# 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

1. P(MPG>38)
2. P(MPG<40)
3. P (20<MPG<50)

Ans : PFA python code for probability calculations #

1. P(MPG>38) : 0.6524
2. P(MPG<40) : 0.7294
3. P (20<MPG<50) : 0.8988

Q2) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution Dataset: Cars.csv

Python code:

1. Check Whether the Adipose Tissue (AT) and Waist Circumference (Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

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

Ans :

1. 90% confidence interval

(Manual calculations)

90% confidence interval = ( 1+ 0.90)/2 = 1.90 /2 = 0.95

Search value 0.95 in Z table.

0.950 is average of 0.9495 and 0.9505

Z score of 0.950 = 1.6 + (0.04 + 0.05)/2 = 1.6 + 0.045

Z score of 0.950 = 1.645

1. 94% confidence interval

(Manual calculations)

94% confidence interval = (1 + 0.94)/2 = 1.94/2 = 0.97

Search value 0.970 in Z table

Z score of 0.970 = 1.8 + 0.08

Z score of 0.970 = 1.88

1. 60% confidence interval

(Manual calculations)

60% confidence interval = (1 + 0.60)/2 = 1.60/2 = 0.80

Search value 0.800 in Z table

Z score of 0. 800 = 0.8 + 0.04

Z score of 0.800 = 0.84

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

Ans :

Sample size, n = 25

n-1 = 24

1. t scores of 95% confidence interval

by using t- table,

t (0.95, 24) = 2.064

1. t scores of 96% confidence interval

by using t- table,

t (0.96, 24) = (2.064 +2.492) /2.01 = 2.2667

1. t scores of 99% confidence interval

by using t- table,

t (0.99, 24) = 2.797

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 → pt(tscore,df)

df → degrees of freedom

Ans :

Mean of sample bulbs= 260 days

Population mean, = 270 days

standard deviation of sample, = 90 days

Sample , n = 18 bulbs

t - statistics for the data is given as follows:



t = = -10 /(

t = -0.471

For probability calculations, the number of degrees of freedom is n - 1, so here you need the t-distribution with 17 (18-1) degrees of freedom.

The probability that t < - 0.471 with 17 degrees of freedom assuming the population mean is true, the t-value is less than the t-value obtained with 17 degrees of freedom and a t score of - 0.471, the probability of the bulbs lasting less than 260 days on average of 0.3218 assuming the mean life of the bulbs is 300 days.

Q6) The time required for servicing transmissions is normally distributed with  = 45 minutes and  = 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

A. 0.3875

B. 0.2676

C. 0.5

D. 0.6987

Ans : B

 = 45 minutes and  = 8 minutes

= (60 – 45) / 8 = 15/8

Z = 1.875 1.88

By using Z table

P(1hr) = 0.9699

Since work started after 10 mins after car is dropped of so X will be (60 -10) mins

= (50 – 45) / 8 = 5/8

Z = 0.625 0.63

By using Z table

P(1hr) = 0.7357

probability that the service manager meet his commitment = 0.7357

Probability that the service manager cannot meet his commitment = (1- 0.7357)

= 0.2643

By Python code

import scipy.stats as stats

p2 = stats.norm.cdf(50, 45, 8)

1-p2

O/p => 0.26598552904870054

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

Q7) The current age (in years) of 400 clerical employees at an insurance claims processing center is normally distributed with mean  = 38 and Standard deviation

 =6. For each statement below, please specify True/False. If false, briefly explain why.

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

Ans :

Mean,  = 38 and

Standard deviation,  =6

1. More employees at the processing centre are older than 44 than between 38 and 44.

We have normal distribution with Mean,  = 38 and Standard deviation,  =6.

Let X be the number of employees.

So according to Question,

Probability of employees greater than age of 44 = P(X>44)

P(X>44) = 1 - P(X 44)

=

= (44 -38) /6 = 6/6 = 1

P(X 44) = P( X 0.8413 , from Z- table

Probability that the employees younger than 44 , P(X ) = 0.8413 = 84.13%

Probability that the employees older than 44 , P(X = 1 - 0.8413 = 0.1587 = 15.87%

So the probability of number of employees between 38 to 44 yrs of age ,

P (38 )= P(X 44) – P(X

= 84.13 – 50 = 34.13 %

Therefore, the statement “More employees at the processing centre are older than 44 than between 38 and 44” is false since 15.87% is less than 34.15%

|  |
| --- |
|  |

1. A training program for employees under the age of 30 at the centre would be expected to attract about 36 employees.

Probability that the employees younger than 30 = P(X )

=

= (30- 38)/6 = -8/6 = -1.3333

P(X ) = P ( Z = -1.33 ) = 0.0918 = 9.18 %

Number of employees under age of 30 = 0.0918 \* 400 = 36.72 36

So, the statement A training program for employees under the age of 30 at the centre would be expected to attract about 36 employees is 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 X1 ~ N(μ, σ2) and X2 ~ N(μ, σ2) are 2 independent random variables then

X1 + X2 N ( , ) and

X1 - X2 N ( , )

|  |
| --- |
|  |
|  | Similarly if Z = aX1 + bX2 , where X1 and X2 are as defined above, i.e Z is linear combination of X1 and X2 ,  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(µ1 + µ2, 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

X ~ N(100, 20^2)

Mean, µ = 100

Std. deviation, σ = 20

P(a<x<b) = 0.99

To Find:

Identify symmetric values for the standard normal distribution such that the area enclosed is .99

From the above details, we have to excluded area of .005 in each of the left and right tails. Hence, we want to find the 0.5th and the 99.5th percentiles Z score values

Using Python

Z value is given as stats.norm.ppf(pvalue)

Z value at 0.5th percentile is given as

                                         Z(0.5) = stats.norm.ppf(0.005)= -2.576

Z value at 99.5 percentile is given as

                         Z(99.5) = stats.norm.ppf(0.995) = 2.576

Z = (x - 100)/20 = > x = 20z+100

      a = -(20\*2.576) + 100= 48.5

      b = (20\*2.576)+100= 151.5

Two values symmetric about mean for the given standard normal distribution are[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

1. Specify a Rupee range (centered on the mean) such that it contains 95% probability for the annual profit of the company.
2. Specify the 5th percentile of profit (in Rupees) for the company
3. Which of the two divisions has a larger probability of making a loss in a given year?

Ans :

Profit1 ~ N(5, 3^2) and

Profit2 ~ N(7, 4^2)

µ1 = 5 and σ1 = 3

µ2 = 7 and σ2 = 4

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

Ans: Range is Rs (99.00810347848784, 980.9918965215122) in Millions

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

Ans: 5th percentile of profit (in Million Rupees) is 170.0

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

Ans: Division1 has a larger probability of making a loss in a given year .

**Hints:**

1. Business Problem
   1. Objective
   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:

* 1. Be thorough with the concepts of Probability, Central Limit Theorem and Perform the calculation stepwise
  2. For True/False Questions, explanation is must.
  3. R & Python code for Univariate Analysis (histogram, box plot, bar plots etc.) for data distribution to be attached

1. All the codes (executable programs) should execute without errors