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

**Name**: Nithin Dsouza

**Batch ID**: 05012021-10AM

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

Dataset: Cars.csv

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

MPG <- Cars$MPG

1. P(MPG>38)
2. P(MPG<40)

c. P (20<MPG<50)

Solution:

1. P (MPG > 38)

P (MPG > 38) = 1 - P (MPG ≤ 38)

P (MPG ≤ 38) = 1 – P (X ≤ 38)

P (X ≤ 38)

Z = X- µ / σ

Z = (38 - 34.42208) / 9.131445

µ and σ values are calculated using R program, codes are attached as well.

µ = 34.42208 & σ = 9.131445

Z = 0.3918241

P (Z ≤ 0.3918241)

By using Z-distribution table,

P (X ≤ 38) = P (Z ≤ 0.3918241) = 0.65173

P (MPG > 38) = 1 – P (X ≤ 38)

P (MPG > 38) = 1 - 0.65173

P (MPG > 38) = 0.34827

1. P(MPG<40)

P(MPG<40)

Z = X- µ / σ

µ and σ values are calculated using R program, codes are attached as well.

µ = 34.42208 & σ = 9.131445

Z = (40- 34.42208) / 9.131445

Z = 0.6108475

P (Z < 0.6108475)

By using Z-distribution table,

P(MPG<40) = P (Z < 0.6108475) = 0.72907

؞ P(MPG<40) = 0.72907

1. P (20<MPG<50)

P (20<MPG<50) = P (x1 < X < x2)

P (x1 < X < x2) = P (X < x2) – P (X < x1)

Now P (X < x2) = P (X < 50)

Z = X- µ / σ

µ and σ values are calculated using R program, codes are attached as well.

µ = 34.42208 & σ = 9.131445

Z = (50 - 34.42208) / 9.131445

Z = 1.705964

By using Z-distribution table,

P(Z<1.705964) = 0.95543

P (X < 50) = 0.95543

Now finding P (X < x1) = P (X < 20)

Z = X- µ / σ

µ and σ values are calculated using R program, codes are attached as well.

µ = 34.42208 & σ = 9.131445

Z = (20 - 34.42208) / 9.131445

Z = -1.579386

By using Z-distribution table,

P (Z< -1.579386) = 0.05821

P (X < 50) = 0.05821

So, P (X < 50) – P (X < 20) = 0.95543 - 0.05821

P (20<MPG<50) = 0.89722

Q2) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution Dataset: Cars.csv
2. Check Whether the Adipose Tissue (AT) and Waist Circumference (Waist) from wc-at data set follows Normal Distribution

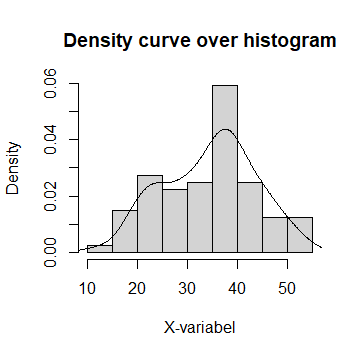
Dataset: wc-at.csv

Solution:

1. Checking Normal distribution of MPG in Cars.csv

Histogram is plotted along with density curve to check normality of MPG

Code is attached in BasicStats(Part-1).r file.

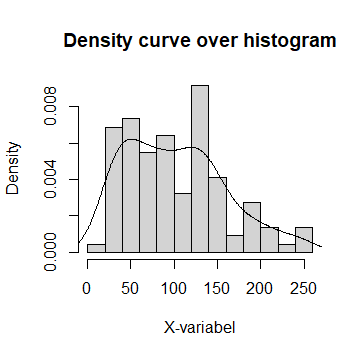


From the Probability Histogram plot we can say that MPG is a not normal distribution.

b. Checking Normal distribution of AT & WC in wc-at.csv

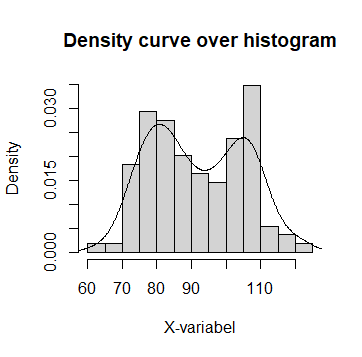
Histogram is plotted along with density curve to check normality of AT

Code is attached in BasicStats(Part-1).r file.



From the Probability Histogram plot we can say that AT is a not normal distribution.

Histogram is plotted along with density curve to check normality of WC.



From the probability Histogram plot we can say that WC is not a normal distribution.

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

Z-score for 90% Confidence interval

From using Z-table, (1-90%) / 2 = 0.05

Now check for 0.05 in Z-table

It lies between -1.64 and -1.65, by taking average -1.645

Z-score for 94% Confidence interval

From using Z-table, (1-94%) / 2 = 0.03

Now check for 0.03 in Z-table

It lies between -1.88 and -1.89, by taking average -1.885

Z-score for 60% Confidence interval

From using Z-table, (1-60%) / 2 = 0.2

Now check for 0.2 in Z-table

It lies between -0.84 and -0.85, by taking average -0.845

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

t-score for 95% Confidence interval of df = 25-1 = 24

From using t-table, df = 24

Now check for 95% CI in t-table

It is 2.064

t-score for 96% Confidence interval of df = 24

From using t-table, df = 24, (1-96%) / 2 = 0.02

Now check for 0.02(two tailed) CI in t-table under df = 24

It is 2.492

t-score for 99% Confidence interval of df = 24

From using t-table, df = 24

Now check for 99% CI in t-table

It is 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

Solution:

x -> 260(sample mean)

µ ->270(population mean)

s -> 90(standard deviation of sample)

n -> 18 (items in sample)

t =

t = (260 – 270) / (90 / √18)

t = -0.4714

Number of probability calculation is n-1 = 18-1 = 17

t < -0.4714 & df = 17

By using hint given,

pt(tscore , df)

pt(-0.4714, 17)

=0.3216741

0.3216741 is the probability that 18 randomly selected bulbs would have an average life of no more than 260 days.

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

Solution:

X = 50 (Time taken to repair 60min – 10min = 50min)

µ = 45

σ = 8

P (X > 50) = 1 – P(X≤50)

Z = (X-µ) / σ

Z = (50 – 45) / 8

Z = 0.625

P (X ≤ 50) = P (Z ≤ 0.625)

P-value from Z-Table: (used z-calculation site)

P(x≤50) = 0.73401

P (X > 50) = 1 – P(X≤50) = 1 - 0.73401

P (X > 50) = 0.26599 ≈ 0.2676 (Options Given)

0.26599 is the probability that the service manager cannot meet his commitment.

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.

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.

Solution:

µ = 38

σ = 6

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

X = number of employees acc. to question

P (employees age > 44)

P(x>44) = 1 - P(x≤44)

Z = (x-µ) / σ

Z = (44 – 38) / 6

Z = 1

P-value from Z-Table:

P(x<44) = 0.84134

P(x>44) = 1 - P(x≤44) = 0.15866 = 15.86%

P (employees age: 38 < X < 44)

P (38 < x < 44) = P(x<44) – P(x<38)

Now,

P(x<44)

Z = (X-µ) / σ

Z = (44 – 38) / 6

Z = 1

P-value from Z-Table:

P(x<44) = 0.84134

And,

P(x<38)

Z = (X-µ) / σ

Z = (38 – 38) / 6

Z = 0

P-value from Z-Table:

P(x<38) = 0.5

So,

P (38 < x < 44) = P(x<44) – P(x<38)

P (38 < x < 44) = 0.84134 – 0.5

P (38 < x < 44) = 0.34134 = 34.13% (Greater)

Therefore,

The Statement “More employees at the processing center are older than 44 than between 38 and 44” is True.

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

P (employee age less than 30) = P (x < 30)

P(x<30)

Z = (X-µ) / σ

Z = (30 – 38) / 6

Z = -1.333333

P-value from Z-Table:

P(x<30) = 0.091211 = 9.12%

So, no. of employees with probability 0.091211 under age 30 is

= 0.091211 \* 400 = 36.4844 (≈36 employees)

The Statement “A training program for employees under the age of 30 at the center 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.

Solution:

X1 and X2 are two independent random variables (iid normal random variable)

X1 + X2 ~ N (µ+µ, σ2+σ2)

X1 + X2 ~ N (2µ + 2 σ2)

Z = aX1 ~ N (aµ, a2σ2)

2X1 ~ N (2µ, 4σ2)

Formula, N (X1 ± X2 ….) = N(X1) ± N(X2)

Therefore,

2X1 – (X1+X2) = N (2µ, 4σ2) ± N (2µ + 2 σ2)

2X1 – (X1+X2) = N (4µ, 6σ2)

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.

1. 90.5, 105.9
2. 80.2, 119.8
3. 22, 78
4. 48.5, 151.5
5. 90.1, 109.9

Solution:

Given, µ = 100, σ = 20

Probability of a & b symmetric about mean is 0.99

Therefore, probability outside a and b is 1-0.99 = 0.01

Since a & b is symmetric about the mean,

Probability left of ‘a’ is -0.005 (0.01/2 = 0.005)

Probability right of ‘b’ is +0.005 (0.01/2 = 0.005)

Using these values, we can find X (random variable which satisfy this probability)

By using Z formula, we can find X,

Z = x-µ / σ

For 0.05 Z value is -2.57(Z-table)

By tweaking Z formula,

X = (Z \* σ) + µ

Xa = -(-2.57 \* 20) +100 = 151.4

Xb = +(-2.57 \* 20) +100 = 48.6

So, option D correct.

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?

Solution: Code File.