|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Discrete |
| Weight of a person | Discrete |
| Weight of Gold | Discrete |
| 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 | Ordinal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ordinal |
| Sales Figures | Ordinal |
| Blood Group | Interval |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Ordinal |
| Religious Preference | Ordinal |
| Barometer Pressure | Interval |
| SAT Scores | Ordinal |
| Years of Education | Interval |

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

Ans:- Three coins are tossed sample space is

S={(HHH),(HHT),(HTH),(THH),(TTH),(HTT),(TTT)}

n(S)=8

Number of two head and one tail is

A={(THH),(HTH),(HHT)}

n(A)= 3

P(A) = n(A)/n(S)

= 3/8

P(A) = 0.375

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:- The sample space of two dice are

S= {(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

There is a 0 sample for both sum is equal to 1 that’s why the probability 0.

1. Less than or equal to 4

Let assume A

A={(1,1),(1,2),(1,3), (2,1),(2,2),(3,1)}

n(A)=6

The probability of A is

P(A) = n(A)/n(S)

= 6/36

= 0.33333

1. Sum is divisible by 2 and 3

Let assume B

B = {(1,1),(1,2),(1,3),(1,5),(2,1),(2,2),(2,4),(2,6),(3,1),(3,3),(3,5),

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

(6,4),(6,6)}

n(B) = 24

P(B) = n(B)/n(S)

= 24/36

= 0.66666

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:- When first ball is drawn so we have 5 option out of 7 balls.

Probability= 5/7

When second ball is drawn so we have 4 option out of 6 balls.

Probability = 4/6=2/3

The probability that none of the balls drawn is blue are = 5/7\*2/3

= 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

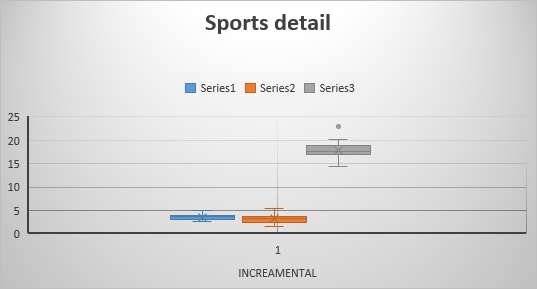
Answer:- 3.09

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.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | Points |  | Score |  | Weight |
| Mean |  | 3.5965625 |  | 3.21725 |  | 17.84875 |
| Median |  | 3.695 |  | 3.325 |  | 17.71 |
| Mode |  | 3.92 |  | 3.44 |  | 17.02 |
| Variance |  | 0.2769476 |  | 0.9274609 |  | 3.0933797 |
| Standard deviation | | 0.5346787 |  | 0.9784574 |  | 1.7869432 |
| Rang |  | 3.69 |  | 3.19 |  | 20 |

bg

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:- the weight of the patient are

108,110,123,134,135,145,167,187,199

Randomly chosen one of the patient is probability 1/9

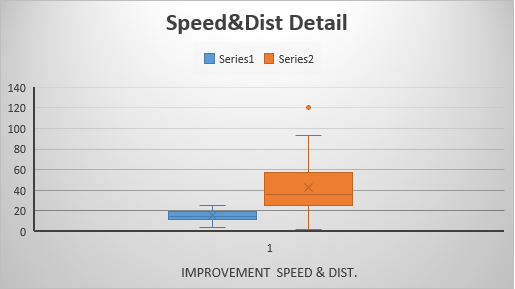
The expected value of weight of that patient are 145.33333333

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

**Cars speed and distance**

**Use Q9\_a.csv**

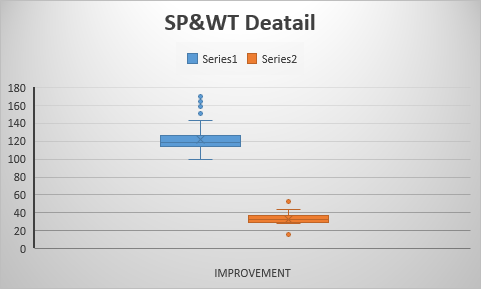
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Column1** | **Column2** | **Column3** | **Column4** | **Column5** |
| SPEED |  |  | DISTANCE |  |
| Skewnes | -0.1175099 |  | Skewness | 0.806895 |
| Kurtosis | -0.5089944 |  | Kurtosis | 0.4050526 |



**SP and Weight(WT)**

**Use Q9\_b.csv**

|  |  |  |
| --- | --- | --- |
| **Column1** | **Column2** | **Column4** |
|  | SP | WT |
| Skewness | 1.6114502 | -0.6147533 |
| Kurtosis | 2.9773289 | 0.9502915 |



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



**ANS:** The histogram has a right skewed data and it has outlier at the point of 350 to 400.



**ANS:** The higher ends of the boxplot has a outliers.

**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:-Sample mean=200

Sample Standard deviation = 30

Sample size=2000

To calculate 94% confidence interval using t distribution t=1.8916

200-1.8916\*30/root of 2000=201.27

To calculate 96% confidence interval using t distribution t=2.0673

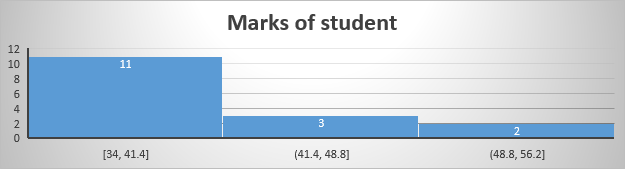
200-2.0673\*30/root of2000=

**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?

|  |  |  |
| --- | --- | --- |
| **Column1** |  | **Column2** |
| Mean |  | 41 |
| Median |  | 40 |
| Standard deviation | | 5.3193828 |
| Variance |  | 28.295833 |



The histogram of Student marks is show positive because is plot in left skewness.

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

Ans:- If Mean equals to median then skewness is zero because it’s look like symmetric data

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

Ans:- If mean > median then skewness show positively skewed because data graph plot left side.

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

Ans:- If median>mean then skewness is show negatively skewed because data graph plot right side.

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

Ans:- Positive kurtosis value indicates (3)meaning thinner peak and wide tails.

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

Ans:- Negative kurtosis value indicates (-3)meaning wider peak and Thinner tails.

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



What can we say about the distribution of the data?

ANS: Normal distribution

What is nature of skewness of the data?

ANS: Right skewness

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

ANS: (Q3-Q1)=(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.

**Ans**: Let’s assume boxplot 1 is boxplot A and boxplot 2 is boxplot B

Now the ranges of boxplot B is more than the boxplot A that means B data has more distributed or more no. of variable data to A data. The median of the both box are overlap or median are same with each other.

Here not shown outliers but boxplots B both ends are out of the range of data points

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)

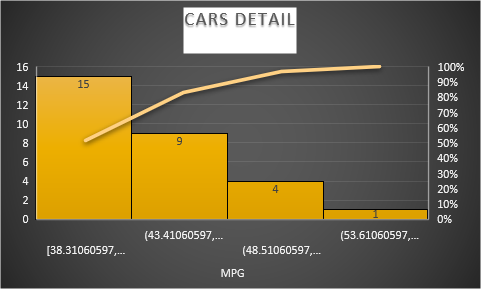
c. P (20<MPG<50)

ANS:- Attached File IPYNB name Q.20 Assignment-level 1

Q 21) Check whether the data follows normal distribution

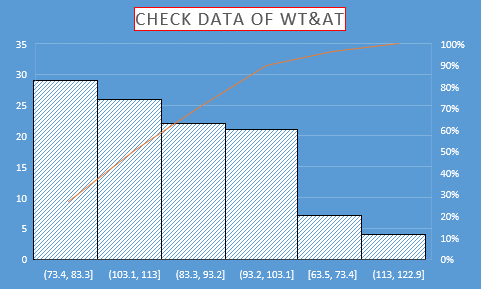
1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv



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

Dataset: wc-at.csv



Q22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval

|  |  |
| --- | --- |
| Confidence interval | Z-score |
| 90% | 1.645 |
| 94% | 1.8807 |
| 60% | 0.84162 |

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

|  |  |
| --- | --- |
| Confidence interval | t-Score |
| 95% | 2.063899 |
| 96% | 2.171545 |
| 99% | 2.79694 |

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:- population average = 270

Sample average= 260

No. of bulbs for testing=18

Sample standard deviation=90

t-distribution :-

t= Xbar-mu/s/sqrt of n

t= 260-270/90/sqrt of 18

t=-0.0249177

In r code pt(tscore,df)

Pt(-0.024,17,lower.tail=TRUE)

0.490566 is the probability that 18 randomly selected bulbs would have an average life of no more than 260 days.