|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Numerical(discrete) |
| Results of rolling a dice | Numerical(discrete) |
| Weight of a person | Numerical(continuous) |
| Weight of Gold | Numerical(continuous) |
| Distance between two places | Numerical(continuous) |
| Length of a leaf | Numerical(continuous) |
| Dog's weight | Numerical(continuous) |
| Blue Color | Categorical |
| Number of kids | Numerical(discrete) |
| Number of tickets in Indian railways | Numerical(discrete) |
| Number of times married | Numerical(discrete) |
| Gender (Male or Female) | Categorical data type |

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 | Ordinal |
| 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)3/8

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

1. Equal to 1

A)0

1. Less than or equal to 4

A)6/36=1/6

1. Sum is divisible by 2 and 3

A)24/36=2/3

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. There are 7 balls originally with 2 of them blue so the probability of the first ball not being blue is 5/7. This leaves 6 balls with 2 blue. The probability of the second ball not being blue assuming that the first wasn’t is 4/6. The probability that neither ball drawn was blue is

(5/7)\*(4/6)=20/42=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)Expected number of candies for a randomly selected child  = 3.09

#Expected number of candies for a randomly selected child

= 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

= 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.

A) Mean Median Mode STD Variance Range

Points 3.59 3.69 3.07 0.53 0.28 2.76-4.93

Score 3.21 3.32 3.44 0.97 0.95 1.513-5.424

Weigh 17.84 17.71 17.02 1.78 3.19 14.5-22.9

**Use Q7.csv file**

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)

Expected Value  =  ∑ ( probability  \* Value )

 ∑ P(x).E(x)

there are 9 patients in clinic

Probability of selecting each patient = 1/9

E(x)  108, 110, 123, 134, 135, 145, 167, 187, 199

P(x)  1/9  1/9   1/9  1/9   1/9   1/9   1/9   1/9  1/9

Expected Value  =  (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

Hence,the expected Value of the Weight of that patient = 145.33 

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

**Cars speed and distance**

**Use Q9\_a.csv**

1. **Skewness of Q9a is** 

**speed -0.117510**

**dist 0.806895**

**kurtosis of Q9a is**

**speed -0.508994**

**dist 0.405053**

**SP and Weight(WT)**

**Use Q9\_b.csv**

1. **Skewness of Q9b is** 

**SP 1.611450**

**WT -0.614753**

**kurtosis of Q9b is**

**SP 2.977329**

**WT 0.950291**

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



A)1. The above given histogram is positively skewed.

2. It has positive kurtosis.

3. Most of the data is distributed in between 50-150.



1. The mean and median are different in this boxplot because of the number of outliers which increase the mean by a lot.

The IQR is also not symmetrical and is positively skewed.

Min-25%-50%(Median)-75%-Max

**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)94%-> (198.738325292158, 201.261674707842) 

96%-> (198.62230334813333, 201.37769665186667)

98%-> (198.43943840429978, 201.56056159570022)

**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?
3. Mean=41.000000 variance=25.529411764705884

Median=40.5 std=5.052664

Hence, majority of the students scored average score.

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

1. There is no skewness, it will be normal distribution.

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

A)Positively skewed.

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

1. Negatively skewed.

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

1. It indicates a distribution where more of the numbers are located in the tails of the distribution instead of around the mean.

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

1. The distribution has lighter tails than the normal distribution.

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



What can we say about the distribution of the data?

1. The distribution is not normal distribution. And median is closer to the upper quartile.

What is nature of skewness of the data?

1. Negatively skewed.

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

1. Q3-Q1=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.

A)\* Both boxplot 1 and 2 have same median.

\* IQR of 1 is less than 2.

\* Both boxplots have no outliers.

\* Both boxplots are normally distributed.

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

A)

* 1. P(MPG>38)= 0.3475939251582705
  2. P(MPG<40)= 0.7293498762151616

c. P (20<MPG<50)= 0.013116469610523374

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

1. Not normally distributed. 
2. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

A)adipose tissue=not normally distributed.

Waist circumference= not normally distributed.

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

A)90%= (-1.281 to 1.281) 

94%= (-1.555 to 1.555)

60%= (-0.253 to 0.253)

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

A) 95%= (-1.711 to 1.711) 

96%= (-1.828 to 1.828)

99%= (-2.492 to 2.492)

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 probability of 18 randomly selected bulbs with average life <260 is 0.3216 i.e., 32%

