|  |  |
| --- | --- |
| 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) | 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: let three coins are tossed {HHH,HHT,HTH,HTT,THH,THT,TTH,TTT}

PROBABILITY = favorable out comes/total out comes

1. P(two heads)=4/8=1/2
2. P(one tail)=7/8=0.875

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
4. a=0

b=6(1,1)(1,2)(1,3)(2,1)(2,2)(3,1)

c=29/36

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?

1. =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

1. Expected number of candies are =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**

1. **POINTS SCORE WEIGH**

**Mean 3.596563 3.217250 17.848750**

**Median 3.695 3.325 17.710**

**Mode 3 3 0**

**Variance 0.285881 0.957379 3.193166**

**Standard deviation 0.534679 0.978457 1.786943**

**Range 2 3.911 8.4**

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?

1. Expected value of weight = 145.3.

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

**Cars speed and distance**

**Use Q9\_a.csv**

**A ) SKEWNESS**

**\*index skew : 0.000**

**\*speed skew : -0.117**

**\*dist skew : 0.806**

**KURTOSIS**

**\*index kurt :-1.200**

**\*speed kurt :-0.508**

**\* dist kurt :0.405**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**A) SKEW NESS:**

**\*sp skew : 1.611**

**\*wt skew : -0.615**

**KURTOSIS:**

* **Sp kurt : 2.977**
* **Wt kurt :0.95**

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



A:

The histogram is positively skewed.

Maximum number of chicks are falling under the weight between 50 to 100.

More then50% of the chicks weights are under the weight limit of 200.



THE BOX PLOT SHOWS THAT THERE ARE OUTLERS AT THE UPPER SIDE AND SKEWED ON RIGHT.

**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?

1. 94% confidence interval is higher limit=201.26&lower limit =198.73
2. 98% confidence interval is higher limit=201.57&lower limit=198.42
3. 96% confidence interval is higher limit=201.38&lower limit=198.61

**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?

A) 1)

\* MEAN : 41

\* MEDIAN = 40.5

\* VARIANCE = 24.111

\* STD = 4.917

2) we can say the student marks are average.

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

1. It has “zero” skew ness.

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

1. The mode occurs at highest frequncey of the highest distribution.

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

1. These ocuurs positive skew.

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

1. Positive excess values of kurtosis (>3) indicate that a distribution is peaked and possess thick tails.

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

1. Negative excess values of kurtosis(<3) indicate that a 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)?

1. \* The data is probability distribution.

\*The skewness is negative skewness.

\*IQR of data is approximately “8”.

Q19) Comment on the below Boxplot visualizations?



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

1. 1. Q1 = 275 Q3 = 250

Minimum = 235, median = 265, maximum = 280

Lower Fence = 312.5

Upper fence = 21.5

2. Q1 = 300 Q3 = 225

Minimum = 200, median = 265, maximum = 325

Lower Fence = 337.5

Upper Fence = 187.5

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 = 66/81=0.815

B =61/81=0.753

C =69/81=0.852

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

ANS)

The skewness value of mpg coloum =-0.18

The kurtosis value of mpg coloum = -0.61

The mpg follows the normal distribution as the values are close to 0.

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

Dataset: wc-at.csv

ANS)

The skewness and kurtosis of waist respectively : 0.13&-1.1

Waist doesn’t follow the normal distribution curve as the kurt value>1

The skewness and kurtosis of AT respectively:0.58&0.29

AT follows the normal distribution curve as values are closer to 0.

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

ANS) Z score of 90% = 1.644

Z score of 94% =1.88

Z score of 60%= 0.84

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

A) For 95% =1.71

For 96% = 1.82

For 99%=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

1. The t value is -0.471

The p value is 0.322