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

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

S={HHH,TTT,HHT,HTH,THH,HTT,THT,TTH}

2 heads and 1 tail-{ HHT,HTH,THH}

The propbabilty =3/8(0.375)

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

Total possible outcome=6^2=36

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| (1,1) | (1,2) | (1,3) | (1,4) | (1,5) | (1,6) |
| (2,1) | (2,2) | (2,3) | (2,4) | (2,5) | (2,6) |
| (3,1) | (3,2) | (3,3) | (3,4) | (3,5) | (3,6) |
| (4,1) | (4,2) | (4,3) | (4,4) | (4,5) | (4,6) |
| (5,1) | (5,2) | (5,3) | (5,4) | (5,5) | (5,6) |
| (6,1) | (6,2) | (6,3) | (6,4) | (6,5) | (6,6) |

1. **Equal to 1**

--- 0 .there is no outcome

1. **Less than or equal to 4**

(1,1),(1,2),(1,3),(2,1),(2,2),(3,1)

Probability= 6/36=0.166

1. **Sum is divisible by 2 and 3**

(1,5),(2,4),(3,3),(4,2),(5,1),(6,6)

Probability=6/36=0.166

**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 Balls=7

n(S)=No.of ways drawing 2 balls = 7C2

(7x6)/(2x1)=21

n(E)= noof way dawing balls ,none of which is blue=5C2

(5x4)/(2x1)=10

P(E)=n(E)/n(S)=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 \* 0.12

= 0.015 + 0.8  + 1.95 + 0.025 + 0.06 + 0.24

=       3.090

**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** | 3.596563 | 3.21725 | 17.84875 |
| **Median** | 3.695 | 3.325 | 17,71 |
| **Mode** | 3.92 | 3.44 | 17.84875 |
| **Variance** | 0.285881 | 0.957379 | 3.193166 |
| **Standard Deviation** | 0.534679 | 0.978457 | 1.786943 |
| **Range** | 2.17 | 3.911 | 8.39 |

**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?**

Expected value = Sum (X \* Probability of X) = (1/9)(108)+ (1/9)(110)+ (1/9)(123)+ (1/9)(134)+ (1/9)(145)+ (1/9)(167)+ (1/9)(187)+ (1/9)(199) = 145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

|  |  |  |
| --- | --- | --- |
|  | **Car Speed** | **Distance** |
| **Skewness** | -0.1175 | 0.8068 |
| **Kurtosis** | -0.5089 | 0.405 |

**SP and Weight(WT)**

**Use Q9\_b.csv**

|  |  |  |
| --- | --- | --- |
|  | **SP** | **Weight** |
| **Skewness** | 1.6114 | -0.6147 |
| **Kurtosis** | 2.9773 | 0.9502 |

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



The histograms peak has right skew and tail is on right. Mean > Median. We have outliers on the higher side.



The boxplot has outliers on the maximum side.

**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?**

Using the **t-distribution**, it is found that:

* The **94%** confidence interval is **(198.73, 201.27).**

*# Avg. weight of Adult in Mexico with 94% CI*

stats**.**norm**.**interval(0.94,200,30**/**(2000**\*\***0.5))

* The **96%** confidence interval is **(198.61, 201.39).**

*# Avg. weight of Adult in Mexico with 98% CI*

stats**.**norm**.**interval(0.98,200,30**/**(2000**\*\***0.5))

* The **98%** confidence interval is **(198.43, 201.57).**

*# Avg. weight of Adult in Mexico with 96% CI*

stats**.**norm**.**interval(0.96,200,30**/**(2000**\*\***0.5))

**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

Variance=24.11

Standard Deviation=4.91

1. **What can we say about the student marks?**

Skew=1.54(Then more weight in the left tail of the distribution.) is positive

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

Zero Skewness

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

 If the mean is greater than the median, the distribution is positively skewed

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

 If the median is greater than the mean, the distribution is negatively skewed

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

 If the distribution is light-tailed and the top curve steeper, like pulling up the distribution, it is called Positive Kurtosis.

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

The data can be heavy-tailed, and the peak can be flatter, almost like punching the distribution or squishing it. This is called Negative Kurtosis

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



**What can we say about the distribution of the data?-** Not distributed normally

**What is nature of skewness of the data?-**Negative skewness

**What will be the IQR of the data (approximately)?** :10-18

**Q19) Comment on the below Boxplot visualizations?**



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

Data in Boxplot1 distributed within range of Boxplot 2 data,share same mean and median

**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)**

stats**.**norm**.**cdf(38,cars**.**MPG**.**mean(),cars**.**MPG**.**std())

:

0.3475939251582705

* 1. **P(MPG<40)**

stats**.**norm**.**cdf(40,cars**.**MPG**.**mean(),cars**.**MPG**.**std())

0.7293498762151616

**c. P (20<MPG<50)**

stats**.**norm**.**cdf(0.50,cars**.**MPG**.**mean(),cars**.**MPG**.**std())**-**stats**.**norm**.**cdf(0.20,cars**.**MPG**.**mean(),cars**.**MPG**.**std())

 0.013000000000000012

**Q 21) Check whether the data follows normal distribution**

1. **Check whether the MPG of Cars follows Normal Distribution**

Yes,it follow normal Distribution

**Dataset: Cars.csv**

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

It does not follow normal Distribution

**Dataset: wc-at.csv**

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

*print('Z score for 90% Conifidence Intervla =',np.round(stats.norm.ppf(.05),4))*

Z score for 60% Conifidence Intervla = -1.6449

*print('Z score for 94% Conifidence Intervla =',np.round(stats.norm.ppf(.03),4))*

Z score for 94% Conifidence Intervla = -1.8808

*print('Z score for 60% Conifidence Intervla =',np.round(stats.norm.ppf(.02),4))*

Z score for 94% Conifidence Intervla = -2.0537

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

*print('T score for 95% Confidence Interval =',np.round(stats.t.ppf(0.025,df=24),4))*

T score for 95% Confidence Interval = -2.0639

*print('T score for 94% Confidence Interval =',np.round(stats.t.ppf(0.03,df=24),4))*

T score for 94% Confidence Interval = -1.974

*print('T score for 99% Confidence Interval =',np.round(stats.t.ppf(0.053,df=24),4))*

T score for 99% Confidence Interval = -1.6796

**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**

Using t distribution

(260-270)/(90/sqrt(18))

=-10/21.21

=-0.471

p\_value**=**1**-**stats**.**t**.**cdf(abs(**-**0.4714),df**=**17)

p\_value

0.32167411684460556