Q1) Identify the Data type for the Following:

|  |  |
| --- | --- |
| 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 |

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

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

Answer -: When three Coins are tossed then the probablity that two heads and one tail are as follows:-

{ ( H,H,T);(T,H,H);(H,T,H) }

n(P)=3/8

Q4) Two Dice are rolled, find the probability that sum is

* Equal to 1
* Less than or equal to 4
* Sum is divisible by 2 and 3

Answer-: When two Dice are rolled then probability of that sum which is

* Equal to 1 :- Their is no possibility of getting 1.
* Less than or equal to 4 :- Possibility of getting sum less than or equal to 4 is

{ (1,1); (1,2); (1,3); (2,1); (2,2); (3,1) }

then n(P)=6/36

= 1/6.

* Sum is divisible by 2 and 3 :-Possibility of getting sum which is divisible by 2 and 3 is

{ (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(P)=24/36

=2/3

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

Let S be the sample space

(D) = Number of ways of drawing 2 balls out of 7

n(S)=7 ( 2+3+2)

=7C2

= (7\*6)/(2\*1)

=42/2

n(S)=21

Total probability is 21.

E= Event of 2 balls,none of which is blue

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

n(D)=5(2+3)

=5C2

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

=20/2

n(P)=10

Therefore, 10 is the probability that none of the balls drawn is blue.

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 a randomly selected child

=1\*0.015 + 4\*0.20 + 3\*0.65 + 5\*0.005 + 6\*0.01 + 2\*0.120

= 0.015 + 0.8 + 1.95 + 0.025 + 0.06 + 0.24

=3.09

Therefore, Expected number of candies for a randomly selected child ins 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:-

Mean, Median, Mode, Variance, Standard Deviation and Range is as follows:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Points | Score | Weigh |
| Mean | 3.596563 | 3.21725 | 17.84875 |
| Median | 3.695 | 3.325 | 17.71 |
| Mode | 3.92 | 3.44 | 17.02 |
| Variance | 0.2858814 | 0.957379 | 3.193166 |
| Standard  Deviation | 0.5346787 | 0.9784574 | 1.786943 |
| Range | 2.76 4.93 | 1.513 5.424 | 1.513 5.424 |

Q8) Calculate Expected Value for the problem below

* 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 of the Weight of a Paitent is:-

Mean = 145.3333.

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

**Cars speed and distance**

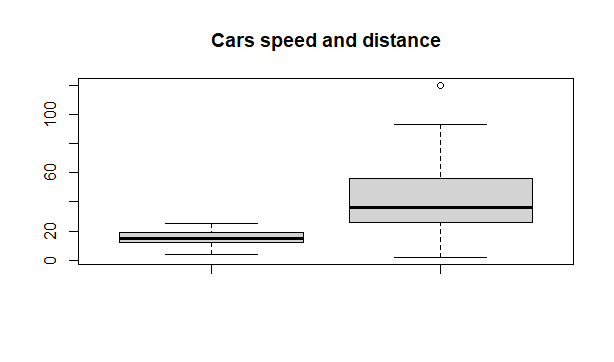
**Use Q9\_a.csv**

**SP and Weight(WT)**

**Use Q9\_b.csv**

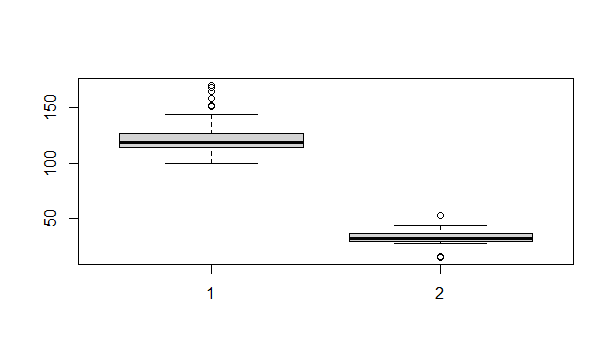
Answer:- Skewness, Kurtosis of the data of Cars speed and distance =

|  |  |  |
| --- | --- | --- |
|  | **Speed** | **Distance** |
| **Skewness** | **-0.1139548** | **0.7824835** |
| **Kurtosis** | **2.422853** | **3.248019** |



**SP and Weight(WT)**

|  |  |  |
| --- | --- | --- |
|  | **SP** | **Weight(WT)** |
| **Skewness** | **1.581454** | **-0.6033099** |
| **Kurtosis** | **5.723521** | **3.819466** |



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

Answer:- The above boxplot tells that the distribution has lots of outliers towards upper extreme.

**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 at 94% confidence interval [134.897 265.103]

Average weight of adult in Mexico at 98% confidence interval [122.725 277.275]

Average weight of adult in Mexico at 96% confidence interval [130.21 269.79]

**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**

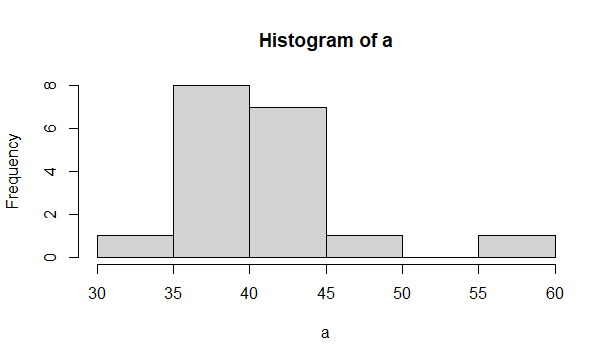
* Find mean, median, variance, standard deviation.
* What can we say about the student marks?

Answers:-

1) Mean, Median, Variance, Standard Deviation of scores obtained by a student

in tests =

|  |  |
| --- | --- |
| **Mean** | **41** |
| **Median** | **40.5** |
| **Variance** | **25.52941** |
| **Standard Deviation** | **5.052664** |



2) Mass of the students marks is between 38-42.

Skewness is 1.52 is positive because mass of marks is in left side of histogram.

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

Answer:- If the distribution is symmetric ,then the mean is equal to the median,And the distribution has zero skewness. Therefore if the mean and median are equal there will be No Skewness.

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

Answer:- The nature of skewness when mean>median is Positively Skewed.

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

Answer:- The nature of skewness when Median>mean is Negatively Skewed.

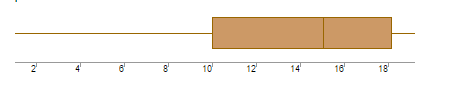
Q16) What does positive kurtosis value indicates for a data ?

Answer:- Positive values of kurtosis indicate that a distribution is peaked and possess thick tails.

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

Answer:- Negative values of kurtosis indicate that a 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?

Answer:- The data is not normally distributed.

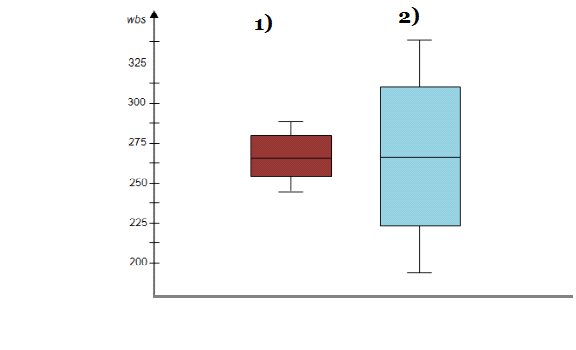
What is nature of skewness of the data?

Answer:-The nature of data is Negatively skewed.

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

Answer:- The IQR of the data is approximately 10 to 18.

Q19) Comment on the below Boxplot visualizations?



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

Answer:- Both the plot conclude that the data in boxplot 1 and boxplot 2 is normally distributed. Boxplot 1 is sample distribution and Boxplot 2 is a population distribution also there are no outliers in boxplot 1 and 2.

Q1 is 25%, Q3=75%, IQR is 50% for both the boxplots. So we can say both

the distribution follow normal distribution(mean=median=mode).

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

* P(MPG>38)
* P(MPG<40)

c. P (20<MPG<50)

Answer:- Probability of MPG>38 is 0.348

Probability of MPG<40 is 0.729

Probability of 20<MPG<50 is 0.214

Q 21) Check whether the data follows normal distribution

* Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Answer:- MPG of Cars does not follows Normal Distribution in Cars.csv

Data set.

* 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 Circumference(Waist) from wc-at data set does not follows the Normal Distribution.

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

Answer:- Z scores of 1) 90% confidence interval is 0.95.

2) 94% confidence interval is 0.97

3) 60% confidence interval is 0.80

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

Answer:- T scores of for sample size of 25 is

1) 95% confidence interval is 2.0638985616280205

2) 96% confidence interval is 2.0537489106318225

3) 99% confidence interval is 2.5758293035489004.

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

Answer:- t\_value = (260 - 270)/(90/np.sqrt(18))

print('critical value = ', np.round(t\_value, 2))

print('probabilty for average life of no more than 260 days is', np.round(stats.t.cdf(t\_value, df=17), 2))

critical value = -0.47

probabilty for average life of no more than 260 days is 0.32.