|  |  |
| --- | --- |
| 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 | Interval |
| 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 | Ratio |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Ordinal |
| Years of Education | Ordinal |

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

Ans:-

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

Probability that two heads and one tail are :

P= {HHT, THH, HTH}

P=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 – (1,3) (3,1) (2,2) (2,1) (1,2) (1,1) = 1/6
3. Sum is divisible by 2 and 3 =

Sum is divisible by 2 = (1,1) (1,3) (1,5) (2,2) (2,4) (2,6) (3,1) (3,5) (4,2) (4,4) (4,6) (5,1) (5,5) (6,6)

Sum is divisible by 3 = (1,2) (1,5) (2,1) (2,4) (3,3) (3,6) (4,5) (6,6)

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:-

Total number of balls = (2 + 3 + 2) = 7

Let S be the sample space.

Then, n(S) = Number of ways of drawing 2 balls out of 7

= 7C2

​= (7\*6)/(2\*1)

=21

Let E = Event of drawing 2 balls, none of which is blue.

n (E)= Number of ways of drawing 2 balls out of (2 + 3) balls.

=

5C2

=(5\*4)/(2\*1)

=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

Ans:-

=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.

**Use Q7.csv file**

**Ans :**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **POINTS** | **SCORE** | **WEIGH** |
| **MEAN** | **3.59** | **3.21** | **17.84** |
| **MEDIAN** | **3.69** | **3.32** | **17.71** |
| **MODE** | **3.92** | **3.44** | **17.02** |
| **STD** | **0.53** | **0.97** | **1.78** |
| **VARIANCE** | **0.28** | **0.95** | **3.19** |
| **RANGE** | **2.17** | **3.91** | **8.4** |

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 :

Probability of selecting each patient = 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)

= 143.88

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans=**

**Skewness =**

Index 0.000000

speed -0.117510

dist 0.806895

**speed distribution is left skewed i.e negative skewness**

**dist distribution is right skewed i.e positive skewness**

**Kurtosis =**

Index -1.200000

speed -0.508994

dist 0.405053

dtype: float64

**speed distributions is platykurtic i.e negative kurtosis**

**dist distribution is leptokurtic i.e positive kurtosis**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans =**

**Skewness=**

**SP=1.6114 = Positive skewness**

**WT= -0.6147= Negative Skewness**

**Kurtosis =**

**SP = 2.9773 = Positive kurtosis**

**WT= 0.9502 = Positive kurtosis**

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



Ans =

Given histogram shows the values of ChickWeight to Frequency.



Above boxplot suggest that the distribution has too many 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?

Ans =

By using formula for t-score statitics

C.I for 94% = 201.262,198.738

C.I. for 96% = 201.378 , 198.622

C.I. for 98% = 201.561, 198.439

**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 = 24.11

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 and has Zero skewness.

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

Ans:-

If mean is greater than the median then the distribution is positively skewed.

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

Ans:-

If median is greater than mean then the distribution is Negatively skewed.

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

Ans:-

When tail on the right side of the distribution is longer then the mean and the median will be greater than the mode.

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

Ans:-

When the tail on the left side of the distribution is longer than the right side then the mean and median will be less than the mode.

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



What can we say about the distribution of the data?

Ans:-It is nota normal distribution.

What is nature of skewness of the data?

= It is Negatively skewed

What will be the IQR of the data (approximately)?   
  
= **Inter Quartile Range =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.

= **1) The median of the two boxplots are same approximately 260.**

**2) The boxplots are not skewed in +ve or –ve direction.**

**3) Outliers doesn’t exist in both of the boxplots.**

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) = 0.3475
  2. P(MPG<40) = 0.7293

c. P (20<MPG<50) =0.89

Ans:-

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans:-

Mean of MPG:-34.42

Median of MPG:-35.12

Hence mean and median are approximately same the MPG of the cars follow normal distribution

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

Dataset: wc-at.csv

Ans:-

|  |  |  |
| --- | --- | --- |
|  | mean | median |
| WC | 91.90 | 90.08 |
| AT | 101.89 | 96.54 |

Both the Adipose TISSUE and WAIST CIRCUMFERENCE data set do follow Normal Distribution.

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

Ans:-

For 90% CI = stats.norm.ppf(0.95)

= 1.6448

For 94%CI = stats.norm.ppf(0.97)

= 1.8807

For 60% CI = stats.norm.ppf(0.80)

=0.8416

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

Ans =

For 95% CI=stats.t.ppf(0.975,24)

=2.063

For 96% CI = stats.t.ppf(0.98,24)

=2.1715

For 99% CI = stats.t.ppf(0.995,24)

=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

Ans=

Given data=

Mean = 270

Sample mean = 260

Sample S.D. = 90

N= 18

Df= n-1 = 18-1 = 17

H0 = Avg. life of bulb >=260 days

Ha = Avg life of bulb<260 dqys

Find t by using t score formula

T = - 0.47

Calculate p value by using

1-stats.t.cdf(abs(t\_scores),df=n-1)

P = 0.3216

Assume significance value = 0.05

If p value is less than significance value (alpha) then reject H0 and accept Ha or vice versa.

As 0.32 is greater than alpha.

Accept H0

The CEO claims are false and the average life of bulb is >260dys.