

Biomeetria praks 5

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 'Praks5' ja
3. kopeerige kogu 'Andmed'-lehel paiknev andmetabel leheküljele 'Praks5'.

Ülesanne 1.

Kas tudengite pikkus ja jalanumber on seotud? Uurige seost *MS Excel*'i funktsioonide abil.

- Leidke tunnuste 'PIKKUS' ja 'JALANR' vaheline lineaarne korrelatsioonikordaja;
- kirjeldage pikkuse ja jalanumbri vahelist seost leitud kordaja alusel;
- testige seose statistilist olulisust:
 - sõnastage null- ja alternatiivne hüpotees (pange need töölehele kirja),
 - kontrollige nende kehtimist (leidke vaatluspaaride arv n , teststatistik t ja nende alusel olulisuse tõenäosus p),
 - pange kirja lõppjärelus.

Ülesanne 2.

Illustreerige tunnuste 'PIKKUS' ja 'JALANR' vahelist seost hajuvus- e punktdiagrammiga (inglise keeles *scatter plot*).

Ülesanne 3.


- Kasutades statistikaprotseduuri *Correlation (Tools/Tööriistad -> Data analysis...)* leidke korruga kõigi andmestikku kuuluvate pidevate arv-tunnuste (pikkus – jalanumber) vahelised lineaarsed korrelatsioonikordajad.
- Milliste tunnuste vahel on kõige tugevam lineaarne seos? Aga kõige nõrgem?
- Kirjeldage mõnda seost (pange vastavad laused ka kirja).

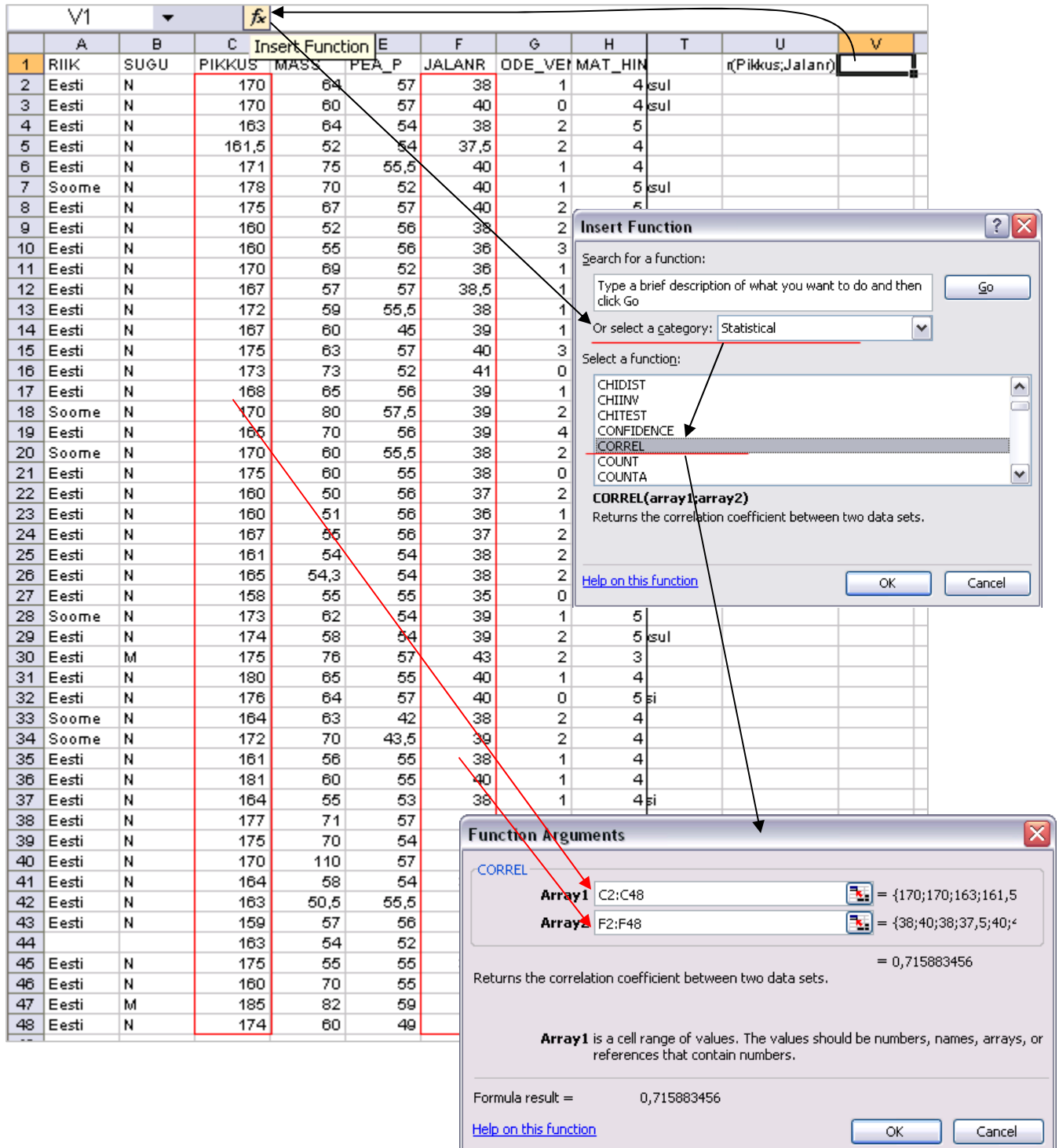
Ülesande 1 tööjuhend

1. Et *MS Excel*'i funktsioonide tulemuseks on enamasti vaid üks kommenteerimata väärtus, on enne millegi välja arvutamist soovitatav kirja panna, mis see on, mida arvutatakse. Näiteks antud juhul on sooviks leida tudengite pikkuste ja jalanumbrite vahelist lineaarset korrelatsioonikordajat – *MS Excel*'i töölehele võiks siis trükkida
 - 'Pikkuse ja jalanumbri vaheline lineaarne korrelatsioonikordaja'
 - või lühemalt '**r(Pikkus;Jalanr)**' – lineaarset korrelatsioonikordajat tähistatakse enamasti tähega 'r'.Seejärel pange kursor tühja lahtrisse, millesse soovite vastavat korrelatsioonikordajat arvutada.

2. Lineaarne korrelatsioonikordaja on arvutatav funktsiooniga CORREL, millel on kaks argumenti – esimese tunnuse väärtuste blokk ja teise tunnuse väärtuste blokk.

Kogenumad *MS Excel*'i kasutajad võivad kohealt sisestada arvutamiskäsu kujul
 =CORREL (C2:C48 ; F2 : F48)

Vähem kogenumatel (eelnevast valemist 100%-liselt aru mittesaanutel) on soovitatav klikkida nupul  ja jätkata vastavalt järgevale joonisele.



The screenshot shows an Excel spreadsheet with columns A through V. Column C is labeled 'PIKKUS' and column F is labeled 'JALANR'. A red box highlights the data ranges C2:C48 and F2:F48. Two dialog boxes are overlaid on the spreadsheet:

- Insert Function:** This dialog box is open, showing the 'Statistical' category selected. The 'CORREL' function is chosen from the list. The description reads: 'CORREL(array1;array2) Returns the correlation coefficient between two data sets.'
- Function Arguments:** This dialog box is open, showing the arguments for the CORREL function:
 - Array1: C2:C48 (representing the array {170;170;163;161,5})
 - Array2: F2:F48 (representing the array {38;40;38;37,5;40;...})
 The calculated result is shown as 0,715883456.

3. Kirjeldage, milline on tudengite pikkuste ja jalanumbrite vaheline seos

- kui tugev (nõrk / keskmise tugevusega / tugev);
- kas positiivne või negatiivne (mida see positiivne või negatiivne tähendab?).

NB! See järeldus tuleneb üksnes seose positiivsusest/negatiivsusest! Lünka tuleks kirjutada sõna „suurem“ või „väiksem“.

r(Pikkus;Jalanr)	0,715883			
Pikkuse ja jalanumbri vahel on seos.				
St, et mida suurem on pikkus, seda on keskmiselt jalanumber.				
Hüpoteeside paar				
H ₀ : (ehk matemaatiliselt)				
H ₁ : (ehk matemaatiliselt)				

4. Pange kirja korrelatsioonikordaja kohta kontrollitav hüpoteeside paar (soovitavalt nii teksti kujul kui ka matemaatiliselt).

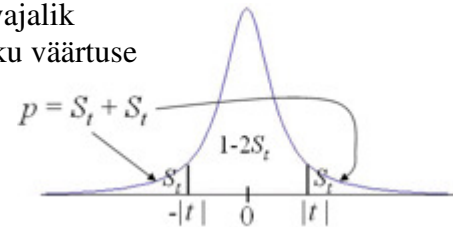
Meeldetuletuseks teooriast – hüpoteeside testimine korrelatsioonikordaja kohta

Korrelatsioonikordaja 0-st erinevuse testimiseks (st seose statistilise olulisuse testimiseks) MS Excel'is tuleb arvutada teststatistik, mis on nullhüpoteesi kehtides t-jaotusega, valemist

$$t = r\sqrt{n-2} / \sqrt{1-r^2} \underset{H_0}{\sim} t_{n-2}$$

Suurus *r* selles valemis on arvatud korrelatsioonikordaja väärtus, *n* aga vaatluspaaride arv (ehk nende tudengite arv, kelle kohta on teada nii pikkuse kui ka jalanumbri väärtus).

Otsuse, kumb hüpoteesidest on õige, vastu võtmiseks vajalik olulisuse tõenäosus *p* kujutab enesest leitud teststatistiku väärtuse poolt ära lõigatud t-jaotuse sabade osakaalu (joonisel pindalade *S_t* summa).



Excel'is on p-väärtus leitav funktsiooniga TDIST (ABS (t) ; n-2 ; 2).

5. Olulisuse tõenäosuse p leidmiseks vajalikke arvutusi on mugav teostada, kui kõik vajalikud suurused on töölehele kirja pandud.

Näiteks kujul:

- a) Lahtri 'n(Pikkus;Jalanr)' järele tuleks kirjutada nende tudengite arv, kelle andmete alusel on korrelatsioonikordaja arvatud (ehk nende tudengite arv, kelle kohta on teada nii nende pikkus kui ka jalanumber).

r(Pikkus;Jalanr)	0,715883			
Pikkuse ja jalanumbri vahel on seos.				
St, et mida suurem on pikkus, seda on keskmiselt jalanumber.				
Hüpoteeside paar				
H ₀ : (ehk matemaatiliselt)				
H ₁ : (ehk matemaatiliselt)				
n(Pikkus;Jalanr)				
t(Pikkus;Jalanr)				
p(Pikkus;Jalanr)				

b) Lahtri 't(Pikkus;Jalanr)' järele tuleb sisestada valem teststatistiku absoluutväärtuse arvutamiseks:

	T	U	V	W	X	Y
1		r(Pikkus;Jalanr)	0,715883			
2						
3		Pikkuse ja jalanumbri vahel on seos.				
4		St, et mida suurem on pikkus, seda on keskmiselt jal				
5						
6		Hüpoteeside paar				
7		H ₀ : (ehk matemaatiliselt)				
8		H ₁ : (ehk matemaatiliselt)				
9						
10						
11		n(Pikkus;Jalanr)	47			
12						
13		t(Pikkus;Jalanr)	=ABS(V1*SQRT(V11-2)/SQRT(1-V1*V1))			

c) Lahtri 'p(Pikkus;Jalanr)' järele tuleb sisestada funktsioon TDIST kolme argumendiga:

- teststatistiku absoluutväärtus |t|,
- (vaatluspaaride arv) – 2, so vastava t-jaotuse parameeter (n – 2), ja
- arv 2 (mis näitab, et testime kahepoolset hüpoteesi – mida see tähendab?).

	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE
1		r(Pikkus;Jalanr)	0,715883									
2												
3		Pikkuse ja jalanumbri vahel on seos.										
4		St, et mida suurem on pikkus, seda on keskmiselt jal										
5												
6		Hüpoteeside paar										
7		H ₀ : (ehk matemaatiliselt)										
8		H ₁ : (ehk matemaatiliselt)										
9												
10												
11		n(Pikkus;Jalanr)	47									
12												
13		t(Pikkus;Jalanr)	6,877909									
14												
15		p(Pikkus;Jalanr)	=TDIST(V13;V11-2;2)									
16												
17												

Function Arguments

TDIST

x V13 = 6,877909042

Deg_freedom V11-2 = 45

Tails 2 = 2

= 1,54641E-08

Returns the Student's t-distribution.

X is the numeric value at which to evaluate the distribution.

Formula result = 1,54641E-08

[Help on this function](#)

OK Cancel

6. Tehke formaalne otsus, kumb püstitatud hüpoteesidest on õige ja miks.

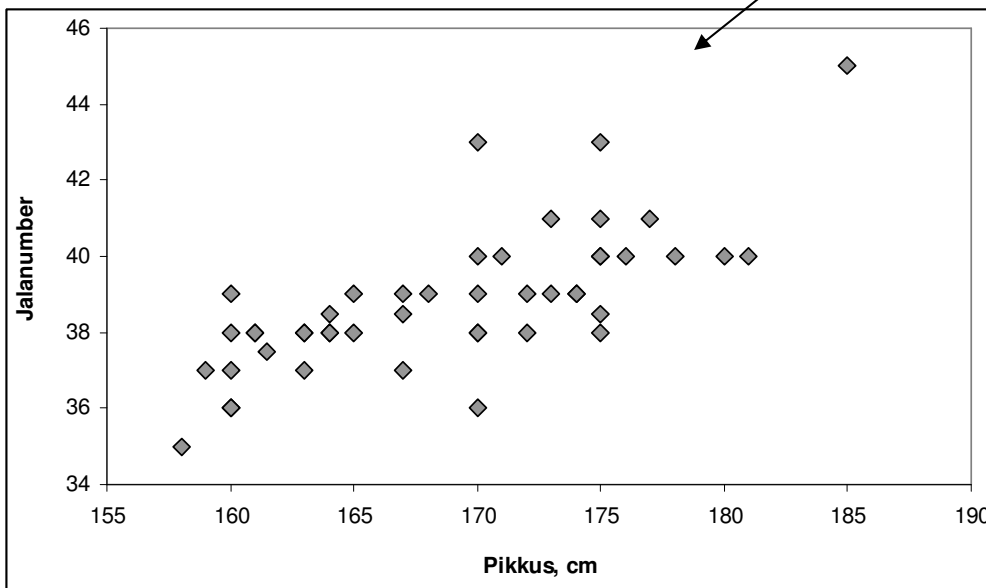
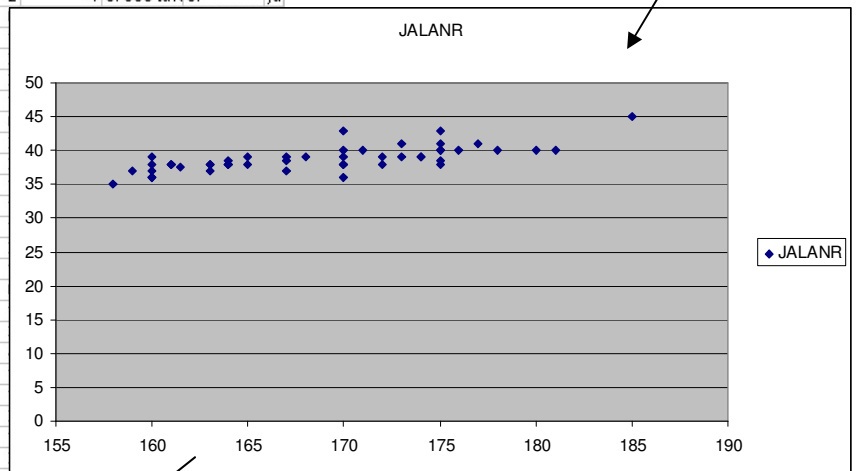
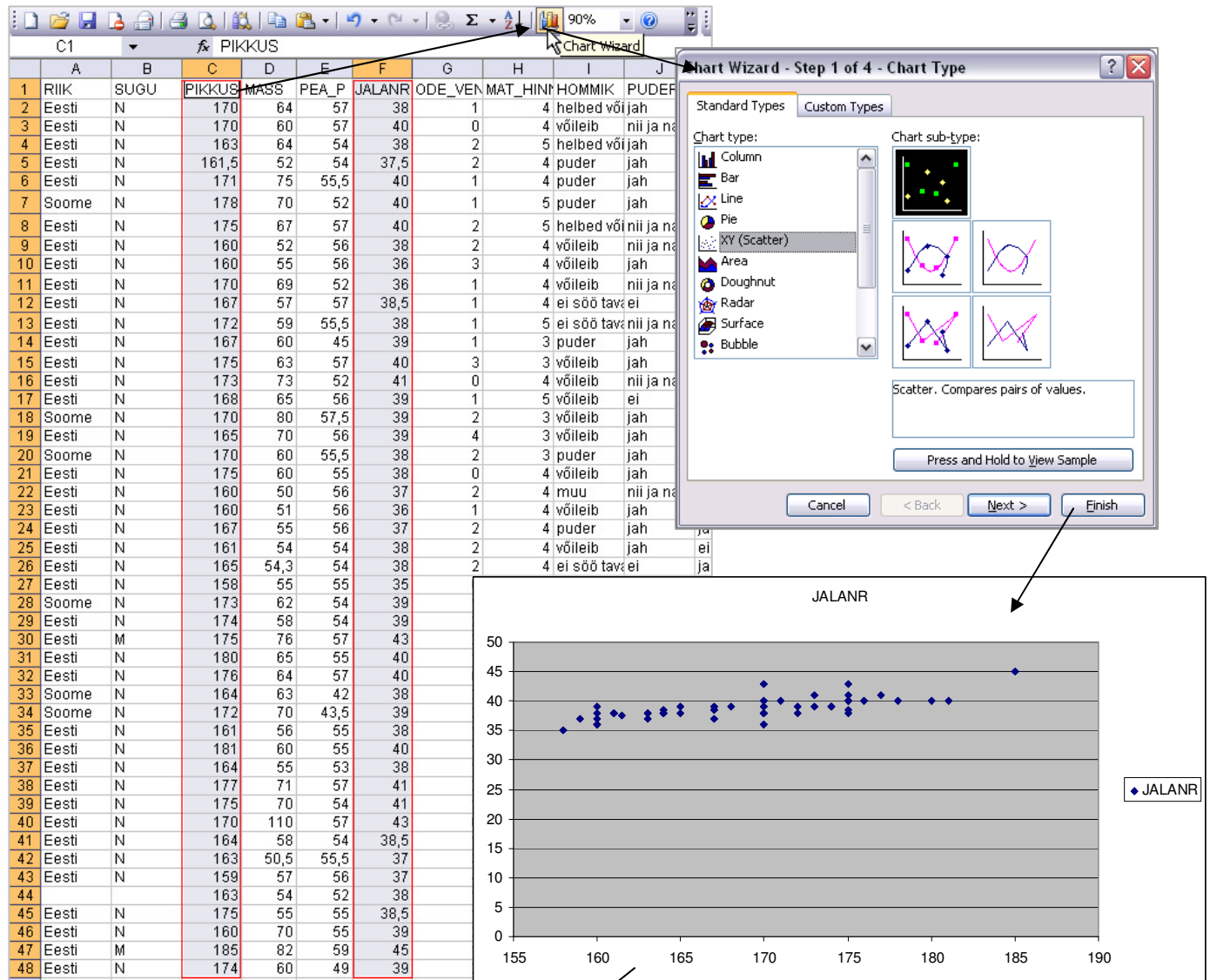
A'la: p(Pikkus;Jalanr) 1,55E-08 < 0,05 => H₁: tudengite pikkus ja jalanumber on seotud
 Märkus. 1,55E-08 = 1,55×10⁻⁸

7. Pange kirja päris lõppjärelus.

A'la: tudengite pikkuse ja jalanumbri vahel on statistiliselt oluline tugev positiivne seos (r = 0,716; p < 0,001).

Ülesande 2 tööjuhend

Illustreerige tunnuste 'Pikkus' ja 'Jalanr' vahelist seost hajuvus- ehk punktdiagrammiga.



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

Ülesande 3 tööjuhend

1. Kasutades statistikaprotseduuri *Correlation (Tools/Tööriistad -> Data analysis... -> Correlation)* leidke korraga kõigi andmestikku kuuluvate pidevate arv-tunnuste vahelised korrelatsioonikordajad.

The screenshot shows the following steps in Excel:

- The **Tools** menu is open, and **Data Analysis...** is selected.
- The **Data Analysis** task pane is open, and **Correlation** is selected.
- The **Correlation** dialog box is open, with the following settings:
 - Input Range:** \$C\$1:\$F\$48
 - Grouped By:** Columns
 - Labels in first row:**
 - Output Range:** \$U\$20
- The correlation matrix is displayed in the background, showing the following values:

	PIKKUS	MASS	PEA_P	JALANR
PIKKUS	1			
MASS	0,46263	1		
PEA_P	0,09441	0,10159	1	
JALANR	0,71588	0,72931	0,17293	1

2. Milliste tunnuste vahel on kõige tugevam lineaarne seos? Aga kõige nõrgem?
Kas jalanumber on tugevamini seotud pikkusega või kehamassiga?
Millise tunnusega on enim seotud peaümberrõõd?

Kirjeldage mõnda seost (pange vastavad laused ka kirja)!