|  |  |
| --- | --- |
| 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 | String |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | Categorical (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 | 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) | Ordinal |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Ordinal |
| Religious Preference | Ordinal |
| Barometer Pressure | Interval |
| SAT Scores | Ratio |
| Years of Education | Ordinal |

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

3/8

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

1. Equal to 1- 0
2. Less than or equal to 4 – 1/6
3. Sum is divisible by 2 and 3 - 5/36

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? 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 = 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**

Points Score

Length:32 Min. :2.760 Min. :1.513

Class :character 1st Qu.:3.080 1st Qu.:2.581

Mode :character Median :3.695 Median :3.325

Mean :3.597 Mean :3.217

3rd Qu.:3.920 3rd Qu.:3.610

Max. :4.930 Max. :5.424

Weigh

Min. :14.50

1st Qu.:16.89

Median :17.71

Mean :17.85

3rd Qu.:18.90

Max. :22.90

mode\_points

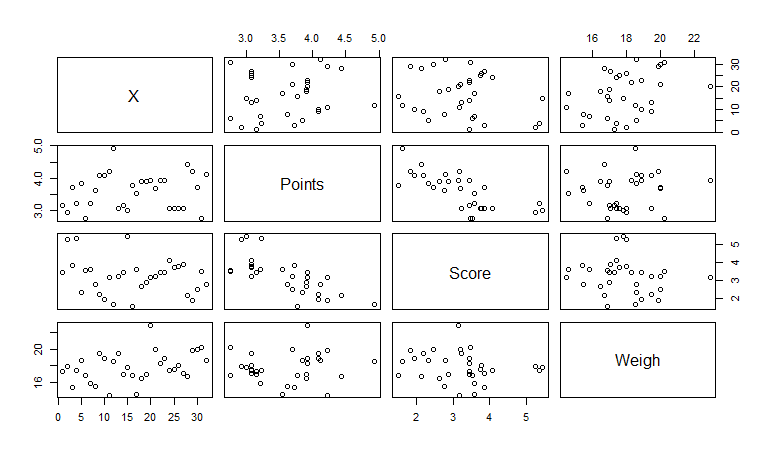
[1] 3.92

mode\_score

[1] 3.44

mode\_weigh

[1] 17.02



Points & Score Are negatively Correlated

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?

set<-c(108, 110, 123, 134, 135, 145, 167, 187, 199)

> mean(set)

[1] 145.3333

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

**Cars speed and distance**

**Use Q9\_a.csv**

skewness(Q9\_a$speed)

[1] -0.1105533

skewness(Q9\_a$dist)

[1] 0.7591268

kurtosis(Q9\_a$speed)

[1] -0.6730924

kurtosis(Q9\_a$dist)

[1] 0.1193971

**SP and Weight(WT)**

**Use Q9\_b.csv**

> skewness(Q9\_b$SP)

[1] 1.552258

> skewness(Q9\_b$WT)

[1] -0.5921721

> kurtosis(Q9\_b$SP)

[1] 2.583072

> kurtosis(Q9\_b$WT)

[1] 0.7257402

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



Positively Skewed & Heavily tailed



The data sets has Some outliers

**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?

1. Confidence Interval = 98%

Standard error = 30 / root 2000 = 0.67

Compute alpha (α): α= 1 – 0.98 = 0.02

Critical probability = 1 – α/2 = 0.99

Degree of freedom (df) = 2000 – 1 = 1999

Critical value for probability = 0.99 & DF = 1999 is 2.328

Computing margin of error = 2.328 \* 0.67 = 1.55

98% Confidence Interval (CI) = 200 +/- 1.55

2. CI = 96%

Standard error = 30 / root 2000 = 0.67

Compute alpha (α): α= 1 – 0.96 = 0.04

Critical probability = 1 – α/2 = 0.98

Degree of freedom (df) = 2000 – 1 = 1999

Critical value for probability = 0.98 & DF = 1999 is 2.055

Computing margin of error = 2.055 \* 0.67 = 1.37

98% Confidence Interval (CI) = 200 +/- 1.37

3. CI = 94%

Standard error = 30 / root 2000 = 0.67

Compute alpha (α): α= 1 – 0.94 = 0.06

Critical probability = 1 – α/2 = 0.97

Degree of freedom (df) = 2000 – 1 = 1999

Critical value for probability = 0.97 & DF = 1999 is 1.882

Computing margin of error = 1.882 \* 0.67 = 1.26

98% Confidence Interval (CI) = 200 +/- 1.26

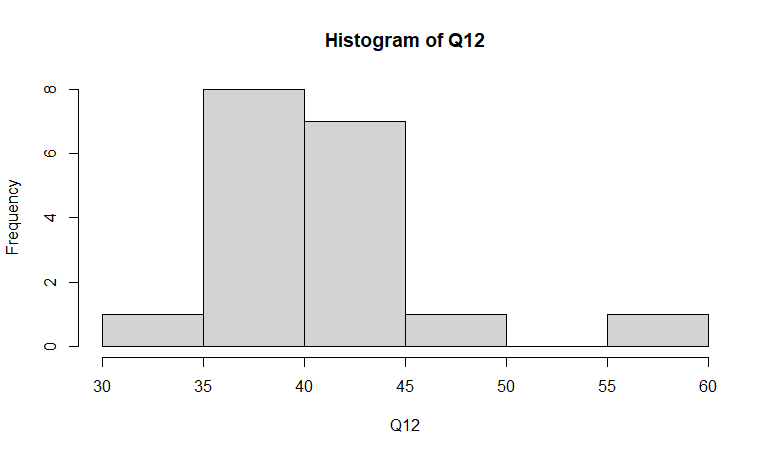
**Q12)** Below are the scores obtained by a student in tests

**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?

Min. 1st Qu. Median Mean 3rd Qu. Max.

34.00 38.25 40.50 41.00 41.75 56.00



> var(Q12)

[1] 25.52941

> sd(Q12)

[1] 5.052664

Positively Skewed

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

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

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

Q16) What does positive kurtosis value indicates for a data ? Positive values of kurtosis indicate that a distribution is peaked and possess thick tails

Q17) What does negative kurtosis value indicates for a data? Negative values of kurtosis indicate that a distribution is flat and has thin tails

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



What can we say about the distribution of the data?

50% of data lies between Q1 = 10’ to Q3=18’

Median = 15’

Negatively Skewed

Mean is not equal to Median is not equal to Mode

What is nature of skewness of the data? Negative

What will be the IQR of the data (approximately)? 10’ – 18’  
  
  
Q19) Comment on the below Boxplot visualizations?



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

Median for both = 262.5

1st dataset is more dense. The values of dataset is close to median compared to in 2nd dataset.

Kurtosis is more in 1st data compared to 2nd data.

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)

> mean(Q20$MPG)

[1] 34.42208

> sd(Q20$MPG)

[1] 9.131445

> pnorm(38,34.42,9.13)

[1] 0.652513

Hence P(MPG>38)= 1 – 0.65 = 0.35 = 35%

* 1. P(MPG<40)

> pnorm(40,34.42,9.13)

[1] 0.7294571

c. P (20<MPG<50) = 0.9 =90%

> pnorm(20,34.42,9.13)

[1] 0.05712119

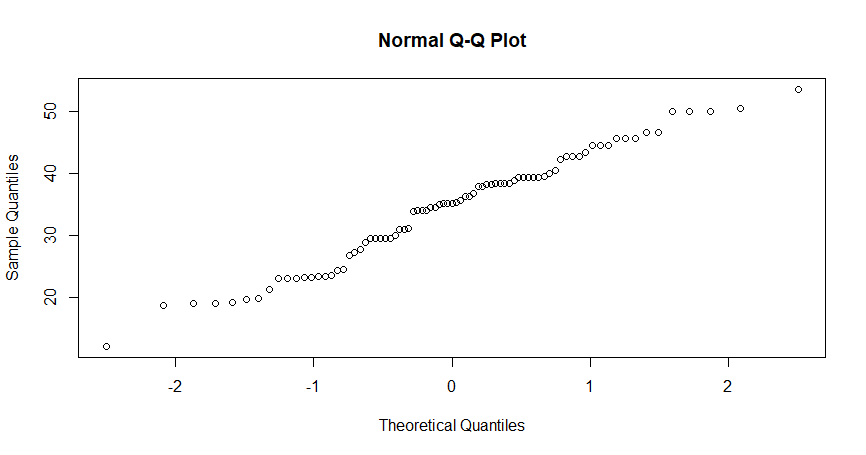
> pnorm(50,34.42,9.13)

[1] 0.956039

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv



Shapiro-Wilk normality test

data: Q21$MPG

W = 0.97797, p-value = 0.1764

> kurtosis(Q21$MPG)

[1] -0.7054604

> skewness(Q21$MPG)

[1] -0.1714104

> mean(Q21$MPG)

[1] 34.42208

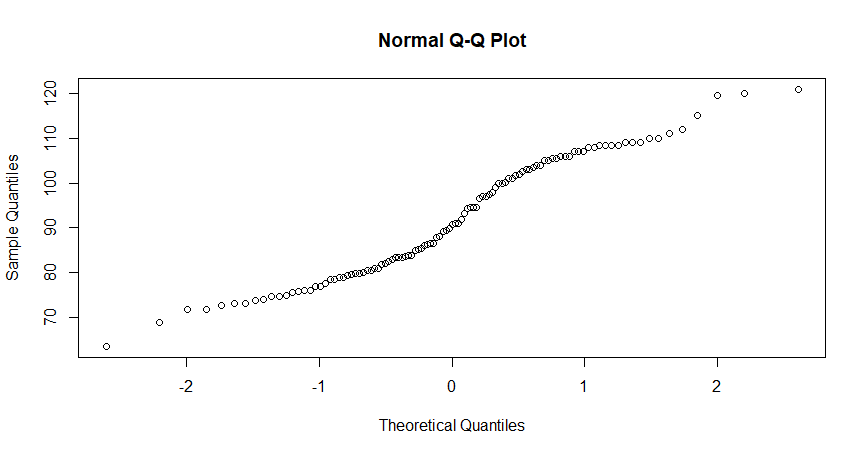
> median(Q21$MPG)

[1] 35.15273

From above results I can conclude data is normally distributed

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

Dataset: wc-at.csv



data: Q21b$Waist

W = 0.95586, p-value = 0.00117

> kurtosis(Q21b$Waist)

[1] -1.141846

> skewness(Q21b$Waist)

[1] 0.130389

> mean(Q21b$Waist)

[1] 91.90183

> median(Q21b$Waist)

[1] 90.8

Data is Normally distributed

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

For 90% CI , Z -Score = 1.645

For 94% CI, Z – Score = 1.88

For 60% CI, Z -Score = 0.842

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

95% CI , T score = 2.064

96% CT, T score = 2.20

99% CI, T score = 2.797

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

T score = (260-270)/(90/root 18) = -0.471

Df = 18-1 = 17

pt(-0.471,17)

[1] 0.321814