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

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

3A) 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 🡪1/6
3. Sum is divisible by 2 and 3 🡪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?

5A) 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

6A) 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**

7A)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Mean | Median | Mode | Variance | S.D. | Range |
| Points | 3.5966 | 3.695 | 3.07,3.92 | 0.2859 | 0.5347 | 2.17 |
| Score | 3.2173 | 3.325 | 3.44 | 0.9574 | 0.9785 | 3.911 |
| Weigh | 17.849 | 17.71 | 17.02,18.9 | 3.1932 | 1.7869 | 8.399 |

* Dispersion of data is more in weigh
* Extreme values are not affected the central tendencies

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?

8A)145.33

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

**Cars speed and distance Use Q9\_a.csv**

**SP and Weight (WT) Use Q9\_b.csv**

|  |  |  |
| --- | --- | --- |
|  | Skewness | Kurtosis |
| Cars speed | -0.1175 | -0.5089 |
| distance | 0.8069 | 0.4050 |
| SP | 1.6115 | 2.9773 |
| Weight | -0.6147 | 0.9503 |

9A)

* Distance and SP are positively skewed
* Cars speed and Weight are negatively skewed
* SP is highly skewed compared to remaining parameters
* SP following approximate normal distribution based on kurtosis

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



Histogram:

* The plot is positively skewed
* Frequency is drawn for every 50 intervals of weight

Boxplot:

* There are few out layers in the box plot
* IQR is shifted (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?

11A) Considering t distribution:

For 94% confidence interval🡪 (198.73, 201.27)

For 98% confidence interval🡪 (198.43, 201.)

For 96% confidence interval🡪 (198.61, 201.39)

**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?

12A) mean= 41, median= 40.5, variance= 25.529, S.D.= 5.052

There are two out layers 49 and 56

Negatively skewed distribution

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

13A) ZERO

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

14A) Positively skewed

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

15A) Negatively skewed

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

16A) Positive kurtosis indicates heavily tailed and variance is

Less

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

17A) Negative kurtosis indicates lightly tailed and variance is high

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



What can we say about the distribution of the data?

A) Positively Skewed with median of 15

What is nature of skewness of the data?

A) Positively Skewed

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

A) 8

Q19) Comment on the below Boxplot visualizations?



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

19A) Boxplot1 has less IQR compared to Boxplot2

Both the plots have same median of 262.5

Min and max values of boxplot1 are 237.5 and287.5

Min and max values of boxplot2 are 200 and 350

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

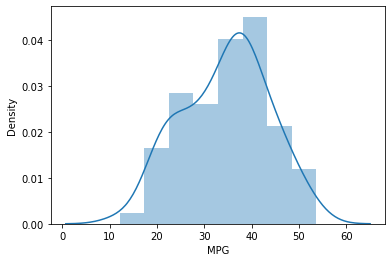
* 1. P(MPG>38) = 0.3475
  2. P(MPG<40) =0.729

c. P (20<MPG<50) =0.8988

Q 21) Check whether the data follows normal distribution

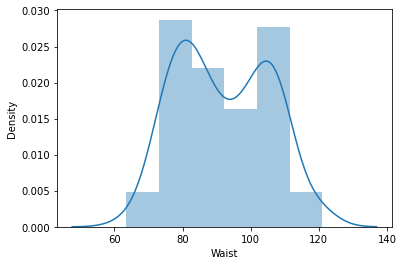
1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

1. Yes, MPG of Cars approximately follows Normal Distribution 
2. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

bA) No



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

22A) Z score for 90% confidence interval🡪1.645

Z score for 94% confidence interval🡪1.88

Z score for 94% confidence interval🡪0.841

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

23A) t score for 95% confidence interval🡪2.06

t score for 96% confidence interval🡪2.171

t score for 99% confidence interval🡪2.796

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

24A) Probability that 18 randomly selected bulbs would have an average life of no more than 260 days is 32.16% Assuming significance value α = 0.05

t= (260-270)/ (90/18\*\*0.5)

p\_value=1-stats.t.cdf(abs(t),df=17)=0.32167