| 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) | Binary |

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

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

probability of one head=½, Probability of two heads=½+½

Probability of one tail=½

=½+½+½=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 2and 3
4. The sum is equal to 1 is zero
5. Less than or equal to 4

(1,3),(2,2),(3,1) therefore n( b) = 3/36 = 1/12

1. When 2 dice are rolled find the probability of getting a sum divisible by 3 - 12.
2. When 2 dice are rolled find the probability of getting a sum divisible by 2 - 0.5

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?

Probability of two balls drawn are blue P(b)=⅖

none of the balls drawn are blue=1-p(b)=1-⅖=0.6

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

= 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

= 3.09

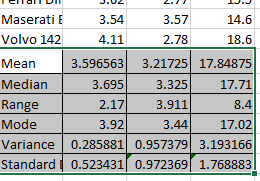
Expected number of candies for a randomly selected child = 3.09

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**

****

**Comments:**

* **There were outliers in the Weights as there variance and Standard deviation**

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?

there are 9 patients

Probability of selecting each patient = 1/9

Ex 108, 110, 123, 134, 135, 145, 167, 187, 199

P(x) 1/9 1/9 1/9 1/9 1/9 1/9 1/9 1/9 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

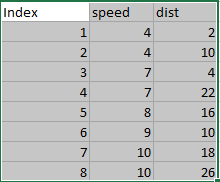
Expected Value of the Weight of that patient = 145.33

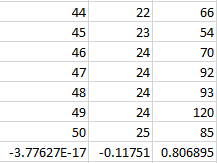
**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

**Cars speed and distance**

**Use Q9\_a.csv**

**Answer: Here i mentioned the sample screen shots and i have given skewness**

****

****

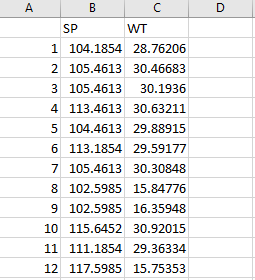
* **Inferences is that distance has positive skewness and Speed and index has negative skewness**

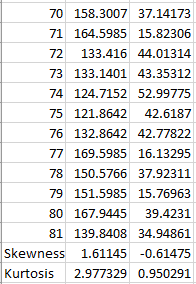
**Kurtosis**

* **Index -1.200000**
* **speed -0.508994**
* **dist 0.405053**
* **Speed distribution is negative kurtosis i.e. flatter than normal distribution)**
* **Distance distribution is positive kurtosis i.e. peaked than normal distribution)**

**SP and Weight(WT)**

**Use Q9\_b.csv**

****

****

* **Inferences is that Speed has positive skewness and Distance has negative skewness**

**Comments:**

* **Weight distribution is Positive kurtosis i.e. peaked than normal distribution)**
* **Speed distribution is positive kurtosis i.e. peaked than normal distribution)**

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



Inferences:

* when the chickWeight$weight is in the range of 50-100 then the frequency is in peak 200
* when ChickWeight$weight is in the range of 350-400 then the frequency is in zero 0
* when ChickWeight$weight is in the range of 0-50 then the frequency is in 80
* when ChickWeight$weight is in the range of 100-150 then the frequency is in zero 130
* The curve looks like it has positive skewness or right side skewed



Inferences:

* There are outliers after the upper extreme or upper whisker

**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?

*# Avg. weight of Adult in Mexico with 94% CI*

stats**.**norm**.**interval(0.94,200,30**/**(2000**\*\***0.5))

(198.738325292158, 201.261674707842)

*# Avg. weight of Adult in Mexico with 98% CI*

stats**.**norm**.**interval(0.98,200,30**/**(2000**\*\***0.5))

(198.43943840429978, 201.56056159570022)

*# Avg. weight of Adult in Mexico with 96% CI*

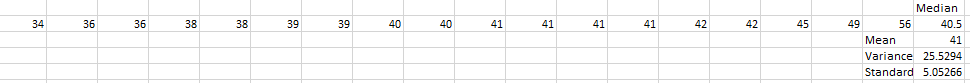
stats**.**norm**.**interval(0.96,200,30**/**(2000**\*\***0.5))

(198.62230334813333, 201.37769665186667)

**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?



Mean:41

Median:40.3

Variance: 25.5204

Standard Deviation:5.05266

2. The average student marks are 41

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

If the distribution is symmetric, then the mean is equal to the median, and the distribution has zero skewness

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

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

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

If the median is greater than the mean, the distribution is negatively skewed.

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

Positive excess values of kurtosis (>3) indicate that a distribution is peaked and possess thick tails.A leptokurtic distribution has a higher peak (thin bell) and taller (i.e. fatter and heavy) tails than a normal distribution.

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

Negative excess values of kurtosis (<3) indicate that a distribution is flat and has thin tails. 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?

Lower Quartile was 10

Median was 15

Upper Quartile was 18

What is nature of skewness of the data?

It is negative skewness

What will be the IQR of the data (approximately)?   
!QR:Q3-Q1: 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.

Inferences:

Boxplot 1 and Boxplot 2 has a median of 262.5

Lower Extreme for Boxplot 1 is 250 and Boxplot 2 is 200

upper extreme for Boxplot 1 is 300 and Boxplot 2 is 350

upper Quartile for Boxplot 1 is 275 and Boxplot 2 is 312.5

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

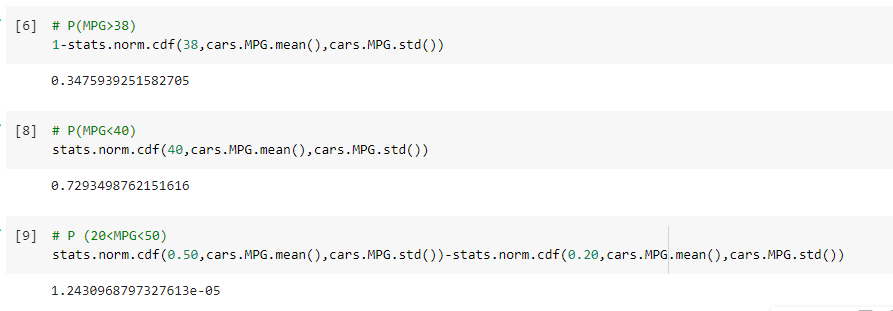
Data \_set: Cars.csv

Calculate the probability of MPG ofCars for the below cases.

MPG<- Cars$MPG

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

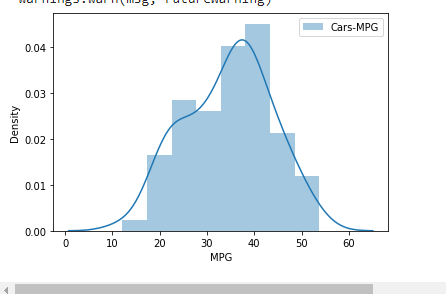
c. P (20<MPG<50)



Q 21) Check whether the data follows normal distribution

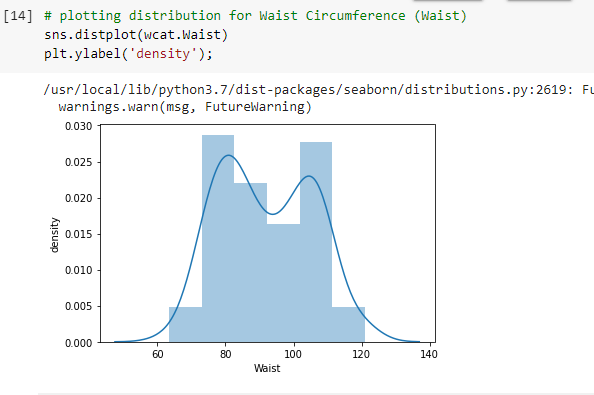
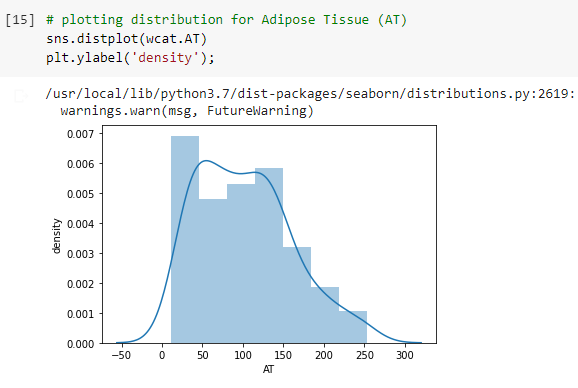
1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

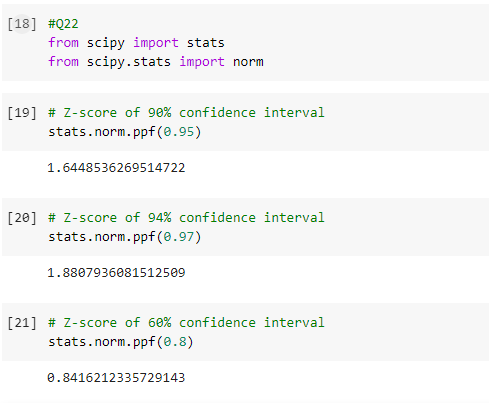


1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

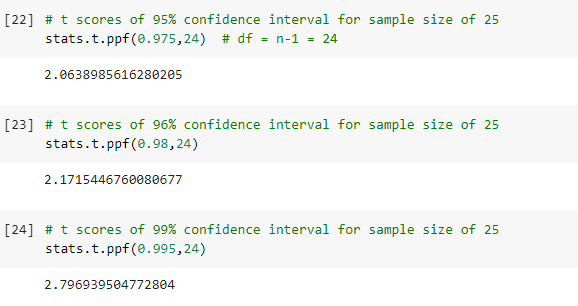
Dataset: wc-at.csv



Q 22) Calculate the Z scoresof 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)

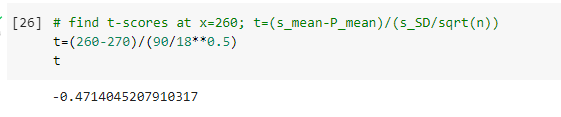
df 🡪 degrees of freedom

x = mean of the sample of bulbs = 260

μ = population mean = 270

s = standard deviation of the sample = 90

n = number of items in the sample = 18



For probability calculations, the number of degrees of freedom is n - 1.

The probability that t < - 0.471 with 17 degrees of freedom assuming the population mean is true, the t-value is less than the t-value obtained With 17 degrees of freedom and a t score of - 0.471, the probability of the bulbs lasting less than 260 days on average of 0.3216 assuming the mean life of the bulbs is 300 days.

