

# STATIC WORKSHEET

Choose only one correct option (Q1 to Q9)

Q1 Ans – True (option A)

Q2 Ans – Central limit theorem (option A)

Q3 Ans – Modeling bounded count data (option B)

Q4 Ans – All of the mentioned (option D)

Q5 Ans – Poisson (option C)

Q6 Ans – True (option A)

Q7 Ans – Hypothesis (option B)

Q8 Ans – 0 (option A)

Q9 Ans – Outliers can not conform to the regression relationship (option C)

Answer the following question Q10 to Q15

Q10 – what do you understand by the term normal distribution ?

Ans- 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 frequent in occurrence than data far from the mean, in graph form, normal distribution will appear as a bell curve.

# STATIC WORKSHEET

- In a normal distribution the mean is zero and the standard deviation is 1. It has zero skew and a kurtosis of 3.
- Normal distributions are symmetrical, but not all symmetrical distributions are normal.
- In reality, most pricing distributions are not perfectly normal.

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

Ans – There are a lot of techniques to treat missing value. I am trying to think what is the best way to organize some of the most commonly used methods, if you use SAS to implement it -

- **Ignore the records with missing values.**

Many tools ignore records with missing values. When the percentage of records with missing values is small, we could ignore those records.

- **Substitute a value such as mean.**

When the percentage is large and also when it makes sense to do something to avoid bias modeling results, substituting a value (e.g. mean, median) is a commonly used way. But this method could cause bias distribution and variance. That's where the following imputation methods come in.

- **Predict missing values.**

Depending on the type of the imputed variable (i.e. continuous, ordinal, nominal) and missing data pattern (i.e. monotone, non-monotone), below are a few commonly used models. If you plan to do it in SAS, there are SAS codes that you can write to identify the missing data pattern.

- o Logistic Regression

# STATIC WORKSHEET

- o Discriminant Regression
- o Markov Chain Monte Carlo (MCMC)
- **Predict missing values - Multiple Imputation.** Although there are pros & cons, MI is considered to be superior to single imputation, and it better measures the uncertainty of the missing values.

In addition, there are a few required **statistical assumptions** for multiple imputation:

1. Whether the data is missing at random (MAR).
2. Multivariate normal distribution, for some of the modeling methods mentioned above (e.g. regression, MCMC)

**Q12 - What is A/B testing?**

**Ans - A/B testing** is one of the most popular controlled experiments used to optimize web marketing strategies. It allows decision makers to choose the best design for a website by looking at the analytics results obtained with two possible alternatives A and B. To understand what A/B testing is about, let's consider two alternative designs: A and B. Visitors of a website are randomly served with one of the two. Then, data about their activity is collected by web analytics. Given this data, one can apply statistical tests to determine whether one of the two designs has better efficacy.

Now, different kinds of metrics can be used to measure a website efficacy. With **discrete metrics**, also called **binomial metrics**, only the two values **0** and **1** are possible. The following are examples of popular discrete metrics.

# STATIC WORKSHEET

Q13 - Is mean imputation of missing data acceptable practice?

Ans - **True**, imputing the mean preserves the mean of the observed data. So if the data are missing completely at random, the estimate of the mean remains unbiased. Since most research studies are interested in the relationship among variables, mean imputation is not a good solution.

Q14 - what is linear regression statics ?

Ans - Linear regression strives to shows the relationship between two variables by applying a linear equation to observed data. One variable is supposed to be an independent variable, and the other is to be a dependent variable. For example, the weight of the person is linearly related to his height.

Q15- What are the various branches of statistics?

Ans - there are 9 types of branches of statistics.

1. Econometric
2. Actuarial
3. Psychometrics
4. Physics statistics
5. Population statistics
6. Official statistics
7. Biostatistics
8. Industrial statistics
9. Computing statistics

# STATIC WORKSHEET