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

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

HHT,THT,THH,TTH.HTT,HTH,HHH,TTT

HHT

HTH

THH

SOLUTION: 3/8

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

Two dice are 6\*6=36

SOLUTIONS:

1. Sum=1 is 0
2. Less than or equal to 4 (1,1),(1,2),(1,3)(2,1),(2,2),(3,1) 6/36=1/6
3. Sum is divisible by 2 and 3 (1,5),(3,3),(4,2),(5,1),6,6)=5/6

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?

1 bag = 2 red, 3green,2 blue = 7 balls

7C2= 7\*6=42/ 2\*1= 2

= 21

5C2(balls leving blue)= 20/2=10

Solutions

= 10/21

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

SOLUTIONS:

Expected number of candies is

=  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

=       3.090

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**

Column 1 is for points

Column 2 is for score

Column 3 is for weigh

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *Column1* |  | *Column2* |  | *Column3* |  |
|  |  |  |  |  |  |
| Mean | 3.5965625 | Mean | 3.21725 | Mean | 17.84875 |
| Standard Error | 0.0945187 | Standard Error | 0.172968 | Standard Error | 0.31589 |
| Median | 3.695 | Median | 3.325 | Median | 17.71 |
| Mode | 3.92 | Mode | 3.44 | Mode | 17.02 |
| Standard Deviation | 0.5346787 | Standard Deviation | 0.978457 | Standard Deviation | 1.786943 |
| Sample Variance | 0.2858814 | Sample Variance | 0.957379 | Sample Variance | 3.193166 |
| Kurtosis | -0.450432 | Kurtosis | 0.416595 | Kurtosis | 0.864931 |
| Skewness | 0.2927802 | Skewness | 0.465916 | Skewness | 0.406347 |
| Range | 2.17 | Range | 3.911 | Range | 8.4 |
| Minimum | 2.76 | Minimum | 1.513 | Minimum | 14.5 |
| Maximum | 4.93 | Maximum | 5.424 | Maximum | 22.9 |
| Sum | 115.09 | Sum | 102.952 | Sum | 571.16 |
| Count | 32 | Count | 32 | Count | 32 |

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?

Sum: 108+ 110+ 123+ 134+ 135+ 145+ 167+ 187+ 199=1308

SOLUTION:

The expected value of the weight of a patient chosen at random will be same as mean of all the patients hence the value of the weight of that patients145.33 pounds

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

**Cars speed and distance**

**Use Q9\_a.csv**

SOLUTON**:**

Skewness of speed is -0.1139548 (left skew) the distribution will be towards right

Kurtosis of speed is 2.422853 positive Kurtosis with wide peak

Skewness of Distance is 0.7824835

Kurtosis of Distance is 3.248019 positive Kurtosis with wide peak

**SP and Weight(WT)**

**Use Q9\_b.csvb**

Skewness of SP is 1.5814 right skew

Kurtosis of SP is 5.7235 positive Kurtosis with high peak

Skewness of Weight is -0.60330 left skew

Kurtosis of Weight is 3.8194 positive Kurtosis with high peak

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



SOLUTION:

Histogram

Data is not a normal distribution as 50-100 weight is having more frequency between 150-200 and 350-400 weight is having less frequency and it could be between 0-5.

0-50 weight having frequency between 50-100

50-100 weight is having more frequency between 150-200

100-150 weight is having more frequency between 100-130 and so on

Data is right skewed hence it is positive skew

BOXPLOT

Can see 7 outliers with positive skewness as data is towards right

Data is not distributed normally

**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:

**X  ±  Zs√n**

**For 94%**

**200  ±  1.88 30√2000**

**Confidence level for 94% is [198,201]**

**For 96%**

**200  ±  2.05 30√2000**

**Confidence level for 96% is [198.6,201.3]**

**For 98%**

**200  ±  2.328 30√2000**

**Confidence level for 98% is [198.4,201.4]**

**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:

Mean =41

Median =40.5

Variance = 25.52941

SD= 5.052664

The students marks range from 34 to 56 where Most of the students score is between 38 to 42

Average students marks is 41

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

No skewness

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

Positive skewed as data will be towards left

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

Negative skewed

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

Distribution will be peaked and has thick tails.

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

Distribution will be flatter and have normal curve

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



What can we say about the distribution of the data?

No outliers, data is towards left so left skew , most of the data range from 10 to 18 , median is between 15-16

What is nature of skewness of the data?

Left skewness

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

IQR=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.

Normal distribution

No outliers

box plot 2 is large in size

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)

Z=X-mue/sd

38-34.42/9.131=0.6517

P(MPG>38) =1-0.6517 =0.3483

* 1. P(MPG<40)

40-34.42/ 9.131= 0.610

Z table = 0.7291

P(MPG<40) = 0.7291

* 1. P (20<MPG<50)

SOLOUTION:

50-34.42/9.131=1.700 z table 0.758

20-34.42/9.131=-1.579 z table -0.582

P (20<MPG<50) = 0.758-0.582 = 0.176

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

MPG of cars don’t follow 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

AT follows normal distribution

Waist don’t follow normal distribution

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

SOLUTION:

90% is 1.65

94% is 1.55

60% is 085

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

SOLUTION:

95% =20

96%=2.17

99%= 2.2

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

t - statistics for the data is given as follows:

t=x=mu/s/sqrt(n)

x = =  260

mu= 270

s = = 90

n = 18

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

t = - 0.471