Q1) Identify the Data type for the Following:

|  |  |
| --- | --- |
| 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 |

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 | Ratio |
| Weight | Interval |
| Hair Color | Nominal |
| Socioeconomic Status | Nominal |
| Fahrenheit Temperature | Ratio |
| Height | Interval |
| Type of living accommodation | Ordinal |
| Level of Agreement | Nominal |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Interval |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Nominal |

|  |  |
| --- | --- |
| Number of Children | Nominal |
| Religious Preference | Ordinal |
| Barometer Pressure | Ratio |
| SAT Scores | Interval |
| Years of Education | Interval |

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

Ans : 3/8

Solution:

P (Two heads and one tail) = N (Event (Two heads and one tail)) / N (Event (Three

coins tossed))

= 3/8 = 0.375 = 37.5%

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

Total Number of possible outcomes =6^2 = 36

1. Equal to 1 = 0 (There are two dices so sum =1 is not possible)
2. Less than or equal to 4 :

Ans : 1/6

Sol :

Possible outcomes for sum less than 4 : (1,1),(1,2),(1,3),(2,1),(3,1),(2,2)

Total outcomes =36

Prob(sum<4) = 6/36 =1/6

1. Sum is divisible by 2 and 3:

Ans : 1/6

Sol : Possible outcomes for Sum is divisible by 2 and 3:

(1,5),(5,1),(2,4),(4,2), (3,3),(6,6)

Total outcomes =36

Prob(sum<4) = 6/36 =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-10/21

Sol :

1. There are a total of 2 + 3 + 2 = 7 balls in the bag.
2. We are asked to find the probability that 2 balls drawn at random are both non-blue.
3. The number of ways to choose 2 balls from 7 balls is 7C2 = 21.
4. The number of ways to choose 2 non-blue balls from 5 non-blue balls is 5C2 = 10.
5. Therefore, the probability that none of the balls drawn is blue is 10/21 = 2/7.

6) 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

Sol:

0.015+0.8+1.95+0.025+0.06+0.24 =3.09

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

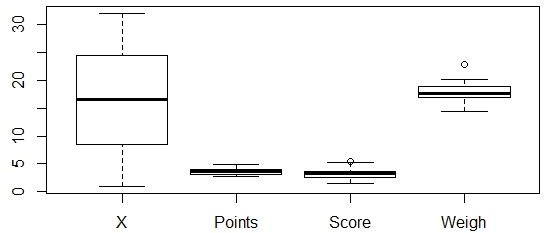
- For Points,Score,Weigh>

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

**Answer-**

Attached Q7\_Solution.ipynb for complete solution

|  |  |  |  |
| --- | --- | --- | --- |
|  | Points | Score | Weight |
| Mean | 3.596 | 3.217 | 17.848 |
| Median | 3.695 | 3.325 | 17.71 |
| Mode | 3.891 | 3.54 | 17.43 |
| Variance | 0.285 | 0.957 | 3.19 |
| Standard  Deviation | 0.534 | 0.978 | 1.786 |
| Range | 2.76,4.93 | 1.513,5.424 | 14.5,22.9 |



Q8) Calculate Expected Value for the problem below

a) 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 : 145.3333

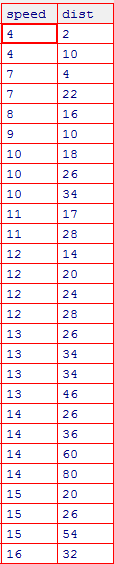
Solution:

Expected value = Sum (X \* Probability of X)

= (1/9)(108)+ (1/9)(110)+ (1/9)(123)+ (1/9)(134)+ (1/9)(145)+ (1/9)(167)+ (1/9)(187)+ (1/9)(199)

=145.3333

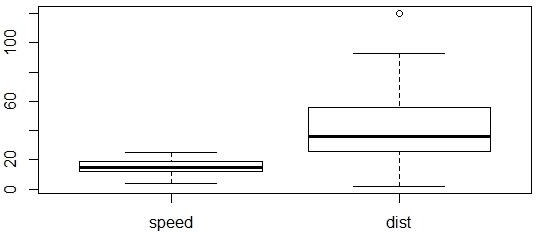
**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

**Cars speed and distance**

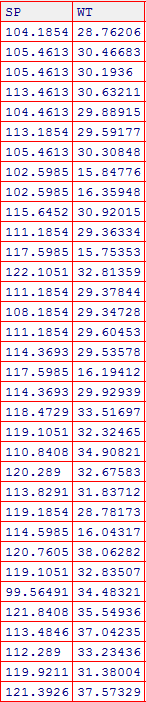
**Solution :**

**Solved in Q9\_a\_Solution.ipynb present in folder Assignment 1.**

|  |  |  |
| --- | --- | --- |
|  | Car speed | Distance |
| **Skewness** | -0.1139548 | 0.7824835 |
| **Kurtosis** | 2.422853 | 3.248019 |



**SP and Weight(WT)**

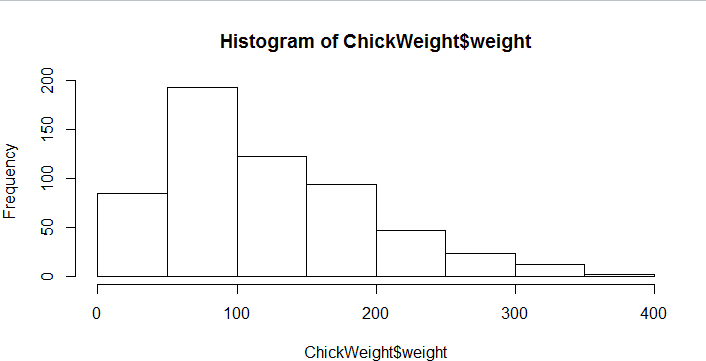


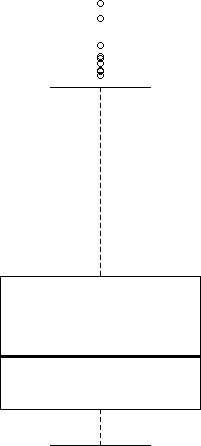
**SOLUTION :**

**Solved in Q9\_b\_Solution.ipynb present in folder Assignment 1.**

|  |  |  |
| --- | --- | --- |
|  | SP | WT |
| Skewness | 1.611450 | -0.614753 |
| Kurtosis | 2.977329 | 0.950291 |

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





**Ans :**

In Histogram , peak has right skew and tail is on right. Mean > Median. We have outliers on the higher side

In boxplot ,we can see outliers present on higher values.

**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 ?

Solution=

Solved in Q11\_Solution.ipynb

|  |  |
| --- | --- |
| **confidence interval** | **Range** |
| confidence interval94% | 198.74,201.26 |
| confidence interval96% | 198.62,201.38 |
| confidence interval98% | 198.43,201.56 |

**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? Solution=

1)Solved in Q12\_Solution.ipynb

|  |  |
| --- | --- |
| Mean | 41 |
| Median | 40.5 |
| Variance | 25.52 |
| Standard deviation | 5.05664 |



2)Mass of students marks between 38-42.

Skewness(1.52) is positive because mass of marks in left side of plot.

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

**Solution-**Data is normalized and there is no skewness.

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

**Solution-**Negative Skewness implies mass of the Distribution concentrated on right side.

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

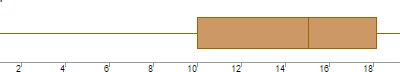
**Solution-**Positive Skewness implies mass of the Distribution concentrated on left side.

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

**Solution-** Positive kurtosis value indicates that thinner peak and wider tails. Q17) What does negative kurtosis value indicates for a data?

**Solution-**Negative kurtosis value indicates that wider peak and thinner tails.

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



What can we say about the distribution of the data?

-Not normally distributed

What is nature of skewness of the data?

The data is a skewed towards left. The whisker range of minimum value is greater than maximum

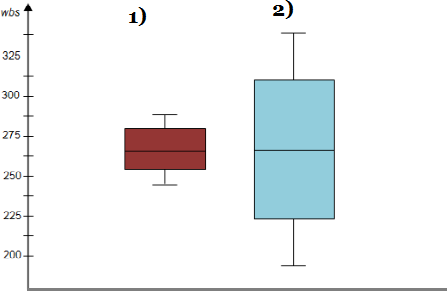
What will be the IQR of the data (approximately)?

THE INTER QUARTILE RANGE = 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.

Answer :

1 . No outliers present.

2 For plots median is same(250 -275)

3 .They are normally distributed with zero to no skewness neither at the minimum or maximum whisker range.

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)

Answer :

Solved in Q20\_Solution.ipynb in Assignment1 folder

1. P(MPG>38)= 0.3475908
2. P(MPG<40)= 0.7293527
3. P (20<MPG<50)=0.01311818

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution Dataset: Cars.csv

Solution :

Solved in Q21\_Solution.ipynb in Assignment1 folder.

**Answer :MPG of Cars follows Normal Distribution**

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

Answer-

**Adipose Tissue (AT) and Waist does not follow Normal Distribution**

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

**Solution-**

Solved in Q22\_and\_23\_Solution.ipynb in Assignment folder

|  |  |
| --- | --- |
| **Confidence interval** | **Z scores** |
| 60% | 0.8416212 |
| 90% | 1.644854 |
| 94% | 1.880794 |

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

**Solution-**

Solved in Q22\_and\_23\_Solution.ipynb in Assignment folder.

|  |  |
| --- | --- |
| **Confidence interval** | **T scores** |
| 95% | 2.063899 |
| 96% | 2.171545 |
| 99% | 2.79694 |

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

Solution:

Population mean = 270 days

Sample mean = 260 days

Sample Std Dev = 90days

Sample(n) = 18 bulbs

Degree of freedom = n - 1 = 18-1 = 17

T-SCORE = -0.471

p-value =0.3216(32%)