|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Discrete |
| Weight of a person | Continuous |
| Weight of Gold | Continuous |
| Distance between two places | Continuous |
| Length of a leaf | Continuous |
| Dog's weight | Continuous |
| Blue Color | Discrete |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | Discrete |

Q1) Identify the Data type for the Following:

Q2) Identify the Data types, which were among the following

Nominal, Ordinal, Interval, Ratio.

|  |  |
| --- | --- |
| Data | Data Type |
| Gender | Nominal |
| High School Class Ranking | Ordinal |
| Celsius Temperature | Interval |
| Weight | Ratio |
| Hair Color | Nominal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation | Ordinal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ratio |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Interval |
| Number of Children | Nominal |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Interval |
| Years of Education | Ordinal |

Q3) Three Coins are tossed, find the probability that two heads and one tail are obtained?

**Ans**: Probability = (Number of favourable outcome)/ (Total number of outcomes)

P(2 head and 1 tail)=N(Events of 2 head and 1 tail )/N(Three coins are tossed)

=**3/8**

Q4) Two Dice are rolled, find the probability that sum is

1. Equal to 1

**Ans: 0/36**

1. Less than or equal to 4

**Ans:1/6=0.166**

1. Sum is divisible by 2 and 3

**Ans:1/6=0.166**

Q5) A bag contains 2 red, 3 green and 2 blue balls. Two balls are drawn at random. What is the probability that none of the balls drawn is blue?

**Ans: Total no of balls=2+3+2=7**

**Two balls are drawn randomly=7C2**

**None of the ball is blue=Total ball -Only blue ball=7-2=5=5C2**

**Probability that none of the ball is drawn blue=5C2/7C2=10/21**

Q6) Calculate the Expected number of candies for a randomly selected child

Below are the probabilities of count of candies for children (ignoring the nature of the child-Generalized view)

|  |  |  |
| --- | --- | --- |
| CHILD | Candies count | Probability |
| A | 1 | 0.015 |
| B | 4 | 0.20 |
| C | 3 | 0.65 |
| D | 5 | 0.005 |
| E | 6 | 0.01 |
| F | 2 | 0.120 |

Child A – probability of having 1 candy = 0.015.

Child B – probability of having 4 candies = 0.20

**Ans: The Expected number of candies for a randomly selected child**

**(x\*P(x))=(1\*0.015)+(4\*0.20)+(3\*0.65)+(5\*0.005)+(6\*0.01)+(2\*0.120)=3.09**

Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

* For Points,Score,Weigh>

Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

**Use Q7.csv file**

**Ans:**

**#for points**

> mean(Q7$Points)

[1] 3.596563

> median(Q7$Points)

[1] 3.695

> mfv(Q7$Points)

[1] 3.07 3.92

> var(Q7$Points)

[1] 0.2858814

> sd(Q7$Points)

[1] 0.5346787

> range(Q7$Points)

[1] 2.76 4.93

diff(range(Q7$Points))

[1] 2.17

**#for score**

mean(Q7$Score)

[1] 3.21725

> median(Q7$Score)

[1] 3.325

> mfv(Q7$Score)

[1] 3.44

> var(Q7$Score)

[1] 0.957379

> sd(Q7$Score)

[1] 0.9784574

> range(Q7$Score)

[1] 1.513 5.424

diff(range(Q7$Score))

[1] 3.911

**#for weigh**

mean(Q7$Weigh)

[1] 17.84875

> median(Q7$Weigh)

[1] 17.71

> mfv(Q7$Weigh)

[1] 17.02 18.90

> var(Q7$Weigh)

[1] 3.193166

> sd(Q7$Weigh)

[1] 1.786943

> range(Q7$Weigh)

[1] 14.5 22.9

diff(range(Q7$Weigh))

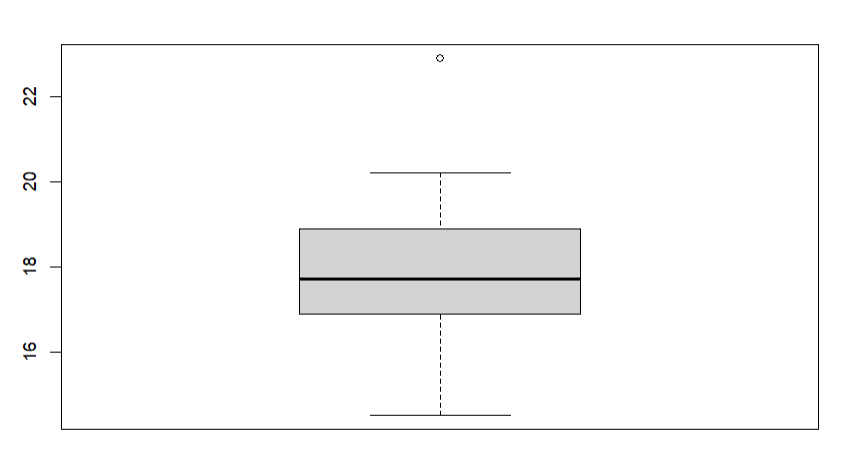
[1] 8.4

|  |
| --- |
| boxplot(Q7[2]) |

boxplot(Q7[3])

|  |
| --- |
|  |

**boxplot(Q7[4])**

****

Q8) Calculate Expected Value for the problem below

1. The weights (X) of patients at a clinic (in pounds), are

108, 110, 123, 134, 135, 145, 167, 187, 199

Assume one of the patients is chosen at random. What is the Expected Value of the Weight of that patient?

**Ans: Expected value=sum(X\*probability(X))**

**={[(1/9)\*(108)]+ [(1/9)\*(110)]+[(1/9)\*(123)]+ [(1/9)\*(134)]+ [(1/9)\*(135)]+ [(1/9)\*(145)]+[(1/9)\*(167)]+ [(1/9)\*(187)]+ [(1/9)\*(199)]}=145.33**

**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

**Cars speed and distance**

**Use Q9\_a.csv**

Ans:

#for speed

skewness(Q9\_a$speed)

[1] -0.1139548 negative skewness means left skew i.e data distribution on right side

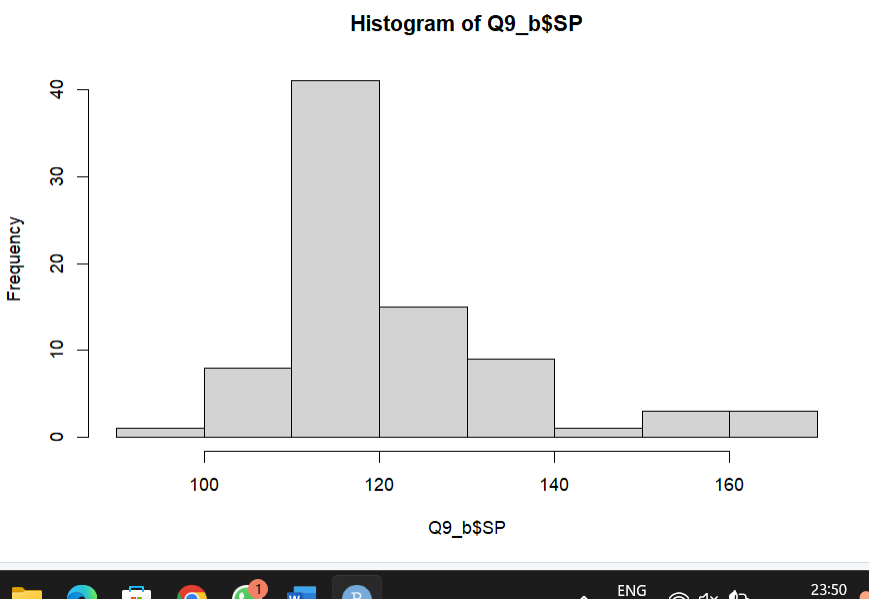
|  |
| --- |
| kurtosis(Q9\_a$speed)  [1] 2.422853 positive kurtosis i.e data normal distribution  > hist(Q9\_a$speed)    #for dist  skewness(Q9\_a$dist)  [1] 0.7824835 positive skewness means right skew i.e data distribution on left side  kurtosis(Q9\_a$dist)  [1] 3.248019 positive kurtosis i.e distribution wide not peak  hist(Q9\_a$dist) |
|  |
|  |

**SP and Weight(WT)**

**Use Q9\_b.csv**

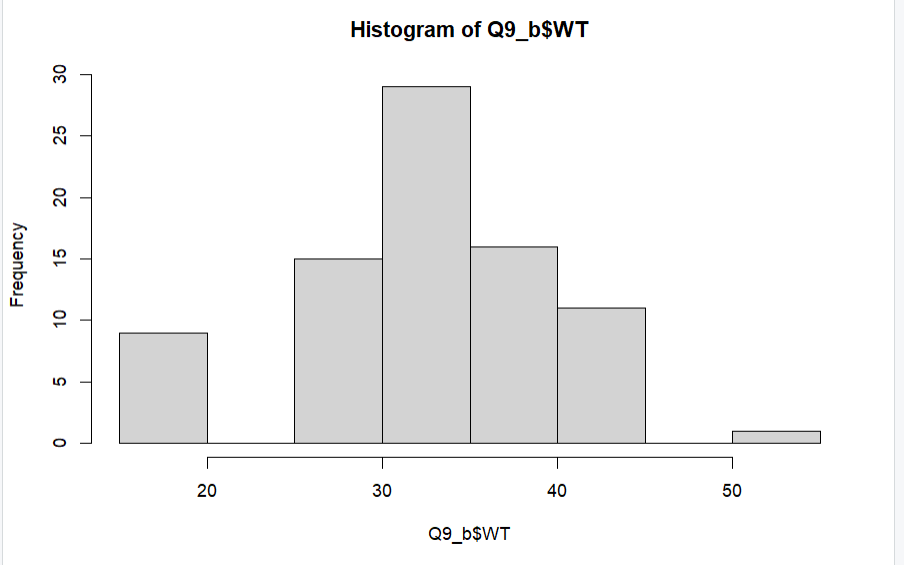
**Ans:#for SP**

|  |
| --- |
| skewness(Q9\_b$SP)  [1] 1.581454  > kurtosis((Q9\_b$SP))  [1] 5.723521  > hist(Q9\_b$SP) |
|  |
| |  | | --- | | > | |

****

**#for WT**

|  |
| --- |
| skewness((Q9\_b$WT))  [1] -0.6033099  > kurtosis(Q9\_b$WT)  [1] 3.819466  > hist(Q9\_b$WT) |
|  |
| |  | | --- | |  | |

****

**Q10) Draw inferences about the following boxplot & histogram**



**Ans: In the above histogram the peak is on the left side and skewed is on the right side. So the above histogram is positively or right skewed. So it contain the outliers which is on the higher side i.e on the right side. Mean>Median.**



**Ans :The above diagram is box plot diagram. This diagram show it contain outliers. The outliers are present in the upper extreme**.

**Q11)** Suppose we want to estimate the average weight of an adult male in Mexico. We draw a random sample of 2,000 men from a population of 3,000,000 men and weigh them. We find that the average person in our sample weighs 200 pounds, and the standard deviation of the sample is 30 pounds. Calculate 94%,98%,96% confidence interval?

Ans:

**Q12)** Below are the scores obtained by a student in tests

**34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56**

1. Find mean, median, variance, standard deviation.

**Ans**: stud=c(34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56)

> mean(stud)

[1] 41

> median(stud)

[1] 40.5

> mode(stud)

[1] "numeric"

> var(stud)

[1] 25.52941

> sd(stud)

[1] 5.052664

1. What can we say about the student marks?

**Ans: We can say that the data doesn’t have any outliers. As the mean is greater than median it is slightly skewed towards right side.**

Q13) What is the nature of skewness when mean, median of data are equal?

**Ans: It’s a Normal Distribution. No skewness present. Perfect symmetric distribution.**

Q14) What is the nature of skewness when mean > median ?

**Ans: Positively skewed or right skewed**

Q15) What is the nature of skewness when median > mean?

**Ans: Negatively skewed or Left skewed**

Q16) What does positive kurtosis value indicates for a data ?

**Ans: It means curve is more peaked and it is Lepto kurtotic. Positive kurtosis.**

Q17) What does negative kurtosis value indicates for a data?

**Ans: Negative kurtosis, the curve is flatter.**

Q18) Answer the below questions using the below boxplot visualization.



What can we say about the distribution of the data?

**Ans: The box plot is not normally distributed , median is towards right side. The box plot does not has any outliers. It is skewed towards left side.**

What is nature of skewness of the data?

**Ans: It is Left skewed data.**

What will be the IQR of the data (approximately)?

**Ans: IQR=Upper quartile(Q3)-Lower quartile(Q1 )=18-10=8**

Q19) Comment on the below Boxplot visualizations?



Draw an Inference from the distribution of data for Boxplot 1 with respect Boxplot 2.

**Ans: The both boxplot are Normally distributed with 0 to no skewness in the upper or lower extreme. The both median are same i.e lies approximately in between 250 to 275 range. Lastly no outliers are present in upper or lower extreme to the both the diagram.**

Q 20) 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)

Ans: P(MPG>38)= 1-pnorm(38,mean(Cars$MPG),sd(Cars$MPG))

=0.3475939

**(here >38 is given so when we do normal distribution to 38 all values till 38 i.e (<38) is generated so for (>38) we should subtract it with 1.)**

* 1. P(MPG<40)

P(MPG<40)= pnorm(40,mean(Cars$MPG),sd(Cars$MPG))

= 0.7293499

* 1. P (20<MPG<50)

P(20<MPG<50)= pnorm(50,mean(Cars$MPG),sd(Cars$MPG))-

(1-pnorm(20,mean(Cars$MPG),sd(Cars$MPG)))

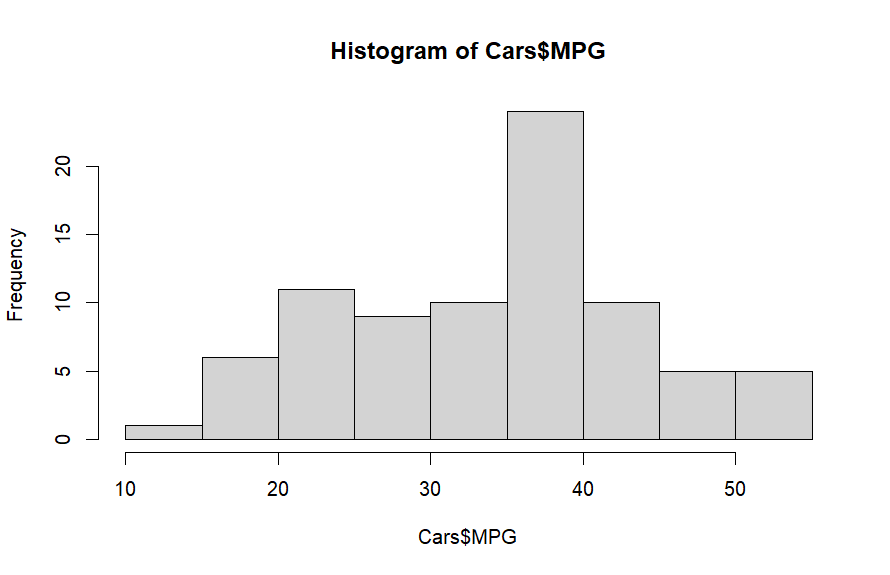
=0.01311647

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans:

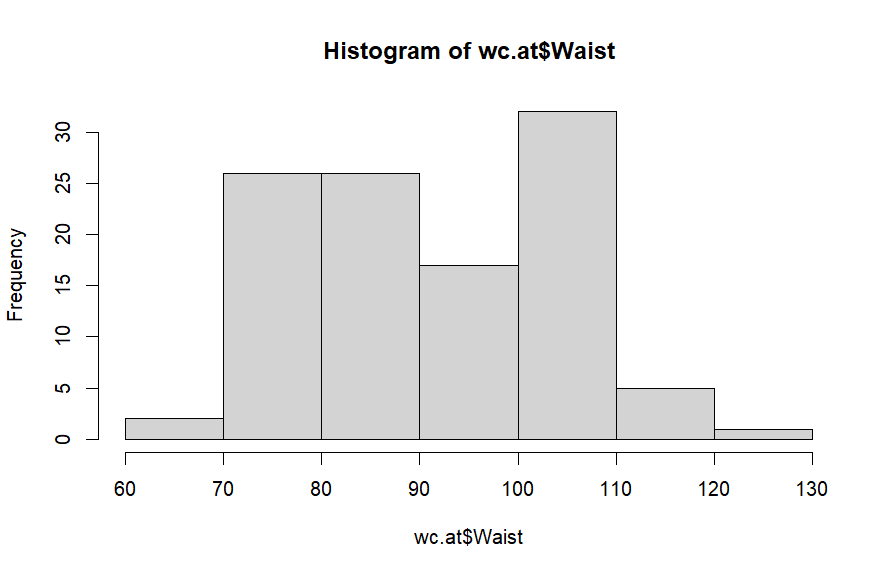


Ans :- No it is not normal distribution. As from the histogram it is clearly mentioned that the curve is not symmetrical and not bell shaped. Rather it is unsymmetrical and flatter in shape

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

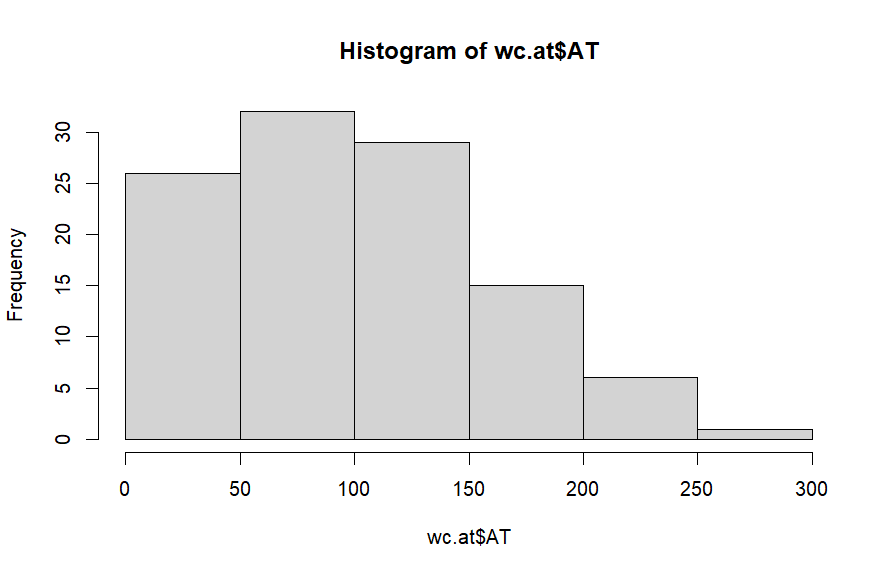
Dataset: wc-at.csv

Ans: For Waist



Waist Circumference(Waist) dataset doesnot form Normal Distribution as the curve is not symmetric and bell shape.

For Adipose



Adipose Tissue (AT) dataset in not Normal Distribution as the plot is clearly positively skewed. And does not form symmetric and bell shape plot.

Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval

Ans: **A=(1+CL)/2 ; z-score formula=qnorm(CI value)**

For 90%

qnorm((1+0.90)/2)

[1] 1.644854

For 94%

> qnorm((1+0.94)/2)

[1] 1.880794

For 60%

> qnorm((1+0.60)/2)

[1] 0.8416212

|  |
| --- |
|  |
|  |
| |  | | --- | |  | |

Q 23) Calculate the t scores of 95% confidence interval, 96% confidence interval, 99% confidence interval for sample size of 25

Ans: n=25 , df=n-1=25-1=24

t-score formula=qt(Confidence interval value,degree of freedom)

For 95%

qt((1+0.95)/2,24)

[1] 2.063899

For 96%

> qt((1+0.96)/2,24)

[1] 2.171545

For 99%

> qt((1+0.99)/2,24)

[1] 2.79694

Q 24**)** 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: Consider it is a normal distribution**

**X(mean)=260**

**µ=270**

**std=90**

**number of sample=18**

**t\_score=(x-µ)/(std/(number of sample^1/2))**

**(260-270)/(90/(18^(1/2)))**

**[1] -0.4714045**

**t\_score = -0.4714045**

**df=n=1=18=1=17**

**pt(t\_score,df)= pt(-0.4714045,17)**

**0.3216725**

**Probability is 32%**