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

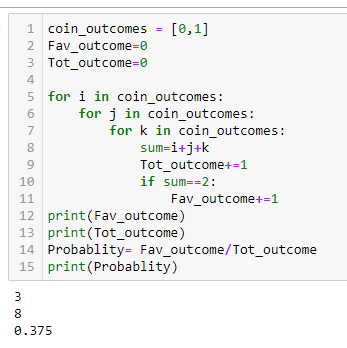
Q3) Three Coins are tossed, find the probability that two heads and one tail are obtained?

Ans:

H: Head, T:Tail

P(2H & 1T) = N ( Event (2H & 1T))/ N ( Event (Three coin tossed))

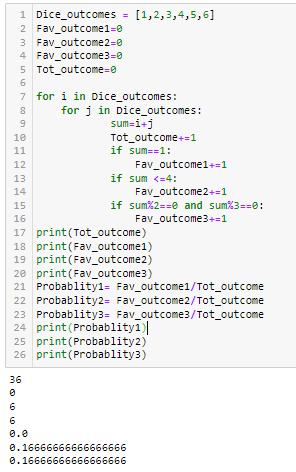
=3/8= 0.375



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

Ans:



Number of possible outcomes for the above event is

N (Event (Two dice rolled)) = 36

1. P (sum is Equal to 1) = 0.
2. P (Sum is less than or equal to 4) = N (Event (Sum is less than or equal to 4))/ N (Event (Two dice rolled))=6/36=0.1666=16.66%
3. P (Sum is divisible by 2 and 3)= N (Event (Sum is divisible by 2 and 3))/ N (Event (Two dice rolled))= 6/36=0.1666=16.66%

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?

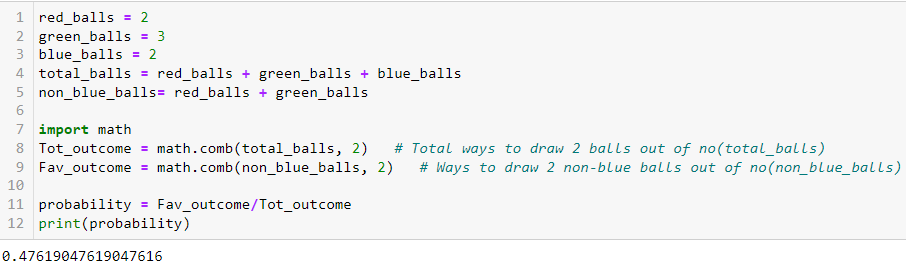
Ans:

Total number of balls =7 balls

Total ways to draw 2 balls out of 7 = 7C2 = (7! / (2!(7 - 2)!)) = 21 ways

Ways to draw 2 non-blue balls out of 5 (2 red and 3 green) = 5C2 = (5! / (2!(5 - 2)!)) = 10 ways

Probability (none of the balls drawn is blue) = (Number of favorable outcomes) / (Total number of possible outcomes) = 10/21=0.4761=47.61%



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 number of candies for a randomly selected child =

Σ (Candies count x Probability)

= 0.015+0.2 x 4+3 x 0.65+0.005 x 5+0.01 x 6+0.120 x 2

= 0.015+0.8+1.95+0.025+0.06+0.24 = 3.09

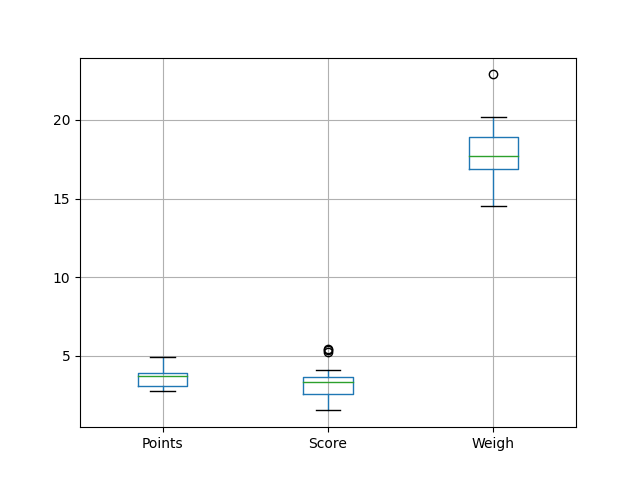
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

* 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**

Ans:

Draw Inferences

Mean for Points = 3.596, Score = 3.217 and Weigh = 17.848

Median for Points = 3.695, Score = 3.325 and Weigh = 17.710

Mode for Points = 3.07, Score = 3.44 and Weigh = 17.02

Variance for Points = 0.285, Score = 0.957, Weigh = 3.193

Standard Deviation for Points = 0.534, Score = 0.978, Weigh = 1.786

Range [Max- Min]:

Points [4.93– 2.76] =2.17

Score [5.424– 1.513] =3.911

Weigh [22.9 – 14.5] = 8.40

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 = Sum (X \* Probability of X)

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

= 145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans:**

Skewness:

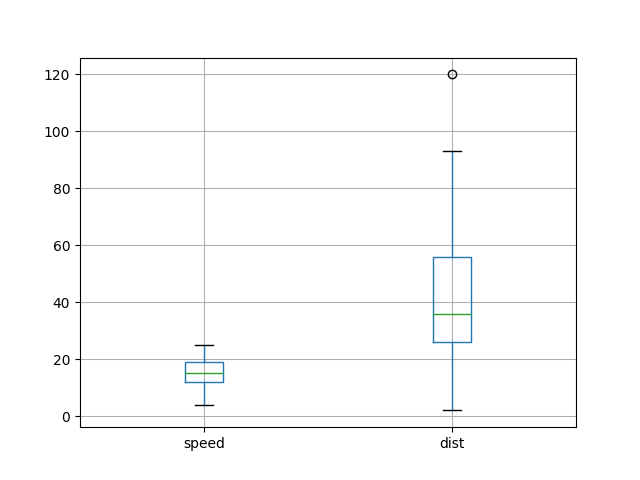
speed -0.117510

dist 0.806895

Kurtosis:

speed -0.508994

dist 0.405053



**SP and Weight (WT)**

**Use Q9\_b.csv**

Ans:

Skewness:

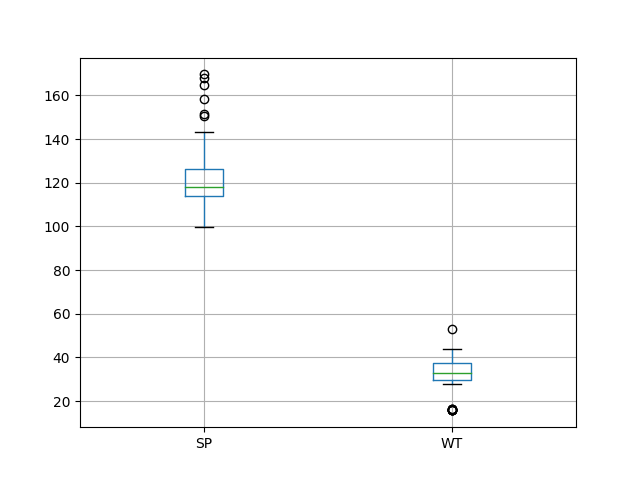
SP 1.611450

WT -0.614753

Kurtosis:

SP 2.977329

WT 0.950291



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



Ans: The histograms peak has right skew and tail is on right. Mean > Median. We have outliers on the higher side.



Ans: The boxplot has outliers on the maximum 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?

Ans:



94.0% Confidence Interval: 198.738 to 201.262 and Zscore 1.881

98.0% Confidence Interval: 198.439 to 201.561 and Zscore 2.326

96.0% Confidence Interval: 198.622 to 201.378 and Zscore 2.054

**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.

Ans:

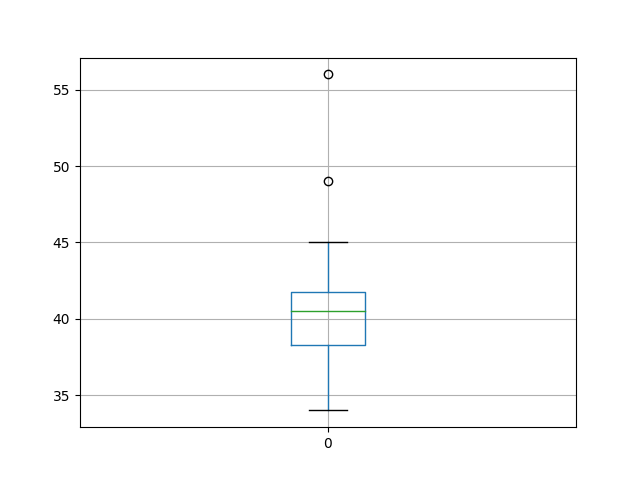
Mean: 41.0

Median: 40.5

Variance: 24.111

Standard Deviation: 4.910

1. What can we say about the student marks?



Ans: Two maximum outliers are there and the data is slightly skewed towards left because median is greater than mean.

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

Ans: No skewness is present we have a perfect symmetrical distribution

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

Ans: Skewness towards right and heavier tail on the right side compared to the left.

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

Ans: Skewness towards left and heavier tail on the left side compared to the right.

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

Ans: A positive kurtosis value indicates a distribution with heavier tails and a sharper central peak (or more peaked than a normal distribution). This is known as leptokurtic distribution.

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

Ans: A negative kurtosis value indicates a distribution with lighter tails and a flatter peak compared to a normal distribution. This type of distribution is known as platykurtic.

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



What can we say about the distribution of the data?

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

What is nature of skewness of the data?

Ans: Skewness towards left and heavier tail on the left side compared to the right because median is greater than mean

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

Ans: The Inter Quantile Range = 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.

Ans: First there are no outliers. Second both the box plot shares the same median that is approximately in a range 262.5 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)
  3. P (20<MPG<50)

Ans:

P(MPG>38) = 0.348

P(MPG<40) = 0.729

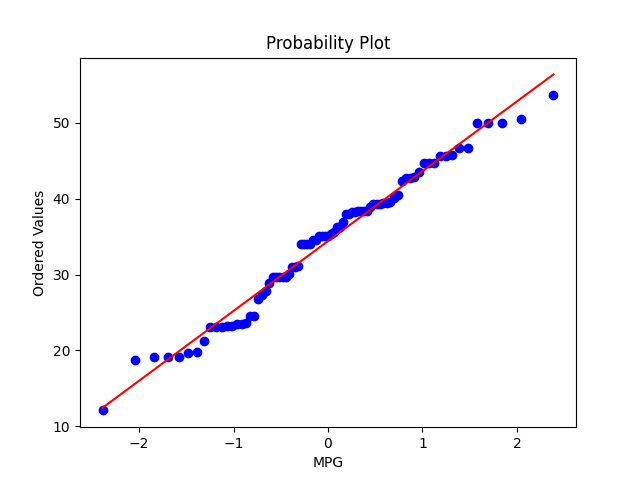
P(20<MPG<50) = 0.013

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans:



The data points closely follow the diagonal line. MPG of cars follows normal distribution

P\_value check:

Shapiro-Wilk test statistic: 0.978

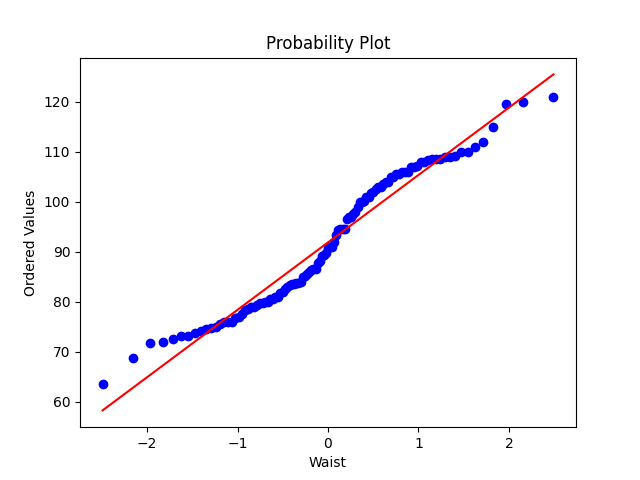
P-value: 0.176

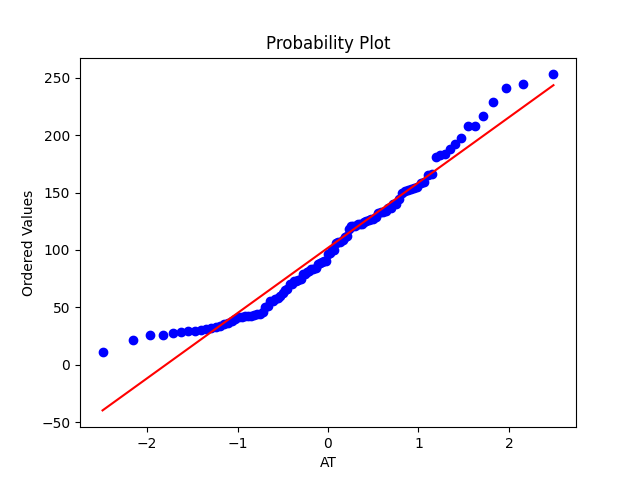
The data appears to be normally distributed. p\_value > 0.05

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

Dataset: wc-at.csv

Ans:





The data points deviate significantly from the line, Adipose Tissue (AT) and Waist does not follow Normal Distribution

P\_value check:

---------:Waist:------------

Shapiro-Wilk test statistic: 0.9559

P-value: 0.001

The data does not appear to be normally distributed. p\_value < 0.05

---------:AT:------------

Shapiro-Wilk test statistic: 0.9523

P-value: 0.001

The data does not appear to be normally distributed. p\_value < 0.05

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

Ans:

Z score for 90% Confidence Interval = -1.6449

Z score for 94% Confidence Interval = -1.8808

Z score for 60% Confidence Interval = -0.8416

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

Ans:

T score for 95% Confidence Interval = -2.0639

T score for 94% Confidence Interval = -1.974

T score for 95% Confidence Interval = -2.7969

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

Ans:

t-score: -0.4714

Probability: 0.3217=32.17 %