

## STATISTILISED ARVUTUSED EXCELIS

Käesolev juhend on mõeldud lisamaterjalina Audentese Kõrgemas Ärikoolis loetavale statistika kursusele. Juhendis antakse infot tabelarvutusprogrammi MS Excel statistiliste funktsioonide kohta. Põhjalikuma informatsiooni saamiseks on järgmised allikad:

1. Microsoft Excel Help (elektroonsel kujul, inglisekeelne)
2. Raamat "Excel 7.0 ('95) Funktsioonid. Tallinn: IT Koolitus, 1996, 218 lk."
3. Kiviste, A.. Matemaatiline statistika MS Excel keskkonnas. GT Tarkvara, Tallinn 1999.

Suur osa Exceli võimalustest põhineb funktsioonidel, mis võimaldavad teostada mahukaid arvutusi ilma matemaatiliste valemite koostamiseta. Funktsioonide sisestamiseks on mugav kasutada

nupurealt nuppu *Paste Function* .

Statistikas kasutatavad funktsioonid asuvad enamasti kategoorias *Statistical* või *Math&Trig*.

Enamike funktsioonide kasutamisel on vaja ära näidata **argumentid**.

**Argumentideks** võivad olla:

- arvude kogum (2;25;57)
- arvusid sisaldavate lahtrite piirkond (A2:A5)
- arvusid sisaldava lahtrite piirkonna nimi (kui sellele on omistatud nimi)

### Kirjeldavas statistikas kasutatavad funktsioonid

**AVERAGE** aritmeetilise keskmise  $\bar{x}$  leidmine

**number1**, number2, argumentid, millest soovitakse leida aritmeetilist keskmist

Arvutamiseks kasutatakse valemit  $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$ .

**COUNT** loendamine

**value1**, value2, argumentid

Loendatakse, mitu arvu on argumentide hulgas või argumentidega määratud lahtrites.

**COUNTIF** etteantud kriteeriumile vastavate väärtuste loendamine

**range** lahtrite piirkond

**criteria** kriteerium, milleks võib olla arv, avaldis, tekst või viide lahtrile, kus asub vastav kriteerium

Lahtrite piirkonnas loendatakse kokku need lahtrid, milles olev väärtus vastab kriteeriumile. Näiteks kui lahtrite piirkond A1:A6 sisaldab arve 1; 2; 2; 1; 2; 3 ja lahtris A8 on väärtus 2, siis COUNTIF(A1:A6;A8) annab vastuseks 3.

**SUM** summeerimine

**number1**, number2, argumentideks võivad olla arvud või lahtrite piirkonnad

**SUMIF** etteantud kriteeriumile vastavates lahtrites olevate väärtuste summeerimine

**range** lahtrite piirkond, mille väärtusi võrreldakse kriteeriumiga

**criteria** kriteerium, milleks võib olla arv, avaldis või tekst

**sum\_range** lahtrite piirkond, milles olevad väärtused summeeritakse. Selle puudumisel summeeritakse piirkonnaga *range* määratud lahtrid.

Näiteks kui lahtrite piirkond A1:A6 sisaldab arve 1; 2; 2; 1; 2; 3, siis SUM(A1:A6;"<3") annab vastuseks 8. Kui lahtrite piirkond A1:A4 sisaldab väärtusi Mart; Jüri; Mart, Tiina ja piirkonnas B1:B6 on nende poolt tehtud kulutused

500; 400; 600; 300, siis SUMIF(A1:A4;"Mart";B1:B4) annab vastuseks 1100.

**GEOMEAN** geomeetrilise keskmise  $\bar{x}_{geom}$  leidmine

**number1**, **number2**, argumendid, millest soovitakse leida geomeetrilist keskmist  
Ignoreeritakse teksti sisaldavaid ning tühje lahtrid. Kui mõni arv  $\leq 0$ , annab veateate #NUM!

Arvutamiseks kasutatakse valemit  $\bar{x}_{geom} = \sqrt[N]{x_1 \cdot x_2 \cdot \dots \cdot x_N}$ .

**HARMEAN** harmoonilise keskmise  $\bar{x}_{harm}$  leidmine

**number1**, **number2**, argumendid, millest soovitakse leida harmoonilist keskmist.  
Ignoreeritakse teksti sisaldavaid ning tühje lahtrid. Kui mõni arv  $\leq 0$ , annab veateate #NUM!

Arvutamiseks kasutatakse valemit  $\bar{x}_{harm} = \frac{1}{\frac{1}{N} \sum_{i=1}^N \frac{1}{x_i}}$

**MAX** kogumist suurima arvu leidmine

**number1**, **number2**, argumendid, millest soovitakse leida maksimumi

**MIN** kogumist vähima arvu leidmine

**number1**, **number2**, argumendid, millest soovitakse leida miinimumi

**MEDIAN** mediaani leidmine

**number1**, **number2**, argumendid, millest soovitakse mediaani leida.

Ignoreeritakse teksti sisaldavaid ja tühje lahtrid. Sõnadega väljendatud järjestikaskaala korral kasutada ei saa.

**MODE** moodi leidmine

**number1**, **number2**, argumendid, millest soovitakse moodi leida.

Ignoreeritakse teksti sisaldavaid ja tühje lahtrid. Sõnadega väljendatud nominaal- ja järjestikaskaala korral kasutada ei saa.

**QUARTILE** kvartiilide leidmine

**array** lahtrite piirkond, kus asuvad andmed

**quart** näitab, mitmendat kvartiili on vaja leida

Parameeter **quart** võib omada järgmisi väärtusi:

0	leitakse miinimum
1	leitakse 1. kvartiil
2	leitakse 2. kvartiil (mediaan)
3	leitakse 3. kvartiil
4	leitakse maksimum

Kui lahtrite piirkond on tühi või sisaldab rohkem kui 8191 väärtust, annab veateate #NUM! Kui mõni lahter ei sisalda arvu, annab veateate #VALUE!

**PERCENTILE** protsentiilide leidmine

**array** lahtrite piirkond, kus asuvad andmed

**k** parameeter, mis näitab, millist protsentiili soovitakse leida. NB!  $0 \leq k \leq 1$ , s.t. kui soovitakse leida 1. protsentiili, siis **k** peab olema 0,01

Kui lahtrite piirkond on tühi või sisaldab rohkem kui 8191 väärtust, annab veateate #NUM! Kui mõni lahter ei sisalda arvu, annab veateate #VALUE!

**VARP** kogumi dispersiooni  $\sigma^2$  leidmine

**number1**, **number2**, argumendid, millest soovitakse leida dispersiooni

Kasutatakse valemit  $\sigma^2 = \frac{\sum_i^n (x_i - \bar{x})^2}{N} = \frac{N \sum x_i^2 - (\sum x_i)^2}{N^2}$

**STDEVP** kogumi standardhälbe  $\sigma$  leidmine  
**number1**, **number2**, argumendid, millest soovitakse leida standardhälvet

Kasutatakse valemite 
$$\sigma = \sqrt{\frac{\sum_i^n (x_i - \bar{x})^2}{N}} = \sqrt{\frac{N \sum x_i^2 - (\sum x_i)^2}{N^2}}$$

## Kombinatoorikas kasutatavad funktsioonid

Funktsioonid FACT ja COMBIN asuvad kategoorias *Math&Trig*, PERMUT kategoorias *Statistical*

**FACT** faktoriaal positiivsest täisarvust  $n$   
**number** argument (arv  $n$  või vastava lahtri aadress)  
 Faktoriaal  $n! = 1 \cdot 2 \cdot 3 \cdot \dots \cdot n$

**PERMUT** variatsioonide arv  $n$  elemendist  $m$  kaupa  
**number** objektide koguarv  $n$   
**number\_chosen** objektide arv igas ühendis (mitme kaupa)  $m$   
 Arvutamiseks kasutatakse valemite  $V_n^m = \frac{n!}{(n-m)!}$ .

**COMBIN** kombinatsioonide arv  $n$  elemendist  $m$  kaupa  
**number** objektide koguarv  $n$   
**number\_chosen** objektide arv igas ühendis (mitme kaupa)  $m$   
 Arvutamiseks kasutatakse valemite  $C_n^m = \frac{n!}{m!(n-m)!}$ .

## Jaotusseaduste uurimisel kasutatavad funktsioonid

**BINOMDIST** tõenäosuse leidmine binoomjaotuse korral  
**number\_s** positiivsete tulemuste arv  $m$   
**trials** katsete arv  $n$   
**propability\_s** positiivse tulemuse tõenäosus  $p$   
**cumulative** tõenäosuse leidmisel 0  
 kumulatiivse tõenäosuse ledmisel 1

Tõenäosuse leidmiseks kasutatakse valemite  $P(m) = C_n^m p^m (1-p)^{n-m}$

Kumulatiivne tõenäosus  $P(x \leq m) = \sum_{k=0}^m P(k) = P(0) + P(1) + \dots + P(m)$

**POISSON** tõenäosuse leidmine Poissoni jaotuse korral  
**X** sündmuste arv  $m$  mingis intervallis  
**mean** keskmine sündmuste arv  $\lambda$  vastavas intervallis  
**cumulative** tõenäosuse leidmisel 0  
 kumulatiivse tõenäosuse ledmisel 1

Tõenäosuse leidmiseks kasutatakse valemite  $P(m) = e^{-\lambda} \frac{\lambda^m}{m!}$

Kumulatiivne tõenäosus  $P(x \leq m) = \sum_{k=0}^m P(k) = P(0) + P(1) + \dots + P(m)$

<b>NORMDIST</b>	jaotustiheduse $f(x)$ ja jaotusfunktsiooni $F(x)$ leidmine normaaljaotuse korral
<b>x</b>	väärtus, mille jaoks tahetakse vastavat funktsiooni leida
<b>mean</b>	kogumi keskvaartus $\bar{x}$
<b>standard_dev</b>	kogumi standardhälve $\sigma$
<b>cumulative</b>	jaotustiheduse leidmisel 0 jaotusfunktsiooni leidmisel 1

Jaotustihedus leitakse valemist  $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}}$

Jaotusfunktsiooni  $F(x)$  leidmiseks kasutatakse keerulisemaid matemaatilisi meetodeid

<b>NORMINV</b>	jaotusfunktsiooni $F(x)$ väärtusele vastav juhusliku suuruse väärtuse $x$ leidmine normaaljaotuse korral, funktsiooni NORMDIST pöördfunktsioon
<b>probability</b>	jaotusfunktsiooni väärtus (tõenäosus) $F(x)$
<b>mean</b>	kogumi keskvaartus $\bar{x}$
<b>standard_dev</b>	kogumi standardhälve $\sigma$

Arvutamiseks kasutatakse numbrilist meetodit.

<b>STDEV</b>	valimi standardhälbe $s$ leidmine
<b>number1, number2,</b>	argumendid, millest soovitakse leida standardhälvet

Kasutatakse valemit  $\sigma = \sqrt{\frac{\sum_i^n (x_i - \bar{x})^2}{N-1}} = \sqrt{\frac{N\sum x_i^2 - (\sum x_i)^2}{N(N-1)}}$

<b>CONFIDENCE</b>	üldkogumi keskvaartuse usalduspiiride leidmine
<b>alpha</b>	olulisuse nivoo, $\alpha = 1 - \beta$ , kus $\beta$ on usaldatavus. Näit. kui usaldatavus $\beta = 0,95$ , siis $alpha$ on $0,05$
<b>standard_dev</b>	kogumi standardhälve $\sigma$
<b>size</b>	kogumi maht $n$

Arvutamiseks kasutatakse valemit  $t_{\beta, \infty} \frac{\sigma}{\sqrt{n}}$ , kus  $t_{\beta, \infty}$  on Studenti tegur. Sobib suurte kogumite korral, kui  $n > 100$ .

Väiksemate kogumite korral tuleb funktsiooniga TINV leida Studenti tegur ja usaldusvahemiku laius arvutada eespool toodud valemi põhjal.

<b>TINV</b>	$t$ -testi parameetri kriitilise väärtuse leidmine
<b>probability</b>	olulisuse nivoo $\alpha$
<b>degrees_freedom</b>	vabadusastmete arv $\nu$

<b>CHIINV</b>	$\chi^2$ - testi parameetri kriitilise väärtuse leidmine
<b>probability</b>	olulisuse nivoo $\alpha$
<b>degrees_freedom</b>	vabadusastmete arv $\nu$

## Korrelatsioon- ja regressioonanalüüs

<b>CORREL</b>	korrelatsioonikoefitsiendi $r$ leidmine
<b>array1</b>	lahtrite piirkond, kus asuvad ühe suuruse $X$ väärtused $x_i$
<b>array2</b>	lahtrite piirkond, kus asuvad teise suuruse $Y$ väärtused $y_i$

Arvutamiseks kasutatakse valemit  $r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n\sigma_x\sigma_y}$ .

<b>COVAR</b>	kovariatsiooni leidmine
--------------	-------------------------

**array1** lahtrite piirkond, kus asuvad ühe suuruse  $X$  väärtused  $x_i$   
**array2** lahtrite piirkond, kus asuvad teise suuruse  $Y$  väärtused  $y_i$   
 Arvutamiseks kasutatakse valemit  $cov = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n}$ .

**SLOPE** regressioonsirge  $y = ax + b$  tõusu  $a$  leidmine  
**known\_y's** lahtrite piirkond, kus asuvad suuruse  $Y$  väärtused  $y_i$   
**known\_x's** lahtrite piirkond, kus asuvad suuruse  $X$  väärtused  $x_i$   
 Arvutamiseks kasutatakse valemit  $a = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - \left( \sum x_i \right)^2}$ .

**INTERCEPT** regressioonsirge  $y = ax + b$  vabaliikme  $b$  leidmine  
**known\_y's** lahtrite piirkond, kus asuvad suuruse  $Y$  väärtused  $y_i$   
**known\_x's** lahtrite piirkond, kus asuvad suuruse  $X$  väärtused  $x_i$   
 Arvutamiseks kasutatakse valemit  $b = \bar{y} - a\bar{x}$ , kus  $a$  leitakse funktsiooni SLOPE järgi

**TREND** lineaarse regressioonsirge  $y = ax + b$  punktide leidmine (silumine)  
**known\_y's** lahtrite piirkond, kus asuvad suuruse  $Y$  empiirilised väärtused  $y_i$   
**known\_x's** lahtrite piirkond, kus asuvad suuruse  $X$  empiirilised väärtused  $x_i$   
**new\_x's** lahtrite piirkond, kus asuvad suuruse  $X$  väärtused  $x_k$ , millele vastavaid regressioonsirgel asuvate punktide  $y$ -koordinaadi väärtusi tahetakse leida  
**const** 0, kui parameeter  $b=0$  (sirge peab läbima nullpunkti)

**STEYX** regressioonmudeli standardvea leidmine  
**known\_y's** lahtrite piirkond, kus asuvad suuruse  $Y$  väärtused  $y_i$   
**known\_x's** lahtrite piirkond, kus asuvad suuruse  $X$  väärtused  $x_i$