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

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

* Solution: - when three coins are tossed the total number of possible combinations are  23  = 8

The combinations are HHH, HHT, HTH, THH, TTH, THT, HTT, TTT

The number of combinations which have two heads and one tail are: - HHT, HTH, TTH which makes 3 in number.

i.e. No of desired outcomes = 3

P(Two Heads and one tail) = 3/8

=**0.038**

**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

* Solution: -

6\*6 = 36 Possible Outcomes

(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

Total number of outcomes = 62 = 36

Number of outcomes sum equal to 1 = 0 (i.e. not possible that sum always exceed to 1)

P (Equal to 1) = 0/36

= **0**

1. Less than or equal to 4

Total Number of outcomes = 62 =36

There are six outcomes where the sum is less than or equal to 4: these are: (1, 1), (1, 2), (1, 3), (2, 1), (2–, 2), (3, 1).

P(Less than or equal to 4) = 6/36

= **0.1666**

1. Sum is divisible by 2 and 3

Total Number of outcomes = 62 =36

There are 5 outcomes where the sum is divisible by 2 and 3:

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

Sum is divisible by 2 and 3 = 5/36

= **0.1388**

**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?**

* Solution: - P (2R, 3G, 2B)

Total No of Balls is 7 and we need to find the probability that none of the drawn ball is blue.

Two balls can be drawn in 7C2 ways i.e. (7 \* 6) / (2 \* 1) = 21

Number of ways of drawing 2 balls such as that none is blue = number of ways of drawing 2 balls and 3 green balls

= 5C2 = (5\*4) / (2\*1) = 10

**Probability is 10/21 = 0.476**

**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

* Answer – Expected number of candies for randomly selected child

= 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**

* **Answer: -**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Points | Score | Weight |
| Mean | 3.596563 | 3.21725 | 17.84875 |
| Median | 3.695 | 3.325 | 17.71 |
| Mode | 3.92 | 3.44 | 17.02 |
| Variation | 0.285881 | 0.95738 | 3.193166 |
| Std. Deviation | 0.534679 | 0.9784574 | 1.786943 |
| Min | 2.76 | 1.513 | 14.5 |
| Max | 4.93 | 5.424 | 22.9 |
| Range | 2.17 | 3.911 | 8.4 |

# Also kindly check the .ipynb file for this question

**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?

* Answer

Expected Value = ∑ (probability \* Value)

 ∑ P(x).E(x)

there are 9 patients

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

= (1/9) ( 108 + 110 + 123 + 134 + 135 + 145 + 167 + 187 + 199)

= (1/9)  (  1308)

= 145.33

**Expected Value of the Weight of that patient = 145.33**

In Simple Words Expected Value is nothing but the average value of the given data.

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

1. **Cars speed and distance**

**Use Q9\_a.csv**

* **Answer: -** Skewness: - Index 0.000000 speed -0.117510 dist 0.806895

Kurtosis: - Index -1.200000 speed -0.508994 dist 0.405053

1. **SP and Weight(WT)**

**Use Q9\_b.csv**

* **Answer: -** Skewness: - SP 1.611450 WT -0.614753

Kurtosis: - SP 2.977329 WT 0.950291

Kindly check .ipynb file for the answer

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



Histogram: -

* The Data is skewed on the right side i.e. data is positively skewed.
* There are no outliers in the data
* The most of the data points are connected in the range 50 – 100 with frequency 200. And least range of weight is 400 somewhere around 0 – 10. So, the expected value the above distribution is 75
* We can see a long tail towards right so it is heavily right skewed.



Boxplot: -

* Median is less than mean right skewed and we have outliers on the upper side of box plot
* Data is distributed on the right and positively skewed.

**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?

* Answer: -

Average Weight of adult in mexico with 94% Confidence Interval

(198.738325292158, 201.261674707842)

Average Weight of adult in mexico with 98% Confidence Interval

(198.43943840429978, 201.56056159570022)

Average Weight of adult in mexico with 96% Confidence Interval

(198.62230334813333, 201.37769665186667)

Please check .ipynb file

**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

Mode – 41

Standard Deviation - 5.05266382858645

Variation - 25.529411764705884

1. What can we say about the student marks?

from above plot we can say that mean of marks of student is 41 which is slightly greater than median. Most of the students got marks in between 41-42, there are two outliers 49,56.

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

* Answer: -

If the mean of the data is equal to the median of the data, then the skewness is zero

No skewness is present we have a perfect symmetrical distribution

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

* Answer: -

If the Mean > Median then the distribution is positively skewed, the values are more concentrated towards the right side, and the left tail is spread out, Hence, the statistical result are bent towards the left hand side. Hence the mean and median are always positive

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

* Answer: -

If the Median > Mean the distribution is negatively skewed, the data points are more concentrated towards right hand side of the distribution. This makes the mean, median bend towards the right. Hence these values are always negative.

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

* Answer: -

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

Positive Kurtosis values indicates that distribution is peaked and possesses thick tail. Extremely positive kurtosis indicates a distribution where more numbers are located in the tails of the distribution instead of around the mean.

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

* Answer: -

A distribution with a negative kurtosis value indicates that the distribution has lighter tails than the normal distribution.

Negative Kurtosis means the curve will be flatter and broader

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



What can we say about the distribution of the data?

* Answer: -

The above Boxplot is not normally distributed the median is towards the higher value

What is nature of skewness of the data?

* Answer: -

The data is a skewed towards left. The whisker range of minimum value is greater than maximum

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

* Answer: -

The Inter Quantile Range = Q3 Upper quartile – Q1 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.

* Answer: -
* There are no outliers in above boxplots.
* Both the boxplot shares the same median that is approximately in a range between 250 to 275.
* They are normally distributed with zero to no skewness neither at the minimum or maximum range.

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) = 0.34759392515827137
  2. P(MPG<40) = 0.7293498762151609

c. P (20<MPG<50) = 1.2430968797327613e-05

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Answer: - MPG of cars follows 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

Answer: -  Adipose Tissue (AT) and Waist does not follow Normal Distribution

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

Answer: - Z – Scores: -

90% Confidence Interval = 1.6448536269514722

94% Confidence Interval = 1.8807936081512509

60% Confidence Interval = 0.8416212335729143

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

Answer: - T – Scores: -

95% Confidence Interval = 2.0638985616280205

96% Confidence Interval = 2.1715446760080677

99% Confidence Interval = 2.796939504772804

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