|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Discrete |
| Weight of a person | Continues |
| Weight of Gold | Continues |
| Distance between two places | Continues |
| Length of a leaf | Continues |
| Dog's weight | Continues |
| 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 | Interval |
| Blood Group | Nominal |
| Time Of Day | Ratio |
| Time on a Clock with Hands | Ratio |
| Number of Children | Ordinal |
| Religious Preference | Nominal |
| 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?

Sol: The probability is 3/8

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

* Equal to 1
* Less than or equal to 4
* Sum is divisible by 2 and 3

Sol:

* There is no outcomes which corresponds sum is equal to one(sum =1) the probability is 0.
* Equal to 4 (1,3)(2,2)(3,1)=3 outcomes , 3/36= 1/12
* The prpbability is 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?

Sol: 10/21 =0.4761

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

Sol:

=1\*0.015 + 4\*0.20 + 3\*0.65 + 5\*0.005 + 6\*0.01 + 2\*0.12

=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

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

**Use Q7.csv file**

Sol:

**POINTS**

Mean: 3.596563

Median: 3.695

Standard deviation: 0.534679

Mode: 3.92

Range: 2.76 4.93

Variance: 0.285881

**SCORE**

Mean: 3.21725

Median: 3.325

Standard deviation: 0.978457

Mode: 3.44

Range: 1.513 5.424

Variance: 0.957379

**WEIGHT**

Mean: 17.84875

Median: 17.71

Standard deviation: 1.786943

Mode: 17.02

Range: 14.5 22.9

variance: 3.193166

* Skewness is zero the dataset is normally distributed.
* There is no outliers cause the dataset is normally distributed.

Q8) Calculate Expected Value for the problem below

* 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?

Sol:

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

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

= (1/9)  (  1308)

=145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

Sol:

**SPEED**

Skew= -0.11751

Kurtosis=-0.57371

**DISTANCE**

Skew= 0.806895

Kurtosis=0.405053

Inferences:

* Speed is negative skewed, it is left skewed
* Distance is positive skewed ,it is right skewed
* Speed is negative kurtosis, the peakness is wide.
* Distance is positive kurtosis, the peaknessis short.

**SP and Weight(WT)**

**Use Q9\_b.csv**

Sol:

**SPEED**

Skew= 1.61145

Kurtosis= 2.977329

**DISTANCE**

Skew= -0.61475

Kurtosis=0.950291

* Speed postively skewed and distance is negatively skewed.
* the both speed and distance has positive kurtosis.

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





Hitogram has positive outliers

* Histogram has positive outliers because there is positive skewness.
* Boxplot has outliers because there is positive skewness.
* Boxplot is also know as whiskers plot.

**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?

Sol:

Population=2000

Mean=200

Standdard deviation=30

Degrees of freedom=1999

* 94% confidence interval

t=1.88

upper limit value of 94%=200+1.88(0.67)=201.26

lower limit value of 94%=200-1.88(0.67)=198.74

* 98% confidence interval

t=2.323

upper limit value of 96%=200+2.33(0.67)=201.56

lower limit value of 96%=200-2.33(0.67)=198.44

* 96% confidence interval

t=2.06

upper limit value of 96%=200+2.06(0.67)=201.38

lower limit value of 96%=200-2.06(0.67)=198.62

**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**

* Find mean, median, variance, standard deviation.
* What can we say about the student marks?

Sol:

Mean= 41

Meadian=40.5

Variance=25.52941

Standard deviation=5.052664

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

Sol:

zero skewness, symmetric.

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

Sol:

The distribuition is positively skewed.

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

Sol:

The distribuition is negatively skewed.

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

Sol:

The distribuition is peaked and possess thick tails.

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

Sol:

The 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?

What is nature of skewness of the data?

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

* The distribuition is not normal
* it is left skewed
* 1st quartile-3rd quartile= 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.

Sol:

* The two boxplots are symmetric in nature.
* There is no outliers in two boxplots.
* Mean=Median=Mode

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

* P(MPG>38)
* P(MPG<40)

c. P (20<MPG<50)

Sol:

Mean= 34.42

Standard deviation=9.13

* P(MPG>38)= 0.35
* P(MPG<40)=0.72935

c. P (20<MPG<50)=0.898869

Q 21) Check whether the data follows normal distribution

* Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Sol:

MPG of cars does not follows normal distribuition.the mean is not equal to meidan

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

Dataset: wc-at.csv

Sol:

Waist follows normal distribution.the mean=median.

Adipoise tissue is not normally distributed.the mean is not equal to meidan

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

Sol:

90% confidence interval,(0.95) Z score = 1.65

94% confidence interval,(0.97) Z score = 1.89

60% confidence interval,(0.8) Z score = 0.84

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

Sol:

* Confidence interval of 95%,24

t= 2.063899

* Confidence interval of 96%,24

t=2.171545

* Cofidence interval of 99%,24

t=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

Hint:

rcode pt(tscore,df)

df degrees of freedom

Sol:

t - Statistics for the data is given as follows;

x = Mean of the sample of bulbs =  260

μ = Population mean = 270

s = Standard deviation of the sample = 90

n = Number of items in the sample = 18

t=-10/23.23

t = - 0.471

Degrees of freedom is n – 1 = 18-1= 17

In R code,

pt (-0.471,17)=0.321

Probability is 32%