|  |  |
| --- | --- |
| 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 | ORDINAL |
| SAT Scores | INTERVAL |
| Years of Education | RATIO |

Q3) Three Coins are tossed, find the probability that two heads and one tail are obtained? **3/8**

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

1. Equal to 1  **0**
2. Less than or equal to 4  **5/36**
3. Sum is divisible by 2 and 3 **5/36**

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? **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\*0.015+4\*0.20+3\*0.65+5\*0.005+6\*0.01+2\*0.120=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.

**Points-** **mean=3.596563, median=3.695 , mode=** **, variance= 0.2858814,SD= 0.5346787, Range=2.76-4.93**

**Score- mean=3.21725, median= 3.325 , mode= , variance= 0.957379,SD= 0.9784574, Range=1.513-5.424**

**Weigh- mean=17.84875, median= 17.71 , mode= , variance=3.193166 ,SD=1.786943 , Range=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?

**Probability of selecting each patient is 1/9**

**108+110+123+134+145+167+187+199=1308**

**1308\*1/9=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

**Speed- skewness= -0.1139548, kurtosis=2.422853**

**Distance- skewness= 0.7824835, kurtosis=3.248019**

SP and Weight(WT)

Use Q9\_b.csv

**SP- skewness= 1.581454, kurtosis=5.723521**

**Weight- skewness= -0.6033099, kurtosis=3.819466**

Q10) Draw inferences about the following boxplot & histogram



**The above histogram is right skewed which means that most of the data falls to the right of the graph’s peak.**

**The above boxplot is left skewed which means that the mean is lesser than the median, which is often less than the mode**

**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?

**Standard error= 30/ root 2000=0.67**

**aplha= 1-0.94=0.06**

**Critical probability= 0.97**

**Degrees of freedom= 999**

**Critical value=1.88**

**Margin of error=1.25**

**94% confidence interval= [198.75, 201.25]**

**Alpha= 1-0.98= 0.02**

**Critical probability= 0.99**

**Degrees of freedom=999**

**Critical value=2.33**

**Margin of error= 1.56**

**98% confidence interval= [198.44, 201.56]**

**Alpha=1-0.96=0.04**

**Critical probability=0.98**

**Degrees of freedom=999**

**Crtical value=2.06**

**Margin of error=1.38**

**96%confidence interval=[198.62, 201.38]**

**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.

**MEAN=41**

**MEDIAN=40.5**

**MODE=41**

**VARIANCE=24.11**

**SD=4.9101**

1. What can we say about the student marks?

**We can say that the average score of the class is 41 with the lowest being 34. Variance is 24 since the marks are spread out from the mean**

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

Q14) What is the nature of skewness when mean > median ? **It is skewed to the right**.

Q15) What is the nature of skewness when median > mean? **It is skewed to the left**

Q16) What does positive kurtosis value indicates for a data ? **it means that the distribution has heavier tails and sharper peaks than the normal distribution**

Q17) What does negative kurtosis value indicates for a data? **It indicated that 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?

**We can say that the distribution has more values plotted on the right than left, where the tail of the distribution is longer on the left side and mean is lower than the median and mode**

What is nature of skewness of the data?

**It is skewed to the left**

What will be the IQR of the data (approximately)? **8**

Q19) Comment on the below Boxplot visualizations?



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

**Both the boxplots have the same median but the upper and lower quartiles of boxplot1 are lesser than that of boxplot2**

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) = **33/81**
  2. P(MPG<40)= **61/81**

c. P (20<MPG<50)= **69/81**

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 does follow normal distribution approximately since mean and median are almost same**

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

Dataset: wc-at.csv

**NO, the data set does not follow normal distribution**

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

**90%= 1+0.9/2=0.85**

**0.80234**

**94%=1+0.94/2=0.97**

**0.83398**

**60%=1+0.6/2=0.8**

**0.78814**

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

**95%- 1.7109**

**99%- 2.4922**

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

**The probability 18 random bulbs would have an average life of no more than 260 days is 32.17%**