Biometry practical 2

Illustrated (imperfect) practical guide

Preparatory work

- 1. Open in MS Excel the questionary data (file analysed already in previous practical),
- 2. rename 'Sheet3' to 'Praks2' (or 'Practical 2') and
- 3. make a copy of the data table (from worksheet 'Andmed') and paste it into the upper left corner of the worksheet 'Praks2'.

Exercise 1.

Construct the frequency table to variable 'HEAD' and illustrate it with histogram. Format the histogram.

Guide

To construct the frequency table to continuous variable there are at least tree possibilities in Excel – statistical procedure *Histogram*, function *FREQUENCY* and *PivotTable*. The first two assume that there is also formed the table containing upper limits of continuous variable values' classes (intervals)

(actually can procedure *Histogram* form these intervals by itself, but mostly are the limits of these automatically generated intervals not rounded and therefore the description of values' distribution is more inconvenient).

Following it is introduced, how to construct the frequency table with procedure Histogram.

- 1. To decide, how many and how big intervals to form, there must be knowledge about the count and range of values of studied variable.
 - The count n = 54 (Did you get the same number? How?) gives the hint, that the values of students head circuits should be divided into 7 or 8 intervals (because $\sqrt{54}\approx7,3$; at the same time this is not absolute truth, if it is more suitable, there can be formed also 6 or 9 intervals).
 - The range of values is given by minimum and maximum value.
 To find these characteristics there are several ways in Excel. The most quick variant is (look at the scheme in next page also):
 - a. select the cells containing values of study variable,
 - b. click with mouse right key on Status Bar under the worksheet and select wanted functions (for example *Min*),
 - c. the values of selected functions are displayed on the Status Bar.

4	А	в	с	D	E	6			0	P	Q
1	GENDER	HEIGHT	WEIGHT	HEAD	JALANR	Cus	tomize Status Bar		SMOKE		
4	w	170	47.5	35.5			Cell Mode	Ready	10		
4	w	170	60	53		~	Cell Mode	Ready	10		
5	w	170	50	55	1	1	Elash Fill Blank Cells		10		
6	w	179	68	58		*			10		
7	W	163	56	40		~	Flash Fill Changed Cells		10		
8	w	177	60		- 1		<u>_</u>		10 10		
10	w	170	75	36		1	Signatures	Off			
11	M	175	74	57					10		
12	w	176	66	57	1	\checkmark	Information Management Policy	Off	10		
13	м	175	64	56					to anymore	e, but I\'ve s	smoked
14	M	190	82	58	- 1	\checkmark	Permissions	Off	10		
15	w	161	50						10		
17	w	176		57			Caps Loc <u>k</u>	Off			
18	w	172	90	58			• –		10		
19	w	158	55	57	1		<u>N</u> um Lock	Off	res		
20	м	189	82	17			o		res		
21	W	169	60	55.5		\checkmark	Scroll Lock	Off	res		
22	W	164	52	56		-	Fixed Device al	0.0	2		
23	w	172	62	26		\checkmark	Eixed Decimal	Off	10		
25	w	169	60	33			Overture Mede		10		
26	w	162	50	50			<u>O</u> vertype Mode		10		
27	w	165	52	50.5	1	. /	End Mode		10		
28	м	170	80	36		~	<u>e</u> na mode		10		
29	M	176	74	56			Macro Recording	Not Recording	res		
30	M	175	73	34			<u>Macro Recording</u>	Not necoraing	res		
32	w	170	60	53			Selection Mode		to anymor	e. but i\/ve r	smoked
33	w	163	62	55		*	Selection Mode		to anymor	e, but I\'ve :	smoked
34	м	181	74	55		. /	Page Number		10		
35	w	168	60	35		*	<u>r</u> age Number		10		
36	W	174	54	55	- 1		Average	55 2027027	10		
37	w	166	68	36		*	Average	33.2031031	to anymon	e, but ivve s	moked
39	w	165	58	36		1	Count	55	10		
40	w	171	75	35		*	<u>_</u>		10		
41	W	165	77	58		\checkmark	Numerical Count	54	10		
42	W	161	55	57			-		to anymor	e, but I\'ve s	smoked
43	M	183	75	75		\checkmark	Minimum	17	10		
45	w	175	60	57		1			6		
46	W	167	80	57.5		/~	Ma <u>x</u> insum	75	10		
47	w	158	70	55	/		C.um	2004	10		
48	м	174	87	57	<u> </u>	\checkmark	<u>p</u> um	2981	10		
49	W	165	61	57			Under al Obstant		res		
50	w	164	38	57		\checkmark	Upload Status		to anymor	a, but I/ve s	moked
52	w	103	63	60	/ 1						
53	W	160	70	57		\checkmark	View Shortcuts		to anymor	e, but I\'ve :	smoked
54	w	162	70	*			Zeene Slider		10		
55	w	172	58	67	1	\checkmark	<u>Z</u> oom Silder		to anymor	e, but i\'ve s	smoked
		×.		Andn	ned	\checkmark	Zoom	56%			
	0	\$	7		RAGE:	55.20	37037 COUNT: 55 NUMERICAL COUNT:	54 MIN: 17	MAX: 7	'S SUI	M: 298 ⁻

• Maximal head girth is 75 and minimal 17 cm. Are these values real? If it is hard to imagine the girth, then maybe it is easier to imagine the diameter. But how to calculate the diameter? How big is the head diameter, if head girth is 17 cm? And is this number really possible?

<u>Hint</u>. The girth is calculated as $P = 2\pi r$, from which the diameter is $d = 2r = P/\pi$ (to calculate the diameter corresponding to girth 17 cm in Excel you can use the formula '=17/PI()'). The result is 5.4 cm. To compare – the diameter of CD is 12 cm and the diameter of usual coffee cup is about 7 cm. So, the head girth 17 cm and corresponding head diameter 5.4 cm are clearly too small. Additionally there is one head girth value 40 cm in data table, which is also too small (corresponding head diameter is 12.7 cm). And the maximal value 75 cm is too big for head girth.

Obviously are values 17, 40 and 75 cm measuring or typing errors.

As **traditional statistical analyses methods are not derived to analyse unreal or very exceptional data**, it is usual to omit these very different values. At the present situation it is the most reasonable to just delete the head girth vales 17, 40 and 75 cm.

And to avoid future confusion these values should be deleted in both worksheets 'Praks2' ('Practical2') and 'Andmed' ('Data').

- Minimal head girth after deleting unreal values is 50 and maximal 62 cm.
- To find the approximate width of intervals the difference between maximal and minimal value must be divided by the desired number of intervals: (62-50)/7= 1.7... cm.

To get round interval limits it is necessary to slightly increase or decrease the calculated value, also the number of intervals can be changed. At the present situation we can fix the width of intervals to **2 cm**.

- To fix the intervals' limits it is natural to start with the first interval and fix it in way that it contains the minimal value. At the present situation the first interval can be for example 48-50 cm.
- Altogether we can form seven 2-centimeter intervals of students head girth: 48-50; 50-52; 52-54; 54-56; 56-58; 58-60; 60-62.

As Excel puts the values equal to interval limit into the lower interval, it is more correct to present the intervals as half-closed intervals (because how to known otherwise into witch interval belongs the value 41 cm, for example):

(48,50], (50,52], (52,54], (54,56], (56,58], (58,60] and (60,62].

NB. But instead these intervals somebody can form also nine 1.5-centimeter intervals of the form

(49;50.5], (50.5;52], (52;53.5], (53.5;55], (55;56.5], (56.5;58], (58;59.5], (59.5;61], (61;62.5];

or seven 2-centimeter intervals of the form

(49,51], (51,53], (53,55], (55,57], (57,59], (59,61], (61,63].

- To give these intervals to Excel there must be constructed a special table containing the upper limits of intervals:
- Remarks:

- between initial data table and any additional tables should be at least one empty column (or row), why?;

- Excel understood the given values as follows:
 - '51' means '≤51';
 - '53' means '51<x≤53'; ...

P 0 Q **MOKE** Head upper limits ho 50 ho 52 ho 54 ho 56 58 ho ho 60 ho

- there is **no need to specify the last interval '62'=(60,62]**, because constructing the frequency table Excel automatically makes one class for all values not belonging into any given intervals. As all head girth values less or equal to 60 belong into specified intervals, the additionally made class will contain only values bigger than 60.

2. Frequency table with procedure *Histogram*: *Data*-tab \rightarrow *Data Analysis*... \rightarrow *Histogram*

			OME	IN	CEDT	DAGE		- EOI		s D		EVIEW	VIEW	N		NIC
					JENT	PAGE				Elach Ell			VIEV	•	Data 4	CVI
			G		A↓ Z A	A Z			∃ ¹⁶⁵ /	Flash Fill	Dereliante			2	Data A	Inalysis
Get	Extern	al R	efresh	-	Z So	rt Fil	ter 🦉	Text t	io	Kemove	Duplicates	· · · · · · · · · · · · · · · · · · ·	Outline	e 爷	Solver	
	Data ≖		All -	3	AΨ		1	Colum	ins 🗟	Data Vali	dation 👻		*			
		Co	onnecti	ons	S	ort & Filte	r			Data To	ols				Analy	/sis
T	1		*	: 7	X v	f _x							/			
4	c	D	E	F	G	н		к	L	м	N O	/	Q	R	s	т
1 W 2	EIGHT HE	AD 55.5	SHOE_SIZE 39	МАТН З	BREAKFAS F other y	PORRIDGE PET	SICK	SPORT yes	SKI yes	CAR BE yes	ER SMOKE 0 no		Head_uppe 50	r_limits		
3	47.5	55	36	3	cereals or y	ves yes	no	yes ves	yes ves	yes	0 n0		52			1
5	50	55	37	4	sandwich y	yes yes	no	yes	yes	no	0 100		36			1
6 7	68 56	58 40	41	4	sandwich y	γes γes γes γes	no	no	no yes	no no	0 no 0 no	_	58			1
8	65	35	40	3	sandwich s	cometimes yes	yes	yes	yes	no /	0 no					1
10	75	56	39	Data	Analy	cic				/						i –
11	74 66	57	42 39	Date	r Anaty	313			-/							
13	64	56	42	Ana	alysis Too	ols			/				smoked			
14	82 50	58 55	46	An	iova: Two	o-Factor V	vithout R	eplication.		^			1		1	
16	85	57	41		variance	1					Can	cel			i	
18	90	58	41	De	scriptive	Statistics				=	Hel				- 1	
19 20	55 82	57	38 43	Ex	ponentia Tost Two	l Smoothin -Sample fr	lg Varian									
21	60	33.5	41	Fo	urier Ana	alysis	n variari	Les							1	
22 23	52	56 56	37	His	togram]					i	
24	66	56	40	Mo Ra	odom Nu	erage Imber Geo	eration			~						
26	50	30	38				cracion				_		i i		1	
27 28	52 80	30.5 36	37	4	cereals or in	ometimesyes to yes	yes	yes	yes	ves	0 no 0 no		1		!	
29	74	56	42		porridge y	yes yes	yes	yes	yes	yes	0.1 yes	_	-			
30	63	34 57	43	3	cereals or y	iometimes yes ies yes	yes es	yes	no yes	no yes	0.1 yes 0 no					
32	60	53	39		other r	no yes	yet	yes	no	yes	0 no anym	iore, but i\'ve	e smoked			
34	74	35	44	4	sandwich y	res yes	no	yes	yes	yes	1 no					
35 36	60 54	35 35	39 40	4	cereals or y cereals or y	res yes res yes	yes	yes	yes yes	yes ves	0 no 0 no	_		_ <u>i</u> _		
37	68	56	39	3	other r	no no	no	yes	yes	yes	0 no anym	iore, but i\'ve	snoked			
38	58	56	39	- 4	sandwich y sandwich n	γes γes no γes	no	yes	no yes	no	0.5 yes 1 no		1			
40	75	33	41	4	sandwich y	γes γes	no	yes	yes	yes	0 no		1	1		
41	55	57	38		porridge y	res γes γes	yes	yes	ves	yes yes	0 no anym	iore, but I\'ve	smoked	+		
43	75	75	43	<u></u> ,	porridge y	es no	no	yes	yles	yes	3 no		1	$\frac{1}{1}$		
45	60	57	42		cereals or y	ves yes	no	no	no	no	0.5 no					
46 47	80 70	57.5 35	41		other y cereals or y	/es γes ∕es γes	no yes	yes yes	yes no	yes no	2 no 0 no	_	i i			
48	87	57	40	4	sandwich s	ometimes yes	yes	yes	no	ves	0.5 no					
50	58	57	39	3	sandwich y	iometimes yes ies yes	yes 🗸	yes	no	yes	0.5 yes 0 no anym	iore, but I\'ve	smoked			
51	80 63	60 60	41	4	cereals or s sandwich of	ometimes yes	(-	4			1 1			
53	70	57	39	4	sandwich s	ometimes yes	Hist	ogram	22				i i			1
54 55	70 58	55	40	4	other s	io yes cometimes yes	Inp	ut							_	
						Data	Inp	out Range	e:		\$D\$1:	\$D\$55	1 🖻	<u>.</u>		ок
		Γ	т /	i	••••••	1		-						3	Ca	ancel
			mer	vais	upper	i iimits	Bin	Range:			\$Q\$1	\$Q\$7	1			
								Labels					1		Ŀ	telp
to I	Excel	tha	t ther	e are	e the v	ariable		Eapers					1			
nes	in the	e fir	st rov	v of	selecte	ed data	Ou	tput optic	ns —							
								Output	lancer		¢T¢1		5	3		
Loc	cality	of t	he up	per l	left co	rner of			(ange:		4141			4		
	-		1	-	outpi	ıt table	0	New Wo	rksheel	t <u>Ply</u> :						
••••••					'P '		0	New <u>W</u> o	rkbook							
								Devet- (الم مراد							
			Ac	lditir	nallv to	ן פווצון ט		P <u>a</u> reto (:	sorted	niscogram	9					
	fr	eau	encie	c alc	any to	ulativa		Cu <u>m</u> ulati	ve Per	centage						
	11	fra		5 a15		ulate d		⊆hart Oi	utput							
		trec	luenc	ies a	re cal	culated										

Expected result of procedure *Histogram*:_____

(**cumulative frequency** shows the percentage, of values less or equal to the upper limit of interval)

	Т	U	V
He	ad_upper_lim	Frequency	Cumulative %
	50	1	1.96%
	52	2	5.88%
	54	4	13.73%
	56	25	62.75%
	58	16	94.12%
	60	2	98.04%
	More	1	100.00%

3. Add in front of the table real interval limits in the form of half-closed intervals.

	Неа	d_upper_lin	Frequency	Cumulative %
(48,50]		50	1	1.96%
(50,52]		52	2	5.88%
(52,54]		54	4	13.73%
(54,56]		56	25	62.75%
(56,58]		58	16	94.12%
(58,60]		60	2	98.04%
(60,62]	I	More	1	100.00%

NB! To select separate cells hold down the 'Ctrl'-key ...



4. Write down at least one sentence using absolute frequencies and one sentence using cumulative frequencies.

Exercise 2.

Use the *PivotTable* to describe the students' weight depending on the porridge eating.

Guide

- 1. Put the cursor into data table (in worksheet 'Praks2').
- 2. *Insert*-tab \rightarrow *PivotTable*
- 3. Place the output table under the data table:



Home

Table

ĴĴ

PivotTable

Insert

Cli

Picture

5. Instead of sum of weights (calculated by Excel by default), calculate the **count of students** in different groups.

X∄	5 • (,> - ∓									PIVO	OTTABL	e tools
FIL	E HOME	INSERT	PAGE LA	YOU	T FORMUL	AS	DATA	REVIEW	VIEW	ADD-INS	ANAL	YZE	DESIGN
Pivot	Active Fi Sum of Fable	eld: WEIGHT Settings Active Field	n Up *	+	→ Group Sele	ection d	Inser Inser	t Slicer t Timeline · Connections Filter	Refresh	Change Data Source v Data	Actions •	民 F 廠 C 唱 R	ields, Items, 8)LAP Tools - elationships Calculations
B58	-	: 🗙 🗸	fx			V	alue Fie	ld Settings		? ×			
57 58 R 59 n	A Row Labels	▼ Sum of W	EIGHT 636		Source Name: ' <u>C</u> ustom Name: Summarize Va	WEIGHT Count lues By	of WEIGH Show V	T alues As				J	K
60 s 61 y	ometimes es Srand Total	:	717		Choose the ty data from the	pe of cal selected	lculation t I field	hat you want t	o use to s	ummarize			
63 64 65 66 67 68					Sum Count Average Max Min Product					Cancel			
69					Number Forma	at	_	OK		Cancel			

6. Calculate additionally average weights.



PivotTable Name:

Active Field:

SHOE_SIZE

Group Selection

F	Result:				MATH BREAKFAST PORRIDGE PET Drag fields between ar	eas below:
Row Labels 🔻	Count of WEIGHT	Average of WEIGHT	StdDev of WEIGHT	Min of WEIGHT:	Max of WEIGHT	Σ Values 🔻
no	9	68.4444444	11.95942213	50	85	
sometimes	11	65.18181818	11.60015674	52	87	
yes	34	64.72058824	10.2033797	47.5	90	
Grand Total	54	65.43518519	10.66206744	47.5	90	
					Row Labels	Σ Values
					PORRIDGE 🔻	Count of WEI 🔻
						Average of 🔻
						StdDev of W 🔻
						Min of WEIGHT2
						Max of WEIG 🔻

8. Round the averages and standard deviations to one decimal place (using corresponding command at *Home*-tab, for example).

F	ile Home	Insert P	Page Layout	Formulas	Data Revie	w View	Add-Ins XL
	<u> </u>	Nori -	11 * A	A =	≡ ≫ *	General	*
Pa	ste 🧹 🖪	<u>I</u> <u>I</u> -	* 🔕 * .	<u>A</u> · ≡ ₹		a · 🦉 · %	•••••••••••••••••••••••••••••••••••••
Clip	board 🕞	Font		Es.	Alignment	G Nur	mber 🕞
C59	•	<i>f</i> _x 68.44444	4444444				
	А	В		С	D	E	F
57			\backslash				
58	Row Labels 🔻	Count of WEIG	HT Average	of WEIGHT	StdDev of WEIGHT	Min of WEIGHT	Max of WEIGHT
59	no		9 6	8.4444444	11.95942213	50	85
60	sometimes		11 6	5.18181818	11.60015674	52	87
61	yes		34 64	4.72058824	10.2033797	47.5	90
62	Grand Total		54 6	5.43518519	10.66206744	47.5	90

Result:

Row Labels 💌	Count of WEIGHT	Average of WEIGHT	StdDev of WEIGHT	Min of WEIGHT:	Max of WEIGHT
no	9	68.4	12.0	50	85
sometimes	11	65.2	11.6	52	87
yes	34	64.7	10.2	47.5	90
Grand Total	54	65.4	10.7	47.5	90

9. Present the table in form where different characteristics are in different rows and for each porridge-eating group corresponds own column:

	Column Li 🔻			
Values	no	sometimes	yes	Grand Total
Count of WEIGHT	9	11	34	54
Average of WEIGHT	68.4	65.2	64.7	65.4
StdDev of WEIGHT	12.0	11.6	10.2	10.7
Min of WEIGHT2	50	52	47.5	47.5
Max of WEIGHT3	85	87	90	90



10. And now divide the table into rows according to the variable 'CAR' values and omit one student who did not know does she or he have a car or not.

Expected	result:
----------	---------

	Column 🔻			
Row Labels 🛛 🖵	no	sometimes	yes	Grand Total
no				
Count of WEIGHT	4	5	13	22
Average of WEIGHT	69.0	62.8	63.6	64.4
StdDev of WEIGHT	11.7	8.0	10.4	9.9
Min of WEIGHT2	58	52	50	50
Max of WEIGHT3	85	73	90	90
yes				
Count of WEIGHT	5	5	21	31
Average of WEIGHT	68.0	70.2	65.4	66.6
StdDev of WEIGHT	13.5	13.8	10.3	11.1
Min of WEIGHT2	50	53	47.5	47.5
Max of WEIGHT3	82	87	82	87
Total Count of WEIGHT	9	10	34	53
Total Average of WEIGHT	68.4	66.5	64.7	65.7
Total StdDev of WEIGHT	12.0	11.3	10.2	10.6
Total Min of WEIGHT2	50	52	47.5	47.5
Total Max of WEIGHT3	85	87	90	90

Hint:

PivotTable Field List							
Choose fields to add to	report:						
GENDER	~						
LENGTH							
WEIGHT							
HEAD							
SHOE_SIZE							
MATH							
BREAKFAST							
PORRIDGE							
PET							
SICK							
SPORT							
SKI							
CAR							
Drag fields between are	eas below:						
🝸 Report Filter	Column Labels						
	PORRIDGE 🔻						
Row Labels	Σ Values						
CAR 🗸	Count of WEI 🔻						
∑ Values ▼	Average of 🔻						
	StdDev of W 🔻						
LE TOOLS	Min of WEIGHT2						
	Max of WEIG 🔻						

11. Try different *PivotTable* layouts (sometimes helps some layout better understand the structure of the table):

х	□ 5 · C · -											PIVOTTABLE TOOLS		
	FILE HOME	INS	ERT	PAGE LAYOUT	FORMULAS	DA	ATA	REVIEW	v v	IEW A	DD-INS	ANALYZE	DESIGN	
[Row Headers	Bande	d Row	/5							
Su					rs 🗌 Bande	Banded Columns			==					
	Layou Show in Compact From					yle Options			PivotTable Styles					
	61 v		Show in	<u>compact rorm</u>	IGHT	IGHT								
	A		Sh <u>o</u> w in	Outline Form	с	D		E		F	G	н	1	
57														
58		-	Show in	<u>T</u> abular Form										
59	Row Labels				etimes	yes	Grand	Total			_			
60	no		<u>R</u> epeat /	All Item Labels										
61	Count of '				5	13		22						
62	Average of		Do <u>N</u> ot I	Repeat Item Labels	62.8	63.6		64.4						
63	StdDev of				8.0	10.4		9.9						
64	Min of WEIGHT3			58	52	50		50						
65	Max of WEIGHT4			85	73	90		90						
66	yes		-											
6/	Count of WEIGHT			5	5	21		31						
68	Average of WEIGHT			68.0	/0.2	65.4		66.6						
69	StdDev of WEIGHT2			13.5	13.8	10.3		11.1						
70				50	53	47.5		47.5						
/1	Total Count of WEIGHT4 82				8/	82		8/						
72	Total Average of WEIGHT 9				10	54		53						
73	Total StdDoy of WEIGHT2 42.0					04.7		05.7						
74	74 TOTAL STUDEV OF WEIGHTZ 12.0					10.2		10.0						
75	Total Max of WEIGHT3 50					47.5		47.5						
/6	TOTAL MAX OF V	VEIGHI	4	85	87	90		90						

12. Can you say something about each number in the last table?

Write down some sentences about similarities or differences of average body weights of students

- depending on car owning,
- depending on porridge eating.