



Review Test Submission: Quiz 1B

| | |
|---------------|--|
| User | Kismat Khatri |
| Course | 202203_Introduction to Data Science_DATS_6101_11 |
| Test | Quiz 1B |
| Started | 10/28/22 5:23 PM |
| Submitted | 10/28/22 6:19 PM |
| Due Date | 10/28/22 11:59 PM |
| Status | Completed |
| Attempt Score | 23.5 out of 25 points |
| Time Elapsed | 56 minutes out of 1 hour |
| Instructions | <ul style="list-style-type: none"> This is a closed-book/closed-notes exam. You will not be permitted to consult course materials or talk with peers. After opening the quiz, you will have 60 minutes to complete and submit it. Note that the timer cannot be paused Be sure to click Submit at the bottom of the quiz when you are finished. |

Results Displayed All Answers, Submitted Answers, Correct Answers, Feedback, Incorrectly Answered Questions

Question 1

2 out of 2 points



After importing a dataset into R as "mlb", you found that the fourth and fifth column names are rather inconveniently set as "Height.inches." and "Weight.pounds."

Write the codes to rename them to "height" and "weight" in the dataframe respectively.

Selected Answer: `colnames(mlb)[4:5]=c("height","weight")`

Correct Answer: [None]

Response Feedback: [None Given]

Question 2

1.5 out of 2 points



What is the standard deviation (either describe it or give correct formula)?

Selected Answer: The average degree of variability in your dataset is represented by the standard deviation. It's the square root of variance. It reveals the average deviation of each score from the mean. The data set is more unpredictable the higher the standard deviation. The formula for standard deviation is given below:

$$\sqrt{\frac{\sum (X - \bar{x})^2}{n - 1}} = s$$

s= sample standard deviation

sample mean = \bar{x}

number of values in the sample = n

\sum = sum of

X= each value

Correct Answer: [None]

Response Feedback: denominator is n

Question 3

1 out of 1 points



What is the correct description of the above distribution?

Selected Answer: ☒ A. Left skewed

- Answers:
- ☒ A. Left skewed
 - ☐ B. Right skewed
 - ☐ C. Chi Squared
 - ☐ D. Reverser J-shaped
 - ☐ E. Gaussian
 - ☐ F. Triangular
 - ☐ G. J-shaped

Question 4

1 out of 1 points



For the distribution above, which statement below about the mean and median is true?

- Selected Answer: ☒ A. Mean < Median
- Answers:
- ☒ A. Mean < Median
 - ☐ B. Not enough information to determine
 - ☐ C. Mean = Median
 - ☐ D. Median < Mean

Question 5

2 out of 3 points



What is the difference between covariance and correlation?

Selected Answer: covariance assesses the degree to which movement in one variable predicts movement in a related one whereas correlation analysis is used to assess the strength of the association (linear relationship) between two variables.

A change in one variable reflects a change in the other, which is referred to statistically as covariance. Covariance describes a systematic link between two random variables. A negative number for the covariance value indicates a negative association, whereas a positive value indicates a positive link. The covariance value can vary from - infinity to + infinity.

In statistics, correlation is a metric that assesses how closely two or more random variables follow one another. The variables are said to be correlated when, during the study of two variables, an analogous movement of one variable reciprocates the movement of the other variable in some manner.

Correct Answer: [None]

Response

Feedback: Both are one and the same, correlation is scaled version of covariance (correlation is unit free, can be directly used in comparisons like $\text{corr}(X1,Y1) > \text{corr}(X2,Y2)$, but it cannot be done directly with covariance)

Question 6

3 out of 3 points



Briefly explain what is "inferential statistics". Give an example.

Provide the answer in your own words (don't copy from other sources).

Selected Answer: A statistical field known as inferential statistics is one that employs analytical tools to make inferences about a population by analyzing samples taken at random about a population are what inferential statistics are intended to do. Measures of inferential statistics are t-test, z test, linear regression, etc. Inferential statistics (such as the sample mean) from the sample data to draw conclusions about the population parameter (e.g., the population mean). For example: Suppose it is known what the average grade of 50 students in a certain nation is. Inferential statistics can be used to estimate the country's average using this sample data.

Correct Answer: ☒

From the lecture, "Making inferences here specifically means using a sample of size n to infer the statistics of the entire population. In particular, using the sample to estimate the population average (μ)."

Defintion: 1.5 point

Example 1.5 point

Response [None Given]
Feedback:

Question 7

2 out of 2 points



The interquartile range.....

Choose all that are true.

Selected Answers: ☒ B. boxplot for visualizoation

☒ B.

☒ D. IQR= Q3 -Q1

Answers: ☐ A. represents the spread from Q1 to Q4

☐ boxplot for visualizoation

☒ B.

☐ bargraph for visualizoation

☐ C.

☒ D. IQR= Q3 -Q1

☐ includes the left quartile value

☐ E.

☐ includes the right quartile value

☐ F.

Question 8

2 out of 2 points



What is the z (critical) value that corresponds to 80% confidence level?

Selected Answer: ☒ C. 1.28

Answers: ☐ A. 0.84

☐ B. 1

☒ C. 1.28

☐ D. 2

☐ E. 2.57

☐ F. There is not enough information to answer the question

Question 9

1 out of 1 points



In order to suppress the R code chunk and the output on the knitted file, what attribute we can set in R code chunk in R Markdown?

Select one

Selected Answer: ☒ D. include = FALSE

Answers: ☐ A. print = FALSE

☐ B. echo = FALSE

☐ C. suppress = TRUE

☒ D. include = FALSE

☐ E. hide = TRUE

☐ F. eval = FALSE

Question 10

5 out of 5 points



Let's say we want to estimate the mean quiz score of students in a specific population. To do that, we collected a random sample size of 64 students with the following information:

- Sample mean score = 75
- Sample standard deviation $s = 15.5$

Consider, confidence level 95%

What will be the confidence interval in this case? Also, interpret the outcome

(For this question, you can look at the formula in case you can't remember)

Selected Answer: Sample mean score = 75
Sample standard deviation s = 15.5
sample size = 64

confidence level 95% = $75 \pm 1.96 * (15.5/\sqrt{64})$
confidence level 95% = (71.2025, 78.7975)

We can be 95% confident that the true mean weight of a quiz score in this population is between 71.2025 score and 78.7975 score.
Another way of saying the same thing is that there is only a 5% chance that the true population mean lies outside of the 95% confidence interval.
That is, there's only a 5% chance that the true population mean weight of a quiz is greater than 78.7975 score or less than 71.2025 score.

Correct Answer: $75 \pm 1.96 * (15.5/\sqrt{64})$
 $75 - 1.96 * (15.5/\sqrt{64}) = 71.20$
 $75 + 1.96 * (15.5/\sqrt{64}) = 78.79$

✔ We can be 95% confident that the true mean score of a student in this population is between 71.20 and 78.79.

Response [None Given]
Feedback:

Question 11

2 out of 2 points



T interval is good for situations where

Selected Answer: ✔ A. the sample size is small and the population standard deviation is unknown

Answers: ✔ A. the sample size is small and the population standard deviation is unknown
B. the sample size is small
C. the sample size is large and the population standard deviation is unknown
D. the sample size is small and the population standard deviation is known
E. the sample size is large and the population standard deviation is known
F. the population standard deviation is unknown

Question 12

1 out of 1 points



Sigma, σ for a sample within the population

Selected Answer: ✔ False

Answers: True
✔ False

Sunday, September 24, 2023 12:37:43 AM EDT

← OK