**Week 2 Quiz**

**LATEST SUBMISSION GRADE**

1.Question 1

Which of the below data sets has the **highest** standard deviation? You do not need to calculate the exact standard deviations to answer this question.

**1 / 1 point**



0, 25, 25, 25, 25, 25, 25



0,1,2,3,4,5,6



0, 100, 200, 300, 400, 500, 600



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

**Correct**

This question refers to the following learning objective(s):

Note that there are three commonly used measures of center and spread:

* center: mean (the arithmetic average), median (the midpoint), mode (the most frequent observation)
* spread: standard deviation (variability around the mean), range (max-min), interquartile range (middle 50% of the distribution)

The dataset with the least repeated observations that are farthest from the center has the most variability, hence the highest standard deviation.

2.Question 2

The distribution of housing prices in a country where 25% of the houses cost below $350,000, 50% of the houses cost below $450,000, 75% of the houses cost below $1,000,000 and there are a meaningful number of houses that cost more than $6,000,000 is most likely

**1 / 1 point**



left skewed



symmetric



uniform



right skewed

**Correct**

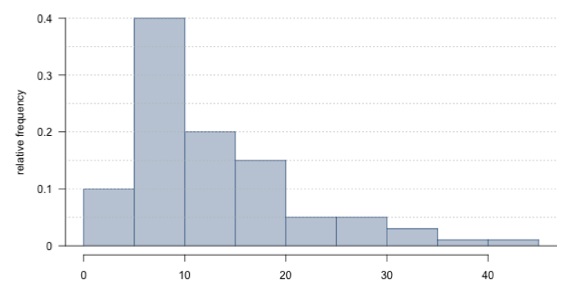
This question refers to the following learning objective(s):

Identify the shape of a distribution as symmetric, right skewed, or left skewed, and unimodal, bimodoal, multimodal, or uniform.

There is a long tail on the right side of the distribution, yielding a right skew.

3.Question 3

Based on the relative frequency histogram below, which of the following statements is supported by the plot?



**1 / 1 point**



There are no outliers in the distribution.



It is not possible to estimate the median without knowing the sample size.



The IQR of the distribution is roughly 10.



The distribution is multimodal.



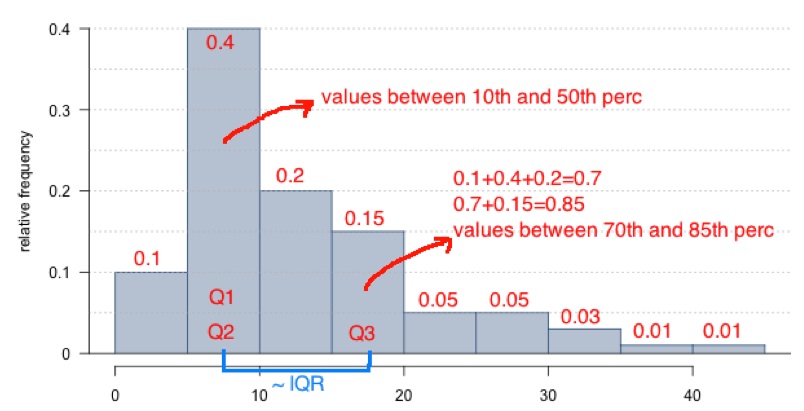
The mean of the distribution is smaller than its median.

**Correct**

This question refers to the following learning objective(s):

Use histograms and box plots to visualize the shape, center, and spread of numerical distributions, and intensity maps for visualizing the spatial distribution of the data.

Using the relative frequency histogram, we can tell that 10% of observations are below 5 (in the first bin), 40% are between 5 and 10, 20% are between 10 and 15, and 15% between 15 and 20. Therefore Q1 is in the second bin (between 5 and 10) and Q3 is in the fourth bin (between 15 and 20). This confirms that the IQR is roughly 10.



4.Question 4

A recent housing survey was conducted to determine the price of a typical home in a city that is mostly middle-class, with one very expensive suburb. The mean price of a house in this city is roughly $650,000. Which of the following statements is **most likely** to be true?

**0 / 1 point**



Majority of houses in this city cost more than $650,000.



Majority of houses in this city cost less than $650,000.



We need to know the standard deviation to answer this question



There are about as many houses in this city that cost more than $650,000 than less than this amount.

**Correct**

This question refers to the following learning objective(s):

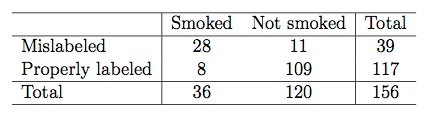
Define a robust statistic (e.g. median, IQR) as a statistics that is not heavily affected by skewness and extreme outliers, and determine when such statistics are more appropriate measures of center and spread compared to other similar statistics.

This is true about the median but not the mean. Since the city is mostly middle-class, with one very expensive suburb, we would expect the distribution to be right skewed, and therefore the mean to be greater than the median.

Option B right Answer

5.Question 5

It is relatively common for fish to be mislabeled in supermarkets and even in restaurants. The table below shows the results of a study where a random sample of 156 fish for sale were collected and genetically tested. The researchers classified each sample as being labeled properly or being mislabeled. What fraction of smoked fish in the sample were mislabeled? Choose the closest answer.



**1 / 1 point**



9%



72%



78%



28%



18%

**Correct**

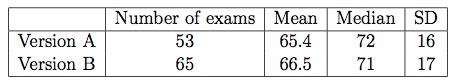
This question refers to the following learning objective(s):

Use contingency tables and segmented bar plots or mosaic plots to assess the relationship between two categorical variables.

Of the 36 smoked fish, 28 are mislabeled: 28/36 ≈ 78%.

6.Question 6

Professors regularly give two versions of an exam. The professor may also provide summary statistics for each version. Suppose the following summary is provided:



A student who took Version A says that he should get an extra point because his exam was harder as evidenced by the lower mean score for Version A, as shown by the mean score. Does the student have a good argument? Pick the best answer below.

**1 / 1 point**



No. The average scores are relatively close when considering the spread of the distributions. The difference might just be due to just chance.



No. The median of Version A is higher.



Yes. Only 53 students took exam Version A while 65 students took exam Version B.



We need to know the shape of the distribution for each version to determine if this argument is valid.



Yes. The difference in the exam scores means that there is a difference in difficulty between the versions.



We need to know the minimum and the maximum for each version to determine if this argument is valid.

**Correct**

This question refers to the following learning objective:

Note that an observed difference in sample statistics suggesting dependence between variables may be due to random chance, and that we need to use hypothesis testing to determine if this difference is too large to be attributed to random chance. Set up null and alternative hypotheses for testing for independence between variables, and evaluate the data support for these hypotheses using a simulation technique.

The mean scores are off by only about 1 point, when the standard deviations of the distributions are much larger than the difference. It is likely that such a small difference (with respect to how variable these distributions are) could be due to chance.