|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete data |
| Results of rolling a dice | Discrete data |
| Weight of a person | Continuous data |
| Weight of Gold | Continuous data |
| Distance between two places | Continuous data |
| Length of a leaf | Continuous data |
| Dog's weight | Continuous data |
| Blue Color | Discrete data |
| Number of kids | Discrete data |
| Number of tickets in Indian railways | Discrete data |
| Number of times married | Discrete data |
| Gender (Male or Female) | Discrete data |

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-Nominal |
| Barometer Pressure | Continuous-Ratio |
| SAT Scores | Continuous-Ratio |
| Years of Education | Discrete-Nominal |

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

Answer-P{HHT}, P{THH}, P{HTH}

1/8+1/8+1/8=3/8

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

Ans a) Probability will be zero as sum cannot be zero(no such event occurs)

b)P{1,2}+P{2,2}+P{1,1}=three possibilities

3/36=1/12

1. P{1,5}, P{2,4}, P{3,3}, P{4,2},P{5,1}, P{6,6}

6/36=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?

Ans=2R,3G,2B

P(2 Red and 3Green)=20/42=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)

Expected number of candies=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 | 3.59656 | 3.21725 | 17.8 |
| median | 3.695 | 3.325 | 17.71 |
| mode | NA | NA | NA |
| std dev | 0.53468 | 0.97845744 | 1.7869 |
| min | 2.76 | 1.513 | 14.5 |
| max | 4.93 | 5.424 | 22.9 |
| range | 2.17 | 3.911 | 8.4 |
| variance | 0.28588 | 0.95737897 | 3.1932 |

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=Expected value =∑P(x).E(x)

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

Expected value of the Weight of that patient=145.3

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

**Cars speed and distance**

**Use Q9\_a.csv**

Ans=Skewness for speed= -0.11751 and kurtosis -0.50899

Skewness is –ve so it is left skewed

Skewness for distance 0.806895 and kurtosis 0.405053

It is right skewed

**SP and Weight(WT)**

**Use Q9\_b.csv**

Skewness for speed -0.34676 and kurtosis 0.275426-Left skewed

Skewness for distance -1.5564 and kurtosis 0.405053- Left skewed

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



Ans=Max data is covered in 50 to 100(weight) and has freq is 200

As it is extended more towards right it is more right skewed



Ans=Median<mean so it is right skewed. Also there are outliers on above 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?

Ans= confidence interval for 94% is 1.882 Confidence interval for 98%= 2.33 Confidence interval for 96% = 2.05

**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=40.11765, median=40.5, variance=25.52, standard deviation=5.052

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

Symmetrical

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

Ans=Distribution is +vely skewed

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

Ans= Distribution is -vely skewed

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

Ans=Distribution is peaked and has thick tails

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

Ans= Distribution is flat and has thin tails

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



What can we say about the distribution of the data?

What is nature of skewness of the data?

Ans=Left skewed

What will be the IQR of the data (approximately)?   
Ans=8

Q19) Comment on the below Boxplot visualizations?



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

Ans=Mean and median is equal, hence distribution is symmetrical

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)

Ans= P(MPG>38)=0.347593

P(MPG<40)=0.729349

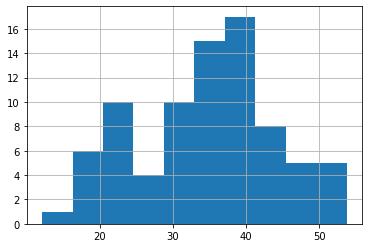
P(20<MPG<50)= 1.2430968

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

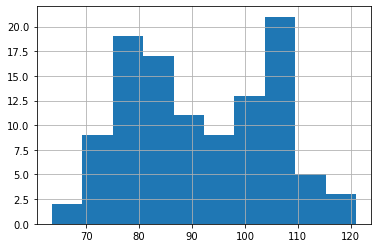
Ans=As per below histogram(constructed using python), we can conclude distribution is fairly normal

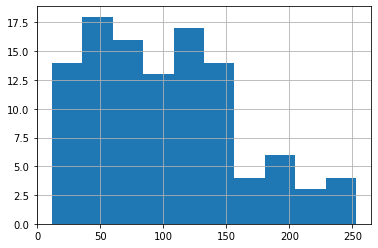


1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

Ans=





Data is symmetrical

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

Ans= Z scores of 90% confidence interval= 1.64485

Z scores of 94% confidence interval=10880793

Z scores of 60% confidence interval=0.84162

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

Ans= t scores of 95% confidence interval=2.060

t scores of 96% confidence interval=

t scores of 99% confidence interval=2.797

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=

t=

=

t=-0.47

p-value=0.32167