Biometry practical 5

Illustrated (imperfect) practical guide

Preparatory work

- 1. Open in MS Excel the questionary data (file analysed already in previous practicals),
- 2. insert new worksheet, rename it as 'Praks5' (or 'Practical5') and
- 3. make a copy of the data table (from worksheet 'Andmed') and paste it into the upper left corner of the new worksheet.

Exercise 1.

Are the students' height and shoe size related? Study this using MS Excel functions.

- Calculate the correlation coefficient between variables 'HEIGHT' and 'SHOE_SIZE';
- describe the relationship on the basis of calculated coefficient;
- test the statistical significance of the relationship:
 - o formulate the null- and alternative hypothesis,
 - test, which of these hypothesis is true (find the sample size n and teststatistic t, and calculate on the basis of these values significance probability p),
 - \circ phrase the final conclusion.

Exercise 2.

Illustrate the relationship between variables 'HEIGHT' and 'SHOE_SIZE' with scatterplot.

Exercise 3.

- Calculate correlation coefficients between all continuous variables in dataset (height -shoe size) using statistical procedure *Correlation* (*Data*-tab -> *Data analysis*... -> *Correlation*).
- Between which variables is the strongest relationship? But the weakest?
- Describe some correlations (write down the sentences describing both the strength and the direction of relationships).

Exercise 1 guide

1. As the result of *MS Excel* functions is usually only one non-commented value, it is useful to write down before calculations what will be calculated.

For example, at the present moment the task is to calculate the correlation coefficient between height and shoe size – into *Excel* worksheet should be typed

- 'Linear correlation coefficient between height and shoe size'
- or more shortly '**r(Height;Shoe_size)**', as the linear correlation coefficient is usually denoted with letter '*r*'.

After that put the cursor into empty cell where you want to calculate the correlation coefficient.

2. Linear correlation coefficient is calculable with function CORREL, which has two arguments – the range of values of the first variable and the range of values of the second variable.

	(=	fr		-				• •		-	=CORREL(B2:B55;E2:E55)
	•	B		D	F		D	0			Less experienced students (who
1	GENDER	HEIGHT	WEIGHT	HEAD	SHOE SIZ	матн		r(Height-Sho	e (178)	<u> </u>	Less experienced students (who
2	w	170	70	55.5	39	3				—-÷	• did not understand the previous
3	w	158	47.5	55	36	3					formula) should click on button
4	w	170	60	53	18	5					Torindia) should enex on batton
5	w	170	50	55	37	4					$f_{\mathbf{x}}$ and continue according to
6	w	179	68	58	41	-	L				
7	w	163	56		37	4	\searrow				the scheme.
· ·	w	177			40	-					
0	w	167.6	60	55							
10	W	102,5		55	30				v -	_	
11	14	175	73	50	13				Insert Fu	nction	on 🤅 🔀 🕹
17	IVI M/	175		57	42				_		
12	vv	1/0	00	5/	29	4	land		Search for a	a functio	ction:
15	IVI M	1/5	04	50	42	4	keu		Type a b	orief des	escription of what you want to do and then click
14	IVI IVI	190	50	20	40	4			Go		
15	W	101	50		3/	4			On colorat		
10	W	1/0	60	5/	41	4			Or select	a <u>c</u> ateg	
1/	W	1/6	58	52	29	5			Select a fun	nction:	:
18	w	172	90	58	41	4			CUICO T	-	
19	w	158	55	57	38	4			CONFIDE	EDT ENCE NO	NORM
20	M	189	82		43	4			CONFIDE	ENCE.T	
21	W	169	60	55,5	41	4			CORREL		
22	W	164	52	56	37	4			COUNT		
23	w	172	62	56	39	4			COUNTA	1 1 A MIZ	
24	W	173	66	56	40	5			COUNTER		
25	W	169	60	55	39	3			CORREL((array)	y1;array2)
26	W	162	50	50	38	3			Returns th	the corre	rrelation coefficient between two data sets.
27	W	165	52	50,5	37	4			_		
28	M	170	80	56	41	4					
29	M	176	74	56	42	5					
30	M	175	73	54	43	4					
31	W	171	63	57	39	5			Help on this	s functio	tion OK Cancel
32	W	170	60	53	39	5	ked				
33	w	163	62	55	38	5	ked				
34	M	181	74	55	44	4					
35	W	168	60	55	39	4					
36	w	174	54	55	40	- -					
37	W	166	68	56	- 39	F	unction	Argument	s		₹ 2
38	W	168	63	53	39		_				
39	W	165	58	56	37		CORREL				
40	W	171	75	55	41			Array	B2-855		$[\mathbf{K}] = \{170; 158; 170; 170; 179; 163; 177; 162\}$
41	W	165	77	58	39			minay	- 02.033		
42	W	161	55	57	38			Array	2 E2:E55		[55] = {39;36;38;37;41;37;40;38;39;42;39
43	м	183	75		43						0.704757504
44	W	169	53	55	38						= U,784/5/584
45	W	175	60	57	42	R	eturns the	e correlation c	oefficient betw	ween tv	two data sets.
46	W	167	80	57,5	41				Array	y1 is a	s a cell range of values. The values should be numbers, names, arrays.
47	w	158	70	55	38					or	or references that contain numbers.
48	м	174	87	57	40						
49	w	165	61	57	39	-					
50	W	164	58	57	39	F	ormula res	alt = 0.7843	757584		
51	W	185	80	60	41						
52	W	177	63	60	40	H	lelp on this	function			OK Cancel
53	W	160	70	57	39						
54	W	167	70	55	40	5					
55	w	172	58	62	39	4	ked				
1.000						-					

• More experienced *Excel* users can type the appropriate command yourself:

Tanel Kaart, Alo Tänavots, Mirjam Vallas

3. Describe the relationship between students' height and shoe size:

- how strong (weak / intermediate / strong),
- Positive or negative (what this positive or negative means?).

NB! This conclusion follows from	f(Height;Shoe_size)	0,784758							
the positivity/negativity of the relationship! Only the word	There is a	relationship	between l	height and	d shoe size				
"bigger" or "smaller" is necessary to	This means, that to	bigger heig	ht correspo	nds sh	noe size on	an average.			
fill the gap in text.	Hypothesis pair								
	H₀: Height and shoe size are not related (or mathematically r = 0)								
·	H ₁ : Height and shoe	size are re	lated (or ma	athematic	ally <i>r</i> ≠ 0)				

4. Write down the hypothesis pair also in text form.

Reminder from theory – hypothesis testing about correlation coefficient

To test in *Excel*, is the correlation coefficient different from zero (is the relationship statistically significant), at first the absolute value of teststatistic (which in case of null hypothesis follows the t-distribution) must be calculated by the formula

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

Quantity r in this formula is the correlation coefficient and n is the sample size (number of students whose height and shoe size were both known).

The decision, which of the hypothesis is true, will be made according to the p-value p, which is calculated as the sum of the areas under the

tails of teststatistic's distribution

(denoted as S_t in figure).

In *Excel* the p-value is calculable with function T.DIST.2T(ABS(t);n-2).



5. The evaluation of significance probability (p-value) is easier to perform, if all necessary

intermediary quantities are pre- calculated and named in <i>Excel</i>	r(Height;Shoe_size) 0,784758											
worksheet.	There is a strong positive relationship between height and shoe size.											
For example:	This means, that to bigger height corresponds bigger shoe size	on an average.										
	Hypothesis pair											
	H_0 : Height and shoe size are not related (or mathematically $r = 0$)											
\ \	H_1 : Height and shoe size are related (or mathematically $r \neq 0$)											
a) Type 'n(Height;Shoe_size)' and count into following cell the number of students whose	n(Height;Shoe_size)											
height and shoe size were both	t(Height;Shoe_size)											
known (only these students are												
used by to calculate the	p(Height;Shoe_size)											
correlation coefficient value).												

b) Type behind the cell 't(Height;Shoe_size)' formula to calculate absolute value of teststatistic:

teststatistic:	СС	DUNT	*	:	\times	\checkmark	<i>fx</i>	=ABS	(R1*SQRT	R11-2)/SQ	RT(1-R1*R1)))		
		Р			Q		1	R	s	т	U			
	1		r(He	eight	;Shoe_	size)	0.784	757584						
	2											1)) ;ize) ar		
	3		There is a strong positive relationship between height and shoe size This means, that to bigger height corresponds bigger shoe size on ar											
	4													
	5													
	6		Нур	othe	sis pai	r								
	7		H ₀ : Height and shoe size are not related (or mathematically <i>r</i> = 0)											
	8		H ₁ : I	Heigł	nt and	shoe	size ar	e relate	d (or math	ematically	r≠0)			
۱ ۱	9													
	10											_		
· · · · · · · · · · · · · · · · · · ·	11		n(H	eight	;Shoe	_size		54						
	12													
	13		t(He	eight	;Shoe	_size)	=ABS(R1*SQR	T(R11-2)/S	QRT(1-R1*	R1))			
	14													
	15		р(н	eight	;Shoe	_size)								

- c) Input behind the cell 'p(Height;Shoe_size)' function T.DIST.2T with two arguments:
 - Absolute value of teststatistic | *t* | and
 - (number of observations) -2, the parameter of the corresponding t-distribution: (n-2).

- (XV	′ <u>f</u> ∗ =T.DIST.2T(R13	;R11-2)								
	Р	Q	R	S	Т	U	V	W	Х	Y	Z
1		r(Height;Shoe_size)	0,784758	(F							
2				Functio	n Argumer	nts					
3		Pikkuse ja jalanumb	ri vahel on	t _T.DIST.2	2T						
4		St, et mida suurem o	on pikkus, s	e		X R13			=	9,130270543	3
5					Deg fre	edom R11	-2		= 1	52	
6		Hüpoteeside paar								0 40404E 40	
7		H ₀ : Pikkus ja jalanun	nber ei ole	S Returns t	he two-tailer	Student's t-di	stribution		=	2,18481E-12	
8		H ₁ : Pikkus ja jalanun	nber on sec	t		5000010500	e :- the :		- L L != L L = -		
9					/		X is the i	numeric value	at which to e	valuate the o	istribution.
10											
11		n(Height;Shoe_size)	54								
12				Formula r	esult = 2,18	481E-12					
13		t(Height;Shoe_size)	9,130271	Help on th	pis function						ОК
14					1	1		1	1		
15		p(Height;Shoe_size)	=T.DIST.2T	(R13;R11-2	2)						

NB! In older *Excel* versions there is no function T.DIST.2T and function TDIST must be used. It has three arguments: the first two are the same as in function T.DIST.2T (|t| and *n*-2), the third argument is number 2 (it determines, that two-side hypothesis $r \neq 0$ is tested).

6. Make a formal decision, which of the hypothesis is right and why.

A'la:	p(Height;Shoe_size)	2,18E-12	< 0,05 =>	H ₁ : students	s' height a	ind shoe si	ize are rela	ated
			∕ _{Rem}	ark. 2,18481	E-12 = 2	2,18481	×10 ⁻¹²	

7. Write down the final conclusion.

A'la: between students height and shoe size there is **strong positive statistically significant** relationship (r = 0.785, p < 0.001).

Exercise 2 guide

Illustrate the relationship between variables 'HEIGHT' and 'SHOE_SIZE' with scatterplot.

File	Home	Inse	rt Pa	age Layout	Fo	ormulas D	ata	R	eview	View	v A	dd-Ins	XL Toolbox		1	
1.7								1 .	<u>.</u>		-		⊡ 🖒			
PivotTable	a Table	Picture	Clip	Shapes Sma	≕ rtArt	Screenshot	Colu	mn	Line	Pie	Bar	Area	Scatter Othe	r Line C		
Tab	les		Art	• Illustrations		*	*		*	* 	* barts	Ŧ	Charter	.*		
• (0	f _x	SHOE	SIZE								\sim	<u> </u>		2		
A	В	с	D	E	F	G		н	I.	1	K		· ·			
1 GENDER	HEIGHT	WEIGHT 70	HEAD	SHOE_SIZ MA	TH	BREAKFAST	PO	RRIDGE	PET	SICK	SPOR	T SKI				
3 W	158	47,5	55,5	36	3	cereals or muesli	yes	, ;	yes	no	yes	yes		R 8		
4 W	170	60	53	38	5	cereals or muesli	yes		yes	no	yes	yes		16 X		
5 W	170	50	55	37	4	sandwich	yes		yes	no	yes	yes				
5 W	179	55	58	37	5	cereals or muesli sandwich	yes		yes	no	yes	no		1		
8 W	105	65	55	40	3	sandwich	son	, netime:	yes yes	ves	ves	ves	1XX			
9 W	162,5	53	55	38	3	porridge	yes		yes	no	yes	no				
10 W	170	75	56	39	5	other	yes		yes	no	no	yes	All Char	t Types		
11 M	175	74	57	42	3	sandwich	yes		yes	no	yes	yes.	-	0.5.00		
13 M	175	64	56	42	4	other	ves	iveorne:	yes	no	yes	ves	yes	0,5 no 0,2 no anvn		
14 M	190	82	58	46	4	other	yes		yes	no	yes	yes	yes	3 no		
15 W	161	50	55	37	4	nothing	no		yes	yes	yes	yes	yes	0 no		
16 W	170	85	57	41	4	cereals or muesli	no					*				
17 W	176	58	52	39	5	cereals or muesli	yes					S	HOE_SIZ	ZE		
19 W	172	55	57	38	4	cereals or muesli	yes	50								
20 M	189	82		43	4	cereals or muesli	no	45							•	
21 W	169	60	55,5	41	4	cereals or muesli	yes	45								
22 W	164	52	56	37	4	other	sor	40	1							
23 W	172	66	56	40	4	cereals or muesli	ves	35	+							
25 W	169	60	55	39	3	other	yes	30	+							
26 W	162	50	50	38	3	porridge	yes	25	+							
27 W	165	52	50,5	37	4	sandwich	sor	20								10L_312L
28 M	170	74	56	41	4	cereals or muesil porridge	no ver	15								
30 M	175	73	54	43	4	other	sor	10								
31 W	171	63	57	39	5	cereals or muesli	yes	5								
32 W	170	60	53	39	5	other	no	0								
35 W	165	74	55	44		cereals or muesil sandwich	ves		0		50	\ 1	00	150	200	
35 W	168	60	55	39	4	cereals or muesli	yes	_	yes	yes	yes	y===	yes	0 110		
36 W	174	54	55	40	5	cereals or muesli	yes		yes	no	no	yes	yes	0 no		
37 W	166	68	56	39	3	other	no		no	no	yes) es	yes	0 no anyn		
39 W	165	58	56	37	4	sandwich	no		ves	no	ves	ves	no	0,5 yes		
40 W	171	75	55	41	4	sandwich	yes		yes	no	yes	yes	yes	0 no		
41 W	165	77	58	39		49								-		1
42 W	161	55	57	38									7			I
43 M	183	75	55	43												I
45 W	175	60	57	42		46 -								\$		I
46 W	167	80	57,5	41												I
47 W	158	70	55	38		a							\diamond			
48 M	174	61	57	40		- ⁴³						\$	\$	\diamond		I
50 W	164	58	57	39		e l						~~				I
51 W	185	80	60	41		ž				<u>_</u>	0000	· •				I
52 W	177	63	60	40		40 -			~			~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~	,			
53 W	160	70	57	39					ັ້	***	~~~~					I
55 W	172	58	62	39		37 -			Š	~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~	Ň					
-								<	٥Č	***	*					I
																I
						34 +			1		1		1			I
						150)		160		170		180	190	200	I
											Ноі	aht co	,			I
											neq	Bucken				1

Exercise 3 guide

1. Calculate correlation coefficients between all continuous variables in dataset (height – shoe size) using statistical procedure *Correlation* (*Data*-tab -> *Data analysis*... -> *Correlation*).

	File	Home	Inser	rt Pa	ge Layou	ıt F	Formula	IS	Data 🕂	Review	View	Add-Ins	XL To	olbox		۵ (?)	- 6	
	B	R	Co	nnection	s Az↓	AZA	Y	7 K K	Clear	₩ ₩	×××× ×××× ××××	式 Data	Validati olidate	on ∗ 🗣 Gr lan Un		Data	Analysis	
Get	External	Refresh	Edi	tlinks	Z A	Sort	Filte	r V	Advanced	Text to Remove Oced Columns Duplicates W What-If Analysis *						- Joive		
	Dala *	AIIŦ	Connecti	ons			Sort &	Filter	Auvanceu	Data Tools Outline G							vsis	
-	(=	fr									Data Analysis							
-	A	B	с	D	F	F		G	Р	0	- Analys	ia Taola			/			
1	GENDER	HEIGHT	WEIGHT	HEAD	SHOE_SIZ	MATH	BREAK	FAST		r(Height;Shoe	Analys				/		ОК	
2	W	170	70	55,5	39		3 other 3 cereals	ormues	6	Dikkuse ia iala	Anova Anova	a: Single H a: Two-Ea	actor ctor Wil	h Replication		<u>^</u>	Cancel	
4	w	170	60	53	38		5 cereals	or mues	6	St, et mida su	Anova	a: Two-Fa	ctor Wil	hout Replicat	on _			
5	w	170	50	55	37		4 sandw	ich			Corre	lation					Help	
7	w	1/5	56	30	37		4 sandw	ich		Ho: Pikkus ja ja	Descr	iptive Stat	istics				<u> </u>	
8	w	177	65	55	40		3 sandw	ich		H1: Pikkus ja ja	Expor	ential Sm	oothing					
9	w	162,5	53	55	38		3 porridg 5 other	e			F-Tes	t Two-San w Opelucia	nple for	Variances				
11	M	175	74	57	42		3 sandw	ich		n(Height;Shoe	Histor	a Analysis gram	,			~		
12	W	176	66	57	39		4 sandw	ich		eluciate ch		0.1705						
13	M	175	64 82	56	42		4 other 4 other			i(Height;Shoe	_size) 9,13	027054						
15	w	161	50	55	37		4 nothir	-										
16	w	170	85	57	41		4 cerea	COL	relatio	n		v						
17	w	176	58	52	39		5 cerea 4 porrid	Inp	ut								H	
19	w	158	55	57	38		4 cereal	Inc	ut Doog	101		* +D+1		-	a 🕒	ок		
20	M	189	82		43		4 cereal	Ťi it	Jul Kang	je:		- \$B\$I	:\$E\$5:					
21	w	169	52	55,5	37		4 other	Gro	ouned By	ed By: 💽 Columns								
23	w	172	62	56	39		4 sandv			O Rows Halp								
24	W	173	66	56	40		5 cereal 3 other				_	<u> </u>	JWIS			telp		
26	w	162	50	50	38		3 porrid	\checkmark	<u>L</u> abels i	in first rov	\checkmark							
27	w	165	52	50,5	37		4 sandv											
28	M	170	74	56	41		5 porrid	Ou	tput opti									
30	м	175	73	54	43		4 other	•	Output	utput Range: 🗸 \$Q\$39 💽								
31	w	171	63 60	57	39		5 cereal 5 other											
33	w	163	62	55	38		5 cereal		New w	w worksneet Ply:								
34	M	181	74	55	44		4 sandv	0	New <u>W</u>	orkbook	1							
36	w	108	54	55	40		5 cereal	L			-j-							
37	W	166	68	56	39		3 other	_			/		-					
38	w	168	63 58	53	39		4 sandw 5 sandw	ich ich			_							
40	W	171	75	55	41		4 sandw	ich			i							
41	W	165	77	58	39		5 sandw	ich										
42	M	161	55	57	38 43		3 porridg	e le										
44	W	169	53	55	38		3 sandw	ich										
45	W	175	60	57	42		5 cereals	s or mues	li -									
47	w	157	70	55			5 cereals	or mues	ii .									
48	м	174	87	57	40		4 sandw	ich										
49	w	165 164	61 58	57	39		3 other 3 sandw	ich (
51	W	185	80	60	41		4 cereals	or mues				HEIGH	IT	WEIGHT	HEAD	SHOE S	SIZE	
52	W	177	63	60	40		2 sandw	ich	HEIGH	т			1				_	
53	W	160	70 70	57	39 40		4 sandw 5 sandw	ich	MILIGH				1					
55	w	172	58	62	39		4 other		WEIGH	IGHI U,51228311 1								
1									HEAD 0,28996283 0,39609 1									
	Result:								SHOE	SIZE		0,78475758 0.698			0,29501 1			

- 2. a) Between which variables is the strongest relationship? But the weakest?
 - b) Is the shoe size more related with height or weight?
 - c) With which body measurement has the strongest relationship head circumference?

Describe some correlations (write down the sentences describing both the strength and the direction of relationships)!