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

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

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

So the total number of events=8

Ans. 3/8 (HHT,HTH,THH)

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

1. Equal to 1
2. Less than or equal to 4
3. Sum is divisible by 2 and 3

Ans.a) 0% probability

b) 6/36=1/6 (1,1),(1,2)(1,2)(1,3)(2,1)(2,2)(3,1)

c) (6/36)= 1/6 (1,5),(2,4),(3,3),(4,2),(5,1)(6,6)

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: it is a case of comination rule so just applying nCr equation

Total no.of events will be nCr= 7C2 = 7!/2!(7-2!)=21

Interested events 5C2= 5!/2!-3!=10

Probability that none of the ball drawn blue is 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:Expected number =E(x) =E (count), x( probability) =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.

Answer: for points

Q7<-read.csv("C:/Users/harsh/Desktop/excel r/Q7.csv")

**mean(Q7$Points)**

**[1] 3.596563**

**median(Q7$Points)**

**[1] 3.695**

**getmode<-function(Q7){**

**uniquv<-unique(Q7)**

**uniquv[which.max(tabulate(match(Q7,uniquv)))]**

**}**

**getmode(Q7$Points)**

**[1] 3.92**

**var(Q7$Points)**

**[1] 0.2858814**

**sd(Q7$Points)**

**[1] 0.5346787**

**range(Q7$Points)**

**[1] 2.76 4.93**

**For score**

**mean(Q7$Score)**

**[1] 3.21725**

**median(Q7$Score)**

**[1] 3.325**

**getmode(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**

**For weigh**

**attach(Q7)**

**> mean(Weigh)**

**[1] 17.84875**

**median(Weigh)**

**[1] 17.71**

**getmode(Q7$Weigh)**

**[1] 17.02**

**var(Weigh)**

**[1] 3.193166**

**sd(Weigh)**

**[1] 1.786943**

**range(Weigh)**

**[1] 14.5 22.9**



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:EX/N= 1308/9 or (108\*0.11+110\*0.11…….+…+…=145.33

[1] 145.3333

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

**Cars speed and distance**



**SP and Weight(WT)**

**Ans :Cars speed and distance**

**Q9<-read.csv("C:/Users/harsh/Desktop/excel r/Q9\_a.csv")**

**install.packages(moments)**

**library(moments)**

**View(Q9)**

**skewness(Q9$speed)**

**-0.1139548**

**skewness(Q9$dist)**

**[1] 0.7824835**

**kurtosis(Q9$speed)**

**[1] 2.422853**

**kurtosis(Q9$dist)**

**3.248019**

**FOR SP AND WEIGHT**

**Q9B<-read.csv("C:/Users/harsh/Desktop/excel r/Q9\_b.csv")**

**> View(Q9B)**

**> View(Q9B)**

**> skewness(Q9B$SP)**

**[1] 1.581454**

**> kurtosis(Q9B$SP)**

**[1] 5.723521**

**> skewness(Q9B$WT)**

**[1] -0.6033099**

**> kurtosis(Q9B$WT)**

**[1] 3.819466**



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



Ans : The histogram shows a positive data distribution majorly the data values are form around 50-100 range and boxplot shows the distribution has lot of outliers towards upper area.

**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**

**n=2000**

**𝑋̅= 200**

**s= 30**

**Confidence Interval Estimate= 𝑋̅ ± Z 𝑠/√𝑛 => 200 ± Z 30√2000**

**94% Confidence: qnorm(0.94)**

**qnorm(0.94)**

**[1] 1.554774**

**200+1.554774\*0.6708=201.0429**

**200-1.554774\*0.6708=198.9571**

**98% Confidence: > qnorm(0.98)**

**200+ 2.053749\*0.6708=**

|  |
| --- |
| **201.3777** |
|  |
| |  | | --- | | **> 200-2.053749\*0.6708**  **[1] 198.6223** | |

**96% Confidence: > qnorm(0.96)**

**qnorm(0.96)**

**[1] 1.750686**

**200+1.750686\*0.6708= 201.1744**

**200-1.750686\*0.6708**

**[1] 198.8256**

**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.
2. What can we say about the student marks?

**Ans : mark<-c(34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56)**

**View(mark)**

**mean(mark)**

**[1] 41**

**median(mark)**

**[1] 40.5**

**var(mark)**

**[1] 25.52941**

**> sd(mark)**

**[1] 5.05266**

**Marks are in a positive distribution. no outliers**

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

Ans : symmetry

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

Ans positive skewed or right skewed

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

Negatively skewed or left skewed

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

The distribution has a high peakdness and less varience

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

Less peakdness and more variation

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



What can we say about the distribution of the data?

Ans :data distribution have a left skewness possess a negative distribution of data

Data is not normally distributed.

What is nature of skewness of the data?

Left skwed

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

IQR=UPPER QUARTILE –LOWER QUARTILE

18-10=8

Q19) Comment on the below Boxplot visualizations?



1.BOTH BOXPLOTS HAVE SAME MEDIAN ,SHOWS EQUAL DISTRIBUTIONS ON 2 SIDES OF BOXPLOT.NO OUTLIERS PRESENT AND NO SKEWNESS

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

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)**
  2. **P(MPG<40)**

**c. P (20<MPG<50)**

**ans a:p(MPG>38)**

**cars<- read.csv("C:/Users/harsh/Desktop/excel r/Cars.csv")**

**View(cars)**

**mpg<-cars$MPG**

**View(mpg)**

**mean(mpg)**

**34.42208**

**sd(mpg)**

**9.131445**

**pnorm(38,34.42,9.13)**

**0.652513**

**1-Pnorm= 1-0.652513**

**0.347487**

**B. P(MPG<40)**

**pnorm(40,34.42,9.13)**

**[1] 0.7294571**

**C. P (20<MPG<50)**

**pnorm(50,34.42,9.13)**

**[1] 0.956039**

**> pnorm(20,34.42,9.13)**

**[1] 0.05712119**

**> 0.956039-0.05712119**

**[1] 0.8989178**

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans

cars<- read.csv("C:/Users/harsh/Desktop/excel r/cars mlr/Cars.csv")

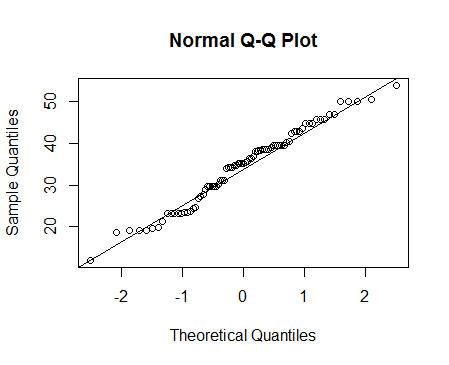
View(cars)

summary(cars)

MPG<-(cars$MPG)

qqnorm(cars$MPG)

qqline(cars$MPG)



almost all the points on line so it follows a normal distribution

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

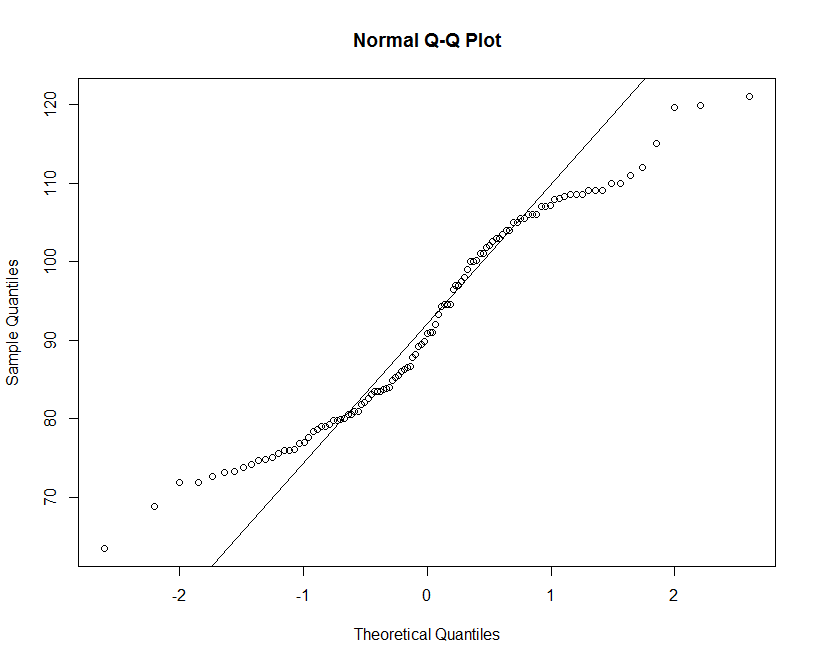
Dataset: wc-at.csv

wc <-read.csv("C:/Users/harsh/Desktop/excel r/Data+sets+of+slr/wc-at.csv")

View(wc)

qqnorm(wc$Waist)

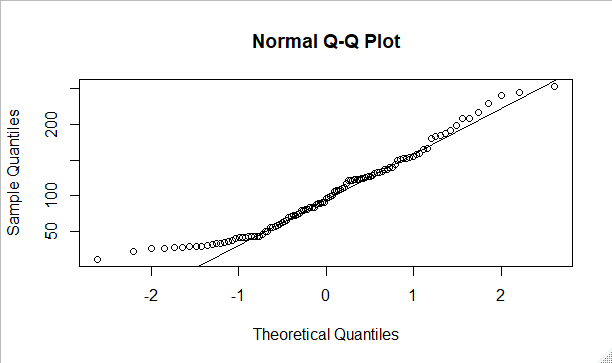
qqline(wc$Waist)



waist data is following normal distribution since almost all the data points are lying on a straight line.

qqnorm(wc$AT)

qqline(wc$AT)



Adipose tissue data is following normal distribution since almost all the data points are lying on a straight line.

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

**90% (1-α/2)**

|  |
| --- |
| **qnorm(0.95)**  **[1] 1.644854** |
|  |
| |  | | --- | | **94% qnorm(0.97)**  **[1] 1.880794**  **60% qnorm(0.8)**  **[1] 0.8416212** | |

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

|  |
| --- |
| **95%= qt(0.975,24)**  **[1] 2.063899** |
| **96%= qt(0.98,24)**  **[1] 2.171545**  **99%= qt(0.995,24)**  **[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 :the size of bulb taken for test=18(n)

average or mean=260 sd=90 µ=270

t= = = -10/21.21= -0.4714

now we have to find p value by pt(tscore,df)

>pt(-0.4714,17)

[1] 0.3216741

Required probability will be 32%.