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

Answer:True

1. 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?

Answer:Central Limit Theorem

1. Which of the following is incorrect with respect to use of Poisson distribution?

Answer: Modeling bounded count data

1. Point out the correct statement.

Answer: The exponent of a normally distributed random variables follows what is called the log- normal distribution

1. \_\_\_\_\_\_ random variables are used to model rates.

Answer: Poisson

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

Answer: False

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

Answer: Hypothesis

1. 4. Normalized data are centered at\_\_\_\_\_\_and have units equal to standard deviations of the original data.

Answer: 5

1. Which of the following statement is incorrect with respect to outliers?

Answer: Outliers cannot conform to the regression relationship.

1. What do you understand by the term Normal Distribution?

Answer: Normal distribution, also known as the Gaussian distribution, is a probability distribution that is symmetric about the mean, showing 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.

1. How do you handle missing data? What imputation techniques do you recommend?

Answer: We can use none. What we should do instead is either the Bayesian approach of simply treating the missing data as latent variables and thus integrate them out, or the more commonly used practice of *multiple imputation*. (The former is effective but impractical for arbitrarily large dimensional problems; in certain settings the latter is simply an approximation of the former anyways.)

*Imputation*, which is the process of filling in the missing data, should account for the uncertainty you have in the values you are imputing. Thus it does not make sense to plug in single values and then to treat the inference as if it were on complete data. This is very similar to the difference between the estimated likelihood and the profile likelihood.

This leads to the process of imputing the missing data multiple times in order to obtain multiple "complete data sets". We can then obtain estimates from each complete data set and use a combining rule that accounts for the variability in the individual imputations.

1. What is A/B testing?

Answer: A/B testing is a basic randomized control experiment. It is a way to compare the two versions of a variable to find out which performs better in a controlled environment.

1. Is mean imputation of missing data acceptable practice?

Answer: Bad practice in general If just estimating means: mean imputation preserves the mean of the observed data Leads to an underestimate of the standard deviation Distorts relationships between variables by “pulling” estimates of the correlation toward zero.

1. What is linear regression in statistics?

Answer: Linear regression is a linear approach for modelling the relationship between a scalar response and one or more explanatory variables (also known as dependent and independent variables).

1. What are the various branches of statistics?

Answer: The two main branches of statistics are descriptive statistics and inferential statistics. Both of these are employed in scientific analysis of data and both are equally important for the student of statistics.