|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete- Countable |
| Results of rolling a dice | Discrete- Countable |
| 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 – Nominal |
| Number of kids | Discrete – Countable |
| Number of tickets in Indian railways | Discrete – Countable |
| Number of times married | Discrete – Countable |
| Gender (Male or Female) | Discrete – 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 | Discrete – Nominal |
| High School Class Ranking | Discrete – Nominal |
| Celsius Temperature | Continuous - Interval |
| Weight | Continuous - Ratio |
| Hair Color | Discrete - Ratio |
| Socioeconomic Status | Continuous - Interval |
| Fahrenheit Temperature | Continuous - Ratio |
| Height | Continuous - Ratio |
| Type of living accommodation | Discrete - Ordinal |
| Level of Agreement | Discrete - Interval |
| IQ(Intelligence Scale) | Discrete - Interval |
| Sales Figures | Discrete - Interval |
| Blood Group | Discrete - Ratio |
| Time Of Day | Continuous - Interval |
| Time on a Clock with Hands | Continuous - Interval |
| Number of Children | Discrete - Interval |
| Religious Preference | Discrete - Ratio |
| Barometer Pressure | Discrete - Interval |
| SAT Scores | Discrete - Ratio |
| Years of Education | Discrete - Nominal |

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

Answer:

When three coins are tossed the total number of possible combinations are 2^3 = 8

The number of combinations which have two heads and one tail are 3 (HHT, HTH, TTH)

P(Two heads and One Tail) = 3/8 = 0.375

(formula: number of ways of event/possibile outcomes)

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

Answer: Total possibility outcomes are 36 (6\*6-two Dice).

1. Probability equal to 1 = 0 because minimum number we should get is 2 when we roll 2 Dice.
2. Probability to get less than or equal to 4 = Below are the pair of outcomes which give sum of 4.

die 1    die 2   sum

  1         1        2

  1         2        3

  1         3        4

  2         1        3

  2         2        4

  3         1        4

So we have in total 6 ways of outcomes give sum of 4.

Answers will be 6/36 = 1/6.

1. Sum is divisible by 2 and 3=

Numbers that can be divided by 2&3 between 2 to 12 are 6,12.

Favorable outcomes are (1 , 5) , (3 , 3) , (4 , 2) , (5 , 1) , (6 , 6)

Probability = 5/36.

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?

Answer: Total number of balls = (2 + 3 + 2) = 7.

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

7C2= (7\*6)/(2\*1)=21.

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

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

Probability=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

Answer:

1 \* 0.015 + 4\*0.20 + 3 \*0.65 + 5\*0.005 + 6 \*0.01 + 2 \* 0.12

3.09

..Defined in python

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.

Answer: Please see attached python code and comments below.

Mean for Points = 3.59, Score = 3.21 and Weigh = 17.84

Median for Points = 3.69, Score = 3.32 and Weigh = 17.71

Mode for Points = 3.07, Score = 3.44 and Weigh = 17.02

Variance for Points = 0.28, Score = 0.95, Weigh = 3.19

Standard Deviation for Points = 0.53, Score = 0.97, Weigh = 1.78

Range for Points [3.59 – 4.93], Score [3.21 – 5.42] and Weigh [17.84 – 22.9].

**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?

Answer: Expected Value is = 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

..Defined in python

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Done in python**

Speed skewness: -0.12

Speed Kurtosis: -0.51

dist skewness: 0.81

dist kurtosis: 0.41

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Done in python**

SP skewness: 1.61

SP Kurtosis: 2.98

WT skewness: -0.61

WT kurtosis: 0.95

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



Done in Python

**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?

Calculated in python using normal distribution

Ans:

94% = (198.738325292158, 201.261674707842)

98%= (198.43943840429978, 201.56056159570022)

96%= (198.62230334813333, 201.37769665186667)

**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:41.0

Median:40.5

Variance: 24.11

standard deviation: 4.91

1. What can we say about the student marks?

The average marks of the students is 41 which is almost near to the median value. Therefore data can be symmetric. Each student marks are deviating at 4.91 from each other. Frequently repeated marks are between 36 to 42 where as data highly implemented with 56, which can be considered as outlier.

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

symmetric

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

Positively skewed

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

Negatively skewed

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

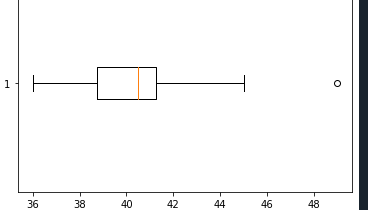
Outlier

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

All the data values are most near to the mean

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





What can we say about the distribution of the data?

Anything above 45 and below 36 can be considered as outlier. 25 percentile is 38.5 and 75% is 41.5 and mean is 41.

What is nature of skewness of the data?

Data is positively skewed.

What will be the IQR of the data (approximately)?   
IQR=11

Q19) Comment on the below Boxplot visualizations?

Even the data is expanded on the second diagram wide, mean is same on the both the data. As mean is in middle of the data, data is normally distributed. Range of 2 is more than 1.



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

1 has lesser range than 2. Whatever data is available in box 1 is also available in box 2. 1 has the high level of agreement each other. It have the same mean of 2.

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)

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Data is slightly negatively skewed.

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

Dataset: wc-at.csv

Waist following normal distribution. At is positively skewed.

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

1.64, 1.88, 0.84

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

2.06, 2.17 and 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

Ans: 0.321 (in favor of alternate hypothesis)