

Biomeetria praks 5

Illustreeritud (mittetäielik) tööjuhend

Eeltöö

1. Avage *MS Excelis* tudengite ankeedivastuseid sisaldav andmestik,
 2. lisage uus tööleht, nimetage see ümber leheküljeks 'Praks5' ja
 3. kopeerige kogu 'Andmed'-lehel paiknev andmetabel lehekülje 'Praks5' ülemisse vasakusse nurka.
-

Ülesanne 1.

Kas tudengite pikkus ja jalanumber on seotud? Uurige seost *MS Exceli* funktsioonide abil.

- Leidke tunnuste 'HEIGHT' ja 'SHOE_SIZE' 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 lõppjärelendus.

Ülesanne 2.

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

Ülesanne 3.

- Kasutades statistikaprotseduuri *Correlation (Data-sakk -> Data analysis...)* leidke korruga kõigi andmestikku kuuluvate pidevate arvtunnuste (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 Exceli* 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 eesmärgiks leida tudengite pikkuse ja jalanumbri vaheline lineaarne korrelatsioonikordaja – *MS Exceli* töölehele võiks siis trükkida


- 'Pikkuse ja jalanumbri vaheline lineaarne korrelatsioonikordaja'
- või lühemalt '**r(Height;Shoe_size)**', sest lineaarset korrelatsioonikordajat tähistatakse enamasti tähega 'r'.

Seejärel pange kursor tühja lahtrisse, millesse soovite vastavat korrelatsioonikordajat arvutada.

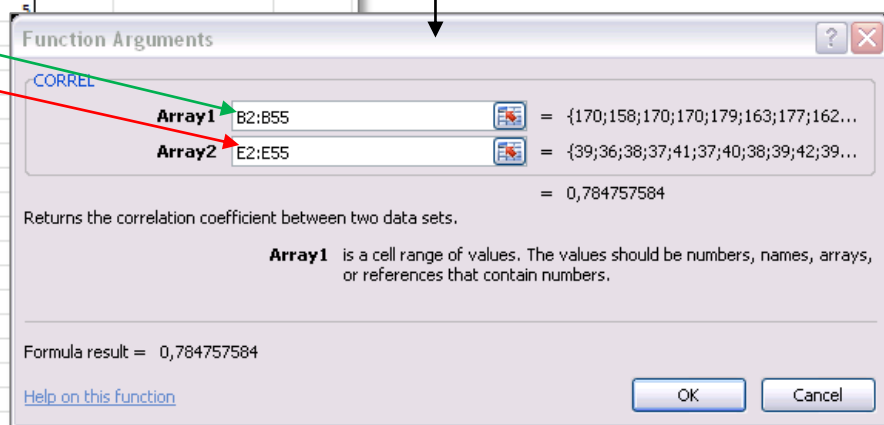
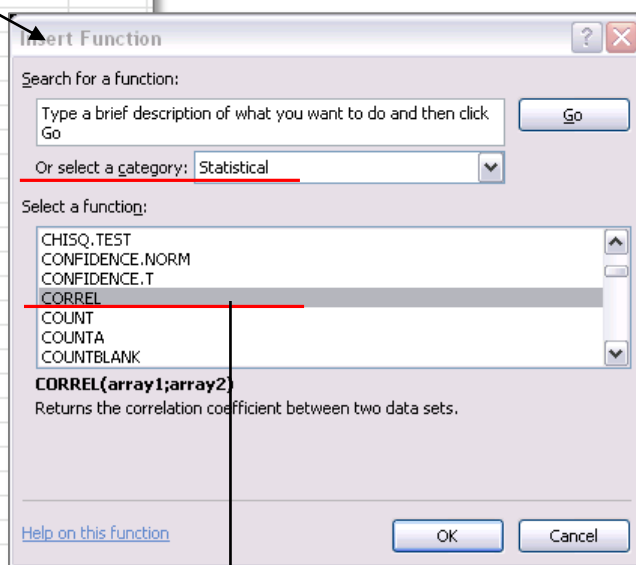
2. Lineaarne korrelatsioonikordaja on arvutatav funktsiooniga CORREL, millel on kaks argumenti – esimese tunnuse väärtuste blokk ja teise tunnuse väärtuste blokk.

- Kogenumad *MS Exceli* kasutajad võivad koheselt sisestada arvutamiskäsu kujul

=CORREL (B2 : B55 ; E2 : E55)

- Vähemkogenumatel (eelnevast valemist mitte arusaanutel) on soovitatav klikkida nupul  ja jätkata vastavalt joonisele.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	GENDER	HEIGHT	WEIGHT	HEAD	SHOE_SIZ	MATH												r(Height;Shoe_size)
2	W	170	70	55,5	39													
3	W	158	47,5	55	36													
4	W	170	60	53	35													
5	W	170	50	55	37													
6	W	179	68	58	41													
7	W	163	56		37													
8	W	177	65	55	40													
9	W	162,5	53	55	38													
10	W	170	75	56	39													
11	M	175	74	57	42													
12	W	176	66	57	39													
13	M	175	64	56	42													
14	M	190	82	58	46													
15	W	161	50	55	37													
16	W	170	85	57	41													
17	W	176	58	52	39													
18	W	172	90	58	41													
19	W	158	55	57	38													
20	M	189	82		43													
21	W	169	60	55,5	41													
22	W	164	52	56	37													
23	W	172	62	56	39													
24	W	173	66	56	40													
25	W	169	60	55	39													
26	W	162	50	50	38													
27	W	165	52	50,5	37													
28	M	170	80	56	41													
29	M	176	74	56	42													
30	M	175	73	54	43													
31	W	171	63	57	39													
32	W	170	60	53	39													
33	W	163	62	55	38													
34	M	181	74	55	44													
35	W	168	60	55	39													
36	W	174	54	55	40													
37	W	166	68	56	39													
38	W	168	63	53	39													
39	W	165	58	56	37													
40	W	171	75	55	41													
41	W	165	77	58	39													
42	W	161	55	57	38													
43	M	183	75		43													
44	W	169	53	55	38													
45	W	175	60	57	42													
46	W	167	80	57,5	41													
47	W	158	70	55	38													
48	M	174	87	57	40													
49	W	165	61	57	39													
50	W	164	58	57	39													
51	W	185	80	60	41													
52	W	177	63	60	40													
53	W	160	70	57	39													
54	W	162	70	55	40													
55	W	172	58	62	39													



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(Height;Shoe_size)	0,784758
Pikkuse ja jalanumbri vahel on seos.	
St, et mida suurem on pikkus, seda on keskmiselt ka jalanumber.	
Hüpoteeside paar	
H ₀ : (ehk matemaatiliselt r = 0)	
H ₁ : (ehk matemaatiliselt r ≠ 0)	

4. Pange korrelatsioonikordaja kohta kontrollitav hüpoteeside paar kirja ka teksti kujul.

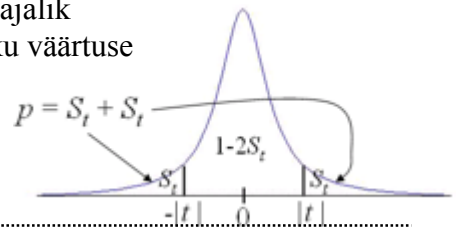
Meeldetuletuseks teooriast – hüpoteeside testimine korrelatsioonikordaja kohta

Testimaks korrelatsioonikordaja erinevust nullist (st testimaks seose statistilist olulisust) 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 2010-s on p-väärtus leitav funktsiooniga T.DIST.2T (ABS (t) ; n-2).

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

Näiteks kujul:

r(Height;Shoe_size)	0,784758
Pikkuse ja jalanumbri vahel on tugev positiivne seos.	
St, et mida suurem on pikkus, seda suurem on keskmiselt ka jala	
Hüpoteeside paar	
H ₀ : Pikkus ja jalanumber ei ole seotud (ehk matemaatiliselt r = 0)	
H ₁ : Pikkus ja jalanumber on seotud (ehk matemaatiliselt r ≠ 0)	
n(Height;Shoe_size)	
t(Height;Shoe_size)	
p(Height;Shoe_size)	

- 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).

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

	P	Q	R	S	T	U
1		r(Height;Shoe_size)	0,784758			
2						
3		Pikkuse ja jalanumbri vahel on tugev positiivne seos.				
4		St, et mida suurem on pikkus, seda suurem on keskmiselt k				
5						
6		Hüpoteeside paar				
7		H ₀ : Pikkus ja jalanumber ei ole seotud (ehk matemaatiliselt				
8		H ₁ : Pikkus ja jalanumber on seotud (ehk matemaatiliselt r ≠				
9						
10						
11		n(Height;Shoe_size)	54			
12						
13		t(Height;Shoe_size)	=ABS(R1*SQRT(R11-2)/SQRT(1-R1*R1))			
14						
15		p(Height;Shoe_size)				

c) Lahtri 'p(Pikkus;Jalanr)' järele tuleb sisestada funktsioon T.DIST.2T kahe argumendiga:

- teststatistiku absoluutväärtus $|t|$ ja
- (vaatluspaaride arv) - 2, so vastava t-jaotuse parameeter $(n - 2)$.

	P	Q	R	S	T	U	V	W	X	Y	Z
1		r(Height;Shoe_size)	0,784758								
2											
3		Pikkuse ja jalanumbri vahel on t									
4		St, et mida suurem on pikkus, se									
5											
6		Hüpoteeside paar									
7		H ₀ : Pikkus ja jalanumber ei ole s									
8		H ₁ : Pikkus ja jalanumber on seot									
9											
10											
11		n(Height;Shoe_size)	54								
12											
13		t(Height;Shoe_size)	9,130271								
14											
15		p(Height;Shoe_size)	=T.DIST.2T(R13;R11-2)								

Function Arguments

T.DIST.2T

X: R13 = 9,130270543

Deg_freedom: R11-2 = 52

Returns the two-tailed Student's t-distribution.

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

Formula result = 2,18481E-12

Help on this function

OK

NB! Vanemates Exceli versioonides puudub funktsioon T.DIST.2T ning kasutada tuleb funktsiooni TDIST. Viimane nõuab kolme argumenti: neist kaks esimest on samad, mis funktsioonil T.DIST.2T ($|t|$ ja $n-2$), kolmas argument on arv 2 (tähistab seda, et testime kahepoolset hüpoteesi $r \neq 0$, mitte seda, kas $r > 0$ või $r < 0$).

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

A'la: p(Height;Shoe_size) 2,18481E-12 < 0,05 => H₁: tudengite pikkus ja jalanumber on seotud

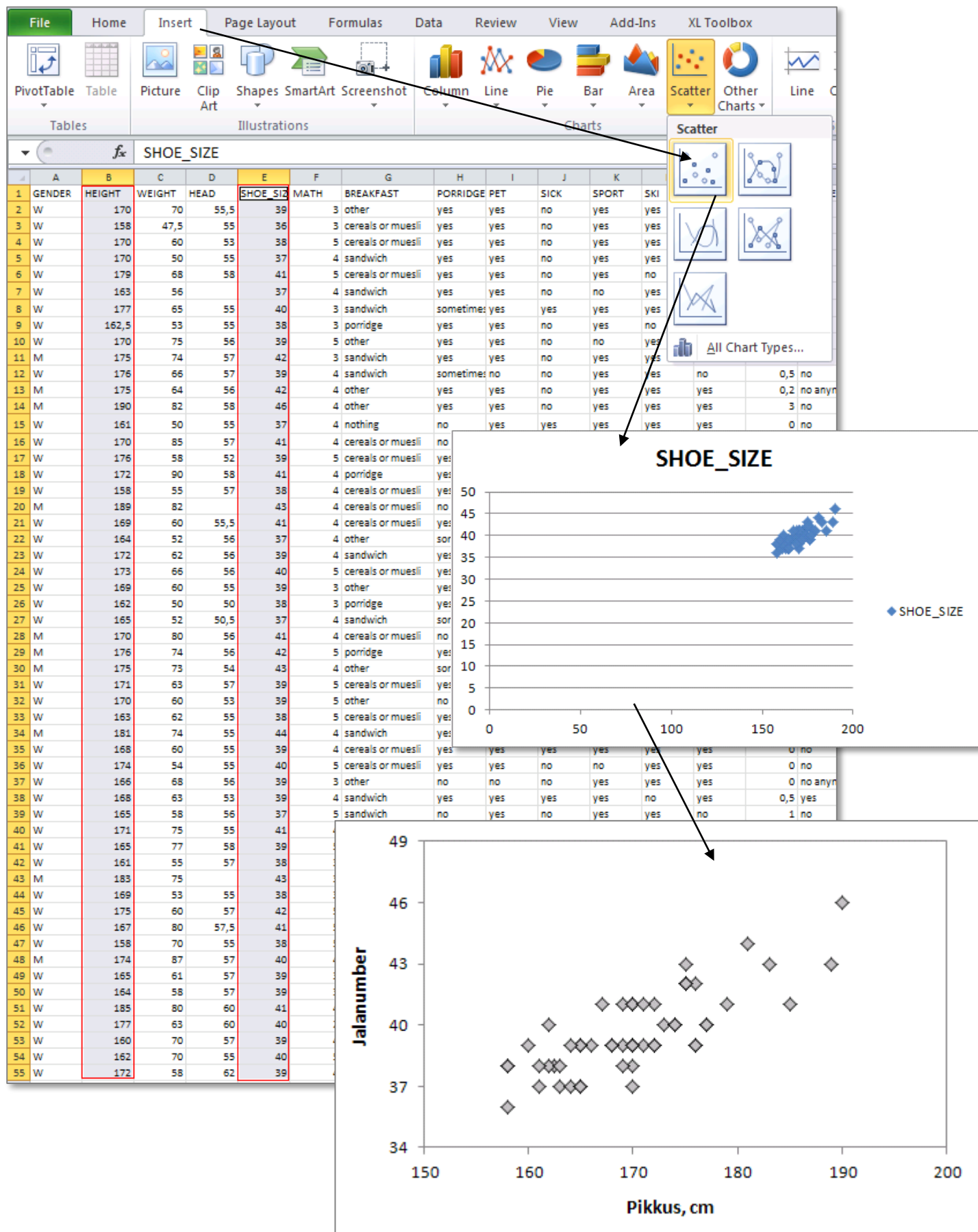
Märkus. 2,18481E-12 = 2,18481... × 10⁻¹²

7. Pange kirja päris lõppjärelus.

A'la: tudengite pikkuse ja jalanumbri vahel on tugev positiivne statistiliselt oluline seos ($r = 0,785$; $p < 0,001$).

Ülesande 2 tööjuhend

Illustreerige tunnuste 'HEIGHT' ja 'SHOE_SIZE' vahelist seost hajuvus- ehk punktdiagrammiga.



Ülesande 3 tööjuhend

1. Kasutades statistikaprotseduuri *Correlation* (Data-sakk -> Data analysis... -> Correlation) leidke korruga kõigi andmestikku kuuluvate pidevate arvturnuste (pikkus -- jalanumber) vahelised korrelatsioonikordajad.

The screenshot shows the Excel interface with the Data Analysis toolpak installed. The 'Data Analysis' dialog box is open, and the 'Correlation' option is selected. The 'Input Range' is set to '\$B\$1:\$E\$55', 'Labels in first row' is checked, and the 'Output Range' is set to '\$Q\$39'. The resulting correlation matrix is shown below:

	HEIGHT	WEIGHT	HEAD	SHOE_SIZE
HEIGHT	1			
WEIGHT	0,51228311	1		
HEAD	0,28996283	0,39609	1	
SHOE_SIZE	0,78475758	0,69898	0,29501	1

Tulemus:

2. a) Milliste tunnuste vahel on kõige tugevam lineaarne seos? Aga kõige nõrgem?
 b) Kas jalanumber on tugevamini seotud pikkusega või kehamassiga?
 c) Millise tunnusega on enim seotud peäumbermõõt?

Kirjeldage mõnda seost (nii seose suunda kui ka tugevust, pange vastavad laused kirja)!