|  |  |
| --- | --- |
| 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:

P = {HHH HHT HTH THH HTT THT TTH TTT}

A = {HHT HTH THH}

P(A) = 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

Ans:

Total possible outcomes = 6^2 =36

1. favorable outcomes (sum is equal to 1) = 0 (i.e. not possible)

probability = 0/36 = 0

b) favorable outcomes (less than or equal to 4) = {1,1 1,2 1,3 3,1 2,1 2,2} = 6

probability = 6/36 =1/ 6

c)favorable outcomes (sum is divisible by 2 and 3) = {1,5 2,4 3,3 4,2 5,1 6,6} = 6

probability = 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. :

Total no. of balls = 2+3+2 = 7

S= sample space

n(s) = number of ways to draw 2 balls out of 7

n(s) = 7C2 = (7\*6) /(2\*1)

= 21

E = Event of draw 2 balls , none of the ball is blue

n(E) = number of ways to draw 2 balls out of ( 2red + 3green)

= 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 candidates for a randomly selected child is 3.09.

Explanation :

= 1\*0.015+4\*0.20+ 3\*0.65+5\*0.65+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 Point , Score , Weigh>

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

**Use Q7.csv file**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Points** | **Score** | **Weigh** |
| Mazda RX4 | 3.9 | 2.62 | 16.46 |
| Mazda RX4 Wag | 3.9 | 2.875 | 17.02 |
| Datsun 710 | 3.85 | 2.32 | 18.61 |
| Hornet 4 Drive | 3.08 | 3.215 | 19.44 |
| Hornet Sportabout | 3.15 | 3.44 | 17.02 |
| Valiant | 2.76 | 3.46 | 20.22 |
| Duster 360 | 3.21 | 3.57 | 15.84 |
| Merc 240D | 3.69 | 3.19 | 20 |
| Merc 230 | 3.92 | 3.15 | 22.9 |
| Merc 280 | 3.92 | 3.44 | 18.3 |
| Merc 280C | 3.92 | 3.44 | 18.9 |
| Merc 450SE | 3.07 | 4.07 | 17.4 |
| Merc 450SL | 3.07 | 3.73 | 17.6 |
| Merc 450SLC | 3.07 | 3.78 | 18 |
| Cadillac Fleetwood | 2.93 | 5.25 | 17.98 |
| Lincoln Continental | 3 | 5.424 | 17.82 |
| Chrysler Imperial | 3.23 | 5.345 | 17.42 |
| Fiat 128 | 4.08 | 2.2 | 19.47 |
| Honda Civic | 4.93 | 1.615 | 18.52 |
| Toyota Corolla | 4.22 | 1.835 | 19.9 |
| Toyota Corona | 3.7 | 2.465 | 20.01 |
| Dodge Challenger | 2.76 | 3.52 | 16.87 |
| AMC Javelin | 3.15 | 3.435 | 17.3 |
| Camaro Z28 | 3.73 | 3.84 | 15.41 |
| Pontiac Firebird | 3.08 | 3.845 | 17.05 |
| Fiat X1-9 | 4.08 | 1.935 | 18.9 |
| Porsche 914-2 | 4.43 | 2.14 | 16.7 |
| Lotus Europa | 3.77 | 1.513 | 16.9 |
| Ford Pantera L | 4.22 | 3.17 | 14.5 |
| Ferrari Dino | 3.62 | 2.77 | 15.5 |
| Maserati Bora | 3.54 | 3.57 | 14.6 |
| Volvo 142E | 4.11 | 2.78 | 18.6 |
|  |  |  |  |
| **Mean** | 3.5965625 | 3.21725 | 17.84875 |
| **Median** | 3.7 | 3.215 | 17.6 |
| **Mode** | 3.92 | 3.44 | 17.02 |
| **Variance** | 0.2858814 | 0.957379 | 3.193166 |
| **Standerd deviation** | 0.5346787 | 0.978457 | 1.786943 |
| **Range** | 2.17 | 3.911 | 8.4 |

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 weight for a random selected patient = 145.33

Explanation :

Expected value = sum(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)

**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(WT)

Kurtosis = 5.72(SP) , 3.87(WT)

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





**ANS:**

The Histogram and Boxplot in fig is positively skewed on right side.

i.e. mean and median of the data is greater than the 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:**

**CI = x + z (s/sqrt(n))**

CI = Confidence interval

X = Sample Mean

S = Sample Standard Deviation

n =sample size

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 (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.

Ans:

Mean = 41

Median = 40.5

Variance = 25.52

Standard deviation = 5.05

2)What can we say about the student marks?

Ans:

Repeatedly 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:

Distribution is symmetric when mean and median is equal.

Skewness of the distribution is zero.

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

Ans:

When mean is greater than median then the distribution is positively skewed.

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

Ans:

When median is greater than mean the distribution is negatively skewed.

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

Ans:

Positive value of kurtosis indicates that distribution is peaked and possesses thick tails and also indicates the distribution where more 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:

Negative kurtosis value indicates that the distribution has lighter tails than the normal distribution.

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



What can we say about the distribution of the data?

Ans:

Distribution is negatively skewed.

What is nature of skewness of the data?

Ans:

Negatively skewed .

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

Ans:

IQR = UQ-LQ

= 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.

Ans:

In this boxplot graph , Boxplot 1 is positively skewed that’s mean mean and median is greater than the mode . In boxplot 2 , it is normal distributed 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)
  3. P (20<MPG<50)

Ans:

a) P(MPG>38) = 0.3475908

1. P(MPG<40) = 0.7293527
2. P (20<MPG<50) = 0.01311818 .

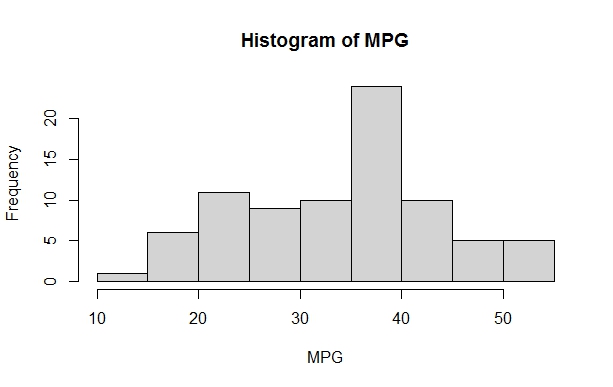
Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

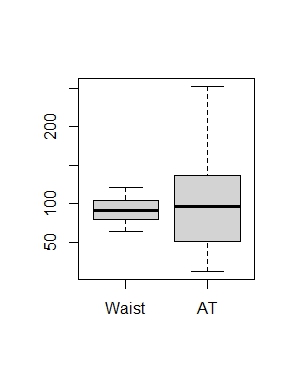
Dataset: Cars.csv

Ans:

The MPG highly Skewness of Cars is skewed = - 0.1746343

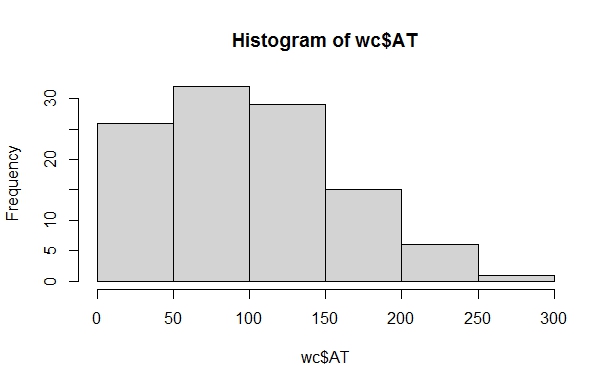


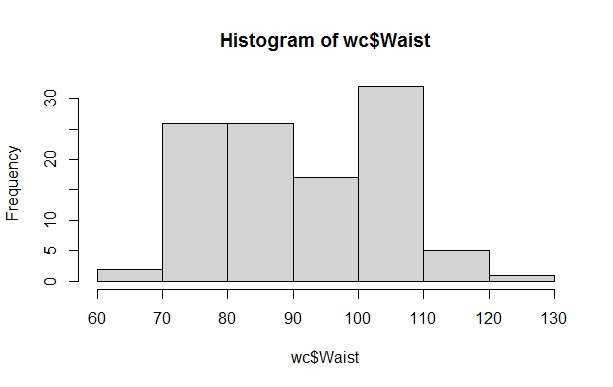
1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

 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

|  |  |
| --- | --- |
| **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 then the t-value obtained with 17 degrees of freedom and 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 .