4 Paskaita. Statistinių hipotezių tikrinimas.

4.1 Statistinės hipotezės sąvoka

Statistinėmis vadiname hipotezes apie stebimojo atsitiktinio dydžio arba kelių dydžių pasiskirstymą. Statistinės bus, pavyzdžiui, hipotezės: stebimasis atsitiktinis dydis yra pasiskirstęs pagal normalųjį dėsnį, dviejų atsitiktinių dydžių vidurkiai yra lygūs, vieno atsitiktinio dydžio dispersija yra didesnė negu kito.

Su tokiais uždaviniais dažnai susiduriame praktikoje. Sakysime, kokiai nors ligai gydyti yra vartojami žinomi vaistai. Rasti nauji vaistai. Ar jie bus efektyvesni už senuosius? Fabrikas gamina elektros lemputes. Siūloma pakeisti technologini procesą. Ar nauju būdu pagamintos lemputės degs ilgiau?

Hipotezės skirstomos į *parametrines* ir *neparametrines.* Pirmiausiai aptarsime parametrines hipotezes, t.y. hipotezes apie nežinomus pasiskirstymo parametrus.

Sakykime, turime pasiskirstymų klasę $\{P_\theta, \theta \in \Theta\}$ ir Θ_0 yra aibės Θ poaibis. Statistinė hipotezė bus teiginys, kad stebimojo atsitiktinio dydžio skirstinys $p(\theta)$ priklauso klasei $\{P_\theta, \theta \in \Theta_0\}$, t.y. $\theta \in \Theta_0$. Tikrinamoji hipotezė paprastai vadinama pagrindine, arba *nuline*, ir žymima H_0. Kartu nagrinėjama ir priešinga jai hipotezė H_1. Ji vadinama *alternatyviąja* hipoteze, arba tiesiog *alternatyva.* Tai bus teiginys, kad $\theta \in \Theta_1 = \Theta \setminus \Theta_0$. Jei aibė Θ_0 yra sudaryta iš vieno taško, tai hipotezė vadinama *paprastąja.* Priešingu atveju ji vadinama *sudėtingąja.* Pavyzdžiui, jei stebime atsitiktinį dydį X, turintį normalųjį skirstinį su nežinomu vidurkiu μ,

1. paprastoji hipotezė su dvipuse alternatyva:

$$\begin{cases} H_0 : \mu = 5, \\ H_1 : \mu \neq 5. \end{cases}$$

2. sudėtingoji hipotezė su vienpuse alternatyva:

$$\begin{cases} H_0 : \mu > 5, \\ H_1 : \mu \leq 5. \end{cases}$$

Matematinė statistika nagrinėja kriterijus arba testus, kurie leistų spręsti, ar stebėjimo duomenys suderinami su nuline hipoteze. Kriterijų sudarymo idėja šitokia. Hipotezei tikrinti naudojama statistika $T(\mathbf{X})$, kuri pasirenkama taip, kad turėtų žinomą skirstinį, kai H_0 yra teisinga. Erdvę $\mathbf{R}^n$ suskaidome į dvi sritis – Borelio aibes: W ir $W_0 = \mathbf{R}^n \setminus W$. Sritį W parenkame taip, kad tikimybė statistikos $T(\mathbf{X}) = T(X_1, ..., X_n)$ realizacijai patekti į tą sritį, kai hipotezė H_0 yra teisinga,

$$P_\theta(T(\mathbf{X}) \in W), \theta \in \Theta_0$$

būtų mažos. Susitariame, kad hipotezė H_0 nesuderinama su stebėjimo duomenimis ir todėl yra atmestina, kai $T(\mathbf{x}) \in W$. Ši sritis paprastai vadinama *kritine.*

Jei $T(\mathbf{x}) \in W_0$, tai laikome hipotezę H_0 suderinama su stebėjimo duomenimis ir todėl priimtina.

Ši procedūra negarantuoja, kad padarytoji išvada bus visada teisinga. Mes galime tik teigti, kad ji bus teisinga su tam tikra tikimybe. Jei tikimybės $P_\theta(T(\mathbf{X}) \in W), \theta \in \Theta_0$ yra mažos, tai praktiškai labai retai atmesime hipotezę H_0, kai ji teisinga. Sakysime, jei tos tikimybės ne didesnės kaip 0,01, tai vidutiniškai ne daugiau kaip vieną kartą iš 100 atmesime teisingą hipotezą H_0. Naudodamiesi statistiniu kriterijumi, galime padaryti dviejų rūšių klaidas: atmesti, kaip jau minėjome, hipotezę H_0, kai ji yra teisinga (*pirmosios rūšies klaida*), arba priimti H_0, nors ji klaidinga (*antrosios rūšies klaida*). Galimi atvejai nurodyti lentelėje.

	Išvada	
	Priimta H_0	Atmesta H_0
H_0 teisinga	Teisingai	I rūšies klaida
H_0 klaidinga	II rūšies klaida	Teisingai

1 pav.: Pirmosios ir antrosios rūšies klaidos.

Pvz. Teismas sprendžia, ar teisiamasis kaltas. H_0 yra hipotezė, kad teisiamasis nekaltas. Teismas gali padaryti 2-jų rūšių klaidas – nuteisti nekaltą (pirmosios rūšies klaida) arba išteisinti kaltą (antrosios rūšies klaida). Manoma, kad pirmosios rūšies klaida yra pavojingesnė (nekaltą pripažinti kaltu). Todėl I rūšies klaidos tikimybė yra fiksuojama. Ji neturi viršyti mažo skaičiaus α.

Dažnai kritinė sritis yra tokio pavidalo: $W = \{T(\mathbf{X}) > c\}$, čia T yra tam kriterijaus statistika, o c yra vadinama *kritine reikšme*.

Geriausia būtų kriterijus sudaryti taip, kad abiejų rūšių klaidų tikimybės būtų minimalios. Deja, kai stebėjimų skaičius yra ribotas, to padaryti negalima. Praplėtus kritinę sritį, antrosios rūšies klaidos tikimybė sumažės, bet padidės pirmosios rūšies klaidos tikimybė, ir atvirkščiai. Todėl paprastai parenkamas rėžis, kurio neturėtų viršyti pirmosios rūšies klaidos tikimybė, ir stengiamasi minimizuoti antrosios rūšies klaidos tikimybę. Kitaip tariant, parenkamas nedidelis skaičius $\alpha \in (0, 1)$, vadinamas *reikšmingumo lygmeniu*, ir reikalaujama, kad pirmosios rūšies klaidos tikimybė būtų ne didesnė už tą skaičių. Paprastai $\alpha = 0,1; 0,05; 0,001$.

$$\max_{\theta \in \Theta_0} P_\theta(T(\mathbf{X}) \in W) \leq \alpha,$$

o antrosios rūšies klaidos tikimybė

$$\beta(\theta) = P_\theta(T(\mathbf{X}) \in W_0), \theta \in \Theta_1,$$

būtų kiek galima mažesnė, t. y. tikimybė

$$1 - \beta(\theta) = P_\theta(T(\mathbf{X}) \in W) = 1 - P_\theta(T(\mathbf{X}) \in W_0), \theta \in \Theta_1,$$

vadinama *kriterijaus galia*, kiek galima didesnė. Kriterijaus galia – tai tikimybė atmesti nulinę hipotezę H_0, kai ji yra klaidinga. Kriterijaus galia leidžia palyginti du kriterijus, turinčius tą patį reikšmingumo lygmenį to paties didumo imčiai. Tarkime, turime 2 kriterijus su kritinėmis sritimis W_1 ir W_2, turinčius tą patį reikšmingumo lygmenį α. Kriterijus su kritine sritimi W_1 vadinamas *tolygiai galingesniu* už kriterijų su kritine sritimi W_2, jei visiems $\theta \in \Theta_1$ jo galia didesnė, t.y.

$$P_\theta(T(\mathbf{X}) \in W_1) \geq P_\theta(T(\mathbf{X}) \in W_2), \theta \in \Theta_1.$$

Kriterijus su kritine sritimi W_1, kuris esant tam pačiam reikšmingumo lygmeniui α visiems $\theta \in \Theta_1$ turi didžiausią galią, vadinamas *tolygiai galingiausiu kriterijumi*, t.y.

$$P_\theta(T(\mathbf{X}) \in W_1) = \max_{W_2} P_\theta(T(\mathbf{X}) \in W_2), \theta \in \Theta_1,$$

čia maksimumas imamas pagal visas kritines sritis W_2 su reikšmingumo lygmeniu α. Kai alternatyva yra paprastoji (Θ_1 turi tik vieną elementą), tolygiai galingiausią kriterijų vadiname tiesiog galingiausiu.

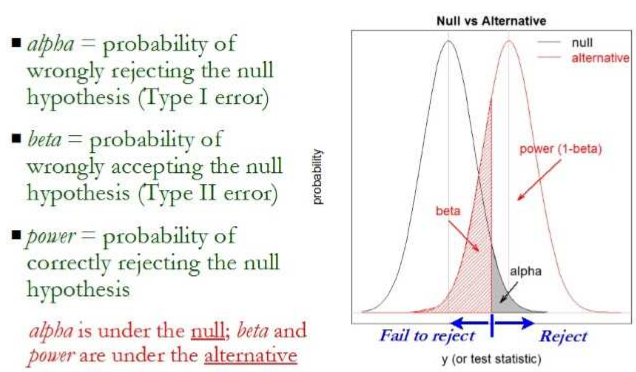
Tarkime, kad nagrinėjame paprastąją hipotezę ir paprastąją alternatyvą:

$$\begin{cases} H_0 : \theta = \theta_0, \\ H_1 : \theta = \theta_1, \end{cases}$$

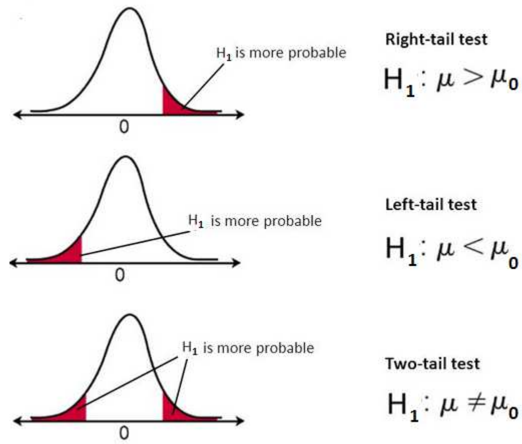
čia $\theta_0 < \theta_1$. Pagal fundamentaliąją Neimano-Pirsono lemą, galingiausias kriterijus, jei jis egzistuoja, yra pagrįstas funkcija

$$\frac{p_{\theta_1}(x)}{p_{\theta_0}(x)} = \frac{p_{\theta_1}(x_1) \dots p_{\theta_1}(x_n)}{p_{\theta_0}(x_1) \dots p_{\theta_0}(x_n)},$$

t.y. tikėtinumo funkcijų realizacijų santykiu.



2 pav.: Pagrindinės hipotezių tikrinimo sąvokos.



3 pav.: Kritinių sričių ryšys su alternatyvomis.

4.2 Hipotezė apie normaliojo skirstinio vidurkio lygybę skaičiui, kai dispersija žinoma

Tarkime, kad stebime normalųjį atsitiktinį dydį $\mathcal{N}(\mu, \sigma^2)$ ir tarkime, kad žinomas σ^2, o nežinomas μ. Reikia patikrinti hipotezę $H_0 : \mu = a$, kai alternatyva yra $H_1 : \mu > a$. Paprastai alternatyva nustatoma pagal apskaičiuotą $\bar{X}$ realizaciją. Pavyzdžiui, jei tikriname hipotezę $H_0 : \mu = 2$, tuo tarpu empirinio vidurkio realizacija yra $\bar{x} = 2,6$, logiška tikrinti šią hipotezę su vienpuse alternatyva $H_0 : \mu > 2$ arba dvipuse $H_0 : \mu \neq 2$. Turime atsakyti į klausimą: ar skirtumas tarp teorinės vidurkio reikšmės ir empirinio vidurkio gali būti laikomas statistiškai reikšmingu? Jei hipotezė H_0 yra teisinga, tai statistika

$$Z = \frac{(\bar{X} - a)\sqrt{n}}{\sigma}$$

turi standartinį normalųjį skirstinį $\mathcal{N}(0, 1)$. Kritinė sritis bus pavidalo

$$\frac{(\bar{x} - a)\sqrt{n}}{\sigma} \geq u.$$

Pirmosios rūšies klaidos tikimybė

$$\alpha = P_{\mu_0}\left(\frac{(\bar{X} - a)\sqrt{n}}{\sigma} \geq u\right) = 1 - \Phi(u).$$

Iš šios lygybės randame $u = z_\alpha$, ta yra standartinio normaliojo skirstinio α lygmens kritinė reikšmė.

Jei alternatyva yra $H_1 : \mu < a$, tai hipotezę atmetame, kai konkreti imtis tenkina nelygybę

$$z = \frac{(\bar{x} - a)\sqrt{n}}{\sigma} \leq -z_\alpha.$$

Panagrinėkime alternatyvą $H_1 : \mu \neq a$. Hipotezę atmetame, kai konkreti imtis tenkina nelygybę

$$z = \frac{|\bar{x} - a|\sqrt{n}}{\sigma} \geq z_{\alpha/2}.$$

1 lentelė: Alternatyvos ir kritinės sritys, kai $H_0 : \mu = a$

Alternatyva H_1	H_0 atmetama	H_0 neatmetama
$\mu \neq a$	$\|z\| > z_{\alpha/2}$	$\|z\| \leq z_{\alpha/2}$
$\mu > a$	$z > z_\alpha$	$z \leq z_\alpha$
$\mu < a$	$z < -z_\alpha$	$z \geq -z_\alpha$

Pavyzdys. Iš normaliosios populiacijos su standartiniu nuokrypiu $\sigma = 5,2$ gauta dydžio $n = 100$ imtis ir paskaičiuotas imties vidurkis $\bar{x} = 27,56$. Esant reikšmingumo lygmeniui $\alpha = 0,05$ patikrinkite hipotezę:

$$\begin{cases} H_0 : \mu = 26, \\ H_1 : \mu \neq 26. \end{cases}$$

Pirmiausia skaičiuojame kriterijaus statistiką ir kritinę reikšmę:

$$z = \frac{(27,56 - 26)\sqrt{100}}{5,2} = 3, \quad z_{\alpha/2} = z_{0,025} = 1,96.$$

Kadangi $|z| > z_{\alpha/2}$, H_0 atmetama.

Esant reikšmingumo lygmeniui $\alpha = 0,05$ patikrinkite hipotezę:

$$\begin{cases} H_0 : \mu = 27, \\ H_1 : \mu \neq 27. \end{cases}$$

$$z = \frac{(27,56 - 27)\sqrt{100}}{5,2} = 1,08 \Rightarrow |z| < z_{\alpha/2}.$$

Šiuo atveju nulinė hipotezė negali būti atmesta.

Pavyzdys. Patikrinsime hipotezę su vienpuse alternatyva. Detalės nominalus svoris yra $0,5$ gr. Yra žinoma, kad detalės svoris turi normalųjį skirstinį su $\sigma = 0,11$ gr. Atsitiktinai patikrinus 121 detalę, gautas vidutinis patikrintų detalių svoris $0,53$ gr. Esant reikšmingumo lygmeniui $\alpha = 0,01$ patikrinkite hipotezę, kad vidutinis detalės svoris atitinka normatyvą:

$$\begin{cases} H_0 : \mu = 0,5, \\ H_1 : \mu > 0,5. \end{cases}$$

$$z = \frac{(0,53 - 0,5)\sqrt{121}}{0,11} = 3, \quad z_\alpha = z_{0,01} = 2,326.$$

Kadangi $z > z_\alpha, (3 > 2,326)$, nulinė hipotezė atmetama, t.y. detalių gamyba neatitinka standartų.

4.3 Reikšmingumo lygmuo ir p-reikšmė

Tarkime, tikriname hipotezę su vienpuse alternatyva:

$$\begin{cases} H_0 : \theta = \theta_0, \\ H_1 : \theta > \theta_0. \end{cases}$$

Tegu $T(x_1, \ldots, x_n)$ yra kriterijaus statistikos realizacija.

Apibrėžimas. Tikimybė, kad kriterijaus statistika T, kai H_0 teisinga, ne mažesnė už *stebimą* realizaciją t^*, vadinama *p-reikšme*.

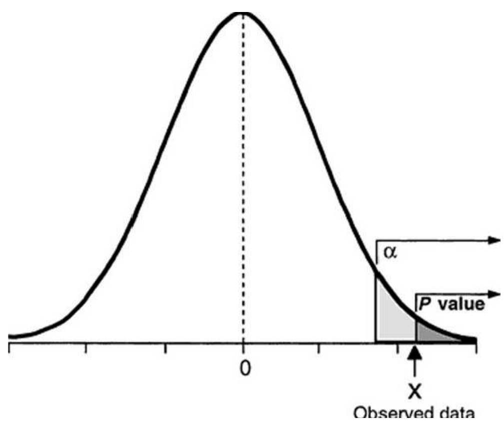
$$p = P(T \geq t^*), \text{ kai } H_0 \text{ teisinga.}$$

Kartais $p-$reikšmės vadinamos *uodegos tikimybėmis*. Pagal p-reikšmę galime iš karto formuluoti išvadą apie tikrinamą hipotezę. Tegu α yra reikšmingumo lygmuo, o p yra p-reikšmė.

1. Jei $p < \alpha$, tai hipotezė H_0 atmetama,

2. Jei $p \geq \alpha$, tai hipotezė H_0 neatmetama.

Suprantama, kad, jei $p < \alpha$, mūsų statistikos realizacija pateko į kritinę sritį, todėl atmetame nulinę hipotezę.



4 pav.: Reikšmingumo lygmuo ir p-reikšmė.