

### Question 1

The variable **HealthStatus** indicates the self-reported health status of each of the Californian adults. Using SPSS, summarise the **HealthStatus** variable and write a paragraph explaining the key features of the data observed in the output *in the style presented in the course materials*.

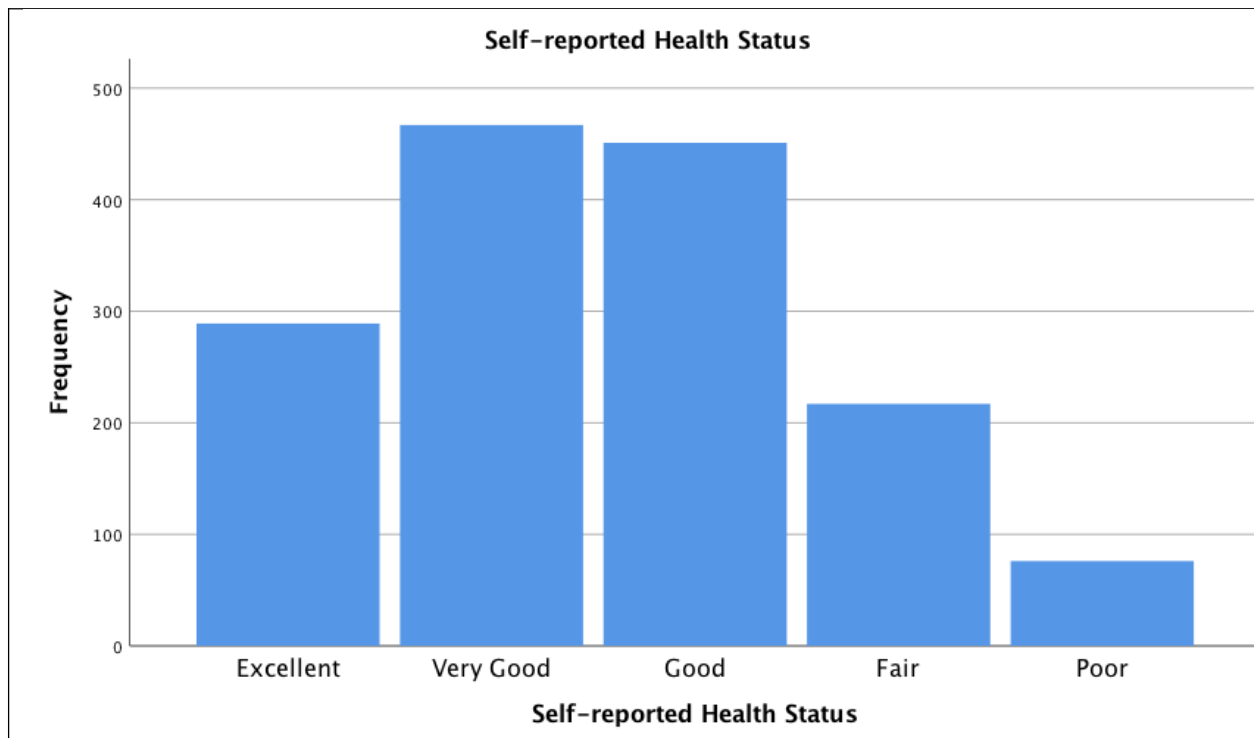
The distribution of self-reported health for a sample of 1500 Californian adults is displayed in *figure 2*. The most typical reported status was very good (31.1%) although good was also high at 30.1%. A small percentage recorded their health status as poor (5.1%)

Include your SPSS output for this question here

#### *Self-reported Health Status*

		<i>Frequency</i>	<i>Percent</i>	<i>Valid Percent</i>	<i>Cumulative Percent</i>
<i>Valid</i>	<i>Excellent</i>	289	19.3	19.3	19.3
	<i>Very Good</i>	467	31.1	31.1	50.4
	<i>Good</i>	451	30.1	30.1	80.5
	<i>Fair</i>	217	14.5	14.5	94.9
	<i>Poor</i>	76	5.1	5.1	100.0
<i>Total</i>		1500.	100.0	100.0	

Figure 1 Distribution of self-reported health status



*Figure 2 Distribution of self-reported health status.*

### Question 2

The variable **Cigarettes** gives an indication of number of cigarettes each adult reported smoking in the previous day. Using SPSS, summarise the **Cigarettes** variable and write a paragraph explaining the key features of the data observed in the output *in the style presented in the course materials*.

The distribution of the number of cigarettes smoked the previous day for a random sample of 1500 is displayed in figure 3. The distribution is positively skewed, with the average cigarettes smoked the previous day being 11.26 cigarettes, standard error of 0.523 cigarettes. Typically, the average cigarettes smoked the previous day fell between 4 and 17 with half the cigarettes smoked the previous day falling within this range.

Figure 1

		Percentiles						
		5	10	25	50	75	90	95
Weighted Average(Definition 1)	Number of cigarettes smoked [previous day]	2.00	2.00	4.00	10.00	17.00	22.60	24.00
Tukey's Hinges	Number of cigarettes smoked [previous day]			4.00	10.00	17.00		
Std. Deviation		7.450						
Minimum		0						
Maximum		26						
Range		26						
Interquartile Range		13						
Skewness		.318						
Kurtosis		-1.073						

Figure 2

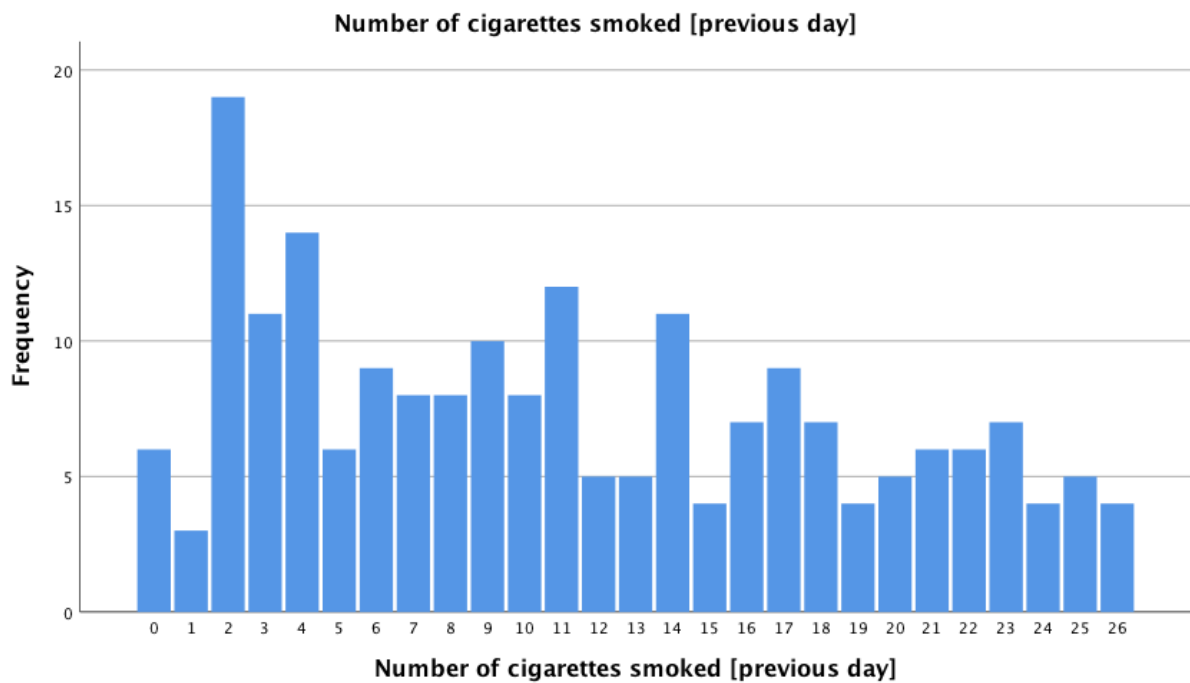


Figure 3

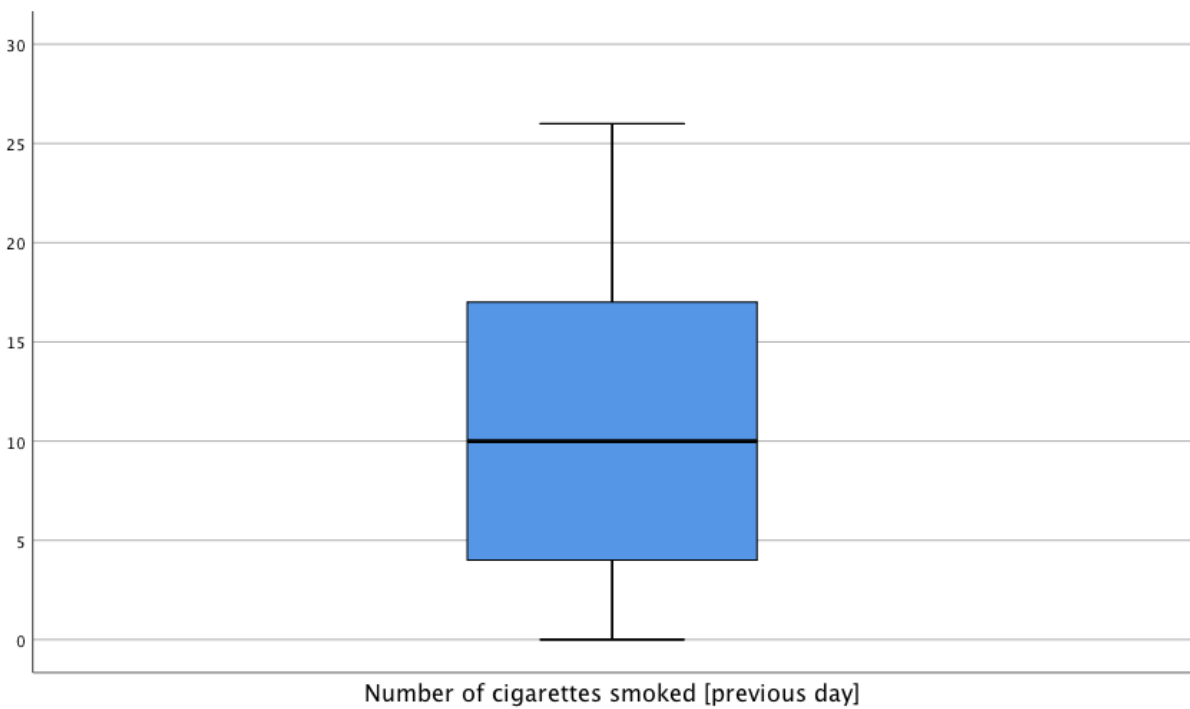
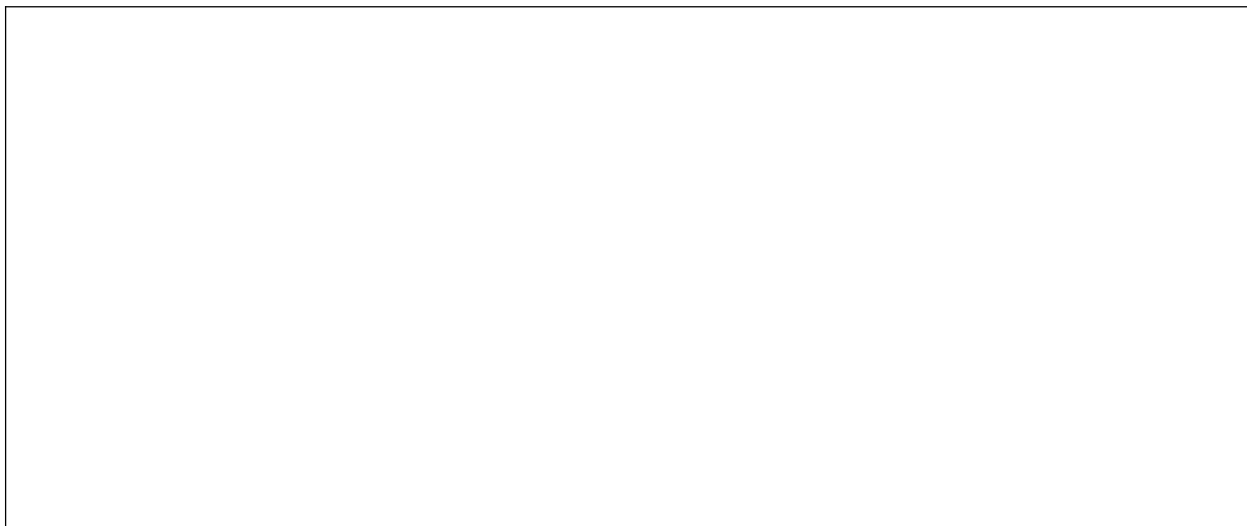


Figure 4



### Question 3

The variable **WalkLeisure** gives an indication of the time each participant spent walking for leisure. This variable was measured in minutes spent walking in the previous week. Using SPSS, produce the relevant output to summarise the **WalkLeisure** variable and write a paragraph explaining the key features of the data observed in the output *in the style presented in the course materials*

The distribution of the number of minutes per day walking for leisure in a sample of 1500 respondents is displayed in figure 3. The distribution is positively skewed with 50% of the minutes walked for leisure in the previous week being between 15 and 40 minutes per day. Some respondents reported rather high number of minutes, of over 70 minutes, the highest being 420 Minutes per day for leisure.

Include your SPSS output for this question here

### Descriptives

			Statistic	Std. Error
Time walked for leisure [mins in previous week]	Mean		33.74	1.161
	95% Confidence Interval for Mean	Lower Bound	31.46	
		Upper Bound	36.02	
	5% Trimmed Mean		28.68	
	Median		25.00	
	Variance		1256.089	
	Std. Deviation		35.441	
	Minimum		5	
	Maximum		420	
	Range		415	
	Interquartile Range		25	
	Skewness		5.016	.080
	Kurtosis		37.198	.160

Percentiles

		Percentiles						
		5	10	25	50	75	90	95
Weighted Average(Definition 1)	Time walked for leisure [mins in previous week]	10.00	10.00	15.00	25.00	40.00	60.00	90.00
Tukey's Hinges	Time walked for leisure [mins in previous week]			15.00	25.00	40.00		

Figure 2

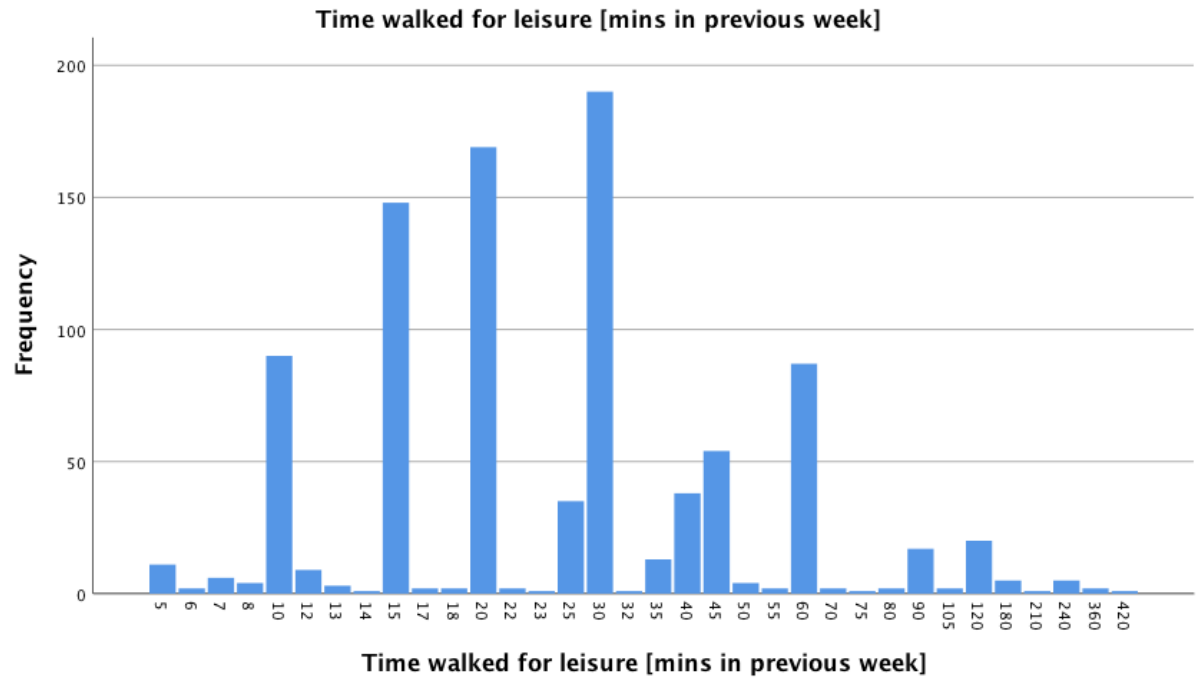


Figure 3

**Question 4 [does not require SPSS]**

Dale is a psychology student who enters two competitions in which puzzles need to be solved as quickly as possible [by correctly fitting all the pieces of a gadget together in each case]. In Competition A, Dale takes 80 seconds to complete the puzzle and in Competition B he takes 84 seconds. Overall completion times for participants in Competition A are normally distributed with a mean of  $\mu = 85$  seconds and a standard deviation  $\sigma = 8$  seconds.

Completion times for participants in Competition B are normally distributed with a mean  $\mu = 90$  seconds and a standard deviation  $\sigma = 5$  seconds.

In which competition was Dale's performance better, relative to others who took part in that competition?

Justify your answer, quoting the relevant statistics as part of your explanation.

This can be determined by the relevant z-scores.

The formula for this result is score – mean divided by the standard deviation

Z-score for competition A time is -0.625. The Z-score for competition B time is -1.2. The aim is to get as close to the mean as possible. The score that was closest to the mean is Competition A with a Z score of -0.625 and as such was the best performance.



**Question 5 [does not require SPSS]**

Australian Consumer Law has mandatory packaging standards relating to biodegradable / compostable products. The Bioware Company manufactures coffee cups and lids, and in order to display a progressive impression to consumers, they have ceased producing plastic coffee cup lids and have replaced these with the production of compostable lids. All coffee cup lids require a hole punched in them for the transfer of air. Despite the use of very accurate equipment in the production of coffee cup lids, there are always some lids in each batch that do not meet the Australian standards, where dents rather than holes are pierced in the lids.

The equipment at the Bioware factory has recently been replaced and the production line manager wants to check that it is performing well. If less than 95% of the compostable coffee cup lids produced meet the Australian standards, the equipment will need to be recalibrated, which involves a lengthy and expensive process. The manager takes a random sample of 1000 compostable coffee cup lids produced during July to investigate the proportion of lids that meet the Australian standards.

- a. What is the population we can draw conclusions about in this study?

*The manager takes a random sample of 1000 coffee cup lids from the new production line. In this instance 1000 is the population.*

To answer questions (b) to (d), consider the sampling distribution shown in *Figure 1*.

- b. What does the highlighted section of the distribution in *Figure 1* represent?

The highlighted column represents the sample mean.

- c. The random sample of 1000 lids taken by the production line manager turned out to have a proportion of 0.929. Does this sample look like it belongs to the sampling distribution displayed in *Figure 1*? Justify your answer.

The random samples of 1000 lids has a proportion of 0.929. This happens to appear in the left hand tail end being in the unlikely region of less than 2.5%.

- d. Given that the sample was randomly selected from all compostable coffee cup lids manufactured by the Bioware Company in July, and given that the sample proportion was measured accurately, what conclusion can we reach from part (c)?

As the proportion is in the left hand tail being less than 2.5% it can be concluded that the new compostable lids produced do not meet the Australian standards resulting in having to recalibrate the equipment