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

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

A3) When three coins are tossed the total number of possible combinations are 2\*2\*2= 8.

These combinations are HHH, HHT, HTH, THH, TTH, THT, HTT, TTT.

The number of combinations which have two heads and one tail are:

HHT, HTH, TTH

Therefore, the Probability of getting two heads and one tails in the toss of three coins simultaneously is

P(Two Heads and One Tail) = Number of desired outcomes/Total no. of Outcomes

= 3/8 = 0.375

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

A4) When two dice are rolled the total number of possible combinations are 6\*6=36

These combinations are (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)

a) the sum is equal to 1 is zero because they starts with (1,1) .... likewise . other than in the dice we are not having zero.

b) the sum is less than or equal to 4 the possible outcomes are (1,1),(1,2),(1,3),(2,1),(2,2),(3,1) therefore n( b) = 6/36 = 1/36

c) the sum is divisible by 2 and 3 the possible outcomes are (1,5), (2,4), (3,3), (4,2), (5,1), (6,6) therefore n(c)= 6/36 = 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?

A5) Total number of balls = ( 2(red) + 3(green) + 2 (blue) ) = 7.

Probability of two balls are drawn at random = 2/7.

Probability of two balls are drawn at random when none of

the balls drawn is blue = 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

A6) 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 \* 0.12

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

= 3.090

**Expected number of candies for a randomly selected child = 3.09**

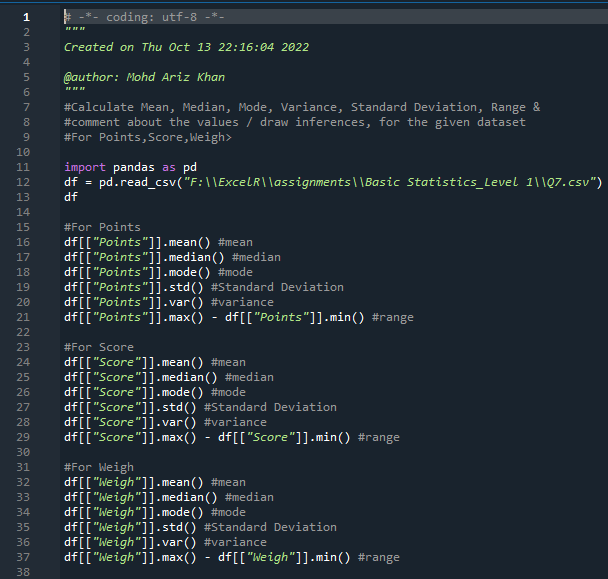
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

* 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**

A7)



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?

A8) Expected Value = ∑ (probability \* Value)

 ∑ P(x).E(x)

There are 9 patients

Probability of selecting each patient = 1/9

Expected Value = (1/9) (108) + (1/9) 110 + (1/9)123 + (1/9)134 + (1/9)135 + (1/9)145 + (1/9(167) + (1/9)187 + (1/9)199

= (1/9) (108 + 110 + 123 + 134 + 135 + 145 + 167 + 187 + 199)

= (1/9) (1308)

= 145.33

Expected Value of the Weight of that patient = 145.33

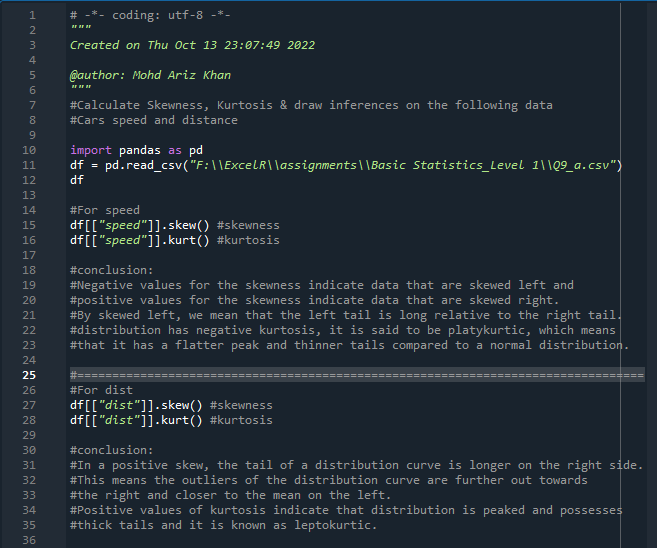
**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

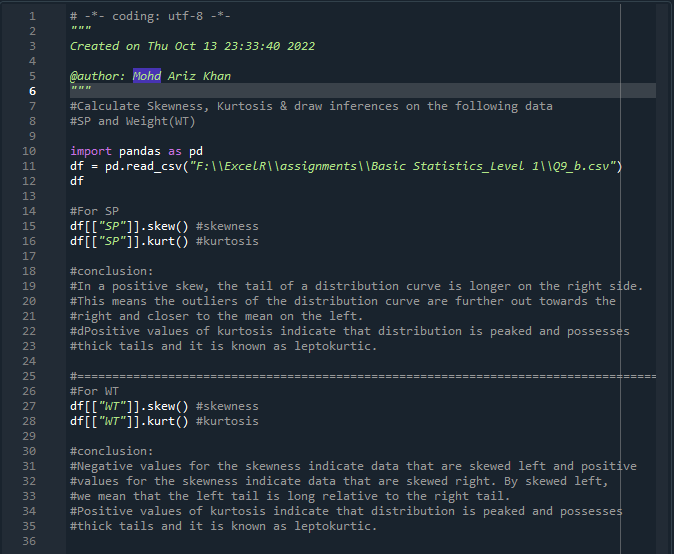
**Cars speed and distance**

**Use Q9\_a.csv**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**A9) a)** ****

**b)** ****

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



A10) Histogram :-

- Chick weight data is right skewed or positively skewed.

- More than 50% Chick Weight is between 50 to 150.

- Most of the chick weight is between 50 to 100.

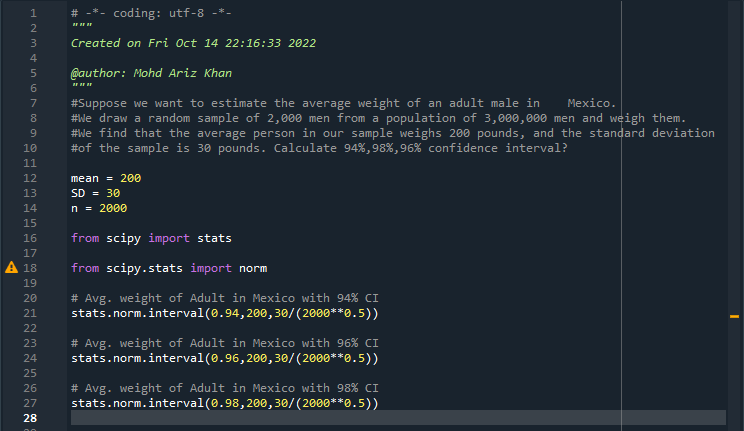
Box Plot :-

-The data is right skewed.

-There are outliers at upper side.

**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?

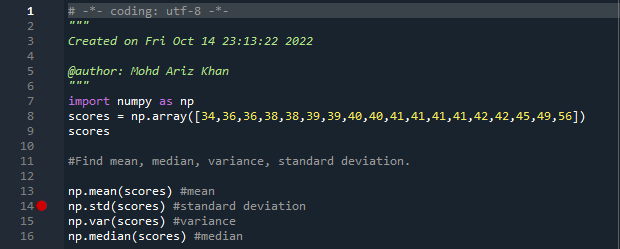
**A11)**



**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?

**A12)** 

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

A13) The mean and median of the data are equal that means the distribution is symmetric. so, the distribution has **zero skewness.**

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

A14) If the mean is greater than the median, the distribution is positively skewed.

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

A15) If the mean is less than the median, the distribution is negatively skewed.

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

A16) 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?

A17) 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?

What is nature of skewness of the data?

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

A18)

The above Boxplot is not normally distributed the median is towards the higher value.

The data is a skewed towards left. The whisker range of minimum value is greater than maximum.

The Inter Quantile Range (IQR) = Q3 Upper quartile – Q1 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.

A19) First there are no outliers. Second both the box plot shares 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

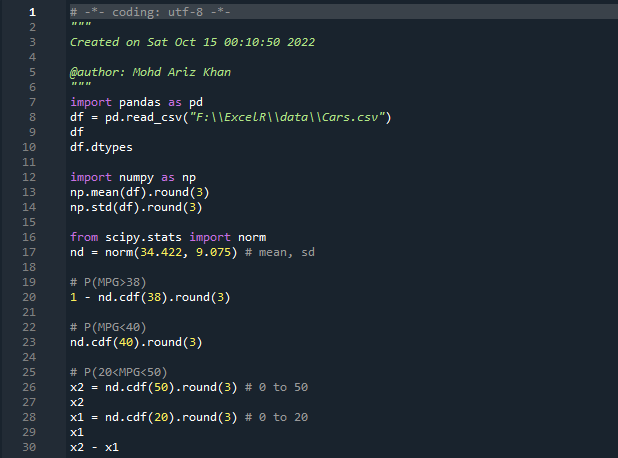
Data \_set: Cars.csv

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)

A 20)

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

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

Dataset: wc-at.csv

A 21)

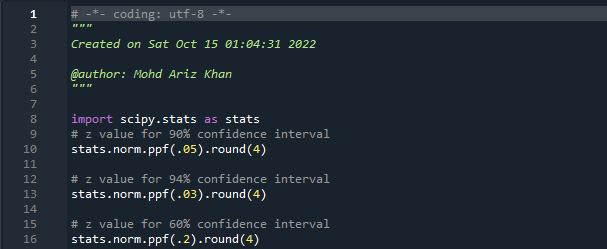
a) MPG of cars follows normal distribution.

b) Adipose Tissue (AT) and Waist Circumference (Waist) does not follow

Normal Distribution.

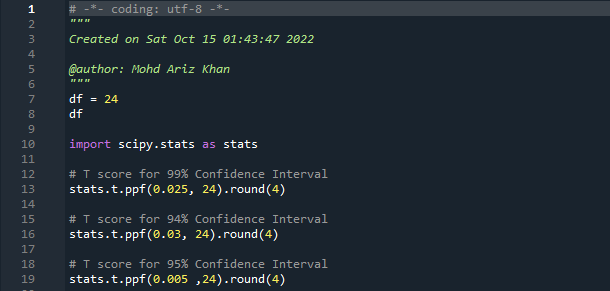
Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval

A22)



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

A 23)



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

A 24)

