

4

DESCRIBING YOUR DATA

© Nick Lee and Mike Peters 2016.

QUESTION 1.

You sent out a questionnaire regarding peoples' television preferences to 30 people. One question asked if they liked a certain television programme and you received 19 valid responses. This data is shown below where a 0 means did not like, 1 means did like and 2 means unsure.

Response 2,0,0,0,1,0,2,0,0,2,0,0,1,1,1,0,0,1,1

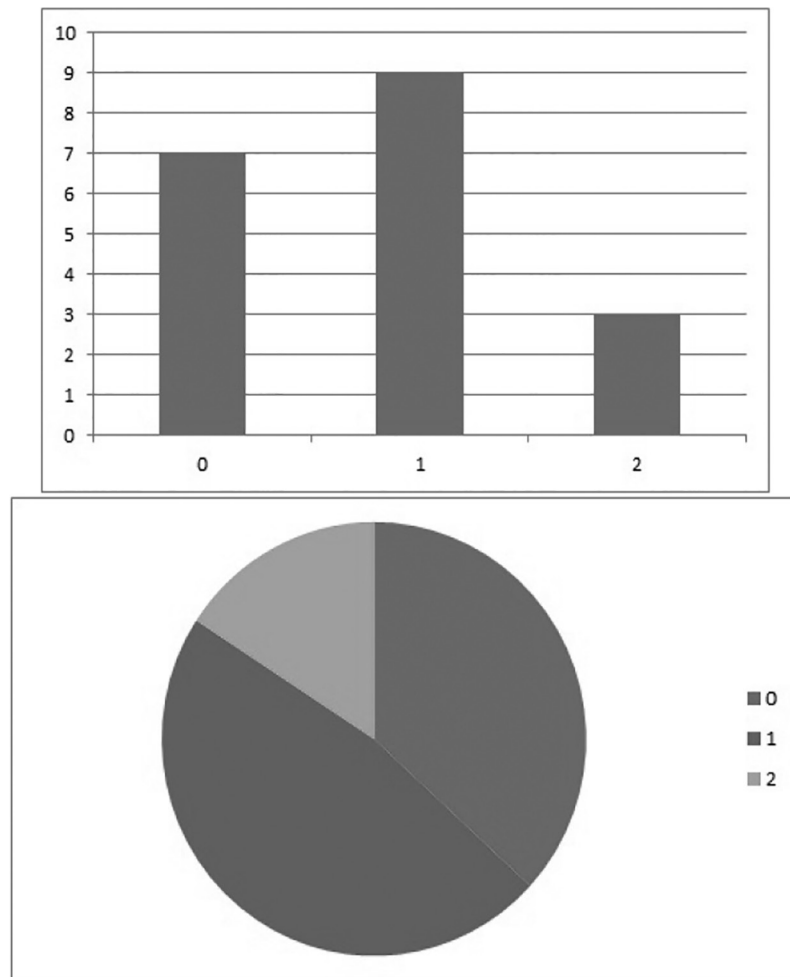
Enter this data into an Excel spreadsheet and create a frequency distribution.

QUESTION 2.

Using the data from question 1, create a bar chart and a pie chart in Excel. Don't forget to add titles and labels!

QUESTION 3.

There are 2 diagrams below, a bar chart and a pie chart. They are the result of a different sample of people who were asked if they liked the same programme as in question 1. As they stand they just represent data and require additional features to make them meaningful. Add these additional features so your boss can make sense of the diagrams.



They both need labels and titles added. It is common to add percentages to pie charts.

QUESTION 4.

Create an instance of SPSS, then create a variable called 'scores' and enter the data shown below.

53	80	43	57	57	53	80	43	57	57
----	----	----	----	----	----	----	----	----	----

QUESTION 5.

You and some of your friends, being good citizens, decided to produce a good pub guide for your town. You visited 10 establishments and tasted their beers and recorded the price you paid. In order to be of service to your fellow citizens you also rated the beer on a scale of ok, good and excellent. The data you recorded is given below. From this data construct a cross-tabulation table (a Pivot table in Excel).

Bar	Quality rating	Price
1	good	3.00
2	ok	2.50
3	ok	3.25
4	excellent	4.00
5	good	4.00
6	ok	2.75
7	good	3.00
8	excellent	3.50
9	good	3.00
10	good	2.50

Once you have created the cross-tabulation table, fill in the boxes below.

Rating	Total
ok	3
good	5
excellent	2

Number of beers costing less than £3.50	3
Number of beers costing more than £4	0
Number of beers rated as good and costing less than £3	1

QUESTION 6.

The table below shows the scores of 15 students who sat a statistics test.

Calculate the mean, mode and median.

student	score
1	24
2	35
3	35
4	46
5	40
6	40
7	34
8	21
9	20
10	35
11	40
12	15
13	50
14	36
15	48

Fill in the values for the mean, mode and median in the box below.

Mean	34.6
Mode	35
Median	35

Give an interpretation of the test in terms of the mean, mode and median. You might want to consider how the 3 values relate to the skewness of the data and how this can be interpreted as to the fairness of the test.

This test was fair because:

All three measures give similar values. This suggests the data is symmetrical about the mean.

QUESTION 7.

Below is the output from an SPSS operation on some data. You have been asked to explain this output but you have not seen the original data.

Using your answers to the following questions, make a guess as to how the data was collected.

The number of items of data is	50
The number of distinct data items is	26
The number of repeated data items is	24
I think the data comes from	a questionnaire and is secondary data

Frequencies

Statistics

Indicator of first matching case as

Primary

N	Valid	50
	Missing	0

Indicator of each first matching case as Primary					
		Frequency	Percent	Valid Percent	Cumulative Percent
→ Valid	Duplicate Case	24	48.0	48.0	48.0
	Primary Case	26	52.0	52.0	100.0
	Total	50	100.0	100.0	

QUESTION 8.

Create an Excel worksheet and enter the following data and calculate the mean, median, standard deviation, variance and skewness of the data.

ID	1	2	3	4	5	6	7	8	9	10
Glasses	0	3	7	15	2	1	0	0	10	5

Using your judgement which, if any, are useful statistics in describing the data?

Measure	Useful/not useful	Useful/not useful because:
Mean	useful	gives an idea of mean consumption
Median	useful	can be compared to the mean to give an idea of the skewness of the data
Standard deviation	useful	if the data is assumed to be normally distributed
Variance	useful	gives an idea of the variation around the mean
Skewness	useful	gives a number indicative of the shape of the data

Plot the data to verify your answers above.

QUESTION 9.

The summary statistics shown below were calculated based on a random sample of 100 people who were asked how far they travelled by car in the first week of June 2015.

Construct a boxplot from the given statistics (you may not need all of the statistics given).

Statistics		
miles		
N	Valid	100
	Missing	0
Mean		71.75
Median		73.50
Mode		74
Std. Deviation		15.238
Range		48
Minimum		47
Maximum		95
Percentiles	25	56.00
	50	73.50
	75	86.75

MINI PROJECT

The following data was collected on the same day, at the same time in two shops owned by the same person in neighbouring towns. Shop A employs 10 people and shop B employs 5 people. Shop A is situated on the high street and shop B is situated in a shopping precinct.

The owner is in financial difficulties and wants to close one of the shops. She has asked you to prepare a report using the data to help her decide which shop to close.

Cust spend A		Cust spend B
55		72
42		79
41		63
75		19
85		34
12		48
88		38
17		89
37		8
24		77
75		49
47		98
93		44
32		10
82		91
31		50
22		15
48		76
47		52
95		93
95		63
32		60
30		72
14		33
53		48
48		60
93		93
23		39
19		48
67		100

You need to decide which statistics would be useful and give a fair picture of the long term financial viability of the shops and hence recommend which shop should be closed.

And finally...

They say a picture paints a thousand words, so how many words does a boxplot paint?