|  |  |
| --- | --- |
| 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 | Ordinal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ratio |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Interval |
| Number of Children | Nominal |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Interval |
| Years of Education | Ordinal |

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

Ans: No. of possible combinations=8

No. of combinations with 2 heads and 1 tail

= p(HHT)+p(HTH)+p(THH)

= 1/8+1/8+1/8

= 3/8=0.375 = 37.5%

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) Probability of sum is equal to 1 =(0/36)=0

b) Probability of sum is less than or equal to 4 i.e., (1,1), (1,2), (1,3), (2,1), (2,2), (3,1) = (6/36) =(1/6) =0.1667= 16.67%

c) Probability that sum is divisible by 2 and 3 i.e., (1,5), (2,4), (3,3), (4,2), (5,1), (6,6) = (6/36) = (1/6) =0.1667= 16.67%

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 no. of balls = 7

Let S be the sample space

Then, n(s)= no. of ways of drawing 2 balls out of 7 = 7c2 = (7\*6)/(2\*1) = 21

Let E = event of drawing 2 balls, none of which is blue i.e., n(E)= no. of ways of drawing 2 balls out of 5 balls = 5c2 = (5\*4)/(2\*1) = 10

P(E)= n(E)/n(S)= 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: 1\*0.015+4\*0.20+3\*0.65+5\*0.005+6\*0.01+2\*0.120

= 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.

**Use Q7.csv file**

Ans: **For points**

Mean=F(X)/N= 3.5965625

Median=N+1/2= 3.695

Mode= Highest repeated value= 3.92, 3.07

Variance= 0.2858814

Standard Deviation= 0.5346787

Range= Max value- Min value=4.93-2.76= 2.17

**For Score**

Mean=3.21725

Median=3.325

Mode=3.44

Variance=0.957379

Standard Deviation=0.9784574

Range=3.91

**For Weigh**

Mean=17.84875

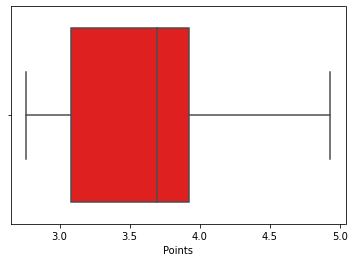
Median=17.71

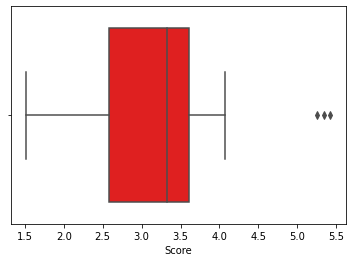
Mode=17.02

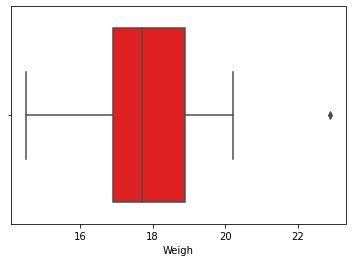
Variance=3.193166

Standard Deviation=1.786943

Range=8.40







**Note:** In each of these case , no variables has Mean = Median = Mode

Thus, as seen in the graph, dataset ‘score’ and ‘weigh’ has outliers.

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: Mean of the Weights = 1308

Expected value = P(X)/X

=1308/9 = 145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans:**

|  |  |  |
| --- | --- | --- |
|  | Skewness | Kurtosis |
| Speed | -0.117 | -0.508 |
| Dist | 0.806 | 0.405 |

* “Dist” is positively skewed whereas, “speed” is negatively skewed
* Thus Dist has distribution of data concentrated on left whereas speed has distribution on the right. As seen in the graph

**SP and Weight(WT)**

**Use Q9\_b.csv**

|  |  |  |
| --- | --- | --- |
|  | Skewness | Kurtosis |
| SP | 1.611 | 2.977 |
| WT | -0.614 | 0.950 |

* “SP” is positively skewed whereas, “WT” is negatively skewed
* Thus SP has distribution of data concentrated on left whereas WT has distribution on the right. As seen in the graph.

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



Ans: 1)The histogram peak has right (positively)skew and tail is on right. Mean>Median and we have outliers on the higher side.



2) Median is less than the Mean. The box plot has outliers on the maximum side.

**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:

**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= 40.5, Variance=25.53, Standard deviation= 5.05

1. What can we say about the student marks?

Ans: Most of the students marks varies between 34-50. Data is slightly skewed to the right because mean is greater than median.

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

Ans: Data is symmetrical and no skewness present.

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

Ans: Data is positively skewed

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

Ans: Data is negatively skewed

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

Ans: The curve of the data is more peaked and it is Leptokurtic.

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

Ans: The curve of the data is flatter and it is Platykurtic.

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



What can we say about the distribution of the data?

Ans: The data is not equally distributed. The median is towards the higher value.

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= 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: 1) There are no outliers in both the plots.

1. Both the boxplots has same median that is approximately in between the range of 250 to 275.
2. Both the graphs are normally distributed with no skewness.

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) #find the MPG less than 38--> P(MPG<38)

nd.cdf(38)

#find the MPG more than 38--> P(MPG>38)

1-nd.cdf(38)

P(MPG>38)= 0.346

1. #find the MPG less than 40--> P(MPG<40)

nd.cdf(40)

P(MPG<40)=0.73

1. #find the MPG less than 20--> P(MPG<20)

nd.cdf(20)

#find the MPG more than 20--> P(MPG>20)

1-nd.cdf(20)

P(MPG>20)=0.94

#find the MPG less than 50--> P(MPG<50)

nd.cdf(50)

P(MPG<50)=0.95

P(20<MPG<50)= P(MPG<50)- P(MPG>20)

= 0.95-0.94 =0.01

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans: 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

Dataset: wc-at.csv

Ans: 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

Ans: Z score for 90% confidence interval= -1.6448, + 1.6448

Z score for 94% confidence interval=-1.8807,+1.8807

Z score for 60% confidence interval=-0.8416,+0.8416

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

Ans: T score for 95% Confidence Interval= -2.0639

T score for 96% Confidence Interval= -2.1715

T score for 99% Confidence Interval= -2.7969

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: Degrees of freedom= 18-1 = 17

T score =-0.4714

P(t\_score) =0.3216