Q1) Identify the Data type for the Following:

Activity	Data Type
Number of beatings from Wife	Discrete data type
Results of rolling a dice	Discrete data type
Weight of a person	Continuous data type
Weight of Gold	Continuous data type
Distance between two places	Continuous data type
Length of a leaf	Continuous data type
Dog's weight	Continuous data type
Blue Color	Discrete data type
Number of kids	Discrete data type
Number of tickets in Indian railways	Discrete data type
Number of times married	Discrete data type
Gender (Male or Female)	Discrete data type

Q2) Identify the Data types, which were among the following

Nominal, Ordinal, Interval, Ratio.

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

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

Sol: As 3 Coins are tossed.

The possible outcomes:

S = {HHH, HTH, HHT, HTT, TTT, THT, TTH, THH}

E = {HTH, HHT, THH}

P(E) = (3/8)

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

a) Equal to 1
b) Less than or equal to 4
c) Sum is divisible by 2 and 3

N(S) = 36

a) P (sum = 1) = 0
b) E = {(1,1), (2,1), (3,1), (2,1), (2,2,), (3,1)}
P(sum<=4) = (6/36) = (1/6)
c) E = {(1,5), (2,4), (3,3), (4,2),(5,1),(6,6)}
P (sum is divisible by 2 and 3) =(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?

Total no of balls =7

No of blue balls (B) = 2

No of possible ways to draw 2 balls = $^{7}C_{2}$

No of ways to draw 2 balls that are not blue = $^{5}C_{2}$

Probability = ($^{5}C_{2}$)/($^{7}C_{2}$) = (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

sol: Expected value from randomly selected =Σ (X*P(X))

= ((1*0.015) +(4*0.20) +(3*0.65) +(5*0.005) +(6*0.01) +(2*0.120))

Σ (X*P(X)) =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

Sol :(Refer ipnyb basic stats 1)

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?

Sol : The probability of one patient out of 9 given observations is =1/9

1/9 [108 +110 + 123 + 134 +135 + 145 + 167 + 187 + 199] = 1/9 [1308] = 145.3333

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

Cars speed and distance

Use Q9_a.csv

Sol : (refer ipnyb basic stats 1)

Skewness : it will give us information about the symmetrical property of our histogram. It is measuring the symmetry of the distribution

Kurtosis : it is measure of the combined weight of a distribution's tails relative to the centre of the distribution or

The sharpness of the peak of a frequency distribution curve

skewness(Q9_a['speed']) = -0.1139548 : left skewed

skewness(Q9_a['dist']) = 0.7824835 : right skewed

kurtosis

kurtosis(Q9_a['speed']) = 2.422853 it refers distribution is sharp peak

kurtosis(Q9_a['dist']) = 3.248019 it refers distribution is sharp peak with excess kurtosis

(refer ipnyb basic stats 1)

SP and Weight(WT)

Use Q9_b.csv

Sol :

(Refer ipnyb for basic stats 1)

skewness(Q9_b['SP']) = 1.581454 : right skewed

skewness(Q9_b['WT']) = -0.6033099 : left skewed

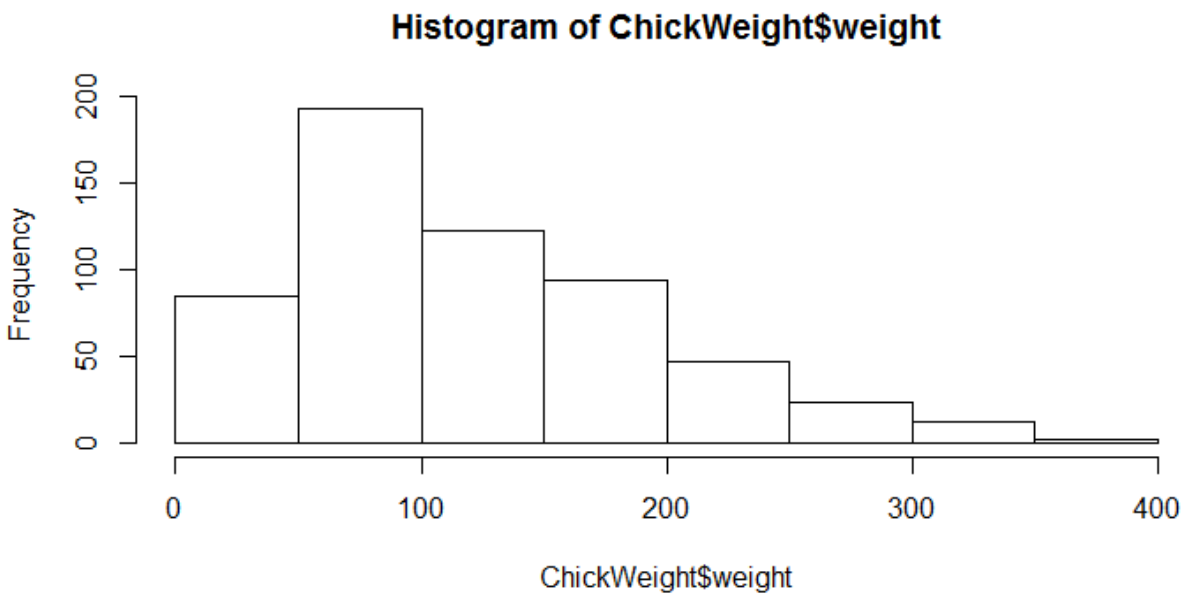
kurtosis

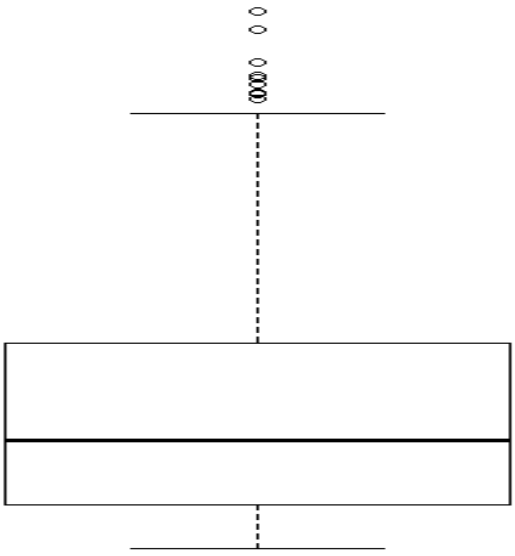
kurtosis(Q9_b['SP']) = 5.723521 it refers distribution is sharp peak and excess kurtosis

kurtosis(Q9_b['WT']) = 3.819466 it refers distribution is sharp peak with positive kurtosis

(refer ipnyb basic stats 1)

Q10) Draw inferences about the following boxplot & histogram





Observations:

Lower whisker = 25%

Upper whisker =25%

IQR means inter quartile range it has 50% of observation (Q3-Q1)

Lower limit = Q1 – 1.5(IQR)

Upper limit = Q3 + 1.5 IQR

Q1 have 25% of observation and Q3 have 25% of observation

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?

Sol : refer ipnyb basic stats

For 94% CI : 201.0429738341421
For 98% CI : 201.37769665186667
For 96% CI : 201.17439591877456

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.

Sol: (refer ipnyb for basic stats 1)

Mean = (738/18) = 41

Median = (40+41)/2 = 40.5

Variance = 42.666667

Standard deviation = 6.531973

2) What can we say about the student marks?

Sol :

Most of the students scored marks between 40 and 45

mean>median , so skewness is always positive

positive kurtosis

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

Sol : **mean = median then skewness = 0 this is perfectly symmetric data**

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

Sol : **mean > median then skewness is right skewed**

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

Sol : median > mean then skewness is left skewed

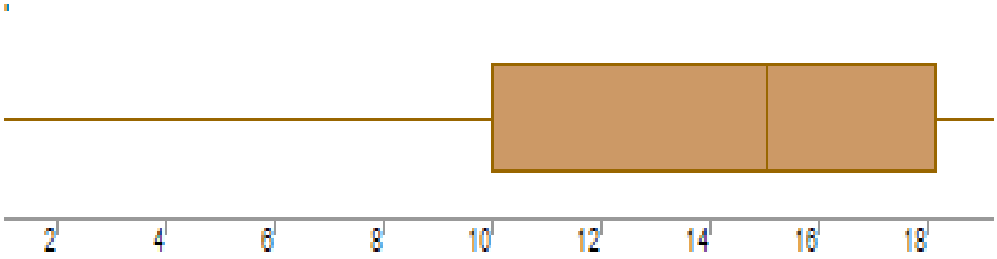
Q16) What does positive kurtosis value indicates for a data?

Sol : The positive kurtosis value indicates sharp peak distribution of data

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

Sol : The negative kurtosis value indicates flat tail distribution

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



What can we say about the distribution of the data?

Sol : mean < median, the skewness is left skewed

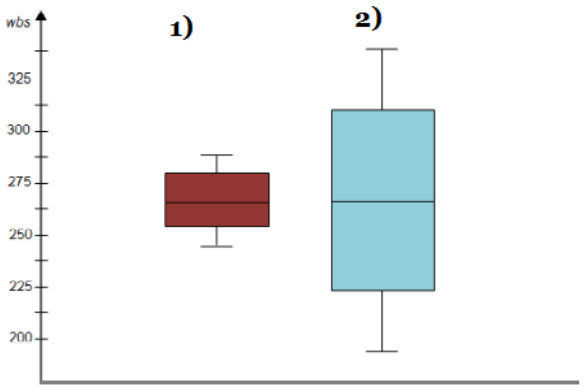
What is nature of skewness of the data?

Sol : Left skewed

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

Sol : IQR = 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.

Sol :

From boxplot 1:

Q3 =275 and Q1 = 250

IQR = Q3 – Q1 = 275- 250 =25

Lower extreme = Q1 – 1.5 * IQR = 250 -1.5 * 25 = 212.5

Upper extreme = Q3 + 1.5 * IQR = 275 + 1.5 * 25 = 312.5

From 2nd boxplot:

Q3 = 300, Q1 = 225

IQR (2) = 300 – 225 = 75

Lower extreme = 225 – 1.5 * 75 =112.5

Upper extreme = 300 + 1.5 * 75 =412.5

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(MPG>38)
b. P(MPG<40)
c. P (20<MPG<50)

Sol: (refer ipnyb basics stats 1)

a. **P(MPG>38) =** 0.652406074841
b. **P(MPG<40) =** 0.7293498762151616
c. **P (20<MPG<50) =** = 0.8988689169682046

Q 21) Check whether the data follows normal distribution

a) Check whether the MPG of Cars follows Normal Distribution
Dataset: Cars.csv

Sol: from ipnyb basic stats 1

Mean = 34.422075728024666
Median = 35.1527

Mean = Median (approximately)

The data is slightly right skewed , can say it follows normal distribution.

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

Sol : refer ipnyb in basic stats 1

Adipose Tissue (mean, median) = (101.89403669724771, 96.54)
Waist Circumference (mean, median) = (91.90183486238533, 90.8)

For given data's mean and median are not equal, so it is not following the normal distribution

---

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

Sol: refer ipnyb basic stats 1:

Z scores of 90% confidence interval = 1.2815515655446004
Z scores of 94% confidence interval = 1.5547735945968535
Z scores of 60% confidence interval = 0.2533471031357997

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

Sol: refer ipnyb basic stats 1

N = 25 , df = N-1 = 24
t scores of 95% confidence interval = 1.7108820799094275
t scores of 96% confidence interval = 1.8280511719596342
t scores of 99% confidence interval = 2.4921594731575762

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

Sol:

X = 260

Mean = 270

N = 18

Std = 90

Df = (N-1) =17

SE = Std / sqrt (N) = 90/ sqrt(18) = 21.213203435596427

t = (X-Mean)/SE = (260-270)/21.213203435596427 = -0.4714045207910317

P(t) = 0.32167253567098364 (refer ipnyb basic stats 1)