|  |  |
| --- | --- |
| 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) | Discretes |

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 | Ratio |
| 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?

A: Three coins are tossed simultaneously.

All the possible outcomes are 8

i.e ;(H,H,H),(T,H,H),(H,T,H),(H,H,T),(T,T,H),(T,H,T),(H,T,T),(T,T,T)

The probability of two heads and one tail is = 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

A: a) The probability that sum is equal to 1 is 0.the lowest combination is 2.

b) Less than or equal to 4

The total possible outcomes is =6^2=36

The possible outcomes are

(1,1),(1,2),(2,1),(2,2)  
 The possibilities are =4/36

1. Sum is divisible by 2 and 3.

Favourable outcomes are (3,3),(2,4),(4,2) ,(1,5),(5,1),(6,6)

Sum is divisible by 2 and 3 is 6/36

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?

A: The total number of balls is = 7

The probability that none of ball drawn is blue =5C2/7C2 = 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

A: Expectation = ∑ pixi

=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+0.006+0.24

=3.036

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.

A. There are 32 observations and 4 variables

**Points:**

Mean=3.5965625

Median=3.695

Variance=0.285881

Standard Deviation=0.5346787

Range: Min value=2.76,Max value=4.93

**Score:**

Mean=3.2172499

Median=3.325

Variance=0.95737

Standard deviation=0.97845

Range=(1.513,5.424)

**Weigh**:

Mean=17.84875

Median=17.71

Variance=3.193166

Standard deviation=1.78694

Range=(14.5,22.9)

Inference: points and Weigh have bimodal values

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?

A.The probability of choosing one patient = 1/9

Expectation = ∑ pixi

=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]

=145.3333

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

**Cars speed and distance**

**Speed:**

Skewness= -0.1175

Kurtosis=-0.5089

**Distance**:

Skewness=0.806894

Kurtosis=0.40505

The data of speed is negatively skewed and kurtosis is <3

The data of distance is positively skewed and kurtosis is <3

**SP and Weight(WT)**

**Use Q9\_b.csv**

**SP:**

Skewness=1.6114

Kurtosis=2.9773

**Weigh**:

Skewness= -0.614

Kurtosis=0.95029

The data of WT is negatively skewed and kurtosis is <3

The data of SP is positively skewed and kurtosis is <3

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



A: **Inferences:**

**Histogram:**  the histogram is peak has right skew and tail is on right

It is positively skewed (>3)

**Boxplot:** Middle line indicates the Median or Inter Quartile Range

Above the Median is Upper Quartile (Q1)

Below the Median is Lower Quartile (Q3)

Upper whisker > Lower whisker

The boxplot has Outlier 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?

A.Mean = 200, SD=30

Confidence interval for 94% = 143.57, 256.42

Confidence interval for 98% = 130.20, 269.424

Confidence interval for 96% = 138.38, 261.612

**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.

A. Mean=41

Median= 40.5

Variance :∑( xi-mean) 2=25.529

Standard deviation =5.052

2)What can we say about the student marks?

A. Since mean is greater than median the skewness of the data is positive.

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

A: The data will be perfectly symmetric.

Skewness = 0

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

A: The data will be right Skewed

Skewness will be Positive skewed

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

A: The data will be left Skewed

Skewness will be Negative skewed

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

A: Positively skewed (>3)

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

A: Negatively skewed (<3)

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)?

a) The Boxplot is not follows the Normal Distribution. It has no Outliers .the median is towards the higher value

1. It is Negatively Skewed.
2. Inter Quartile Range = 8

Upper Quartile is 18 and Lower Quartile is 10.

Q19) Comment on the below Boxplot visualizations?



Draw an Inference from the distribution of data for Boxplot 1 with respect Boxplot 2.

A. A: For Boxplot1 and Boxplot2 Median (or) Inter Quartile Range is Equal.

There are no outlier. The Range between 250 to 275

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)

c. P (20<MPG<50)

A. a) P (MPG>38) = 0.3475

b) P (MPG<40) = 0.7293

c ) P (20<MPG<50) = 0.8988

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**A.** No, the data of MPG is not normally distributed because the mean, median and mode are not equal and the skewness is also not equal to 0.

Mean =34.42, Median=35.15, Mode=29.629 & skewness = -0.17

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

Dataset: wc-at.csv

1. No, either of the data in the respective columns are not normally distributed because the mean, median and mode are not equal and the skewness of the data is also not equal to 0.

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

60% confidence interval

A: Z scores of 90% confident interval is 1.644.

Z score of 94% confidence interval is 1.880.

Z score of 60% confidence interval is 0.841.

Q 23) Calculate the t scores of 95% confidence interval, 96% confidence interval,

99% confidence interval for sample size of 25

A.For 95% confidence interval is 2.0638

For 96% confidence interval is 2.1715

For 99% confidence interval is 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

A: T score = 260-270/(90/18^0.5) = -0.47

P(X<260)=stats.t.cdf(-0.47,17)=0.32

\* Refer: “Assignment.1.ipynb” file

**Submitted by**

**DASARI RACHANA**

**dasarirachana789@gmail.com**