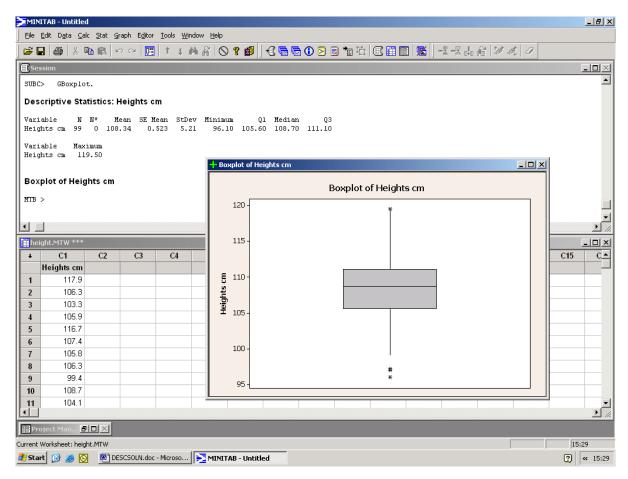
MRes in Medical Statistics MMB8028

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 and then selecting Boxplot of data and then clicking on <u>OK</u> (twice).



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 (Tr Mean is a trimmed mean, a quantity we will not use, that attempts to combine the advantages of a mean and a median)

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.

MINIT	AB - Untit	led														_ 8 ×
Eile E	dit D <u>a</u> ta g	<u>c</u> alc <u>S</u> tat	Graph Edi	or <u>T</u> ools <u>v</u>	√indow <u>H</u> el	p										
- 	6	h 🛱	n ce	n t i	AA	0 ? 🗗	+3		<mark>> 🗟 📬</mark> 昭		<u>民</u>		8	1.0		
SUBC>	ion	am;														_ U ×
SUBC> SUBC>	N; NMiss:	ing.														
Desc	riptive S	tatistics	Hbs		Display I	Descriptive	Statistic	5			×					
	15 ble Max:	0 15.2		ean StDe 377 1.46		Hbs sex		¥ariable Hbs	es:		2					
Hbs	17	.800						-			-					
MTB >								Puunari	ables (optio	noll:						
								sex	anies (uhuni							Ţ
•								Sex			<u>^</u>					•
Worl	ksheet 1 *	**									-					- U ×
+	C1	C2	C3	C4	-							C12	C13	C14	C15	C16_
	Hbs	sex		()												
6	12.9	1			_											
7	13.2	1														
8	16.6	2				Select	1									
9	15.8	2			_	aciect		<u>S</u>	tatistics	G <u>r</u> aphs.						
10	17.8	2		2				_		2 - 2 2 - 1		-				
11 12	15.8 16.0	2			- <u> </u>	lelp			<u>0</u> K	Cancel				-		
12	15.9	2			<u></u>	-			1		-					
14	17.0	2		.)				2			-					
15	16.4	2				-										
16	10.4					1										•
E Proj	ect Man	BOX														
Calculate o	descriptive s	tatistics and	l display in S	ession windov	,									Editable	1	14:16
2 Start		()	DESCSOLN.c	loc - Microso.		ITAB - Unti	itled	1							2	x 🏂 14:16

Doing this gives the following output in the session window^{\dagger}.

MTB > De	escribe	е 'Н	lbs ' ;									
SUBC> By 'sex';												
SUBC>	Mean;											
SUBC>	SEMear	.;										
SUBC>	StDevi	ati	on;									
SUBC>	QOne;											
SUBC>	Mediar	.;										
SUBC>	QThree	;										
SUBC> Minimum;												
SUBC>	Maximu	ım ;										
SUBC>	N;											
SUBC>	NMissi	ng.										
Descriptiv	ve Statis	tics	: Hbs									
Variable	e sex	Ν	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3		
Hbs	1	7	0	13.929	0.283	0.750	12.900	13.200	14.200	14.600		
	2	8	0	16.413	0.250	0.706	15.800	15.825	16.200	16.900		
Variable	e sex		ximu									
Hbs	1		4.80									
	2	1	7.80	0								

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

[†] 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.

MINITAB - Untitled							_8>
<u>File E</u> dit D <u>a</u> ta <u>C</u> alc <u>S</u> tat <u>G</u> raph E <u>d</u> itor <u>I</u>	ools <u>W</u> indow <u>H</u> elp						
🖆 🖬 🐇 📲 🛍 🛏 🛄	t + M & 🛇 ? d	+3 🖬 🖬 🖬 🖉 🖬 🕨	e ti 🗉 📰 🗉	S - Z -	L€ ≫ «	0	
E Session							_ _ ×
15/11/2004 14:25:17							
Welcome to Minitab, press Fl for h	elp.						
MTB >							
I	Normal Distribution			×	1		
		© Probability density					
		Cumulative probability	lity				
		C Inverse cumulative	probability				
		<u>M</u> ean: 1,78					-
		Standard deviation:	0.172				• //
Worksheet 1 ***							- 🗆 ×
+ C1 C2 C3		C Input column:	[C13	C14	C15 C16
		Optional storage:				10 10	
1		Input constant:	1.65				
3	19- 	Optional storage:	11.00			-	
4	Select	opuonai siorage.					
5							
6	Help		<u>0</u> K	Cancel			
7 -							
9							
10							
11							
							• //
Project Man 🗗 🗆 🔀							
Calculate probabilities from a normal distribution	1				[14:26
🐮 Start 🔀 🍠 🔯 💆 DESCSOLN.doc - M	1icroso 🔁 MINITAB - Un	titled					<table-cell> < 🏷 14:26</table-cell>

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:

MINITAB - Untitled								_ 8 ×
Eile Edit Data Calc Stat Graph Editor To	ols <u>W</u> indow <u>H</u> elp							
🚰 🖬 🎒 X 🖻 🖀 🕫 🕫 🖪 1	I M & 0 ? d		2 2 to 12 2 2	B S -Z - Z - 	L 🔐 💓	1.0		
Session		4						_ _ ×
								•
15/11/2004 14:31:38 —								
Welcome to Minitab, press Fl for he	lp.							
Inverse Cumulative Distribution Fu	nction							
No	ormal Distribution			×	1			
Normal with mean = 10 and standar		<u>Generate</u> 100	rows of da	ta				
P(X<= x) x 0.05 8.35515								
0100 010010		Store in column	(s):					
MTB >		c1		<u> </u>				
								-
								• //
Worksheet 1 ***				*				-OX
+ C1 C2 C3					C13	C14	C15	C16_
		<u>M</u> ean: 10			-			
2		Standard deviat	ion: 1	_		-		
3	1							
4	Select							
5	Help		<u>0</u> K	Cancel				
6								
7 8								
9								
10								
								<u> </u>
Project Man						- 11 I I		
Generate data from a normal distribution 3 Start [] 6 6 6 6 6 6 6		wal-d				Editable	1.0	14:34 « 🗞 14:34
	MINITAB - UN	iciciea					L.	« » 14:34

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 μ =10, σ =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.

MINITAB - Untitled								_ 8 ×
Eile Edit Data Calc Stat Graph Editor Too	ols <u>W</u> indow <u>H</u> elp							
1 🖂 🖓 🐇 🖻 🖀 🗠 🖂 🧮		+C 🖶 🔂 🖸 🛙	1 1 1 1		6 20 4	\$ 0		
Session				14				-0×
								-
15/11/2004 14:31:38		2						
Welcome to Minitab, press Fl for he:	lp.							
Inverse Cumulative Distribution Fu	nction							
No	rmal Distribution			×	1			
Normal with mean = 10 and standar		Generate 100	rows of dat	a				
P(X<=x) x 0.05 8.35515		Touring in the	iono or dui	-				
0100 010010		Store in column(s):						
MTB >		c1		<u> </u>				
								-
								•
Worksheet 1 ***				-				-UX
+ C1 C2 C3					C13	C14	C15	C16_
		Mean: 10	-					
2		Standard deviation	1	1				
3								
4	Select							
5	Help		ОК	Cancel				
6								
7 8		5						
9		19						
10								
11								- -
								• //
Project Man								
Generate data from a normal distribution	1					Editable		4:34
🛃 Start 🔞 🈹 🔯 👼 DESCSOLN.doc - Mic	roso 🔁 MINITAB - Un	titled					<u>?</u> «	🏂 14:34

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, 10.009 and 0.990 (shown below) but your values will vary from these as everyone will have a different random sample from the population (but see the comment at the end of this answer).

The table shows that 4 of the 100 values are below 8.35515, so 4% 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.

Descriptive Statistics: C1

Variable Ν* SE Mean StDev Minimum Ν Mean Q1 Median Q3 100 0 10.009 0.0990 0.990 6.327 9.346 10.082 10.672 C1 Variable Maximum 12.057 C1 MTB > XTABS C2; Layout 1 0; SUBC> SUBC> Counts; SUBC> DMissing C2. **Tabulated statistics: C2** Rows: C2 Count 0 96 1 4

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 Cl	N 10000	N* 0	Mean 10.004		StDev 1.004	Minimum 6.398	Q1 9.331	Median 9.989	Q3 10.680				
Variable Cl	Maximu 13.69												

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.

(note that because this question uses randomly generated values, the answers above will not, in general, be the same as those shown here. However, if before you generate any random numbers, you type the following command in the Session window:

MTB > *Base 321*

All

100

then the first two sets of 100 numbers and the subsequent 10,000 should be the same as used here. You do not have to type the Base command when using random numbers in MINITAB, it is just that if you do you it sets the random number generator to a reproducible state.}