Q.1. Solution :

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

Q.2.

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 | 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 | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Ratio |
| Years of Education | Interval |

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

All the events are independent with the same probability of 1/2. The probability of getting two heads and a tail is the same as the probability of getting three heads, three tails, two tails and a head, … =1/8=0.125

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

a) sum equal to 1 = 0 {i.e. not possible that sum always exceed to 1}

Required probability=0/36 =0

b) sum equal to 4 = 3 {i.e. (1,3)(2,2)(3,1)}

required probability =3/36 =1/12

c) sum less than 13 = 36 {i.e. there are maximum sum is 12 of(6,6)}

required probability =36/36 =1

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?

There are 7 balls originally with 2 of them blue so the probability of the first ball not being blue is 5/7. This leaves 6 balls with 2 blue. The probability of the second ball not being blue assuming that the first wasn’t is 4/6. The probability that neither ball drawn was blue is (5/7)\*(4/6)=20/42=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

Expected number of candies for a randomly selected child

= 1 \* 0.015  + 4\*0.20  + 3 \*0.65  + 5\*0.005  + 6 \*0.01  + 2 \* 0.12

= 0.015 + 0.8  + 1.95 + 0.025 + 0.06 + 0.24

=  3.090

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

* For Points , Score ,Weight>

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

Solution:

MEAN:

Points 3.596563

Score 3.217250

Weigh 17.848750

MEDIAN:

Points 3.695

Score 3.325

Weigh 17.710

MODE:

Points Score Weigh

0 3.07 3.44 17.02

1 3.92 NaN 18.90

Standard deviation:

Points 0.534679

Score 0.978457

Weigh 1.786943

VARIANCE:

Points 0.285881

Score 0.957379

Weigh 3.193166

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?

Expected Value=(108+110+123+134+135+145+167+187+199)/9

=145.333.

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

a) Cars speed and distance

Use Q9\_a.csv

Solution:

Skewness:

speed -0.117510

it means the data in the speed is fairly symmetrical and it is not normaly distributed.and the negative skewness refers that the data is having longer or flatter tail on the left side.

dist 0.806895

here the dist data is having the negative skewness which tells that the is having longer or flatter tail on the right side. which also refers that the data will be greater than the median.

Kurtosis:

speed -0.508994

dist 0.405053

here in the both the variables which are speed and dist are having the kutosis value less than 3.which means that dataset is having the lighter tails than a normal distribution.

b)SP and Weight(WT)

Use Q9\_b.csv

Skewness:

SP 1.611450

Here the data are highly skewed .according to the thumbs rule.

WT -0.614753

it means the data in the speed is fairly symmetrical and it is not normaly distributed.and the negative skewness refers that the data is having longer or flatter tail on the left side.

KURTOSIS:

SP 2.977329

WT 0.950291

here in the both the variables which are speed and dist are having the kutosis value less than 3.which means that dataset is having the lighter tails than a normal distribution.

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



The weight between 50-100 are having the highest no in the histogram and age less than 50 is having less no ,but the weight grater than 100 are getting decreased in numbers.



The box plot is saying that there are outlier in the data after the 4th quartile value.

**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?

**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?

MEAN: The sum of all the data entries divided by the number of entries

So:

(**34+36+36+38+38+39+39+40+40+41+41+41+41+42+42+45+49+56)/18=41**

**Median=** The value that lies in the middle of the data when the data set is ordered. If the data set has an odd number of entries, then the median is the middle data entry. If the data has an even number of entries, then the median is obtained by adding the two numbers in the middle and dividing result by two.

So:

Middle values are 41 and 40.

(40+41)/2= 40.5

Mode: The data entry that occurs with the greatest frequency

So: 41 is the value which is having the highest frequency.

Standard deviation:

The standard deviation measure variability and consistency of the sample or population.

After the calculation we got he standard deviation value is =4.58

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

Skewness can be positive or negative depending upon the situation of mean median value. But there will be no skewness when the median and mean values are equal.

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

The nature of skewness will be positively skewed if mean is greater than median.

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

The nature of skewness will be negatively skewed if median is greater than mean.

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

A positive value of kurtosis tells that you have heavy tails on the data which means a lot of data in your tails.

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

A negative value of kurtosis tells that you have light tails which means that very little data in your tails.

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



What can we say about the distribution of the data?

:::::Most of data like of the data is containing in the Ist quartile .like from 1 to 10.and the 2nd quartile and 3rd quartile value containing the value 10 to 18.and maximum value is 19 in this case. there are no outliesrs.and the median of the data lies towards the right side.and on to 15 in this case.

What is nature of skewness of the data?

::::Here the median is right most part o the data so the skewness is towards the left side means most of the values are towards the left.

What will be the IQR of the data (approximately)?   
::::After analyzing the ist quartile and 3rd quartile value the IQR will be = 11(approx.)

Q19) Comment on the below Boxplot visualizations?



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

Here in the 1st boxplot the median line is downward from the centre so it can be called as rightly skewed, whereas in the 2nd box plot the median line is in the centre so it can be said as there is no skewness. and in the 1st box plot the values falls in every quartile is much lesser data as compare to the 2nd boxplot .

There are no outliers in the both the boxplots. In both the plots the data is not uniformly distributed. Because in the 2nd boxplot the data in the 3rd quartile and 2nd quartile seems to be of equal frequency whereas in the 1st boxplot this is not the case.

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)
  3. P (20<MPG<50)

For case a) p(MPG>38)=0.40

For case b) p(MPG<40)=0.75

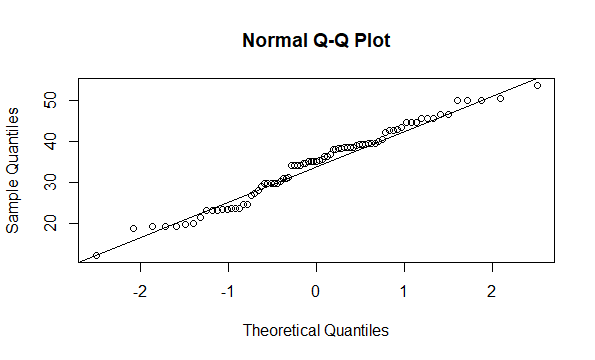
For case c)p(20<MPG>50)=0.85

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

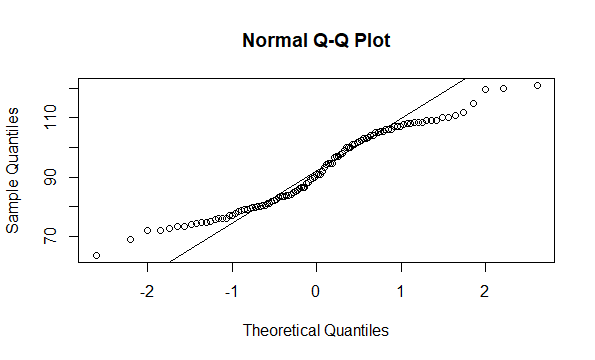
I will be checking whether MPG follows the normal distribution or not using the qqplot.

So analysing this the data in the mpg Approximately showing straight line so we can say that the MPG follows the normal distribution.

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

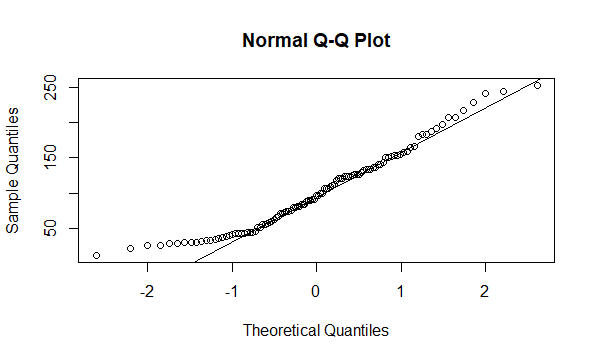
Dataset: wc-at.csv

For Waist:



As we can see it Is not following the normal distribution because data points are not forming the straight line.

For AT:

 as we can see here most of the data points are falling towards the straight line so we can say that it is following the normal distribution.

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

We can calculate the z score of the sample given.in here there is no sample given so we need to give the value for 90% =1.64,for 94%= 6% and for 60% value is 0.9.we need to put these values in to the formula of z score with the population sample respectively.

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

For the 95% confidence interval with sample size 25 value of z score will be=1.96

For the 96% confidence interval with sample size 25 value of z score will be=2.1

For the 99% confidence interval with sample size 25 value of z score will be=2.3

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== 32.17%