|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Qualitative Data>Countable |
| Results of rolling a dice | Qualitative Data >Countable |
| Weight of a person | Continuous Data>Interval |
| Weight of Gold | Continuous Data >Interval |
| Distance between two places | Continuous Data |
| Length of a leaf | Continuous Data >Interval |
| Dog's weight | Continuous Data >Interval |
| Blue Color | Nominal Data |
| Number of kids | Qualitative Data >Countable |
| Number of tickets in Indian railways | Qualitative Data >Countable |
| Number of times married | Qualitative Data >Countable |
| Gender (Male or Female) | Nominal Data |

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 data |
| High School Class Ranking | Ordinal |
| Celsius Temperature | Interval |
| Weight | Interval |
| Hair Color | Nominal |
| Socioeconomic Status | Nominal |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation | Nominal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ordinal |
| Sales Figures | Interval |
| Blood Group | Nominal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Interval |
| 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?

Two heads and one tail obtained in three coins, possible outcomes are:

[HHT,THH,HTH] = 3/8, total outcomes are 8, out chances are 3, assuming it’s a fair coin.

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

1. Equal to 1 = Probability is ZERO
2. Less than or equal to 4 = 6/36 or 1/6
3. Sum is divisible by 2 and 3 = 24/36 or 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 probability of blue balls is 2/7

So probability of balls not being blue is, total probability – blue probability

= 1-(2/7)

=5/7

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 candies for a selected child is

1 candy multiplied by its probability = 1\*0.015 and so on add all probabilities, which will give you total number of candies.

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

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

=  **3.090**

The Expected number of candies for a randomly selected child is **3.090**

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.

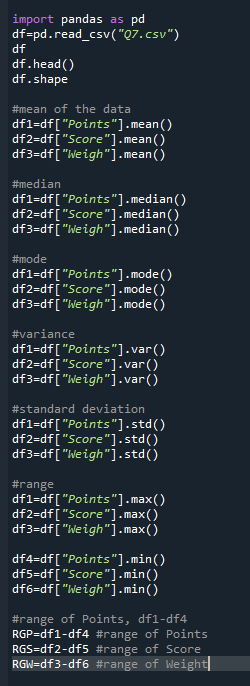
**Use Q7.csv file**

**Answer:**

|  |  |  |  |
| --- | --- | --- | --- |
| Mean | 3.5965625 | 3.21725 | 17.84875 |
| Median | 3.695 | 3.325 | 17.71 |
| Mode | 3.92 | 3.44 | 17.02 |
| Variance | 0.285881351 | 0.957378968 | 3.193166129 |
| Standard Deviation | 0.534678736 | 0.978457443 | 1.786943236 |
| Maximum Value | 4.93 | 5.424 | 22.9 |
| Minimum Value | 2.76 | 1.513 | 14.5 |
| Range(max-min) | 2.17 | 3.911 | 8.4 |

As per the values drawn we can see that Mean<Median<Mode so we can draw a conclusion that the data is negatively skewed and is not in normal distribution.

SD shows the numbers are far apart from each other and variance shows deviation from the mean.



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?

Answer :

Avg weight of the patients calculated is 145.3 Pounds. So, the expected value of the weight of the patient selected is 145.3 pounds. Each patient has a 1/9 chance of getting picked.

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

**Cars speed and distance**

**Use Q9\_a.csv**

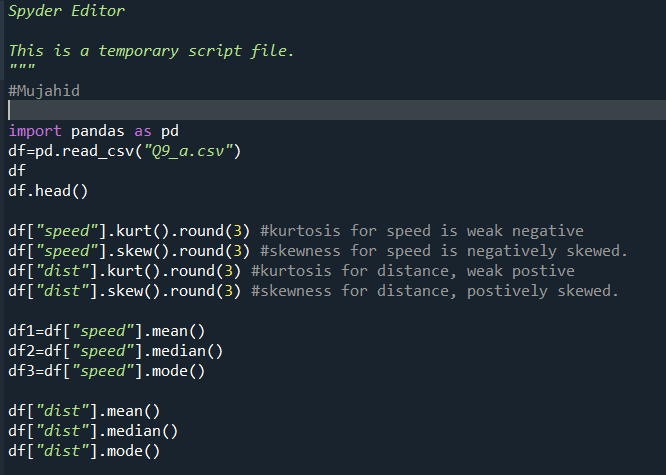
**Answer :**

df["speed"].kurt().round(3) = -0.509 – Speed kurtosis

df["speed"].skew().round(3) = -0.118 – speed skewness

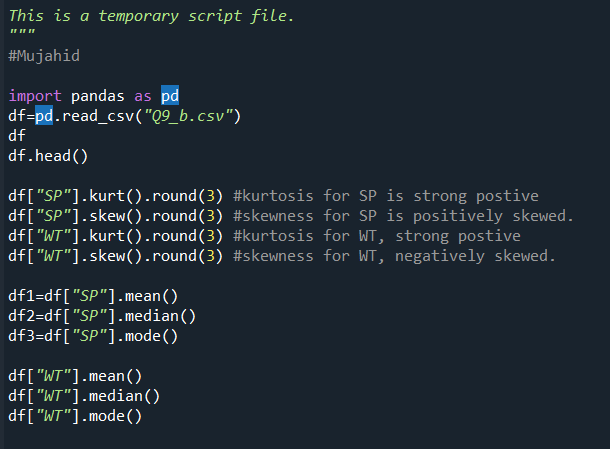
df["dist"].kurt().round(3) = 0.405 - Distance kurtosis

df["dist"].skew().round(3) = 0.807 - distance skewness



**SP and Weight(WT)**

**Use Q9\_b.csv**



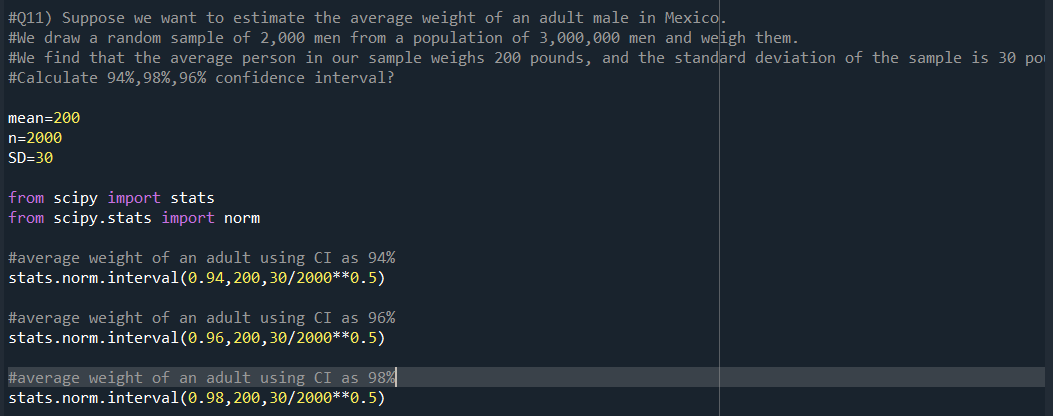
**Q10) Draw inferences about the following boxplot & histogram**



**Histogram** – We can see that majority(>50%) of the data is present under 50-150 range, and highest ChickWeight is available between 50-100 range. The data is positively skewed where mode<median<mean

**BoxPlot :** Box plot has a lot of outliers on the right side and is positively skewed.

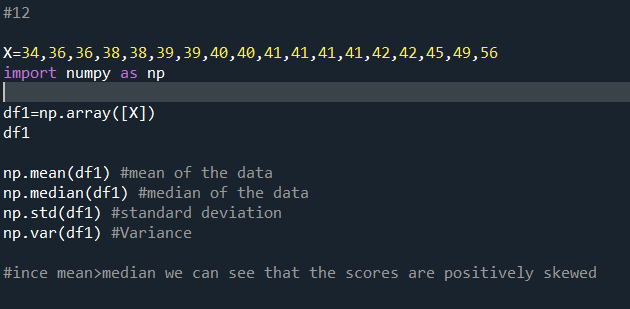
**Q11)** Suppose we want to estimate the average weight of an adult male in Mexico. We draw a random sample of 2,000 men from a population of 3,000,000 men and weigh them. We find that the average person in our sample weighs 200 pounds, and the standard deviation of the sample is 30 pounds. Calculate 94%,98%,96% confidence interval?



**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?



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

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

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

Q16) What does positive kurtosis value indicates for a data ? Positive values of kurtosis indicate that distribution is peaked and possesses thick tails. Leptokurtic distributions have positive kurtosis values. A Leptokurtic distribution where more of the numbers are located in the tails of the distribution instead of around the mean.

Q17) What does negative kurtosis value indicates for a data? Negative values of kurtosis indicate that a distribution is flat and has thin tails. Platykurtic distributions have negative kurtosis values. A platykurtic distribution is flatter (less peaked) when compared with the normal distribution, with fewer values in its shorter (i.e. lighter and thinner) tails.

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



What can we say about the distribution of the data? The above Boxplot is not normally distributed the median is towards the higher value.

What is nature of skewness of the data? The data is a skewed towards left. The whisker range of minimum value is greater than maximum.

Upper whisker length = Q3+(1.5\*IQR) = 18+(1.5\*8) = 30

Lower whisker length = Q1-(1.5\*IQR) = 10-(1.5\*8) = -2

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

Q3=18

Q1=10

IQR = 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.

There are no outliers in the boxplot. Both of the box plots share the same median that is approximately in a range between 275 to 250 and they are normally distributed with zero to no skewness neither at the minimum or maximum whisker range.

Q 20) Calculate probability from the given dataset for the below cases

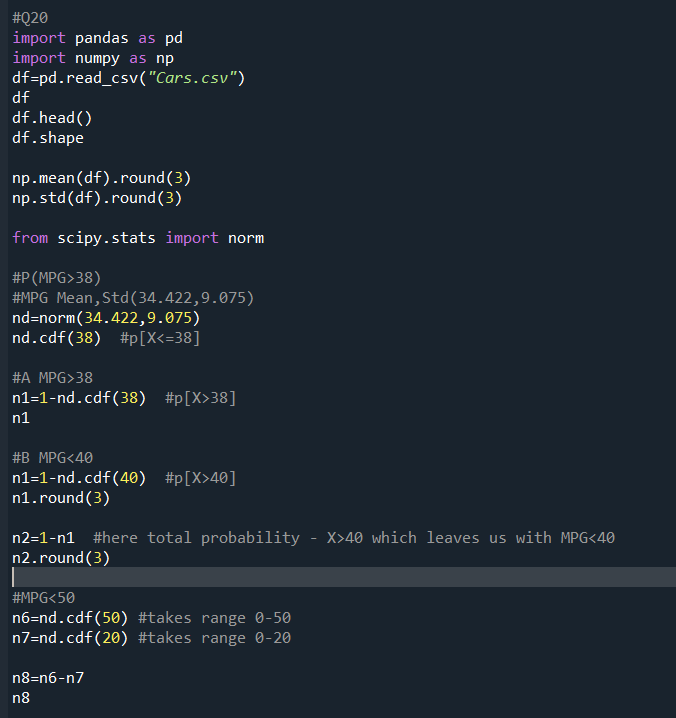
Data \_set: Cars.csv

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

MPG of cars follows normal distribution. Confirming using QQ Plot



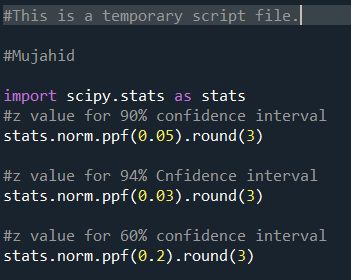
1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

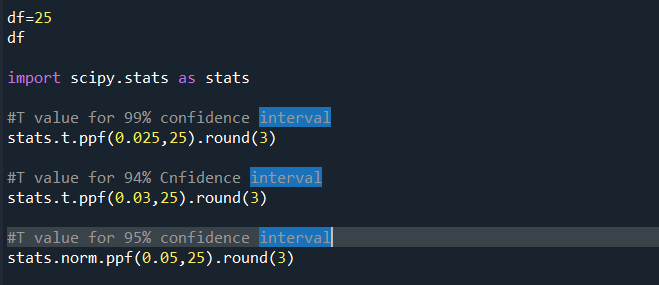
Adipose Tissue (AT) and Waist Circumference(Waist) follows normal distribution. Confirming using QQ Plot



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



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



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

