Statistics Worksheet 1 Answers

Multiple Choice Questions

Q1: Bernoulli random variables take (only) the values 1 and 0.

A. a) True

Q2: Which of the following theorem states that the distribution of averages of iid variables, properly normalized, becomes that of a standard normal as the sample size increases?

A. a) Central Limit Theorem

Q3: Which of the following is incorrect with respect to use of Poisson distribution?

A. b) Modeling bounded count data

Q4: Point out the correct statement.

A. c) The square of a standard normal random variable follows what is called chi-squared distribution

Q5: _____ random variables are used to model rates

A. c) Poisson

Q6: 10. Usually replacing the standard error by its estimated value does change the CLT.

A. b) False

Q7: 1. Which of the following testing is concerned with making decisions using data?

A. b) Hypothesis

Q8: 4. Normalized data are centered at_____and have units equal to standard deviations of the original data.

A. a) 0

Q9: Which of the following statement is incorrect with respect to outliers?

A. c) Outliers cannot conform to the regression relationship

Subjective Questions

Q10: What do you understand by the term Normal Distribution?

Normal distribution, also known as Gaussian distribution, is a probability distribution that is symmetric about the mean.
It shows that data near the mean are more frequent in occurrence than data far from the mean. In graph form, normal distribution
will appear as a bell curve.

Q11: How do you handle missing data? What imputation techniques do you recommend?

Handling missing data is crucial in data analysis to maintain the integrity of the dataset. There are several techniques to handle
missing data, including:
1. Deletion Methods: Removing rows or columns with missing values.
2. Imputation Methods:
   - Mean/Median Imputation: Replacing missing values with the mean or median of the column.
   - Mode Imputation: Replacing missing values with the most frequent value (for categorical data).
   - Predictive Imputation: Using algorithms like regression, k-NN, or other machine learning models to predict and replace
     missing values.
3. Advanced Methods:
   - Multiple Imputation: Replacing missing values multiple times to create several complete datasets and then combining results.
   - Interpolation and Extrapolation: Using mathematical methods to estimate missing values.

Q12: What is A/B testing?

A/B testing, also known as split testing, is an experiment where two or more variants (A and B) of a page, feature, or product
are compared to determine which one performs better. It involves randomly showing these variants to different users and using
statistical analysis to compare performance metrics such as conversion rates, click-through rates, or other key performance indicators.

Q13: Is mean imputation of missing data acceptable practice?

Mean imputation of missing data is a common practice but it has its drawbacks. While it is simple and maintains the sample size,
it can distort the data distribution, underestimate the variability, and potentially introduce bias. It is often better to use
more sophisticated imputation methods that take into account the relationships between variables.

Q14: What is linear regression in statistics?

Linear regression is a statistical method used to model the relationship between a dependent variable and one or more independent
variables. The model assumes that the relationship between the variables is linear and can be represented by a straight line,
described by the equation Y = b0 + b1X + e, where Y is the dependent variable, X is the independent variable, b0 is the intercept,
b1 is the slope, and e is the error term.

Q15: What are the various branches of statistics?

The various branches of statistics include:
1. Descriptive Statistics: Deals with the description and summarization of data through numerical calculations, graphs, and tables.
2. Inferential Statistics: Involves making predictions or inferences about a population based on a sample of data.
3. Probability Theory: Studies the likelihood of occurrence of events.
4. Biostatistics: Applies statistical methods to biological and health sciences.
5. Econometrics: Applies statistical methods to economic data.
6. Psychometrics: Concerned with the theory and techniques of psychological measurement.
7. Social Statistics: Application of statistical methods to social science data.