Research Methods 2

Week 3: Exercise Sheet 1, numerical summaries

Solution sheet

Question 1.

When you click on $D\underline{a}$ ta and then Sort, you are confronted with the following dialogue box.

≥ n.04	ITAB - Un	titled								×
Pla	Edit Data	Cak Stat	graph ta	itor <u>Taala Window Help</u>						
			10 64	TIMES) 🕈 🗊 🖉 🔂 👘 🛈 🖸 🗑	114 2 1 1 1 1	6 -2-2 4 6 2	1.0		
0.	alan		-						1	
_	0	5/07/200	05 23:13:2	9	8					5
Uelo ETTE	ome to Mi > Retriev	nitab, j e "C:\DO	DE 650 F1	iort			🛛 🛛 🖂	OBICOLOGYA	VEEK3\UEI	EK 3. HT
Retr DOCU	ieving wo MENTS/WOR	Eksheet	from fi	Cl Quesi	Sert column(s):					52.525
COCC Dork	ngeleer was	SQUEEDCE	. NTU' en 03/12				100			
ELLE :	>				By column:	- Destandi	na l			
					By column:	E Descendir	ng.			
					By column:	Descendi	ng			
100			_		By column:	🗖 Descendir	ng			
-	THE LITER	***			Store sorted data in:				6	
	C4	(M			R New worksheet		_	244	C15	CHEA
•	Ouest	Q	5		Name:		(Optional)	CH	CD	C16-2
1	15.7	8 6			C Original column(s)					
2	15.1	-			Column(s) of current	worksheet				
3	16.3						12			
- 4	16.0			Select						
5	16.6									
6	16.0			Help		OK	Cancel			
1	16.4									
8	10.0									
10	16.1									
e'ie	10.0		1	从一款一款			1. 1. 1			3
	opert. If		1							
Satione	or more calu	mins of data	a						2	+ti.6
- s	tart	Co weed		E-been 😂	😫 Correns sheet 1 so	🙆 Research Hethode	MINETAR - Unkided	EN 🤇	0.848	A 23:14

If the boxes are filled in according to the hint, then the following screen will be obtained,

> 111	ITAB - Une	itled									i.	- 6 🗙
Ele	ER Ogia	(ak 2.00	graph Egit	or Isoks Window Help								
			0.0	TINKO	010 00	000	1111		4448	11.0		
C 50												
Weld MTR	ome to His	5/07/200	15 23:13:29	rt.					×	2: 080001.06W		KE3. NT
Bett	ieving wor	tisheet	TEOR EL		Sort columni	el:					1-600110-1000	
disco Wank	LOGY/MEEKS sheet was	s/MEESCS.	HSC 600. HTV: on 03/12		Ques1				3			
e .					By column: By column: By column: By column:	Ques1		Descendin Descendio Descendio Descendio	9 10 10			
1	EK3.MTW				Store sorted	data in:				ĺ.,		
٠	C1 Ower1	62	3		New wor	ksheet			- JOptinnit	C14	C15	C16 ^
1	15.7				C Original	column(s)						
2	15.1				@ Column	s) of cyrrent v	worksheet					
3	16.3				- C2				100			
4	16.0			Select					25			
5	16.5										-	-
	16.0			Help				QK	Cancel		-	
-	16.4						-				-	
	18.1		202									
10	16.5						-			-	-	
100	10.0										-	
						_	_	_				
	olect d											1.15
	and a start	and or oak		Const.	100	where the second		1 Martine La	Name and		1.1.2.4	

and clicking on OK gives

∑ MIN	ITAB - Unti	tled														- 🕫 🔀
Pie 1	tdit Data g	ak Stat	Graph Edi	tor <u>Tools</u>	Window the	lei										
1	1 8 X		001	TI	AR	011	1 1 1	🖱 🛈 🖸 I	a 1 a 3		圓濕	-1-20	68 2	1.0		
(503	alan														2	
OffCO Dock Rest strate Strate Strate	LOGY/GEEKS sheet was ults for: W > Sort 'Or > By 'Or >	AMERCI.M Javed on MEEKJ.MI Mal' (2) Mal' (2)	יזסי 03/12/2	001 14:2°	1: 03											8
5			_	_	_	_	_		_	_	_	_	_	_	_	215
III WE	EK3.MTW	***													. 6	
.+	CI	C2	C3	C4	cs	06	a	08	a	C10	C11	C12	C13	C14	C15	C16 🗠
1	15.7	14.5														
z	15.1	14.9														
3	16.3	15.1														
4	16.0	15.7														
5	16.6	16.7														
6	16.0	15.B														
7	16.4	15.B														
8	16.0	16.8														- 1
9	18.1	15.9														
10	16.5	16.D					_									
eife	35.0	10.0						1				1	1	/ /		3
III PO	opect. 🗗		6													
Current's	Worksheet / W	EEK3.MTW													2	115
🐮 si	tart	😂 wieła		2	Need		1 Contra	e abaat 1 as	0	esearch Meth	hode	MONETAR	- United	EN 🤇		🗞 23:15

Scrolling down the data window until you reach row 9, you find the value in C2 is 15.9 g/dl. Carrying on until row 18 gives you 16.4 g/dl and in row 27 the value in C2 is 16.9 g/dl. There are 35 values in each of these columns. So there are 17 values less than the value in row 18 of C2 and 17 values bigger than this

value, so the value in row 18 is the middle value and therefore is the median. In terms of the formula given in Appendix 1, the sample size is n = 35, and the median is the $\frac{1}{2}(n + 1)$ largest value, i.e. the $\frac{1}{2}(35 + 1) = 18^{th}$ largest value, which is 16.4 g/dl..

Similarly, the lower quartile is the $\frac{1}{4}(n + 1)$ th largest value, i.e. the $\frac{1}{4}(35 + 1) = 9^{th}$ value in the sample put in ascending order, which is 15.9 g/dl. Also the upper quartile is the $\frac{3}{4}(35 + 1) = 27^{th}$ value in ascending order in the sample, which is 16.9 g/dl.

Question 2.

If you choose <u>Stat</u> from the main menu bar in Minitab and then select <u>Basic</u> Statistics and then <u>Display Descriptive Statistics</u>... you will be presented with the screen show below.

∑nn	IITAB - Un	titled													- • 🗙
Pie	Edit Data	Cek Stat	Graph Ed	tor <u>Tools</u>	Window <u>H</u>	hilp									
			1) (r	T 1	AR	0 11	1	n 🖻 🛈 🕯	2 🗐 🖬 🖄		-1-2	ah 🕅 🗹	1.0		
	salion														
ONCO Dozk	LOGY/GEES	(3), WEER(3., 1 , staved or	mo; 03/12/:	2001 14:21	0:03										2
Res	uits for: \	WEEK3.M	TW		Displa	y Descrip	tive Stati	ities			X				
19778 19778 C 19778 C	> Soct '(> By '(hasz1' C2; hasz1'.	9		C1 C2	Quer	s1.	Variable	051	6					
										8					
								By varia	ables joption	all:					52
								(×
¢															2
₩W	ERG MTW							a			10			i	- 🗆 🛛
.+	CI	C2	C3	C4							C12	C13	C14	C15	C16 🖄
15	16.7	16.7										-			
26	15.8	16.7			-										
27	15.8	16.9					- 11 ·					-			
28	16.4	17.1				Selee	4.1	St	atistics	Graphs					
29	15.7	17.1								-					
30	16.5	17.3				Help		1	0K	Cancel					
31	16.3	17.4				1	1	11			2				1
32	17.1	17.5													
33	16.4	17.9				-	-								
34 6	16.9	18.1					4								3
	opect,		dealer in f												2.40
-actual	tart	statistics and	uspiay in 5	ession vendo	wardet.		- El	ene abaset de		a areh Matheole	Sugar	I.I. Internet		1.0.4	1116
	care	- Allans		0	and the second se		al co	COLUMN T T	- E -	and an	2 Multi	Aa - 0 100e0	En	A 10 19-3	

Double clicking on Ques1 in the left hand box will place Ques1 in the <u>Variables</u>: box. Clicking on <u>O</u>K gives the following screen.

∑ MP	ITAB - Unt	itled														- 6 🛛
Pie	Edit Dyta	Cek Stat	Graph Ed	tor <u>T</u> eals	Window their	p										
		白色	0.00	T 1	ARC	980	1 1 1	🖱 🛈 🖸	🗐 🖬 🤅		圖派	-Z-Z-	4812	1.0		
Addition Control Contro Control <thcontrol< th=""> <th< td=""><td></td></th<></thcontrol<>																
SUBC SUBC SUBC SUBC	> Niniw > Naxiw > N; > Miss criptive S	un; un; ing. Ratistics:	Ques1													8
Vari Quer Vari Oues BTB	able H 1 35 able Hax 1 18	N" Ra 0 16.4 1800	ara 581 M 31 0.	0emos 550e 140 0.83	ev Minimu 28 14.50	an (D 15.90	11 Bedien 10 16.400	93 9 16.900								
171238																×
5											_				0	1
III W	IK3 MTW			1.			11. I I I I I I I I I I I I I I I I I I	1.11				10				
-+	C1 Outer1	0	G	ci	G	06	Ċ/	08	G	.C10	011	C12	C13	C14	CIS	C16 2
25	16.2	16.7														
26	15.8	16.7														
27	15.8	16.9														
28	16.4	17.1														
29	15.7	17.1														
30	16.5	17.3														
31	16.3	17.4					-						-			
32	17.1	17.5														
33	16.4	17.9														
34	16.9	18,1										-				
6	10.0	10.0						1		_						3
-		Contraction 1														1000
HI PI	opect 6		Q												19.83 - Ories	
Current	Worksheet / W	EB(3.MTW												Editable	2	9119
1 5	tart	😂 ved 3		0	weed-ct		E cercis	a sheet 1 so	- 🔁 i	tenearch Meth	sode	MUNITAN	- Untilded	EN 🦷) 	A 23:19

The results of the calculation are shown in the Session window. Several quantities have been calculated. The ones of primary interest are: the Median, which is 16.4 g/dl, as obtained previously, and the quartiles, which Minitab labels Q1, for the lower quartile and Q3 for the upper quartile. These are, respectively, 15.9 g/dl and 16.9 g/d, again agreeing with the results of question I.

Some of the other results calculated by this command may already be familiar, and in any case soon will be. The mean and standard deviation (labelled StDev) will be described next week and the Standard Error of the mean (SE Mean) will be considered the week after.

The only other values given are the maximum and minimum values in the sample, which together with the median and quartiles allows us to quote the *five number summary* for this sample as:

14.5 15.9 16.4 16.9 18.2 (all g/dl).

Question 3.

By following the instructions in the hint, a copy of 'Ques1' is placed in C2 and the final element of the latter changed to 81.2. The method of question 2 requires you to click on <u>S</u>tat from the main menu bar and then select <u>B</u>asic Statistics and then <u>D</u>isplay Descriptive Statistics.... You then select 'Ques1' as described in the answer to question 2 and then repeat the process to select C2 as well. This will leave the dialogue box looking as follows:

≥ 10	IITAB - Unt	itled												- 🕫 🗙
Pla	Edit Date 1	Cek Stat	Graph Ed	tor <u>Tools</u> <u>Win</u>	dow <u>tt</u> elp									
	1 @ X		in ce	T 1 1	110 8	1	🖸 🗇 🖉	3 13 社		-1-20	18 2	1.0		
E = -	ssion													
SUBC SUBC SUBC	> Noxim > N; > MMigg:	ing.												-
Des	criptive S	tatistics:	Quest		Disalas Dasada	due Diet	a store		1					
0				Incon Phillips	publick nearth	tive stati								
Quez	1 35	0 16.4	31 0.	140 D.828	C1 0.08	s1.	Variables:							
Vari Oues	able Hax 1 18	inux . 300					Ques1 Ci	2	0.8					
8778 8778	> Copy 'On >	es1' c2.					By variabl	les (optiona	4:					
														*
										-				1.05
1 W	ERG. MTW	***	-											- 🗆 🛛
-+	C1	C2	C3	લ						C12	C13	C14	C15	C16 ^
	15.7	16.7												
2	15.1	15.1												
3	16.3	16.3				···· 11							-	
4	15.0	16.0			Selo	# 1	Stati	istics	Graphs	1				
5	16.6	16.5												
6	16.0	16.0			Help	1	1	OK	Cancel					
7	15.4	16.4	_										-	
8	16.0	16.0				-								
9	18.1	18.1												
10	16.5	16.5												-
eit	100	+1.0		12 1	-	1				1				
-		Contractor of		_	_		_	_	_				_	1000
	opect of		7										10.11	
Calculate	e descriptive s	tatistics and	display in S	ession window						-			2	3.23
- S	tart	Can weeks		💭 H40		E Co	rene sheet 1 so	- 🙆 Rese	arch Methode	MONETAN	- Untilded	EN	0	A. 23:23

You should then click on <u>O</u>K, which will give the following screen:

∑ MIN	ITAB - Unt	itled														
Pla	Edit Data	Cak Stat	Graph Edit	or Icola Wi	ndow <u>th</u> elp					a start to						
1		0.8	a) (a 🚦	1 1	ARG	?	10	🖱 🛈 🖸	🗐 🖬 🕅	100		-2-Z-	4812	1.0		
C Se	ssion												1110000		6	183
SUBC SUBC	> N; > Miss	ing.														
Des	criptive S	tatistics:	Quest,	C2												
Vari Quer C2	abie N 1 35 35	H* He 0 16.4 0 18.	an 58 M 31 0. 23 1	ean StDev 140 0.828 .85 10.95	Kiniaun 14.500 14.50	01 15.900 15.90	Nedian 16.400 16.40	03 16.900 16.90								
Vari Ques Cl	able Max 1 18 8	18.08 .200 1,20														
BILB	>															- 1
6	0															- 2
# w	TK3 MTW	***													6	In B
	C1	0	C 2	C1	C5	CE.	17	69	0	C10	C11	(012	00	611	C15	C16.
	Quest								 .							C.I.C
1	15.7	15.7				_										
z	15.1	15.1														
3	16.3	16.3														
4	16.0	16.0														_
5	16.6	16.5														
б	16.0	16.0														
7	16.4	16.4														
8	16.0	16.0														
9	18.1	18.1														
10	16.5	16.5														
eir	100	+7.0						1								
		Contraction of the														
E PI	opect 15		G													
ment	Worksheet / W	EB(3.MTW												Editable	2	3.27
1 8	tart	Chan 😋		ڬ m	ndet i		Conche	sheet I so	- 🖸 R	esearch Med	ode	MONTAN	- Untilled	EN 🤇	1.44	A 1327

It can be seen that the median and quartiles (Q1 and Q3) are unchanged by the perturbation to the largest value in the data. However the mean and standard deviation (StDev) have changed markedly.

This is a feature of the median and quartiles. Once you have placed the sample in ascending order, the median and quartiles are (crudely speaking) located a quarter, a half and three quarters of the way up the sample. If the sample size, n, is such that the point $\frac{3}{4}(n+1)$ of the way up the sample is not the largest value, then changes to the largest value will have no effect on the median or quartiles of the sample. Indeed, all the values above the point $\frac{3}{4}$ of the way up the sample could be changed without any effect on the quartiles or median. The lack of sensitivity of these summaries to changes in the data can be both an advantage and a disadvantage, depending on the circumstances. Note, by way of a contrast, how the mean and standard deviation are very sensitive to changes in the data.

End of solution sheet