AP Stats Name_____

Univariate Data Analysis Review Answers

Multiple Choice

1. D

The median is NOT always greater than the mean. The median can be greater than, less than, or equal to the mean. The median is only greater than the mean if a distribution is skewed to the right.

2. C

The mean being much higher than the median indicates the distribution is skewed strongly to the right. The median is resistant to skew so it is not increased as much as the mean.

3. E

Both sets are symmetric about 30, so they have the same mean and median. Both sets have the range 50-10=40. Set A has a higher percentage of values further from the mean than does set B, so set A has a greater variance. Histograms give relative frequencies, not actual numbers.

4. E

Here's a problem where the 68-95-99.7 rule is useful. It's possible to solve this using standard normal values, but it's much faster using the rule. To set it up, notices that the 97.5th percentile is the value two standard deviations above the mean. The 16th percentile is the value one standard deviation below the mean.

5. B

The key to this problem is figuring out what approximately is the standard deviation. If 47 is at the 11th percentile, then 47 its value as z is about -1.23. Therefore, 47 is 1.23 standard deviations below the mean. This suggests that the standard deviation is about 6.18. Calculating the x-value for age 57 gives 0.39 which puts the age 57 in the 65.07th percentile.

6. B

This is a straightforward standard normal calculation problem. Find P(x >10) for N(9, 2.5)

7. D

Increasing each score by 20 increases the mean to 520 and leaves the standard deviation unchanged at 100. Then increasing each result by 10 percent increases both the mean and standard deviation by 10 percent to 520 + 0.10(520)=572 and 100+0.10(100)=110.

8. A

The median is at the 50th percentile so that 50% of the data are less than or equal to it and the same amount is greater than or equal to it. The other two responses assume that the distribution is symmetric, so that Q1 is at 260 and Q3 at 280. There is note enough information to make this assumption.

9. D

When the extreme values are removed, the range, standard deviation, and variance will all *decrease*. The median, or middle value, remains the same if one extreme value at each end is removed, and it is *possible* that the mean remains unchanged.

10. E

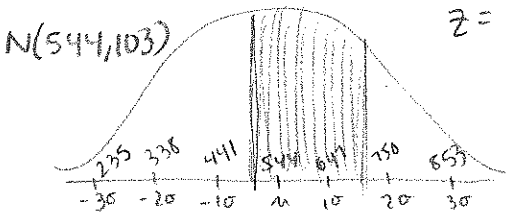
Median and interquartile range are more accurate measure of central tendency and variability when the distribution is skewed, as it would be by potential outliers.

always include (if possible): { picture, work, answer, context }

Free Response

1. The Graduate Record Examinations are widely used to help predict the performance of applicants to graduate schools. The range of possible scores on a GRE is 200-900. They psychology department at a university finds that the scores of its applicants on the quantitative GRE are approximately Normal with a mean of 544 and a standard deviation of 103.

(a) What percent of applicants scored between 500 and 700? $P(500 < x < 700)$



$z = \frac{x-\mu}{\sigma} = \frac{500-544}{103} = -0.4272$ $P(-0.4272 < z < 1.5146)$

$= \frac{700-544}{103} = 1.5146$ $= 0.6004$

or 60.04% of applicants scored between 500 and 700 on the GRE.

(b) What minimum score would a student need in order to score better than 77% of those taking the test? $P(x > ?)$ $\text{invNorm}(0.77) = 0.7388 = z$

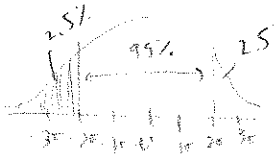
$z = \frac{x-\mu}{\sigma}$ $0.7388 = \frac{x-544}{103}$ $x =$ 620.10 is the minimum score needed to score better than 77% of those students taking the GRE.

2. A company is considering implementing one of two quality control plans for monitoring the weights of automobile batteries that it manufactures. If the manufacturing process is working properly, the battery weights are approximately Normally distributed with a specified mean and standard deviation.

Quality Control Plan A calls for rejecting a battery as defective if its weight falls more than 2 standard deviations below a specified mean. reject if $-2 < z$

Quality Control Plan B calls for rejecting a battery as defective if its weight falls more than 1.5 interquartile ranges below the lower quartile of the specified population. IQR = middle 50%

(a) What proportion of batteries will be rejected by Plan A assuming the manufacturing process is under control?



Using the 68-95-99.7% rule (±1σ ±2σ ±3σ), we know that 5% of data falls outside ±2σ of the mean. We only care about weights that fall below -2σ, so we can divide 5% by 2 to get 2.5%

Approximately 2.5% of batteries will be rejected by plan A, assuming the manufacturing process is under control

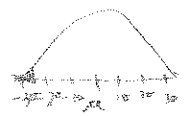
(b) What proportion of batteries will be rejected by Plan B assuming the manufacturing process is under control? the IQR is the middle 50% of data. First, we should find the z-scores that correspond to the IQR boundaries (25%ile and 75%ile)

invNorm(0.25) = -0.6745 invNorm(0.75) = 0.6745 the distance between these z-scores is $0.6745 \cdot 2 = 1.3490 = IQR$.

To reject batteries, their weight must be 1.5IQR less than Q1, or less.

Q1 = 25%ile $\Rightarrow z = -0.6745$ $Q1 - 1.5IQR$

$-0.6745 - 1.5(1.3490) = -2.6980 = z$



$P(z < -2.6980) = 0.003488$ or 0.3488% of batteries will be rejected by plan B, assuming the manufacturing process is under control.

3. A consumer advocate conducted a test of two popular gasoline additives, A and B. There are claims that the use of either of these additives will increase gasoline mileage in cars. A random sample of 30 cars was selected. Each car was filled with gasoline and the cars were run under the same driving conditions until the gas tanks were empty. The distance traveled was recorded for each car.

Additive A was randomly assigned to 15 of the cars and additive B was randomly assigned to the other 15 cars. The gas tank of each car was filled with gasoline and the assigned additive. The cars were again run under the same driving conditions until the tanks were empty. The distance traveled was recorded and the difference in the distance with the additive minus the distance without the additive for each car was calculated.

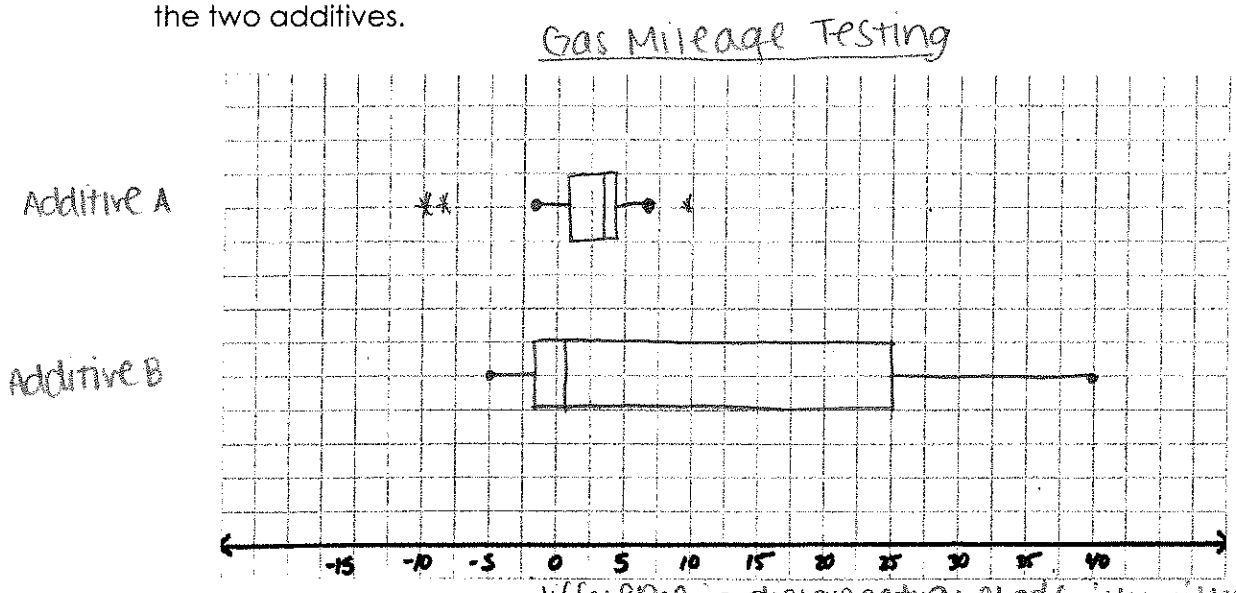
The following table summarizes the calculated differences. Note that negative values indicate less distance was traveled with the additive than without the additive.

Additive	Values below Q1	Q1	Median	Q3	Values above Q3
A	-10, -8, -2	1	3	4	5, 7, 9
B	-5, -3, -3	-2	1	25	35, 37, 40

IQR: 15 | 3 | 4.5 | 27 | 40.5

Outliers for A: $1 - 4.5 = -3.5 \Rightarrow -10, -8$ $4 + 4.5 = 8.5 \Rightarrow 9$ Outliers for B: $-2 - 40.5 = -42.5$ none $25 + 40.5 = 65.5$ none

(a) On the grid below, display parallel boxplots (showing outliers, if any) of the differences of the two additives.



(b) Two ways that the effectiveness of a gasoline additive can be evaluated are by looking at either

- The proportion of cars that have increased gas mileage when the additive is used in those cars
  OR
- The mean increase in gas mileage when the additive is used in those cars

(i) Which additive, A or B, would you recommend if the goal is to increase gas mileage in the highest proportion of cars? Explain your choice. I would recommend Additive A because more than 75% of the cars experienced a positive increase in mileage whereas less than 75% of cars with Additive B experienced a positive increase in gas mileage.

(ii) Which additive, A or B, would you recommend if the goal is to have the highest mean increase in gas mileage? Explain your choice. I would recommend Additive B. The boxplot for B is skewed right which will pull the mean toward the larger values. The mean for B will be substantially greater than 1. The distribution of A has less variability and is skewed left, so the mean will be less than the median of 3.