**Statistics for Decision Making, Sep 2018 (DSE-Hyderabad)**

1. Load the data (recruitment\_data.csv) file and perform the following, [7 Marks]
2. Delete all missing value rows and report the reduced data frame
3. Check the variable ‘sales\_quota\_pct’, for its normality, justify with suitable plots
4. Check whether the sales\_quota\_pct vary significantly with respect to the mode of requirement (Referral candidates Vs Applied Online)
5. Check the proportion of Attrition vary significantly or not, with respect to all mode of recruitment
6. Load the data (prima-indians-diabetes.csv) file and do the following, [5 Marks]
7. Assign the following column names for (V1- V9) in the order of following header names, [No. of times Pregnant, glucose concentration, Blood pressure, Skin Thickness, Insulin, Body Mass Index, Diabetes Pedigree Function, Age, Class]
8. Class-0 is attached with Healthy, Class-1 is attached with Diabetes. Check whether Body Mass Index is having a significant effect on Diabetes. Justify with statistical evidence.
9. Also, check whether ‘Age’, factor affecting the Disease, justify with suitable test.
10. Load the ‘survey’ data set from MASS library and test whether the students smoking habit is independent of their exercise level. [5 Marks]
11. Load the ‘Cars93’ data set from MASS library and verify the following [5 Marks]
12. Considering the ‘Price’ attribute, check whether the mean price of the car significantly vary with respect to ‘Air Bags’. Justify your answer with suitable hypothesis test.
13. Also check the mean mileage of the car, vary significantly with respect to ‘Drive Train’.
14. For the given quantitative samples, compute the following. [8 Marks]

34,67,40,72,37,33,42,62,49,32,52,40,31,19,68,55,57,54,37,32,

54,38,20,50,56,48,35,52,29,56,68,65,45,44,54,39,29,56,43,42.

1. First Quartile
2. Third Quartile
3. Inter Quartile Range (IQR)
4. Inner fence
5. Outer fence
6. A valve manufacturing company reports 95% of their products are defect free, based on their past 10 years data. Generate suitable probability distributions for 20 samples picked randomly from the conveyor belt of the manufacturing unit and compute the maximum probability and draw your inference. [5 marks]
7. A health drink manufacturing company claims their average fill volume of their bottle is 235ml across their factories with a standard deviation of 5ml. [5 marks]
8. What is the probability of randomly picked bottle measuring less than 230ml?
9. What is the probability of randomly picked bottle measuring more than 235ml?
10. Give an inference and comments based on the results of a & b.
11. Clinical lab data reports 65% of the patients are having high BP and 40% having high blood glucose. What is the probability that a patient have high blood glucose, given that patient is having BP. [5 Marks]
12. A weather forecast model predicts thunderstorm 70% of the time. When a domain expert (meteorologist) inputs were taken 80% success was achieved and 20% went unsuccessful. Given a meteorologist consultation was done, what is the probability that the model predicts thunderstorm. [5 Marks]
13. Load HR.txt data and verify the following with suitable hypothesis test [10 Marks]
14. Attrition is having dependency with Gender
15. Attrition is having dependency with respect to Department
16. Average Monthly Income is Gender biased or not
17. Average Monthly Income vary with respect to Department

Give a meaningful inference with proper statistical evidence.

----------------------------------------------------------------------------------------