|  |  |
| --- | --- |
| 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 | Discrete (Nominal) |
| High School Class Ranking | Discrete (Ordinal) |
| Celsius Temperature | Continuous (Interval) |
| Weight | Continuous (Ratio) |
| Hair Color | Discrete (Ratio) |
| Socioeconomic Status | Continuous (Interval) |
| Fahrenheit Temperature | Continuous (Interval) |
| Height | Continuous (Ratio) |
| Type of living accommodation | Discrete (Ordinal) |
| Level of Agreement | Discrete (Interval) |
| IQ(Intelligence Scale) | Discrete (Interval) |
| Sales Figures | Discrete(Interval) |
| Blood Group | Discrete(Nominal) |
| Time Of Day | Continuous(Interval) |
| Time on a Clock with Hands | Continuous(Interval) |
| Number of Children | Discrete (Ordinal) |
| Religious Preference | Discrete (Nominal) |
| Barometer Pressure | Continuous(Interval) |
| SAT Scores | Continuous(Ratio) |
| Years of Education | Discrete(Ordinal) |

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

Ans - Since three Coins are tossed simultaneously, the total outcomes are 8 i.e.

(HHH), (HTH),(HHT), (HTT),(TTT),(TTH),(THH),(THT)

H- Heads and T- Tails

No of outcomes with two heads and one tail are = P(HTH)+P(HHT)+(THH)

= 1/8+1/8+1/8

=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 Outcomes= 62 =36

1. No possible outcomes can have sum which is equal to 1, therefore the probability is 0/36= 0
2. (2,2), (3,1) and (1,3) outcomes are possible, therefore the probability is 3/36 = 1/12
3. (3,3), (4,2), (2,4), (1,5),(5,1),(6,6) are the possible outcomes, therefore the possibility is 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- Let the balls be named as R, G and B where R is Red, B is Blue and G is Green

P(2R,3G,2B)

Total number of Balls are= 2+3+2 =7

Number of ways the 2 balls can be drawn out of 7 is

7C2 = (7\*6)/ (2\*1) =21

Number of ways the 2 balls can be drawn where none of the balls drawn is blue is (2+3)

5C2 =(5\*4)/ (2\*1) = 10

Therefore, the Probability of drawing two balls which are not blue is = 5C2/ 7C2

= 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 is

= A+B+C+D+E+F

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

|  |  |  |  |
| --- | --- | --- | --- |
|  | 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 |

**Ans-**

|  |  |  |  |
| --- | --- | --- | --- |
| **Type** | **For Points** | **For Score** | **For Weight** |
| **Mean** | **3.5865** | **3.217** | **17.84875** |
| **Median** | **3.69** | **3.325** | **17.71** |
| **Mode** | **3.92,3.07 (Bi-modal)** | **3.44** | **17.02** |
| **Standard Deviation** | **0.5346787** | **0.9784574** | **1.786943** |
| **Variance** | **0.2858814** | **0.957379** | **3.193166** |
| **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 = Summation (probability \* Value)**

**Summation of P(x).E(x)**

**Probability for selecting each patient is =1/9**

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

**= 145.33**

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

**Cars speed and distance**

**Ans- For Q9\_a**

**Skewness according to the given data**

**Speed= -0.1139548**

**Distance= 0.7824835**

**Kurtosis according to the given data is**

**Speed=2.422853**

**Distance=3.248019**

* **From the above data the Speed has Negative Skewness and since the data is less than 3, therefore the distribution has flat and thin tails, so Negative Kurtosis.**
* **From the above data the Distance has Positive Skewness and since the data is greater than 3, therefore the distribution has peaked out and thick tails, so Positive Kurtosis.**

**For Q9\_b**

**Skewness according to the given data**

**SP= 1.581454**

**WT=-0.6033099**

**Kurtosis according to the given data is**

**SP=5.723521**

**WT=3.819466**

* **From the above data the SP has Positive Skewness and since the data than 3, therefore the distribution has Peak out at start and thick tails, so Positive Kurtosis.**
* **From the above data the WT has Negative Skewness and since the data is greater than 3, therefore the distribution has peaked out and thick tails, so Positive Kurtosis.**

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



**Ans- From the above Histogram, most of the data is peaked at the range of 50-100 and the maximum frequency is 200.**

**And the least range is from 350-400 and frequency is between 0-10.**

**The Skewness for the above is Right skewed i.e. Positive Skewed because the tail of the data is towards right.**



**Ans- The boxplot represent that the data is positive skewed. Since we have outliers on the upper side of the box it has maximum range. Median is less than mean.**

**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?

**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- **Total number of tests= 18**

1. **Mean= 738/18= 41**

**Median = 40+41/2=40.5**

**Variance=25.52941**

**Standard Deviation= 5.052664**

1. **The average score of the students is 41, the variation from the average mean is 5.052664**

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

Ans- The **Distribution of is Symmetric and has Zero skewness.**

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

Ans- **the Distribution is Positively Skewed**.

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

Ans- **The Distribution is Negatively Skewed**.

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

Ans**- Distribution is peaked and possess thick tails**

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

Ans- **Distribution is flatter and possess light tails**

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



What can we say about the distribution of the data?

Ans**- Most of the data lies above 10 i.e 50% of data, and remaining data are less.** And Above 15, 40% of data lies in it.

What is nature of skewness of the data?

Ans- **since Median is greater than Mean, Nature of Skewness is Left Skewed.**

What will be the IQR of the data (approximately)?   
Ans – **The Interquartile range is approximately =18-10= 8 (Q3-Q1).**

Q19) Comment on the below Boxplot visualizations?



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

Ans- **1) Whiskers level of Boxplot 1 is low than that of Box Plot 2.**

**` 2) Both the Boxplot have same mean and median. Hence 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

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

c. P (20<MPG<50)

Ans- **Total entries of MPG= 81**

**By using Rstudio and using subset for filtering data.**

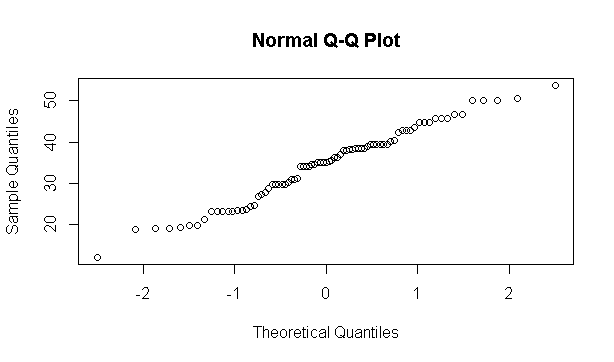
1. **Probability= 33/81**
2. **Probability= 61/81**
3. **Probability= 69/81**

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans-



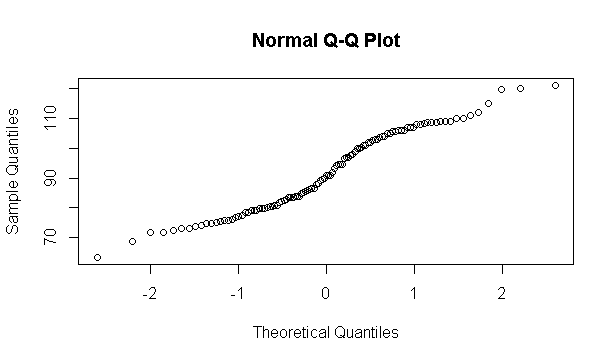
**Mean = 34.42208**

**Median= 35.15273**

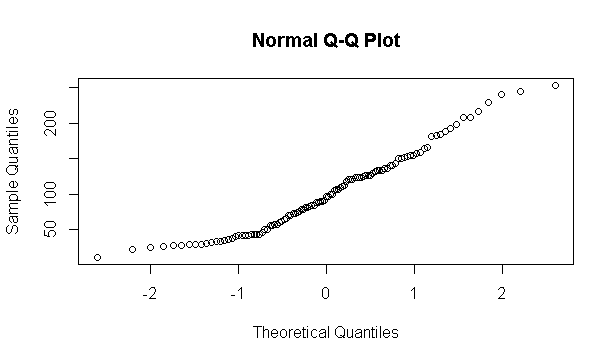
**From the Above qqplot the distribution of data is steadily increasing, therefore the distribution from qqplot and values can be called normal distributed**

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

Ans-



**From the above plot for WT, the data can be said normally Distributed, since the data is steadily increasing**.



**For the above graph for AT, the distribution can be said to be normalized**

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

Ans- **Using Qnorm in Rstudio-**

**z-score for 90% confidence = 1.645**

**z-score for 94% confidence= 1.55474**

**z-score for 60% confidence= 0.25334**

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

Ans- **by Computing Mean and Standard deviation of sample(2,3,5,6,9)=25 is**

**Mean=5**

**SD=2.738613**

**Zscore for 95%=1.96**

**Zscore for 96%=2.05**

**Zscore for 99%=2.57**

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- **Let tscore= x-population mean/sd/root(n)**

**I.e, tscore=x(260)-population mean(270)/sd(90)/root of n(18)**

**Tscore= -10/30/3\*root2**

**Tscore= -0.471**

**Therefore, Df= N-1= 18-1=17 degree**

**By using rcode= pt(tscore,df)**

**The probability of bulb having average life is = 0.321814**