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

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

A-probability of two heads=2/3=0.667(67%)

B-probability of one tail=1/3=0.333(33%)

Q4) Two Dice are rolled, find the probability that sum is

Explanation: When Two dice are rolled total number of events=6\*6=36

(1,1) (1,2) (1,3) (1,4) (1,5) (1,6)

(2,1) (2,2) (2,3) (2,4) (2,5) (2,6)

(3,1) (3,2) (3,3) (3,4) (3,5) (3,6)

(4,1) (4,2) (4,3) (4,4) (4,5) (4,6)

(5,1) (5,2) (5,3) (5,4) (5,5) (5,6)

(6,1) (6,2) (6,3) (6,4) (6,5) (6,6)

1. Equal to 1: It is not possible because here there are two dice rolled. and suppose both dice got (1,1) value. because value on dice are (1,2,3,4,5,6) .so we can say that minimum value on dice is 1. So probability that sum is not equal to 1.
2. Less than or equal to 4:(Number of interested events(1,1) (1,2) (1,3) (2,1) (2,2) (3,1)=6)/(Total number of events=6\*6=36)=6/36=0.166
3. Sum is divisible by 2 and 3:number of interested events (1,1) (1,3) (1,5) (2,2) (2,4) (2,6) (3,1) (3,3) (3,5) (4,2) (4,4) (4,6) (5,1) (5,3) (5,5) (6,2) (6,4) (6,6) =18/36=0.5

(1,2) (1,5) (2,1) (2,4) (3,3) (3,6) (4,2) (4,5) (5,1) (5,4) (6,3) (6,6) =12/36=0.33

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?

Probability=5C2 / 7C2=5\*4/7\*6=10/21=0.476, (P=0.476)

Explanation: Total number of balls in this beg is 2+3+2=7.Since 2 balls can be drawn out of 7 in 19C2 ways, Then total number of cases are 19C3

And If none of the drawn balls is blue, then all the 2 balls must be out of the red and green balls, out of 2+3=5 balls. Hence the number of interested cases for this event is 5C3.And total number of cases are 19C3.

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

Explanation: The formula for calculating the expected value for a discrete random variable X, denoted by µ, is:

µ= ∑x. p(x)

Expected Value of candies for a randomly selected child A, and child B

µ=x1.p(x1)+x2p(x2) +x3.p(x3)+x4p(x4)+x5.p(x5)+x6.p(x­6)

µ =1\*0.015+4\*0.20+3\*.65+5\*.005+6\*.01+2\*.120

µ =.015+.8+1.95+.025+.06+.24

µ =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.



**Explanation: For Points**

Mean=3.596, Median=3.695, Mode=3.07, Variance=0.285, Standard Deviation =0.534 Range= Max.value - Min.value= 4.93-2.76 = 2.17

**Explanation: For Score**

Mean= 3.22, Median= 3.325, Mode=3.44 , Variance= 0.957379, Standard Deviation= 0.9784574 , Range = Max.Value-Min.Value= 5.424-1.513 =3.911

**Explanation: For Weigh**

Mean= 17.84875 , Median= 17.71,Mode= 17.02,Variance= 3.193166,

standard Deviation= 1.786943, Range= Max.Value-Min.Value

Range = 22.9-14.5= 8.4

* Points Mean Closes to Score Mean and Weigh Mean

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?

Solution: The Expected Value of the weight of that patient

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

E(X) =1308/9 = 145.33

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

**Cars speed and distance**



**Skewness for Speed and Distance:**

Solution: Skewness for speed =-0.1139548, Skewness value is negative.so it is left skewed and data concerned right side.

Skewness for distance= 0.7824835,Skewness value is positive. So it is right skewed and data concerned left side .

**Kurtosis for Speed and Distance :** Solution: kurtosis for speed= 2.422853,kurtosis value ipositive(min.).So its called a positive kurtosis and positive kurtosis implies sharper peaked.

Kurtosis for Distance= 3.248019,Kurtosis value is positive(max.).So it is called a positive kurtosis and positive kurtosis implied sharper peaked.

**SP and weight(WT) :**

****

**Skewness for SP and WT:**

**Solution:** Skewness for SP= 1.581454, Skewness value is Positive. So it is right skewed and data concerned on left side.

Skewness for WT=-0.6033099 , Skewness value is Negative. So it is left skewed and data concerned on right side.

**Kurtosis for SP and WT:**

**Solution**: Kurtosisfor SP**=** 5.723521, kurtosis value is positive. So its called a positive kurtosis and positive kurtosis implies sharper peaked.

Kurtosis for WT= 3.819466, kurtosis value is also positive(min). So it is called positive kurtosis and positive kurtosis implies sharper peaked

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



**Explanation:**

* The most the data concerted in the 50-150 with frequency 200.
* Skewness: We can say that long tail on right side and data concentrate on left side (It is also called a Third Moment Of Business Decision)
* Kurtosis: Sharper peaked in this data and it is called a positive kurtosis (It is also called a fourth moment business decision)



**Explanation:** In box plot the box contains the middle 50% of the data points and each of the two whisker contain 25% of the data .we have outlier(Major,Minor) on upper extreme of box plot and there is less data points between Q1 and lower extreme. Positive skewness ehich means that long tail on right side and data concentrate on left side.

**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 ?

**Solution**: In this question we have not given population standard deviation, then we will use t distribution for Calculate 94%,98%,96% confidence interval .

Given Population (N)= 3,000,000 ,sample size( n)=2,000 ,Sample Mean (x̄)=200, sample standard deviation (s)=30, Degree of freedom=n-1=2000-1

1999, Standard Error=s/=30/30/44.7213=0.671

1**.For 94% Confidence interval**

α =1-0.94=0.06

(t-value for 1999 by using t-table)

t 1-α,n-1=t 0.94,1999=,

+ t 0.94,1999\*s/=200+1.881861\*0.671=200+1.26272873=201.262729

- t 0.94,1999\*s/=200- 1.881861\*0.671=200-1.26272873=198.737271

[198.737271, 201.104466]

2**.For 98% confidence interval**

α =1-0.98=0.02

(t-value for 1999 by using t-table at 98%)

t 1-α,n-1=t 0.98,1999= 2.328215

+ t 0.98,1999\*s/=200+2.328215\*0.671=200+1.56223227=201.562232

- t 0.98,1999\*s/=200-2.328215\*0.671=200-1.56223227=198.437768

[198.437768, 201.562232]

**For 96% confidence interval**

α =1-0.96=0.04

(t-value for 1999 by using t-table at 96%)

t 1-α,n-1=t 0.96,1999=2.05509

+ t 0.96,1999\*s/=200+2.05509\*0.671=200+1.37896539=201.378965

- t 0.96,1999\*s/=200-2.05509\*0.671=200-1.37896539=198.621035

[198.621035, 201.378965]

**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?

Solution: Mean=41, Median= 40.5, Variance=24.111, Standard Deviation=4.910

We say that the scores obtained by a student in tests in increasing order and most Occurring value is 41 and the smallest marks is 34 and highest marks is 56.

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

Explanation: No skewness because skewness is the measurement of asymmetry in the distribution . but when Mean, Median of data are equal so the distribution is symmetric.

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

Explanation: Positive Skewness

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

Explanation: Negative Skewness

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

Explanation:  positive kurtosis value indicates that the distribution has heavier tails and a sharper peak than the normal distribution.

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

Explanation :  negative kurtosis value indicates that the distribution has lighter tails and a flatter peak 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?

**Explanation:** Let’s assume above box plot is about age’s of the students in a school.in Box Plot Box contain middle of the 50% data of the student’s age above 10yrs.and upper whisker contain 25% data of the student’s age above 18yrs,and lower whisker contain 25% data of the student’s age less 10yrs. Upper Quartile(Q3)=18, Lower Quartile(Q1)= 10

2.What is nature of skewness of the data?

**Explanation:** Left Skewed, Data concentrate on right side. Median is greater than Mean.

3.What will be the IQR of the data (approximately)?   
**Explanation:** Upper Quartile(Q3)=18, Lower Quartile(Q1)= 10

(**Inter Quartile Range**)**IQR = Q3-Q1**

**IQR =** 18-10

**IQR =**8

Q19) Comment on the below Boxplot visualizations?



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

**Explanation**: Here we assume they are two salesmen who sold mobiles of different companies to customers in the last one year.

The Box contain 50% of the data and the two Whisker contains 25%,25% data.

Plot-1:Upper Quartile(Q3)=280.5,Lower Quartile(Q1)=256.5, IQR(inter Quartile Range)=Q3-Q1=280.5-256.5=24

Major Outlier on upper extreme =Q3+3(IQR)=352.5>then major outlier

Minor Outlier on upper extreme= Q3+1.5(IQR)=316.5

Major Outlier on lower extreme=Q1-3(IQR)=184

Minor outlier on lower extreme=Q1-1.5(IQR)=220

Plot-2: Upper Quartile(Q3)=315.5, Lower Quartile(Q1)=223.6,

IQR(inter Quartile Range)=Q3-Q1=315.5-223.6=91.9

Major Outlier on upper extreme =Q3+3(IQR)=591.2

Minor Outlier on upper extreme= Q3+1.5(IQR)=453.35

Major Outlier on lower extreme=Q1-3(IQR)=-52.1

Minor outlier on lower extreme=Q1-1.5(IQR)=85.75

1. There is no outlier for upper quartile and lower quartile.
2. Symetrical/Normal Distribution of data.
3. Both the Box plots median are same.

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)= interested events=26/total number of events=81

=26/81=0.3209

* 1. P(MPG<40)=interested events =61/total events=81

=61/81=0.75308

* 1. P (20<MPG<50)=interested events=69/total events=81

=69/81=0.85185

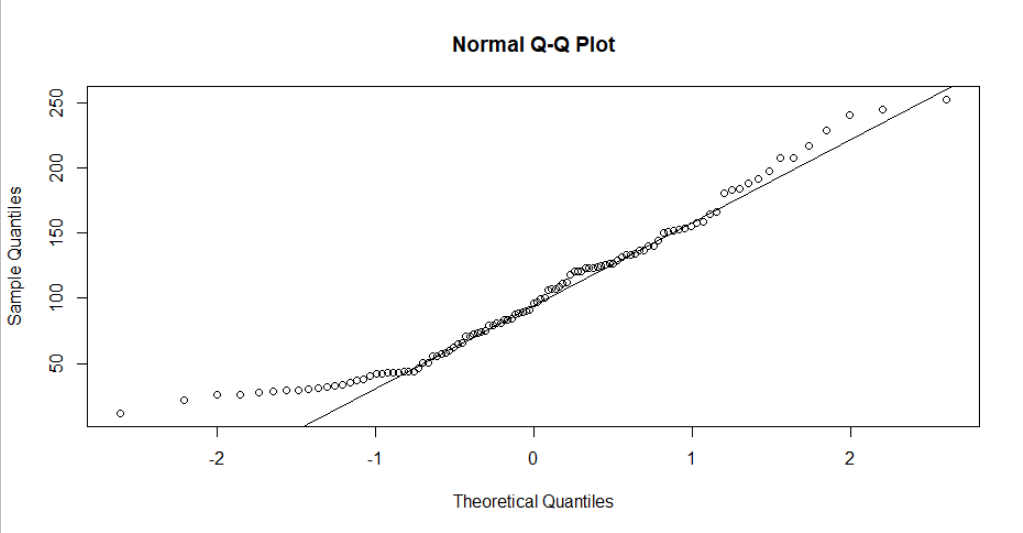
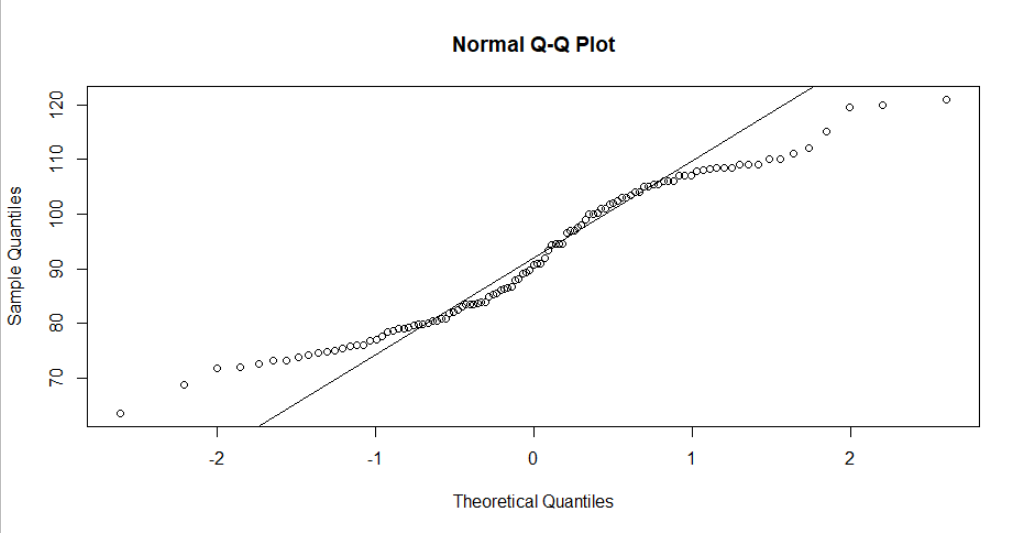
Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv



1. **Explanation**: the definition of Normal Quantile (Q-Q) plot: if all the data points lie on the line(Q-Q line) or nearby this line so it is called the particular data set is normally distributed. yes the data follows Normal distribution. Because when we apply this data on normal quantile (Q-Q) plot.so we see that all data points lie on the lines (Q-Q line) or nearby this line (Q-Q line).by the definition of Normal Quantile (Q-Q) plot.
2. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution ,Dataset: wc-at.csv



1. **Explanation:** the definition of Normal Quantile (Q-Q) plot: if all the data points lie on the line(Q-Q line) or nearby this line so it is called the particular data set is normally distributed. yes the data set of Adipose Tissue (AT) follows Normal distribution. Because when we apply this data set on normal quantile (Q-Q) plot.so we see that all data points lie on the lines (Q-Q line) or nearby this line (Q-Q line).by the definition of Normal Quantile (Q-Q) plot.And similarly the data set of waist Circumference(Waist) follows Normal distribution. Because when we apply this data set on normal quantile (Q-Q) plot.so we see that all data points lie on the lines (Q-Q line) or nearby this line (Q-Q line).by the definition of Normal Quantile (Q-Q) plot.

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

Solution23: Z scores of 90% confidence interval =1.65

Z scores of 94% confidence interval=1.89

Z scores of ,60% confidence interval=0.13

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

Solution23: t scores of 95% confidence interval for sample size of 25= 2.063899

t scores of 96% confidence interval for sample size of 25= 2.171545

t scores of 96% confidence interval for sample size of 25=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

Solution: Given, sample mean (260, Population Mean(α)=270, sample number(n)=18, standard deviation(s)=90, df 🡪 degrees of freedom= (n-1)=(18-1)=17

Find T value t = = = = -0.4714

rcode 🡪 qt(tscore,df)qt(0.4714,17)

we use this code of t value with (n-1=18-1=17) degree of freedom. probability will be=.072819

the probability that 18 randomly selected bulbs would have an average life of no more than 260 days =1-.0728=0.927 or 92.7%

Therefore, if CEO claim of 270 days average bulb lifespan is true. There is 92.7% chance that the 18 randomly selected bulbs would have an average life of no more than 260 days.