



Valsts administrācijas
skola

Datu analīze – 1.līmenis

Biruta Sloka, Ieva Brence
1.nodarbība

Mācības finansētas no Eiropas Sociālā fonda projektiem

Nr.3.4.2.0/15/I/001 "Valsts pārvaldes cilvēkresursu profesionālā pilnveide labāka regulējuma izstrādē mazo un vidējo komersantu atbalsta jomā" un

Nr.3.4.2.0/15/I/002 "Valsts pārvaldes cilvēkresursu profesionālā pilnveide korupcijas novēršanas un ēnu ekonomikas mazināšanas jomā"

NACIONĀLAIS
ATTĪSTĪBAS
PLĀNS 2020



EIROPAS SAVIENĪBA

Eiropas Sociālais
fonds

I E G U L D Ī J U M S T A V Ā N Ā K O T N Ē



Valsts administrācijas
skola

Mācību plānojums, 1.nodarbība

9.00-10.30	1.daļa, Statistikas mērķis, statistikas metodes, statistikas pielietošanas iespējas, datu avoti; datu grupēšana, statistikas praktiskais pielietojums
10.30-10.45	Kafijas pauze
10.45-12.15	2.daļa, Kvalitatīvie un kvantitatīvie dati un to apstrādes metodes; Populācija (ģenerālkopa) un izlase; Skaitīšanas un izlases apsekojums, Statistikas pielietojums; Datu veidi (t.sk. individuālie dati, administratīvo reģistru dati), mikrodati un makrodati; laikrindas
12.15-13.00	Pārtraukums
13.00-14.30	3.daļa, Lokācijas mēri: aritmētiskais vidējais; moda, mediāna
14.30-14.45	Kafijas pauze
14.45-16.15	4.daļa, Variācijas mēri: dispersija, standartnovirze, vērtību amplitūda; starpkvarrtiņu amplitūda



Valsts administrācijas
skola

Statistikas mērķis

Statistikas metodes

Statistikas metodes izmanto statistikas datu analīzei un pamatotu lēmumu pieņemšanas sagatavošanai

- grafiskās metodes
- dažādu veidu diagrammas, tajā skaitā intensitātes kartes
- laika rindu (angliski *time series*) analīzi, ieskaitos sezonālītātes analīzi
- aprakstošo statistiku (angliski - *descriptive statistics*)
- dažādas daudzdimensiju analīzes (angliski *multivariate analysis*) metodes



Valsts administrācijas
skola

Statistikas mērķis

Statistikas metodes

aprakstošā statistika (angliski - *descriptive statistics*)

- vidējie (pakāpju vidējie, no kuriem plašāk izmanto aritmētisko vidējo), struktūras vidējie (moda un mediāna)
- variācijas jeb izkliedes rādītāji: (variācijas apjoms jeb vērtību amplitūdu (angliski *range*), dispersija (angliski *variance*), standartnovirze (angliski *standard deviation*), aritmētiskā vidējā standartklūda (angliski *standard error of mean*), vidējā lineārā jeb vidējā absolūtā novirze (angliski *mean absolute deviation*))
- asimetrijas rādītāji (angliski *skewness*)
- ekscesa rādītāji (angliski *kurtosis*)



Valsts administrācijas
skola

Statistikas mērķis

Statistikas metodes

dažādas daudzdimensiju analīzes (angliski *multivariate analysis*) metodes:

- regresiju un korelāciju analīze (angliski *regression and correlation analysis*)
- dispersiju analīze (angliski *variance analysis jeb ANOVA – no Analysis of variance*)
- kovariāciju analīze (angliski *covariance analysis*)
- faktoru analīze (angliski *factor analysis*)
- diskriminantu analīze (angliski *discriminant analysis*)
- klasteru analīze (angliski *cluster analysis*),
u.c.



Valsts administrācijas
skola

Datu avoti

- **administratīvie datu avoti** — reģistri, datubāzes, informācijas sistēmas un citi informācijas avoti
- **dati** — fiziskās personas dati, dati par privāto tiesību juridisko personu un šādu personu apvienību, kā arī dati par valsts institūciju
- **klasifikācija** — sistematizēts objektu sadalījums pēc noteiktām pazīmēm, kurā katram objektam vai objektu grupai ir piešķirts kods
- **respondents** — privātpersona vai valsts institūcija, kurai pieprasa datus par statistisko vienību



Valsts administrācijas
skola

Statistiskās pielietošanas iespējas

Statistika galvenokārt izmantojama analīzei un
pamatotai lēmumu pieņemšanai

Piemēri

CSP -

<http://www.csb.gov.lv/>

Latvijas Banka – Mājsaimniecību finanšu un patēriņa
statistika

[https://www.bank.lv/statistika/dati-
statistika/mfps/raditaji](https://www.bank.lv/statistika/dati-statistika/mfps/raditaji)

Eiropas Centrālā Banka

[https://www.ecb.europa.eu/pub/economic-
research/research-
networks/html/researcher_hfcn.en.html](https://www.ecb.europa.eu/pub/economic-research/research-networks/html/researcher_hfcn.en.html)

Jautājums kursu dalībniekiem par statistikas
izmantošanas pieredzi



Valsts administrācijas
skola

Datu grupēšana

Variācijas rindas

- Statistikas pētījuma objekts vienmēr ir **masveida objekti** un **parādības**, piemēram, kādas valsts pārvaldes institūcijā strādājošie, kādas augstskolas studenti, uzņēmuma strādnieki, kādas nozares uzņēmumi, visas valsts iedzīvotāji utt.
- Statistikas pētījuma objekts ir **kopa** jeb **kopums**, ko veido **kopas vienības** jeb elementi (cilvēki, uzņēmumi utt.)
- Lai iegūtu ziņas par interesējošo kopu, ir jānovēro tās vienības un jāreģistrē dati par svarīgākajām pazīmēm, kas raksturo kopas vienību un līdz ar to visu kopu
- Datu savākšanas procesu sauc par **statistisko novērošanu**. Statistiskās novērošanas gaitā parasti savāc datus par vairākām, nereti daudzām **pazīmēm**
- Vienkāršākās pētījuma metodes paredz datus apstrādāt un analizēt par katru pazīmi atsevišķi. Tādēļ pieņemam, ka par katru kopas vienību ir reģistrēts viens rādītājs, visbiežāk - skaitlis, kas raksturo vienu novērojamo pazīmi



Datu grupēšana

Nepārtrauktām (varianti ir intervāli) variācijas rindām būtiski ir izvēlēties intervāla plašumu: lai variantu nav pārāk daudz vai pārāk maz. Bieži izmanto **Sterdžesa formulu**:

$$\Delta = \frac{x_{max} - x_{min}}{1 + 3,2 \times \lg n}, \text{ kur}$$

Δ – intervāla plašums

x_{max} – lielākā (maksimālā) vērtība variācijas rindā

x_{min} – mazākā (minimālā) vērtība variācijas rindā

n – novērojumu skaits variācijas rindā

$\lg n$ – novērojumu skaita variācijas rindā decimāllogaritms



Valsts administrācijas
skola

Datu grupēšana – relatīvais biežums

- $w_i = \frac{f_i}{\sum_{i=1}^k f_i}$, kur
- w_i - relatīvais biežums jeb īpatsvars variantam i kur
$$i = 1, \dots, k \quad \text{jeb } (i = \overline{1, k})$$
$$k - \text{variantu skaits variācijas rindā}$$
- $\sum_{i=1}^k f_i = n$ - kopējais novērojumu skaits variācijas rindā
- visu relatīvo biežumu jeb īpatsvaru summa ir 1 ($\sum_{i=1}^k w_i = 1$)



Valsts administrācijas
skola

Lokācijas mēri – aritmētiskais vidējais

- **Aritmētiskais vidējais** – visbiežāk izmantotais pakāpju vidējais
 - par citiem pakāpju vidējiem (harmonisko vidējo; ģeometrisko vidējo: kvadrātisko vidējo un pakāpju vidējo mažoritāti) skatiet O.Krastiņš (1998), Statistika un ekonometrija, CSP, pieejams [Google http://ezis.appspot.com/Statistika/d.00_ievads.htm](http://ezis.appspot.com/Statistika/d.00_ievads.htm)
 - visiem aritmētiskā vidējā variantiem aprēķināšanas piemēri (izdales materiāls) ar kursu dalībnieku aktīvu piedalīšanos piemēru analīzē par darbinieku fiziskajām aktivitātēm

Vienkāršais aritmētiskais vidējais

- $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$, kur
- $\bar{x}$ – aritmētiskais vidējais
- x_i – faktiskie novērojumi, kur $i = 1, \dots, n$ jeb $i = \overline{1, n}$
- n – novērojumu skaits



Valsts administrācijas
skola

Lokācijas mēri – aritmētiskais vidējais

Svērtais aritmētiskais vidējais, kur par statistiskajiem svāriem izmantoti variantu absolūtie biežumi (f_i)

$$\bar{x} = \frac{\sum_{i=1}^k x_i \times f_i}{\sum_{i=1}^k f_i}, \quad \text{kur}$$

$\bar{x}$ – aritmētiskais vidējais (svērtais, par svāriem izmantoti absolūtie biežumi)

x_i – faktisko novērojumu varianti, kur $i = 1, \dots, k$ jeb $i = \overline{1, k}$

k – variantu skaits

f_i – faktisko novērojumu absolūtie biežumi jeb absolūtie statistiskie svāri,

kur $i = 1, \dots, k$ jeb $i = \overline{1, k}$

$$\sum_{i=1}^k f_i = n$$

n – novērojumu skaits



Lokācijas mēri – aritmētiskais vidējais

Svērtais aritmētiskais vidējais, kur par statistiskajiem svariem izmantoti variantu relatīvie biežumi (w_i)

- $\bar{x} = \frac{\sum_{i=1}^k x_i \times w_i}{\sum_{i=1}^k w_i},$ jeb
- $\bar{x} = \sum_{i=1}^k x_i \times w_i,$ jo
 $\sum_{i=1}^k w_i = 1,$ kur
- $\bar{x}$ – aritmētiskais vidējais (svērtais, par svariem izmantoti relatīvie biežumi)
- x_i – faktisko novērojumu varianti kur $i = 1, \dots, k$
jeb $i = \overline{1, k}$
- k – variantu skaits
- w_i – faktisko novērojumu relatīvie biežumi (īpatsvari) jeb relatīvie statistiskie svari,
kur $i = 1, \dots, k$ jeb $i = \overline{1, k}$



Valsts administrācijas
skola

Lokācijas mēri – moda

***Moda* ir variācijas rindā visbiežāk sastopamā vērtība**

Modu ļoti vienkārši atrast diskrētā variācijas rindā – variants ar vislielāko biežumu, piemēram, visbiežāk izmaksātā alga – mūsu piemērā par diviem uzņēmumiem –
1.uzņēmumā tie bija 900 eiro, bet
2.uzņēmumā 400 eiro (aritmētiskais vidējais abos uzņēmumos bija 900 eiro)



Lokācijas mēri – moda

Modu nepārtrauktā variācijas rindā aprēķina, izmantojot formulu:

$$M_o = x_{M_o} + \Delta \frac{(f_{M_o} - f_{M_o-1})}{(f_{M_o} - f_{M_o-1}) + (f_{M_o} - f_{M_o+1})}$$

M_o – moda

x_{M_o} – modālā intervāla apakšējā robeža
modālais intervāls ir intervāls ar
vislielāko biežumu

Δ – modālā intervāla plašums

f_{M_o} – modālā intervāla absolūtais biežums

f_{M_o-1} – pirmsmodālā intervāla absolūtais
biežums

f_{M_o+1} – pēcmodālā intervāla absolūtais biežums



Valsts administrācijas
skola

Lokācijas mēri – mediāna

Mediāna dala ranžētu variācijas rindu uz pusēm – puse jeb 50% no ranžētās variācijas rindas vērtībām ir mazākas nekā mediāna un puse jeb 50% no ranžētās variācijas rindas vērtībām ir lielākas nekā mediāna.

Mediānu variācijas rindām ar nepārskaita novērojumu skaitu atrod:

nepāra skaitlis **$n=2m+1$** , tad
mediāna $M_e = x_{m+1}$

Mediānu variācijas rindām ar pārskaita novērojumu skaitu atrod:

pārskaitlis **$n=2m$** ,
tad mediāna $M_e = \frac{x_m + x_{m+1}}{2}$



Lokācijas mēri – mediāna

Mediānu nepārtrauktām variācijas rindām atrod:

$$M_e = x_{M_e} + \Delta \times \frac{\frac{\sum_{i=1}^k f_i}{2} - \sum_{i=1}^{M_e-1} f_i}{f_{M_e}}, \quad \text{kur}$$

M_e – Mediāna

x_{M_e} – mediānas intervāla apakšējā robeža (mediānas intervāls ir intervāls, kur uzkrātie biežumi pirmo reizi pāršniedz pusi no kumulatīvajiem jeb uzkrātajiem biežumiem)

Δ - mediānas intervāla plašums (mediānas intervāla augšējās un apakšējās robežas starpība)

$\sum_{i=1}^k f_i$ – novērojumu skaits variācijas rindā

$\sum_{i=1}^{M_e-1} f_i$ - pirmsmediānas intervāla uzkrātais biežums

f_{M_e} - mediānas intervāla biežums



Lokācijas mēri – kvartile

Kvartiles dala ranžētu variācijas rindu četrās daļās

- pirmā kvartile norāda, ka par to mazākas ir ceturtā daļa jeb 25% no visiem novērojumiem ranžētā variācijas rindā, bet trīs ceturtdaļas jeb 75% no visiem novērojumiem ranžētā variācijas rindā ir lielāki nekā pirmā kvartile
- otrā kvartile ir vienāda ar variācijas rindas mediānu
- trešā kvartile norāda, ka par to mazākas ir trīs ceturtdaļas jeb 75% no visiem novērojumiem ranžētā variācijas rindā, bet viena ceturtdaļa jeb 25% no visiem novērojumiem ranžētā variācijas rindā ir lielāki nekā trešā kvartile



Valsts administrācijas
skola

Lokācijas mēri – kvintile

- **Kvintiles** dala ranžētu variācijas rindu piecās daļās – pirmā kvintile norāda, ka par to mazākas ir piektā daļa jeb 20% no visiem novērojumiem ranžētā variācijas rindā, bet četras piektdaļas jeb 80% no visiem novērojumiem ranžētā variācijas rindā ir lielāki nekā pirmā kvintile
- otrā kvintile norāda, ka par to mazākas ir divas piektdaļas jeb 40% no visiem novērojumiem ranžētā variācijas rindā, bet trīs piektdaļas jeb 60% no visiem novērojumiem ranžētā variācijas rindā ir lielāki nekā otrā kvintile
- trešā kvintile norāda, ka par to mazākas ir trīs piektdaļas jeb 60% no visiem novērojumiem ranžētā variācijas rindā, bet divas piektdaļas jeb 40% no visiem novērojumiem ranžētā variācijas rindā ir lielāki nekā trešā kvintile
- ceturtā kvintile norāda, ka par to mazākas ir četras piektdaļas jeb 80% no visiem novērojumiem ranžētā variācijas rindā, bet viena piektdaļa jeb 20% no visiem novērojumiem ranžētā variācijas rindā ir lielāki nekā ceturtā kvintile



Izkliedes mēri – dispersija

Dispersija (angliski *variance*) ir visvairāk izmantotais variācijas jeb izkliedes rādītājs: datu noviržu no aritmētiskā vidējā kvadrātu aritmētiskais vidējais

$$s^2 = \overline{(x - \bar{x})^2}$$

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}, \text{ kur}$$

s^2 – dispersija (angliski *variance*)

x_i – novērojums variācijas rindā ($i = \overline{1, n}$) vai $i = 1, \dots, n$

n – novērojumu skaits variācijas rindā
(piemērā par iestādes n darbinieku
sadalījumu pēc fiziskām nodarbībām $n = 146$)



Izkliedes mēri – dispersija

Svērtā dispersija (par svariem izmantoti variantu absolūtie biežumi f_i)

$$s^2 = \frac{\sum_{i=1}^k (x_i - \bar{x})^2 \times f_i}{\sum_{i=1}^k f_i}$$

s^2 – dispersija (angliski *variance*)

x_i – variants i (diskrētā variācijas rindā) vai
varianta viduspunkts (nepārtrauktā
variācijas rindā) ($i=\overline{1, k}$) jeb $i=1, \dots, k$

f_i – variantā i absolūtais biežums ($i=\overline{1, k}$) jeb
 $i=1, \dots, k$

$\sum_{i=1}^k f_i = n$ – visu novērojumu skaits



Izkliedes mēri – dispersija

Svērtā dispersija (par svariem izmantoti variantu relatīvie biežumi w_i)

$$s^2 = \sum_{i=1}^k (x_i - \bar{x})^2 \times w_i$$

w_i – varianta i relatīvais biežums ($i = \overline{1, k}$) jeb
 $i = 1, \dots, k$

$$\sum_{i=1}^k w_i = 1 \quad - \text{relatīvo biežumu summa}$$

Ar izmantotā piemēra palīdzību (par iestādes N darbinieku fiziskajām aktivitātēm)
pārliecināties, ka visu dispersiju vērtības ir
vienādas



Valsts administrācijas
skola

Izkliedes mēri – dispersija

Ļoti bieži dispersiju aprēķina ar momentu metodi (ērti lietojama, ja aritmētiskais vidējais nav vesels skaitlis), šo formulu bieži izmanto arī paskaidrojumos dažādās datorprogrammās ar *Help*

$$s^2 = \overline{x^2} - \bar{x}^2$$

s^2 – dispersija (angliski *variance*)

$\overline{x^2}$ – novērojumu x_i kvadrātu aritmētiskais vidējais

$\bar{x}^2$ – novērojumu x_i aritmētiskā vidējā kvadrāts

Vārdiem to var izteikt – dispersija ir datu kvadrātu aritmētiskā vidējā un aritmētisko vidējo kvadrātu starpība

$$\overline{x^2} = \frac{\sum_{i=1}^n x_i^2}{n}$$

$$s^2 = \frac{\sum_{i=1}^n x_i^2}{n} - \left(\frac{\sum_{i=1}^n x_i}{n} \right)^2$$



Valsts administrācijas
skola

Izkliedes mēri – dispersija

Svērtā dispersija – par svariem izmantoti absolūtie biežumi

$$\overline{x^2} = \frac{\sum_{i=1}^k x_i^2 \times f_i}{\sum_{i=1}^k f_i}$$

$$s^2 = \frac{\sum_{i=1}^k x_i^2 \times f_i}{\sum_{i=1}^k f_i} - \left(\frac{\sum_{i=1}^k x_i \times f_i}{\sum_{i=1}^k f_i} \right)^2$$



Izkliedes mēri – dispersija

Svērtā dispersija – par svariem
izmantoti relatīvie biežumi

$$\overline{x^2} = \sum_{i=1}^k x_i^2 \times w_i$$

$$s^2 = \sum_{i=1}^k x_i^2 \times w_i - \left(\sum_{i=1}^k x_i \times w_i \right)^2$$



Valsts administrācijas
skola

Izkliedes mēri – dispersija

Dispersijas koriģētas vērtējums

$$s'^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{v} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

$$s'^2 = s^2 \times \frac{n}{n-1}$$



Valsts administrācijas
skola

Izkliedes mēri – standartnovirze

Standartnovirze ir kvadrātsakne
no dispersijas

$$s = \sqrt{s^2}$$

Standartnovirzes koriģētas
vērtējums

$$\acute{s} = \sqrt{s^2}$$



Valsts administrācijas
skola

Izkliedes mēri – vērtību amplitūda

Vērtību amplitūda jeb variācijas apjoms (*angliski Range*) ir starpība starp lielāko un mazāko vērtību

$$\mathbf{R} = x_{max} - x_{min}$$

Sagrupētiem datiem

Diskrētām variācijas rindām

$$\mathbf{R} = x_k - x_1, \quad \text{kur}$$

x_1 – pirmais variants diskrētā variācijas rindā

x_k – pēdējais variants diskrētā variācijas rindā



Valsts administrācijas
skola

Izkliedes mēri – vērtību amplitūda

Nepārtrauktām variācijas rindām

$$R = x_k - x_1, \quad \text{kur}$$

x_1 – pirmā varianta intervāla viduspunkts
nepārtrauktā variācijas rindā

x_k – pēdējā varianta intervāla
viduspunkts nepārtrauktā variācijas rindā
vai

$$R = x_{ka} - x_{1z}, \quad \text{kur}$$

x_{1z} – pirmā varianta intervāla apakšējā
jeb zemākā robeža nepārtrauktā
variācijas rindā

x_{ka} – pēdējā varianta intervāla augšējā
robeža nepārtrauktā variācijas rindā



Valsts administrācijas
skola

Izkliedes mēri – Starpkvartiļu amplitūda

Starpkvartiļu amplitūda

$$R_{Q1} = x_{Q1a} - x_{Q1z}$$

$$R_{Q2} = x_{Q2a} - x_{Q2z}$$

$$R_{Q3} = x_{Q3a} - x_{Q3z}$$



Valsts administrācijas
skola

Izkliedes relatīvie mēri – variācijas koeficients

Variācijas novērtēšanai un salīdzināšanai dažādām pazīmēm ar dažādām mērvienībām, izmanto relatīvo variācijas rādītāju - variācijas koeficientu:

$$v = \frac{s}{\bar{x}} ;$$

To parasti izsaka procentos – pareizinot šo koeficientu ar 100

$$v = \frac{s}{\bar{x}} \times 100 , \text{ kur}$$

v- variācijas koeficients

$\bar{x}$ – *aritmētiskais vidējais*

s- standartnovirze



Valsts administrācijas
skola

Paldies par Jūsu uzmanību!

{Kontaktinformācija – Biruta Sloka, e-pasts Biruta.Sloka@lu.lv; 29244966}

{Kontaktinformācija – Ieva Brence, e-pasts Ieva.Brence@riseba.lv; 26363506}

Mācības finansētas no Eiropas Sociālā fonda projektiem

Nr.3.4.2.0/15/I/001 "Valsts pārvaldes cilvēkresursu profesionālā pilnveide labāka regulējuma izstrādē mazo un vidējo komersantu atbalsta jomā" un

Nr.3.4.2.0/15/I/002 "Valsts pārvaldes cilvēkresursu profesionālā pilnveide korupcijas novēršanas un ēnu ekonomikas mazināšanas jomā"

NACIONĀLAIS
ATTĪSTĪBAS
PLĀNS 2020



EIROPAS SAVIENĪBA

Eiropas Sociālais
fonds

I E G U L D Ī J U M S T A V Ā N Ā K O T N Ē