|  |  |
| --- | --- |
| 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) | Interval |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Interval |
| Number of Children | Ordinal |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Interval |
| 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=Possible outcomes

={HHH,HHT,HTT,HTH,THT,TTH,THH,TTT}

No of possible outcomes of three coins tossed simultaneously=8

Possible outcomes for getting two heads and one tail={HHT,HTH,THH}

No of possible outcomes for getting two heads and one tail=3

Probability of favorable outcomes=

no of possible outcomes for getting two heads and one tail /no of possible outcomes

of three coins tossed =3/8=0.37

Probability for getting two heads and one tail = 37%

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 Possible outcomes for two dice rolled =

{(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)}

No of possible outcomes for two dice rolled = 36

1. No of Possible outcomes for sum equal ot 1 = 0
2. Possible outcomes for sum less than or equal to 4 =

{(1,1),(1,2),(1,3),(2,1),(2,2),(3,1)} = 6 = 6/36 = 0.16

Probability of getting sum less than or equal to 4 = 16%

1. Possible outcomes for getting sum divisible by 2 & 3 = {(1,5),(2,4),(3,3),(4,2),(5,1),(6,6)} = 6 = 6/36 = 0.16

Probability of getting sum divisible by 2 & 3 is 16%

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:** Total no of balls is 2 Red + 3 Green + 2 Blue = 7 Balls

If r balls are drawn at random then the no of ways of drawing r balls out of n balls is

**nCr=n!/r!(n-r!)**

Hence if two balls are drawn at random then the no of ways of drawing two balls out

Of 7 balls is 7C2 = (7x6x5!)/(2x1)(5!) = (7x6)(2) = 21

For the two balls drawn at random the no of drawing two balls out of which none is

Blue is 5C2 = (5x4x3!)/(2x1)(3!) = 10

Therefore,the probability that none of the balls drawn is blue is 10/21 = 0.42

Probability of none of the balls drawn is blue is 47%

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:** Probability of expected no of candies for randomly selected child

= (1x 0.015)+(4 x 0.2)+(3 x 0.65)+(5 x 0.005)+(6 x 0.01)+(2 x 0.12) = 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.

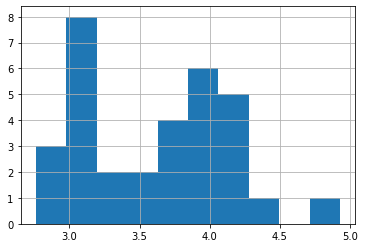
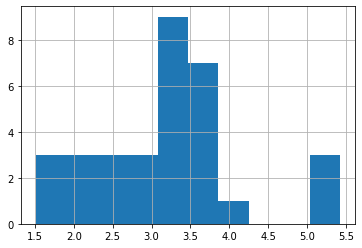
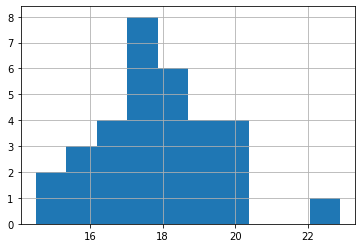
**Use Q7.csv file**

**Ans:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Data/Measures** | **Mean** | **Median** | **Mode** | **Variance** | **Standard**  **Deviation** | **Range** |
| **Points** | 3.596563 | 3.695 | 3.07,3.92 | 0.285881 | 0.534679 | 2.17 |
| **Score** | 3.217250 | 3.325 | 3.44 | 0.957379 | 0.978457 | 3.911 |
| **Weigh** | 17.848750 | 17.710 | 17.07,18.90 | 3.193166 | 1.786943 | 8.399 |

‘Points’ has less standard deviation compared to ‘score’ & ‘weigh’. It is recommended to have a lower standard deviation/variance for better insights & prediction quality of data.

**Points**  **Score Weigh**

 **** ****

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 Probability that one patient is chosen at random is = 1/9

Expected value = Sum of Probabilities \* values)

Expected value = (1/9)(108)+(1/9)(110)+(1/9)(123)+(1/9)(134)+(1/9)(145)

+ (1/9)(167)+(1/9)(187)+(1/9)(199) = 145.333

Therefore, expected value of patient is 145.333

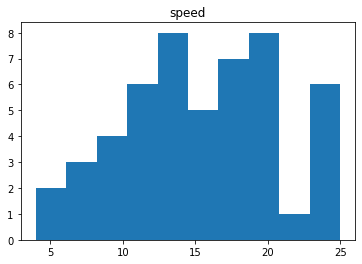
Q9) Calculate Skewness, Kurtosis & draw inferences on the following data

Cars speed and distance

**Use Q9\_a.csv**

**Ans:**

|  |  |  |
| --- | --- | --- |
| **Data/ Calculation** | **Skewness** | **Kurtosis** |
| **Speed** | -0.117510 | -0.508994 |
| **Weigth** | 0.806895 | 0.405053 |

****

Here we can say that the data is almost normally distributed.

****

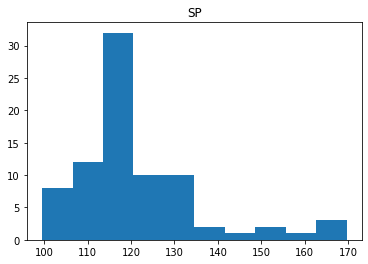
Here we can say that the data tends to shift towards right. Hence it's a right skewed or positively skewed data.

**SP and Weight(WT)**

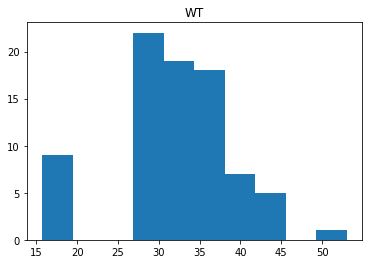
**Use Q9\_b.csv**

**Ans:**

|  |  |  |
| --- | --- | --- |
| **Data/ Calculation** | **Skewness** | **Kurtosis** |
| **Speed** | 1.611450 | 2.977329 |
| **Weight** | -0.614753 | 0.950291 |

****

Here, the distribution of the data is closer to the right, making it right-skewed.

****

Here also the spread of data is towards right hence it's a right skewed data. Also there is an outlier between 15-19.

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



**Ans:** a) The Histogram as per above graph is right/positively skewed. Skewness shows the direction of outliers, which means from right side of mean distribution there are outliers.

1. The Boxplot as per above graph, generally divided in 5 parts and also depicts that there are outliers (extreme values) on the top side (in circles). The box in boxplot describes about the Q1, Q2 & Q3 values. Q1 is the lower quartile, which has 25% of data. Q2 is the median or has 50% of data is below this. And Q3 is the upper quartile. The distribution is positively skewed as median is closer to the bottom of the box.

**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% = 201.2616747

Confidence Interval for 96% = 201.3776966

Confidence Interval for 98% = 201.560561

**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.

**Ans:** Mean:41.3125

Median:41.0

Variance:27.9625

Standard deviation:5.28795

1. What can we say about the student marks?

**Ans:** Mean of the marks of the students 41 is slightly greater than median 40.5

Marks of most of the students lie between 41 to 42 and 49 & 56 are outliers.

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

**Ans:** The distribution is symmetrical and has no skewness if the data's mean and median are equal.

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

**Ans:**The distribution is positively skewed if the mean is higher than the median.

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

**Ans:**The distribution is negatively skewed if the median is higher than the mean.

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

**Ans:**Positive kurtosis value indicate that distribution is peaked and has thick tails.

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

**Ans:**Negative kurtosis value indicate that distribution is peaked and has thick tails.

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



What can we say about the distribution of the data?

**Ans:** The above distribution is not normally distributed the median tends towards

the higher value. Almost half of the data lies between the interval of 10-18.

What is nature of skewness of the data?

**Ans:** The data is left skewed since the whisker range of the upper quadrant is higher than that of the lower quadrant. Median will be greater than mean.

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

**Ans:** The IQR = The upper quartile – the lower quartile = 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:** Both the boxplot shares the same median which approximately lies between

the range of 250-275. They are both normally distributed. Also there are no

outliers.

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) = 0.3475939
  2. P(MPG<40) = 0.7293498

c. P (20<MPG<50) = 0.898868

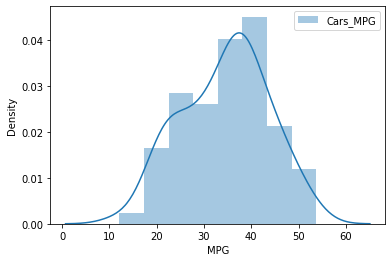
Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Mean of MPG = 34.4220757

Median of MPG = 35.1527269



Here, since the mean< median the distribution is negatively skewed.

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

Dataset: wc-at.csv

**Ans:** Mean

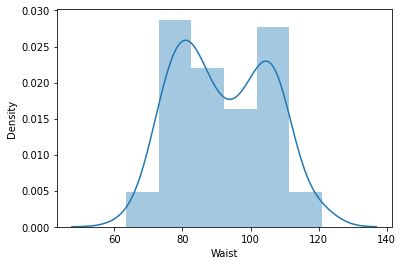
Waist = 91.901835

AT = 101.894037

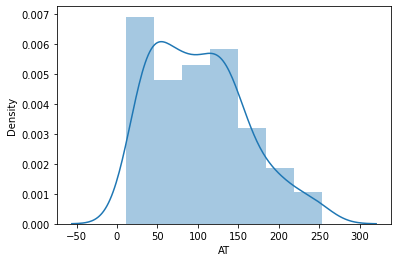
Median

Waist = 90.80

AT = 96.54



Here, since the mean>median the data is positively skewed.



Here, since the mean>median the data is positively skewed.

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

**Ans:** Confidence Interval for 90% = 1.64485

Confidence Interval for 94% = 1.88079

Confidence Interval for 60% = 0.84162

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

**Ans:** Confidence Interval for 95% = 2.0638

Confidence Interval for 96% = 2.1715

Confidence Interval for 99% = 2.7969

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:** Assume, H0= Average life of bulb >=260

H1=Average life of bulb < 260

Here, population mean=270 & sample mean=260

sample std= 90

df = 18-1= 17

Hence the probability that the selected bulbs would have an average life of no more than 260 days is 0.32.