Distributions

LLE – Mathematics and Statistics Skills

1. Consider the following dataset:

$$34, 35, 35, 36, 40, 40, 40, 40, 41, 42, 44, 47$$

(a) Find the mean of the dataset above. (Round to two decimal places if necessary)

> **Solution**
> The mean is calculated as the sum of all values divided by the number of values ($\frac{\sum x}{N}$). $\frac{475}{12} = 39.58$.

(b) Find the median of the dataset above.

> **Solution**
> Since there is an even number of values (12), the median is the average of the two middle values (the 6th and 7th values): $\frac{40+40}{2} = 40$.

(c) Find the range of the dataset above.

> **Solution**
> The range is the largest value minus the smallest value: $47 - 34 = 13$.

2. For the data in Question 1, we will change the final three data points ($42, 44,$ and 47) to $50, 65,$ and 100.

The new dataset is: $34, 35, 35, 36, 40, 40, 40, 40, 41, 50, 65, 100$.

(a) Recalculate the median of the new dataset.

> **Solution**
>
> The median is the average of the 6th and 7th values: $\frac{40+40}{2} = 40$. The median is not affected by these 3 extreme values.

(b) Recalculate the mean of the new dataset. (Round to two decimal places if necessary)

> **Solution**
>
> The new sum is 586. The new mean is $\frac{586}{12} = 48.83$.

(c) What do you notice about the values for the median and mean between Question 1 and Question 2?

> **Solution**
>
> The median remained the same (40), but the mean increased significantly (from 39.58 to 48.83). This demonstrates that the mean is sensitive to extreme values (outliers), while the median is more resistant to them."

(d) What type of skew is shown in the data for Question 2?

○ Negative (left)　　⊗ **Positive (right)**　　○ Symmetrical

> **Solution**
>
> The mean (48.83) is significantly higher than the median (40). This indicates that the extreme values (outliers) are pulling the mean toward the positive (right) end of the distribution, which is a Positive Skew.

3. Consider the following three sets of data:

   Set A: $12, 13, 15, 15, 16, 16, 16, 17, 17, 19, 20, 20$

   Set B: $3, 5, 5, 9, 16, 21, 22, 30, 33$

   Set C: $21, 24, 25, 26, 26, 27, 27, 27, 27, 28, 29, 30$

   (a) For each set of data, calculate the summary statistics of the median, the mean, and the range.

Set A (N=12): Mean = 16.33, Median = 16.0, Range = 8

Set B (N=9): Mean = 18.22, Median = 16.0, Range = 30

Set C (N=12): Mean = 26.42, Median = 27.0, Range = 9

(b) Compare and contrast the three datasets based on their summary statistics.

Central Tendency (Mean/Median): Set A and Set B have similar median/mean values (around 16-18), while Set C has a much higher central tendency (around 26-27).

Dispersion (Range): Set B has the largest range (30), indicating the greatest spread of data. Set A has the smallest range (8), indicating the most homogeneous (least spread) data.

Skew: Set A and Set C have very similar mean and median values, suggesting they are relatively symmetrical. Set B's mean (18.22) is slightly higher than its median (16.0), indicating a slight positive skew due to the high end of the data.

4. Summary statistics are provided for three sets of data:

Set A: Mean = 45.4, Median = 33, SD = 6.8, Skew = 1.6

Set B: Mean = 34.2, Median = 33.5, SD = 2.3, Skew = 0.21

Set C: Mean = 33, Median = 33.6, SD = 0.9, Skew = -0.12

(a) Which dataset had the least dispersion (least spread of data)?

Dispersion is measured by the Standard Deviation (SD). Set C has the smallest SD (0.9), meaning the data points are clustered most tightly around the mean.

(b) Which dataset had the most skew, and in what direction was the skew?

> **Solution**
>
> Set A had the largest absolute skew value (1.6), and since the value is positive, it shows a Positive Skew.

(c) Compare and contrast the three datasets.

> **Solution**
>
> Set A has a much higher mean value than the other two datasets, which have similar means. The medians of all of the datasets are very similar.
>
> Set A is highly skewed to the right, whereas the other two datasets are more symmetrical.
>
> Set C is the least spread of the three datasets, with the lowest standard deviation.

5. Consider the following histograms and statistics:
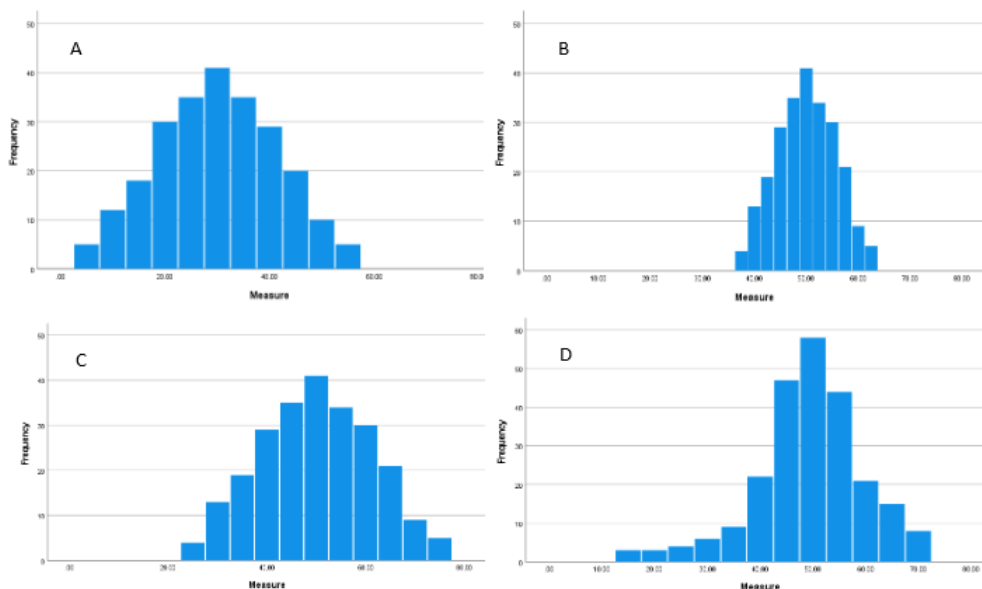
Statistic	Set 1	Set 2	Set 3	Set 4
Mean	49.2	29.9	50	49.9
Median	50	30	50	50
Standard Deviation	10.5	11.4	5.7	11.4
Skew	-0.64	0	0	0

Match each of the histograms to one of the sets of statistics provided above.

Analysis and Matching:

1. Look at Mean/Median/Skew: One set of data has a Mean of 29.9 and a Median of 30. This is significantly lower than the other sets (which are near 50). This must be the histogram that is clustered more to to the left of the others on the horizontal axis, histogram A.

2. Look at Skew: One set has a Skew of -0.64. This indicates a Negative Skew (Skewed Left). This must be the histogram with the long tail extending to the left, histogram D.

3. Look at Standard Deviation (Dispersion): Two remaining sets are symmetrical (Skew=0). One has a low SD (5.7) and one has a high SD (11.4). The low SD set must match the histogram that is the least dispersed. The high SD set must match the histogram that is the **most spread out.

Solution:
* Histogram A Matches Mean = 29.9, Median = 30, Set 2.
* Histogram B Matches Mean = 50, SD = 5.7 Set 3.
* Histogram C: Matches Mean = 49.4, SD = 11.4 Set 4.
* Histogram D: Matches Skew = -0.64 Set 1.