DAUGAVPILS UNIVERSITĀTE
Dabaszinātņu un matemātikas fakultāte
Matemātikas katedra

Studiju kurss

Eksperimentu plānošana un analīze

1.lekcija

Docētājs: Dr. P. Daugulis

2008./2009.studiju gads

Saturs

1. Ievads

Studiju kurss "Eksperimentu plānošana un analīze"

Docētājs - Pēteris Daugulis, Ph.D., DU vadošais pētnieks
Tālr.: DU Dabaszinātņu un matemātikas fakultātes matemātikas katedras telefons,
E-pasts: peteris.daugulis@du.lv

Webvieta lekciju materiāliem, mājasdarbiem un citai informācijai:

http://www.de.dau.lv/matematika/

Pārbaudes formas:

- rakstiski mājasdarbi,
- rakstisks eksāmens.

Kontroldarba un eksāmena darba izpildes laikā atļauts izmantot personīgos lekciju konspektus, docētāju sagatavotus metodiskos materiālus un vispārīga rakstura mācību grāmatas. Visi uzdevumi ir jāpilda pilnīgi patstāvīgi.

Galīgā vērtējuma veidošanās:
- mājasdarbi - 70%,
- eksāmena darbs - 30%.

Literatūra:
- Montgomery, D.C. *Design and Analysis of Experiments*, Wiley; 6th edition, 2004.
- Montgomery, D.C. *Applied Statistics and Probability for Engineers*, John Wiley & Sons, 4th edition, 2006.
- Dean, A., Voss, D. *Design and Analysis of Experiments*, Springer, 1999.
- Walpole, R.E., Myers, R.H., Myers, S.L., Ye Keying *Probability & Statistics for Engineers & Scientists*, Pearson Education International, 2002.
- Papildliteratūra - augstākā matemātika.
- Internet resursi - www.wikipedia.org.

Lekcijas mērķis:

- apskatīt eksperimentu plānošanas un analīžes vispārējos principus,
- atkārtot statistikas pamatus.

2. Eksperimentu plānošanas un analīzes pamatprincipi

2.1. Pamatjēdzieni

Eksperiments- organizēts tests vai testu virkne, kas satur mērījumus, kurā pētāmās sistēmas vai procesa parametri tiek mainīti ar nolūku izpētīt kāda interesējoša lieluma izmaiņas un to cēloņus.

Novērojums - mērījumi bez mērķtiecīgas parametru maiņas.

Eksperimenti ļauj efektīvāk veikt secinājumus.

2.1. piemērs. Kādā ražotnē ir novērots, ka viens darbagalds ražo sliktas kvalitātes produkciju. Tam var būt vairāki iemesli - darbagalda defekti, darbinieku kvalifikācija, paaugstināta temperatūra dotajā vietā. Bez papildus eksperimenta šī novērojuma rezultātu nevar analizēt tālāk.

Eksperimenti parasti tiek veikti, lai

- pētītu sistēmas vai procesus, kurus nav iespējams pētīt ar matemātiskās modelēšanas palīdzību;
- eksperimentāli pamatotu matemātiskās modelēšanas rezultātus.

Pētāmais process ir atkarīgs no šādiem lielumiem:

- *ievadobjekta (input)* (bieži - materiāla objekta), kas ir atkarīgs no procesa parametriem - *faktoriem* $x_1, ..., z_1, ...$;
- *izvadobjekta (output)* (procesa rezultātā pārveidota ievadobjekta), kuru raksturo parametri $y_1, ..., y_k$;
- kontrolējamiem faktoriem $x_1, ..., x_n$;
- nekontrolējamiem faktoriem $z_1, ..., z_m$.

Svarīgākie jautājumi, uz kuriem ir jādod atbilde pirms eksperimenta plānošanas:

- kāds ir eksperimenta mērķis?
- kādi rezultāti ir jāsasniedz?

- vai mērķi var sasniegt ar pieejamajiem resursiem?

Parasti eksperiments tiek veikts šādu iemeslu dēļ:

- noskaidrot kāda izvadobjekta parametra variāciju (izmaiņu, izkliedes) galvenos cēloņus;
- noskaidrot kontrolējamo parametru vērtības, kas nodrošina izvadobjekta interesējošā parametra maksimālās vai minimālās vērtības (optimizācija);
- salīdzināt izvadobjekta interesējošā parametra vērtības ar dažādām citu kontrolējamo parametru vērtībām;
- noskaidrot, kādi kontrolējamie faktori visspēcīgāk ietekmē izvadobjekta parametrus;
- noskaidrot, ar kādām kontrolējamo parametru vērtībām izvadobjekta parametru vērtību izmaiņas ir minimālas;
- izveidot matemātisku (kvantitatīvu) modeli nākotnes paredzēšanai (citi eksperimenti, novērojumi, ražošanas procesu iznākumi u.c.).

Plānojot eksperimentus ir vēlams ievērot šādus principus:

- ir vēlams apzināt visus iespējamos interesējošā izmērāmā lieluma variāciju cēloņus un veikt eksperimentus, kuros vienlaicīgi mainās vairāki variāciju cēloņu parametru, tas parasti samazina eksperimentu skaitu (*faktoriālie eksperimenti*);
- ir vēlams plānot eksperimentus tā, lai maksimizētu rezultātu precizitāti un minimizētu resursu patēriņu (laiks un nauda);
- ir vēlams vienlaicīgi plānot eksperimentu un eksperimenta rezultātu analīzi.

2.2. Pamatprincipi

Svarīgākās eksperimentu plānošanas tehnoloģijas ir

- salīdzināšana,
- replikācija,
- blokošana,
- randomizācija,
- ortogonalitāte.

2.2.1. Salīdzināšana

Mērāmos lielums nav iespējams (un dažreiz nav nepieciešams) noteikt absolūti precīzi. Lietderīgāk ir salīdzināt savā starpā vairākus eksperimentu rezultātus.

2.2.2. Replikācija

Replikācija ir vairāku neatkarīgu eksperimentu veikšana ar **vienādiem kontrolējamiem** un, **iespējams, dažādiem nekontrolējamiem** variāciju cēloņu parametriem.

2.2. piemērs. 4 cilvēkiem vienu reizi tiek dotas vienas un tās pašas zāles - replikācijas piemērs, jo cilvēki ir dažādi un katram cilvēkam zāles tiek dotas vienu reizi.

1 cilvēkam tiek dotas vienas un tās pašas zāles 4 reizes - tas nav replikācijas piemērs, jo eksperimenti nav neatkarīgi.

2.2.3. Blokošana

No vienas puses - variāciju cēloņu parametrus ir vēlams pētīt plašos diapazonos. No otras puses - ja parametru diapazoni ir plaši, to vērtības ir nekontrolējamas un to ietekme nav pētīšanas fokusā, tad ir grūti salīdzināt rezultātus.

Blokošana - novērojumu sadalīšana grupās (*blokos*), kurās eksperimentu parametri (kontrolējamie un nekontrolējamie) ir tuvi.

2.2.4. Randomizācija

Randomizācija (nejaušināšana) - nejaušības ieviešana eksperimentu plānošanā (nejaušas eksperimentu kārtības vai nejaušu parametru izmantošana).

Randomizācija ir nepieciešama, lai izvairītos no sistemātiskas vai personiskas subjektivitātes eksperimentā.

2.3. piemērs. Medicīniskā eksperimentā eksperimentators var apzināti vai neapzināti izvēlēties tādus cilvēkus, kas varētu labāk pamatot vēlamos eksperimenta rezultātus.

Ir dotas 3 krāsa un 4 veidu materiāli. Ir jānosaka, kurai krāsai ir īsākais žūšanas laiks. Ja eksperimentā katra krāsa tiek pēc kārtas pārbaudīta uz visiem materiāliem, tad dažādu faktoru, piemēram, eksperimentatora noguruma, atmosfēras apstākļu vai materiālu īpašību dēļ var rasties kļūdas.

Randomizācijā ir vēlams izmantot nejaušo skaitļu ģeneratorus.

2.2.5. Ortogonalitāte

Veicinot faktoriālos salīdzināšanas eksperimentus, ir lietderīgi

- reprezentēt parametru kopas kā vektorus,
- dažādos eksperimentos izvēlēties ortogonālus vektorus.

bigskip

2.2.6. Eksperimentu analīze

Eksperimentu rezultāti var tikt analizēti

- grafiski,
- analītiski, statistiski.

Mūsdienās eksperimentu rezultāti tiek analizēti ar datorprogrammu palīdzību - Excel, SPSS, MINITAB.

3. Eksperimentu plānošanas vadlīnijas

3.1. Eksperimentu plānošanas, izpildes un analīzes algoritms

3.1.1. Problēmas formulēšana

Nepieciešams

- precīzi noformulēt pētāmo problēmu, apskatot visus aspektus,
- iesaistīt visas ieinteresētās puses;
- sagatavot konkrētu jautājumu vai apakšproblēmu sarakstu, kas tiks pētīti eksperimenta gaitā,

3.1. piemērs. Pētot kādu ražošanas kvalitātes problēmu, nepieciešams iesaistīt speciālistus no vairākām nozarēm - inženieri, kvalitātes nodrošinātāji, mārketings, administratori u.c.

3.1.2. Kontrolējamo parametru - faktoru, to līmeņu un diapazonu izvēle

Eksperimenta parametri (ievadparametri) - jebkuri parametri, kas var izmainīt eksperimenta izvadparametrus vai eksperimenta procesus.

Parametri var būt

- diskrēti,
- nepārtraukti.

Eksperimentu parametri dalās divās grupās:

- *potenciālie dizaina parametri, potenciālie faktori (potential design/treatment factors)* - tos eksperimentators var mainīt, to ietekme ir pētījumu mērķis;
- *blakusparametros (nuisance factors)*.

Potenciālie dizaina parametri dalās sīkāk

- *dizaina parametros, faktoros (design/treatment factors)* - tie tiek mainīti,
- *konstantajos parametros*- tie tiek uzturēti konstantā līmenī,
- *pieļaujami mainīgos parametros*- tiem tiek ļauts mainīties noteiktās robežās.

Blakusparametri dalās sīkāk

- *kontrolējamos parametros* - ar tiem nodarbojas blokošana, *blokošanas faktori*,
- *nekontrolējamos parametros (covariates)* - tos nevar kontrolēt, bet, iespējams, var mērīt,
- *trokšņa parametros* - tos ir grūti kontrolēt un mērīt, tie var mainīties eksperimenta gaitā, tiem apzināti ļauj būt mainīgiem, nebloķē.

Nepieciešams

- noteikt visus parametrus - sastādīt izsmeļošu sarakstu,

- atdalīt potenciālos faktorus un blakusparametrus,
- sadalīt potenciālos faktorus apakšgrupās,
- sadalīt blakusparametrus apakšgrupās.

Pēc paramatru klasifikācijas nepieciešams izvēlēties faktoru pētāmos diapazonus un līmeņus.

Vēlams izmantot pēc iespējas mazāk faktoru un to līmeņu skaitu. To noteikšanai ir jāizmanto teorētiskās zināšanas par pētāmo sistēmu vai procesu un iepriekšējā pieredze.

3.1.3. Izejas parametru noteikšana

Jāizvēlas parametru, kas sniedz pietiekoši daudz informācijas par pētāmo problēmu.

Visbiežāk izejas parametrs ir kādas mērījumu grupas vidējā vērtība un/vai standartnovirze.

Ir jāņem vērā mērījumu kļūda - dizaina faktoru līmeņu starpībām ir jābūt pietiekoši lielām.

3.1.4. Eksperimenta plāna (dizaina) izvēle

Jārisina dažādi jautājumi par eksperimenta plānu:

- kas ir *eksperimentālā vienība*,
- jāizvēlas *eksperimenta plāns/dizains*- kārtībā, kādā eksperimentālajām vienībām tiek piekārtoti faktoru līmeņi,
- jānosaka nepieciešamie mērījumi, eksperimentālās procedūras,
- replikāciju skaits,
- blokošanas veids,
- randomizācijas algoritms,
- eksperimentu secība.

3.1.5. Eksperimenta izpilde

Eksperimenta izpildes laikā ir nepieciešams sekot, lai viss notiek saskaņā ar plānu.

Vēlams veikt pilotekspeimentus, lai

- pārbaudītu mērījumu prezicitāti,
- papraktizētos.

3.1.6. Datu statistiskā analīze

Datu analīzei jāizmanto statistiskās metodes.

Ir daudz atbilstošas programmatūras - Excel, SPSS, MINITAB u.c.

3.1.7. Secinājumi un rekomendācijas

Pēc datu analīzes var veikt šādas darbības:

- jāizdara praktiski secinājumi,
- jāformulē rekomendācijas,
- jāprezentē (vēlams, grafiski) iepriekšminētais,
- ja nepieciešams, var veikt otrās paaudzes eksperimentus un secinājumu apstiprināšanas eksperimentus.

Pētnieciskā darbība ir iteratīva - eksperimenti un to rezultātu teorētiska noformēšana mainās cikliski. Vajadzētu plānot eksperimentu sēriju vai programmu.

3.2. Īsa vēsture

3.2.1. Lauksaimniecības ēra

R.A.Fišers (1920.-1930.gadi) izstrādāja randomizācijas, replikācijas un blokošanas principus, faktoriālo dizainu, dispersiju analīzi. Pētījumi tika veikti ar mērķi maksimizēt lauksaimniecības produkciju.

Svarīgākās monogrāfijas tika izdotas 1958.g un 1966.g.

3.2.2. Industriālā ēra

Tika veikti pētījumi, kas atbilda rūpnieciskās ražošanas procesu specifikai. Tika izstrādāta reakcijas virsmas (response surface) metodoloģija (Box, Wilson). Pētījumi tika veikti ar mērķi maksimizēt rūpnieciskās ražošanas produkciju.

3.2.3. Taguči ēra

Lauksaimniecības un industriālajās ērās akcents tika likts uz vidējo vērtību pētīšanu, jo eksperimenti bija vērsti uz produkcijas maksimizēšanu vai resursu minimizēšanu.

Japānā G.Taguči vadībā tika veikti pētījumi, kas vērsti uz ražošanas kvalitātes un tās stabilitātes paaugstināšanu - izejas parametru standartnovirzes samazināšanu:

- lai padarītu ražošanas procesus stabilus attiecībā uz grūti kontrolējamiem ārējiem faktoriem,
- lai padarītu produkciju stabilu attiecībā uz mainīgiem izejvielu parametriem,
- lai atrastu tādas parametru vērtības, kas samazina standartnovirzes un panāk vēlamo vidējo vērtību.

Taguči izstrādāja frakcionālus (nepilnus) faktoriālos dizainus, izmantoja ortogonalitāti.

4. Statistikas pamati

4.1. Ievads

4.1.1. Pamatdefinīcijas

Eksperimentos parasti iegūst vairākus mērījumus, kas var būt

- diskrēti/kvalitatīvi (jā-nē, $0-1$, krāsas),
- nepārtraukti/kvantitatīvi (reāli skaitļi).

Sīkāk analizēsim kvantitatīvos datus.

Termini:

- *statistiskā rinda* (dati sakārtoti iegūšanas kārtībā),
- *variāciju rinda* (dati sakārtoti nedilstošā kārtībā).

Variāciju rindu

$$x_1 \leq x_2 \leq ... \leq x_n$$

var raksturot divu tipu lielumiem:

- statistiskajiem raksturlielumiem,
- struktūras raksturlielumiem.

Svarīgākie statistiskie raksturlielumi:

- (aritmētiskā) vidējā vērtība

$$\bar{x} = \frac{x_1 + ... + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n},$$

- dispersija

$$\sigma^2 = \frac{(x_1 - \bar{x})^2 + ... + (x_n - \bar{x})^2}{n} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n},$$

- standartnovirze $\sigma = \sqrt{\sigma^2}$,
- variācijas koeficients

$$V = \frac{\sigma}{\bar{x}} \cdot 100\%.$$

Svarīgākie struktūras raksturlielumi:

- relatīvā biežuma funkcija - noteiktā intervālā esošu rindas elementu skaits vai relatīvais biežums, speciālgadījums - elementu skaits ar noteiktu vērtību,
- kumulatīvā (uzkrātā) biežuma funkcija

$$x_i \to \frac{i}{n},$$

- $100k$ procentu procentīles funkcija p_k - variāciju rindas elements, kuram kumulatīvais biežums pirmo reizi pārsniedz k,
- *kvartīles (quartiles)*:1.kvartīle - $p_{0.25}$, 2.kvartīle (*mediāna*) - $p_{0.5}$, 3.kvartīle - $p_{0.75}$,
- *starpkvartīļu diapazons (interquartile range, IQR)* -

$$p_{0.75} - p_{0.25},$$

- *moda* - visbiežāk sastopamā vērtība.

4.1. piezīme. SPSS → Analyze → Descriptive Statistics → Explore

4.1.2. Datu reprezentācija

Var izmantot šādus datu reprezentācijas veidus:

- *"zaru un lapu diagramma" (stem-and-leaf diagram)*
- *histogramma (relatīvā/parastā un kumulatīvā)*,
- *"kastes un ūsu diagramma" (box-and-whiskers diagram)*- kaste ir starpkvartīļu diapazons, ūsas ir no mazākās/lielākās vērtības, kas ir intervālā $1.5 \cdot IQR$, parādīta mediāna.

4.2. piezīme. SPSS → Graphs → Interactive.
SPSS → Analyze → Descriptive Statistics → Explore

4.1.3. Varbūtību sadalījumi

Jāatkārto varbūtību teorijas termini:

- varbūtiskā telpa (Ω, σ, P), diskrētās un nepārtrauktās,
- gadījuma lielums $\xi : \Omega \to \mathbb{R}$.

Gadījuma lieluma ξ *varbūtības sadalījuma funkcija*

$$F_\xi(x) = P(\xi \leq x).$$

Gadījuma lieluma ξ *varbūtības blīvuma funkcija* $p(x)$:

- ja ξ ir diskrēts gadījuma lielums, tad

$$p(x) = P(\xi = x) \text{ un } \sum_i p(x_i) = 1.$$

- ja ξ ir nepārtraukts gadījuma lielums, tad

$$p(x) = F'(x) \text{ un } F(x) = \int_{-\infty}^{x} p(t)dt \text{ un } \int_{-\infty}^{+\infty} p(t)dt = 1.$$

Ja $\Phi(x_1, ..., x_m)$ ir m argumentu funkcija, tad var definēt gadījuma lielumu $\Phi(\xi_1, ..., \xi_m)$.

Var definēt m dimensiju gadījuma lielumu $(\xi_1, ..., \xi_m)$.

Gadījuma lieluma ξ *matemātiskā cerība (pirmās kārtas moments)* $\mu = E(\xi) = \bar{\xi}$ (vidējā vērtība ilgtermiņā):

- ja ξ ir diskrēts gadījuma lielums, tad

$$\mu = \sum_i x_i p(x_i).$$

- ja ξ ir nepārtraukts gadījuma lielums, tad

$$\mu = \int_{-\infty}^{+\infty} tp(t)dt.$$

Gadījuma lieluma ξ *dispersija (otrās kārtas moments)* $\sigma^2 = D(\xi) = E[(\xi - \bar{\xi})^2]$:

- ja ξ ir diskrēts gadījuma lielums, tad

$$\sigma^2 = \sum_i (x_i - \bar{x})^2 p(x_i).$$

- ja ξ ir nepārtraukts gadījuma lielums, tad

$$\mu = \int_{-\infty}^{+\infty} (t - \bar{\xi})^2 p(t) dt.$$

Gadījuma lieluma ξ augstākie momenti:

- *asimetrija (trešās kārtas moments)*

$$As(\xi) = \frac{\int_{-\infty}^{+\infty} (t-\mu)^3 f(t) dt}{D(\xi)^{3/2}},$$

- *ekscess (ceturtās kārtas moments)*

$$Ex(\xi) = \frac{\int_{-\infty}^{+\infty} (t-\mu)^4 f(t) dt}{D(\xi)^2} - 3.$$

Par gadījuma lieluma *p-kvantili* x_p sauc tādu $x \in \mathbb{R}$, kas apmierina vienādojumu

$$P(\xi \leq x) = p \text{ vai } F_\xi(x) = p \text{ vai } \int_{-\infty}^{x} f(t) dt = p.$$

Kvartiles definē tāpat kā variāciju rindas gadījumā - $x_{0.25}, x_{0.5}, x_{0.75}$.

Divus gadījuma lielumus ξ_1, ξ_2 sauc par neatkarīgiem, ja

$$P(\xi_1 \leq x, \xi_2 \leq y) = P(\xi_1 \leq x)P(\xi_2 \leq y).$$

Definēsim

$$E[(\xi_1 - \bar{\xi_1})(\xi_2 - \bar{\xi_2})] = Cov(\xi_1, \xi_2).$$

Pamatīpašības:

- $E(\xi_1 + \xi_2) = E(\xi_1) + E(\xi_2)$,
- $D(\xi_1 + \xi_2) = D(\xi_1) + D(\xi_2) + 2Cov(\xi_1, \xi_2)$,
- ja ξ_1, ξ_2 ir neatkarīgi, tad

$$Cov(\xi_1, \xi_2) = 0,$$

$$D(\xi_1 + \xi_2) = D(\xi_1) + D(\xi_2),$$
$$E(\xi_1 \xi_2) = E(\xi_1) E(\xi_2),$$

- $E(c\xi) = cE(\xi)$,
- $D(c\xi) = c^2 D(\xi)$,
- $E(\xi^2) = E(\xi)^2 + D(\xi)$.

Par gadījuma lieluma ξ *normēto sadalījumu* sauc

$$\xi^* = \frac{\xi - E(\xi)}{\sqrt{D(\xi)}}.$$

Var redzēt, ka $E(\xi^*) = 0$ un $D(\xi^*) = 1$.

Svarīgi speciālgadījumi:

- *normālais sadalījums*

$$N(\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2},$$

- *standarta normālais sadalījums* - $N(0, 1)$.

Ja $\xi \sim N(\mu, \sigma^2)$, tad $z = \frac{\xi - \mu}{\sigma} \sim N(0,1)$.

4.1. teorēma. (*Centrālā robežteorēma*) Gadījuma lielumi $\xi_1, ..., \xi_n$ ir neatkarīgi ar vienādām sadalījuma funkcijām, kuriem $E(\xi_i) = \mu$ un $E(\xi_i) = \sigma^2$. Definēsim $\rho_n = \xi_1 + ... + \xi_n$ un $\zeta_n = \frac{\rho_n - n\mu}{\sqrt{n\sigma^2}}$. Tad

$$\lim_{n\to\infty} \zeta_n \sim N(0,1)$$

šādā nozīmē. Ja $F_n(x)$ ir gadījuma lieluma ζ_n sadalījuma funkcija un $\Phi(x)$ ir $N(0,1)$ sadalījuma funkcija, tad

$$\lim_{n\to\infty} \frac{F_n(x)}{\Phi(x)} = 1.$$

Svarīgākie fakti par normālo sadalījumu:

- $\xi \sim N(\mu, \sigma^2) \implies a\xi + b \sim N(a\mu + b, a^2\sigma^2)$,
- $\xi_i \sim N(\mu_i, \sigma_i^2) \implies \xi_1 \pm \xi_2 \sim N(\mu_1 \pm \mu_2, \sigma_1^2 + \sigma_2^2)$.

4.3. piezīme. Dota x vērtība, atrast $F_\xi(x)$ vērtību, piemēram, standarta normālam sadalījumam:

SPSS (ievadīt stūra šūnā jebkuru skaitli)→ Transform → Compute Variable → (nosaukt mainīgo jebkurā vārdā)→ CDF& Noncentral CDF→ Cdf.Normal → (ievadīt parametrus).

Dota $F_\xi(x)$ vērtība, atrast x vērtību, piemēram, standarta normālam sadalījumam:

SPSS (ievadīt stūra šūnā jebkuru skaitli)→ Transform → Compute Variable → (nosaukt mainīgo jebkurā vārdā)→ Inverse DF→ Idf.Normal → (ievadīt parametrus).

Bieži vien ir svarīgi risināt šādas problēmas standarta normālajam sadalījumam $z \sim N(0,1)$: atrast tādas z-vērtības z_- un z_+, lai $F(z_-) = \alpha/2$ un $F(z_-) = 1 - \alpha/2$. Citiem vārdiem, sakot,

$$\int_{z_-}^{z_+} f(t)dt = \alpha.$$

Redzam, ka $z_- = -z_+$.

4.1.4. Statistiskie modeļi

Par kāda lieluma *statistisko modeli* sauc tā izteikšanu vienkāršu gadījuma lielumu (ar zināmiem varbūtību sadalījumiem) funkcijas formā.

4.1. piemērs. $\xi = \alpha + \beta$, kur $\alpha \sim N(\mu_1, \sigma_1^2)$, $\beta \sim N(\mu_2, \sigma_2)$.

Eksperimentu un novērojumu rezultātu *statistiskā analīze* -

- statistisko modeļu izstrāde,
- pamatošana,
- to parametru noteikšana.

4.2. Izlases un to raksturlielumi

Gadījuma lieluma ξ visu lielumu kopu sauc par *ģenerālkopu*.

Gadījuma lieluma ξ dažu (n) izmērītu vai novērotu lielumu kopu sauc par *izlasi*.

Statistika (*inferatīvā statistika*, *inferential statistics*) pēta, kā iegūt informāciju par gadījuma lielumu, ja ir zināma tā izlase.

Dažas no statistikas problēmām:

- **sadalījuma funkcijas noteikšana**, dota gadījums lieluma ξ izlase, tuvināti jāatrod sadalījuma funkciju $F_\xi(x)$ vai sadalījuma blīvumu,
- **sadalījuma funkcijas parametru noteikšana,** dota gadījums lieluma ξ izlase, ir zināms $F_\xi(x)$ veids:

$$F_\xi(x) = F(x, \theta_1, ..., \theta_k),$$

tuvināti jāatrod lielumi $\theta_1, ..., \theta_k$,

- **statistisko hipotēžu pārbaude,** dota gadījuma lieluma ξ izlase, pieņemam, ka ir spēkā noteikti apgalvojumi par $F_\xi(x)$ vai $F_\xi(x) = F(x, \theta_1, ..., \theta_k)$ parametriem $\theta_1, ..., \theta_k$, pārbaudām, vai izlases dati nav pretrunā ar pieņēmumiem.

Sakārtotai n lielumu izlasei atbilst n gadījuma lielumu $X_1, X_2, ..., X_n$ virkne ar šādām īpašībām:

- X_i ir neatkarīgi gadījuma lielumi,
- $\forall\ i\ X_i$ sadalījuma funkcijas ir vienādas un sakrīt ar ξ sadalījuma funkciju.

Par *statistiku* sauc gadījuma lielumu virknes $X_1, ..., X_n$ funkciju (arī gadījuma lielumu).

Svarīgākās statistikas:

- izlases vidējā vērtība

$$\overline{X} = \frac{X_1 + ... + X_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n},$$

- izlases dispersija

$$S^2 = \frac{(X_1 - \overline{X})^2 + ... + (X_n - \overline{X})^2}{n-1} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}.$$

Par gadījuma lieluma ξ *parametru* θ sauc jebkuru tā sadalījuma funkcijas raksturlielumu (piemēram, momentus - vidējo vērtību, dispersiju u.c.).

Par θ *novērtējumu (estimator, punktveida novērtējums, pointwise estimator)* $\widehat{\theta}$ sauc jebkuru statistiku, ko izmanto, lai novērtētu θ.

θ novērtējumu $\hat{\theta}$ sauc par *nenobīdītu (unbiased)*, ja

$$E(\widehat{\theta}) = \theta.$$

4.2. piemērs. $\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ ir nenobīdīts novērtējums parametram

$E(\xi)$, jo

$$E(\overline{X}) = E(\frac{\sum_{i=1}^{n} X_i}{n}) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n} \cdot nE(\xi) = E(\xi).$$

Bet arī $\frac{X_1+X_2}{2}$ un pat X_i ir nenobīdīti novērtējumi.

4.3. piemērs. Izlases dispersija $S^2 = \frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{n-1}$ ir nenobīdīts dispersijas $D(\xi)$ novērtējums:

$$E(S^2) = E\Big[\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2\Big] = \frac{1}{n-1}E\Big[\Big(\sum_{i=1}^{n} X_i^2\Big) - n\overline{X}^2\Big] =$$

$$= \frac{1}{n-1}\Big[\sum_{i=1}^{n}\Big(E(\xi)^2 + D(\xi)\Big) - n\Big(E(\xi)^2 + \frac{1}{n}D(\xi)\Big)\Big] =$$

$$= \frac{1}{n-1}(n-1)D(\xi) = D(\xi).$$

θ novērtējumu $\hat{\theta}$ sauc par *efektīvu (minimum variance unbiased estimator)*, ja

- tas ir nenobīdīts - $E(\widehat{\theta}) = E(\xi)$,
- tā dispersija $D(\widehat{\theta})$ ir minimālā iespējamā.

4.4. piemērs. Salīdzināsim divus $E(\xi)$ novērtējumus - $\overline{X}$ un X_1, tie abi ir nenobīdīti:

$$D(X_1) = D(\xi),$$

$$D(\overline{X}) = D\Big[\frac{\sum_{i=1}^{n} X_i}{n}\Big] = \frac{1}{n^2}\sum_{i=1}^{n} D(X_i) =$$

$$\frac{1}{n^2}\cdot nD(\xi) = \frac{1}{n}D(\xi) = \frac{1}{n}D(X_1).$$

Redzam, ka novērtējumam $\overline{X}$ dispersija ir mazāka, tāpēc tas ir precīzāks.

Par θ novērtējuma $\hat{\theta}$ *standartkļūdu* sauc $\widehat{\theta}$ standartnovirzi $\sqrt{D(\widehat{\theta})}$.

Par θ novērtējuma $\hat{\theta}$ *vidējo kvadrātisko novirzi* $MSE(\widehat{\theta})$ sauc lielumu $E[(\widehat{\theta}-\theta)^2]$.

Ja $\widehat{\theta}$ ir θ novērtējums, tad

$$MSE(\widehat{\theta}) = E[(\widehat{\theta}-\theta)^2] = E[(\widehat{\theta}-\overline{\theta}+\overline{\theta}-\theta)^2] = D(\widehat{\theta}) + (\theta - E(\widehat{\theta}))^2.$$

Lietderīgi ir izmantot novērtējumu ar mazāko MSE.

4.5. piemērs. Ir divi θ novērtējumi $\widehat{\theta}_1$ un $\widehat{\theta}_2$, kuriem $E(\widehat{\theta}_1) = \theta$, $E(\widehat{\theta}_2) = 1.1\theta$, $D(\widehat{\theta}_1) = 60$, $D(\widehat{\theta}_2) = 50$. Redzam, ka

$$MSE(\widehat{\theta}_1) = 60 + 0 = 60,$$

$$MSE(\widehat{\theta}_2) = 50 + (0.1\theta)^2.$$

Ja $60 > 50 + 0.01\theta^2 \implies \theta < \sqrt{1000} \approx 31$, tad jālieto $\widehat{\theta}_2$.

4.2.1. Maksimālās ticamības metode

Ir dots nepārtraukts gadījums lielums ξ ar sadalījuma blīvumu $f(x,\theta)$ un izlase $X_1,...,X_n$. Uzdevums - atrast pēc iespējas precīzāku θ novērtējumu.

Definēsim *ticamības funkciju*

$$L(\theta) = f(x_1,\theta)f(x_2,\theta)...f(x_n,\theta).$$

Meklēsim tādu θ, kas maksimizē $L(\theta) \Longrightarrow$

$$\frac{\partial L}{\partial \theta} = 0.$$

4.6. piemērs. $X_1,...,X_n$ ir neatkarīgi gadījuma lielumi ar normālu sadalījuma funkciju - $X_i \sim N(\mu,\sigma^2)$. Atradīsim statistikas μ un σ

novērtēšanai ar maksimālās ticamības metodi:

$$L = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\Big[-\frac{1}{2}\sum_{i=1}^{n}\Big(\frac{x_i-\mu}{\sigma}\Big)^2\Big].$$

Iegūsim divus vienādojumus:

$$\frac{\partial L}{\partial \mu} = 0 \iff \sum_{i=1}^{n} x_i - n\mu = 0,$$

$$\frac{\partial L}{\partial \sigma} = 0 \iff \sum_{i=1}^{n}(x_i-\mu)^2 - n\sigma^2 = 0.$$

4.2.2. Novērtējuma brīvības pakāpju skaits

Par kāda sadalījuma funkcijas parametra θ novērtējošās statistikas $\widehat{\theta}$ *brīvības pakāpju skaitu* ν sauc veselu skaitli, kas ir vienāds ar

- savstarpēji neatkarīgo gadījuma lielumu skaitu, no kuriem ir atkarīga statistika $\widehat{\theta}$ vai

- izmantoto gadījuma lielumu skaita un starprezultātu starpību.

Svarīgāko statistiku brīvības pakāpju skaiti:

- $\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ - n,
- $S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$ - $n-1$, jo gadījuma lielumus $X_1, ..., X_n$ saista nosacījums
$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}.$$

4.2.3. Statistikas sadalījums

Ja $X_1, X_2, ..., X_n$ ir izlase un $\phi(X_1, ..., X_n)$ ir statistika - gadījuma lielums, tad var pētīt šīs statistikas sadalījuma funkciju $F_\phi(x)$ vai $f_\phi(x)$.

4.2. teorēma.

1. Ja X ir nepārtraukts gadījuma lielums ar sadalījuma funkciju $f(x)$, φ ir bijektīva funkcija, $\varphi^{-1} = \psi$ un $Y = \varphi(X)$, tad Y sadalījuma funkcija

$$g(y) = f(\psi(y))|\psi'(y)|.$$

2. Ja X ir nepārtraukts gadījuma lielums ar sadalījuma funkciju $f(x)$, φ ir funkcija, kurai eksistē m inversās funkcijas $\psi_1, ..., \psi_m$, $Y = \varphi(X)$, tad Y sadalījuma funkcija

$$g(y) = \sum_{i=1}^{m} f(\psi_m(y))|\psi_i'(y)|.$$

4.2.4. Svarīgāko statistiku sadalījuma funkcijas

Svarīgākais un biežāk sastopamais sadalījums - normālais sadalījums.

Pārskaitīsim dažus svarīgākos sadalījums.

Vidējās vērtības sadalījums:

- $X_1, ..., X_n$ ir neatkarīgi normāli sadalījumi $N(\mu, \sigma^2)$,
- $\overline{X} = \frac{X_1 + ... + X_n}{n}$.

Tad $\overline{X} \sim N(\mu, \sigma^2/n)$.

Seko, ka
$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1).$$

Divu vidējo vērtību starpības sadalījums

- $X_{11}, ..., X_{1n}$ ir neatkarīgi normāli sadalījumi $N(\mu_1, \sigma_1^2)$,
- $X_{21}, ..., X_{2m}$ ir neatkarīgi normāli sadalījumi $N(\mu_2, \sigma_2^2)$,
- $\overline{X_1} = \frac{X_{11} + ... + X_{1n}}{n}$,
- $\overline{X_2} = \frac{X_{21} + ... + X_{2n}}{n}$.

Tad

$$\overline{X_1} - \overline{X_2} \sim N(\mu_1 - \mu_2, \sigma_1^2/n + \sigma_2^2/m).$$

Seko, ka

$$\frac{(\overline{X_1} - \overline{X_2}) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n + \sigma_2^2/m}} \sim N(0,1)$$

χ^2-sadalījums

k normālu neatkarīgu $N(0,1)$ gadījuma lielumu $z_1, ..., z_k$ kvadrātu summa

$$\chi_k^2 = z_1^2 + ... + z_k^2,$$

χ_k^2-sadalījums ar k brīvības pakāpēm.

Seko, ka ja $S^2 = \frac{\sum_{i=1}^n (X_i - \overline{X})}{n-1}$, kur $X_i \sim N(\mu, \sigma^2)$, tad

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2.$$

t-sadalījums

ja $z \sim N(0,1)$ un $s_k^2 \sim \chi_k^2$, tad gadījuma lieluma

$$t_k = \frac{z}{\sqrt{\frac{s_k^2}{k}}}$$

sadalījumu sauc par *t-sadalījumu ar k brīvības pakāpēm,*.

Seko, ka ja $X_1, ..., X_n$ ir neatkarīgi normāli sadalījumi $N(\mu, \sigma^2)$, tad

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}.$$

F-sadalījums

Ja $u \sim \chi_n^2$ un $v \sim \chi_m^2$ ir divi neatkarīgi gadījuma lielumi, tad gadījuma lieluma $\frac{u/n}{v/m}$ sadalījumu sauc par F-sadalījumu ar n skaitītāja brīvības pakāpēm un m saucēja brīvības pakāpēm, apzīmē ar $F_{n,m}$.

Seko, ka ja S_1^2 un S_2^2 ir n un m lielu izlašu dispersijas, kas ir ņemtas no ğenerālkopām ar dispersijām σ_1^2 un $\sigma_2 2$, tad

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{n-1,m-1}.$$

4.4. piezīme. Dota x vērtība, atrast $F_\xi(x)$ vērtību, piemēram, t, χ^2 vai F sadalījumam.

SPSS (ievadīt stūra šūnā jebkuru skaitli)→ Transform → Compute Variable → (nosaukt mainīgo jebkurā vārdā)→ CDF& Noncentral CDF→ Cdf.Nosaukums → (ievadīt parametrus).

Dota $F_\xi(x)$ vērtība, atrast x vērtību, piemēram, t, χ^2 vai F sadalījumam.

SPSS (ievadīt stūra šūnā jebkuru skaitli)→ Transform → Compute Variable → (nosaukt mainīgo jebkurā vārdā)→ Inverse DF→ Idf.Nosaukums → (ievadīt parametrus).

4.2.5. Normālā sadalījuma pieņēmuma pārbaude

Lai noteiktu, vai dotā izlase apmierina normālā sadalījuma likumu, var izmantot *normālā sadalījuma pārbaudi (normal probability plot)* vai tā variāciju $Q - Q$ *plot*.

Normalā sadalījuma pārbaude:

1. atliek punktus $(x_i, \frac{i-0.5}{n})$ vai $(\Phi(x_i), \frac{i-0.5}{n})$, kur x_i ir normēts,
2. salīdzina ar taisni, kas atbilst $N(0,1)$.

4.7. piemērs. SPSS → Analyze → Descriptive Statistics → P-P plots vai Q-Q plots.

4.2.6. Statistiskie intervāli

Ir dots nepārtraukts gadījums lielums ξ ar sadalījuma blīvumu $f(x, \theta)$ un izlase $X_1, ..., X_n$.

Pieņemsim, ka ir izvēlēta statistika $\widehat{\theta}$ - parametra θ novērtējums. Tā kā $\widehat{\theta}$ ir gadījuma lielums, tad tā aprēķināšana vienai izlasei dod tuvinātu θ vērtību. Ir vēlams novērtēt $\widehat{\theta}$ precizitāti.

Pamatideja - aizvietot $\widehat{\theta}$ vērtību ar intervālu - *statistisko intervālu*, kurā ar lielu varbūtību var atrasties nezināmā parametra θ precīzā vērtība.

Izmanto 3 veidu statistiskos intervālus:

- *ticamības intervāls* - nosaka robežas, kas satur īsto θ vērtību ar noteiktu varbūtību,
- *prognozēšanas intervāls* - nosaka robežas, kas satur novērojamo $\widehat{\theta}$ vērtību ar noteiktu varbūtību,
- *tolerances intervāls* - nosaka robežas, kas satur noteiktas ģenerālkopas daļas θ vērtību ar noteiktu varbūtību,

$100(1-\alpha)$ ticamības intervāli:

- divpusējie intervāli $[u, l]$: $P(u \leq \widehat{\theta} \leq v) = 1 - \alpha$,

- apakšējais vienpusējais intervāls $[v, +\infty)$: $P(\widehat{\theta} \leq v) = 1 - \alpha$,
- augšējais vienpusējais intervāls $(-\infty, u]$: $P(u \leq \widehat{\theta}) = 1 - \alpha$.

Divpusējo intervālu robežas nav noteiktas viennozīmīgi. Ir jāizmanto papildus nosacījumi - simetrija:

$$P(u \leq \widehat{\theta} \leq v) = 1 - \alpha \rightsquigarrow \begin{cases} P(v < \widehat{\theta}) = \frac{\alpha}{2} \\ P(\widehat{\theta} < u) = \frac{\alpha}{2}. \end{cases}$$

Ticamības intervāla noteikšanas algoritms:

1. Atrast gadījuma lieluma $\widehat{\theta}$ sadalījuma funkciju, normēt to.
2. Atrisināt intervālu robežu noteikšanas vienādojumus.

4.8. piemērs. $\xi \sim N(\mu, \sigma^2)$, $\theta = \mu$, $\widehat{\theta} = \overline{X}$.
Ir zināms, ka $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$.

Pāriesim uz normēto sadalījumu

$$Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{(\overline{X} - \mu)\sqrt{n}}{\sigma} \sim N(0,1).$$

Apzīmēsim gadījums lieluma Z p-kvantiles funkciju ar x_p:

$$P(Z < x_p) = p.$$

$$P(Z < u) = \frac{\alpha}{2} \implies u = x_{\frac{\alpha}{2}},$$
$$P(v < Z) = \frac{\alpha}{2} \implies v = x_{1-\frac{\alpha}{2}},$$

Apzīmēsim $x_{\frac{\alpha}{2}} \to -z_{\frac{\alpha}{2}}$, tad $x_{1-\frac{\alpha}{2}} \to z_{\frac{\alpha}{2}}$, jo $N(0,1)$ sadalījuma blīvuma funkcija $e^{-\frac{x^2}{2}}$ ir pāra funkcija.

Redzam, ka

$$-z_{\frac{\alpha}{2}} < Z < z_{\frac{\alpha}{2}} \implies -z_{\frac{\alpha}{2}} < \frac{(\overline{X} - \mu)\sqrt{n}}{\sigma} < z_{\frac{\alpha}{2}} \implies$$

$$\overline{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}.$$

Cik lielam ir jābūt n, lai ticamības intervāls nepārsniegtu dotu lielumu $2l$:

$$z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq l \implies n \geq \left(\frac{z_{\frac{\alpha}{2}} \sigma}{l}\right)^2.$$

4.5. piezīme. SPSS (ievadīt pirmajā kolonnā izlasi)→ Analyze → Descriptive Statistics (nosūtīt izlases vārdu uz Dependant List box) → Statistics→ (definēt $1-\alpha$, pēc noklusēšanas $100(1-\alpha)\% = 95\%$) → Continue → OK.

$100(1-\alpha)$ prognozēšanas intervāls - nosaka robežas, kas satur novērojamo $\widehat{\theta}$ vērtību ar noteiktu varbūtību.

Novērtēsim viena novērojuma X_0 intervālu, ja ir zināmi vairāki iepriekšējie novērojumi $X_1, ..., X_n$, $\forall i \ X_i \sim N(\mu, \sigma^2)$.

Pētīsim gadījuma lielumu $X_0 - \overline{X} \sim N(0, \sigma^2 + \sigma^2/n)$. Seko, ka

$$Z = \frac{X_0 - \overline{X}}{\sigma\sqrt{1 + 1/n}} \sim N(0, 1).$$

Rīkojoties analoģiski iepriekšējam gadījumam, no nosacījuma

$$P(u \leq Z \leq v) = 1 - \alpha$$

iegūstam, ka

$$\overline{x} - z_{\alpha/2}\sigma\sqrt{1 + 1/n} \leq x_0 \leq \overline{x} + z_{\alpha/2}\sigma\sqrt{1 + 1/n}.$$

4.3. Hipotēžu pārbaude

4.3.1. Pamatdefinīcijas

Statistiskā hipotēze - pieņēmums vai apgalvojums par ģenerālkopas sadalījuma funkciju vai tās parametriem.

Pamatjēdzieni:

- *nulles hipotēze* H_0 - pamatpieņēmums, visbiežāk divu lielumu vienādība, tā ir jānoraida vai nav jānoraida,
- *alternatīvā hipotēze* H_1 - parasti nulles hipotēzes noliegums,
- *1.veida kļūda* - nulles hipotēzes noraidīšana gadījumā, kad tā ir pareiza,
- *2.veida kļūda* - nulles hipotēzes nenoraidīšana, kad tā ir aplama, tiek izmantota, ja alternatīvā hipotēze ir konkrēta,
- α (*nozīmības līmenis*) - 1.veida kļūdas varbūtība, $1-\alpha$ - *drošības līmenis*,
- β - 2.veida kļūdas varbūtība, $1-\beta$ - *jauda*,

- *P-vērtība* - mazākā α vērtība, ar kuru tiek noraidīta nulles hipotēze,
- *testa statistika* - statistika, kas tiek izmantota, lai pieņemtu lēmumu par hipotēzi,
- *kritiskais intervāls* - statistikas vērtību kopa, kas liek noraidīt nulles hipotēzi, gali - *kritiskās vērtības.*

4.3.2. α-vērtības pieeja

Ja H_0 ir formā $\theta = \theta_0$, tad H_0 ir jānoraida ar nozīmības līmeni α, ja θ $100(1-\alpha)\%$ ticamības intervāls satur θ_0.

Hipotēžu pārbaudes algoritms:

1. Formulēt hipotēzes H_0 (vēlams, vienādības formā) un H_1.
2. Izvēlēties α (visbiežāk, 0.05 vai 0.01).
3. Izvēlēties testa statistiku (parasti normētu), atrast tās sadalījumu, atrast testa statistikas vērtību.

4. Atrast kritisko intervālu.
5. Noraidīt H_0, ja testa statistikas vērtība ir kritiskajā intervālā.

4.9. piemērs. 100 cilvēku vidējais dzīves ilgums ir 68.5 gadi, standartnovirze $\sigma = 6.6$ gadi. Vai var apgalvot ar 95% varbūtību, ka vidējais dzīves ilgums ir lielāks kā 66 gadi?

$H_0 : \mu = \mu_0 = 66,$
$H_1 = \mu > \mu_0 = 66,$
$\alpha = 0.05$
Testa statistika: $z = \frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$.
Nosacījums: $P(\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} > v) = 1 - \alpha$.
Kritiskais apgabals: $v = z_\alpha$. $\overline{X} > \mu_0 + \frac{z_\alpha \sigma}{\sqrt{n}} = 66 + \frac{1.09 \cdot 6.6}{10} = 66.72$.
Seko, ka H_0 ir jānoraida, vidējais vecums ar 95% varbūtību ir lielāks nekā 66.

4.10. piemērs. SPSS→ Analyze → Descriptive Statistics → Explore → (novirzīt mainīgā nosaukumu uz Dependent List) → Statistics (ievadīt vajadzīgo $100(1-\alpha)$) vērtību → OK.

4.3.3. P-vērtības pieeja

P-vērtība - mazākā α vērtība, ar kuru ir jānoraida H_0 dotajai izlasei.

Lai atrastu P vērtību, ir jāatrod integrāļi no dotās statistikas sadalījuma blīvuma pa intervāliem, kurus ierobežo testa statistikas vērtība un hipotēzes nosacījumi.

4.11. piemērs. Ja $Z \sim N(0,1)$ un izlases Z-vērtība ir z_0, tad

$$\begin{cases} P = \int_{-\infty}^{-z_0} f(t)dt + \int_{z_0}^{+\infty} f(t)dt = 2(1-\Phi(z_0)) \text{ - divpusējam testam} \\ P = \int_{-\infty}^{z_0} f(t)dt = \Phi(z_0) \text{ - apakšējam testam} \\ P = \int_{z_0}^{+\infty} f(t)dt = 1-\Phi(z_0) \text{ - augšējam testam.} \end{cases}$$

Pārbaudot hipotēzes, ir lietderīgi noteikt arī P-vērtību.

5. 1.mājasdarbs

1.1 Dots, ka $Z \sim N(0,1)$. Atrodiet

(a) $P(z \geq 0.33)$;

(b) $P(z \leq 0.33)$;

(c) u tādu, ka $P(z \geq u) = 0.33$.

1.2 100 studentiem tika mērīts augums, izrādījās, ka vidējā vērtība ir 175 cm, standartnovirze - 12.5 cm. Pieņemot, ka augumu sadalījums ir normāls, atrodiet 97% ticamības intervālu studenta augumam.

1.3 (Montgomery) Šķidra mazgāšanas līdzekļa standarta viskozitātei ir jābūt vienādai ar 800 (nosacītās vienībās). Standarta novirze $\sigma = 25$, sadalījums ir normāls. Tiek veikti eksperimenti ar 16 paraugiem, izrādās, ka viskozitātes vidējā vērtība ir 812.

(a) Formulējiet hipotēzes par vidējām vērtībām.

(b) Pārbaudiet hipotēzes ar $\alpha = 0.05$.

(c) Atrodiet P-vērtību.