

## Biomeetria praks 6

### Illustreeritud (mittetäielik) tööjuhend

#### Eeltöö

1. Avage *MS Excel*'is oma kursuse ankeedivastuseid sisaldav andmestik,
2. lisage uus tööleht (*Insert / Lisa -> Worksheet / Tööleht*), nimetage see ümber leheküljeks 'Praks6' ja
3. kopeerige kogu 'Andmed'-lehel paiknev andmetabel leheküljele 'Praks6'.

#### Ülesanne 1.

- Illustreerige tunnuste 'PIKKUS' ja 'JALANR' vahelist seost hajuvus- ehk punktdiagrammiga.
- Jälgige, et x-telg vastaks jalanumbritele ja y-telg pikkustele. Vajadusel kujundage joonis ümber.
- Lisage joonisele lineaarne regressioonisirge, prognoosimaks pikkust jalanumbri alusel, samuti regressioonivõrrand ja viimase alusel leitavate prognooside täpsust kirjeldav determinatsioonikordaja  $R^2$ .
- Prognoosige leitud võrrandi alusel, keskmiselt kui pikk on jalanumbrit 40 omav tudeng.

#### Ülesanne 2.

- Teostage statistikaprotseduuri Regression (Tools/Tööriistad -> Data analysis...) abil lineaarne regressioonanalüüs prognoosimaks tudengite pikkust jalanumbri alusel.
- Kirjutage protseduuri tulemuste põhjal välja lineaarne regressioonivõrrand (ehk regressioonimudel) kujul
$$Pikkus = a + b \times Jalanumber,$$
kus  $a$  ja  $b$  asemel on *Excel*'i poolt välja arvatud kordajate väärtused.
- Kui suur on keskmiselt pikkuste vaheline erinevus tudengitel, kelle jalanumbrid erinevad 2 võrra?
- Kas leitud regressioonivõrrand on statistiliselt oluline? Põhjendus!
- Kirjeldamiseks prognooside täpsust, sõnastage üks lause kas mitmese korrelatsioonikordaja ( $R$ ) või determinatsioonikordaja ( $R^2$ ) kohta.

## Ülesande 1 tööjuhend

1. Illustreerige tunnuste 'PIKKUS' ja 'JALANR' vahelist seost hajuvus- ehk punktdiagrammiga.

**Joonisel peab x-telg vastama jalanumbritele ja y-telg pikkustele. Vajadusel kujundage joonis ümber (vt allpool toodud juhendit).**

	A	B	C	D	E	F	G	H	I	J
	RIIK	SUGU	PIKKUS	MASS	PEA_P	JALANR	ODE_VEN	MAT_HINN	HOMMIK	PUDE
1	Eesti	N	170	64	57	38	1	4	helbed või jah	
2	Eesti	N	170	60	57	40	0	4	võileib	nii ja
3	Eesti	N	163	64	54	38	2	5	helbed või jah	
4	Eesti	N	161,5	52	54	37,5	2	4	puder	jah
5	Eesti	N	171	75	55,5	40	1	4	puder	jah
6	Soome	N	178	70	52	40	1	5	puder	jah
7	Eesti	N	175	67	57	40	2	5	helbed või nii ja	
8	Eesti	N	160	52	56	38	2	4	võileib	nii ja
9	Eesti	N	160	55	56	36	3	4	võileib	jah
10	Eesti	N	170	69	52	36	1	4	võileib	nii ja
11	Eesti	N	167	57	57	38,5	1	4	ei söö taval	ei
12	Eesti	N	172	59	55,5	38	1	5	ei söö taval	nii ja
13	Eesti	N	167	60	45	39	1	3	puder	jah
14	Eesti	N	175	63	57	40	3	3	võileib	jah
15	Eesti	N	173	73	52	41	0	4	võileib	nii ja
16	Eesti	N	168	65	56	39	1	5	võileib	ei
17	Soome	N	170	80	57,5	39	2	3	võileib	jah
18	Eesti	N	165	70	56	39	4	3	võileib	jah
19	Soome	N	170	60	55,5	38	2	3	puder	jah
20	Eesti	N	175	60	55	38	0	4	võileib	jah
21	Eesti	N	160	50	56	37	2			
22	Eesti	N	160	51	56	36	1			
23	Eesti	N	167	55	56	37	2			
24	Eesti	N	161	54	54	38	2			
25	Eesti	N	165	54,3	54	38	2			
26	Eesti	N	158	55	55	35	0			
27	Soome	N	173	62	54	39	1			
28	Eesti	N	174	58	54	39	2			
29	Eesti	M	175	76	57	43	2			
30	Eesti	N	180	65	55	40	1			
31	Eesti	N	176	64	57	40	0			
32	Soome	N	164	63	42	38	2			
33	Soome	N	172	70	43,5	39	2			
34	Eesti	N	161	56	55	38	1			
35	Eesti	N	181	60	55	40	1			
36	Eesti	N	164	55	53	38	1			
37	Eesti	N	177	71	57	41	2			
38	Eesti	N	175	70	54	41	1			
39	Eesti	N	170	110	57	43	0			
40										

**Chart Wizard - Step 1 of 4 - Chart Type**

Standard Types Custom Types

Chart type: XY (Scatter)

Chart sub-type: Scatter, Compares pairs of values.

Press and Hold to View Sample

Next >

**Chart Wizard - Step 2 of 4 - Chart Source Data**

Data Range Series

Data range: =Praks6!\$C\$1:\$C\$48;Praks6!\$F\$1:\$F\$48

Series in:  Rows  Columns

Series: JALANR

Name: =Praks6!\$F\$1

X Values: =Praks6!\$C\$2:\$C\$48

Y Values: =Praks6!\$F\$2:\$F\$48

Add Remove

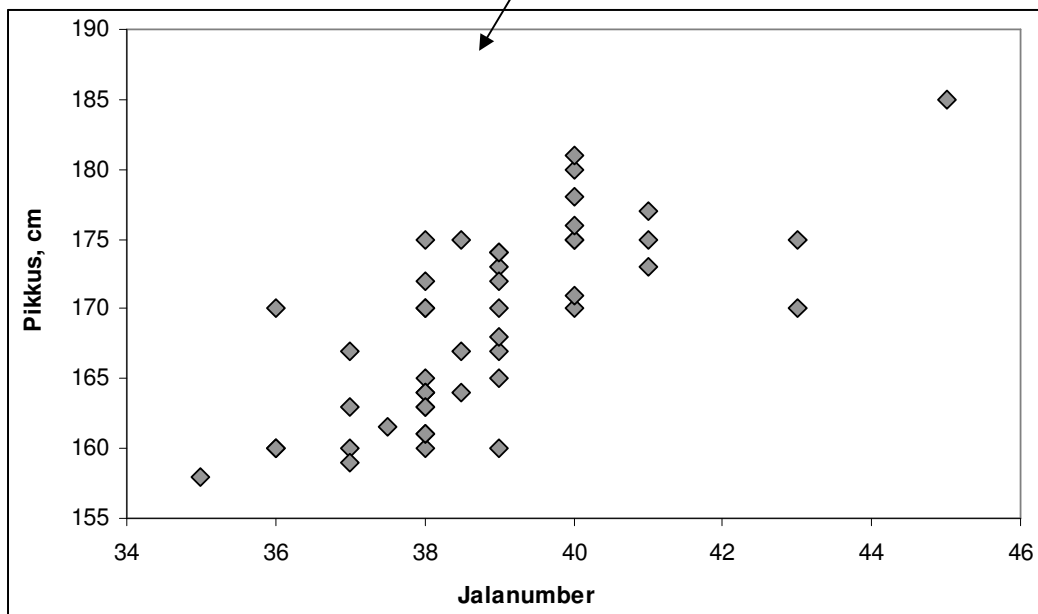
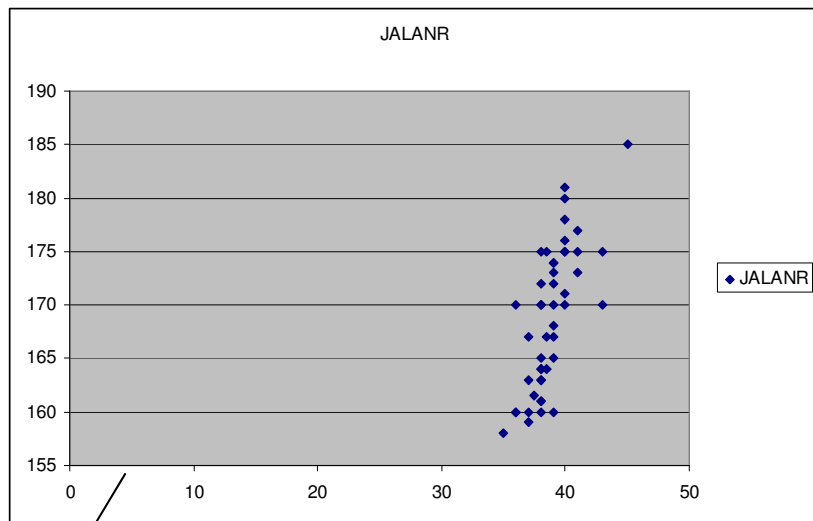
**Chart Wizard - Step 3 of 4 - Chart Options**

Name: =Praks6!\$F\$1

X Values: =Praks6!\$F\$2:\$F\$48

Y Values: =Praks6!\$C\$2:\$C\$48

- Pealkiri ja legend ×
- Taust valgeks
- Ruudujooned ×
- Pealkirjad x- ja y-teljele
- Telgede ühikud ja nimed kirjasuuruses 10
- X-telje ulatus 34-46 märgiste vahega 2
- X-telje ulatus 155-190 märgiste vahega 5
- Sümbolid suuremaks (ja seest näiteks helehalliks)

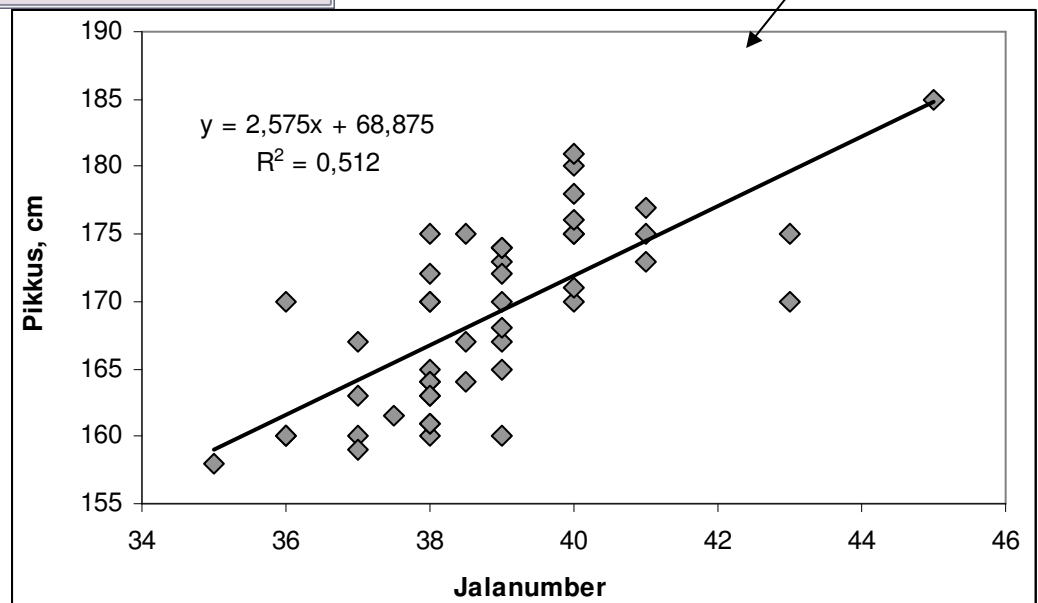
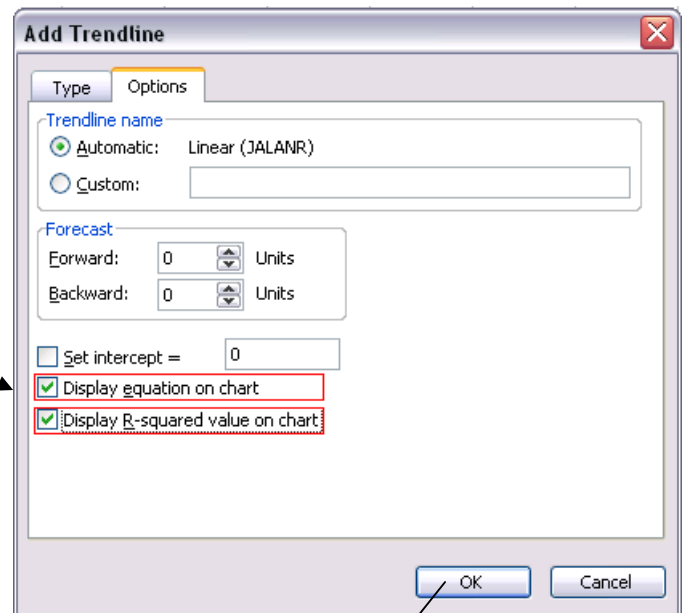
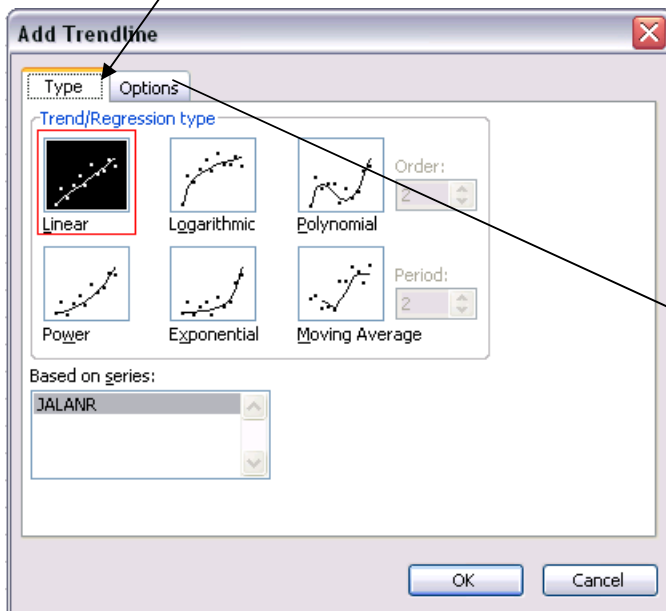
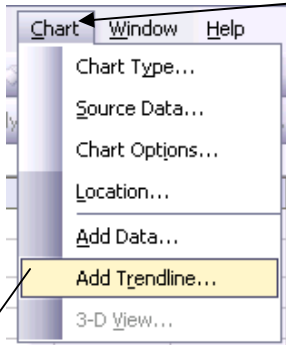
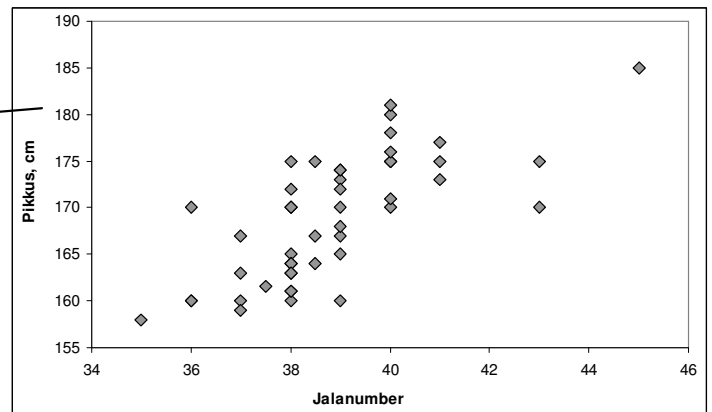


2. Lisage tunnuste 'PIKKUS' ja 'JALANR' hajuvusdiagrammile **regressioonisirge**, prognoosimaks pikkust jalanumbri alusel.

Lisage joonisele ka **regressioonivõrrand** ja viimase alusel leitavate prognooside täpsust kirjeldav **determinatsioonikordaja  $R^2$** .

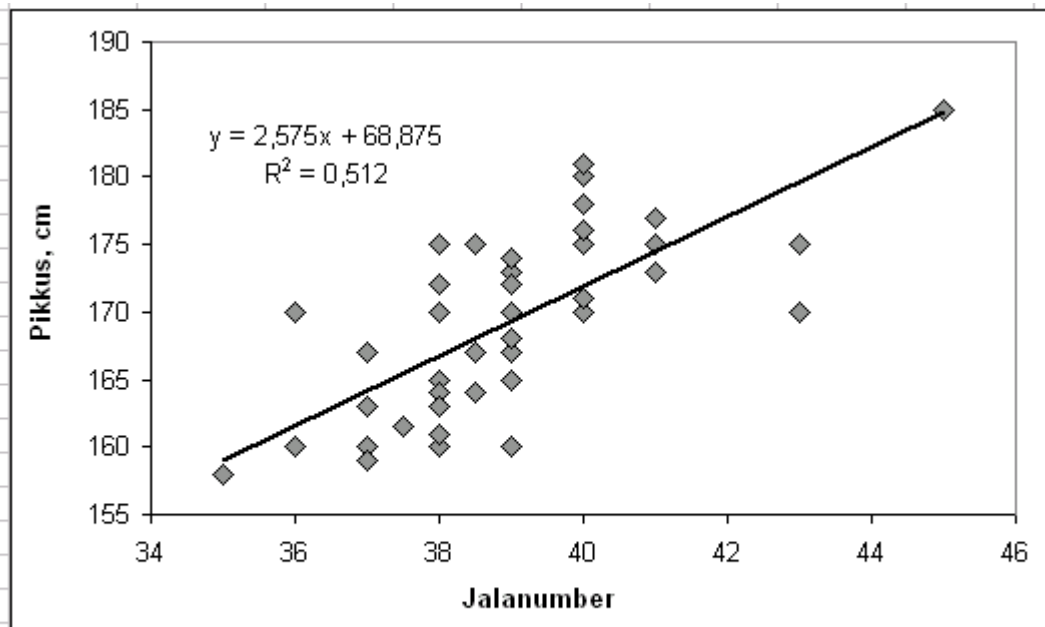
Chart/Diagramm

-> Add Trendline/Lisa Trendijoon



3. Prognosis leitud võrrandi alusel, keskmiselt kui pikk on jalanumbrit 40 omav tudeng.

Selleks pange joonise alla kirja *Excel*'i poolt välja arvutatud regressioonivõrrand, asendades lihtsalt suuruse  $x$  arvuga 40. ☺



Jalanumbrit 40 omava tudengi pikkus on keskmiselt  $=2,575*40+68,875$

## Ülesande 2 tööjuhend.

1. Teostage statistikaprotseduuri Regression (Tools/Tööriistad -> Data analysis...) abil lineaarne regressioonanalüüs prognoosimaks tudengite pikkust jalanumbri alusel.

**NB!** Protseduuri *Regression* rakendamisel ei tohi analüüsitavaid tunnuseid sisaldada puuduvaid väärtuseid. Vajadusel tuleb analüüsi tarvis teha abitabel ilma puuduvaid väärtuseid sisaldavate ridadeta.

### Regressioonanalüüsi tulemus

SUMMARY OUTPUT						
<i>Regression Statistics</i>						
Multiple R		0,71588346				
R Square		0,51248912				
Adjusted R Sq		0,50165555				
Standard Error		4,74509981				
Observations		47				
<i>ANOVA</i>						
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>	
Regression	1	1065,132314	1065,132	47,30563	1,54641E-08	
Residual	45	1013,21875	22,51597			
Total	46	2078,351064				
<i>Coefficients</i>						
	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>ower 95,0%</i> <i>pper 95,0%</i>
Intercept	68,875	14,55383669	4,732429	2,23E-05	39,56206862	98,187931
JALANR	2,575	0,374387039	6,877909	1,55E-08	1,820945807	3,3290542

2. Kirjutage protseduuri tulemuste põhjal välja lineaarne regressioonivõrrand (ehk regressiooni-mudel) kujul

$$Pikkus = a + b \times \text{Jalanumber},$$

kus  $a$  ja  $b$  asemel on *Excel*'i poolt välja arvatud kordajate väärtused.

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	68,875	14,55383669	4,732429	2,23E-05	39,56206862	98,187931
JALANR	2,575	0,374387039	6,877909	1,55E-08	1,820945807	3,3290542

3. Kui suur on keskmiselt pikkuste vaheline erinevus tudengitel, kelle jalanumbrid erinevad 2 võrra?

Vastus:  $2 \times b$  (aga arvuliselt?). **Pange arvuline vastus kirja täislauselga.**

4. Kas leitud regressioonivõrrand on statistiliselt oluline? Põhjendus!

ANOVA					
	df	SS	MS	F	Significance F
Regression	1	1065,132314	1065,132	47,30563	1,54641E-08
Residual	45	1013,21875	22,51597		
Total	46	2078,351064			

=  $p$

**Märkus.** Regressioonivõrrandi statistiline olulisus tähendab seda, et leitud regressioonivõrrand kujul

$$Pikkus = a + b \times \text{Jalanumber}$$

võimaldab pikkust täpsemalt prognoosida võrreldes konstantse võrrandiga

$$Pikkus = a.$$

Ehk siis, statistiliselt olulise regressioonivõrrandi korral võimaldab jalanumbri arvestamine pikkust täpsemalt prognoosida võrreldes konstateeringuga, et kõigi tudengite pikkused on ühesugused (ja võrdsed suurusega  $a$ ).

Hüpotheside paar, mille testimiseks vajaliku  $p$ -väärtuse väljastab *Excel* tabelisse ANOVA, on kujul:

$H_0$ : regressioonivõrrand ei ole statistiliselt oluline

$H_1$ : regressioonivõrrand on statistiliselt oluline

ehk

$H_0$ : leitud võrrand ei ole parem võrreldes konstantse võrrandiga

$H_1$ : leitud võrrand on parem võrreldes konstantse võrrandiga

ehk matemaatiliselt

$H_0$ :  $Pikkus = a$

$H_1$ :  $Pikkus = a + b \times \text{Jalanumber}$

Reaalselt rakendada on põhjust vaid statistiliselt olulist regressioonivõrrandit.

5. Sõnastage üks lause regressioonivõrrandist saadavate prognooside täpsuse kohta kas mitmese korrelatsioonikordaja ( $R$ ) või determinatsioonikordaja ( $R^2$ ) baasil.

SUMMARY OUTPUT	
<i>Regression Statistics</i>	
Multiple R	0,71588346
R Square	0,51248912
Adjusted R Sq	0,50165555
Standard Error	4,74509981
Observations	47

Mitmene korrelatsioonikordaja  $R$  mõõdab uuritava tunnuse ja tema prognoositud väärtuste vahelist korrelatsiooni. Mida suurem, seda parem!

Determinatsioonikordaja  $R^2$  näitab, kui suure osa uuritava tunnuse varieeruvusest võrrandist saadud prognoosid ära kirjeldavad,  $0 \leq R^2 \leq 1$ . Esitatakse enamasti protsentides. Mida suurem, seda parem!

Mudeli standardviga  $SE$  on prognoosijääkide standardhälve. Näitab tegelike ja prognoositud väärtuste vahelist keskmist erinevust (mudeli keskmist viga). Mida väiksem, seda parem!