|  |  |
| --- | --- |
| 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 | Ratio |
| 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 | 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: Possible outcomes:

S = {(H,H,H), (H,H,T), (H,T,H), (H,T,T), (T,H,H), (T,H,T), (T,T,H), (T,T,T)}

Total number of outcomes = n(S) = 8

Let A be the event that two heads and one tail are obtained

A = {(H,H,T), (H,T,H), (T,H,H)}

n(A) = 3

Hence, the probability of getting two heads and one tail = n(A) / n(s) = 3/8

Thus, the probability of getting two heads and one tail on tossing three coins at once is equal to 3/8.

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
4. Ans: Two Dice are rolled:
5. Sample Space (S) = {(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6)
6. (2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6)
7. (3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6)
8. (4, 1), (4, 2), (4, 3), (4, 4), (4, 5) (4, 6)
9. (5, 1), (5, 2), (5, 3), (5, 4), (5, 5), (5, 6)
10. (6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6)}.
11. n(S) = 36
12. 1) Let A be the event that sum is equal to 1
13. But, The minimum possible sum is (1, 1) = ( 2 ).
14. n(A) = 0
15. Therefore P(A) = n(A)/ n(S) = ( 0 )/( 36 ) = 0
16. Hence, the probability that sum is equal to 1 is 0.
17. 2) Let B be the event that the sum is less than or equal to 4
18. B = {(1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (3, 1)}
19. n(B) = 6
20. Therefore, P(B) = n(B)/n(S) = 6/36 = 1/6
21. Hence, the probability less than or equal to 4 is 1/6.
22. 3) Let C be the event that the Sum is divisible by 2 and 3
23. C = {(1 ,5), (3 , 3), (4 , 2), (5 , 1), (6 , 6)}
24. n(C) = 5
25. Therefore, P(C) = n(C)/n(S) = 5/36
26. Hence, the probability that sum is divisible by 2 and 3 is 5/36.

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 = (2 + 3 + 2) = 7

Let S be the sample space

Then, n(S) = Number of ways of drawing 2 balls out of 7

n(S) = 7C2 = 7! / 2! \* 5!

= (7×6) / (2×1)

n(S)=7C2 = 21

Let A be the event of 2 balls, none of which is blue.

n(A) = Number of ways of drawing 2 balls out of (2 + 3) balls

n(A) = 5C2 = 5! / 2! \* 3!

= (5×4) / (2×1)

n(A) = 5C2 =10

Therefore, P(A) = n(A) / 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

= 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

= 3.090

= 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**

**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 | 3.91 |
| **Std Derivation** | 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 = ∑ (probability \*value)

Probability of selecting each patient = 1/9

Expected value

= (1/9)108+(1/9)110+(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 follow :

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

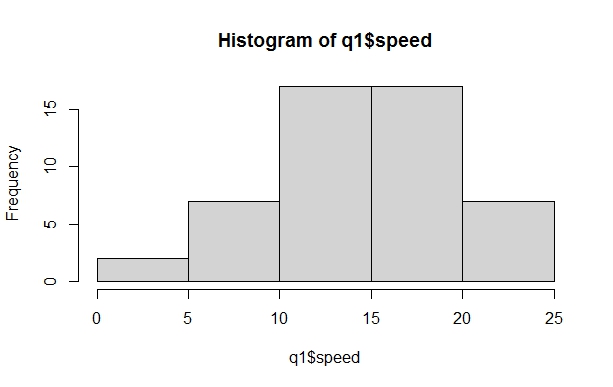
Kurtosis=2.42(car speed),3.24(distance)

Skewness and Kurtosis=of SP and weight (WT) data are as follow:

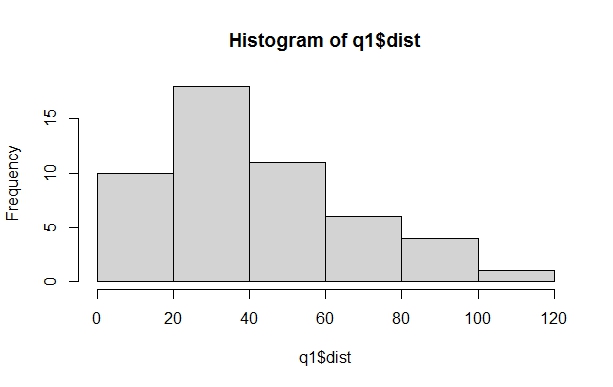
Skewness=1.55(SP), -0.59(weight)

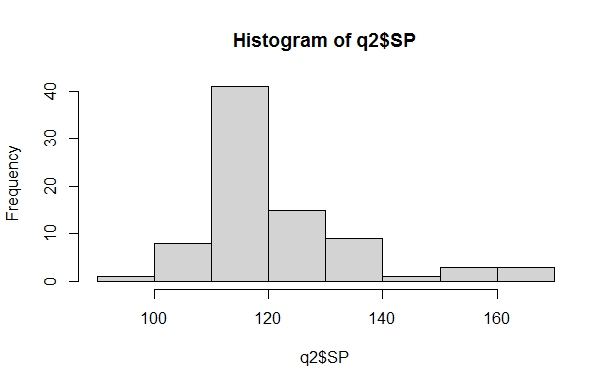
Kurtosis=5.72(SP),3.87(weight)

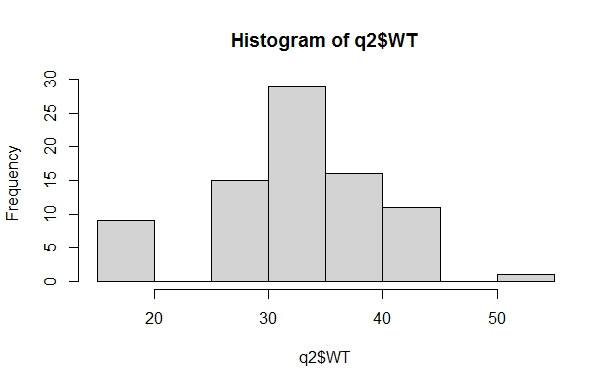
**Inferences of Q9\_a file-**

****

**Inferences of Q9\_b file-**

****

****

****

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



Ans: The most of the data points are concentrated in the range 50-100 with frequency 200.

And least range of weight is 400 somewhere around 0-10.

So the expected value the above distribution is 75.

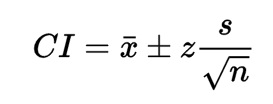
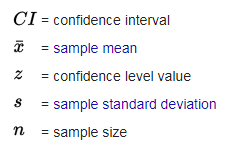
Skewness- we can notice a long tail towards right so it is heavily right skewed



Ans: In this box plot, the data is concentrated downwards and Q2-Q1 < Q3-Q2 which shows positive skewness and has some outliers above the upper quartile.

**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 (198.739,201.62) (z=1.8808)
2. The 96% confidence interval is (198.622,201.378) (z=2.0537)
3. The 98% confidence interval is (198.439,201.561) (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. Repeatably obtained are 36,38,40,41 and 42 Skewness =1.42 i.e., Positive.

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

**Ans: -**

Symetrical

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

Ans: Right skewed

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

Ans: - Left Skewed

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

Ans: - The data is normally distributed and kurtosis value is 0.

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

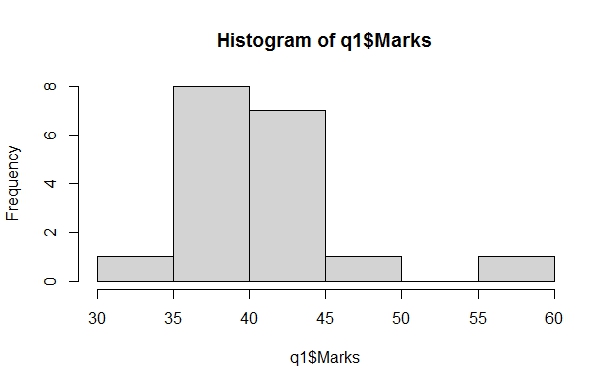
Ans: - The distribution of the data has lighter tails and a flatter peaks than the normal distribution.

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



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

Ans:- The whisker on left and whisker on right on boxplot is not equal so the distribution is not symmetric.



2.What is nature of skewness of the data?

Ans:- Left skewed, median is greater than mean

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

Ans:- Inter Quartile Range = Upper Quartile – Lower Quartile => 18 - 10 =8

Approximately= 8

Q19) Comment on the below Boxplot visualizations?



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

Ans: - Ans: By observing both the plots whisker’s level is high in boxplot 2, mean and median are equal hence 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

* 1. P(MPG>38)
  2. P(MPG<40)

c. P (20<MPG<50)

Ans: -

a) P(MPG>38) = 1-pnorm (38,34.422,9.13144) = 0.348

b) P(MPG>40) = 1-pnorm (40,34.422,9.13144) = 0.7293527

c)P(MPG>50) = 1-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

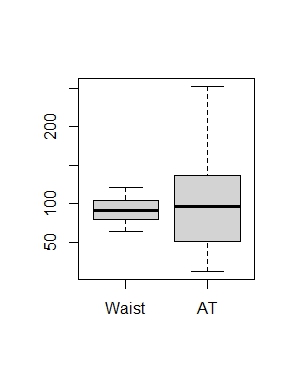
1. Check Whether the Adipose Tissue (AT) and Waist Circumference (Waist) from wc-at data set follows Normal Distribution

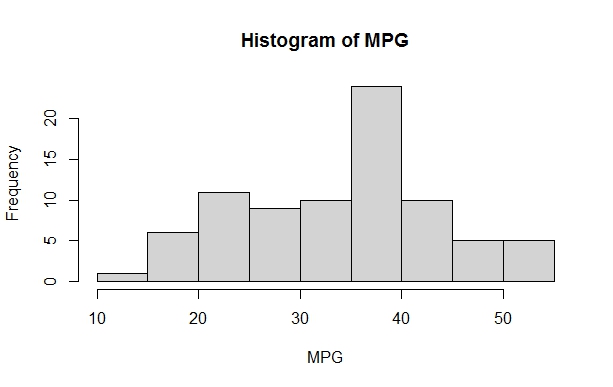
Dataset: wc-at.csv

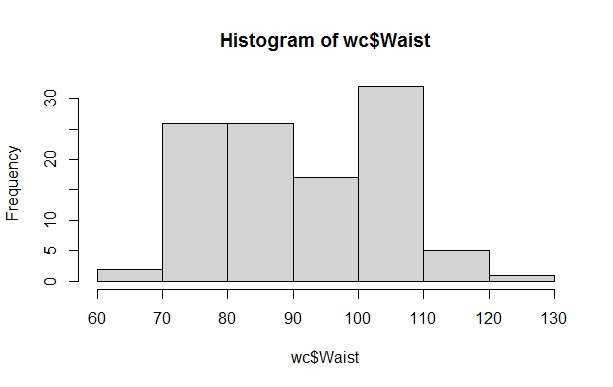
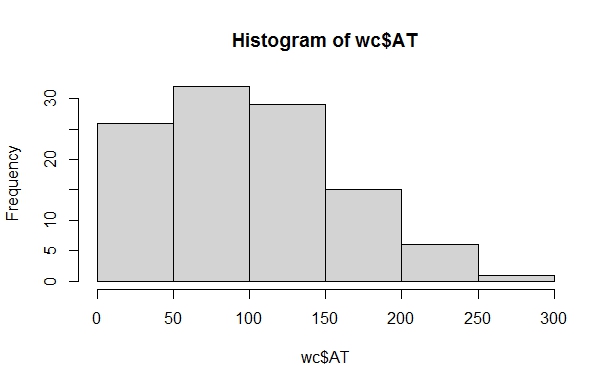
Ans:

a) The MPG of Cars is highly skewed.

Skewness = -0.174634

b) 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.