MDDE 602 RESEARCH METHODS IN DISTANCE EDUCATION

CHARITY FAKINLEDE

STUDENT ID: 2918004

ASSIGNMENT 2

Good work with this rewrite. Your explanations reflect a greater understanding of the problems.

20/20
Question 1: Although there are only 30 scores for the variable “i601” (MDDE 601 Final Grade Interval), does the distribution of scores for this variable appear to reflect a normal distribution? Please explain how you determine whether or not the scores reflect a normal distribution.

The distribution of the scores for “i601” (MDDE 601 Final Grade Interval) does not reflect a normal distribution because the three measures of central distribution – mean, median and mode are not equal to each other (mean is 79.40, median is 82.50 and mode is 74). A normal distribution is symmetrical, has a bell-shaped curve and the three measures of distribution (mean, median, mode) are equal to each other or are said to coincide.

“i601” is an unsymmetrical (skewed) distribution and the statistics show a negative skewness of -1.159. Because more of the scores are concentrated at the higher end and a few are found at the lower end of the scale.

In addition, the kurtosis is 1.919, which means there is a peakness of 1.919. If there is a peakness then the distribution cannot be said to be a normal distribution since a normal distribution is bell-shaped curve rounded off at the top.

Yes, kurtosis refers to the shape of the curve. Good rewrite.

Question 2: How does the bar graph for this output differ from the histogram in question 1? Does this bar graph reflect a normal distribution? Is the type of presentation of data important in the communication of findings? Please explain your answers.

The bar graph displayed discrete ungrouped data and percentages of the scores, while the histogram in question one (1) used grouped, interval data to show the frequency of the scores.

The bar graph does not reflect a normal distribution. Bar charts are used for discrete data while histograms are used for interval or ratio data.
The histogram is more effective in communicating the findings of this data than the bar chart. The bar chart did not group the data, it collapsed the data, and ignored some data points therefore making the graphical presentation difficult to understand.

Using different graphical representation with different scaling for the same data can give misleading interpretations or confusing messages.

*Yes, the key is that the data is interval, not discrete, so the bar chart isn’t the tool to use.*

**Question 3:** Explain how the values for the mean, median and mode for the continuous variable MDDE 601 Final Grade Interval relate to each other for this set of scores. For this distribution is the mean, median or mode, the best measure of central tendency? Explain your answer.

The mean, median and mode are measures of central tendency in a normal distribution (symmetrical distribution). The mean is lower than the median, because the mean has been dragged down by the negative scores at the tail. The mode is far less than the mean and the median, therefore it can not be the best measure.

The median is the best measure for a skewed distribution. Negative skewedness means majority of the scores are concentrated at the higher end of the scale.

Also the median will always exist in a number set, it does not change easily, it is more stable and it is not affected by manipulations of extreme scores. Unlike the mean which is more sensitive to manipulation and the removal of extreme numbers as demonstrated in the table below.

<table>
<thead>
<tr>
<th>Manipulations</th>
<th>Mean</th>
<th>Median</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original data</td>
<td>79.4</td>
<td>82.5</td>
<td>74</td>
</tr>
<tr>
<td>Remove one number from each extreme</td>
<td>80.036</td>
<td>82.5</td>
<td>74</td>
</tr>
<tr>
<td>Remove two numbers from each extreme</td>
<td>80.423</td>
<td>82.5</td>
<td>74</td>
</tr>
<tr>
<td>Remove three numbers from each extreme</td>
<td>80.458</td>
<td>82.5</td>
<td>74</td>
</tr>
</tbody>
</table>

*Good answer*
**Question 5(a):** What is the relationship between the mean values and the standard deviation values for each of the 3 variables? Explain your answer. Explain homogeneity and heterogeneity across these distributions.

Standard deviation is the unit measure of dispersion from the centre (mean). Each of the three variables shows some level of dispersion from the mean.

Variable 3 is least spread and more tightly fit, that is, more scores are concentrated within a shorter range of a unit measure of 8.876 scores, variable 2 with a unit measure of 13.161 is the most spread of the three variables while variable 1 has a unit measure of dispersion of 11.227 scores from the mean.

The smallest standard deviation of MDDE603 indicates that final grades in MDDE603 are more similar. With a tighter fit around the mean one sees high and low scores packed within a shorter range than the scores of either MDDE601 or MDDE602 classes.

Homogeneity and heterogeneity can be compared across these distributions. A variable is homogeneous when the standard deviation is closer to zero, thus showing similarity, minimal spread or tighter fit of the scores within a short range. These three variables are heterogeneous to a degree since their standard deviations are not close to zero and their level of homogeneity is determined by their standard deviation (spread) from the mean.

Variable 3 is least heterogeneous or more homogeneous that is, the scores are more similar. Variable 2 is the most heterogeneous or least homogeneous that is, having the least similarity in scores of the three distributions.

**Question 5(b):** Can one determine whether or not the mean values for each of the 3 variables are significantly different from each other as a result of this analysis?
For one to determine the significant difference of the mean values of each of the three variables presented by this data more statistical analysis need to be carried out.

**Question 7**: Do any of the three scattergrams reflect a moderate or strong linear relationship? Discuss the strengths and relationships presented by the three scattergrams.

Scattergram 1 shows a linear relationship, there is a line running from the low left to an upper right corner, but the line is away from the middle, it is skewed towards the right. There is a **strong positive relationship** which means the higher the age the higher the marks scored. There is a high level of precision as the points were not too widely spread away from the line.

Scattergram 2 also shows a form of linear relationship with a positive direction. There is a **moderate** relationship between age and final grade in 602. There is a moderate level of precision as the points were widely scattered away from the line.

Scattergram 3 shows randomness with a vertical straight line parallel to the Y axis. This implies there is **no relationship**, the variables are independent of each other, no particular direction (negative or positive) and the points were widely spread around the vertical line which meant low precision level.

**Very good**

2/2