|  |  |
| --- | --- |
| 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) | **Nominal** |

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 | **Ordinal** |
| Level of Agreement | **Ordinal** |
| IQ(Intelligence Scale) | **Interval** |
| Sales Figures | **Ratio** |
| Blood Group | **Nominal** |
| Time Of Day | **Interval** |
| Time on a Clock with Hands | **Interval** |
| Number of Children | **Nominal** |
| Religious Preference | **Nominal** |
| Barometer Pressure | **Ratio** |
| SAT Scores | **Interval** |
| Years of Education | **Interval** |

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

S = (HHH,HHT,HTH,THH,HTT,THT,TTH,HHH)

Interested event = (HHT,HTH,THH)

Therefore, Reqd. Probability = **3/8**

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

1. Equal to 1 = Ans : Zero
2. Less than or equal to 4 = Ans : 6/36
3. Sum is divisible by 2 and 3 = Ans : 6 / 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?

Total no of events: ncr7c2=21

None of the balls is blue=5c2=10

Ans : Probability of 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

Expected Value = ∑xP(x) = 1\*0.015+ 4\*0.20 +3\*0.65 +5\*0.005 + 6\*0.01 +2\*0.120

**Ans = 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 | Score | Weigh |
| Mean | **3.59** | **3.21** | **17.85** |
| Median | **3.69** | **3.325** | **17.71** |
| Variance | **0.28** | **0.95** | **3.19** |
| StandardDeviation | **0.53** | **0.97** | **1.78** |
| Range | **2.76 - 4.93 = 2.17** | **1.513 - 5.424** = **3.91** | **14.5 - 22.9**  **= 8.39** |

**Use Q7.csv file**

Answer: Solution in Assignment-1-Q7(Basic Statistics Level-1).ipynb

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 = E(x ) = 1308 / 9 = 145.33**

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

**Cars speed and distance**

**Use Q9\_a.csv**

Answer: Solution in Assignment-1-Q9\_a(Basic Statistics Level-1).ipynb

**SP and Weight(WT)**

**Use Q9\_b.csv**

Answer: Solution in Assignment-1-Q9\_b(Basic Statistics Level-1).ipynb

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



* The Histogram Distribution is Right Skewed.



* The Boxplot is suggests that the distribution has lots of outliers towards upper extreme.

**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?

**94% Confidence:**

X-bar = 200

Sd = 30

n = 2000

Interval Estimate = X-bar ± Z\*Sd/sqrt(n)

=200 ± 1.88\*30/sqrt(2000)

= [ **198.74 – 201.26 ]**

>**98% Confidence:**

X-bar = 200

Sd = 30

n = 2000

Interval Estimate = X-bar ± Z\*Sd/sqrt(n)

=200 ± 2.33\*30/sqrt (2000)

= [ **198.44-201.56 ]**

>**96% Confidence:**

X-bar = 200

Sd = 30

n = 2000

Interval Estimate = X-bar ± Z\*Sd/sqrt(n)

=200 ± 2.05\*30/sqrt (2000)

= [**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.

* M**ean = 41**
* **Median = 40.5**
* **Variance = 24.11**
* **Standard Deviation = 5**

2) What can we say about the student marks?

* Student’s marks are Normally Distributed (because mea , median and mode value are near)

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

**Skewness = 0, Symmetric**

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

**Right Skewed (positive Skewness)**

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

**Left Skewed (Negative Skewness)**

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

**A distribution with a positive kurtosis value indicates that the distribution has heavier tails than the normal distribution.**

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

**A distribution with a negative kurtosis value indicates 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?

**Not a Normal Distribution**

What is nature of skewness of the data?

**Left Skewed**

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

**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)- Shorter Boxplot indicates that data is around center values.

b)- Long Boxplot indicates that data implies more variable data.

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)

Answer: Solution in Assignment-1-Q20 (Basic Statistics Level-1).ipynb

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

ANS = Normal Distribution because mean and median both are same

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

Dataset: wc-at.csv

ANS = not a Normal Distribution because mean and median both are approximately different .

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

**Ans, Z-score at 90% CI = 1.6448**

**Z-score at 94% CI = 1.8808**

**Z-score at 60% CI = 0.8416**

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

**, T-score at 95% CI = 2.0639**

**T-score at 96% CI = 2.1715**

**T-score at 99% CI = 2.7969**

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, n= 18, x bar = 260, s = 90**

**t score = (x bar - µ)/(s/sqrt(n))**

**=(260 – 270)/(90/sqrt(18))**

**=-10/21.23**

**=-0.47**

**Ans : 32.17%**