14/11/22 Assignment 1 – BASIC STATICS LEVEL 1

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	Nominal
Number of kids	Discrete
Number of tickets in Indian railways	Discrete
Number of times married	Discrete
Gender (Male or Female)	Categorical

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)	Interval
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	Ratio
SAT Scores	Interval
Years of Education	Ratio

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

Let S be the sample space then

S = (h,t),(h,t), (h,t) then the event E is {hht,hth,thh}

Considering X be as random variable denoting the two heads and one tail.

P(X=2) = probability of occurrence of 2 heads and 1 tail.

   =P(hht)+P(hth)+P(thh)

   =1/2 x ½ x ½ +1/2 x ½ x ½ + 1/2 x ½ x ½ =3/8

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

   a) Equal to 1 - 0
   b) Less than or equal to 4 - 6
   c) Sum is divisible by 2 and 3 – 2/3

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?

There are total of 2+3+2=7 balls,

2 balls can be selected in 7C2= 21 ways

2 balls can be selected from the Red-Green set in 5C2 = 10 ways

Probability=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

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

= 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.

	Mean	Median	Mode	Standard Deviation	Variance	Range
Points	3.596563	3.695	3.92	0.534679	0.285886	2.17
Score	3.217250	3.325	3.44	0.978457	0.957378	3.911
Weight	17.848750	17.710	17.02	1.786943	3.193166	8.4

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?

Probability of selecting each patient = 1/9

Expected value = $\sum$ (probability * Value)

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

SPEED	Skewness	Kurtosis
DISTANCE	-0.117510	-0.508994
	0.806895	0.405053

FOR SPEED - Skewness value is negative so it is Negative skewed. Since magnitude is slightly greater than 0 it is slightly left skewed.

FOR DISTANCE - Right skewed or positive skewed. Slight magnitude to right

Q10) Draw inferences about the following boxplot & histogram
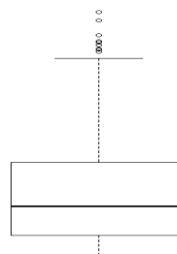
The histograms peak has right skew and tail is on right. Mean > Median. We have outliers on the higher side.

The median is closer to the bottom of the box, and the whisker is shorter on the lower end of the box, So the distribution is positively skewed (skewed right).

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?

For 94% confidence interval (198.738325292158, 201.261674707842)
For 98% confidence interval (198.43943840429978, 201.56056159570022)
For 96% confidence interval (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.

MEAN	MEDIAN	VARIANCE	STANDARD DEVIATION
41.00	40.5	25.5294	5.0526

2) What can we say about the student marks?

There is total 18 data. Minimum marks are 34 and maximum marks are 56. 75% students have 42 marks and below.

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

Perfect symmetrical distribution

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

Distribution is Positive skewed.

Q15) What is the nature of skewness when median > mean?
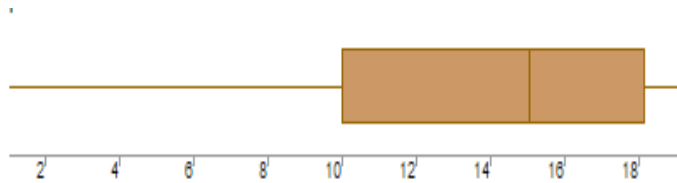
Distribution is Negative skewed.

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

Distribution is peaked and has thick tail

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

Distribution is flat and has thin tail

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

What can we say about the distribution of the data?

Median is closer to the top of the box plot and whisker is shorter on the upper end of the box. Therefore, the distribution is negatively skewed.

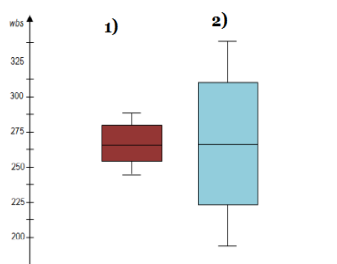What is nature of skewness of the data?

Distribution is negatively skewed.

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

Inter Quartile Range = Third Quartile - First Quartile = 18.2-10 = 8.2

Q19) Comment on the below Boxplot visualizations?

Draw an Inference from the distribution of data for Boxplot 1 with respect Boxplot 2.

a)- Shorter Boxplot indicates that data is remain around center

   values.

b)- Long Boxplot indicates that data implies more variable data.

Q 20) Calculate probability from the given dataset for the below cases.
Calculate the probability of MPG of Cars for the below cases.

a. P(MPG>38) = 33/81
b. P(MPG<40) = 20/81
c. P(20<MPG<50)= 69/81

Q 21) Check whether the data follows normal distribution

    a)   Check whether the MPG of Cars follows Normal Distribution
        Distribution of MPG data is Right Skewed in nature.

    b)   Check Whether the Adipose Tissue (AT) and Waist Circumference (Waist) from wc-at data set follows Normal Distribution
        1) Distribution of AT is Right Skewed in nature.

        2)Distribution of Waist data is Negative Kurtosis in nature.

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

A = (1+CL)/2 = (1+0.90)/2 = 0.95 ➔Z score = 1.65

A = (1+CL)/2 = (1+0.94)/2 = 0.97 ➔Z score = 1.89

A = (1+CL)/2 = (1+0.60)/2 = 0.8 ➔Z score = 0.84

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

➔ t scores of 95% confidence interval:

Df = n-1 = 25 − 1 = 24, Therefore t score = 2.064

➔ t scores of 96% confidence interval = 1.8280

➔ t scores of 99% confidence interval = 2.797

 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 : 32.16%