

LISA 1

D. STATISTILISED ARVUTUSED

LISA

Statistilised arvutused

STATISTILISED ARVUTUSED

KESKMINE, RUUTKESKMINE HÄLVE, DISPERSIOON

Keskväärtaus:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad (\text{L. 1})$$

kus n – mõõtmiste arv. x_i – i -ndal mõõtmisel saadud mõõtetulemus.

Kirjeldamaks mõõtetulemuste hajumist, st keskmist erinevust keskväärtausest, kasutatakse mõistet dispersioon.

ÜKSIKTULEMUSE DISPERSIOON

$$\sigma_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2. \quad (\text{L. 2})$$

σ_x - üksiktulemuse standardhälve.

Kui teeme palju seeriaid, milles igaühes on n mõõtmist, siis saame aritmeetilise keskmise dispersiooniks:

$$\sigma_n^2 = \frac{1}{n} \sigma_x^2. \quad (\text{L. 3})$$

Dispersioon kirjeldab vaid juhuslikku viga, mitte süstemaatilist viga.

Paljude n mõõtmisest koosnevate seeriatega korral on üksikmõõtmise dispersiooniks:

Statistilised arvutused

$$\sigma_s^2 = \frac{1}{n-1} < \sum_{i=1}^n (s_i - \bar{s})^2 >. \quad (\text{L. 4})$$

Aritmeetilise keskmise dispersiooni parimaks hinnanguks nimetatakse suurust

$$\sigma_{\bar{x}}^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n(n-1)}. \quad (\text{L. 5})$$

Üksikmõõtmise standardhälve iseloomustab andmete hajumist (keskmise ümber), aritmeetilise keskmise standardhälve iseloomustab mõõdetava suuruse tegelikku väärtust. Suure hulga mõõtmiste korral

- vahemikku $[\bar{x} - \sigma_x, \bar{x} + \sigma_x]$ mahub 70% andmetest;
- vahemikku $[\bar{x} - 2\sigma_x, \bar{x} + 2\sigma_x]$ mahub 95% andmetest;
- vahemikku $[\bar{x} - 3\sigma_x, \bar{x} + 3\sigma_x]$ mahub 99% andmetest.

Aritmeetilise keskmise standardviga iseloomustab mõõdetava suuruse tegelikku väärtust järgmiselt - suure hulga mõõtmiste korral

- vahemikus $[\bar{x} - \sigma_{\bar{x}}, \bar{x} + \sigma_{\bar{x}}]$ on tegelik mõõdetav väärtus ligikaudu 70%-lise usaldusväarsusega (tõenäosusega), (täpsemalt – 68,3%);
- vahemikus $[\bar{x} - 2\sigma_{\bar{x}}, \bar{x} + 2\sigma_{\bar{x}}]$ on tegelik mõõdetav väärtus ligikaudu 95%-lise usaldusväarsusega (tõenäosusega), (täpsemalt – 95,4%);
- vahemikus $[\bar{x} - 3\sigma_{\bar{x}}, \bar{x} + 3\sigma_{\bar{x}}]$ on tegelik mõõdetav väärtus ligikaudu 99%-lise usaldusväarsusega (tõenäosusega), (täpsemalt – 99,7%).

Suurust $\sigma_{\bar{x}}$ (aritmeetilise keskmise standardhälve) kasutatakse ka vea, täpsemalt – mõõtemääramatuse hindamiseks.

LISA

Statistilised arvutused

Juhul, kui tegemist on lõpliku arvu (nagu alati) mõõtmistega, kusjuures mõõtmiste arv ei ole väga suur, allub mõõtmiste jagunemine aritmeetilise keskmise ümber, Studenti jaotusele. Mõõtemääramatus leitakse järgmiselt:

$$\Delta x = t_{n,p} \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n(n-1)}}, \quad (\text{L. 6})$$

kus $t_{n,p}$ on Studenti tegur, mis leitakse tabelist. n – mõõtmiste arv, p – usaldusväärsus (füüsikas tavaliselt 95%).

Studenti koefitsendid:

n	p		
	90%	95%	99%
2	6,31	12,7	63,7
3	2,92	4,30	9,92
4	2,35	3,18	5,84
5	2,13	2,78	4,60
6	1,92	2,57	4,03
7	1,89	2,45	3,71
8	1,86	2,36	3,50
10	1,83	2,26	3,25
15	1,76	2,14	2,98
50	1,68	2,01	2,68
100	1,66	1,98	2,63

MS Excelis vajaminevad funktsioonid:

AVERAGE(...) – aritmeetiline keskmine

SUM(nr1;nr2;...) – SUM(B4:B7) – summa

STDEV(nr1;nr2;...) – STDEV(B4:B7) – standardhälve

DEVSQ(nr1;nr2;...) – DEVSQ(B4:B7) – summa üle üksikmõõtmiste hälve keskmisest.

VÄHIMRUUTUDE MEETOD

Vaatame kahe füüsikalise suuruse vahelist sõltuvust. Kõige lihtsamal juhul võib selleks olla sirge. Olgu need suurused $y = f(x)$. Ehk . Ehk $y = kx + b$, kus k ja b on konstandid, antud juhul eksperimendi tulemustest määratavad suurused.

Üldkujulise lineaarse funktsiooni $y = kx + b$ korral kasutatakse vähimruutude meetodid suuruste k ja b leidmiseks. St. paljude mõõtmiste korral peab suurus $SS = (\Delta y_1)^2 + (\Delta y_2)^2 + \dots + (\Delta y_n)^2$ olema minimaalne. Antud suuruse minimaalsuse nõudest tulenevad konstandide k ja b leidmiseks järgmised avaldised.

$$k = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}, \quad (\text{L. 7})$$

$$b = \frac{n \sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i - \sum_{i=1}^n x_i \sum_{i=1}^n x_i y_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}, \quad (\text{L. 8})$$

kus n on mõõtmiste arv.

Mõlema suuruse k ja b jaoks saame samuti standardhälbed, need on leitavad järgmistest seostest:

$$\sigma_m = \frac{\sigma \sqrt{n}}{\sqrt{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}}, \quad (\text{L. 9})$$

$$\sigma_b = \frac{\sigma \sqrt{\sum_{i=1}^n x_i^2}}{\sqrt{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}}, \quad (\text{L. 10})$$

$$\text{kus } \sigma = \sqrt{\frac{1}{n-2} \sum (y_i - mx_i - b)^2}. \quad (\text{L. 11})$$

Selliste arvutuste lihtsustamiseks võib teha järgmised tabeli.

LISA

Statistilised arvutused

i	x_i	y_i	x_i^2	$x_i y_i$
1	0			
2	7			
...				
n	$\sum_{i=1}^m x_i =$	$\sum_{i=1}^m y_i =$	$\sum_{i=1}^m x_i^2 =$	$\sum_{i=1}^m x_i y_i =$

Tabelarvutusprogrammi MS Exceli abil võib kasutada ka järgmist võimalust – andmete järgi graafiku joonistamisel – Chart Wizard – XY-Scatter. Kui punktid olemas, siis klikk graafikul mõne punkti peal, manüüst Chart – Add Trendline (või hiire parema nupuga ”Add Trendline”).

OPTIONS

- Display Equation on Chart kuvab graafikule ka vastava regressioonijooni võrrandi;
- Display R-squared on Chart kuvab graafikule R^2 väärtuse, mis iseloomustab, kui hästi andmed selle võrrandiga kokku sobivad. Ideaaljuhul on $R^2=1$. $0 < R^2 < 1$. R on Pearsoni korrelatsioonikordaja (vt allpool).

EXCELI FUNKTSIOONID

INTERCEPT(y_i'd; x_i'd) – Intercept(B7:B15;A7:A15) – leiab lõikepunkti y-teljega. ehk b väärtuse

SLOPE(y_i'd;x_i'd) – Slope(B7:B15;A7:A15) – leiab graafiku tõusu – k väärtuse.(ka LINEST(y_i'd;x_i'd)

Vt. ka Trend.

LISA

Statistilised arvutused

ANDMETEVAHELINE KORRELATSIOON

Andmete vahelise korrelatsiooni uurimiseks võib kasutada erinevaid meetodeid. Juhul, kui sõltumatu muutuja (x) ning sõltuva muutuja (y) vahel on lineaarne seos, mida võib üldjuhul kirjeldada seosega $y=kx+b$, kus k ja b on konstandid, kasutatakse seose 'headuse' määramiseks Pearsoni korrelatsioonikordajat $r_{xy}=r$, mis võib omandada väärtusi -1 kuni 1 . Mida suurem on Pearsoni korrelatsioonikordaja absoluutväärtus $|r|$, st mida lähedasem on see ühele, seda suurem on kindlus, et suuruste x ja y vahel on tõepoolest lineaarne seos.

Seega – Pearsoni korrelatsioonikordaja näitab, kas katsest saadud suuruste x ja y kogumid vastavad teooria poolt ennustatavale hüpoteesile, et nende vahel on lineaarne seos. Täpsemalt – et antud seos võiks kehtida. Kui andmeid on liiga vähe või mõõtmised ebatäpsed, võib r olla ligikaudu 1 ka juhul, kui x ja y vahel on näiteks ruutsõltuvus või eksponentsiaalne seos (iga graafik võib väga väikeses piirkonnas läheneda sirgele).

Pearsoni korrelatsioonikordaja võib leida järgmiselt:

$$r_{xy} = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \sqrt{n \sum_{i=1}^n y_i^2 - (\sum_{i=1}^n y_i)^2}} . \quad (L. 12)$$

Selle jaoks võib teha järgmise tabeli, mille abil on arvutusi lihtsam teha.

x	y	xy	x^2	y^2
...				
...				
$\sum_{i=1}^n x_i =$	$\sum_{i=1}^n y_i =$	$\sum_{i=1}^n x_i y_i =$	$\sum_{i=1}^n x_i^2 =$	$\sum_{i=1}^n y_i^2 =$

Kasutada võib ka mõnda tabelarvutusprogrammi, näiteks MS Excelis on selle jaoks funktsioon `Pearson(...;...)` (vt MS Excel help'i)