|  |  |
| --- | --- |
| 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) | Interval |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Interval |
| Number of Children | Ordinal |
| 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:** 3/8

**Explanation:** (HHH, **HHT, HTH, THH**, TTH, THT, HTT, TTT)

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:** a) 0 b) 1/6 c) 1/6

**Explanation:**

1. The probability of getting sum=1 is zero. i.e. P(1) = 0.
2. The probability of getting<=4 is 1/6.

When we roll two dice, the possibility of getting <= 4 is (1, 1), (1, 2), (1, 3), (2, 1), (2, 2), and (3, 1).So,

* The number of favorable outcomes = 6
* Total number of possibilities = 36
* Probability = The number of favorable outcomes / Total number of possibilities = 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:** The probability of getting no blue balls is 10/21.

**Explanation:**

Total number of balls= (2+3+2) = 7

Let S be the sample space.  
Then, n(S) = Number of ways of drawing 2 balls out of 7  
=7C2 = (7×6) / (2×1) = 21

Let E = Event of drawing 2 balls, none of which is blue.  
∴ n(E)= Number of ways of drawing 2 balls out of (2 + 3) balls.  
=5C2​ = (5×4) / (2×1)​ = 10

∴ P(E) = n(E) / n(S) = 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:**

Expected number of candies for a randomly selected child = 3.09

**Explanation:**

Expected number of candies for a randomly selected child

= 1 \* 0.015 + 4\*0.20 + 3 \*0.65 + 5\*0.005 + 6 \*0.01 + 2 \* 0.12

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

= 3.09

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

* For Points,Score,Weight

Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

**Use Q7.csv file**

**Ans:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **Point** | **Score** | **Weight** |
| **Mean** | 3.60 | 3.22 | 17.85 |
| **Median** | 3.70 | 3.33 | 17.71 |
| **Mode** | 3.914 | 3.49 | 17.65 |
| **Variance** | 0.29 | 0.96 | 3.19 |
| **Std Deviation** | 0.53 | 0.98 | 1.79 |
| **Range** | 2.76, 4.93 | 1.513, 5.424 | 14.5, 22.9 |

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 of the weight for a randomly selected patient = 145.33

**Explanation:**

Expected Value = ∑ (probability \* Value)

Probability of selecting each patient = 1/9

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

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

= (1/9) (1308)

= 145.33

**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**:

Skewness and Kurtosis of Car speed and distance is as follows:

Skewness = -0.111 (car speed), 0.759(distance)

Kurtosis = 2.42(car speed), 3.24(distance)

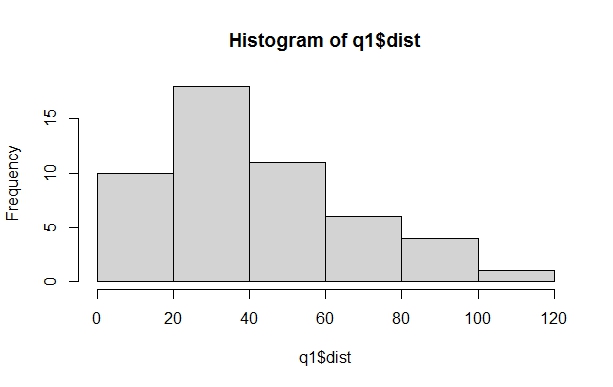
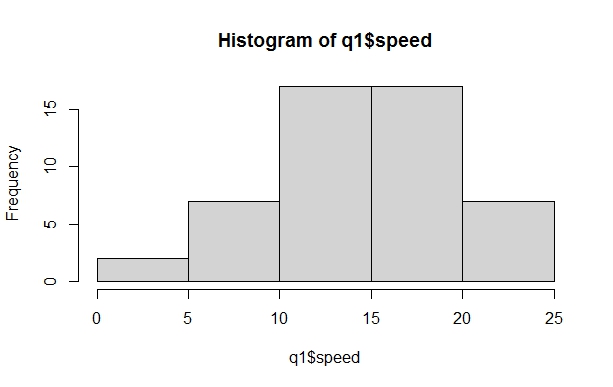
Skewness and Kurtosis of SP and Weight (WT) data are as follows:

Skewness = 1.55 (SP) , -0.59(weight)

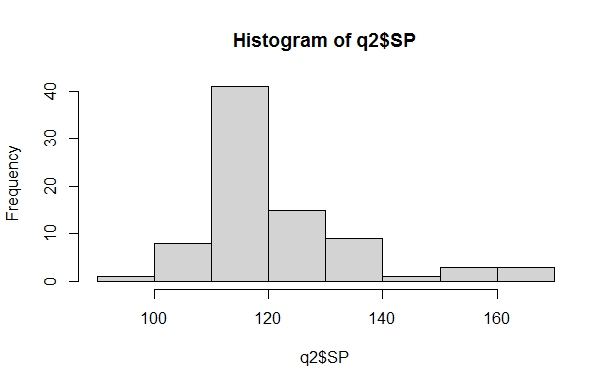
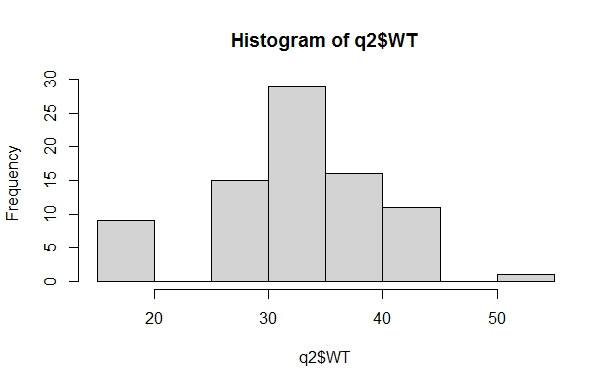
Kurtosis = 5.72 (SP), 3.82(weight)

**Explanation:**

**Inferences of Q9\_a file**

****

**Inferences of Q9\_b file**

** **

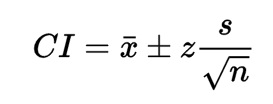
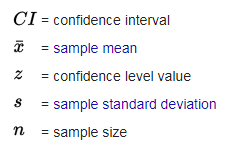
**Q10) Draw inferences about the following boxplot & histogram**



**Ans:** The histogram and boxplot figure is positively skewed on right side. i.e. Mean and median of the data is greater than mode.

**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: x = 200, s = 30, n = 2000

1. The 94% confidence interval is (143.58, 256.42 ) (z= 1.8808)

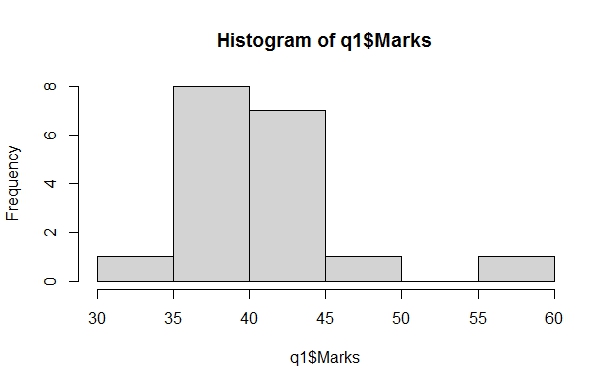
2. The 96% confidence interval is (138.39, 261.61) (z= 2.0537)

3. The 98% confidence interval is (130.21, 269.79) (z= 2.3263)

**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 = 41, Median = 40.5,

Variance = 25.529,

Standard Deviation = 5.05

1. Repeatedly obtained are 36,38,40,41& 42.

Skewness = 1.42 i.e. positive.

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

**Ans:** If the mean is equal to the median as well as the mode, hence the skewness is zero. If the distribution is symmetric, the mean equals to median, and the skewness of the distribution is zero.

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

**Ans:** If the mean is greater than the median, then distribution is positively skewed.

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

**Ans:** If the mean is less than the median, the distribution is negatively skewed.

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

**Ans:** Positive values of kurtosis indicate that distribution is peaked and possesses thick tails. An extreme positive kurtosis indicates a distribution where more of the numbers are located in the tails of the distribution instead of around the mean.

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

**Ans:** If a distribution has negative kurtosis, it is said to be platykurtic**,** which means that it has a flatter peak and thinner tails compared to a normal distribution. This simply means that more data values are located near the mean and less data values are located on the tails. Negative kurtosis is the uniform distribution, which has no peak at all and is a completely flat distribution.

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



1)What can we say about the distribution of the data?

**Ans:**

The distribution in which more values are concentrated on the right side (tail) of the graph is called Negatively Skewed Distribution, while the left tail of the distribution graph is longer.

2)What is nature of skewness of the data?

**Ans:**

The Mean of negatively skewed data will be less than the Median.

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

**Ans:**

The IQR describes the middle 50% of values when ordered from lowest to highest. The Interquartile Range (IQR) = Q(3) – Q(1). In above example of data the IQR = (18-10).

Q19) Comment on the below Boxplot visualizations?



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

**Ans:**

In above figure of boxplot. Boxplot 1 is positively skewed that’s means Mean and Median is greater than Mode. And In boxplot 2 is shows normal distribution that means the skewness for a normal distribution is zero, and any symmetric data should have a skewness near zero.

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)

c. P (20<MPG<50)

**Ans.** a) 0.348 b) 0.73 c) 0.013

**Explanation:**

a. P(MPG>38) = 1- pnorm(38,34.422,9.13144) = 0.3475908

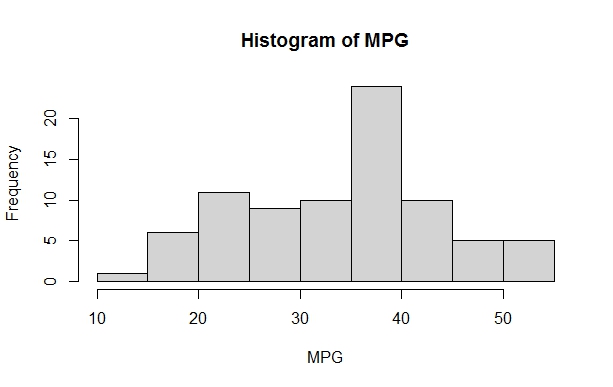
b. P(MPG<40) = pnorm(40,34.422,9.13144)= 0.7293527

c. P (20<MPG<50) =pnorm(50,34.422,9.13144) –

(1-pnorm(20,34.422,9.13144))= 0.01311818

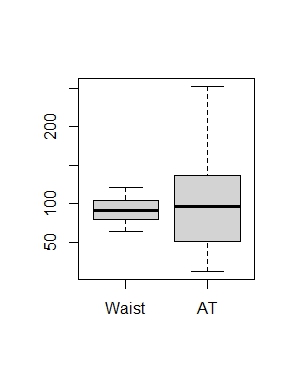
Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

 Dataset: Cars.csv

**Ans:** The MPG of Cars is highly skewed.

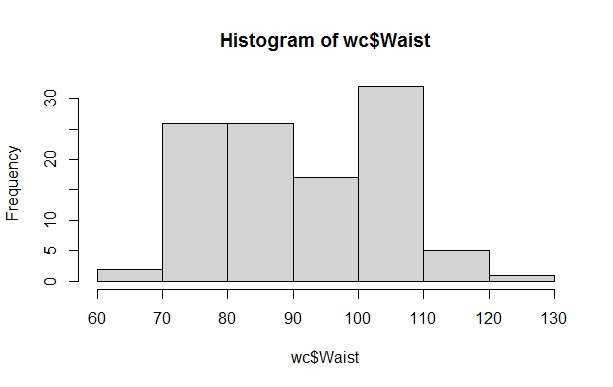
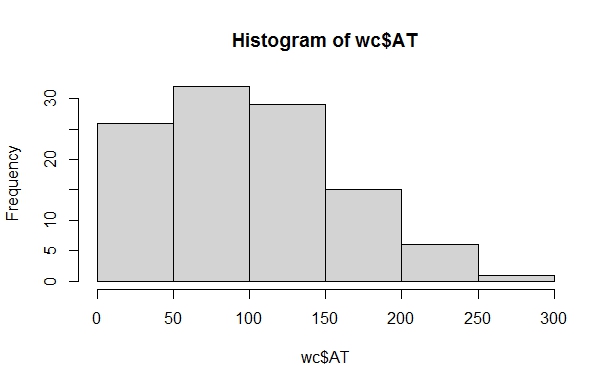
Skewness= -0.1746343



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 Waist data set follows Normal Distribution but AT doesn’t follow Normal Distribution.



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

**Ans:**

Z score of 60% CI = 0.84

Z score of 90% CI = 1.645

Z score of 94% CI = 1.881

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

**Ans:**

|  |  |
| --- | --- |
| **Confidence Interval** | **T Score** |
| 95% | 2.06 |
| 96% | 2.17 |
| 99% | 2.79 |

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:**

For probability calculations, the number of degrees of freedom is n - 1, so here you need the t-distribution with 17 degrees of freedom.

The probability that **t < - 0.471 with 17 degrees of freedom** assuming the population mean is true, the t-value is less than the t-value obtained With 17 degrees of freedom and a t score of - 0.471, the probability of the bulbs lasting less than 260 days on average of **0.3218** assuming the mean life of the bulbs is 300 days.

