|  |  |
| --- | --- |
| 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 Scale |
| Weight | Ratio Scale |
| 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?

**ANS: 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

**ANS: a) 0**

**b) 1/6**

**c) 1/12**

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: 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

**ANS: 3.225**

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**

**ANS:**

**Points:** Mean – 3.59 ,MEDIAN – 3.69, MODE- 3.92 ,VARIANCE- 0.28, STD=0.53, RANGE: MIN-2.7, MAX-4.9

**SCORE:** Mean – 3.21 ,MEDIAN – 3.32, MODE- 3.44 ,VARIANCE- 0.95, STD=0.97, RANGE: MIN-1.51,MAX-5.42

**WEIGH:** Mean – 17.84 ,MEDIAN – 17.71, MODE- 17.02 ,VARIANCE- 3.19, STD=1.78, RANGE: MIN-14.5,MAX-22.9

In **POINTS** we see that the mode>median>mean so it is **negatively skewed.** And the variance is low so we can say that the POINTS is less dispersed from its mean it is close to mean.

In **SCORE** we see that the mode>median>mean so it is **negatively skewed.** And the variance is little high so we can say that the SCORE is little dispersed from its mean it is little far to mean.

In **WEIGH** we see that the mean>median>mode so it is **positively skewed.** And the variance is high so we can say that the WEIGH is dispersed from its mean it is far from mean.

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?

**ANS:**

Expected Value  =  ∑ ( probability  \* Value )

 ∑ P(x).E(x)

there are 9 patients

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

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**

**ANS:**

**SPEED** - skewness : -0.117 , kurtosis: -0.508

**DISTANCE -** skewness : 0.806 , kurtosis: 0.405

In **SPEED** the **skewness is negative**, that means the data points are more concentrated towards the right-hand side of the distribution. This makes the mean, median, and mode bend towards the right. Hence these values are always negative. In this distribution, Mode > Median > Mean. Where as the **kurtosis is also negative,** that means **the distribution is flat and has thin tails**. Platykurtic distributions have negative kurtosis values. A platykurtic distribution is flatter (less peaked) when compared with the normal distribution, with fewer values in its shorter (i.e. lighter and thinner) tails.

In **DISTANCE** the **skewness is positive** , that means the values are more concentrated towards the right side, and the left tail is spread out. Hence, the statistical results are bent towards the left-hand side. Hence, that the mean, median, and mode are always positive. In this distribution, Mean > Median > Mode. Where the **kurtosis is also positive,** that means **distribution is peaked and possesses thick tails**.

**SP and Weight(WT)**

**Use Q9\_b.csv**

**ANS:**

**SP** - skewness : 1.611 , kurtosis: 2.97

**WT -** skewness : -0.614 , kurtosis: 0.950

In **SP** the **skewness is positive** , that means the values are more concentrated towards the right side, and the left tail is spread out. Hence, the statistical results are bent towards the left-hand side. Hence, that the mean, median, and mode are always positive. In this distribution, Mean > Median > Mode.Whereas the kurtosis is less than 3 (beta values is <3) so it is Platykurtic. . A platykurtic distribution is flatter (less peaked) when compared with the normal distribution, with fewer values in its shorter (i.e. lighter and thinner) tails.

In **WT** the **skewness is negative**, that means the data points are more concentrated towards the right-hand side of the distribution. This makes the mean, median, and mode bend towards the right. Hence these values are always negative. In this distribution, Mode > Median > Mean. Whereas the **kurtosis is positive,** that means **distribution is peaked and possesses thick tails**.

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



**ANS:**

In the above **HISTOGRAM ,**we see that it is positively skewd where the mean>median>mode. that means the values are more concentrated towards the right side, and the left tail is spread out.And the **SECOND BIN** of the histogram is the **mode.**

In the above **BOX-PLOT** , we see that the **whisker** is more in the upper side therefore it is a positively skewd i.e mean>median>mode. After the upper extreme we see there are some **outliers**.

**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:**

**94% : (143.54415570565965, 256.45584429434035)**

**98%: (130.15355671679083, 269.84644328320917)**

**96%:(138.34730111522666, 261.6526988847733)**

**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?

**ANS:** mean: 41, median: 40.5,variance:25.52, std: 5.05

Seeing the above information we can say that the marks of the students varies a lot from the mean. And also the mode marks is 41 as the frequency of 41 is more.

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

**ANS:** If the mean and median of the data are equal then the distribution is symmetric so no skewness or zero skewness.

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

**ANS:** If the mean is greater than the median, the distribution is positively skewed.

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

**ANS:** If the median is greater than mean ,the distribution is negatively skewed.

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

**ANS:** positive values of kurtosis indicate that **distribution is peaked and possesses thick tails**. Leptokurtic distribution has a positive kurtosis.

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

**ANS:**  indicate that **a distribution is flat and has thin tails**. Platykurtic distributions have negative kurtosis values.

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)?

ANS: The distribution of the data is negatively skewed . And the median is near about 15. so the nature of the skewness is negatively skewed. As the whisker towards the lower side is elongated The IQR of the data will be approximately 8. i.e upper quartile- lower quartile (18-10).

Q19) Comment on the below Boxplot visualizations?



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

**ANS:** Here we see that the both the boxplot 1 & 2 have the same median. But the data in boxplot1 is more dense as compared to the boxplot2 since the area of the box in boxplot 2 is more(more dispersed)(IQR is more for Boxplot2 than boxplot1). Boxplot 2 has a normal distribution since the median is exactly in the middle of the box.

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)
  3. P (20<MPG<50)

**ANS:** a. 0.34759393

b. 0.72934988

c. 1.24309688e-05

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**ANS:** MPG of Cars does not follows Normal Distribution. Mean , median and mode are not equal. **(see python file)**

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

Dataset: wc-at.csv

**ANS:** Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set does not follows Normal Distribution. Mean , median and mode are not equal. **(see python file)**

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

**ANS:** 90% - 1.6448536269514722

94% - 1.8807936081512509

60% - 0.8416212335729143

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

**ANS:** 95% - 2.0638985616280205

96% - 2.1715446760080677

99% - 2.796939504772804

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:** 0.321 (see python file)