| 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 | nominal |
| 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 | Ordinal |
| 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 | Interval |
| Blood Group | Nominal |
| Time Of Day | Ratio |
| Time on a Clock with Hands | Ratio |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Ratio |
| Years of Education | Interval |

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

**Ans:** S={HHH,TTT,HHT,HTH,TTH,THH,THT,HTT}

Even getting two heads and one tail is {HHT,THH,HTH}

Hence, Probability is **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

**Ans:**

a) If two dices were rolled, then total possible cases =36

Total Favourable cases (Having sum =1)=0

As minimum sum is 2 for outcome(1,1)

Hence, probability is 0

b)The sum is less than or equal to 4 the possible outcomes are (1,1), (1,2), (1,3), (2,1), (2,2), (3,1).

Hence, probability is **(6/36) =** **1/6.**

c) no of sums divisible by 2&3 are 2 i.e 6,12

Possible ways for 6 to be the sum are (1, 5), (2, 4), (3, 3), (4, 2), (5, 1)

Possible ways for 12 to be the sum are (6, 6)

Total possible ways are 6 and no of possible sums are 36

Hence, Probability is **(3/36)=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:** The probability of none of the balls drawn is blue is = [5C2/7C2]= 10/21

| 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 |

**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 A – probability of having 1 candy = 0.015.

Child B – probability of having 4 candies = 0.20

**Ans:** 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= 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.

**Ans:**

| Info | Points | Score | Weigh |
| --- | --- | --- | --- |
| mean | 3.596563 | 3.217250 | 17.848750 |
| Std | 0.534679 | 0.978457 | 1.786943 |
| Med | 3.695000 | 3.325000 | 17.710000 |
| Var | 0.285881 | 0.957379 | 3.193166 |
| Mode | 3.07,3.92 | 3.44 | 17.02,18.90 |
| Range(min) | 2.760000 | 1.513000 | 14.500000 |
| Range(max) | 4.930000 | 5.424000 | 22.90000 |

# INFERENCE:

\*skewness of points and score are negative and skewness of weigh is positive

\*mean is greater than median only in weigh and in points and score the median is greater than mean

\*we have one outlier in weigh and score column.

\*it has 32 the rows and 4 columns

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 Weight of that patient?

**Ans:**

Expected value = ∑(probability\*value)

probability=1/9

Expected value=1/9(108+110+123+134+135+145+167+187+199) =145.33

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

**Cars speed and distance**

**Ans :**

**Cars speed &distance:**

## Kurtosis: skewness:

speed -0.50899 speed -0.117510

dist 0.405053 dist 0.806895

# INFERENCE:

**1)**speed has negative skewness.

2)distance has positive skewness.

3)distance has no outliers.

**SP and Weight(WT)**

## Kurtosis: skewness:

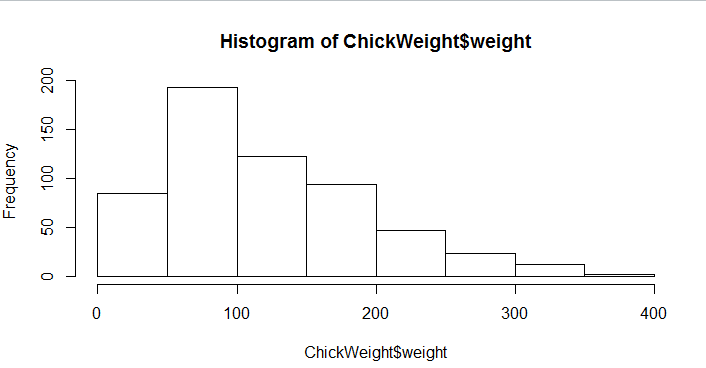
Speed 2.977329SP Speed 1.611450

WT 0.950291 WT -0.614753

# INFERENCE:

1. Speed is positively skewed.
2. Weight is negatively skewed.
3. Speed has positive kurtosis.
4. WT has positive kurtosis.

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





# INFERENCE:

1)From histogram we came to know the data is positively skewed.

2)From box plot we came to know there are more than two 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?

**ANS:**

94%=(143.57619175546247, 256.42380824453755)

98%=(130.2095637787748, 269.7904362212252)

96%=(138.38753268104531, 261.61246731895466)

**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.

**Ans:**

Mean = 41; Median = 40.50; Variance=25.5; Standard Deviation=5.05.

2)What can we say about the student marks?

**Ans:**

* Most of students have scored 35 to 40 (Higher frequency observed).
* 49 & 56 marks are the outliers for the given test data.
* 41 is the mode.

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

**Ans:** If the mean and median of data are equal then the skewness is zero.

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

**Ans:** If the mean is greater than the median then the skewess is positive.

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

**Ans:** If the median is greater than mean then it is negatively skewed.

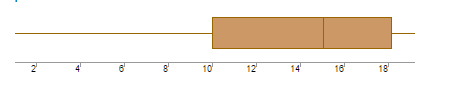
**Q16) What does positive kurtosis value indicates for a data ?**

**Ans:** An extreme positive kurtosis indicates a distribution where more of the values are located in the tails of the distribution rather than around the mean

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

**Ans:** A distribution is flatter (less peaked) when compared with the normal distribution, with fewer values in its shorter (i.e. lighter and thinner) tails.

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



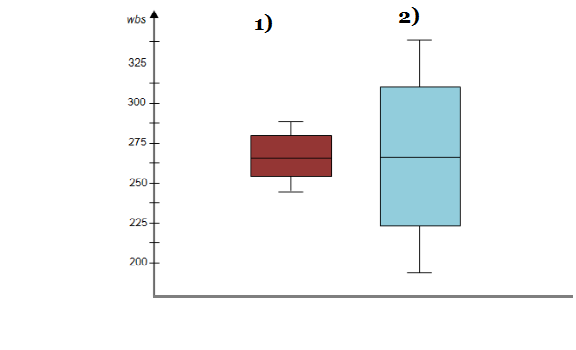
What can we say about the distribution of the data?

**ANS:** Most of the data lies between 10 to 18.

What is nature of skewness of the data?

**ANS:** The data is negatively skewed.

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

**ANS:** IQR=Q3-Q1=8  
  
**Q19)** Comment on the below Boxplot visualizations? 

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

**ANS:**

1)median for both the boxplot is equal.

2)both the boxplot data has mean=median which means it follows normal distribution

3)since both the data are normally distributed it has zero skewness.

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

Data \_set: Cars.csv

Calculate the probability of MPG ofCars for the below cases.

MPG<- Cars$MPG

**Ans:**

* 1. P(MPG>38)

0.34759394041453007

* 1. P(MPG<40)

0.7293498604157946

* 1. P (20<MPG<50)

0.8988689076273199

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**ANS:**

* The MPG of cars dataset follows (approximate) 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

**Ans:**

* Waist Circumference(Waist) does not follows normal distribution.
* Adipose Tissue (AD) follows (approximate) normal distribution.

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

**Ans:**

90%=1.644

94%=1.8807

60%=0.8416

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

**Ans:**

95%=2.0638985616280205

96%=2.1715446760080677

99%=2.796939504772804

**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:**

Probability for 18 Randomly selected bulbs of average life of no more than 260 days are 0.32167411.