# MRes in Medical Statistics MMB8028

### School of Mathematics and Statistics

### Practical session on Minitab, descriptive statistics and the Normal distribution: outline solutions

1.

The file HEIGHT.MTW is opened by clicking on <u>File</u>  $\rightarrow$  <u>Open</u> Worksheet and then selecting the file in the usual Windows manner. When the box stating that a "*copy of the content of this file will be added to the current project*" appears, click <u>OK</u> The screen below shows the results of selecting <u>Stat</u>  $\rightarrow$  <u>Basic Statistics</u>  $\rightarrow$  <u>Display Descriptive Statistics</u>... and entering C1 (or 'Heights cm') in the <u>Variables</u>: box. The boxplot is obtained by clicking on the <u>Graphs</u>... button in the Display Descriptive Statistics box and then selecting Boxplot of data and then clicking on <u>OK</u> (twice).

-	b - Untitled	1.00									-									-					
	dit Data <u>C</u> alc																								
	😂 🕺 🔁 I											206	1 d. 0	7											
		- 🖆 🕨	t 🛷 🕂	p ≠		•	$   \times  \mathbf{Q} $	N 1	TOC	- / • 10	1														
Sessi	on																								
	eving workshe			\TEACHING	MDPHD\Int	trol\heigh	ht.MIW'																		
orka	neet was save	a on 15/1	11/2004																						
esu	Its for: heigh	t.MTW																							
TB > JBC>	Describe 'He GBoxplot.	ights cm'	9																						
	riptive Statis	tics: Heic	ints cm																						
arial				StDev 1	Minimum	01 Me	edian	03																	
	ts cm 99 0																								
arial	ole Maximu ts cm 119.5	20						(																	
e a gui	-0 GM 119-0							4	Boxplot of	Heights cm															
oxp	lot of Heights	s cm									Bo	xplot of H	eights cm												
гв >									120 -				-												
_	_												Ĩ												
-	0.							-11	115 -																'
_	ht.MTW ***																								ıl
÷	C1 Heights cm	C2	C3	C4	C5	C6	C7	Η.									C17	C18	C19	C20	C21	C22	C23	C24	
1	117.9					7			110 - Subject 105 -					_								1			
2	106.3								TE I																
3	103.3							1 3	<b>1</b> 105-				1												
1	105.9																								
5	116.7																								
5	107.4								100 -																
7	105.8																								
8	106.3												14% +												
•	99.4								95 -																
0	108.7																J								
1	104.1																								
2	110.2																								
	111.0																								
	106.9																								
1	110.0																								
4 5																									
4 5 6	103.5																								
3 4 5 6	103.5							_							_	_		_	_	_	_			_	ź
4 5 6																									
4 5 6 Proj.																									1
4 5 6 Proj.		MTW				<b>.</b>												_					4 × 12 1	4.3 12:5	

The mean (labelled Mean) and standard deviation (labelled StDev) appear in the output in the session window. The median is labelled as such and the lower and upper quartiles are labelled Q1 and Q3 respectively. Note that this command gives other information, such as the standard error and the minimum and maximum.

The boxplot could have been obtained independently by clicking on <u>Boxplot...</u> under the <u>G</u>raph item on the menu.

2.

The data can be typed into the data window and the column for haemoglobin concentrations is named as Hbs. Applying the method for finding means used in question 1 gives a mean of 15.253 g/dl and an SD of 1.461 g/dl. If the second column of data is entered in a column which has name 'sex' then the way to obtain separate means and SDs for the sexes is to click OK on the screen as set out below.

initab - Untitl Edit Dat	_			tor Tools	Window	Heln Assis	tant				pro containe i	strategy the	-	-		1000							_	0
								) 🖻 🗟 1	16		K   -2 -		0											
		- 61	k # +	- p ⊉			- <b>X</b>	RINI		. • []														
Session																								
																								_
lcome to M B > Transp BC> Afte B > Descri	dinitab pose C1 er. ibe 'Hb	, press -C15; s'.						_																
scriptive	Statis	tics: Hb	s																					
riable N	N* 5 0	Mean 15.253	SE Mean	StDev 1.461	Minimum 12.900 :	Q1 M	edian 5.800 16	Q3																
iable Ma											isplay Descriptive	Statistics	-	×										
1	17.800										C1 Hbs C2 sex	Variables:												
>											uz sex	Hbs		^										
														Ŧ										
												By variables (o		_										
										_		sex		^	_									
orksheet 1 '	***									-		I		-	-									
C1		C2	C3	C4	C5	C6	C7	C8	C9	С					C16	C17	C18	C19	C20	C21	C22	C23	C24	Ť
Hbs 14		sex									1													-
14	+.2 4.4	1									Select	Statistics.	<u>G</u> raphs											+
14		1									Help	QK	Cancel											T
14	4.8	1																						+
12		1								-		_		_	-									+
13		1																						T
	5.6	2																						
15		2																						+
15		2																						+
16		2																						+
15		2																						
17		2																						+
16	5.4	2						-																+
oj 🗊 🛛		2													_									
-			undan dia Ca																				Editable	
ate descriptiv	ve statist	ics and de	COLUMN TWO IS NOT	ssion windo		<u>i</u>	(Terr			-				-	-	-		-		-		<b>C</b> 11 <b>A</b>		le 13:
						<b>T</b>	<u>19</u>															EN 🔺 📜	02	

Doing this gives the following output in the session window<sup> $\dagger$ </sup>.

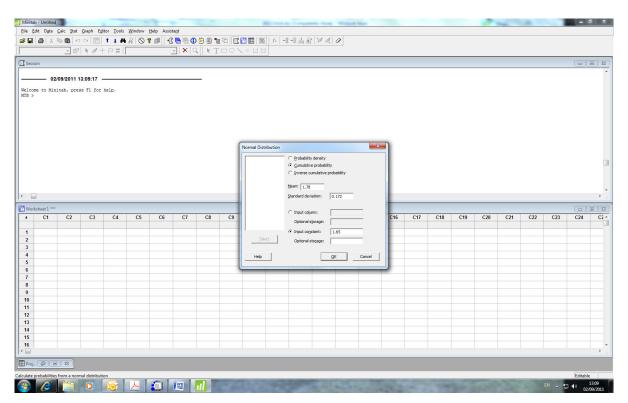
Variable Hbs	1	7	0	13.929	0.283	0.750	Minimum 12.900 15.800	13.200	14.200	14.600	
Variable Hbs	sex 1 2	1		0							

All the descriptive statistics are now shown separately for the sexes.

<sup>&</sup>lt;sup>†</sup> Note that the output shown (here and in other examples) is only displayed if the commands prompt is enabled.

#### 3.

Clicking on the path suggested in the question leads to the following dialogue box, which should be filled in as shown.



On clicking OK the following is written to the Session window.

MTB > CDF 1.65; SUBC> Normal 1.78 0.172.

## **Cumulative Distribution Function**

```
Normal with mean = 1.78 and standard deviation = 0.172 x P( X <= x ) 1.65 0.224880
```

This shows that a proportion 0.22488, or about  $22\frac{1}{2}$ % of this population lies below 1.65 log units.

### 4.

The answer to this question is found in a way that is almost the same as for question 3. The dialogue box is that shown in the answer to question 3 but the <u>Inverse cumulative</u> probability button, rather than the <u>Cumulative</u> probability button should be checked. The value 77.5% must be entered in the Input constant box as 0.775, i.e. as a probability or proportion, not a percentage. Doing this and clicking on <u>O</u>K gives the following output in the Session window:

MTB > InvCDF 0.775; SUBC> Normal 1.78 0.172.

### **Inverse Cumulative Distribution Function**

```
Normal with mean = 1.78 and standard deviation = 0.172
P( X <= x ) x
0.775 1.90993
```

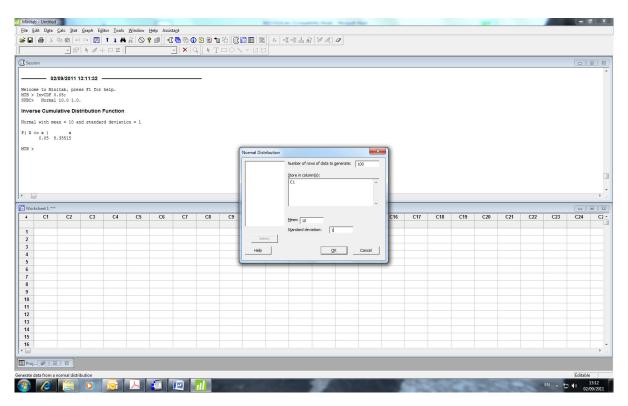
Thus 77.5% of patients in the population have a value of the variable measuring phototoxicity that is less than approximately 1.91 log units.

The proportion above this value is 1-0.775 = 0.225, i.e. the same proportion that is below 1.65 (cf. question 3). From the symmetry of the Normal distribution this implies that 1.91 is the same distance above the mean as 1.65 is below it. As the mean is 1.78, this is indeed the case, 1.78-1.65 = 1.91-1.78 = 0.13.

5.

Using the method of question 4, with an Input constant box of 0.05 gives a threshold of 8.35515.

Following the procedure for generating random numbers given in the question leads to the following screen:



Clicking on <u>OK</u> produces 100 Normal numbers from a *population* with mean 10 and SD 1 (i.e. this is a special case where we know  $\mu$ =10,  $\sigma$ =1).

Computing descriptive statistics using the methods of questions 1 and 2, and then using the method suggested in the question to determine the number of values less than the threshold and computing a classification table leads to the following session window.

	B € • • • ■			-	$   \times   Q $	N T □	• / OE															
Session																						
	Statistics: C1																					
	N* Mean SE Me 0 9.6704 0.09					Q3 2453																
riable Maxi																						
B > let c2= B > XTABS C2 BC> Layout BC> Counts BC> DMissi	t 1 0; s;																					
bulated sta	tistics: C2																					
owa: C2																						
Count																						
92																						
8																						
8 11 100																						
8 11 100 TB >																						
8 11 100 TB >																						
11 100 TB >																						
B > Worksheet 1 ***		C4	C5	C6	C7	C8	C9 (	10 C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	
8 11 100 TB > Worksheet 1 *** • C1	C2 C3	C4	C5	C6	C7	C8	C9 (	10 C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23		
B > Worksheet 1 **** • C1 4 9.1623	C2 C3	C4	C5	C6	C7	C8	C9 (	10 C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23		
8 11 100 18 > Worksheet 1 **** 4 9.1623 5 9.5459	C2 C3 3 0 9 0	C4	C5	C6	C7	C8	C9 (	10 C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23		_
B > Worksheet 1 *** 4 9.1623 5 9.5459 6 9.9048	C2         C3           3         0           9         0           8         0	C4	C5	C6	C7	C8	C9 (	10 C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23		_
8 11 100 TB > ₩orksheet1 *** 4 9.1623 5 9.5459 6 9.9048 7 9.5058	C2         C3           3         0           9         0           3         0           3         0           3         0           3         0	C4	C5	C6	C7	C8	C9 (	10 C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23		_
8 10 100 18 > ₩0rksheet 1 *** • C1 4 9.1623 5 9.5459 6 9.9048 7 9.5058 8 9.6481	C2         C3           3         0           9         0           8         0           1         0	C4	C5	C6	C7	C8	C9 (	10 C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23		
Bill 8 100 TB > Worksheet 1 **** • C1 4 9.1623 5 9.5459 6 9.9048 7 9.5058 8 9.6481 9 9.7421	C2         C3           3         0           9         0           8         0           1         0	C4	C5	C6	C7	C8	C9 (	10 C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23		_
8         8           III         100           IIB >            Worksheet         1************************************	C2         C3           3         0           4         0           5         0           6         0           1         0           3         0           3         0	C4	C5	C6	C7	C8	C9 (	10 C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23		_
8         6           Worksheet         ************************************	C2         C3           3         0           9         0           8         0           1         0           3         0           3         0           5         0	C4	C5	C6	C7	C8	C9 (	10 C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23		_
6         7         9           Workshedt         ************************************	C2         C3           3         0           9         0           8         0           1         0           3         0           5         0           5         0	C4	C5	C6	C7	C8	C9 (	10 C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23		
6         6           TB >            Workshedt 1         ****           •         C1           •         C1           •         S           •	C2         C3           3         0           3         0           3         0           1         0           3         0           3         0           5         0           5         0           5         0           5         0	C4	C5	C6	C7	C8	C9 (	10 C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23		_
6         6           TB >            Worksheet 1         ***           C1            4         9.1623           5         9.5459           6         9.9048           7         9.5658           8         9.6481           9         9.7421           10         9.2673           11         9.2068           12         9.3146           13         10.7245           14         8.1524           15         10.5319	C2         C3           3         0           9         0           8         0           1         0           3         0           5         0           5         0           5         0           4         1	C4	C5	C6	C7	C8	C9 (	10 C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23		
6         6           TB >            •         C1           •         C1           •         C1           •         C1           •         9.1623           5         9.5459           6         9.9048           7         9.5668           8         9.6481           9         9.7421           10         9.2698           12         9.3146           13         10.7245           14         8.1524	C2         C3           3         0           3         0           3         0           1         0           3         0           3         0           5         0           5         0           5         0           5         0	C4	C5	C6	C7	C8	C9 (	10 C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23		
6         6           TB >            Worksheet 1         ***           C1            4         9.1623           5         9.5459           6         9.9048           7         9.5658           8         9.6481           9         9.7421           10         9.2673           11         9.2068           12         9.3146           13         10.7245           14         8.1524           15         10.5319	C2         C3           3         0           3         0           3         0           4         1           5         0	C4	C5	C6	C7	C8	C9 (	10 C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23		

Notice that the *sample* mean and SD, *m* and *s*, are not exactly 10 and 1, but vary from these values by sampling variation. Their values in the sample I generated are, respectively, 9.6704 and 0.9591 (shown below) but your values will vary from these as everyone will have a different random sample from the population.

The table shows that 8 of the 100 values are below 8.35515, so 8% rather than 5% fall below the 5% threshold. This is only to be expected - the number falling below 8.35515 will, *on average*, be 5%, it just will not be guaranteed to be 5% in every sample. This can be appreciated by repeating the instruction to form a column of 100 numbers from a Normal population of mean 10 and SD 1. The proportion below the 5% threshold in this sample will, in general, be different to that obtained from the first sample you drew. For my second sample it was 3.

If the instruction to form a column of 100 values is replaced by a similar instruction to form 10000 values then the descriptive statistics obtained are:

### **Descriptive Statistics: C1**

Variable N N\* Mean SE Mean StDev Minimum Q1 Median Q3 C1 10000 0 10.004 0.0100 1.004 6.398 9.331 9.989 10.680 Variable Maximum C1 13.692

Note that the sample mean and standard deviation, 10.004 and 1.004, are closer to their population counterparts, 10 and 1, in this much larger sample.

Using exactly the same methods as before shows that 487 of the 10,000 values are below 8.3551, i.e. 4.87% of this sample is below the 5% point.