Statistics worksheet

1) a) true
2) A) Central Limit Theorem
3) B) Modeling bounded count data
4) –
5) C) Poisson distribution
6) B) false
7) B) Hypothesis
8) A) 0
9) c) Outliers cannot conform to the regression relationship
10) 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.
11) Missing data can be dealt with a variety of ways. I believe the most common reaction is to ignore it. Choosing to make no decision, on the other hand, indicates that your statistical programme will make the decision for you. Your application will remove things in a list wise sequence most of the time. Depending on why and how much data is gone, list wise deletion may or may not be good idea.

    Another common strategy among those who pay attention is imputation. Imputation is the process of substituting an estimate for missing values and

analysing the entire data set as if the imputed values were the true observed values. The techniques are

Mean imputation

Substitution

Hot deck imputation

Cold deck imputation

Regression imputation

Stochastic regression imputation

Interpolation and extrapolation

Single or Multiple Imputation

12) A/B testing is basically statistical hypothesis testing, or, in other words statistical inference. It is an analytical method for making decisions that estimates population parameters based on sample statistics.

13) True, imputing the mean preserve the mean of observed data. So if the data ara missing completely at random , the estimate of the mean remain unbiased

14) Linear regression is a linear approach for modelling the relationship between a scalar and one or more explanatory variables.

15) Three real branches of statistics

1) **Data collection**
2) **Descriptive statistics**
3) **Inferential statistics**