|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Discrete |
| Weight of a person | Continuous variable |
| Weight of Gold | Continuous variable |
| Distance between two places | Continuous variable |
| Length of a leaf | Continuous variable |
| Dog's weight | Continuous variable |
| 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 | Interval |
| 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 | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Nominal |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Interval |
| Years of Education | Interval |

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

Ans: 3 out of 8 (HHH,HHT,HTH,THH,HTT,THT,TTH,TTT)

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

1. Equal to 1

Ans: 0

1. Less than or equal to 4

Ans: Interested Event=(11,12,13,21,22,31)

Therefore Reqd. Probability = 6/36

1. Sum is divisible by 2 and 3

Ans: Interested Event = (15,24,33,42,51,66)

Therefore Reqd. Probability = 6/36

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 events 7c2 =21

None of the balls is blue 5c2

None of the balls drawn blue :

P = 5C2 / 7C2 = (5 \* (5-1) /2) / (7 \* (7-1)/2) = 20/42 = 47.61

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: **E(X) = ∑X \* P(X)**

X = Number of candies, P(X) = probability associated

Expected value = 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 datasetFor Points,Score,Weigh>

Find Mean, Median, Mode, Variance, Standar/

d Deviation, and Range and also Comment about the values/ Draw some inferences.

**Use Q7.csv file**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Points | Score | Weight |
| Mean | 3.596563 | 3.229281 | 17.84875 |
| Median | 3.695 | 3.325 | 17.71 |
| Mode | 3.92 | 3.44 | 17.02 |
| Variance (Population) | 0.276948 | 0.893394 | 3.09338 |
| Std Deviation (Population) | 0.526258 | 0.945195 | 1.758801 |
| Range | 2.76-4.93 | 1.513 -5.424 | 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:** Expected value of weight of patient =**EV = ∑X/n**

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

= 1308/9

= 145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans:**

|  |  |  |
| --- | --- | --- |
|  | Speed | Distance |
| Skewness | -0.11751 | 0.7824835 |
| Kurtosis | 2.422853 | 3.248019 |

**Speed:** The speed distribution is negative i.e. left skewed.

Distribution with negative kurtosis value indicates that the distribution has heavier tails then normal distribution with flat peak

**Distance:** The distance distribution is positive i.e. right skewed.

Distribution with positive kurtosis value indicates that the distribution has heavier tails then normal distribution with Sharp peak.

**SP and Weight(WT)**

**Use Q9\_b.csvxz**

**Ans:**

|  |  |  |
| --- | --- | --- |
|  | SP | WT |
| Skewness | 1.61145 | -0.61475 |
| Kurtosis | 5.72321 | 3.819466 |

**SP:** The sp distribution is positive i.e. right skewed.

SP Distribution with positive kurtosis value indicates that the distribution has heavier tails then normal distribution with Sharp peak.

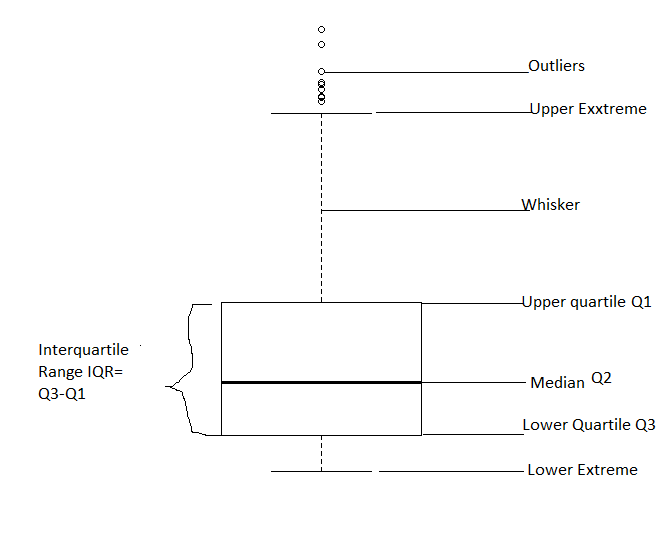
**WT:** The wt distribution is negative i.e. left skewed.

WT Distribution with positive kurtosis value indicates that the distribution has heavier tails then normal distribution with Sharp peak.

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



**Ans:** The above histogram infers that most of the weights are between 50-100 with mean of about 75 approx. Weight gradually decreases, with minimal in the range of 300-100. The distribution is right skewed. Mean>Median



The above boxplot suggests that the distribution has lots of outliers towards 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: 94% Coinfidence Interval

x-bar = 200, sd= 30, n=2000

Interval Estimate = X-bar ± Z\*Sd/sqrt(n)

=200 ± 1.88\*30/sqrt(2000)

=**198.74 – 201.26**

>**98% Confidence:**

Interval Estimate = X-bar ± Z\*Sd/sqrt(n)

=200 ± 2.33\*30/sqrt (2000)

=**198.44-201.56**

>**96% Confidence:**

Interval Estimate = X-bar ± Z\*Sd/sqrt(n)

=200 ± 2.05\*30/sqrt (2000)

=**198.62-201.38**

**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.

|  |  |
| --- | --- |
| **Mean** | **41** |
| **Median** | **40.5** |
| **Variance** | **25.52941** |
| **Standard Deviation** | **5.052664** |

1. What can we say about the student marks?

**Ans:** The average marks scored by students is 41.Mostly the students have marks in the range 36 to 42.Very few have marks above 45.Marks follows normal distribution.Mean > Median, This implies that the distribution is slightly skewed towards right. No outliers are present.

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

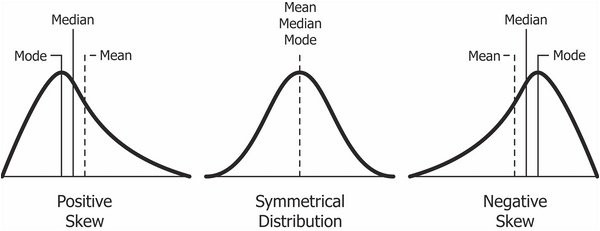
Ans: It’s a perfect symmetrical distribution, when mean, median of data are equal. Skewness =0

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

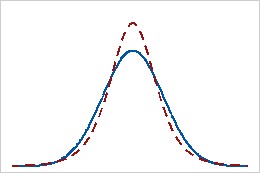
Ans: Mean > median the distribution is positively (Right) skewed.

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

Ans: For Median > mean the distribution is negatively(left) skewed.



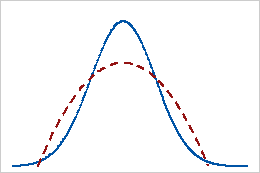
Q16) What does positive kurtosis value indicates for a data ?

Ans: Distribution with positive kurtosis value indicates that the distribution has heavier tails then normal distribution with Sharp peak.

Dotted Line: positive kurtosis value.

Solid line: Normal distribution

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

Ans: The distribution with negative kurtosis value indicates the distribution has lighter tail than normal distribution i.e Flat peak.

Dotted Line: Negative kurtosis value.

Solid line: Normal distribution

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



What can we say about the distribution of the data?

Ans: Not a Normal Distribution

What is nature of skewness of the data?

Ans:Left Skewed

What will be the IQR of the data (approximately)? Ans**:** IQR = 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 are Normally Distributed

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.

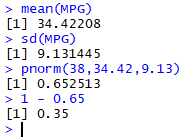
MPG <- Cars$MPG

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

c. P (20<MPG<50)

Ans:

* 1. (MPG>38)



P(MPG>38) = **0.35**

* 1. P(MPG < 40)



P(MPG<40) = **0.73**

* 1. P(20 < MPG < 50)

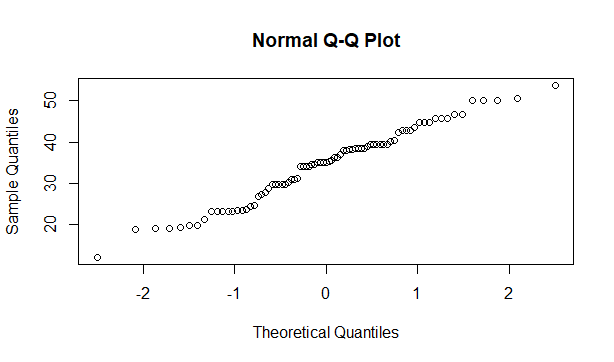


P(20<MPG<50) = 0.898

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

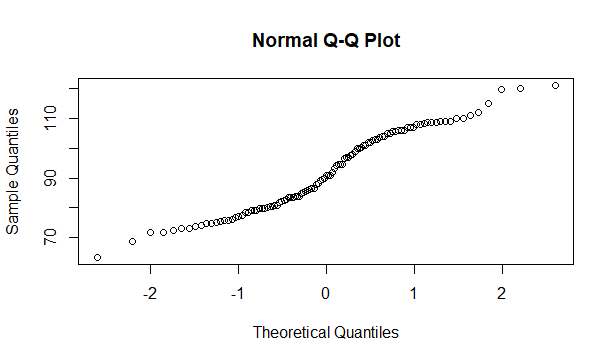
Dataset: Cars.csv



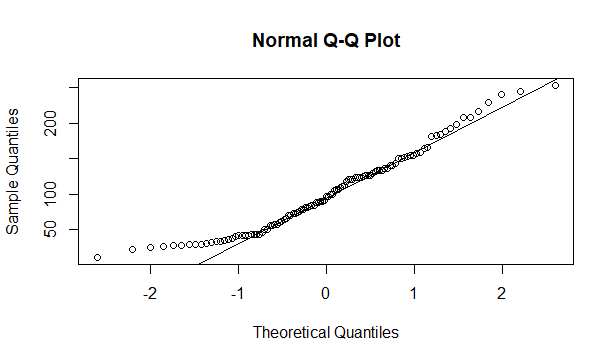
Follows Normal Distribution as indicated by the qqplot.

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

Dataset: wc-at.csv



Wc-at $waist follows Normal Distribution



Wc-at$AT follows Normal Distribution

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

Ans:

**90% Confidence:**

****

**94% Confidence:**



**60% Confidence:**



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

Ans:

**95% Confidence:**



**96% Confidence:**

****

**99% Confidence:**

****

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: df🡪 degrees of freedom

µ = 270, n= 18, x bar = 260, s = 90

t score = (x bar - µ)/(s/sqrt(n))

=(260 – 270)/(90/sqrt(18))

=-10/21.23

=-0.47

Required Probability = 0.32

