

Basic statistics_level1/Assignment 1Sahana M-September 4th batch**Q1) Identify the Data type for the Following:**

Ans:

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

Q2) Identify the Data types, which were among the following**Nominal, Ordinal, Interval, Ratio.**

Ans:

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	Ordinal
Level of Agreement	Ordinal
IQ (Intelligence Scale)	Ratio
Sales Figures	Ratio
Blood Group	Nominal

Basic statistics_level1/Assignment 1Sahana M-September 4th batch

Time Of Day	Ordinal
Time on a Clock with Hands	Interval
Number of Children	Nominal
Religious Preference	Nominal
Barometer Pressure	Interval
SAT Scores	Interval
Years of Education	Ordinal

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

Ans:

$$\begin{aligned}P(\text{Two heads and one tail}) &= N(\text{Event (Two heads and one tail)}) / N(\text{Event (Three coins tossed)}) \\&= 3/8 = 0.375 = 37.5\%\end{aligned}$$

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

- a) Equal to 1**
- b) Less than or equal to 4**
- c) Divisible by 2 and 3**

Ans:

Number of possible outcomes for the above event is

$$N(\text{Event (Two dice rolled)}) = 6^2 = 36$$

a.) $P(\text{sum is Equal to 1}) = '0'$ zero null nada none.

$$\begin{aligned}b.) P(\text{Sum is less than or equal to 4}) &= N(\text{Event (Sum is less than or equal to 4)}) / N(\text{Event (Two dice rolled)}) \\&= 6 / 36 = 1/6 = 0.166 = 16.66\%\end{aligned}$$

$$\begin{aligned}c.) P(\text{Sum is divisible by 2 and 3}) &= N(\text{Event (Sum is divisible by 2 and 3)}) / N(\text{Event (Two dice rolled)})\end{aligned}$$

$$= 6 / 36 = 1/6 = 0.16 = 16.66\%$$

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

$$N(\text{Event (2 balls are drawn randomly from bag)}) = 7! / 2! * 5!$$

$$= (7*6*5*4*3*2*1) /$$

$$(2*1) * (5*4*3*2*1)$$

$$N(\text{Event (2 balls are drawn randomly from bag)}) = (7*6) / (2*1) = 21$$

$$\text{If none of them drawn 2 balls are blue} = 7 - 2 = 5$$

$$N(\text{Event (None of the balls drawn is blue)}) = 5! / 2! * 3! = (5*4) / (2*1)$$

$$= 10$$

$$P(\text{None of the balls drawn is blue}) = N(\text{Event (None of the balls drawn is blue)}) /$$

$$N(\text{Event (2 balls are drawn randomly from$$

$$\text{bag)})$$

$$= 10 / 21$$

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:

Basic statistics_level1/Assignment 1

Sahana M-September 4th batch

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

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.

Please refer to the attachment q7.ipynb file in mail.

Ans:

Mean for Points = 3.59, Score = 3.21 and Weigh = 17.84

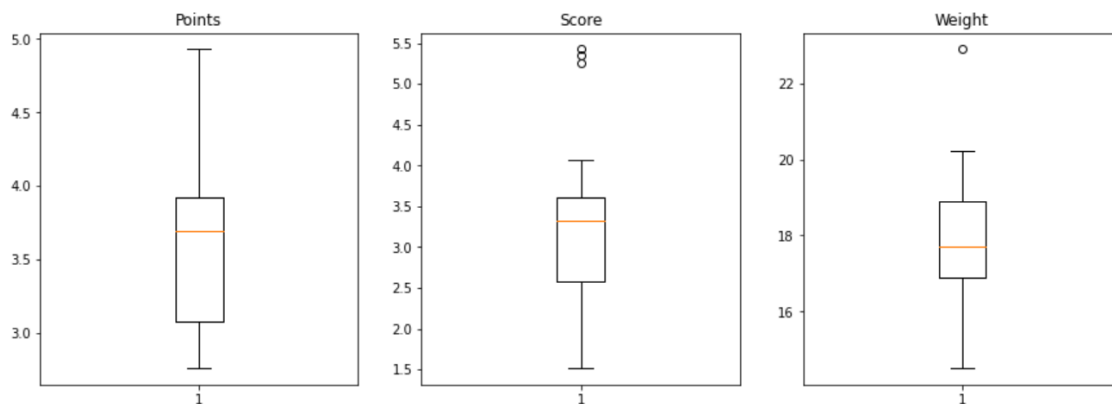
Median for Points = 3.69, Score = 3.32 and Weigh = 17.71

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

Variance for Points = 0.28, Score = 0.95, Weigh = 3.19

Standard Deviation for Points = 0.53, Score = 0.97, Weigh = 1.78

Range [Min-Max] for Points [3.59 – 4.93], Score [3.21 – 5.42] and Weigh [17.84 – 22.9]



Q8) Calculate Expected Value for the problem below

**a) 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)(108)+ (1/9)(110)+ (1/9)(123)+ (1/9)(134)+ (1/9)(145)+ (1/9)(167)+
(1/9)(187)+ (1/9)(199)

= 145.33

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

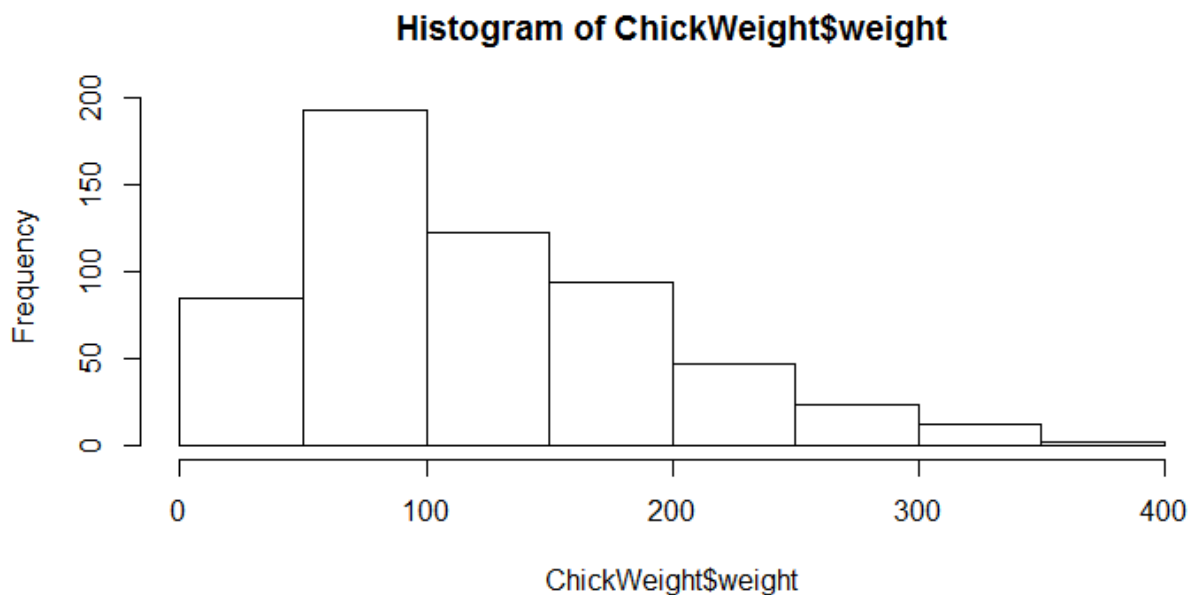
Cars speed and distance

Use Q9_a.csv

Ans:

Please refer to the attachment q9.ipynb file in mail attachment.

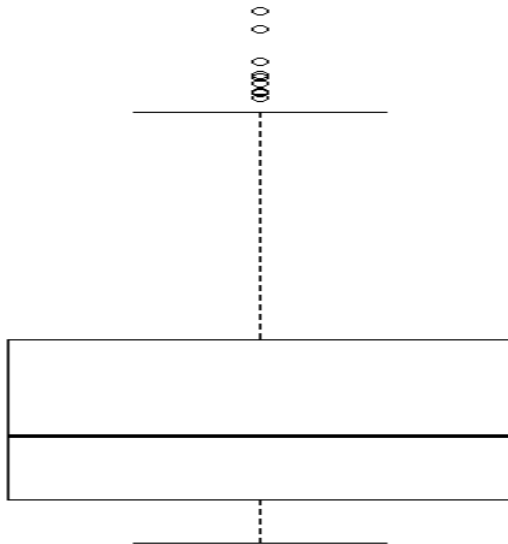
Q10) Draw inferences about the following boxplot & histogram



Basic statistics_level1/Assignment 1

Sahana M-September 4th batch

Ans: The histograms peak has right skew and tail is on right. $\text{Mean} > \text{Median}$. 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:

For 94% confidence interval Range is [198.73 – 201.26]

For 98% confidence interval range is [198.43 – 201.56]

For 96% confidence interval range is [198.62 – 201.37]

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, Median =40.5, Variance =25.52 and Standard Deviation =5.05

2) What can we say about the student marks?

Ans: we don't have outliers and the data is slightly skewed towards right because mean is greater than median.

Q13) What is the nature of skewness when mean and 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 and tail is towards Right

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

Ans: Skewness and tail is towards left

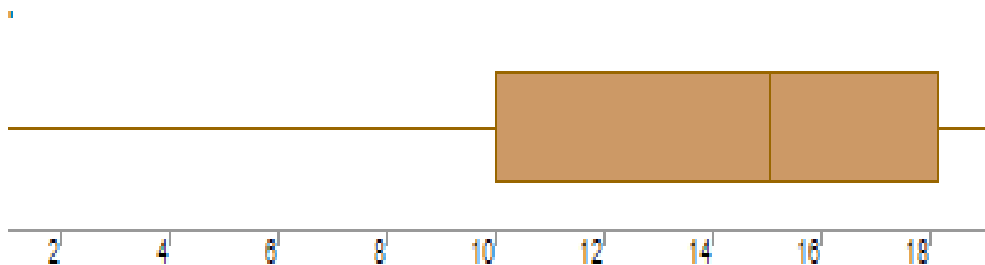
Q16) What does positive kurtosis value indicates for a data?

Ans: Positive kurtosis means the curve is more peaked and it is Leptokurtic

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

Ans: Negative Kurtosis means the curve will be flatter and broader

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

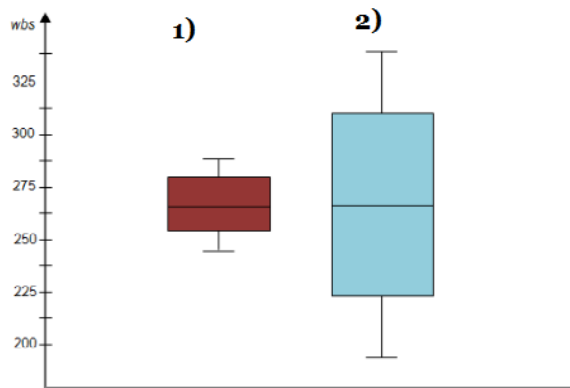
What will be the IQR of the data (approximately)?

Basic statistics_level1/Assignment 1

Sahana M-September 4th batch

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 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

- a. $P(\text{MPG} > 38)$
- b. $P(\text{MPG} < 40)$
- c. $P(20 < \text{MPG} < 50)$

Basic statistics_level1/Assignment 1

Sahana M-September 4th batch

Ans:

The probability of $P(\text{MPG} > 38)$:

```
from scipy import stats
import scipy
1 - stats.norm.cdf(38 , df.MPG.mean() , df.MPG.std())
```

The probability of $P(\text{MPG} > 40)$:

```
stats.norm.cdf(40 , df.MPG.mean() , df.MPG.std())
```

The probability of $P(20 < \text{MPG} < 50)$:

```
stats.norm.cdf(50 , df.MPG.mean() , df.MPG.std()) - (1 - stats.norm.cdf(20
, df.MPG.mean() , df.MPG.std()))
```

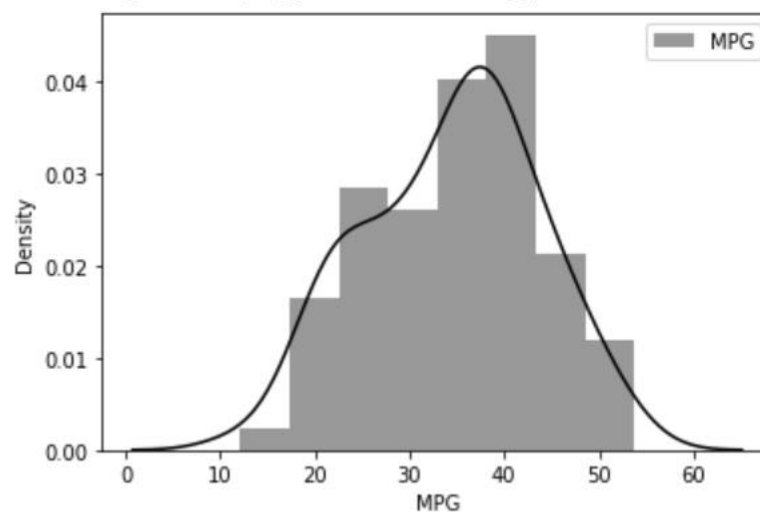
Q 21) Check whether the data follows normal distribution

a) Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans:

a.) MPG of cars follows normal distribution method.



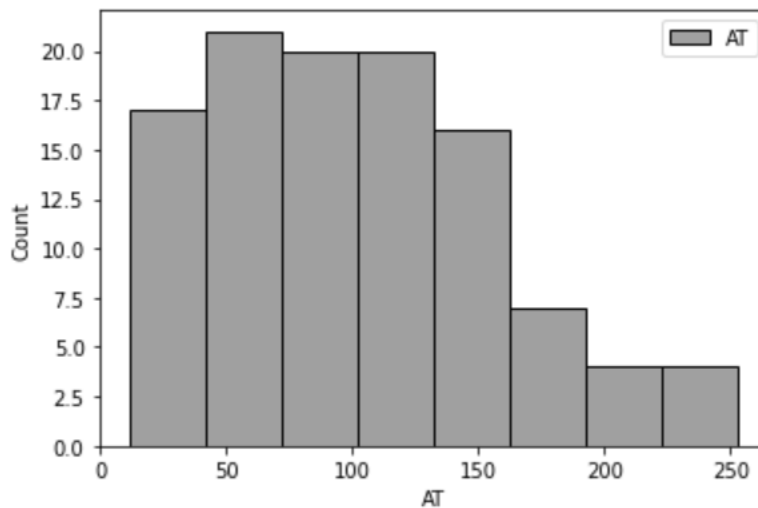
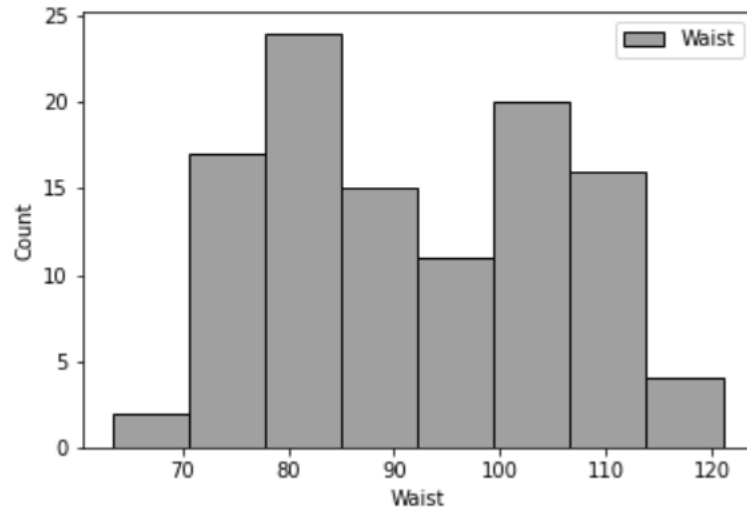
b) Check Whether the Adipose Tissue (AT) and Waist Circumference (Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

Ans: Adipose Tissue (AT) and Waist does not follow Normal Distribution
Below are the plotgrams for AT and Waist.

Basic statistics_level1/Assignment 1

Sahana M-September 4th batch



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

Ans: Assuming that the given data is two tailed test

90% falls under 5% of significance level which divided by 1 is 0.05 so the value for **0.05 from the ZTable is 1.645**

90% falls under 3% of significance level which divided by 1 is 0.03 so the value for **0.03 from the ZTable is 1.555**

Basic statistics_level1/Assignment 1

Sahana M-September 4th batch

90% falls under 20% of significance level which divided by 1 is 0.2 so the value for **0.2 from the ZTable is 0.253**

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

Hint:

rcode \rightarrow pt(tscore,df) , df \rightarrow degrees of freedom

$$T = (260 - 270) / (90 / \sqrt{18}) = -0.471404$$

$$p_value \text{ is } 1 - \text{stats.t.cdf}(\text{abs}(-0.4714), \text{df}=17) = 0.32167$$

$$p_value \text{ is } \text{stats.t.sf}(\text{abs}(-0.4714), \text{df}=17) = 0.32167$$