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

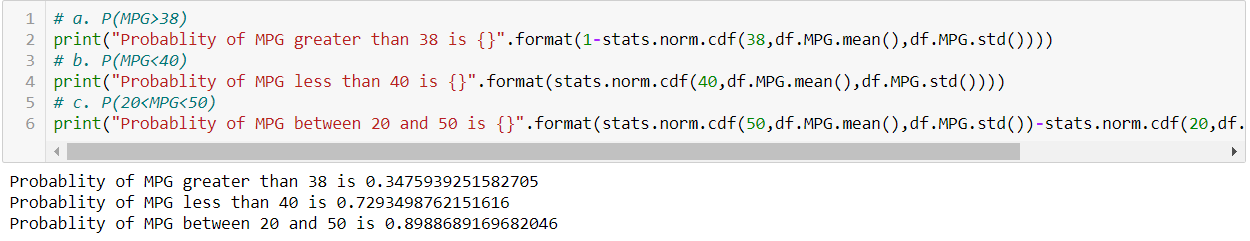
Q1) Calculate probability from the given dataset for the below cases

Data\_set: Cars.csv

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

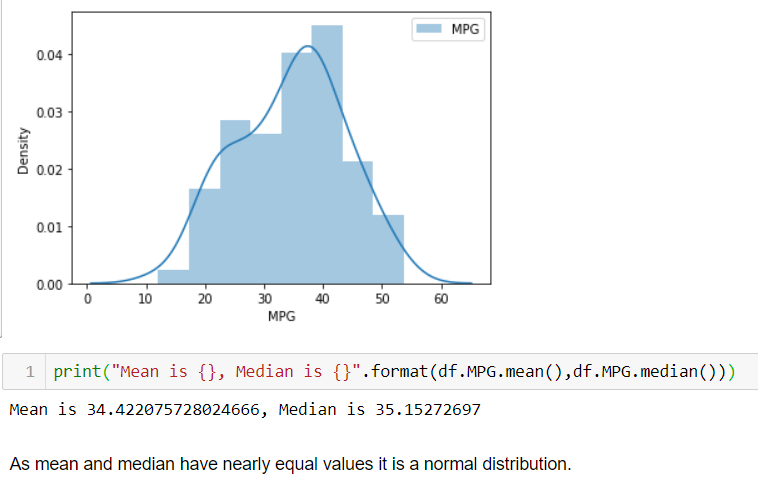
MPG <- Cars$MPG

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



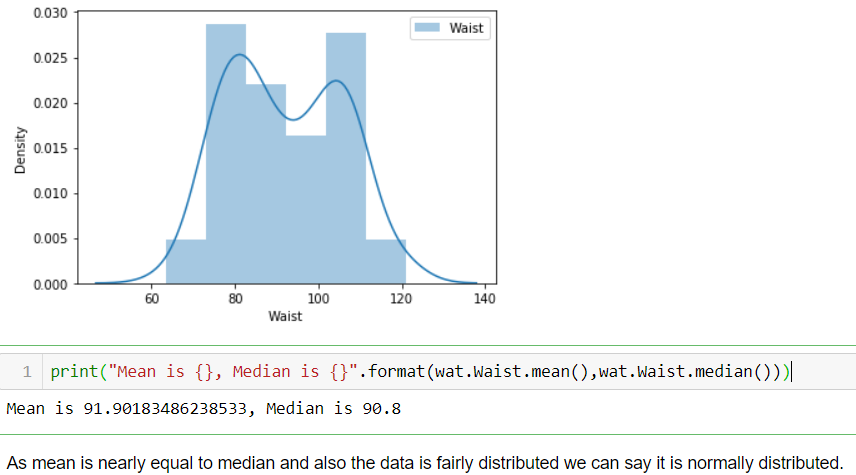
Q2) Check whether the data follows normal distribution

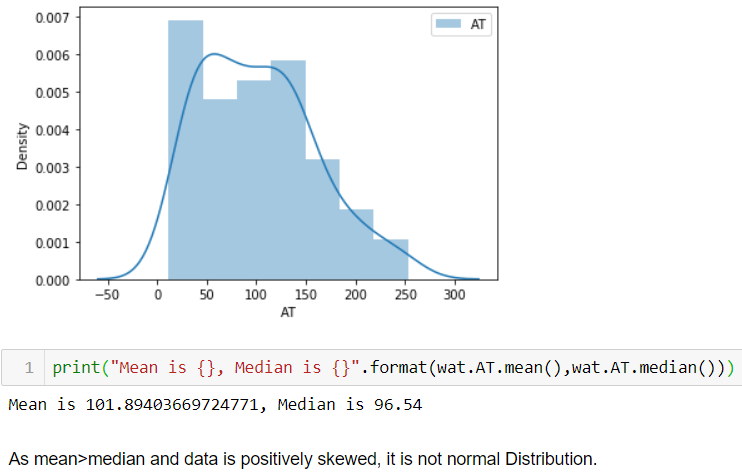
1. Check whether the MPG of Cars follows Normal Distribution Dataset:Cars.csv



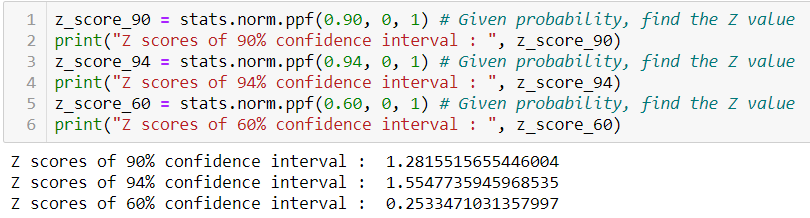
1. Check Whether the Adipose Tissue (AT) and Waist Circumference (Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

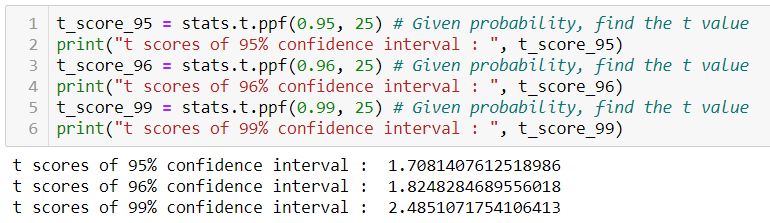




Q3) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval.



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



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→degreesoffreedom

T - statistics for the data is given as follows:

https://tex.z-dn.net/?f=t%3D%5Cdfrac%7Bx-%5Cmu%7D%7B%5Cfrac%7Bs%7D%7B%5Csqrt%20n%7D%7D

x = mean of the sample of bulbs =  260

μ = population mean = 270

s = standard deviation of the sample = 90

n = number of items in the sample = 18

https://tex.z-dn.net/?f=t%3D%5Cdfrac%7B260-270%7D%7B%5Cfrac%7B90%7D%7B%5Csqrt%2018%7D%7D

t = - 0.471

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

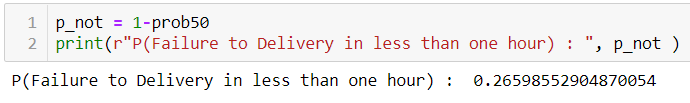
Q6) The time required for servicing transmissions is normally distributed with = 45 minutes and = 8 minutes. The service manager plans to have work begin on the transmission of a customer’s car 10 minutes after the car is dropped off and the customer is told that the car will be ready within 1 hour 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.

Mean = 38

SD = 6

Z score = (Value - Mean)/SD

Z score for 44 = (44 - 38)/6  = 1  => p-score = 84.13 %

=> People above 44 age = 100 - 84.13 = 15.87% ≈ 63    out of 400

Z score for 38 = (38 - 38)/6 = 0 => 50%

Hence People between 38 & 44 age = 84.13 - 50 = 34.13 % ≈ 137 out of 400

Hence, more employees at the processing center are older than 44 than between 38 and 44 is F**ALSE**

Z score for 30 = (30 - 38)/6 =  -1.33  =  9.15  %   ≈ 36 out of 400

Hence A training program for employees under the age of 30 at the center would be expected to attract about 36 employees - **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.

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 ) .

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

Since we need to find out the values of a and b, which are symmetric about the mean, such that the probability of random variable taking a value between them is 0.99, we have to work out in reverse order.

The Probability of getting value between a and b should be 0.99.

So, the Probability of going wrong, or the Probability outside the a and b area is 0.01 (ie.1-0.99).

The Probability towards left from a = -0.005 (ie. 0.01/2).

The Probability towards right from b = +0.005 (ie. 0.01/2).

So, since we have the probabilities of a and b, we need to calculate X, the random variable at a and b which has got these probabilities.

By finding the Standard Normal Variable Z (Z Value), we can calculate the X values.

Z=(X- μ) / σ

For Probability 0.005 the Z Value is -2.57 (from Z Table).

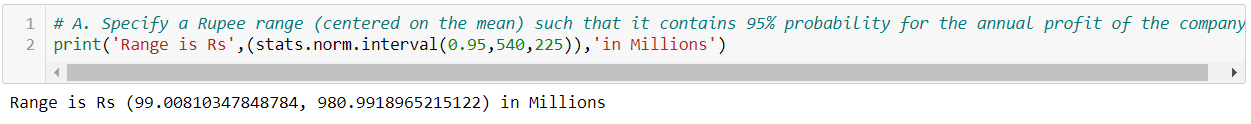
Z \* σ + μ = X

Z (-0.005) \*20+100 = -(-2.57) \*20+100 = 151.4

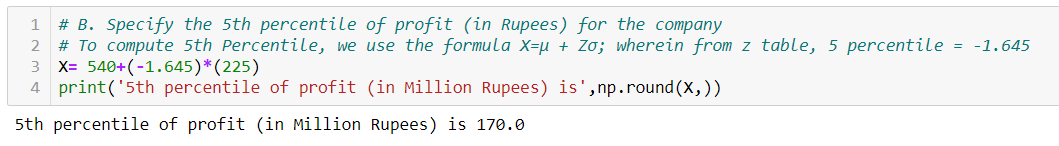
Z (+0.005) \*20+100 = (-2.57) \*20+100 = 48.6

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

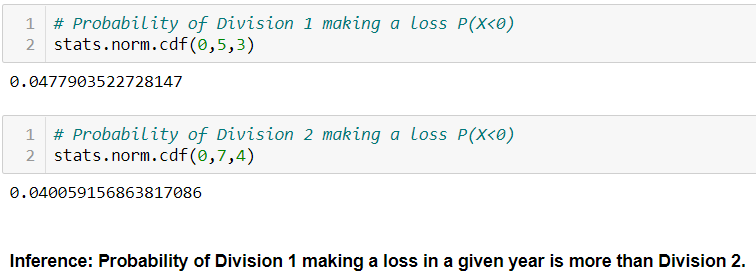
1. Specify a Rupee range (centered on the mean) such that it contains95% probability for the annual profit of the company.



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



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



**Hints:**

1. BusinessProblem
   1. Objective
   2. Constraints (ifany)
2. For each assignment the solution should be submitted in the belowformat
3. Research and Perform all possible steps for obtainingsolution
4. For Basic Statistics explanation of the solutions should be documented in black and white along with thecodes.

One must follow these guidelines as well:

* 1. BethoroughwiththeconceptsofProbability,CentralLimitTheoremandPerformthe calculation stepwise
  2. For True/False Questions, explanation ismust.
  3. R & Python code for Univariate Analysis (histogram, box plot, bar plots etc.) fordata distribution to beattached

1. All the codes (executable programs) should execute withouterrors