|  |  |
| --- | --- |
| 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 | **Nominal** |
| Number of kids | **Discrete** |
| Number of tickets in Indian railways | **Discrete** |
| Number of times married | **Discrete** |
| Gender (Male or Female) | **Ordinal** |

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 | **Ordinal** |
| High School Class Ranking | **Ordinal** |
| Celsius Temperature | **Ratio** |
| Weight | **Ratio** |
| Hair Color | **Nominal** |
| Socioeconomic Status | **Ordinal** |
| Fahrenheit Temperature | **Ratio** |
| Height | **Ratio** |
| Type of living accommodation | **Nominal** |
| Level of Agreement | **Ordinal** |
| IQ(Intelligence Scale) | **Ordinal** |
| Sales Figures | **Ratio** |
| Blood Group | **Nominal** |
| Time Of Day | **Interval** |
| Time on a Clock with Hands | **Ratio** |
| Number of Children | **Interval** |
| Religious Preference | **Nominal** |
| Barometer Pressure | **Ratio** |
| SAT Scores | **Ordinal** |
| Years of Education | **Ratio** |

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

**Total Possible Events=8(HHH,HHT,HTT,TTT,TTH,THH,HTH,THT)**

**Random Variable=3**

**3/6=0.5**

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. **0**
5. **6/36=1/6 = 0.1667**
6. **6/36=1/6 = 0.1667**

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?

**Total no. of Events = 7C2 = 21**

**Random Variable = 5C2 = 10**

**10/21 = 0.4762**

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

**Expected Value(E)=(1\*0.015)+(4\*0.20)+(3\*0.65)+(5\*0.005)+(6\*0.01)+(2\*0.12)**

**= 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**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Mean** | **Median** | **Mode** | **Variance** | **STD** | **Range** |
| **Points** | 3.596563 | 3.695 | **3.07** | 0.285881 | 0.534679 | **2.17** |
| **Score** | 3.217250 | 3.325 | **3.44** | 0.957379 | 0.978457 | **3.911** |
| **Weigh** | 17.848750 | 17.710 | **17.02** | 3.193166 | 1.786943 | **8.39** |

**Inference: As the variance of “Weigh” is more there is possibility of outliers.**

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?

**(108+110+123+134+135+145+167+187+199)/9 = 145.33**

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

**Cars speed and distance**

**Use Q9\_a.csv**

|  |  |  |
| --- | --- | --- |
|  | **Skewness** | **Kurtosis** |
| **Speed** | **-0.117** | **-0.508** |
| **dist** | **0.806** | **0.405** |

* **‘speed’ is negatively skewed so the distribution of the data is towards the right side.**
* **‘dist’ is positively skewed so the distribution of the data is towards the left side.**

**SP and Weight(WT)**

**Use Q9\_b.csv**

|  |  |  |
| --- | --- | --- |
|  | **Skewness** | **Kurtosis** |
| **SP** | **1.611** | **2.977** |
| **WT** | **-0.614** | **0.950** |

* **‘SP’ is positively skewed so the distribution of the data is towards the left side.**
* **‘WT’ is negatively skewed so the distribution of the data is towards the right side.**
* **Both ‘SP’ and ‘WT’ has positive kurtosis.**

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



* It’s a positively skewed distribution
* Majority of chickweight is in the range of 50 to 100
* It may indicate under weighed and over weighed



* It’s a positively skewed distribution
* Outliers are present

**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?

n=2000

X\*=200

STD=30

For 94% Confidence Interval, Z=1.880794

For 96% CI, Z=2.0537

For 98% CI, Z=2.3263

|  |  |  |  |
| --- | --- | --- | --- |
|  | 94% | 96% | 98% |
| + | 201.26 | 201.37 | 201.56 |
| - | 198.73 | 198.62 | 198.43 |

**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?

|  |  |
| --- | --- |
| Mean | 36.667 |
| Median | 41 |
| Variance | 42.88 |
| Standard Deviation | 6.548 |

* Outliers are present in the data.
* Data is positively distributed.
* Majority of the marks obtained is between 35 to 45.

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

Symmetrical normal distribution with no skewness.

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

Positively skewed

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

Negatively skewed

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

High peakedness and less variance.

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

Lower peakedness and more variance

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



What can we say about the distribution of the data?

Not a Normal Distribution

What is nature of skewness of the data?

Negatively skewed

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

IQR=Upper Quartile-Lower Quartile=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.

* Both boxplots have the same medians around 265.
* Outliers doesn’t exist in both box plots.
* Data is not skewed in either of the box plots and is 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

Mean of MPG=34.4220757

* 1. P(MPG>38) = 0.4074
  2. P(MPG<40) = 0.7531

c. P(20<MPG<50) = 0.8518

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

MPG is Normally Distributed

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

Dataset: wc-at.csv

Adipose Tissue follows Normal Distribution

Waist Circumference doesn’t follow Normal Distribution

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

At 90% = Z = 1.69

94% = Z = 1.88

60% = Z = 2.49

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

At 95% = t = 2.064

96% = t = 2.169

99% = t = 2.789

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

µ=270, ẋ=260, SD=90, n=18, DOF=n-1=18-1=17

t-score=-0.471

P(t)=0.321639=32%