Elton Sampson

PFA Worksheet Set 1

Statistics Worksheet

---

1. True
2. a) Central Limit Theorem
3. c) Modelling contingency tables
4. a) The exponent of a normally distributed random variables follows what is called the log-normal distribution
5. c) Poisson
6. a) True
7. a) Probability
8. a) 0
9. d) none of the mentioned
10. By the term Normal Distribution, I understand that the data is distributed in a manner that is expected, follows a regular and symmetric pattern, and can be easily understood, predicted or analysed.
11. I would handle missing data based on the situation. In some instances, deleting the records with the missing data would not significantly alter the integrity and as such this can be done. However, most instances would require some form of imputation since most python libraries do not work with missing data values. Missing data can skew the results of analysis and give unreliable results. The imputation methods i would employ would include, Putting in a missing category, mode imputation, standard deviation imputation.
12. A/B testing is a randomized form of research methodology whereby experiments are done using two variables.
13. Yes, mean imputation methodology is acceptable.
14. Linear regression in statistics can be considered as the line of best fit that summarizes the distribution of data as plotted with the dependent variable on the X-axis and the independent on the Y.
15. The branches of statistics are *Descriptive Statistics* which deals with the presentation and collection of data and Inferential Statistics which deals with the conclusions that can be drawn from a descriptive analysis.