|  |  |
| --- | --- |
| 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 | Continuous |
| 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 | Intervals |
| Celsius Temperature | Ratio |
| Weight | Ratio |
| Hair Color | Nominal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Ratio |
| Height | Ratio |
| Type of living accommodation | Nominal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Interval |
| Blood Group | Nominal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Ratio |
| Number of Children | Ratio |
| 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) Total number if tosses = 3

Total number of outcome as head = 2

Total number of outcome as tail = 1

Probability = No. of outcomes/total no. of cases

Probability that two heads are = 2/3 = 0.67

Probability that one tail is = 1/3 = 0.33

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

Ans) Total number of rolled dices = 36

a) Probability Equal to 1 = No of favorable outcomes/Total number of cases

= 1/36 = 0.027

b) Total number of the chances for the sum less than or equal to 4 = 6

Probability = No of favorable outcomes/Total number of cases

= 6/36 = 0.16

c) Total number of sum is divisible by 2 = 18

Total number of sum is divisible by 3 = 12

Probability of divisible by 2 = 18/36 = 0.5

Probability of divisible by 3 = 12/36 = 0.33

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 number of balls drawn = 2

Total number of balls exept blue = 5

Probability of balls drawn = 2/5 = 0.4

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) Expected number of candies for a randomly selected child = candies count \* Probability

= 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

Expected number of candies for a randomly selected child = 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**

** Mean for points, score, weigh are as follows**

**Points 3.596563**

**Score 3.217250**

**Weigh 17.848750**

** Median for points, score, weigh are as follows**

**Points 3.695**

**Score 3.325**

**Weigh 17.710**

** Mode for points, score, weigh are as follows**

**Points 3.07, 3.92**

**Score 3.44**

**Weigh 17.02, 18.90**

**Points and weigh have two modes because they have the maximum occurance.**

** Variance for points, score, weigh are as follows**

**Points 0.285881**

**Score 0.957379**

**Weigh 3.193166**

** Standard Deviation for points, score, weigh are as follows**

**Points 0.534679**

**Score 0.978457**

**Weigh 1.786943**

** Range for points, score, weight are as follows**

**Points 2.17**

**Score 3.911**

**Weigh 8.4**

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?

Ans) The expected value of the weight of that patient = total of weight/total number of patient.

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

= 1308/9

= 145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans)**Formulae to calculate skewness = (3(mean-median)/standard deviation)

Formulae to calculate kurtosis = ((mean-mode)/standard deviation)^4

9a) Skewness for Index, speed, distance are as follows

Index 0.0000

Speed -0.117510

Distance 0.806895

Kurtosis for Index, speed, distance are as follows

Index -1.200000

Speed -0.508994

Distance 0.405053

9b) Skewness for SP and WT are as follows

SP 1.611450

WT -0.614753

Kurtosis for SP and WT are as follows

SP 2.977329

WT 0.950291

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





Ans) The histogram peak has right skew and tail is on right. Mean > median. We have outliners on the higher peak.

The Boxplot has outliners 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) To calculate confidence interval for 94%, 96%, 98%, python code needs to be followed.

c\_94 = stats.t.interval(alpha = ,df =,loc =,scale = /np.sqrt())

Print(c\_94)

Confidence interval for 94% are (198.73, 201.26)

Confidence interval for 96% are (198.62, 201.37)

Confidence interval for 98% are (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**

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

Ans) 1)Mean for the abbove obtained scores is 41.0

Median for the above obtained scores is 40.5

Variance for the above obtained scores is 24.11

Standard deviation for the above obtained scores is 4.91

2)We don’t have outliners and the data is slightly skewned towards right because mean is greater than median.

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

Ans) No skewness is present we have a perfect symmetrical distribution

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

Ans: Skewness and tail is towards right.

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

Ans) Skewness and tail is towards left.

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

Ans)Positive kurtosis means the curve is more peaked and it is Leptokurtic.

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

Ans) Negative kurtosis means the curve will be flatter and broader.

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



What can we say about the distribution of the data?

Ans) The above Boxplot is not normally distributed, the median is towards the higher value

What is nature of skewness of the data?

Ans) The data is skewed towards left. The whisker range is minimum, value is greater than maximum.

What will be the IQR of the data (approximately)?   
  
Ans) The inter Quartile Range = Upper quartile – Lower 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.

Ans) First there are no outliners. Second both the box shares the same median that is approximately in a range between 275 to 250 and 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

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

c. P (20<MPG<50)

Ans) Probability of MPG > 38 is 0.348

Probability of MPG < 40 is 0.729

Probability of 20<MPG<50 is 0.013

Q 21) Check whether the data follows normal distribution

* Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans) MPG of cars follows normal distribution.

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

Dataset: wc-at.csv

Ans) Adipost Tissue and waist circumference doesn’t follows the normal distribution.

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

Ans) Z score of 90% confidence interval is 1.28

Z score of 94% confidence interval is 1.55

Z score of 60% confidence interval is 0.25

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

T score of 95% Confidence interval for sample size 25 is 1.71

T score of 96% Confidence interval for sample size 25 is 1.82

T score of 99% Confidence interval for sample size 25 is 2.49

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) Formulae = ((x – pop mean)/(sample standard deviation/square root of sample size))

=(260-270)/(90/sqrt(18))

=(-10)/(90/4.24)

=(-10)/(21.22)

=-0.471