|  |  |
| --- | --- |
| 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, 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 | **Interval** |
| Time on a Clock with Hands | **Interval** |
| Number of Children | **Ratio** |
| 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?

**ANS : 3/8= 0.375**

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) zero**

**b) 3/36 = 0.0833**

**c) 6/36 = 0.167**

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 : 5c2/7c2 = 10/21=0.476**

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 : 3.09** { Sigma I(1to n)=pi \*xi }

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

**ANS**  **: MEAN for Points =3.596563..**

**Score =3.217250**

**Weigh =17.848750**

**MEDIAN for Points = 3.695**

**Score =3.325**

**Weigh = 17.710**

**MODE for Points = 3.07 - 3.92**

**Score =3.44**

**Weigh =17.02 - 18.90**

**VARIANCE for Points=0.28588..**

**Score=0.957379…**

**Weigh=3.19316**

**STANDARD DEVIATION for Points = 0.534679**

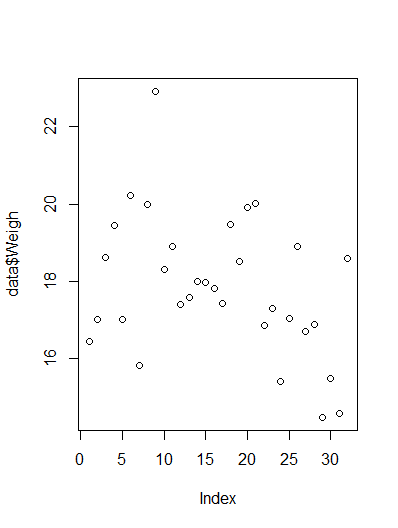
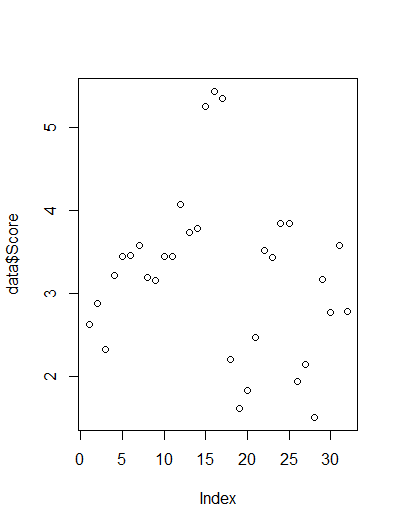
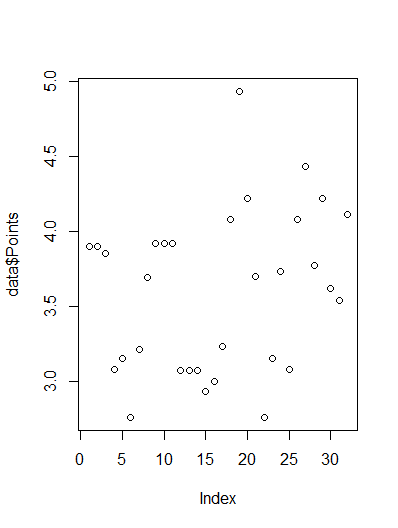
**Score = 0.978457**

**Weigh= 1.786943**

**RANGE for Points = 2.76 - 4.93**

**Score = 1.513 - 5.424**

**Weigh = 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 : 1/9(108+123+134+135+145+167+187+199)**

**=145.33…**

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

Cars speed and distance

Use Q9\_a.csv

**ANS : Skewness of speed = -0.1105533**

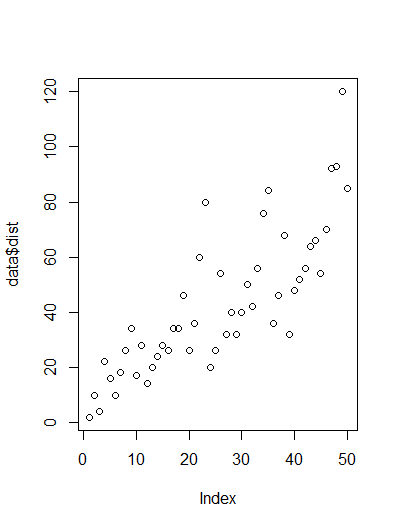
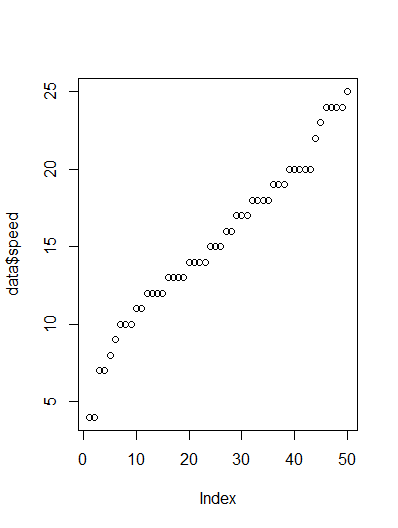
**Distance = 0.7591268**

**Kurtosis of speed = 2.422853**

**Distance = 3.248019**

**Inferences:**

1. **Skewness of Speed is negative in nature, which lets us know that the data is left skewed and also the median > mean.**
2. **Skewness of Distance is positive in nature, which lets us know that the data is right skewed and also the mean > median.**
3. **The Kurtosis values of both the data of Speed and Distance are positive in nature, which means the curve from the given data is sharp in nature and the values in the extremes have thin tails.**



SP and Weight(WT)

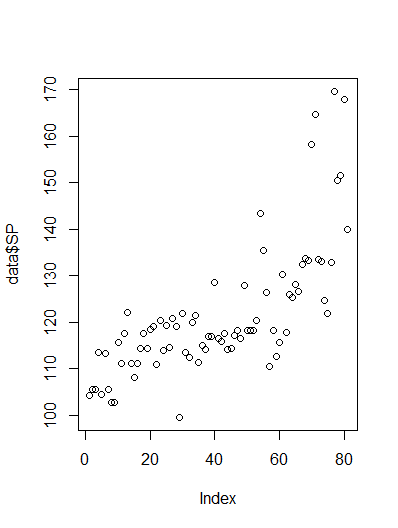
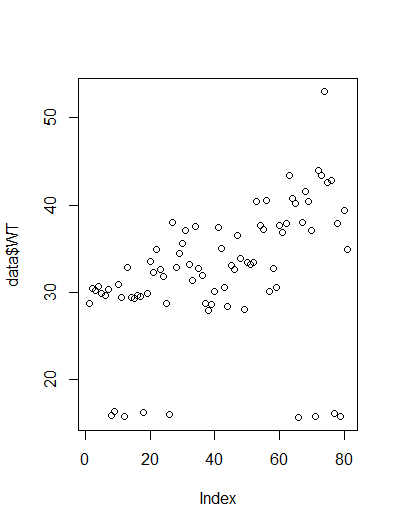
Use Q9\_b.csv

**ANS : Skewness of SP = 1.552258**

**WT = -0.5921721**

**Kurtosis of SP = 5.723521**

**WT = 3.819466**

****

**Inferences:**

1. **Skewness of WT is negative in nature, which lets us know that the data is left skewed and also the median > mean.**
2. **Skewness of SP is positive in nature, which lets us know that the data is right skewed and also the mean > median.**
3. **The Kurtosis values of both the data of SP and WT are positive in nature, which means the curve from the given data is sharp in nature and the values in the extremes have thin tails.**

Q10) Draw inferences about the following boxplot & histogram

**ANS : The skewness for the below histogram is right skewed in nature which depicts that the mean > median.**

**The peak of the below histogram has a sharp therefore the kurtosis is positive .**



**Boxplot :**



**Ans :**  **The data in the histogram shows the different trends of chick weight. The data in the histogram is right skewed as the “frequency” is diminishing towards right. The maximum number of chicks weight ranges between 50 and 100.**

**The box plot is also right skewed as the box is towards the left. There are outliers in the given data set which are represented as circles out side the box plot. These may be inspected before analysis.**

**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 : Given, Mean (μ)=200, Standard deviation (σ)=30**

**confidence interval, Zscore = 1.555**

**Max value x= μ+σz**

**Min value x= μ-σz**

**Class interval for 94% = 153.57619175546247, 246.42380824453755**

**Class interval for 98% = 130.2095637787748, 269.7904362212252**

**Class interval for 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 = sum of all obs./ total obs.= 738/18=41**

**Median = 40+41/2 = 81/2 = 40.5**

**SD = √28.13 = 5.303**

|  |  |  |
| --- | --- | --- |
| **S.NO** | **DEVIATION**  **(obs-mean)** | **DEVIATION^2** |
| **1** | **-6.5** | **42.25** |
| **2** | **-4.5** | **20.25** |
| **3** | **-4.5** | **20.25** |
| **4** | **-2.5** | **6.25** |
| **5** | **-2.5** | **6.25** |
| **6** | **-1.5** | **2.25** |
| **7** | **-1.5** | **2.25** |
| **8** | **-0.5** | **0.25** |
| **9** | **-0.5** | **0.25** |
| **10** | **0.5** | **0.25** |
| **11** | **0.5** | **0.25** |
| **12** | **0.5** | **0.25** |
| **13** | **0.5** | **0.25** |
| **14** | **2.5** | **0.25** |
| **15** | **2.5** | **6.25** |
| **16** | **5.5** | **30.25** |
| **17** | **9.5** | **90.25** |
| **18** | **16.5** | **272.25** |
| **19** | **total** | **506.5** |

2)What can we say about the student marks ?

**ANS : Since mean is greaterthan median the skewness of the data is positive.**

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

**ANS : It is perfectly symmetric.**

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

**ANS : Data is Right-skewed**

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

**ANS : Data is Left-skewed**

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

**ANS : Right-skewed (The peak of the normal distribution curve is sharpe) .**

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

**ANS : Left-skewed (The peak of the normal distribution curve is flat) .**

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



What can we say about the distribution of the data?

**ANS : 1. The IQR lies mostly between 10 & 18.**

**2. Median is close to Upper quartile.**

What is nature of skewness of the data?

**ANS : Median is closer to Upper Quartile . Therefore , the data is negatively skewed (left skewed).**

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

**ANS : Q**3**-Q**1 **= 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 : The median of both the boxplots are on the 50% margin of the IQR therefore there is no skewness and the data is normally distributed. We can observe that there are no outliers in both the data sets.**

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)

**ANS :**  **a. P(MPG>38) = 0.347**

**b. P(MPG<40) = 0.729**

**c. P(20<MPG<50) = 0.898**

**Refer “ASSIGNMENT-01-Q20 BASIC STATISTICS (LEVEL-1).ipynb” file**

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**ANS : No, the data of MPG is not normally distributed because the mean, median and mode are not equal and the skewness is also not equal to 0.**

**Mean =34.42, Median=35.15, Mode=29.629 & skewness = -0.17**.

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

Dataset: wc-at.csv

**ANS : No, either of the data in the respective columns are normally distributed because the mean, median and mode are not equal and the skewness of the data is also not equal to 0.**

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

**ANS : Confidence Level Z-Score 90% = 1.644**

**Confidence Level Z-Score 94% = 1.750**

**Confidence Level Z-Score 60% = 0.841**

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

**ANS : Confidence Level T-Score 95% = 2.063**

**Confidence Level T-Score 96% = 2.171**

**Confidence Level T-Score 99% = 2.796**

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 : T-Score = 260-270/(90/18^0.5) = -0.47**

**P(X<260) = Stats.t.cdf(-0.47,17) = 0.32**

**Refer “ASSIGNMENT-01-Q24-BASIC-STATISTICS(LEVEL-1).ipynb” file**