|  |  |
| --- | --- |
| 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 | Discrete |
| Blue Color | Discrete |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Continuous |
| 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 | Ordinal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Interval |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Interval |
| Years of Education | Ratio |

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

Total outcomes when 3 coins are tossed = 8

HHH, HHT, HTH, THH, TTT, TTH, THT, HTT = 8

Possibility of 2 heads and one tail is 3.

P (2H & 1T) = 3/8

We can also us probability distribution function of a binomial distribution formula to calculate the probability.

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

(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)

Total number of outcomes = = 36

1. Equal to 1 = 0

Favourable Outcome of (sum equal to 1) = 0

because the dice minimum outcome starting from (1,1). There is no **0** value in Dice

**Answer is 0**

1. Less than or equal to 4

Total number of outcomes = (1,1), (1,2),(1,3), (2,1) (2,2), (3,1) = 6 outcomes.

i.e., 6/36 = 1/6

**Answer is 1/6**

1. Sum is divisible by 2 and 3

{2,3,4,5,6,7,

3,4,5,6,7,8

4,5,6,7,8,9

5,6,7,8,9,10,

6,7,8,9,10,11,

7,8,9,10,11,12}

Total favourable outcomes are 6

Probability = 6/36 = 1/6

**Answer is 1/6**

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?

(R,R) (G,G,G) (B,B)

Total number of balls (n)= 7

Number of things to be chosen(r) = 2

Using nCr = n! / r! \* (n – r)! = 7!/ 2! \* (7-2)!

= 7!/ 2! \* 5!

Total favorable outcome = 21

Event of drawing 2 balls none of which is blue (n) = 2R+3G = 5

Number of things to be chosen(r) = 2

nCr = n! / r! \* (n – r)! = 5!/ 2! \* (5-2)!

= 5! / 2! \* 3!

Total possibility outcomes = 10

**Answer : Probability = 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 number of candies for a randomly selected child =

P(E) = 1\*0.015+4\*0.20+3\*0.65+5\*0.005+6\*0.01+2\*0.120

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

**Answer : P(E) = 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**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Points | Score | Weigh |
| Mazda RX4 | 3.9 | 2.62 | 16.46 |
| Mazda RX4 Wag | 3.9 | 2.875 | 17.02 |
| Datsun 710 | 3.85 | 2.32 | 18.61 |
| Hornet 4 Drive | 3.08 | 3.215 | 19.44 |
| Hornet Sportabout | 3.15 | 3.44 | 17.02 |
| Valiant | 2.76 | 3.46 | 20.22 |
| Duster 360 | 3.21 | 3.57 | 15.84 |
| Merc 240D | 3.69 | 3.19 | 20 |
| Merc 230 | 3.92 | 3.15 | 22.9 |
| Merc 280 | 3.92 | 3.44 | 18.3 |
| Merc 280C | 3.92 | 3.44 | 18.9 |
| Merc 450SE | 3.07 | 4.07 | 17.4 |
| Merc 450SL | 3.07 | 3.73 | 17.6 |
| Merc 450SLC | 3.07 | 3.78 | 18 |
| Cadillac Fleetwood | 2.93 | 5.25 | 17.98 |
| Lincoln Continental | 3 | 5.424 | 17.82 |
| Chrysler Imperial | 3.23 | 5.345 | 17.42 |
| Fiat 128 | 4.08 | 2.2 | 19.47 |
| Honda Civic | 4.93 | 1.615 | 18.52 |
| Toyota Corolla | 4.22 | 1.835 | 19.9 |
| Toyota Corona | 3.7 | 2.465 | 20.01 |
| Dodge Challenger | 2.76 | 3.52 | 16.87 |
| AMC Javelin | 3.15 | 3.435 | 17.3 |
| Camaro Z28 | 3.73 | 3.84 | 15.41 |
| Pontiac Firebird | 3.08 | 3.845 | 17.05 |
| Fiat X1-9 | 4.08 | 1.935 | 18.9 |
| Porsche 914-2 | 4.43 | 2.14 | 16.7 |
| Lotus Europa | 3.77 | 1.513 | 16.9 |
| Ford Pantera L | 4.22 | 3.17 | 14.5 |
| Ferrari Dino | 3.62 | 2.77 | 15.5 |
| Maserati Bora | 3.54 | 3.57 | 14.6 |
| Volvo 142E | 4.11 | 2.78 | 18.6 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Mean** | **Median** | **Mode** | **Variance** | **Standard Deviation** | **Range** |
| **Points** | 3.59 | 3.695 | 3.92 | 0.29 | 0.53 | 2.17 |
| **Score** | 3.21 | 3.325 | 3.44 | 0.95 | 0.98 | 3.91 |
| **Weigh** | 17.8 | 17.71 | 17.02 | 3.19 | 1.79 | 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?

Expected Value = ∑ (probability \* Value)

Probability of selecting each patient = 1/9

Expected Value = (

**Answer: Expected value = 145.33**

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

**Cars speed and distance**

**Use Q9\_a.csv**

Skewness for Speed = -0.12

Kurtosis for Speed = -0.51

**A picture containing text, diagram, screenshot, plot

Description automatically generated**

Skewness for dist = -0.81

Kurtosis for dist = 0.41

A picture containing text, diagram, screenshot, plot

Description automatically generated

**SP and Weight(WT)**

**Use Q9\_b.csv**

Skewness for SP = 1.61

Kurtosis for SP = 2.98

A picture containing diagram, text, screenshot, plot

Description automatically generated

Skewness for WT = -0.61

Kurtosis for WT = 0.95

A picture containing text, diagram, screenshot, plot

Description automatically generated

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





**Answer:** The histograms peak has right skew and tail is on right.

Mean and Median is greater than Mode.

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?

**For 94% Confidence interval**

**Range= [198.54, 201.45]**

**For 98% Confidence Interval**

**Range = [198.43, 201.56]**

**For 96% Confidence Interval**

**Range = [198.62, 201.37]**

**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?

**Answer 1 :**

Mean = 41

Median = 40.5

Mode = 41

Variance =5.05

Standard Deviation = 25.52

**Answer 2**:

Here we can see Mean>Median, so the distribution is a slightly skewed towards right. And no outliers are present.

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

**Answer**: No Skewness, Symmetric

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

**Answer**: +ve Skewness(Right Side Skew)

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

**Answer**: -ve Skewness(Left side Skew)

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

**Answer**: It means the curve is more peaked and it is Leptokurtic.

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

**Answer**: It means the curve is less peaked and it is Platickurtic.

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



What can we say about the distribution of the data?

**Answer**: It is not normally distributed and the median is towards the higher value. Which means Negavtively Skewed.

What is nature of skewness of the data?

**Answer**: Left Skewed(-ve)

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

IQR = Upper Quartile – Lower Quartile

= 18 – 10

IQR = 8

Q19) Comment on the below Boxplot visualizations?



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

**Answer**:

As we can see Boxplot 1 and Boxplot2 median values are approximately same which is approx 260.

Both the Boxplot 1 and 2 are not normally distributed and there are no outliers.

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)

**Answer**: 0.35

* 1. P(MPG<40)

**Answer**: 0.729

c. P (20<MPG<50)

**Answer**: 0.898

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

MPG follows Normal distribution A picture containing diagram, plot, text, screenshot

Description automatically generated

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

Dataset: wc-at.csv

WC follows normal distribution

A picture containing text, diagram, plot, screenshot

Description automatically generated

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

**Answers**:

For 90% confidence interval Z score is -1.64

For 94% confidence interval Z score is -1.88

For 60% confidence interval Z score is -0.84

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

**Answers:**

For 95% confidence interval t score is 2.06

For 96% confidence interval t score is 2.17

For 99% confidence interval t score is 2.79

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

**Answer**:

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

t-score= -0.47

Required probability = 0.32=32%