|  |  |
| --- | --- |
| 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 | Ordinal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ratio |
| Sales Figures | Interval |
| Blood Group | Nominal |
| Time Of Day | Ratio |
| Time on a Clock with Hands | Ratio |
| Number of Children | Ordinal |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Ratio |
| Years of Education | Interval |

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

3/8

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

1. Equal to 1 ->0
2. Less than or equal to 4-> 1/6
3. Sum is divisible by 2 and 3 -> 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?

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 ->3.09

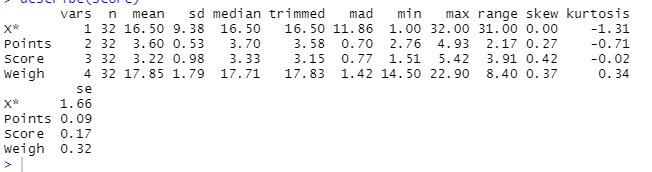
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

* For Points,Score,Weigh>

Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

Answer:-





**Use Q7.csv file**

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

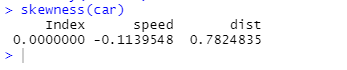
Assume one of the patients is chosen at random. What is the Expected Value of the Weight of that patient?

Ans ->145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**



**Answer:- As we can see that the skewness is negative for the speed that means speed in negatively skewed. Negative skewness implies that the mass of the distribution is concentrated on the right.**

**Distance is positively skewed**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**SP WT**

**1.5814537 -0.6033099**

**Inference: -a) most of the car speed is less than the average speed that is 121.54**

**b) for weight its negative skewness which means most of the car’s weight is more than the average weight of the cars that is 32.41**

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



The frequency of $100 is the most.



There are outliers because of which the mean will get affected. That’s why we should find the median. The upper quartile has more numbers.

**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?->Confidence interval 94%-> 200.028 & 199.971

98%->200.034 & 199.965

96%-> 200.030 & 199.969

**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.

Mean – 41

Median-40.5

Variance-25.52

Sd-5.05

1. What can we say about the student marks?

Since the mean and median value are almost the same which means that there are no outliers. The avg marks of all the students in 41. There is no big difference in the marks of the students.

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

**Skewness is 0 that is data on left side of the curve is equal to the right side**

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

**The distribution is positively skewed**

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

**The distribution is negatively skewed**

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

**The distribution is peaked and possess thick tails**

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

**Distribution has lighter tails than the normal distributions**

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



What can we say about the distribution of the data?

**Median- 15.4**

**1st quartile-10.2**

What is nature of skewness of the data?

**Negative skewness**

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

Q19) Comment on the below Boxplot visualizations?



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

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

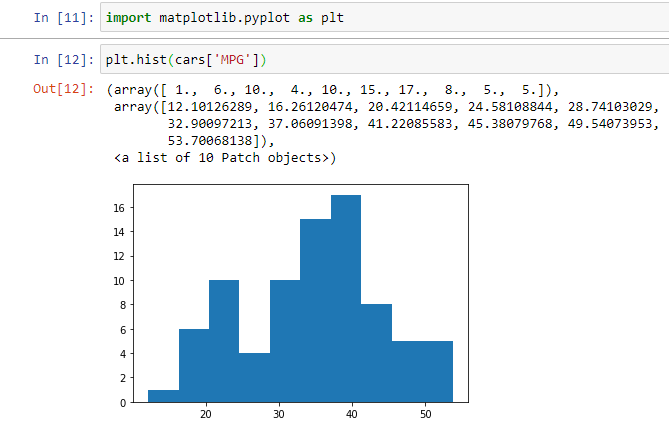
* 1. P(MPG>38)
  2. P(MPG<40)

c. P (20<MPG<50)

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

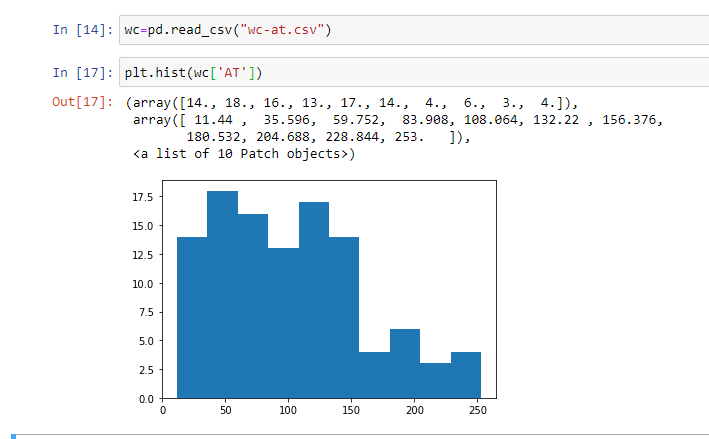
Dataset: Cars.csv



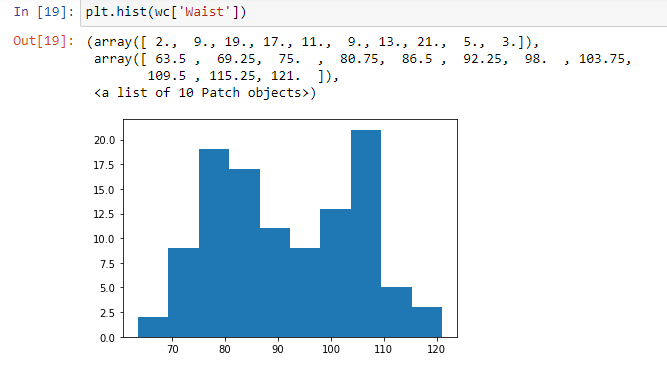
The above histogram doesn’t look symmetric therefore we can say the MPG data in dataset car is not normally distributed

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

Dataset: wc-at.csv

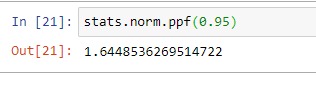


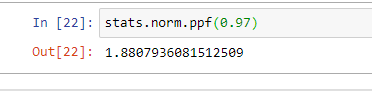
**AT is not normally distributed**

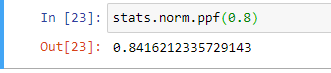


**Waist is normally distributed**

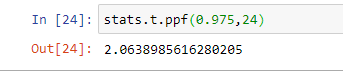
Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval

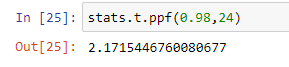
Z score for 90% -> 

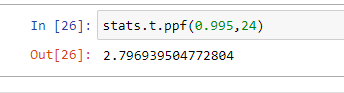
Z score for 94%-> 

Z score for 60%-> 

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

95% -> 

96%-> 

99% -> 

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

**0.07**