Q1) Identify the Data type for the Following:

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

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

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

Ans: P (X=2H and 1T) = 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

Ans:

a) P(X=1) =0/36 =0
b) P(X<=4) =6/36 = 1/6
c) P (X= 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?

Ans: P (X!= none of the balls drawn is blue)=10/24 = 5/12

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=
(1*0.015)+(4*0.2)+(3*0.65)+(5*0.005)+(6*0.01)+(2*0.12)

⇨ 0.015+0.8+1.95+0.025+0.06+0.24

⇨ 3.125

So, the expected number of candies for a randomly selected child is 3.125.

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:

	Points	Score	Weigh
Mean	3.596563	3.21725	17.84875
Median	3.695	3.325	17.71
Mode	numeric	numeric	numeric
Variance	0.2858814	0.957379	3.193166
Standard Deviation	0.5346787	0.9784574	1.786943
Range	2.76-4.93	1.512-5.424	14.5-22.9

mean(Q7$Points) median(Q7$Points) mode(Q7$Points) var(Q7$Points) sd(Q7$Points) range(Q7$Points)	mean(Q7$Score) median(Q7$Score) mode(Q7$Score) var(Q7$Score) sd(Q7$Score) range(Q7$Score)	mean(Q7$Weigh) median(Q7$Weigh) mode(Q7$Weigh) var(Q7$Weigh) sd(Q7$Weigh) range(Q7$Weigh)

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 of mean=(Sum of all weights)/(Total number of patients)

Total number of patients = 9

Sum of all weights= 108+110+123+134+135+145+167+187+199

= 1308

Expected value of mean= 1308/9

= 145.33

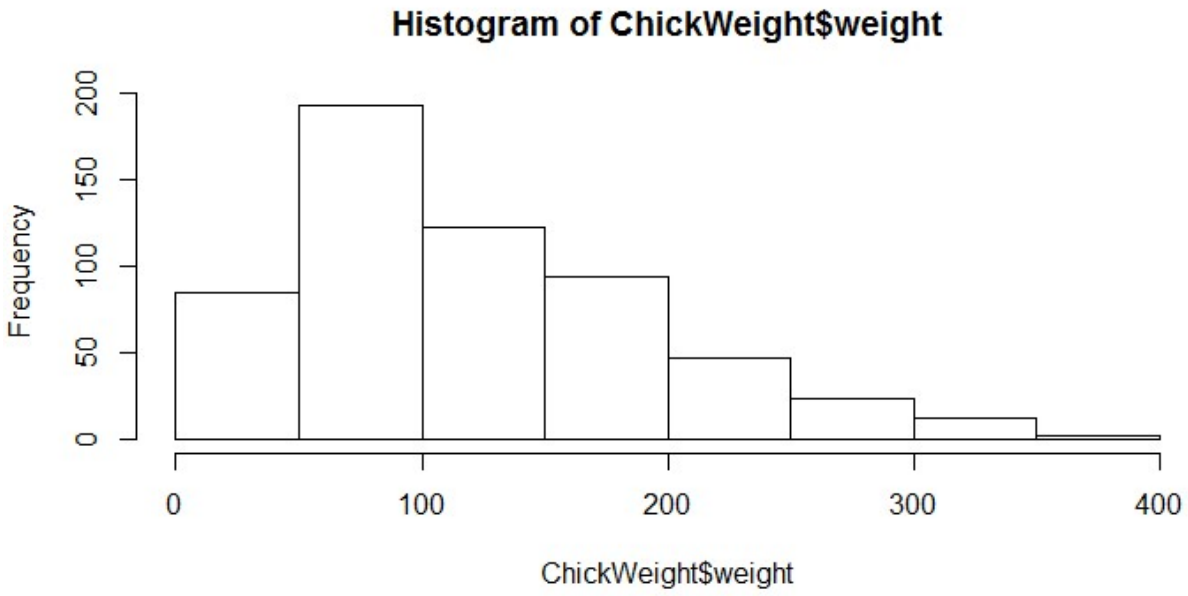
Code:

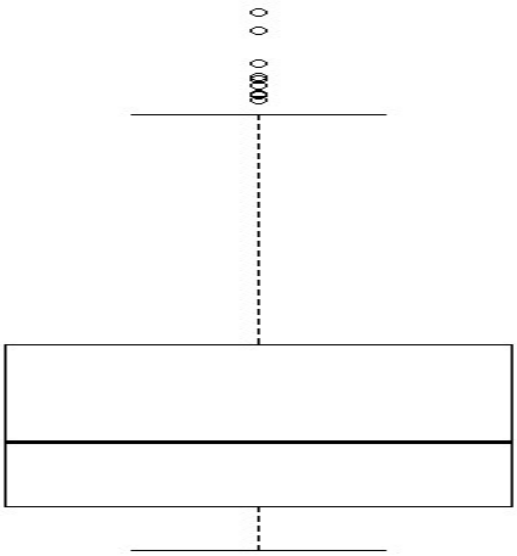
```
X=c(108,110,123,134,135,145,167,187,199)
expected_value=mean(X)
cat("Expected value of weight:", expected_value)
```

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

Cars speed and distance **Use Q9_a.csv**	**SP and Weight(WT)** **Use Q9_b.csv**
Ans: install.packages(“moments”) library(“moments”) skewness(Q9_a$speed) -0.1139548 kurtosis(Q9_a$speed) 2.422853 skewness(Q9_a$dist) [1] 0.7824835 kurtosis(Q9_a$dist) [1] 3.248019	Ans: install.packages(“moments”) library(“moments") skewness(Q9_b$SP) 1.581454 kurtosis(Q9_b$SP) 5.723521 skewness(Q9_b$WT) -0.6033099 kurtosis(Q9_b$WT) 3.819466

Q10) Draw inferences about the following boxplot & histogram





Ans

Histogram:

In the above mentioned Histogram, the data points are plotted with a difference of 50 units each with a minimum and maximum range of 0-400 units on ChickWeight$weight (X-axis), and the Frequency(Y-axis) contains the data points with an difference of the 50 units each with a minimum and maximum range of 0-200 units. So, the minimum range of frequency of 0-100 occurred in the range of 350-400, hence we observed that it is a right skewed.

Boxplot:

By looking at the boxplot, we can see that there are some outliers present, we also observe that the median is less than mean as there are less data points between Q1 and below points range, hence we can consider it as the right skewed.

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:

Given

sample_mean=200

```
sample_sd=30

sample_size=2000

population_size=3000000


confidence_levels=c(0.94,0.96,0.98)  #confidence level

se=sample_sd/sqrt(sample_size)        # standard error


z-score=qnorm(1-(1-confidence_levels)/2)

me=z_scores*se                        # margin error

#it is going to calculate the confidence intervals

confidence_intervals=data.frame(confidence_level=confidence_levels,lower_boun
d=sample_mean-me,upper_bound=sample_mean+me)

# print the confidence_intervals

print(confidence_intervals)
```

confidence_level	lower bound	upper bound
0.94	198.7383	201.2617
0.96	198.6223	201.3777
0.98	198.4394	201.5606

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?

Ans:

1)Mean:

a=(34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56)

a

mean(a)

Output: 41

Median:

median(a)

Output: 40.5

Mode:

mode(a)

Output: “numeric”

Variance:

var(a)

Output: 25.52941

Standard Deviation:

sd(a)

Output: 5.052664

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

Ans:

When the mean and median of a dataset are equal, it typically indicates that the data is symmetrically distributed. This symmetry suggests that the skewness of the data is close to zero.

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

Ans:

When the mean is greater than the median (or) median is lesser than the mean, then it is going to be right skewness.

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

Ans:

When the median is greater than the mean (or) mean is lesser than the median, then it is going to be considered as left skewness.

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

Ans:

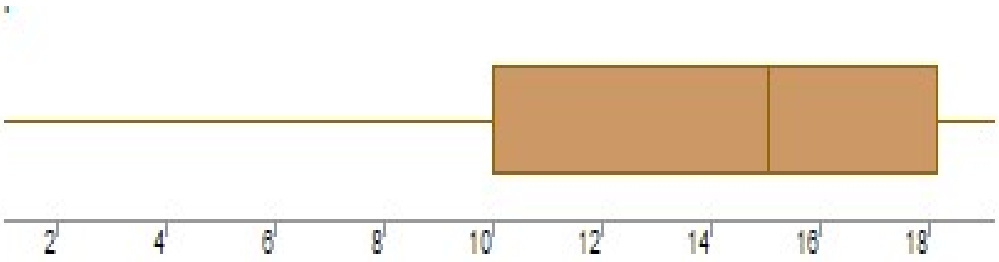
A positive kurtosis value indicates that a dataset has heavy tails with more peaks.

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

Ans:

A negative kurtosis value indicates that a dataset has tinner tails.

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



i. What can we say about the distribution of the data?

Ans:

The data is actually distributed to the right-side which it contains most of the data values was in between the 10-18, hence the plotting was also be form in that range only, where the median lies around at the 15.3 (Q1)

ii. What is the nature of skewness of the data?

Ans:

In the boxplot, most of the data was lying in the range of 10-18, hence it will be considered as the left skewed, as the data values in upper quadrant contains more than the low quadrant.

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

Ans:

Q1=10.2

Q2=15.3

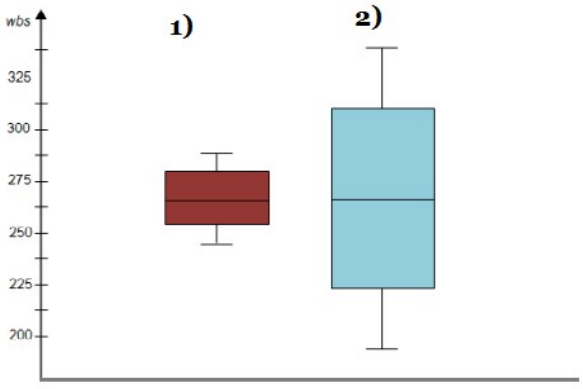
Q3=18.2

IQR=Q3-Q1

=18.2-10.2

=8

Q19) Comment on the below Boxplot visualizations?



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

Ans:

In the above given 2 boxplots, we can observe that mean and median are equal, even though the whisker level is high in boxplot 2, hence we can consider the distribution is symmetrical.

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)

Ans:

View(Cars)

Cars

MPG← Cars$MPG

MPG

probability_a = length(MPG[MPG>38])/length(MPG)

probability_a

0.4074

probability_b = length(MPG[MPG<40])/length(MPG)

probability_b

0.7530

probability_c = length(MPG[MPG>20&MPG<50])/length(MPG)

probability_c

0.8518

Q 21) Check whether the data follows normal distribution

a) Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

(file attached)

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

Dataset: wc-at.csv

(file attached)

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

Ans:

Z-scores for 90% confidence interval:

Z_90=qnorm(0.95)

Z_90

1.644854

Z-scores for 94% confidence interval:

Z_94=qnorm(0.97)

Z_94

1.880794

Z-scores for 60% confidence interval:

Z_60=qnorm(0.8)

Z_60

0.8416212

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

Ans:

Given that,

T-score of 95% of confidence interval:

N=25

confidence interval=c(0.95)

t_scores=qt(1-(1-confidence_interval)/2,df=n-1)

t_scores

2.063899

T-score of 96% of confidence interval:

N=25

confidence interval=c(0.96)

t_scores=qt(1-(1-confidence_interval)/2,df=n-1)

t_scores

2.171545

T-score of 99% of confidence interval:

N=25

confidence interval=c(0.99)

t_scores=qt(1-(1-confidence_interval)/2,df=n-1)

t_scores

2.79694

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

(file attached)