Basic Statistics (Module – 4 (Part – 1))

Q1) Calculate probability from the given dataset for the below cases

Data_set: Cars.csv

Calculate the probability of MPG of Cars for the below cases.

MPG <- Cars$MPG

a. P(MPG>38)

Ans: Mean = 34.42208   Standard deviation = 9.131445

P(MPG<38) = (38 - 34.42208) / (9.131145)

P(MPG<38) = 0.39

From Z-Table Z score P(MPG<38) = 0.65

P(MPG>38) = 1-P(MPG<38) = 1 - 0.65 = 0.35

b. P(MPG<40)

Ans: Mean = 34.42208   Standard deviation = 9.131445

P(MPG>40) = (40 - 34.42208) / (9.131445)

P(MPG>40) = 0.61

From Z-Table score = P(MPG>40) = 0.7291

P(MPG<40) = 1 – P(MPG>40) = 1 - 0.7291 = 0.389

c. P (20<MPG<50)

Ans: Mean = 34.42208   Standard deviation = 9.131445

P(MPG>20) = (20 - 34.42208) / (9.131445) = -1.579

P(MPG>50) = (50 – 34.42208) / (9.131445) = 1.705

From Z-Table score P(MPG>20) = 0.0582   P(MPG>50) = 0.95994

P(20<MPG<50) = P(MPG>50)-P(MPG>20) = 0.95994 – 0.0582 = 0.90174

Q2) Check whether the data follows normal distribution

a) Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans: MPG of cars follows Normal distribution because it's having bell shaped curve

b) Check Whether the Adipose Tissue (AT) and Waist Circumference (Waist) from wc-at data set follows Normal Distribution
Dataset: wc-at.csv
Ans: Adipose Tissue (AT) does not follows Normal Distribution because it does not have bell shaped curve. Waist Circumference (Waist)s does not follows Normal Distribution because it does not have bell shaped curve

Q3) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval

Ans: Z scores of 90% = -1.644 Z score of 94% = -1.88 Z score of 60% = -2.05

Q4) Calculate the t scores of 95% confidence interval, 96% confidence interval, 99% confidence interval for sample size of 25

Ans: sample size = n-1 = 25-1 = 24

t scores of 95% = 2.064 t scores of 96% = 2.492 t scores of 99% = 2.492

Q5**)** A Government company claims that an average light bulb lasts 270 days. A researcher randomly selects 18 bulbs for testing. The sampled bulbs last an average of 260 days, with a standard deviation of 90 days. If the CEO's claim were true, what is the probability that 18 randomly selected bulbs would have an average life of no more than 260 days

Hint:

rcode ➐ pt(tscore,df)

df ➐ degrees of freedom

Ans: number of samples (n) = 18

Sample mean(x) = 260

Population mean(μ) = 270

Standard deviation(σ) = 90

P(x>260) = x-μ/ σ = 260 – 270 / 90 = -0.1111

From Z-table P(x>260) = 0.4562

P(x<260) = 1 – P(x>260) = 1- 0.4562 = 0.5438

The probability that 18 randomly selected bulbs would have an average life of no more than 260 days is 0.

Pt (-0.1111,17) = 0.4564

Q6) The time required for servicing transmissions is normally distributed with μ = 45 minutes and σ = 8 minutes. The service manager plans to have work begin on the transmission of a customer's car 10 minutes after the car is dropped off and the customer is told that the car will be ready within 1 hour from drop-off. What is the probability that the service manager cannot meet his commitment?

A. 0.3875
B. 0.2676
C. 0.5
D. 0.6987

Ans: mean(μ) = 45

Standard deviation(σ) = 8

Sample mean (x) = (total servicing time – work begin time) = 60 – 10 =50min

P(x>50) = x - μ / σ = 50 – 45 / 8 = 0.625

From the Z table P(x>50) = 0.7357

P(x<50) = 1 – P(x>50) = 1 – 0.7357 = 0.2643

probability that the service manager cannot meet his commitment is 0.2643 ≈ 0.2676

Q7) The current age (in years) of 400 clerical employees at an insurance claims processing center is normally distributed with mean μ = 38 and Standard deviation σ =6. For each statement below, please specify True/False. If false, briefly explain why.

A. More employees at the processing center are older than 44 than between 38 and 44.
B. A training program for employees under the age of 30 at the center would be expected to attract about 36 employees.

Ans: μ = 38 σ = 6 n = 400

A)

P(x<44) = 44 – 38 / 6 = 1

From Z table P(x<44) = 0.8413

P(x>44) = 1 – P(x<44) = 1-0.8413 = 0.1587

P(x>38) = 38 – 38 / 6 = 0

From Z table P(x>38) = 0.5000

P(x<38) = 1 - P(x>38) = 1 – 0.5000 = 0.5

P(x>44) = 44 – 38 / 6 = 1

From Z table P(x>44) = 0.8413

P(x<44) = 1 – P(x>44) = 1 – 0.8413 = 0.1587

P(38<x<44) = 0.1587 – 0.5 = -0.3413

Statement is true P(x>44) > P(38<x<44) More employees at the processing Center are older than 44

B) P (x< 30) = 30 – 38 / 6 = -1.33

From the Z table P(x>30) = 0.0918

True, because percentage of employees under age of 30 is 9.12%.
Number of employees under 30 = 400 * 9.12% = approximate 36
Therefore, the statement B of the question is also TRUE

Q8) If X1 ~ N(μ, σ2) and X2 ~ N(μ, σ2) are iid normal random variables, then what is the difference between 2 X1 and X1 + X2? Discuss both their distributions and parameters.

Ans- As we know that if X ~ N (μ1, σ1^2), and Y ~ N (μ2, σ2^2) are two independent random variables then X + Y ~ N (μ1 + μ2, σ1^2 + σ2^2), and X − Y ~ N (μ1 − μ2, σ1^2 + σ2^2). Similarly, if Z = aX + by, where X and Y are as defined above, i.e Z is linear combination of X and Y, then Z ~ N (aμ1 + bμ2, a^2σ1^2 + b^2σ2^2).
Therefore, in the question
2X1~ N (2 u,4 σ^2) and X1+X2 ~ N (μ + μ, σ^2 + σ^2) ~ N (2 u, 2σ^2)
2X1-(X1+X2) = N (4μ,6 σ^2)

Q9) Let X ~ N(100, 20^2) its (100, 20 square).Find two values, a and b, symmetric about the mean, such that the probability of the random variable taking a value between them is 0.99.

A. 90.5, 105.9
B. 80.2, 119.8
C. 22, 78
D. 48.5, 151.5
E. 90.1, 109.9

Ans- D. 48.5, 151.5

Q10) Consider a company that has two different divisions. The annual profits from the two divisions are independent and have distributions Profit1 ~ N (5, 3^2) and Profit2 ~ N (7, 4^2) respectively. Both the profits are in $ Million. Answer the following questions about the total profit of the company in Rupees. Assume that $1 = Rs. 45

A. Specify a Rupee range (centered on the mean) such that it contains 95% probability for the annual profit of the company.

Ans – range [-170991896, 710991896]

B. Specify the 5th percentile of profit (in Rupees) for the company

Ans – Rs 230460330

C. Which of the two divisions has a larger probability of making a loss in a given year?

Ans – first division Profit1 ~ N (5, 32) with 0.0477 probability

Hints:

1. Business Problem
    1.1. Objective
    1.2. Constraints (if any)
2. For each assignment the solution should be submitted in the below format
3. Research and Perform all possible steps for obtaining solution
4. For Basic Statistics explanation of the solutions should be documented in black and white along with the codes.

    One must follow these guidelines as well:

    4.1. Be thorough with the concepts of Probability, Central Limit Theorem and Perform the calculation stepwise
    4.2. For True/False Questions, explanation is must.
    4.3. R & Python code for Univariate Analysis (histogram, box plot, bar plots etc.) for data distribution to be attached
5. All the codes (executable programs) should execute without errors