Name (Last, First):	Student ID:
Lopez Sepulveda,	U59702827

Assignment 4

METCS544A3A4_F2024

Instructions:

1. **This assignment has no specific R programming questions, but you're encouraged to use R to plot and graph the data and calculate relevant statistic summaries.**
2. For answering programming questions, please use Adobe Acrobat to edit the pdf file in two steps **[See Appendix: Example Question and Answer]**:
    a. Copy and paste your R code as text in the box provided (so that your teaching team can run your code);
    b. Screenshot your R console outputs, save them as a .PNG image file, and paste/insert them in the box provided.
    c. Show all work - credit will not be given for code without showing the code in action by including the screenshot of R console outputs.
3. To answer non-programming questions, please type or handwrite your final answers clearly in the boxes. Show all work - credit will not be given for numerical solutions that appear without explanation in the space above the boxes.
4. **[Total 99 pts = 96 + 3 Extra Credit pts]**

Grading Rubric

Each question is worth 3 points and will be graded as follows:

3 points: Correct answer with work shown

2 points: Incorrect answer but attempt shows some understanding (work shown)

1 point: Incorrect answer but an attempt was made (work shown), or **correct answer without explanation (work not shown)**

0 points: Left blank or made little to no effort/work not shown

Reflective Journal [3 pts]

(Copy and paste the link to your live Google doc in the box below)

https://drive.google.com/drive/folders/1_8qcBjQVMfZggF42UYJuHQzMoBcyAy0Q?usp=drive_link

Exploring One Variable Data (96 pts)

1. **(12 pts)** You are given a dataset representing the ages of a group of students in a school. Your task is to calculate various descriptive statistics for this dataset.

   18, 20, 22, 19, 21, 20, 23, 18, 20, 24, 19, 22, 21, 20, 23

   a. Mean (average)
   b. Median
   c. Standard Deviation
   d. Based on your answers above, would you expect the distribution of your data to be skewed or symmetric? Explain why.

Answer:

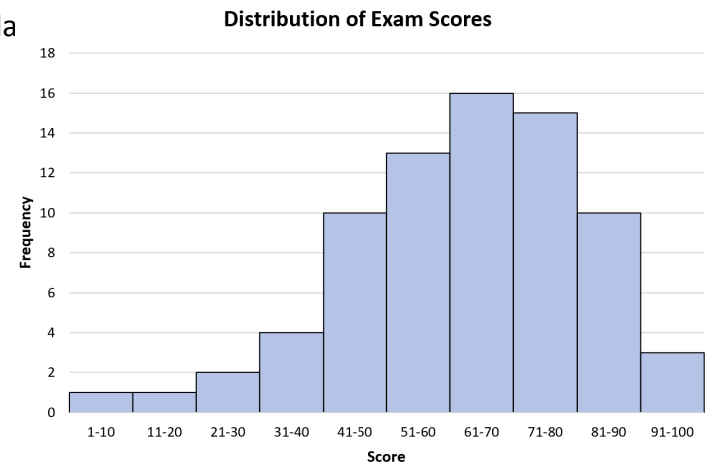
a. 20.66667
b. 20
c. 1.838737
d. I expect it to be symmetric because the Mean and Median are both very similair in range and the standard deviation is equally spaced between the min and max of the dataset

2. **(3 pts)** In a class of 20 students, 19 of them took an exam in class and 1 student took a make-up exam the following day. The professor graded the first batch of 19 exams and found an average score of 84 points with a standard deviation of 9.1 points. The student who took the make-up the following day scored 75 points on the exam.
   What is the new average? (Hint: Think about the formula for the mean and what information we know that we can plug in. You can solve for the information we don't know.)

Answer:

The new average is 83.55 because we multiplied the average by the student count to find the original total value then added the last students score and divided by the new

3. **(9 pts)** Consider the following displa



a. Please identify (1) what type of display is shown, (2) what variable is being displayed and (3) is that variable quantitative or categorical?

b. Describe the shape of the distribution?

c. For center and spread, which measures would you use and why?

Answer:

a. This is a bar graph, The variable being displayed is score frequency based on a range and the score is categorical because it is a range of scores and the frequency is quantative
2. The graph is scewed to the left
3. We would use median and IQR because they are less susceptible to outliers in the graph on the left

4. **(3 pts)** Below are density plots for job satisfaction of three groups all with the same mean (50). Of the three groups (blue, red, or green), which has the highest standard deviation and why?

Answer:

The green curve has the highest standard deviation because it is the widest graph

5. **(12 pts)** Three landmarks of baseball achievement are Ty Cobb's batting average of 0.420 in 1911, Ted Williams's 0.406 in 1941, and George Brett's 0.390 in 1980. These batting averages cannot be compared directly because the distribution of major league batting averages has changed over the years. The distributions are quite symmetric, except for outliers such as Cobb, Williams, and Brett. While the mean batting average has been held roughly constant by rule changes and the balance between hitting and pitching, the standard deviation has dropped over time. Here are the facts:

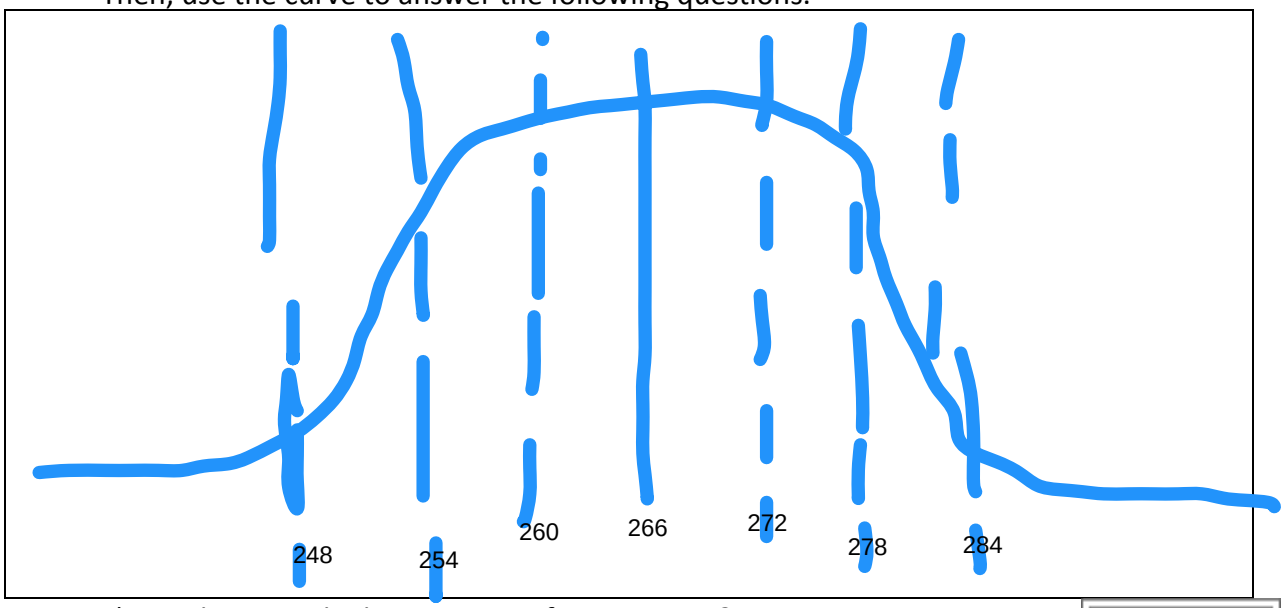
Decade	Mean	Standard Deviation
1910s	0.266	0.0371
1940s	0.267	0.0326
1970s	0.261	0.0317

Find the standardized scores for Cobb, Williams, and Brett. Who was the best hitter?

Answer:

Ted Williams had a higher z-score then the rest at 4.26 so he was farther ahead on the curve allowing him to be considered the best hitter. The other scores are Ty Cobb with 4.15 and George Brett with 4.07. We found the z score by Subtracting the mean from the batting average in each year and dividing it by the standard deviation.

6. **(18 pts)** The length of human pregnancies from conception to birth varies on an approximate normal distribution with mean 266 days and a standard deviation of 6 days. Draw a normal curve below and mark three standard deviations out on either side. Then, use the curve to answer the following questions.



a) How long are the longest 16% of pregnancies?

272 days

b) How short are the shortest 16% of pregnancies?

260 days

c) Between how many days does the middle 95% of all pregnancies fall?

254-278 days

d) Between how many days does the middle 99.7% of all pregnancies fall?

248-284 days

e) A pregnancy of 254 days corresponds to what percentile of pregnancies?

shorter than 97.72% of

7. **(6 pts)** The scores of an introductory statistics class at Simmons University is normally distributed with a mean of 72 and a standard deviation of 10.

(a) If a student has a z-score of 1.4, what score does she have?

(b) If a student has a z-score of -1.23, what score does she have?

Answer:

a. (1.4*10) + 72 = 86

8. **(12 pts)** A college professor plans to give a test and to curve the results. The raw test scores are normally distributed with a mean of 58 and a standard deviation of 9. She wants the grades to be divided up as follows:

D	C	B	A
Bottom 25%	50^{th} percentile or lower before a D	Between an A and a C	Top 10%

What will be the raw score averages she gives to each of the letter grades?

Answer:

Between 0 and	between 51.93 and	between 58 and 69.52	Greater

9. **(3 pts)** with z-scores above 2.5 on an IQ test are sometimes classified as geniuses. If IQ scores have a mean of 100 and a standard deviation of 15 points, what IQ score do you need to be considered a genius?

Answer:

(2.5*15) + 100 = 137.5

137.5

10. **(6 pts)** An incoming freshman took her college's placement exams in French and mathematics. In French, she scored 82 and in math 86. The overall results on the French exam had a mean of 72 and a standard deviation of 8, while the mean math score was 68, with a standard deviation of 12. On which exam did she do better compared with the other freshmen?

Answer:

z-score French = (82-72)/8 = 10/8 = 5/4 = 1.25

She scored better in Math because her z-score is greater

11. (12 pts) A company's customer service hotline handles many calls relating to orders, refunds, and other issues. The company's records indicate that the median length of calls to the hotline is 4.4 minutes with an IQR of 2.3 minutes.
Hint: These questions are asking you what happens when shifting and scaling the data.

a) If the company were to describe the duration of these calls in seconds instead of minutes, what would the median and IQR be?

Answer:

You just multiply both by 60

The median call length in seconds is 264 seconds.

b) In an effort to speed up the customer service process, the company decided to streamline the series of push button menus customers must navigate, cutting the time by 24 seconds. What will the median and IQR of the length of hotline calls become?

Answer:

The median changed but IQR did not

The new median length of hotline calls is 240 seconds.
The IQR remains 138 seconds.

THE END