Investigating and Comparing Data Distributions

Question:

The parallel boxplots in Figure 1 have been constructed to compare and contrast the distribution of *BMI* for males and females in this sample.

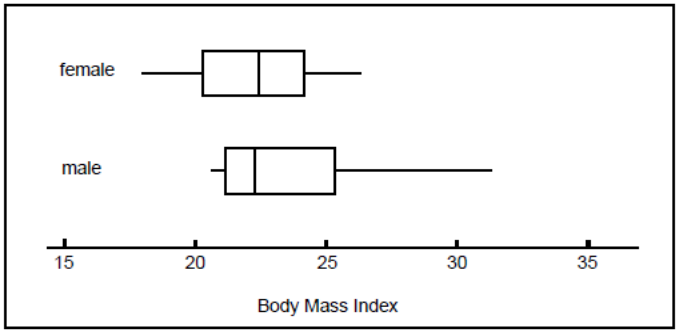

Figure 1. Parallel boxplots comparing *BMI* for males and females

a. Use the parallel boxplots to identify and name two **similar** properties of the *BMI* distributions for males and females.
b. Use the parallel boxplots to compare and contrast the *BMI* distributions for males and females
c. The median *BMI* for males is 22.5, the mean BMI for males is 23.9. Of the **mean** or **median**, which measure gives a better indication of the typical *BMI* for males? Explain your answer.

2+ 2 +2 = 6 marks

Step	Method/ Hint	Answer	Marks allocation (where applicable)
	PART (a) of the question:		
Step 1	➢ Identify the key elements found in a box plot : **The 5 Number Summary** ➢ State the values for both Male and Female.	APPROXIMATE values	

	Male	Female
Minimum – smallest number	20.5	17
Lower Quartile – the 25the percentile	21	20
Median – the middle number or 50th percentile	22.5	22.5
Upper Quartile – the 75th percentile	25.5	24
Maximum – the largest number	31.5	26

	➢ Display this information in a table.	NO OUTLIERS PRESENT in either boxplot	
Step 2	➢ Identify descriptions that can be made using the box plots: **Shape and Distribution**	APPROXIMATE values Male / Female Skewness: Positively skewed / Approx. Symmetrical Range: 31.5-20.5 = 11 / 26 – 17 = 9 IQR: 25.5-21 = 4.5 / 24-20 = 4	
Step 3	Answer the question- WHAT IS SIMILAR:	Median and the IQR are similar for both the Male and Female BMI data displayed in the Boxplots. Both the medians have a value of 22.5 and the IQR evaluate to 4.5 for male and 4 for female.	1 mark -(Median) 1 mark -(IQR) 2 marks in total- for stating what is similar and what are their values

Step	Method/ Hint	Answer	Marks allocation (where applicable)
	PART (b) of the question:		
Step 1	Identify what is similar and dissimilar about the values and distribution of the box plots to compare and contrast	The minimum BMI value for females is 17 and the maximum is 26 while the minimum BMI value for males is 21 and the maximum is 32. Whilst these are quite different, the range of BMI values for males is only slightly higher than the range of BMI values for females. The female BMI data shows that each quartile is similar in spread so the box plot is approximately symmetrical in distribution whilst the male BMI data shows that the upper 50% of values is more spread than the lower 50%, so the box plot is positively skewed.	1 mark -constrasting values of minimum and maximum and comparing the range 1 mark -comparing each distribution to show contrast in their shape

Step	Method/ Hint	Answer	Marks allocation (where applicable)
	PART (c) of the question:		
Step 1	- Identify the two different measures of central tendencies: **Mean and Median** - Identify the difference in Mean and Median and what affects either calculation for the average ➢ MEDIAN: Middle number, when data values are in ascending order, is not affected by outliers or extreme values ➢ MEAN: Calculated average (all values added and then divided by how many there are). Extreme values and outliers are included within this calculation, which skews the centre.	The median gives the better indication of the typical *BMI* as it is not affected by extreme values as is the mean. The upper 50% of the male BMI values shows a larger spread as compared to the lower 50% of the data.	1 mark – identifying that it is the median 1 mark – explaining that it is due to the inclusion of extreme values

Total = 6 marks