STATISTINIAI KRITERIJAI

1. Statistiniai palyginimai ir statistinės hipotezės

Jau ne kartą minėta, kad tyrinėtojui neretai prisieina ne vien tik aprašyti empirinius statistinius duomenis, bet ir juos įvairiai palyginti. Duomenys, o tiksliau – tirto požymio ar požymių reikšmių empiriniai pasiskirstymai gali būti lyginami tiek tarpusavyje (t.y. vieni su kitais), tiek ir su teoriniais pasiskirstymais. Kita vertus, tiesiogiai lyginti galima tiek atskirus „rūpimų" pasiskirstymų parametrus, tiek ir visus juos „ištisai", – tai priklauso nuo lyginimo tikslo: nuo to, ką tokiu palyginimu tyrėjas nori patikrinti ar atskleisti.

Įprastas ir kartu gana universalus statistinių palyginimų bei kitų į juos panašių sprendimų „mechanizmas" remiasi dviem savo turiniu priešingomis *prielaidomis*: viena prielaida postuluoja lyginamųjų objektų (pasiskirstymų konkrečių parametrų arba jų visumos) esminį tapatumą, o kita – jų esminį skirtingumą. Tos prielaidos dažniausiai yra vadinamos *statistinėmis hipotezėmis*. Pirmoji iš jų, teigianti, kad lyginamieji objektai iš esmės nesiskiria, o iš duomenų pastebimi skirtumai yra sąlygoti vien tiktai „grynos" atsitiktinybės ir paaiškinami kaip tik jos, atitiktinybės, veikimu, dažniausiai yra vadinama *nuline hipoteze* (nes ji postuluoja esminio skirtumo nebuvimą, kitaip sakant – nulinį skirtumą; ją įprasta žymėti H_0), o antroji, jai priešinga, teigianti, jog lyginamieji objektai skiriasi iš esmės ir esamo jų skirtumo paaiškinti vien atsitiktinybės veikimu neįmanoma, vadinama *alternatyvine hipoteze* arba tiesiog *alternatyva* (alternatyvą įprasta žymėti H_1). Tad įvairiais statistiniais metodais bei procedūromis ir siekiama nustatyti, **kuri** iš šių prielaidų yra labiau pagrįsta ir **priimtina**. Be abejo, ir čia išvados apie vienos šių prielaidų pranašumą prieš kitą, apie tai, kuri iš jų labiau atitinkanti realybę ir laikytina priimtinesne, irgi yra tikimybinio pobūdžio, ir tų išvadų tikėtinumas (atitikimo tikrovei tikimybė) iš principo negali tapti absoliutus, šimtaprocentinis. Todėl priimant vienokį ar kitokį H_0 bei H_1 liečiantį sprendimą visuomet išlieka galimybė suklysti ir egzistuoja tam tikra klaidų rizika. Skiriamos dviejų rūšių klaidos:

- Pirmos rūšies klaida: atmesti nulinę hipotezę (H_0) tada, kai ji iš tikro yra teisinga
- Antros rūšies klaida: neatmesti nulinės hipotezės (H_0) tada, kai ji iš tikro yra klaidinga

Natūralu, jog siektina minimizuoti abiejų rūšių klaidas, tačiau tai gerokai keblus uždavinys, susijęs su statistikos „vidinių mechanizmų vidiniais prieštaringumais". Todėl praktiškai labiau yra paisoma ir vengiama pirmos rūšies klaidos ir kartu siekiama, kad antros rūšies klaidos tikimybė išliktų kaip galima mažesnė.

2. Bendrasis statistinių kriterijų supratimas

Minėtų statistinių prielaidų – nulinės hipotezės ir alternatyvos – „įvertinimo instrumentai" yra vadinamieji *statistiniai kriterijai*. Tai specialios, matematinės statistikos teoretikų suformuluotos bei atitinkamai pagrįstos ir daugeliu praktinių taikymų patikrintos **taisyklės**, numatančios, *kuo* ir *kaip* remiantis galima priimti sprendimą dėl vienos kurios iš „konkuruojančių" statistinių hipotezių (H_0 ar H_1) persvaros prieš kitą. Praktiškai visuomet statistiniai kriterijai remiasi tam tikrais pagalbiniais dydžiais (vediniais, transformacijomis), kurie bendru atveju yra vadinami tų *kriterijų statistikomis* (KS); jos paprastai apskaičiuojamos iš tyrimo metu gautų empirinių duomenų, o jų apskaičiavimo taisyklės ir formulės būna parinktos taip, kad pagal gautąsias konkrečias KS reikšmes būtų galima priimti pagrįstus sprendimus apie H_0 bei H_1. Dėl to KS apskaičiavimo būdai dažniausiai parenkami taip, kad tie vediniai turėtų gerai ištirtus tikimybinius skirstinius (pvz., standartinį normalųjį, chi-kvadrat, Stjudento ar Fišerio skirstinius). Lyginant konkretų šių vedinių (kriterijų) dydį (empirines reikšmes) su iš teorijos žinomais atitinkamų pasiskirstymų tikimybės tankio funkcijos $f(x)$ arba pasiskirstymo funkcijos $F(x)$ „etalonais" ir galima daryti atitinkamas statistines išvadas apie H_0 arba H_1 priimtinumą.

> Jeigu, pavyzdžiui, kokių nors dviejų lyginamųjų parametrų skirtumas yra transformuojamas taip, kad šios transformacijos rezultatas (kriterijaus statistika, KS) turi *standartinį normalųjį* pasiskirstymą, tai gauta vidurkiui (nuliui) artima KS reikšmė rodytų, kad labiau priimtina yra *nulinė hipotezė* (teigianti, jog lyginamųjų parametrų skirtumas yra grynai atsitiktinis, o šiaip jie iš esmės nesiskiria), o nuo vidurkio smarkiai nukrypstanti (sakysim, viršijanti 3 ar net 4, t.y. nutolstanti nuo vidurkio per 3–4 „sigmas") reikšmė – kad kur kas labiau priimtina *alternatyva* (teigianti, jog lyginamųjų parametrų reikšmės skiriasi iš esmės, ir jų skirtumo paaiškinti vien atsitiktinybe nebeįmanoma).

Statistiniai kriterijai, kuriais remiantis yra tikimybiškai įvertinamos hipotezės apie vienus ar kitus empirinių pasiskirstymų *parametrus*, yra vadinami ***parametriniais kriterijais***, o tie kriterijai, kuriais įvertinamas hipotezių, nesusijusių su konkrečiais parametrais, priimtinumas, vadinami ***neparametriniais kriterijais***. Abi šios statistinių kriterijų atmainos iš esmės nesiskiria, jų ir „prigimtis", ir logika yra labai panaši, o skiriasi tik lyginamieji objektai: vienu atveju – atskiri, konkretūs pasiskirstymų parametrai, o kitu – pasiskirstymų „visuma" (kartais sakoma – bendra pasiskirstymų forma). Be to, didžiuma konkretiems tikslams pritaikytų statistinių kriterijų turi nusistovėjusius konkrečius pavadinimus, besisiejančius arba su juos „sukonstravusių" statistikos teoretikų pavardėmis (pvz., Frydmano kriterijus, Vilkoksono kriterijus ir pan.), arba kokiais nors kitais tam ar kitam kriterijui esmingais momentais, sakysim, su konkrečios KS prigimtimi (pvz., serijų kriterijus) ar pasiskirstymu (pvz., chi-kvadrato kriterijus). Taip pat pabrėžtina, jog kai kada, ypač – šnekamojoje kalboje, neakcentuojamas ir tarpais net neišlaikomas skirtumas tarp paties statistinio kriterijaus ir jam reikalingos statistikos (KS); kitaip tariant, vienu ar kitu būdu apskaičiuota KS pavadinama stačiai kriterijumi.

Gana dažnai statistiniai kriterijai konstruojami taip, kad jų KS turėtų būtent ***standartinį normalųjį*** pasiskirstymą. Tokias KS „ įprasta žymėti *u* arba *z* (todėl ir jos pačios kai kada pavadinamos tiesiog *u*-kriterijumi arba *z*-kriterijumi). Šį pasiskirstymą turinčios KS populiarios dėl dviejų priežasčių: viena, šiaipjau gerai ištirtas yra pats standartinis normalusis pasiskirstymas, pasakytum, išsami jo teorija, o antra – į standartiniam normaliajam pasiskirstymui „paklūstančius" dydžius *u* gana nesunku yra transformuoti daugelį empirinių pasiskirstymų parametrų skirtumų (pvz., dviejų santykinių dažnumų skirtumą, santykinio dažnumo ir teorinės tikimybės skirtumą, dviejų empirinių vidurkių skirtumą, empirinio ir teorinio vidurkių skirtumą ir t.t.), ypač – apskaičiuotų iš palyginti didelių imčių. Todėl vadinamasis „*u*-kriterijus" yra pasidaręs gana universalus ir filologų šiaipjau yra mėgiamas...

Be standartiškai normaliai pasiskirsčiusių KS, statistiniams kriterijams yra naudojamos KS, atitinkančios vadinamuosius *chi-kvadrato* (labiau įprasta žymėti χ^2), *Stdjudento* arba *Fišerio* pasiskirstymus; dar kitokius pasiskirstymus atitinkančios KS – retesnės. Pažymėtina, kad visi paminėtieji pasiskirstymai yra tolydieji ir visi vienaip ar kitaip susiję su standartiniu normaliuoju pasiskirstymu, o vienas nuo kito jie skiriasi kiekvienas „savo" argumento (atsitiktinio dydžio *x*) „prigimtimi" bei galimų reikšmių aibe ir – bene visų labiausiai – savitais *tikimybės tankio funkcijos f*(*x*) pavidalais, kitaip sakant, kiekvienas jų turi vis skirtingą *f*(*x*) priklausomybės nuo *x* pobūdį, kurį galima išreikšti atitinkamomis formulėmis (apie tai plačiau kalbėta 9 paskaitos konspekte). O bet kurio iš jų *pasiskirstymo funkcija F*(*x*) išlieka iš esmės to paties tipo: ji būna *pirmykštė* atitinkamai tikimybės tankio funkcijai ir jos reikšmės bet kuriam *x* reikšmių intervalui randamos tame intervale integruojant jos išvestinę, t.y. atitinkamą tikimybės tankio funkciją; kitais žodžiais – *pasiskirstymo* funkcija bet kurio konkretaus iš šių pasiskirstymų atveju randama (galima būtų pridurti – ir panaudojama) iš esmės taip pat, kaip ir standartinio normaliojo pasiskirstymo atveju. Išsamiau apie χ^2, Stjudento bei Fišerio pasiskirstymus galima pasiskaityti ir atitinkamas formules rasti didžiumoje tikimybių teorijos ir/ar matematinės statistikos knygų (pvz., populiarioje J. Kruopio knygoje).

Praktiniam tyrinėtojo darbui svarbu:

- Suvokti statistinio palyginimo (ar patikrinimo) uždavinį, aiškiai įsisąmoninti, kas ir dėl ko lygintina ar tikrintina
- Konkrečiai suformuluoti abi statistines hipotezes – H_0 ir H_1 (t.y. nulinę hipotezę ir alternatyvą)
- Motyvuotai parinkti statistinį kriterijų, kuris gerai atitiktų tyrimo duomenų specifiką (tiriamojo požymio reikšmių pobūdį bei jų empirinį pasiskirstymą, imties dydį)
- Iš turimų duomenų korektiškai apskaičiuoti pasirinkto kriterijaus statistikos (KS) empirinę reikšmę (ji kitaip dar vadinama kriterijaus statistikos *realizacija*)
- Pagal gautą empirinę KS reikšmę pagrįstai nuspręsti, kuri iš suformuluotų statistinių hipotezių priimtinesnė

Kadangi labai dažnai apskaičiuotoji empirinė KS reikšmė yra ne kas kita, kaip vieną ar kitą žinomą tikimybinį pasiskirstymą turinčio „dirbtinio" (specialiai 'sukonstruoto' pagal kriterijaus taisykles) atsitiktinio dydžio realizacija, tai svarbu gerai įsisąmoninti, kuriam konkrečiam tikimybiniam pasiskirstymui kuri KS „atstovauja". Sprendimas gi dėl vienos iš hipotezių pranašumo prieš kitą priimamas laikantis tam tikro nuoseklumo ir logikos. Yra du pagrindiniai metodai, leidžiantys pagal KS spręsti apie H_0 ar H_1 priimtinumą: vienas jų, atėjęs iš „prieškompiuterinių" laikų, grindžiamas vadinamosiomis *kritinėmis reikšmėmis* bei *kritinėmis sritimis*, o antrasis – vėlesnis, vis labiau įsigalintis lygia greta su statistiniams skaičiavimams skirtų kompiuterinių programų plitimu – remiasi vadinamosiomis *p-reikšmėmis*.

3. Statistinių hipotezių (H_0 bei H_1) vertinimo logika

Ji apskritai išplaukia iš tų prielaidų formulavimo bei iš pirmos rūšies klaidos supratimo, o taip pat – iš kriterijaus statistikos (KS) parinkimo. Pastarąją gi paprastai stengiamasi parinkti taip, kad tuo atveju, **kai H_0 yra teisinga**, KS įgytų kokį nors „standartizuotą" tikimybinį skirstinį (ar kad ją būtų galima pagrįstai aproksimuoti juo); kitaip tariant – tasai pasiskirstymas turi leisti visoms galimoms KS=*x* reikšmėms patikimai apskaičiuoti tikimybės tankio

$f_{(x)}$ ir tikimybės pasiskirstymo funkcijos $F_{(x)}$ reikšmes. Tad jeigu empirinė KS reikšmė vis dėlto papuola į tas x reikšmių zonas, ties kuriomis tikimybės tankis $f_{(x)}$ pasidaro menkas, nežymus, tai tokio jos „nukrypimo" tikroji priežastis greičiausiai yra ne menkai betikėtinas atsitiktinumas, bet pamatinės nuostatos – kad H_0 *esanti teisinga* – neatitikimas tikrovei, nepagrįstumas. Vadinasi, apie H_0 pagrįstumą sprendžiama pagal empirinės KS reikšmės tikėtinumą (kurį, pasinaudojant tikimybės pasiskirstymo funkcija $F_{(x)}$, apskaičiuoti nesunku): pakankamai tikėtina KS reikšmė rodo, kad H_0 galima laikyti pagrįsta, o menkai tikėtina – kad pagrįsta jos laikyti neišeina, tiesiog yra nelogiška.

Vis dėlto bene pats svarbiausias su šiais samprotavimais susijęs klausimas būtų toks: ar galima laikyti, kad empirinės (t.y. iš turimų duomenų apskaičiuotos) KS reikšmės tikėtinumas tiesiogiai išreiškia ir pačios nulinės hipotezės (H_0) tikėtinumą? O gal tai – iš esmės vienas ir tas pat dalykas? Gaila, bet teorinėje statistikos literatūroje šis klausimas taip „tiesmukai" paprastai nekeliamas, neformuluojamas ir nesvarstomas... Tačiau jau iš pirmo žvilgsnio atrodytų, kad abu šie 'tikėtinumai' yra susiję tiesioginės priklausomybės ryšiu: tikėtinesnė empirinė KS reikšmė rodytų, kad ir H_0 yra tikėtinesnė, o menkai tikėtina empirinė KS – kad menkai tikėtina taip pat ir H_0. Remiantis net bendro pobūdžio samprotavimais galima rasti gana svarių argumentų, leidžiančių abu tuos 'tikėtinumus' laikyti tapačiais, lygiais vienas kitam. Tiesa, kalbant griežtesne kalba, čia turbūt derėtų sakyti ne *tikėtinumas*, bet *tikimybė*: tikimybė, kad H_0 esant teisingai, empirinė KS reikšmė bus **ne mažesnė** už dabar gautą jos reikšmę, yra iš esmės tapati tikimybei, kad H_0 yra teisinga (atitinka tikrovę): juk specialiai taip yra „sukonstruotas" KS apskaičiavimo „mechanizmas"! Trumpai užrašius:

$$p_{(\text{KS}>=x)} = p_{(\text{nulinė hipotezė teisinga})}$$

Jeigu nulinę hipotezę atmestume, tai pirmos rūšies klaidos tikimybė būtų, savaime suprantama, lygi tikimybei, jog nulinė hipotezė esanti teisinga.

Tačiau kol kompiuteriai ir statistinių skaičiavimų programos nebuvo išplitę, tol tyrinėtojams praktikams („taikytojams") apskaičiuoti tą $p_{(\text{KS}>=x)}$ (t.y. tikimybę, jog pasirinkto kriterijaus statistika įgis reikšmę, ne mažesnę už tą, kurią sąlygojo atlikto eksperimento duomenys) bei, tuo pačiu, tikimybę, kad nulinė hipotezė yra teisinga, – būdavo gana neparanku ir sudėtinga. Todėl nuo seno įsitvirtino ir iki šiol labai paplitęs tebėra truputį kitoks statistinių prielaidų H_0 ir H_1 vertinimo metodas, nereikalaujantis tiesioginio tikimybės skaičiavimo ir pagrįstas vadinamosiomis kritinėmis reikšmėmis bei kritinėmis sritimis.

4. Statistinių hipotezių vertinimas pagal kritines statistikų (KS) reikšmes

Aptariant jį bus reikalingos dar dvi specifiškos sąvokos: *reikšmingumo lygmuo* (įprasta žymėti α) ir *patikimumas* ($1-\alpha$). Abi jos reiškia iš esmės tą patį: *iš anksto* pasirinktą tam tikrą tikimybės „slenkstį" (ribą); reikšmingumo lygmuo α – tai riba, kurios neturi peržengti (viršyti) pirmos rūšies *klaidos tikimybė*, kai H_0 atmetame, o patikimumas – tai riba, kurią turi viršyti *teisingo sprendimo* tikimybė atmetant nulinę hipotezę. Natūralu, kad aprioriškai renkamasi palyginti nedidelis reikšmingumo lygmuo bei, tuo pačiu, didelis patikimumas: kuo reikšmingumo lygmuo mažesnis (didesnis patikimumas), tuo labiau „įtikinamas" tampa H_0 atmetimas. Tradicinės šių „tikimybinių slenksčių" reikšmės yra:

riba:	10%	5%	1%	0,1%
α:	0.1	0.05	0.01	0.001
$1-\alpha$:	0.9	0.95	0.99	0.999

Kurį iš jų kuriuo konkrečiu atveju rinktis – tyrinėtojas sprendžia pats; filologai šiaipjau gana dažnai renkasi 5% ribą ($\alpha = 0.05$).

Kai yra žinomi reikšmingumo lygmens (bei patikimumo) „dydžiai" bei konkrečios kriterijaus statistikos (KS) tikimybinis skirstinys, tai galima apskaičiuoti ir konkrečią KS reikšmę, ties kuria „susiduria" α ir $1-\alpha$. Ta reikšmė yra vadinama *kritine reikšme*. Faktiškai tai yra KS skirstinio kvantilis, – pasirinktąjį tikimybinį „slenkstį" α atitinkanti „ribinė" KS reikšmė, dalijanti visą galimų KS reikšmių aibę į dvi dalis: tikimybė KS reikšmei papulti į vieną iš jų yra didesnė už α, į kitą – mažesnė. Ši pastaroji KS reikšmių aibės dalis (poaibis) neretai yra vadinamas tiesiog *kritine sritimi*. Paprastai visų paminėtųjų tradicinių reikšmingumo lygmenų (α) kritinės reikšmės didžiumai tikimybinių skirstinių (būdingų įvairioms kriterijų statistikoms) apskaičiuojamos iš anksto ir pateikiamos lentelių pavidalu statistikos knygų prieduose. Tad statistinių kriterijų taikymas labai supaprastėja: suformuluojama H_0 bei H_1 ir pagal atitinkamas taisykles bei formules apskaičiuojama empirinė KS reikšmė. Jeigu ji atitinkamą kritinę reikšmę *pasiekia* ar *viršija*, tai H_0 atmetama su reikšmingumo lygmeniu (t.y. pirmos rūšies klaidos tikimybe), neviršijančia α, o jei nesiekia – H_0 neatmetama.

Imkime kol kas abstraktų parametrinės hipotezės pavyzdį: tarkime, jog reikia palyginti dviejų imčių (empirinių pasiskirstymų) vidurkius ir įvertinti prielaidas (hipotezes) apie jų tapatumą ar skirtingumą.

Nulinė hipotezė: abiejų imčių vidurkiai iš esmės yra lygūs, stebimi jų skirtumai tėra tiktai atsitiktinybės padarinys.
Alternatyva: šių imčių vidurkiai iš esmės nelygūs, skirtingi.

Imčių empirinių vidurkių skirtumas pagal nustatytas formules „perskaičiuotas“ į kriterijaus statistiką (KS) – pagalbinį atsitiktinį dydį *u*, turintį standartinį normalųjį tikimybinį skirstinį (kaip konkrečiai tai padaryti – kiek vėliau). Gauta empirinė *u* reikšmė yra 1,88. Ką ji „reiškia“? Kadangi kriterijaus statistika *u* turi standartinį normalųjį tikimybinį skirstinį, tai žiūrime, kokios yra apskaičiuotos kritinės reikšmės, atitinkančios „populiariuosius“ reikšmingumo lygmenis standartinio normaliojo skirstinio atveju. Iš statistikos knygose pateikiamų lentelių galima suvaikyti, kad šiuo sykiu – kai alternatyva formuluojama būtent šitaip – jos esančios tokios:

Reikšmingumo lygmuo (α)	Patikimumas ($1-\alpha$)	Kritinė reikšmė
0.10	0.90	1.64
0.05	0.95	1.96
0.01	0.99	2.58
0.001	0.999	3,29

Nesunku pastebėti, kad gautoji *u* reikšmė „papuola į tarpą“ tarp kritinių reikšmių, atitinkančių reikšmingumo lygmenį 0.1 ir 0.05. Kitaip sakant, atmesti nulinę hipotezę būtų galima su didesniu nei 90%, bet mažesniu negu 95% patikimumu, o pirmos rūšies klaidos rizika būtų tarp 5% ir 10%. Jeigu tokia rizika dar yra per didelė (kitaip tariant, jeigu reikia kur kas didesnio patikimumo), nulinės hipotezės neatmetame.

Čia jau reikėtų ypatingo akcento: ką reiškia – nulinės hipotezės neatmetame? Ogi vien tik tai, kas „tiesiai“ pasakyta: kad ji, kaipo ***prielaida***, – teigianti, jog lyginamųjų dydžių skirtumus sąlygoja vien tiktai atsitiktinybė, – paliekama ir toliau „galioti“. Ir nieko daugiau. Pati didžiausia daugelio „taikytojų“ (o neretai ir statistikos knygų…) klaida yra ta, kad dažnai tokiais atvejais hipotezė, prielaida, tarsi nejučiom pakeičiama jau nebe hipoteze, nebe prielaida, o tvirtinimu, tarpais net kategorišku teigimu (bet iš esmės klaidingu), esą lyginamieji dydžiai (parametrai) šiaipjau nesiskirią (su vienokiu ar kitokiu patikimumu). Tai – iš principo klaidingas tvirtinimas, akivaizdi netiesa! Empirinė (kitaip tarus, apskaičiavimais gauta) *u* reikšmė iš tikro teleidžia tokiais atvejais tik sakyti, kad nulinė hipotezė išlieka *tam tikru mastu tikėtina* ir kad tas jos tikėtinumas dar nėra toks menkutis, kad praktiniais sumetimais jo būtų galima išvis nepaisyti.

Aiškiau formuluojant mintį, galima būtų teigti, kad kritinę ribą *pasiekusi* ar *pražengusi* kriterijaus statistikos (KS) reikšmė leidžia nulinę hipotezę *atmesti* (t. y. tvirtinimą, jog esminio skirtumo nėra, *laikyti nepagrįstu*) su pirmos rūšies klaidos rizika, neviršijančia reikšmingumo lygmens α. Vadinasi, tuo pačiu „įsigalioja“ alternatyva (hipotezė, postuluojanti skirtumo esmingumą): ji *priimama* kaipo pagrįsta ir pasitvirtinanti. Tai – gana griežtas ir radikalus statistinis sprendimas. Bet kritinės ribos *nesiekianti* KS reikšmė dar jokiu būdu *neleidžia priimti* nulinės hipotezės, neduoda pagrindo *patvirtinti* jos postuluojamo teiginio apie skirtumo nebuvimą, priimti jo kaip jau *įrodyto*, nebe hipotetiško dalyko, nes tokiu atveju labai išaugtų *antros rūšies klaidos* – priimti hipotezę, kai ji iš tiesų yra klaidinga – rizika. Būtina atsiminti, jog *neatmetus* nulinės hipotezės, t. y. nustačius, jog eksperimentų duomenys, kaip kartais korektiškiau pasakoma, „jai neprieštarauja“, tiek pat radikalus sprendimas dažniausiai neįmanomas: nulinė hipotezė ir toliau lieka *būti tik hipoteze*, o jos *tikėtinumas* jokiu būdu *netampa lygus patikimumui* ($1-\alpha$), kaip kartais gali pasirodyti neatidžiam žvilgsniui. Tokiu atveju pamatuotai tegalima tik pasakyti, jog nulinės hipotezės *tikėtinumas viršija* α. Ir tik labai retais atvejais – kuomet gautoji KS reikšmė yra tokia mažytė, kad pagrįstai galima įrodyti, jog *alternatyvos* tikėtinumas tapo mažesnis už pasirinktąjį reikšmingumo lygmenį bei, tuo pačiu, nulinės hipotezės tikėtinumas viršijo atitinkamą patikimumą – tiktai tuomet nulinę hipotezę galima iš tiesų *priimti*, o alternatyvą – *atmesti*. Grįžtant prie mūsų pavyzdžio, *u*=1,88 teleidžia sakyti, kad nulinės hipotezės tikėtinumas yra didesnis už 5%, bet mažesnis už 10% (tuo tarpu alternatyvos tikėtinumas didesnis už 90%, bet mažesnis už 95%). Todėl visi, kas čia sakytų, kad tų vidurkių skirtumas esąs „nereikšmingas“ su 95% patikimumu, – sakytų dažniausiai, – daugiau kaip 90 atvejų iš šimto, – netiesą ir darytų antros rūšies klaidą: nors tam tikra galimybė, kad vidurkių skirtumas yra nereikšmingas, išlieka iš tikrųjų, bet jos tikėtinumas (patikimumas) tėra mažesnis kaip 10%, (bet didesnis už 5%). Nėra čia, tiesą sakant, kitko – nėra 95% patikimo pamato tą galimybę (nereikšmingumo, skirtumo nebuvimo galimybę) *paneigti* (bet tam jau yra 90% patikimas pamatas).

Kritinių reikšmių mechanizmas yra labai parankus ir labai išplitęs dėl savo paprastumo: pakanka įsiminti vos keletą „būdingųjų“ kritinių reikšmių keliems „standartiniams“ reikšmingumo lygmenims (patikimumams) ir su jomis lyginti visas apskaičiuotas kriterijaus statistikos (KS) reikšmes. Tačiau tas lyginimas, net ir formuluojant išvadas visiškai korektiškai, išlieka ganėtinai grubus ir „šuoliškas“.

5. *p*-reikšmė – priemonė statistinių hipotezių tikėtinumui įvertinti

Dabartinės „kompiuterinės“ skaičiavimo priemonės leidžia nesunkiai atlikti ir „priešingus“ apskaičiavimus: pagal gautąją empirinę kriterijaus statistikos (KS) reikšmę apskaičiuoti būtent ją atitinkančią konkrečią „empirinio“ reikšmingumo lygmens reikšmę, kuri irgi yra ne kas kita, kaip maksimali pirmos rūšies klaidos tikimybė, apskaičiuota remiantis imties (ar imčių; bendru atveju – empiriniais) duomenimis: tai tikimybė, kad, nulinei hipotezei esant teisingai, KS, būdama tam tikrą tikimybinį skirstinį turintis atsitiktinis dydis, *pasieks ar viršys* dabar gautąją reikšmę (tad atmesdami H_0 rizikuotume tokiu pat laipsniu suklysti!). Šią tikimybę įprasta vadinti tiesiog *p-reikšme*.

Apskaičiuoti konkrečią p-reikšmę su *Excel*, kai yra žinoma empirinė KS reikšmė ir jos tikimybinis skirstinys, gana paranku pasinaudojant šiame kurse itin dažnai minimos pasiskirstymo funkcijos $F(x)$ specifika ir prisiminus, jog tikimybė, kad tolydusis atsitiktinis dydis (o KS kaip tik toks ir yra!) bus *ne mažesnis* už kokią nors ribinę reikšmę s yra lygi 1–F(s):

$$P_{(\mathrm{TAD} >= s)} = 1 - F(s)$$

Ta pačia proga prisimintina, jog tikimybė, kad tolydusis atsitiktinis dydis:

- bus ne didesnis už kokią nors ribinę reikšmę s yra lygi F(s): $P_{(\mathrm{TAD} <= s)} = F(s)$
- įgis reikšmę iš intervalo nuo a iki v, yra lygi $F(v)$ ir $F(a)$ skirtumui: $P_{(a <= \mathrm{TAD} <= v)} = F(v) - F(a)$

Taigi, ir p-reikšmė apskaičiuojama analogiškai. Tada, kai KS tikimybinis skirstinys yra asimetrinis, o pačios KS reikšmės negali būti neigiamos (pvz., kai KS atitinka chi-kvadrato ar Fišerio skirstinį):

$$p\text{-reikšmė } [= \alpha_{\text{emp.}}] = 1 - F_{(\mathrm{KS})}$$

Kiek sudėtingiau yra tada, kai KS tikimybinis skirstinys būna simetriškas (pvz., standartinis normalusis arba Stjudento skirstinys), o pati KS gali įgauti tiek teigiamas, tiek ir neigiamas reikšmes. Nesileidžiant į įrodymus, tada:

$$p\text{-reikšmė } [= \alpha_{\text{emp.}}] = 2*(1 - F_{(|\mathrm{KS}|)})$$

Vadinasi, svarbu tik žinoti, kaip reikiamo tikimybinio skirstinio atveju apskaičiuotina $F(x)$. Pridurtina, jog daugelis statistiniams skaičiavimams skirtų specializuotų programų (pvz., *SPSS*, *Statistica* ir pan.) taip pat „geba“ apskaičiuoti p-reikšmę. O kadangi skirtingiems tikimybiniams skirstiniams F(x) (bei, tuo pačiu, p-reikšmė!) su *Excel* apskaičiuojama skirtingai, tai čia pateikiama suvestinė jų apskaičiavimo lentelė:

KS skirstinys	***$F_{(KS)}$ apskaičiuojama:***	***p-reikšmė apskaičiuojama:***
Standartinis normalusis	= NORMSDIST(KS)	= 2 * (1-NORMSDIST(ABS(KS)))
Stjudento (su n laisvės laipsnių)	= IF(KS<0; TDIST(ABS(KS); n; 1); 1-TDIST(KS; n; 1))	= TDIST(ABS(KS); n; 2) arba = 2 * TDIST(ABS(KS); n; 1)
Chi-kvadrato (su n laisvės laipsnių)	= 1-CHIDIST(KS; n)	= CHIDIST(KS; n)
Fišerio (su n_1 ir n_2 laisvės laipsnių)	= 1-FDIST(KS; n_1; n_2)	= FDIST(KS; n_1; n_2)

Tad praktiniam taikymui pakanka žinoti pasirinktosios kriterijaus statistikos (KS) tikimybinio skirstinio tipą ir laisvės laipsnių skaičių (-ius; jeigu, savaime suprantama, laisvės laipsniai tą skirstinį apibūdina) bei empirinę, iš tyrimo duomenų apskaičiuotą pačią KS reikšmę: to pakanka, kas su *Excel* būtų galima apskaičiuoti atitinkamą p-reikšmę.

Kaip gautąja p-reikšme pasinaudoti ir pagal ją spręsti apie nulinės hipotezės (H_0) ar alternatyvos (H_1) priimtinumą? Galimi du variantai.

1. Kadangi p-reikšmė faktiškai yra empirinis reikšmingumo lygmuo, tai ją pagrįstai galima lyginti su minėtaisiais tradiciniais reikšmingumo lygmens „slenksčiais“ (0,1; 0, 05; 0,01; 0,001). Logika čia akivaizdi savaime: p-reikšmė leidžia nulinę hipotezę *atmesti* su „tradiciniu“ reikšmingumo lygmeniu (t.y. pirmos rūšies klaidos tikimybe), *ne mažesniu* už ją pačią.

Grįžkime prie ankstesnio pavyzdžio, kur buvo gauta KS = u = 1,88. Kadangi šios KS tikimybinis skirstinys yra standartinis normalusis, tai:

p-reikšmė = 2 * (1-NORMSDIST(ABS(1,88))) = 0,060107948

Tradicinis ne mažesnis už ją reikšmingumo lygmuo yra 0,1, tad H0 pagristai galime atmesti tik su šiuo tradiciniu reikšmingumo lygmeniu, t..y. su patikimumu, lygiu 0,9 (arba 90%).

Einant šiuo keliu, iš esmės išlaikoma tradicinė statistinių hipotezių vertinimo schema bei metodika, tad matyt dėl to būtent šitoks p-reikšmės panaudojimo būdas dažniausiai rekomenduojamas ir dabartinėse statistikos knygose (plg. *Čekanavičius V.*, *Murauskas G.* Statistika ir jos taikymai, I, V., 2000, p. 145).

2. Prisiminkime tikimybinę reikšmingumo lygmens (α) bei patikimumo ($1-\alpha$) prasmę. Abu šie dydžiai yra tikimybės, kurių prasmę apytiksliai galima „sutelkti“ kad ir tokioje lentelėje:

	Reikšmingumo lygmuo (α) prilygsta tikimybei, kad:	*Patikimumas* ($1-\alpha$) Prilygsta tikimybei, kad:
	• atmesdami H_0 padarytume klaidą (1 rūšies) • priimdami H_0 pasielgtume teisingai	• atmesdami H_0 pasielgtume teisingai • priimdami H_0 padarytume klaidą (2 rūšies)
Iš čia:	• H_0 yra teisinga (atitinka realybę)	• H_0 yra neteisinga (teisinga H_1)

Ši „išvada“ gal ir per daug tiesmuka, tačiau iš esmės svarbi: reikšmingumo lygmuo, be viso kito, dar išreiškia ir tikimybę, su kuria pagal duomenis, iš kurių apskaičiuota KS, galima H_0 laikyti esant teisinga. Be abejo tai tinka ir empiriniam reikšmingumo lygmeniui – *p*-reikšmei: *p*-reikšmė, apskaičiuota pagal vienokią ar kitokią KS, kartu yra ir tų duomenų, kurių pagrindu ši KS buvo gauta, atžvilgiu formuluojamos nulinės hipotezės (H_0) teisingumo , jos atitikimo „duomenų tikrovei“ tikimybė! Todėl galimybė apskaičiuoti šią tikimybę (*p*-reikšmę) iš tiesų yra kur kas svarbiau, negu galimybė apskaičiuotąją KS reikšmę tiesiog palyginti su keliomis iš anksto apibrėžtomis kritinėmis reikšmėmis. Žinodami konkrečią *p*-reikšmę kartu žinome ir tikėtinumo, kad nulinė hipotezė atitinka mūsų eksperimento duomenis, mastą arba laipsnį (tikimybė apskritai juk ir yra tikėtinumo mastas, laipsnis), kitaip sakant – *p*-reikšmė apibūdina nulinės hipotezės *priimtinumo laipsnį*. Jeigu jis menkas, tai nulinę hipotezę atmetame (kartu – menkai terizikuodami suklysti) ir priimame („įteisiname“) alternatyvą, o jei nemenkas – neatmetame, bet dažniausiai – paliekame ir toliau „būti hipoteze“. Čia dar kartą akcentuotina, kad statistiniai kriterijai prielaidas radikaliam bei „skaidriam“ sprendimui sudaro tik tada, kai nulinę hipotezę galima *atmesti* dėl menko jos priimtinumo (tikėtinumo) laipsnio. O kalbėti apie neatmestą, paliktą galioti (t.y. būti hipoteze) nulinę hipotezę derėtų atsargiai, apgalvotai ir motyvuotai.

Kartais išvis pravartu atsisakyti statistinių hipotezių „atmetimo / priėmimo“ tradicinio mechanizmo ir pasitenkinti tuo, kad galimybė apskaičiuoti *p*-reikšmę tuo pačiu atveria galimybę įvertinti *abiejų* hipotezių – ir H_0, ir H_1 – priimtinumo laipsnį: nulinės hipotezės priimtinumo laipsnis prilygsta pačiai *p*-reikšmei, o alternatyvos – yra lygus „vienetas minus *p*-reikšmė“. Tad palyginę tuos dydžius iš karto pamatome, *kuri* iš tų hipotezių yra priimtinesnė ir *kiek* viena iš jų savo priimtinumu nusveria kitą.

> Vėl sugrįžkime prie pavyzdžio. Nustatėme, kad tada, kai standartiškai normaliai pasiskirsčiusios KS reikšmė yra 1,88, gaunama p-reikšmė, lygi 0,060107948, t.y. maždaug 6%. Toksai pat būtų ir nulinės hipotezės tikėtinumas (priimtinumo laipsnis). Ar tokio dydžio jis jau yra pakankamai menkas, kad jo būtų galima išvis nepaisyti ir nulinę hipotezę atmesti – klausimas lieka „atviras“, ir atsakymas į jį yra nebe statistikos, bet kitokių samprotavimų „sfera“. Tačiau kartu verta atkreipti dėmesį į tai, kad alternatyvos tikėtinumas (priimtinumo laipsnis) yra apie 94%; vadinasi, alternatyva šiuo atžvilgiu nusveria nulinę hipotezę apie 94/6 = 15, 6 karto. Todėl net jeigu tokio tikėtinumo nulinę hipotezę paliktume ir toliau galioti, „būti hipoteze“, alternatyvos pranašumas prieš ją vis viena būtų labai ženklus.