|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete / Count |
| Results of rolling a dice | Discrete / Count |
| 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 | Categorical |
| Number of kids | Discrete / Count |
| Number of tickets in Indian railways | Discrete / Count |
| Number of times married | Discrete / Count |
| Gender (Male or Female) | Categorical |

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 | Ratio |
| Blood Group | Nominal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Interval |
| Years of Education | Ratio |

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

Ans). 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)0

b)3/36 or 0.083 2=1/36 3=2/36 4 3/36

c)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?

Ans) Total number of balls = (2 + 3 + 2) = 7  
Let S be the sample space.  
Then, n(S) = Number of ways of drawing 2 balls out of 7  
= **7C2**​  
=**(**7 x 6) **/** (2 x 1)​  
=21  
 Let E = Event of drawing 2 balls, none of which is blue.  
 n(E)= Number of ways of drawing 2 balls out of (2 + 3) balls **= 5C2**  
 = (5 x 4) **/** ​(2 x 1)​  
 =10  
 P(E)=n(E) / n(S) ​= 10 / 21​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) Expected number of candies for a randomly selected child

= 1 x 0.015 + 4 x 0.20 + 3 x 0.65 + 5 x 0.005 + 6 x 0.01 + 2 x 0.12

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

= 3.09.

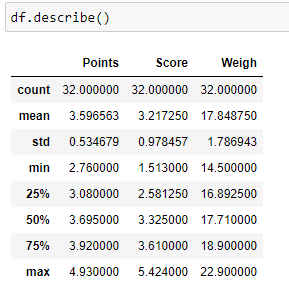
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

* 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)**

****

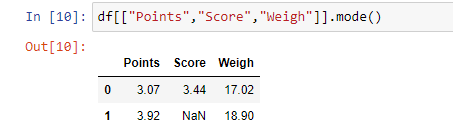
From the above fig the mean for points, score, Weigh are 3.59,3.21,17.84 respectively.

The median of the given data set is given below



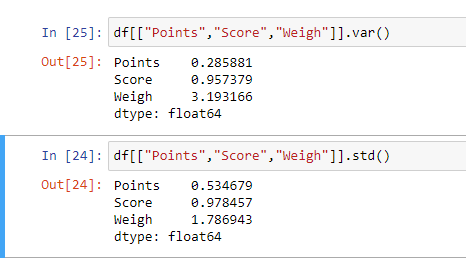
From the above fig you can see that the mean and median of the points, score, Weigh are similar, so we can say that the data follows normal distribution.

The mode of the given data set is given below



From the above fig the mode of Points, Score, Weigh are given as the data is continuous in nature. Mode shouldn’t be used as the measure of central tendency.

The variance and the standard deviation of the data is given below



The Range of the data = max(x)-min(x)

The Range of Points is 2.17

The Range of Score is 3.91

The Range of Weigh is 8.40

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?

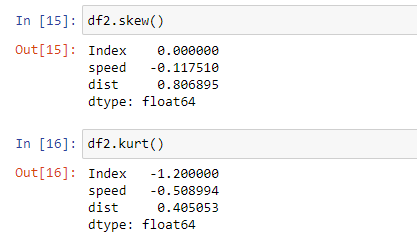
Ans) The mean of Weights is: 1308 / 9 = 145.33

If one of the patients is chosen at random the expected value of the patient is 145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

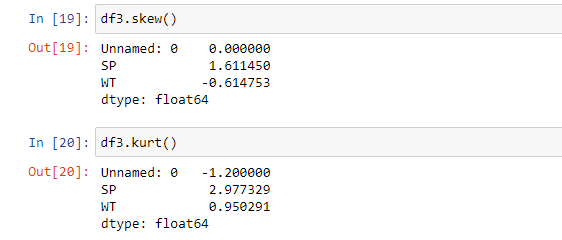
****

From the above fig the skewness of speed and distance is -0.117 and 0.8068 respectively. and the Kurtosis of speed and distance -0.508 and 0.405 respectively.

As the value of skewness and kurtosis is closer to zero. The distribution curve is normally distributed.

**SP and Weight(WT)**

**Use Q9\_b.csv**

****

From the above fig the skewness of SP and WT is 1.611 and -0.614 respectively. The skewness of SP is Positively skewed and skewness of WT is close to zero, so its normally distributed.

The kurtosis of SP is 2.977 and WT is 0.950. So, the curve is Normally distributed

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



From the above histogram, the is not symmetric in nature. The right tail is longer. So, the distribution of the curve is positive-skewed. There may have outliers on the upper fence.



From the above boxplot we can see that there are outliers present on the upper fence of the boxplot.By these values it will have great influence on the mean of the data. The data will be 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?

Ans) Sample size n=3000000

Sample Mean = x =200

Standard Deviation = s= 30

Standard error =

α=1-(confidence level / 100)

Confidence interval = x ± z /standard error

Confidence interval at 94% = α = 0.06 z(c)=1.56

Confidence interval at 94% = 200± 1.56 X 0.67

Confidence interval at 94% = 200± 1.04

Confidence interval at 98% = α = 0.02 z(c)=2.06

Confidence interval at 98% = 200± 2.06X 0.67

Confidence interval at 98% = 200± 1.38

Confidence interval at 96% = α = 0.04 z(c)=1.76

Confidence interval at 96% = 200± 1.76 X 0.67

Confidence interval at 96% = 200± 1.17

**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) 1. The mean of scores is 41

The median of the scores is 40.5

The variance of the scores is 25.529

The standard deviation of the scores is 5.052

2. From the above data, the value of skewness is 1.686 and the value of kurtosis is 3.953. From this we can say that the distribution of the data is positively skewed and its positively kurtosis

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

Ans) If the mean and median of the data are equal , then the nature of the skewness is normal distribution

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

Ans) The nature of skewness is Right Skewed or Positive Skewed

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

Ans) The nature of skewness is Left Skewed or Negative Skewed

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

Ans) When the data has positive kurtosis it indicates the spread of the data is narrow.

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

Ans) The data has positive kurtosis it indicates the spread of the data is broad.

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



What can we say about the distribution of the data?

Ans) From the above of fig, In the distribution of the data the tails are not symmetric in nature. It has a longer tail on the left. There may be outliers present on the lower fence.

What is nature of skewness of the data?

Ans) The nature of Skewness of the data is negatively Skewed or left skewed

What will be the IQR of the data (approximately)?   
Ans) IQR= Q3-Q1

Q1= 10

Q3=18

IQR= 18-10 = 8

The IQR of the data is 8 approximately.

Q19) Comment on the below Boxplot visualizations?



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

Ans) By comparing both boxplots, the Boxplot 1 is smaller than the boxplot 2.

From this we can say that the kurtosis of boxplot 1 is higher than the boxplot 2.

The spread of the data in boxplot2 is greater than boxplot1

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)

1. 0.347
2. 0.729
3. 0.898

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans) The skewness value of MPG is -0.177 and kurtosis value is -0.611

As the value of skewness is closer to zero it follows normal distribution.

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

Dataset: wc-at.csv

Ans) The Skewness value of Adipose Tissue (AT) and Waist Circumference (Waist) are 0.584 and 0.134 respectively. As the skewness of both AT and Waist are closer to zero, It follows normal distribution.

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

Ans) Z score at 90% confidence interval = 1.281

Z score at 94% confidence interval = 1.554

Z score at 60% confidence interval = 0.253

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

Ans) T score of 95% confidence interval at degree of freedom = 24 = 1.710

T score of 96% confidence interval at degree of freedom = 24 = 1.828

T score of 99% confidence interval at degree of freedom = 24 = 2.492

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) the probability that 18 randomly selected bulbs would have an average life of no more than 260 days is 51.02%