau nurodytų būdu dialogo langelį **Color Legend** (17.25 pav.).
- Dialogo langelio kortelėje **Colors** pažymėkite kategoriją (**Categories**), kurios spalvą norite pakeisti, ir paletėje **Category Colors** pasirinkite reikiamą spalvą. Išskleidžiamajame sąrašė **Swatch Type** (pavyzdžiai) galite pasirinkti pageidaujamą legendos simbolio tipą, sąrašė **Border Color** — simbolio kraštelio spalvą, o sąrašė **Size** — simbolio dydį.
- Jeigu spalvas sąlygoja skalės tipo kintamasis, kortelė **Colors** yra kitokia (17.26 pav.). Nurodykite joje pradinę spalvą (**Start Color**) ir galinę spalvą (**End Color**), taip pat pasirinkite spalvų kitimo kryptį — vertikalią ar horizontalią.
- Dialogo langelio kortelėje **Title** (17.27 pav.) galite pasirinkti legendos antraštės vietą ir užrašo lygiavimo tipą. Jeigu norite keisti legendos tekstą, jį dukart spragtelėkite pele.



17.25 pav. Dialogo langelio **Color Legend** kortelė **Colors** diskrečioms kategorijoms



17.26 pav. Kortelė **Colors** tolydžių spalvų kategorijoms



17.27 pav. Dialogo langelio **Color Legend** kortelė **Title**

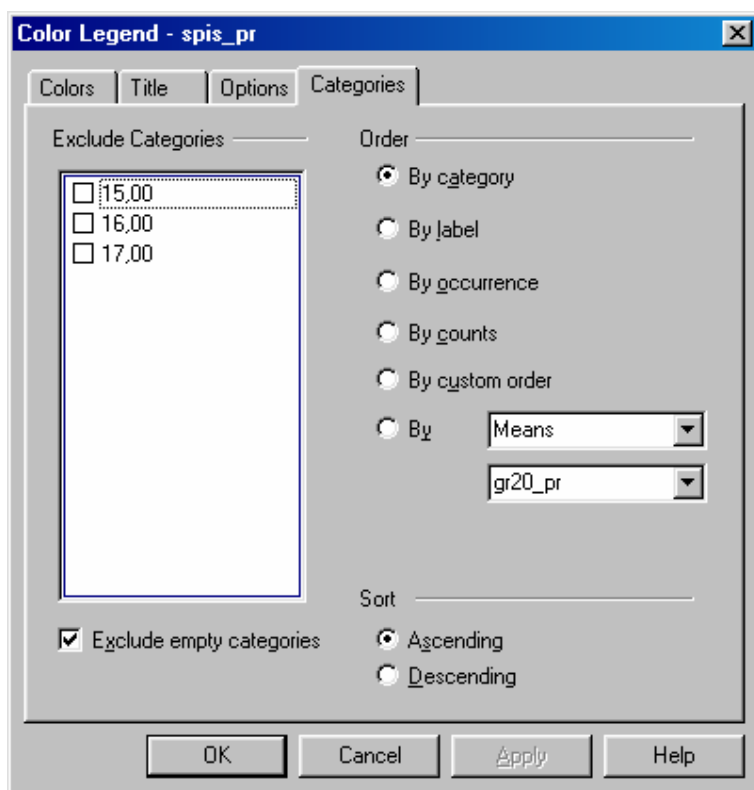
- Dialogo langelio kortelėje **Categories** (17.28 pav.) galite pakeisti kategorijų eiliškumą (**Order**), kategorijų sąrašę pažymėti kategorijas, kurios nebus rodomos diagramoje, rikiuoti kategorijas didėjančia arba mažėjančia tvarka.
- Legendos rėmeliui pakeisti pažymėkite legendą, spragtelėdami ją pele, ir nurodykite komandas **Format → Frame...** Atsidariusiame dialogo langelyje pasirinkite legendos rėmelio linijos tipą, spalvą ir storį, taip pat legendos ploto nuspalvinimo stilių ir spalvą.

17.2.9. Kaip keisti diagramos ašis?

Dialoginių grafikų ašių savybes galima keisti skalės ašių (**Scale Axis**) ir kategorijų ašių (**Category Axis**) dialogo langeliuose. Šių dialogo langelių kortelėse galite:

- Keisti ašių padalų formatą ir diapazoną.
- Nurodyti ašių padalų žymių (*axis labels*) pakartojimo dažnį, vietą ir formatą.
- Nurodyti ašių pavadinimų šriftą ir lygiavimo tipą.
- Nurodyti koordinatinio tinklelio tipą.
- Parinkti antrą koordinatinę ašį dvimatėms diagramoms.

Panagrinėsime skalės tipo ašių parametrų nustatymą. Kategorijų ašių parametrų nustatymo principai yra panašūs.



17.28 pav. Dialogo langelio **Color Legend** kortelė **Categories**

Jeigu norite pakeisti ašių savybes:

- Aktyvuokite diagramą, ją dukart spragtelėdami pele.
- Dukart spragtelėkite pele ašį arba jos padalos žymę.

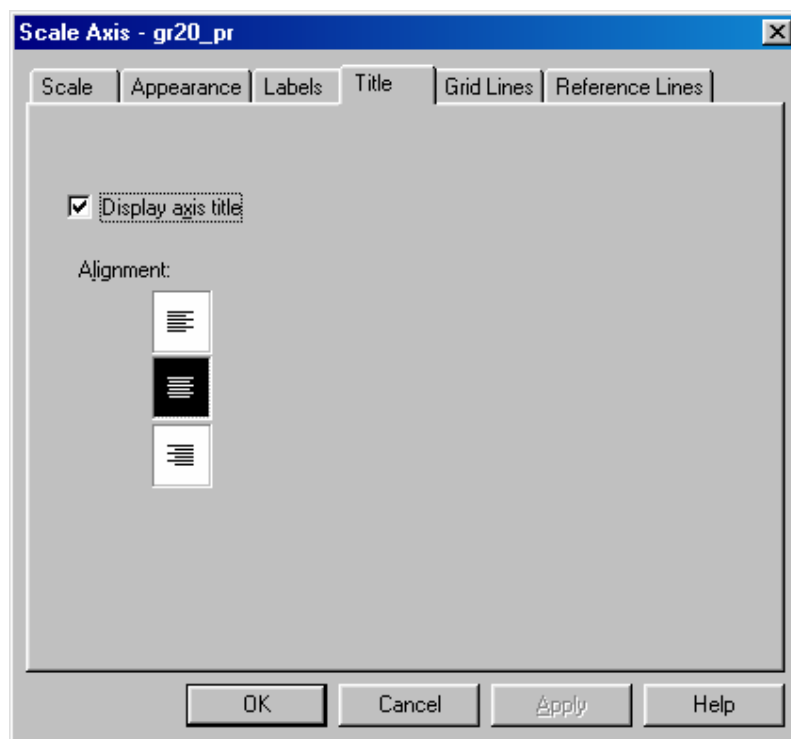
Arba

- Nurodykite komandas **Format** → **Axis** → (ašies pavadinimas).

Arba

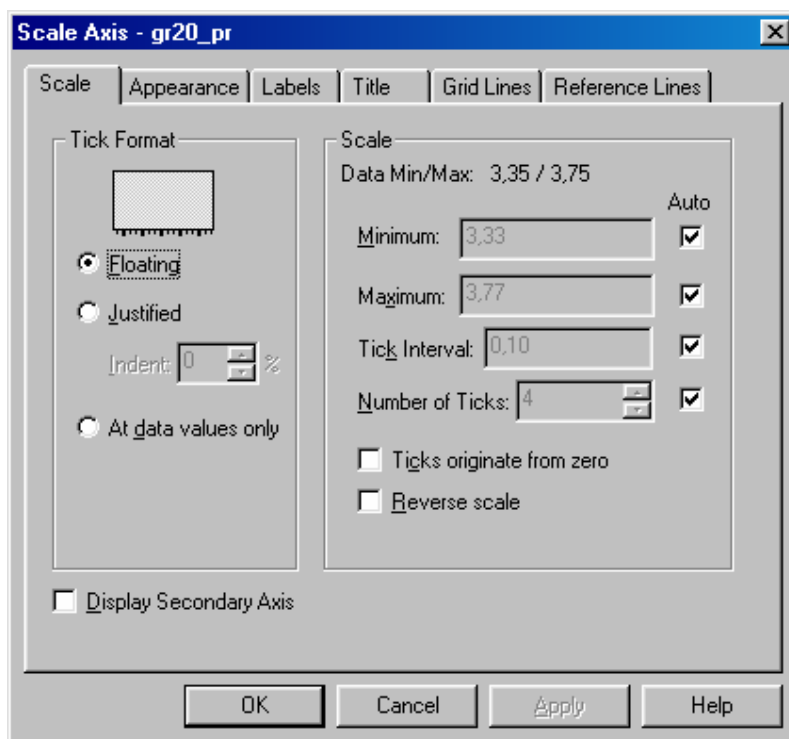
- Atidarykite **Chart Manager** dialogo langelį, pažymėkite reikiamą ašį ir spragtelėkite šio dialogo langelio mygtuką **Edit**.
- Pasirinkite reikiamą dialogo langelio **Scale Axis** kortelę **Scale** (skalė), **Appearance** (išvaizda), **Labels** (žymės), **Title** (pavadinimai), **Grid Lines** (koordinatinis tinklelis), **Reference Lines** (nuorodų linijos).
- Pasirinkite tinkamą variantą ir spragtelėkite **Apply** (komanda bus įvykdyta, bet dialogo langelis liks atidarytas) arba **OK** (komanda bus įvykdyta, o dialogo langelis — uždarytas).
- Dialogo langelio kortelėje **Titles** (17.29 pav.) mygtukais **Alignment** Jūs galite pasirinkti ašies pavadinimo vietą: ašies pradžioje, viduryje ar gale, o panaikinę laukelio **Display axis title** žymę — nerodyti ašies

pavadinimo. Ašies pavadinimą į kitą vietą galite taip pat nuvilkti pele. Pavadinimui redaguoti, dukart spragtelėkite jį pele arba naudokite teksto įrankį (*Text tool*).

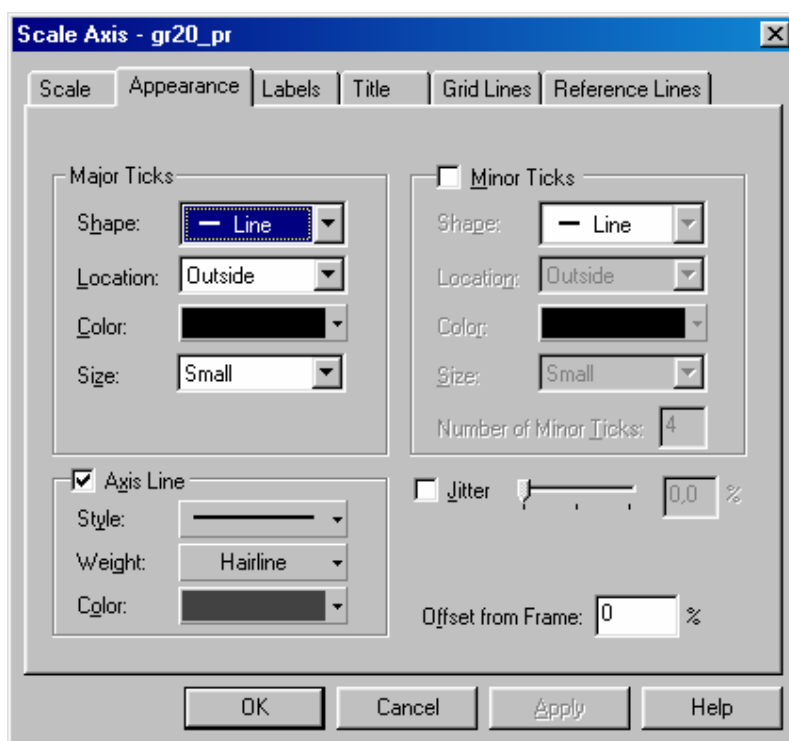


17.29 pav. Dialogo langelio *Scale Axis* kortelė *Title*

- Dialogo langelio kortelėje *Scale* (17.30 pav.) galite:
 - Pasirinkti skalės padalų formatą *Tick Format*;
 - Nustatyti pageidaujamas skalės pradžios (*Minimum*) ir pabaigos (*Maximum*) reikšmes, intervalą tarp padalų (*Tick Interval*) arba padalų skaičių (*Number of Ticks*). Prieš nustatydami patys išvardytus parametrus, panaikinkite žymę atitinkamuose *Auto* laukeliuose.
 - Nurodyti, kad pirma padala kirstų koordinačių pradžią (*Ticks originate from zero*).
 - Pasirinkti vietoje didėjančios mažėjančią skalę (*Reverse scale*).
- Dialogo langelio kortelėje *Appearance* (17.31 pav.) galite:
 - Pasirinkti skalės stambių padalų (*Major Ticks*) arba smulkių padalų (*Minor Ticks* — jos bus rodomos pažymėjus laukelį *Minor Ticks*), formą (*Shape*), padėtį (*Location*), spalvą (*Color*) ir dydį (*Size*).
 - Pasirinkti ašies linijos stilių (*Style*), storį (*Weight*) ir spalvą (*Color*).

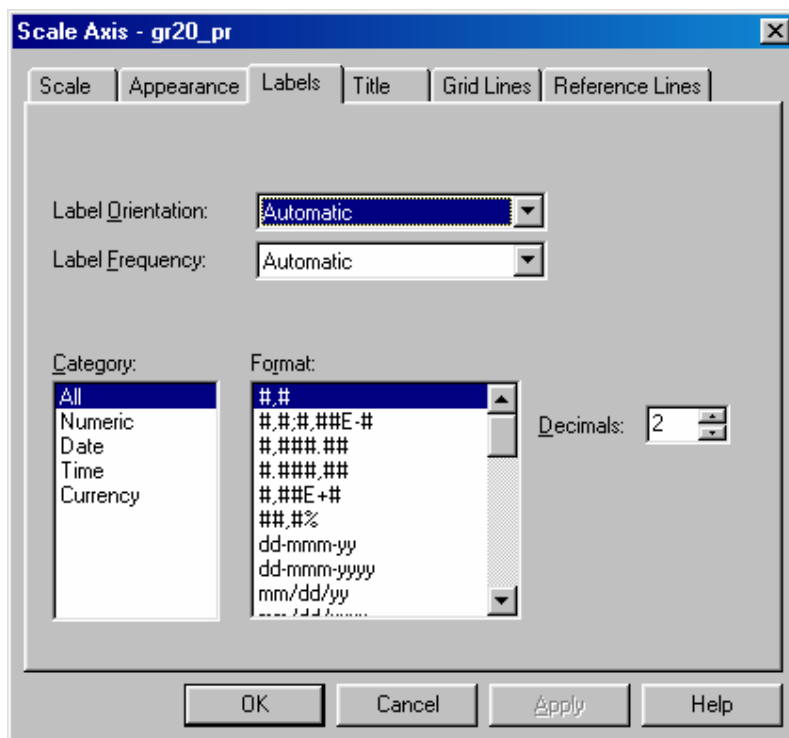


17.30 pav. Dialogo langelio **Scale Axis** kortelė **Scale**



17.31 pav. Dialogo langelio **Scale Axis** kortelė **Appearance**

- Dialogo langelio kortelėje **Labels** (17.32 pav.) galite:
 - Pasirinkti ašies padalų žymių orientaciją (**Label Orientation**) ir jų pasikartojimo dažnį (**Label Frequency**). Dirbant **Automatic** režimu programa parenka šiuos parametrus taip, kad žymės neždengtų viena kitos.
 - Pasirinkti ašies padalų žymių kategoriją (**Category**) ir formatą (**Format**), taip pat ženklų po kablelio skaičių (pasirinkus **Numeric** kategoriją).



17.32 pav. Dialogo langelio **Scale Axis** kortelė **Labels**

- Dialogo langelio kortelėse **Grid Lines** ir **Reference Lines** galite pasirinkti atitinkamai koordinatinio tinklelio ir nuorodų linijų stilių, spalvą ir storį.

17.2.10. Dialogo langelis **Chart Properties**

Be išvardytųjų diagramų redagavimo galimybių, galite pasinaudoti taip pat dialogo langelyje **Chart Properties** (diagramų savybės) pateiktomis galimybėmis.

- Aktyvuokite diagramą, ją dukart spragtelėdami pele.
- Nurodykite komandas **Format → Chart Properties...**
- Dialogo langelyje **Chart Properties** galite pasirinkti:

- Kortelėje **Colors** — pakeisti nustatytas diskrečių kategorijų spalvas, pasirinkti naujas spalvas spalvų paletėje arba naują spalvų diapazoną, kai kategorijas apibrėžia skalės kintamasis.
- Kortelėje **Styles** — pakeisti nustatytus diskrečių kategorijų stilius, pasirinkti naujus stilius iš siūlomų stilių sąrašo.
- Kortelėje **Sizes** — pasirinkti kitus duomenų sekas vaizduojančių simbolių dydžius.
- Kortelėje **Text** — nurodyti kiekvieno diagramos tekstinio objekto (diagramos antraštės, paantraštės, ašių pavadinimo, ašių padalų žymių, legendos, esmėžodžių, pastabų) šrifto parametrus. Diagramos tekstinį objektą pasirinkite šių objektų išskleidžiamajame sąrašė.
- Kortelėje **Filled Objects** — pasirinkti diagramos objektų, turinčių kontūrines linijas apibrėžtus plotus (stulpeliai, paviršiai, duomenų sritis ir t.t.), ploto spalvą ir spalvinimo stilių, kontūrinių linijų stilių, spalvą ir storį. Diagramos objektą pasirinkite objektų išskleidžiamajame sąrašė.
- Kortelėje **Lines & Symbols** — nustatyti įvairių pagalbinių linijų (paklaidos intervalų, glodinimo ir regresijos kreivių, normaliojo skirstinio kreivių ir t. t.) ir simbolių (sklaidos diagramos simbolių, įvairių laukelių) parametrus. Linijas ir simbolius, kurių parametrus numatėte pakeisti, pasirinkite atitinkamame išskleidžiamajame sąrašė.
- Kortelėje **Axis** — nustatyti koordinatinių ašių linijų ir padalų parametrus.

17.2.11. Spalvų ir stilių įrankiai

Be aprašytųjų diagramos objektų spalvų ir stilių nustatymo būdų, pasirinktų diagramos objektų spalvas ir stilius galima nustatyti taip pat naudojantis atitinkamais įrankiais, kurie atsiranda aktyvavus diagramą. Spalvų įrankiai:



Fill Color (užpildyti spalva) tam tikrai sričiai nuspalvinti.



Border Color (kraštinių spalva) srities kraštinėms nuspalvinti


- Aktyvuokite diagramą, ją dukart spragtelėdami pele.
- Pažymėkite diagramos objektą, kurį norite nuspalvinti, spragtelėdami jį pele.
- Spragtelėję **Fill Color** ar **Border Color** įrankį, matysite, kokia spalva yra nuspalvintas pažymėtas objektas. Jeigu norite pasirinkti kitą spalvą, spragtelėkite įrankio rodyklę ir spalvų paletėje pasirinkite pageidaujamą spalvą. Daugiau spalvų galite pasirinkti, spragtelėję paletės mygtuką **Custom Colors...**

Stilių įrankiai:



Fill Style (užpildymo stilius) tam tikros srities nuspalvinimo stiliui

pasirinkti.

 **Symbol Style** (simbolio stilius) simbolio tipui pasirinkti.


Jeigu norite pakeisti nuspalvinimo stilių:


- Pažymėkite diagramos objektą, kurį norite nuspalvinti tam tikru stiliumi, spragtelėdami jį pele.
- Spragtelėkite įrankio **Fill Style** rodyklę ir stilių paletėje pasirinkite pageidaujamą stilių. Daugiau stilių galite pasirinkti spragtelėję paletės mygtuką **More Styles...**


Jeigu norite pakeisti simbolio stilių:

- Pažymėkite diagramos objektą, kurio stilių norite pakeisti, spragtelėdami jį pele.
- Spragtelėkite įrankį **Symbol Style** ir stilių paletėje pasirinkite pageidaujamą simbolio stilių. Daugiau stilių galite pasirinkti, spragtelėję paletės mygtuką **More Styles...**


Kiti įrankiai:

 **Line Style** (linijų stilius).

 **Connector Style** (jungiančių linijų stilius).

 **Symbol Size** (simbolio dydis).

 **Line Weight** (linijos storis).

 **Text Tool** (teksto įrankis). Jeigu norite diagramos lauke parašyti kokį nors tekstą, spragtelėkite pele šį įrankį, paskui spragtelėkite toje vietoje, kur turi prasidėti tekstas — atsiras teksto įvedimo žymeklis.

17.2.12. Kaip pakeisti šriftą?

- Aktyvuokite diagramą, ją dukart spragtelėdami pele.
- Pažymėkite diagramos tekstą, kurio šriftą norite pakeisti, spragtelėdami jį pele.
- Nurodykite komandas **Format → Text...** ir atsidariusiame dialogo langelyje **Text** nustatykite pageidaujamus šrifto parametrus.

DALYKINĖ RODYKLĖ

- aprašomoji statistika — 68, 83
 atstumo ir panašumo matai — 238
 diagrama
 diagrama-medis — 91
 sklaidos — 154, 159, 267
 stulpelinė — 75, 103, 254
 stulpelinė sklaidos — 90
 Q-Q — 91
 diagramų redaktorius — 10, 11
 dinaminių lentelių redaktorius — 10
 dispersija — 73, 84, 86, 88, 117
 dispersinė analizė
 blokuotųjų duomenų — 69, 191, 224
 daugiamatė — 215
 dviejų faktorių — 201
 vieno faktoriaus — 69, 180
 duomenų agregavimas — 43
 duomenų atranka — 44
 duomenų rangavimas — 33
 duomenų redaktorius — 9
 duomenų rikiavimas — 38
 duomenų transformavimas — 23
 hipotezė — 67, 116
 histograma — 66, 75, 262
 imties plotis — 84, 86, 88, 117
 intervalas
 vidurkio pasikliautinis — 88
 išskirtys — 88
 įverčiai
 M-įverčiai — 88, 89
 kategorinis transformavimas — 35
 klasterinė analizė — 237
 koeficientas
 asimetrijos — 74, 84, 86, 88
 determinacijos r^2 — 156
 eksceso — 74, 84, 86, 88
 eta — 112
 kappa — 112
 kontingencijos — 110
 koreliacijos — 105
 dalinės koreliacijos — 148
 dalinis Eta-kvadrato — 196, 208
 determinacijos — 156
 Kendalo koreliacijos — 148
 Pirsono koreliacijos — 108, 146
 Spirmeno koreliacijos — 108, 148
 Kramerio V — 110
 regresijos lygties — 155
 rizikos laipsnio — 112
 ϕ — 110
 kontrastas — 185, 204
 kriterijus
 apriorinis — 183, 185
 Bonferroni — 184
 Durbino-Vatsono — 164
 Hotelling'o pėdsakas — 197, 217, 231
 LSD — 184
 Mauchly — 197, 231
 neparametrinis — 124
 Pillai pėdsakas — 197, 217, 231
 pseudodeterminacijos — 170
 Post-Hoc — 182, 206
 Roy'o — 197, 217, 231
 S-N-K — 184
 Scheffe — 206
 Tukey'o — 184
 Wald'o — 168
 Wilks'o lambda — 197, 217, 231
 kriterijaus galia — 196
 kvartilai — 72
 laiko eilutės — 35
 matavimo skalė
 intervalų — 66
 nominalinė — 65
 rangų — 65

- santykių — 66
 mediana — 73, 86, 88, 117
 moda — 73
 lentelės
 dažnių — 68, 71
 mobilišios — 56
 požymių dažnių — 70, 99
 pataisa
 apatinio režio — 197
 Greenhouse-Geisser'o — 197
 Huynh-Feld'o — 197
p-reikšmė — 68
 procentiliai — 73, 88
 regresijos lygtis — 155, 164
 rodinys — 10, 50
 sferiškumas — 191
 skirstinys
 normalusis — 66
 skirtingumo matai — 150
 standartinis nuokrypis — 68, 73, 84, 86, 88, 117
 standartinė paklaida — 69, 73, 84, 86, 88, 117
 statistinis kriterijus — 67
 sutapimo matai — 150
 testas
 Box'o M — 215
 Chi-kvadrato — 106, 124, 142
 Cochran'o ir Mantel-Haenszel'o — 114
 Friedman'o — 69, 124, 137
 Kendall'o W-testas — 124, 139
 Kolmogorovo ir Smirnov'o — 66, 88, 91, 124, 125, 141
 Kruskal'o ir Wallis'o H-testas — 69, 124, 134
 Levene — 92, 188, 215
 Mann'o ir Whitney U-testas — 69, 124, 125
 McNemar'o — 106, 124, 132
 Mozes'o — 124, 126
 Cochran'o Q-testas — 124, 140
 Shapiro ir Wilk'o — 66, 88, 91
 Stjudento t-testas — 69, 118
 Stjudento porinis t-testas — 69, 121
 Wald'o ir Wolfowitz'o — 124, 125
 Wilcoxon'o — 69, 124, 129
 ženklų — 124, 131
 trendas — 186
 vidurkis — 68, 73, 84, 86, 88, 117

LITERATŪRA

1. Gonestas E., Strielčiūnas R. R. *Taikomoji statistika*. Vadovėlis kūno kultūros ir sporto specialybių studentams. Kaunas: LKKA, 2003.
2. Čekanavičius V., Murauskas G. *Statistika ir jos taikymai 1*. Vilnius: TEV, 2000.
3. Čekanavičius V., Murauskas G. *Statistika ir jos taikymai 2*. Vilnius: TEV, 2002.
4. Sakalauskas V. *Duomenų analizė su STATISTICA*. Vilnius: Margi raštai, 2003.
5. Бююль А., Цефель П. *SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей*. Москва * Санкт-Петербург * Киев, DiaSoft, 2002.
6. *SPSS® Base 11.0 User's Guide*. Chicago, SPSS Inc., 2001.
7. *SPSS® Advanced Models*. Chicago, SPSS Inc., 2001.
8. *SPSS® Interactive Graphics 10.0*. Chicago, SPSS Inc., 2001.