|  |  |
| --- | --- |
| 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 | ordinal |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation | Nominal |
| Level of Agreement | Nominal |
| IQ(Intelligence Scale) | Ordinal |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| 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= When three coins are tossed, the possible outcomes are TTT, THT, TTH, THH, HTT, HHT, HTH, and HHH therefore we get probability as 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 = If we throw 2 Dice we get the possible outcome as

**{**(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 = 0/36

b = 1/6

c = 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?

Ans=Probability of not drawing any blue balls from a bag: 5/7

Q6) Calculate the Expected number of candies for a randomly selected child

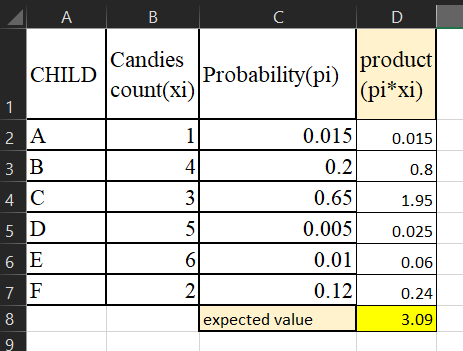
Below are the probabilities of count of candies for children (ignoring the nature of the child-Generalized view)

|  |  |  |
| --- | --- | --- |
| 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=



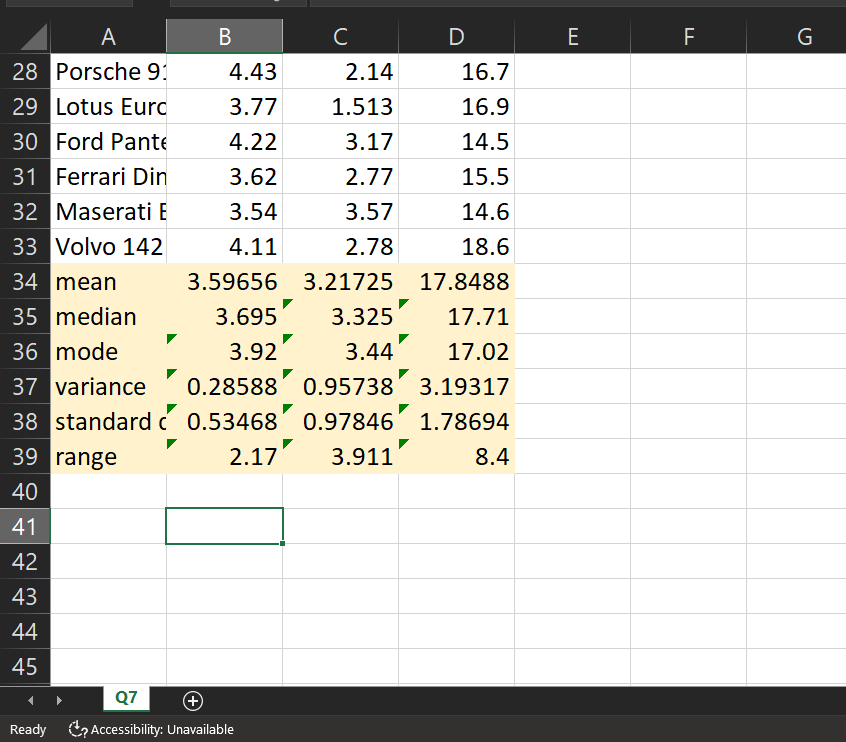
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

* For Points,Score,Weigh>

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

**Using q7 csv file ans is =**

* For Points: The data is relatively close to the mean, with a small variance and standard deviation, indicating consistency.
* For Score: The mode being 3.44 suggests a common value, while the range shows variability in the data.
* For Weigh: The data is more spread out compared to Points and Score, with a larger range and standard deviation, indicating greater variability.

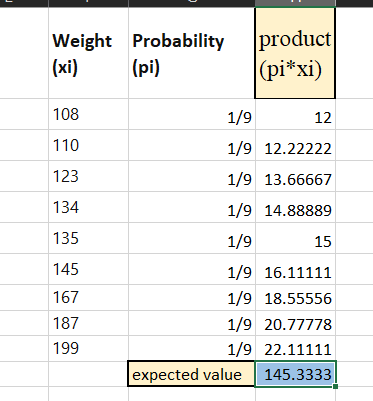
****

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?



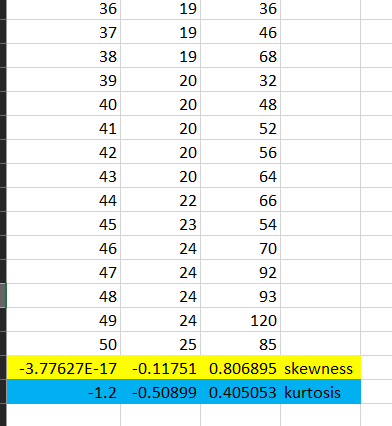
Ans= Expected value is 145.33 pounds

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

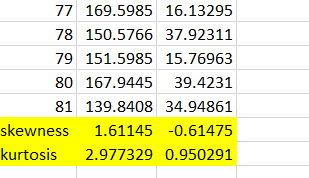
**Cars speed and distance**

**Use Q9\_a.csv**

* A positive skewness indicates a right-skewed distribution, where the tail is on the right side.
* A negative skewness indicates a left-skewed distribution, where the tail is on the left side.

****

**SP and Weight(WT) Use Q9\_b.csv**

****

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

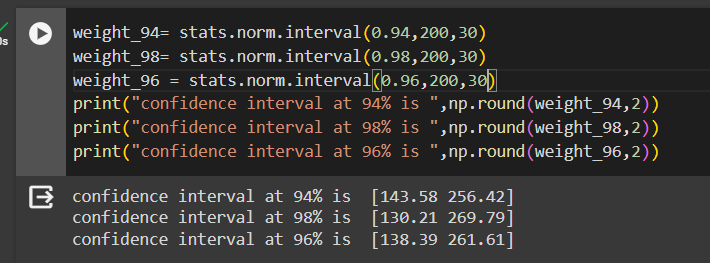


* the histogram is right skewed
* with increase of chickweight the frequency decreases



* there are many outliers in the box plot
* ie. the presence of outliers suggests that there are extreme values in the dataset.
* the box plot is right skewed ie.its positive

**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?

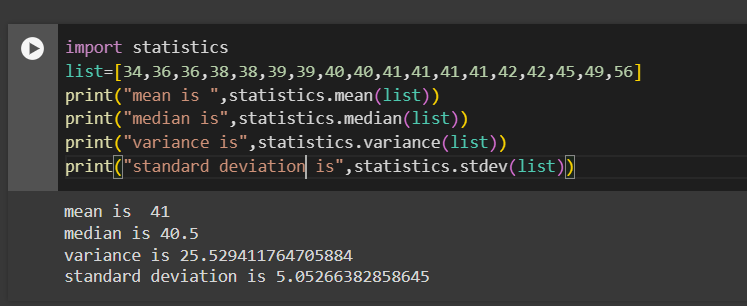
Ans=

**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=the data is relatively spread out



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

Ans= the nature of the curve is symmetric

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

Ans=right skewed

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

Ans = left skewed

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

Ans = it indicated more outliners and peaked distribution like bell curved

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

Ans= it indicates less outliners flatter graph as compared to normal distribution

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



What can we say about the distribution of the data?

Ans = the distribution is left skewed ie. there are outliners present the median is between 15-16.

What is nature of skewness of the data?

Ans = left skewed

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

Ans =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 box plot have same median but the size ie.the values the box plot 1 takes is less than box plot 2 , they have different inter quartile range.

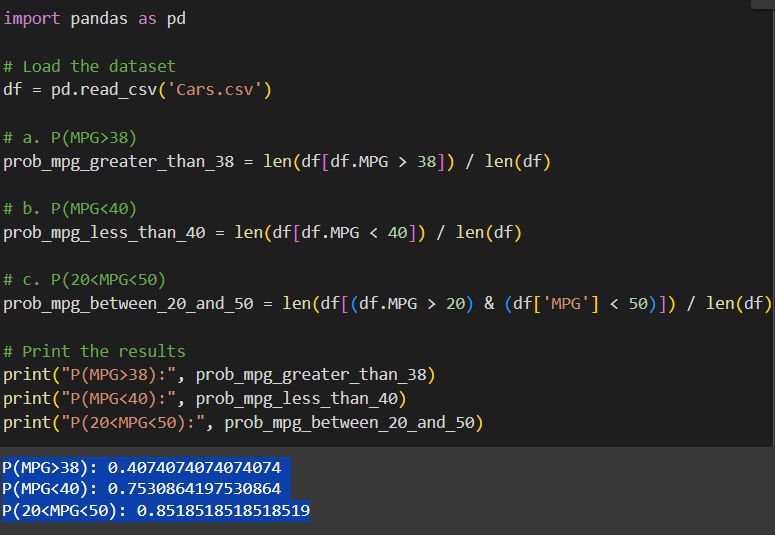
Q 20) Calculate probability from the given dataset for the below cases

Data \_set: Cars.csv

Calculate the probability of MPG of Cars for the below cases.

MPG <- Cars$MPG

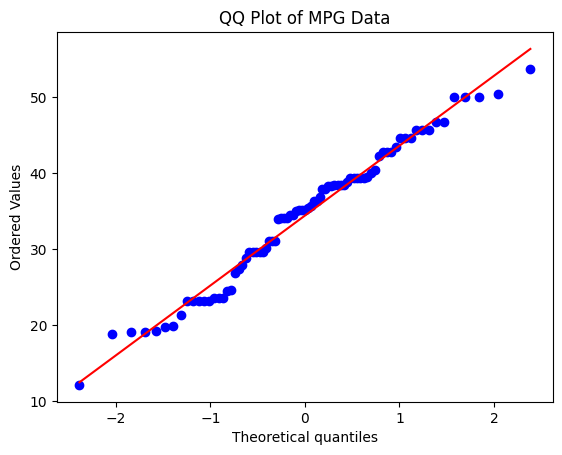
* 1. P(MPG>38)
  2. P(MPG<40)
  3. P (20<MPG<50)

Ans= 

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

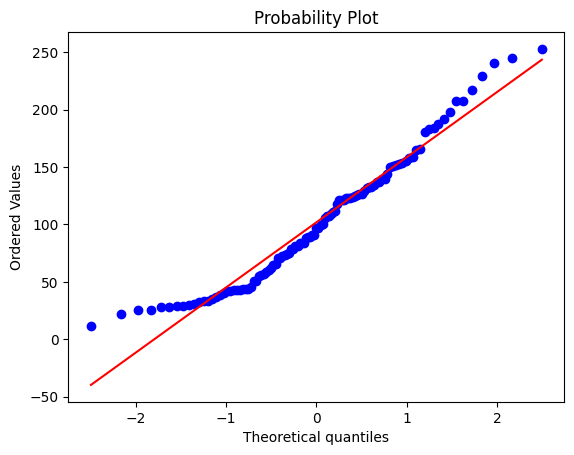
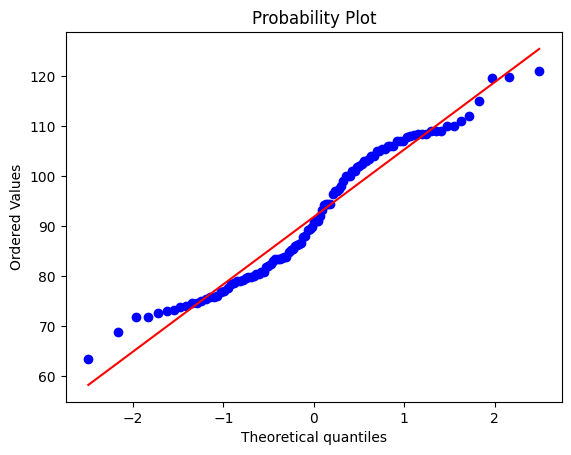
Dataset: Cars.csv



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

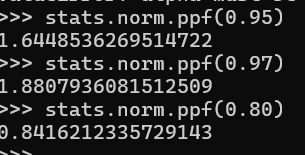
Dataset: wc-at.csv

Ans=

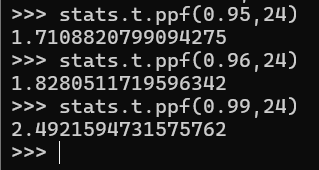
Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval

Ans= by using function in python stats.norm.ppf() and considering data normally distributed



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

Ans=



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

Ans= 32.1% the t value calculated is -0.471

Therefore, based on the sample data, there is a 32.1% chance that the average life of the bulbs would be less than or equal to 260 days, assuming the CEO’s claim is accurate.

