

Biomeetria praks 6

Illustreeritud (mittetäielik) tööjuhend

Eeltöö

1. Avage MS Excel'is oma kursuse ankeedivastuseid sisaldav andmestik,
2. lisage uus tööleht, nimetage see ümber leheküljeks 'Praks6' ja
3. kopeerige kogu 'Andmed'-lehel paiknev andmetabel lehekülje 'Praks6' ülemisse vasakusse nurka.

Ülesanne 1.

- Illustreerige tunnuste 'PIKKUS' ja 'JALANR' vahelist seost hajuvus- ehk punktdiagrammiga.
- Jälgige, et x-telg (horisontaalne telg) vastaks jalanumbritele ja y-telg (vertikaalne telg) pikkustele. Vajadusel kujundage joonis ümber.
- Prognoosimaks pikkust jalanumbri alusel, lisage joonisele lineaarne regressioonisirge, 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 (Data-sakk -> 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.

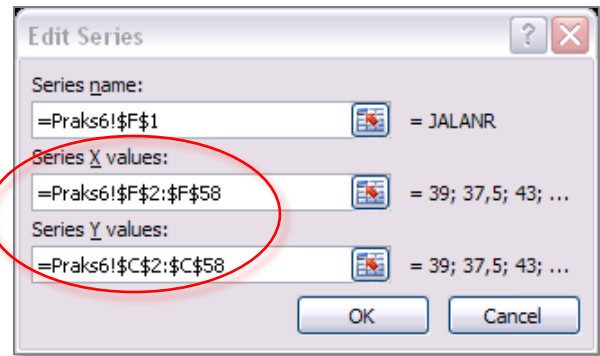
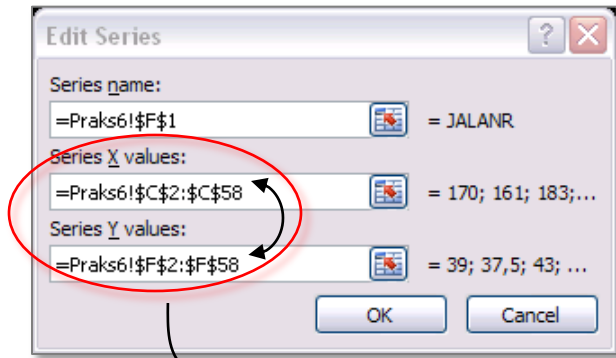
Ülesande 1 tööjuhend

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).

The image shows a Microsoft Excel spreadsheet with a data table and a scatter plot. The data table has columns for 'RIIK', 'SUGU', 'PIKKUS', 'MASS', 'PEA_P', 'JALANR', and various other variables. A scatter plot titled 'JALANR' is shown, with the x-axis labeled 'JALANR' and the y-axis labeled 'PIKKUS'. The plot shows a positive correlation between the two variables. A text box on the plot says: "Teljed on valepidi! Prognoosimaks pikkust jalanumbri alusel, peab pikkus olema y-teljel. Telgede vahetamiseks ...". Below the plot, there are two dialog boxes: 'Select Data Source' and 'Chart Tools - Design'. The 'Select Data Source' dialog box shows the chart data range as '=Praks6!\$C\$1:\$C\$58;Praks6!\$F\$1:\$F\$58' and the legend entries as 'JALANR'. The 'Chart Tools - Design' dialog box shows the chart layout and the legend entries.

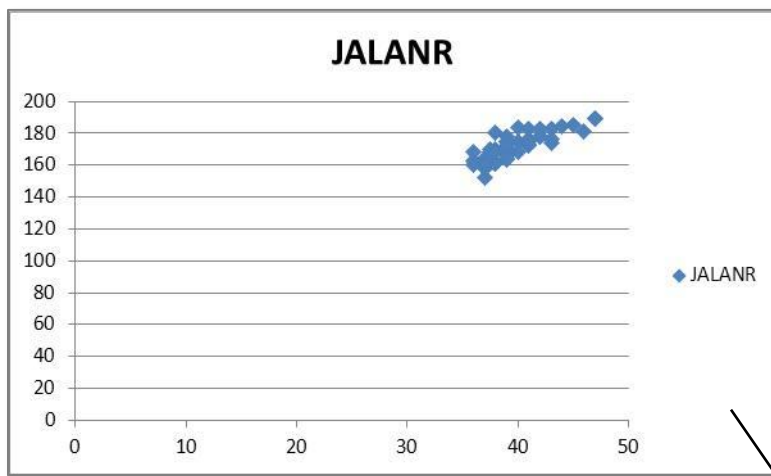
	A	B	C	D	E	F	G	H	I	J	K	L	M
1	RIIK	SUGU	PIKKUS	MASS	PEA_P	JALANR	ODE_VEN	MAT_HIN	HOMMIK	PUDER	LEMMIK	HAIGE	SPORT
2	Eesti	N	170	63	57	39	1	5	võieib	jah	ei	ei	jah
3	Eesti	N	161	57	59	37,5	2	4	võieib	jah	jah	jah	jah
4	Eesti	M	183	80	48	43	2	3	puder	jah	ei	ei	jah
5	Eesti	N	183	81	58	42	2	5	võieib	jah	jah	jah	ei
6	Eesti	M	174	74	56	43	0	4	võieib	nii ja naa	jah	ei	jah
7	Eesti	N	171	70	57	39	3	4	puder	jah	ei	ei	jah
8	Eesti	N	161	59	55	37	1	5	võieib	nii ja naa	jah	jah	jah
9	Eesti	N	172	80	56	41	2	4	võieib	nii ja naa	jah	ei	jah
10	Eesti	N	183,5	68	55,5	40	3	4	ei söö tav	nii ja naa	jah	jah	jah
11	Eesti	N	171	57	56	39	0	4	võieib	jah	jah	jah	jah
12	Soome	N	175	68	56	40	2	4	ei söö tav	ei	jah	ei	jah
13	Eesti	N	168	70	57	40	2	5	võieib	nii ja naa	jah	ei	jah
14	Eesti	N	152	50	52	37	1	3	võieib	jah	jah	ei	jah
15	Eesti	N	164	55	54	37	2	4	helbed võ	jah	jah	ei	jah
16	Eesti	N	160	60	55	37	2	3	helbed võ	jah	jah	ei	jah
17	Eesti	N	163,3	67	56	39	3	5	puder	jah	jah	ei	jah
18	Eesti	N	163	53	56	37	1	4	võieib	jah	jah	ei	jah
19	Soome	N	162	62	55	37	2	1	võieib	jah	jah	ei	jah
20	Soome	N	158	65	56	37	4	3	võieib	jah	jah	ei	jah
21	Eesti	N	170	57	57	39	8	4	ei söö tav	jah	jah	ei	jah
22	Eesti	N	170	58	56	37,5	2	3	võieib	jah	jah	ei	jah
23	Eesti	N	162	57	52	38	3	4	võieib	jah	jah	ei	jah
24	Eesti	N	169	63	56	39	1	4	võieib	ei	jah	ei	jah
25	Eesti	N	176	67	54	41	0	5	võieib	jah	jah	ei	jah
26	Eesti	N	173	64	54,5	41	1	5	helbed võ	jah	ei	ei	jah
27	Eesti	N	170	63	54	38	2	4	võieib	jah	jah	ei	jah
28	Eesti	N	167	73	56	39	1	3	võieib	nii ja naa	jah	ei	jah
29	Eesti	N	167	58	56	39	2	4	muu	ei	jah	ei	jah
30	Eesti	M	178	80	59	42	1	4	võieib	jah	jah	ei	jah
31	Eesti	M	189	71	60	47	2	3	helbed võ	nii ja naa	jah	ei	jah
32	Eesti	M	181	61	55	46	1	4	võieib	ei	jah	ei	jah
33	Eesti	M	189	71	60	47	2	3	helbed võ	nii ja naa	jah	ei	jah
34	Eesti	N	166	64	57	39	1	4	puder	jah	jah	ei	jah
35	Soome	N	169	65	57	39	1	4	puder	jah	jah	ei	jah
36	Soome	N	180	65	58	38	3	5	võieib	ei	jah	ei	jah
37	Soome	N	183	80	59	41	4	4	võieib	jah	jah	jah	ei
38	Soome	N	175	60	56	39	1	4	puder	jah	jah	ei	ei
39	Eesti	N	178	77	57	39	1	4	võieib	jah	jah	ei	ei
40	Eesti	M	175	64	52	39	2	4	võieib	nii ja naa	jah	ei	jah
41	Eesti	M	178	65	57	42	2	5	võieib	nii ja naa	jah	ei	jah
42	Eesti	N	168	52	50	37,5	2	5	võieib	nii ja naa	jah	ei	ei
43	Eesti	N	160	62	53	36	2	4	võieib	nii ja naa	jah	ei	jah
44	Eesti	M	185	80	57	45	2	4	võieib	nii ja naa	jah	ei	jah
45	Eesti	N	175	85	59	41	2	4	võieib	nii ja naa	jah	ei	ei
46	Eesti	N	167	65	55	39	2	5	võieib	nii ja naa	jah	ei	ei
47	Eesti	N	176	74	57	43	2	5	võieib	nii ja naa	jah	ei	ei
48	Eesti	N	161	62	56	38	2	5	võieib	nii ja naa	jah	ei	ei
49	Soome	N	163	52	55	36	2	5	võieib	nii ja naa	jah	ei	ei
50	Eesti	N	161	62	56	38	2	5	võieib	nii ja naa	jah	ei	ei
51	Eesti	N	161,5	55	54	38	2	4	võieib	nii ja naa	jah	ei	jah
52	Eesti	N	163	53	53	37	2	4	võieib	nii ja naa	jah	ei	ei
53	Eesti	N	172	65	56	39	2	5	võieib	nii ja naa	jah	ei	ei
54	Eesti	M	184	83	58	44	1	5	võieib	nii ja naa	jah	ei	ei
55	Eesti	N	164	75	60	39	0	5	võieib	nii ja naa	jah	ei	ei
56	Eesti	N	168	57	56	36	5	5	võieib	nii ja naa	jah	ei	ei
57	Eesti	N	168	60	56	38	2	5	võieib	nii ja naa	jah	ei	ei
58	Eesti	N	160	49	52	37	2	5	võieib	nii ja naa	jah	ei	ei



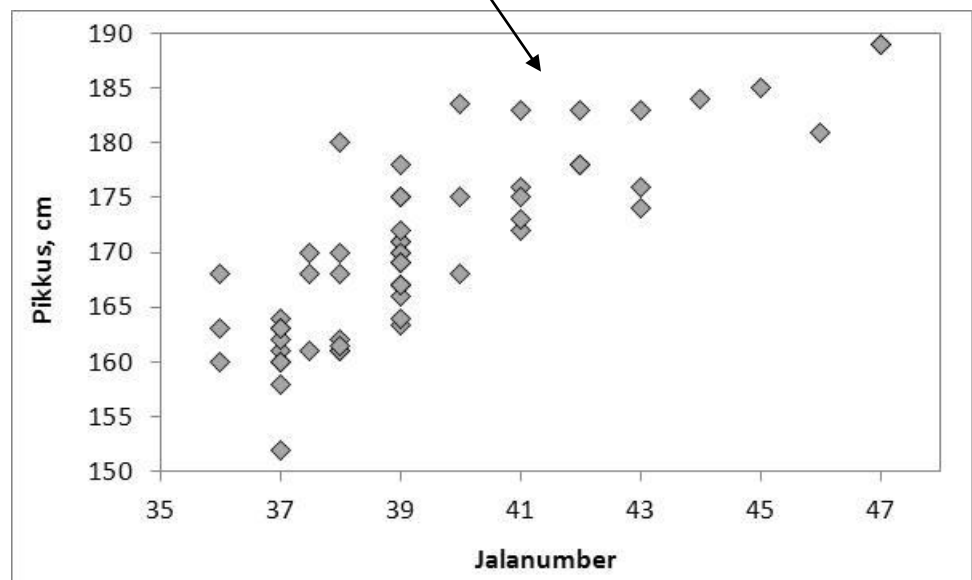
Vahetage x- ja y-telje väärtused.

Tulemus:

pikkuse väärtused on y- ja jalanumbri väärtused x-teljel.



Kujundage joonis!



2. Prognosimaks pikkust jalanumbri alusel, lisage tunnuste 'PIKKUS' ja 'JALANR' hajuvusdiagrammile **regressioonisirge**.

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

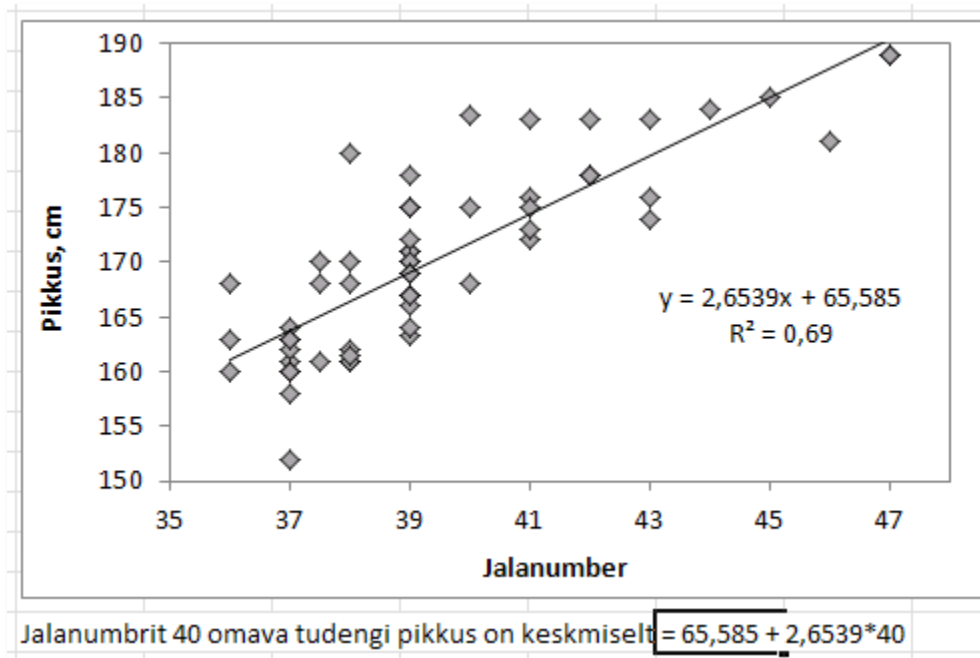
The image shows a Microsoft Excel interface with a scatter plot of 'Pikkus, cm' (Height, cm) on the y-axis and 'Jalanumber' (Foot number) on the x-axis. A linear trendline is applied to the data. The 'Chart Tools' ribbon is active, and the 'Format Trendline' task pane is open. A context menu is open over the trendline, and the 'Format Trendline...' option is selected. The 'Format Trendline' dialog box is open, showing 'Trendline Options' with 'Linear' selected. The 'Display Equation on chart' and 'Display R-squared value on chart' options are checked and highlighted with an orange box.

Lineaarse trendijoone lisamiseks

Lisavalikute tarvis

Märkige, saamaks joonisele regressioonivõrrandit ja R^2 väärtust

3. Prognoosige 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. ☺



Ülesande 2 tööjuhend.

- Teostage statistikaprotseduuri Regression (Data-sakk -> Data analysis...) abil lineaarne regressioonanalüüs prognoosimaks tudengite pikkust jalanumbri alusel.

The screenshot shows the Microsoft Excel interface with the 'Data Analysis' task pane open. The 'Regression' option is selected. The 'Input Y Range' is set to '\$C\$1:\$C\$58' and the 'Input X Range' is set to '\$F\$1:\$F\$58'. The 'Labels' checkbox is checked. The 'Output Range' is set to '\$W\$20'. The 'Confidence Level' is set to 95%.

	A	B	C	D	E	F	G	H	I
1	RIIK	SUGU	PIKKUS	MASS	PEA_P	JALANR	ODE_VEN	MAT_HIM	
2	Eesti	N	170	63	57	39	1	5	asta jooksul
3	Eesti	N	161	57	59	37,5	2	4	asta jooksul
4	Eesti	M	183	80	48	43	2	3	ju jooksul
5	Eesti	N	183	81	58	42	2	5	ju jooksul
6	Eesti	M	174	74	56	43	0	4	asta jooksul
7	Eesti	N	171	70	57	39	3	4	asta jooksul
8	Eesti	N	161	59	55	37	1	5	ju jooksul
9	Eesti	N	172	80	56	41	2	4	ju jooksul
10	Eesti	N	183,5	68	55,5	40	3	4	ju jooksul
11	Eesti	N	171	57	56	39	0	4	ju jooksul
12	Soome	N	175	68	56	40	2	4	asta jooksul
13	Eesti	N	168	70	57	40	2	5	ii aasta tagasi
14	Eesti	N	152	50	52	37	1	3	ju jooksul
15	Eesti	N	164	55	54	37	2	4	asta jooksul
16	Eesti	N	160	60	55	37	2	3	asta jooksul
17	Eesti	N	163,3	67	56	39	3	5	päeva jooksul
18	Eesti	N	163	53	56	37	1	4	asta jooksul
19	Soome	N	162	62	55	37	2	1	päeva jooksul
20	Soome	N	158	65	56	37	4	3	asta jooksul
21	Eesti	N	170	57	57	39	8	4	asta jooksul
22	Eesti	N	170	58	56	37,5	2	3	ju jooksul
23	Eesti	N	162	57	52	38	3	4	asta jooksul
24	Eesti	N	169	63	56	39	1	4	asta jooksul
25	Eesti	N	176	67	54	41	0	5	ju jooksul
26	Eesti	N	173	64	54,5	41	1	5	asta jooksul
27	Eesti	N	170	63	54	38	2	4	asta jooksul
28	Eesti	N	167	73	56	39	1	3	ju jooksul
29	Eesti	N	167	58	56	39	2	4	ju jooksul
30	Eesti	M	178	80	59	42	1	4	asta jooksul
31	Eesti	M	189	71	60	47	2	3	ju jooksul
32	Eesti	M	181	61	55	46	1	4	päeva jooksul
33	Eesti	M	189	71	60	47	2	3	ju jooksul
34	N	166	54	57	39	1	4	päeva jooksul	
35	Soome	N	169	65	57	39	1	4	asta jooksul
36	Soome	N	180	65	58	38	3	5	asta jooksul
37	Soome	N	183	59	41	4	4	4	asta jooksul
38	Soome	N	175	60	56	39	1	4	asta jooksul
39	Eesti	N	178	77	57	39	1	4	ju jooksul
40	Eesti	M	175	64	52	39	2	4	ju jooksul
41	Eesti	M	178	65	57	42	1	4	ju jooksul
42	Eesti	N	168	52	50	37,5	1	4	päeva jooksul
43	N	160	62	53	36	2	4	päeva jooksul	
44	Eesti	M	185	80	57	45	3	5	asta jooksul
45	Eesti	N	175	85	59	41	3	5	asta jooksul
46	Eesti	N	167	65	55	39	0	5	asta jooksul
47	Eesti	N	176	74	57	43	3	4	ii aasta tagasi
48	Eesti	N	161	62	56	38	2	4	ju jooksul
49	Soome	N	163	52	55	36	2	4	päeva jooksul
50	Eesti	N	161	62	56	38	2	4	ju jooksul
51	Eesti	N	161,5	55	54	38	2	4	ju jooksul
52	Eesti	N	163	53	53	37	2	4	ju jooksul
53	Eesti	N	172	65	56	39	2	5	asta jooksul
54	Eesti	M	184	83	58	44	1	5	ju jooksul
55	Eesti	N	164	75	60	39	0	4	päeva jooksul
56	Eesti	N	168	57	56	36	5	4	päeva jooksul
57	Eesti	N	168	60	56	38	2	3	ju jooksul
58	Eesti	N	160	49	52	37	2	4	ju jooksul

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

Regressioonanalüüsi tulemus:

SUMMARY OUTPUT								
<i>Regression Statistics</i>								
Multiple R	0,830676							
R Square	0,690022							
Adjusted R Square	0,684386							
Standard Error	4,781898							
Observations	57							
<i>ANOVA</i>								
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>			
Regression	1	2799,59514	2799,6	122,432	1,314E-15			
Residual	55	1257,65995	22,8665					
Total	56	4057,25509						
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95,0%</i>	<i>Upper 95,0%</i>
Intercept	65,58481	9,49945587	6,90406	5,4E-09	46,547474	84,6221	46,5475	84,6221
JALANR	2,653925	0,23985086	11,0649	1,3E-15	2,173253	3,1346	2,17325	3,1346

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.

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	65,584809	9,499455874	6,90406	5,42E-09	46,54747401	84,62214
JALANR	2,65392492	0,239850865	11,0649	1,31E-15	2,173253048	3,134597

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

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

4. Kas leitud regressioonivõrrand on statistiliselt oluline? Põhjendus!

ANOVA					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	1	2799,595136	2799,595	122,4319	1,31399E-15
Residual	55	1257,659951	22,86654		
Total	56	4057,255088			

= *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*).

Hüpoteeside 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	0,83067561
R Square	0,69002197
Adjusted R Square	0,684386
Standard Error	4,78189759
Observations	57

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!