|  |  |
| --- | --- |
| 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 | Continuous |
| Number of times married | Discrete |
| Gender (Male or Female) | Continuous |

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

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

Ans:3coins tossed, the possible outcomes are:8

HHH,HHT,HTH,HTT,THH,THT,TTH,TTT

Interested events:3

Probability :3/8

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

Ans:possible outcome are:36

(1,1) (1,2) (1,3) (1,4) (1,5) (1,6)

(2,1) (2,2) (2,3) (2,4) (2,5) (2,6)

(3,1) (3,2) (3,3) (3,4) (3,5) (3,6)

(4,1) (4,2) (4,3) (4,4) (4,5) (4,6)

(5,1) (5,2) (5,3) (5,4) (5,5) (5,6)

(6,1) (6,2) (6,3) (6,4) (6,5) (6,6)

1. Equal to 1

Ans: p(E)=n(E)/n(S)=0/36=0

1. Less than or equal to 4

Ans: Three possible outcomes give a sum less than 4: E = {(1,1),(1,2),(2,1)}, hence.P(A) =3/36

Three possible outcomes give a sum equal to 4: E = {(1,3),(2,2),(3,1)}, hence.P(B)=3/36

P(AorB)=P(A)+P(B) = 3/36+3/36=6/36=1/6

1. Sum is divisible by 2 and 3

ANS : Sum is divisible by 2 and 3

Total number of events=36

Interested number of events={(1,5),(2,4),(3,3),(4,2),(5,1),(6,6)}=6

P(sum is divisible by 2 and 3)=6/36=1/6

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:possible outcomes are:7C2=21

Interested events :5C2=10(2C2+3C2+2C1.3C1=1+3+6)

Probability=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: Ans: EV=sigma x\*p(x)

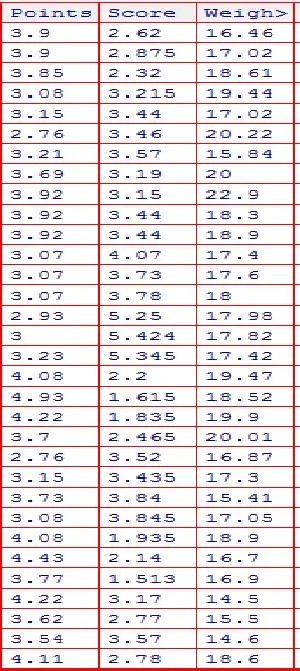
=(1\*0.015 + 4\*0.20 + 3\*0.65 + 5\*0.005 + 6\*0.01 + 2\*0.120)=3.09

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:**

Mean=sum by total

Median=middle most value of a sorted data

Mode= most frequency of data

|  |  |  |  |
| --- | --- | --- | --- |
|  | Points | Score | Weight |
| Mean | 3.596563 | 3.21725 | 17.84875 |
| Median | 3.695 | 3.325 | 17.71 |
| Mode | 3.92 | 3.44 | 17.02 |
| Variance | 0.285881 | 0.957379 | 1.786943 |
| SD | 0.534679 | 0.978457 | 1.786943 |
| Range | 2.17 | 3.911 | 8.4 |

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?

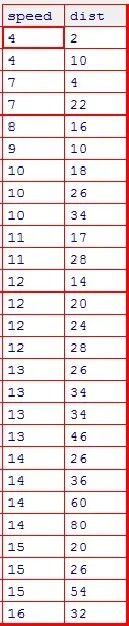
**ANS:** **EV=sigma x.p(x)**

**Ev=x1.p(x1) + x2. p(x2) + ….+ xn.p(xn)**

**Here EV=1/9(108+110+123+134+135+145+167+187+199)=1308/9=145.33**

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

**Cars speed and distance**

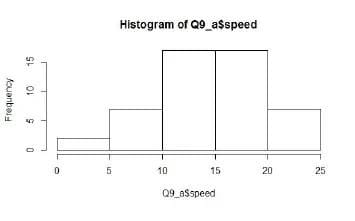
****

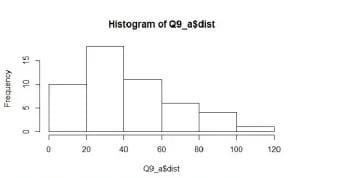
**Use Q9\_a.csv**

**By using histograms in R I can concluded that**

**Skewness of car speed : slight left/negative skewed**

**Skewness of car distance : right/positive skewed**

****

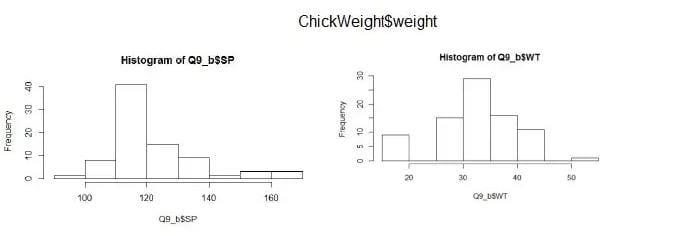
****

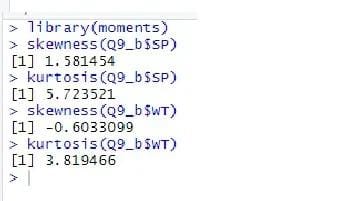
**9.b)**

**By using histograms in R I can concluded that**

**Skewness of sp : right/positive skewed**

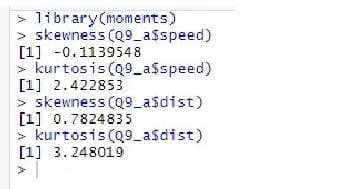
**Skewness of WT : slight normal distributed**

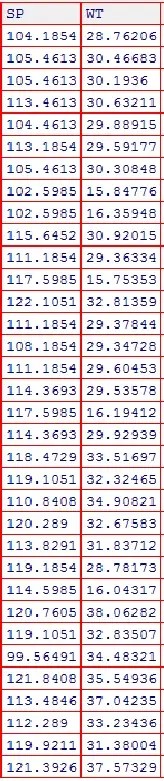
****

****

**SP and Weight(WT)**

**Use Q9\_b.csv**

****

****

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



**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: X+/-(Z1-α.σ/sqrt(n)

Degrees of freedom= 2000-1= 1999

Confidence interval= 94%

(1-σ/2)= 1-0.03) =0.97

for confidene interval for 94% is 1.882

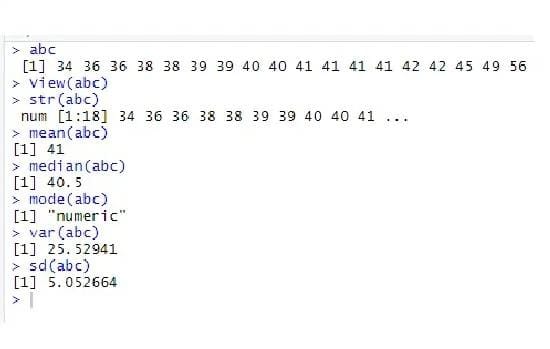
Confidence interval for 98%= 2.33

Confidence interval for 96% = 2.05

**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?



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

Ans: normally distributed

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

ANS: positive /right skewness

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

Ans : negative/last skewnwss

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

Ans: higher peakedness and lower tail

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

Ans: lower peakedness and higher tail

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



What can we say about the distribution of the data?

Ans: median >mean i.e left skew

What is nature of skewness of the data?

Ans: left skew

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

Ans:(q3-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: both the boxplots approximately normally distributed and the median ofboth box plots are same point (i.e.262.5)

Q 20) Calculate probability from the given dataset for the below cases

Data \_set: Cars.csv

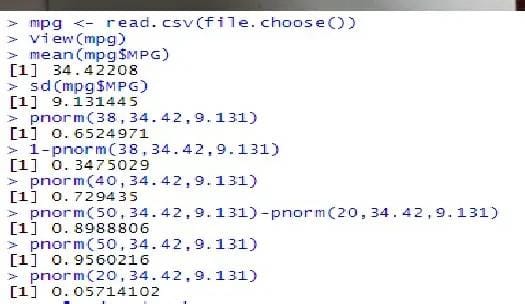
Calculate the probability of MPG of Cars for the below cases.

MPG <- Cars$MPG

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

c. P (20<MPG<50)

Ans:

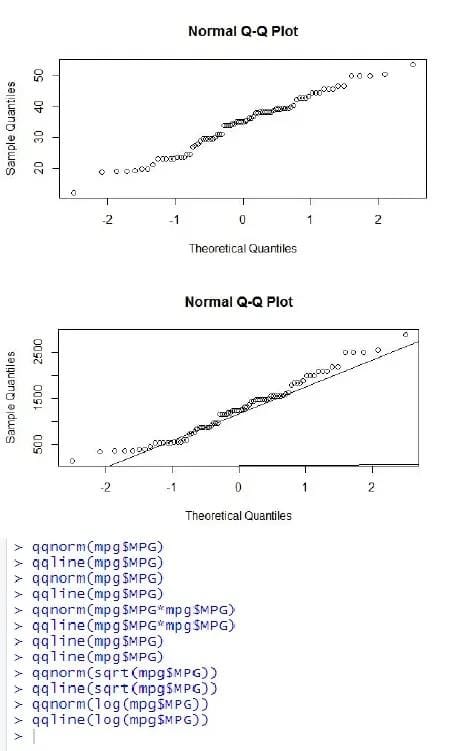


Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans:

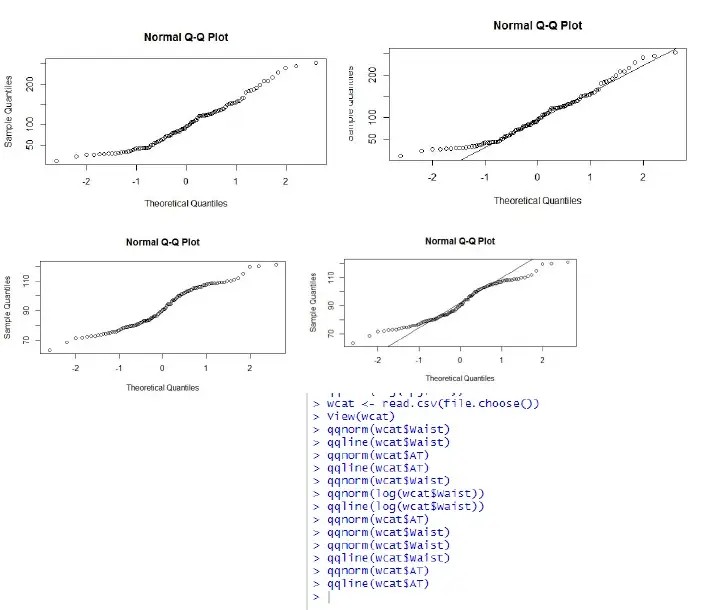


1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

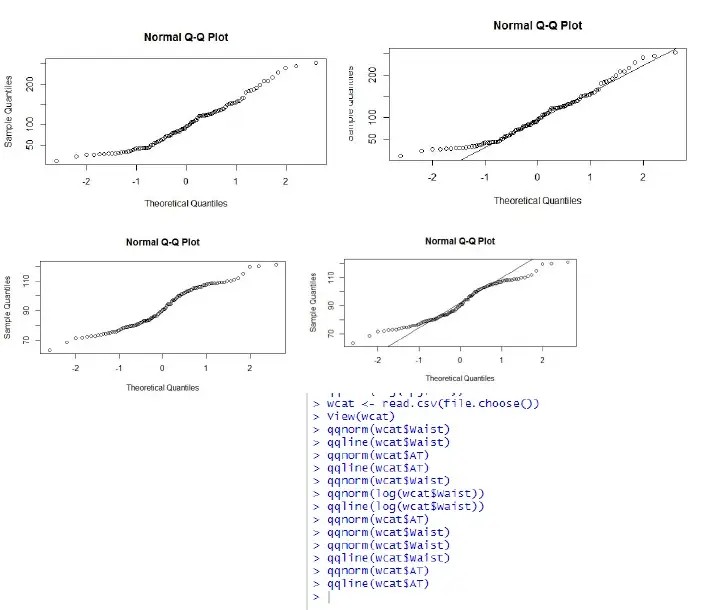
Dataset: wc-at.csv

Ans:

At plots



Waist plots



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

Ans:

Z score of 90% confidence interval is 1.65

Z score of 94% confidence interval is 1.55

Z score of 60% confidence interval is 0.85

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

Ans:

For 95%= 1.96

For 96%= 2.5

For 99% = 2.47

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: mu=270

n=18

xbar=260

sigma=90

z=x-mu/sigma=260-270/90=-0.11

pnorm(-0.11)=0.4562

p=45%

T=x-mu/s/sqrt(n)

=260-270/90/sqrt(18)=-0.4714

Pt-(0.4714, 17)=0.3216

P=32%