|  |  |
| --- | --- |
| 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, c, 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 | Nominal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Interval |
| 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 | Ratio |
| SAT Scores | Interval |
| Years of Education | Ratio |

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

Possible outcomes for the three coins tossed : (HHH, HHT, HTH, THH, HTT, THT, TTH, TTT )

Probability that two heads and one tail is : (HHT, HTH, THH)

Hence P = 3/8

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

S = { (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) }

Possible total number of outcomes = 6 x 6 = 36

1. Equal to 1 : When the sum is equal to '1' the Probability will be Zero.

P(E) = n(E) / n(S) = 0 / 36 = 0

1. Less than or equal to 4

P(AorB)=P(A)+P(B)

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

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?

Total number of events {(R,R),(R,G),(R,B),(G,R),(G,G),(G,B),(B,R),(B,G),(B,B)}=9

Interested number of events =4

Probability =4/9

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

Expected number of candies for a randomly selected child is given by

∑Xp(X)=0.015+0.8+1.95+.025+.06+.24=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: mean(Points): 3.596563

mean(Score): 3.21725

mean(Weigh): 17.84875

median(Points): 3.695

median(Score): 3.325

median(Weigh): 17.71

mode is NA as the given data set is numeric

variance(Points): 0.2858814

variance(Score): 0.957379

variance(Weigh): 3.193166

Standard deviation(Points): 0.5346787

Standard deviation(Score): 0.9784574

Standard deviation(Weigh): 1.786943

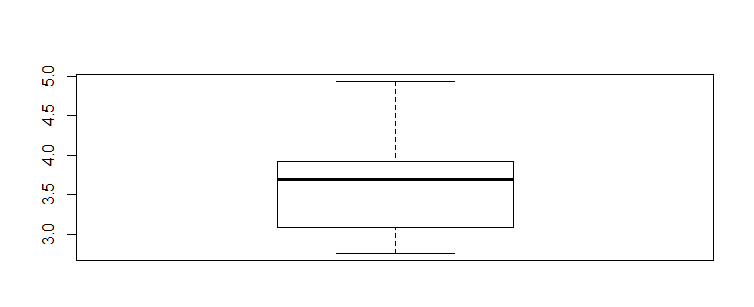
Range:

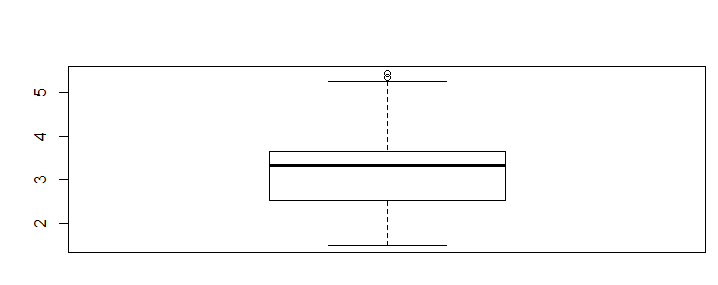
Points:2.76 4.93

Score: 1.513 5.424

Weigh: 14.5 22.9

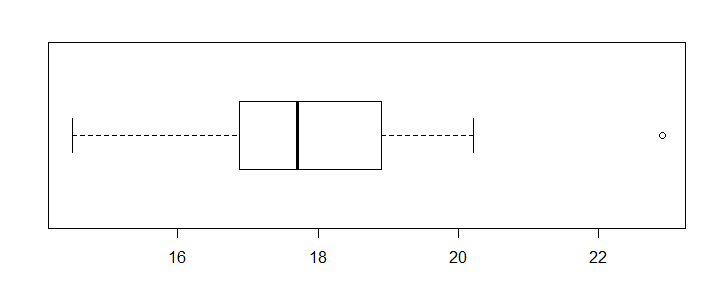
Boxplot for points , score and weigh





Outliers are there in score

$out

[1] 5.424 5.345

$out

[1] 22.9

Q8) Calculate Expected Value for the problem below

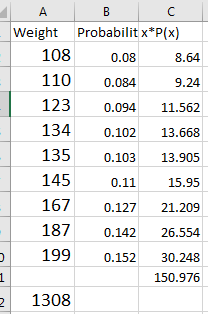
1. The weights (X) of patients at a clinic (in pounds), are

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?

Answer : The mean is:Expected Value=(108 + 110 + 123 + 134 + 135 + 145 + 167 + 187 + 199)/9 = 145.333.

Expected value =Sum(x.p(x)) = 150.976



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

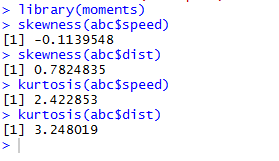
**Cars speed and distance**



**SP and Weight(WT)**

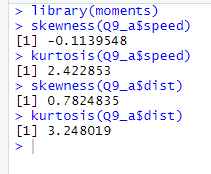


**Answer :**



Skewness for speed= is -0.113, skewness value is negative so it is left skewed. Since magnitude is slightly greater than 0 it is slightly left skewed

And for distance= 0.7824835, right skewed (Positive) slight magnitude to right.

****

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



The maximum of the data points is present in the range 50-100 with frequency 200.

minimum range of weight is 400 and its frequency around 0-10.

Skewness- we can notice a long tail towards right so it is heavily right skewed.

Boxplot : Median is less than mean right skewed and we have outlier on the upper side of box plot and there is less data points between Q1 and bottom point.

**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 ?

Sol: 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?

Answer: Mean= 41, Median= 40, variance= 24.111, Standard deviation= 4.910

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

Answer: Symmetrical

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

Right skewed

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

Left Skewed

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

The data is notmally distributed and kurtosis value is 0.

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

The distribution of the data has lighter tails and a flatter peaks than the normal distribution.

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



What can we say about the distribution of the data?

Answer: Let’s assume above box plot is about ages of the students in a school.

50% of the people are above 10 yrs. old and remaining are less.

And students whose age is above 15 are approx. 40%.

What is nature of skewness of the data?

Left skewed, median is greater than mean.

What will be the IQR of the data (approximately)?   
Approximately= -8

Q19) Comment on the below Boxplot visualizations?



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

Sol: By observing both the plots whisker’s level is high in boxplot 2, mean and median are equal hence distribution is symmetrical.

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

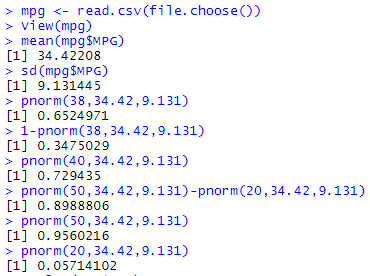
Data \_set: Cars.csv

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

MPG<- Cars$MPG

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

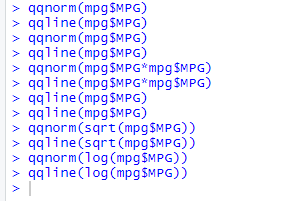
c. P (20<MPG<50)

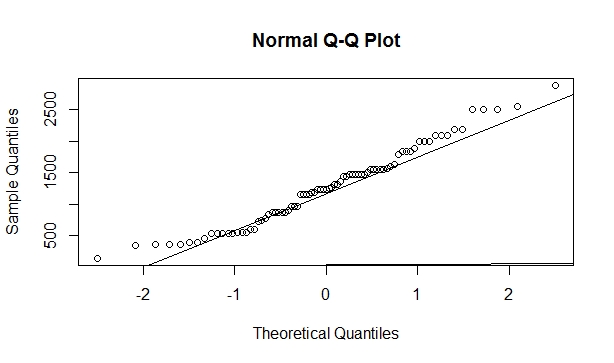
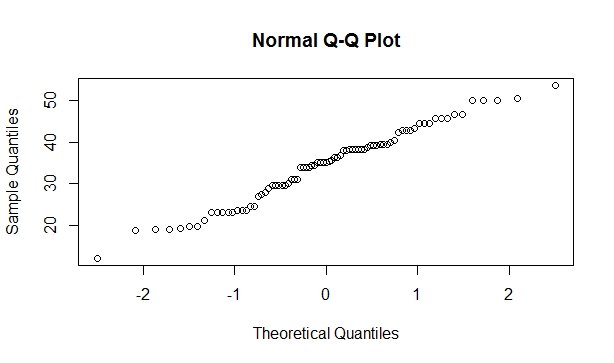


Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

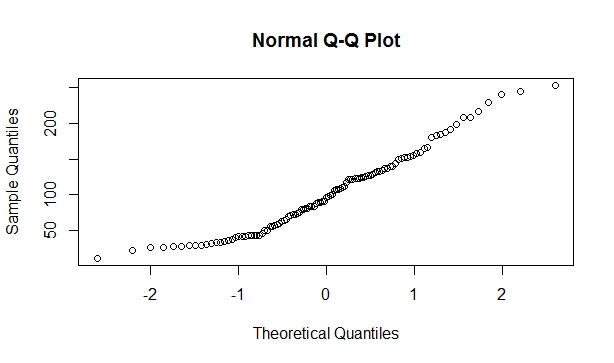


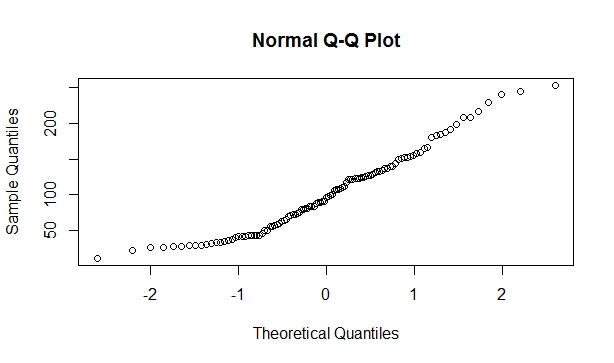


1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

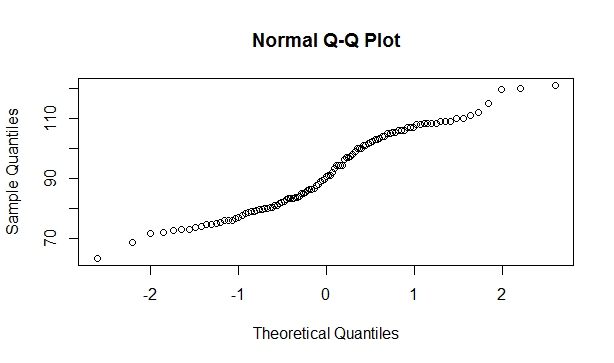
Dataset: wc-at.csv

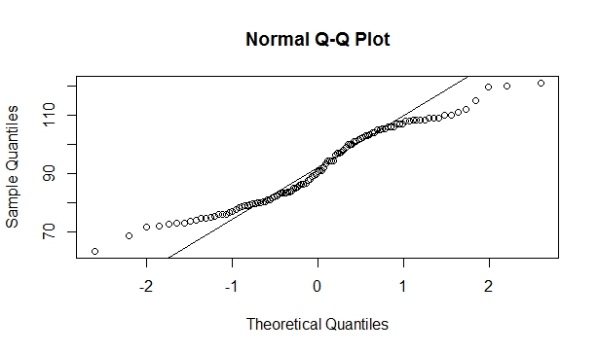
At plots

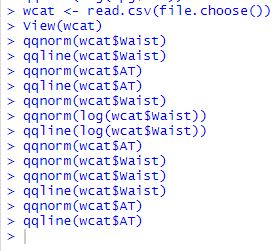




Waist plots







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

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

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

Answer :

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%