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

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

ANS**: Number with 2 heads :3**

**Total outcomes: 8**

**3/8 (or) 0.375**

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 probability**

**b) 6/36**

**c) probability divisible by 2: 6/36**

**probability divisible by 3: 6/36**

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

Statement: two balls are drawn randomly: 7c2

sample=7\*6/2\*1

=21

Actual condition: two balls are drawn that none of the balls drawn is blue

E=5C2

=5\*4/2\*1

=10

**SAMPLE /P(E)= 10/21**

**Ans: probability is 10/21= 0.47**

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**

Detailed explanation :

Expected number of candies for a randomly selected child =

1\*0.015+4\*0.20+3\*0.65+4\*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**

**ANS :**

**PROBLEM IS SOLVED USING Rstudio**

|  |
| --- |
|  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
| |  |  |  |  | | --- | --- | --- | --- | |  | **Points** | **score** | **weigh** | | **Mean** | 3.59 | 3.22 | 17.85 | | **Median** | 3.69 | 3.33 | 17.71 | | **Variance** | 0.29 | 0.96 | 3.19 | | **Standard deviation** | 0.53 | 0.98 | 1.79 | | **Range** | 2.76-4.93 | 1.593-5.424 | 14.5-22.9 | |  |
|  |  |
|  |  |
| Inference drawn   * The means is useful for spotting tendency we can compare means over time period to identify the tendencies, it is the measure of central tendency. * The median is used to divide sample data in half,it is middle score. In some of the extreme cases we use median statistics. * The mean is impacted in some cases * The median is not impacted by the extreme cases . |  |
| 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: **Sum= 108+110+123+134+135+145+167+187+199= 1308**  **Mean = 145.33**  **Expected value = 150** |  |
|  |  |
| **Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**  **Cars speed and distance**  **Use Q9\_a.csv**  **ANS:** for speed column  > skewness(speed)# left skewed  [1] -0.1139548  > kurtosis(speed)# negative kurtosis  [1] 2.422853    Inference  By observing the histogram we can say that distribution is skewed towards left so skewness is negative, distribution of mean is less than median, kurtosis value is less than 3, by evident from the histogram are distribution has broad peak and thin tails .  **For dist**  > skewness(dist)# postive , right skewed  [1] 0.7824835  > kurtosis(dist)# positive kurtosis  [1] 3.248019    By observing the histogram we can say that distribution is skewed towards right so skewness is positive, distribution of mean is greater than median, kurtosis value is morethan 3, by evident from the histogram are distribution has sharp peak and wide tails .  **SP and Weight(WT)**  **Use Q9\_b.csv**  **ANS: for sp**  > skewness(SP)# positive / right skewed  [1] 1.581454  > kurtosis(SP)# positive kurtosis  [1] 5.723521  > hist(SP)        By observing the histogram we can say that distribution is skewed towards right so skewness is positive, distribution of mean is greater than median, kurtosis value is more than 3, by evident from the histogram are distribution has sharp peak and wide tails.  For WT  > skewness (WT)# negative / left skewed  [1] -0.6033099  > kurtosis (WT)# positive kurtosis  [1] 3.819466  > hist (WT) |  |
| By observing the histogram we can say that distribution is skewed towards left so skewness is negative, distribution of mean is less than median, kurtosis value is more than 3, by evident from the histogram are distribution has sharp peak and wide tails. |  |
|  |  |
| Inference:  By observing the histogram we can say that it is the most of the data points are concerated in the range 50-100 with frequency 200.And least range of weight is 400 somewere around 0-10.And we can say that it is right skewed, positive skewness, it has long right tail, mean is greater than median. |  |
|  |  |
|  |  |
|  |  |
|  |  |
| Inference:  By observing the above boxplot we can say that the distribution has lost of outliers towards upper extreme. |  |
| 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:  **> 94%**  **X-bar=200**  **Std = 30**  **N=200**  **Interval estimate = x-bar­ Z\*std/sqrt(n)**  **200 1.88\*30/sqrt(200)**  **= 198.74 – 201.26**  **>98%**  **X-bar = 200**  **Std = 30**  **N = 200**  **Interval estimate = x-bar­ Z\*std/sqrt(n)**  **200 2.33\*30/sqrt(200)**  **= 198.44 – 201.56** |  |
|  |  |
| **>96%**  **X-bar = 200**  **Std = 30**  **N = 200**  **Interval estimate = x-bar­ Z\*std/sqrt(n)**  **200 2.05\*30/sqrt(200)**  **= 198.62 – 201.38** |  |
|  |  |
| 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**: Mean = 41**  **Median= 40.5**  **Standard deviation = 5.05**  **Variance = 25.53** |  |
| By observing the above data we can say that the mean is greater than median ,its is right skewed , no outliers are present. |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

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

ANS : **Symmetrical , skewness 0**

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: **sharp peak thick tails.**

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

ANS **: broad peak thin tails.**

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

ANS :



What can we say about the distribution of the data?

What is nature of skewness of the data?

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

ANS **: let’s assume above box plot is about age’s of the children**

**50% of the people are above 10 years old and remaining are less**

**And students who’s age is above 15 are approx. 40%**

**It is not a normal distribution.**

1. **Left skewed , median is greater than mean**
2. **IQR=8**  
   Q19) Comment on the below Boxplot visualizations?



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

ANS: **by observing the above box plot whisker’s level is high in box plot 2 , mean and median are same hence distribution is symmetrical , hence both are normally distributed.**

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)**

**> mean(MPG)**

**[1] 34.42208**

**> sd(MPG)**

**[1] 9.131445**

**> pnorm(38,34.42,9.13)**

**[1] 0.652513**

**> 1-0.65**

**[1] 0.35**

**B) P(MPG<40)**

**pnorm(40,34.42,9.13)**

**[1] 0.7294571**

**c) P (20<MPG<50)**

**pnorm(50,34.42,9.13)-pnorm(20,34.42,9.13)**

**[1] 0.8989178**

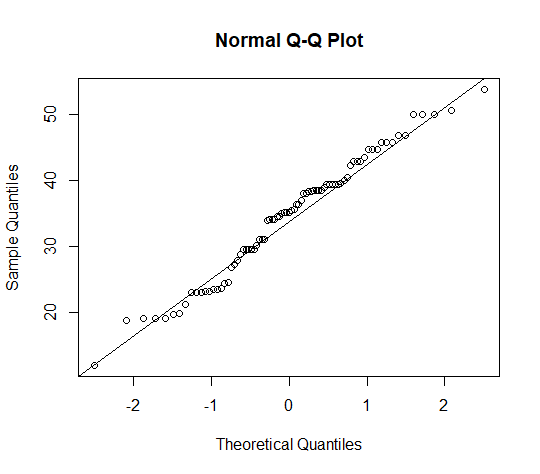
Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

ANS: qqnorm(MPG)

> qqline(MPG)



Inference : it follows normal distribution

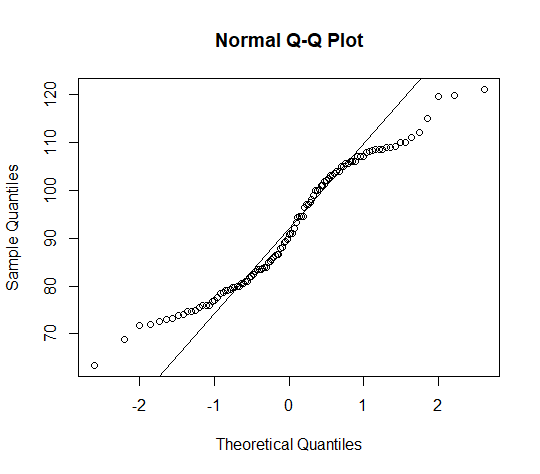
1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

ANS: attach(wc\_at\_1\_)

> qqnorm(Waist)

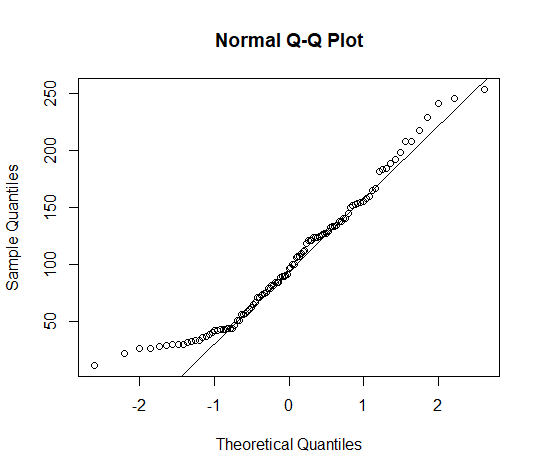
> qqline(Waist)



Inference : it follows normal distribution

qqnorm(AT)

> qqline(AT)



Inference : it follows normal distribution

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

ANS: qnorm(0.95)

[1] 1.644854

> qnorm(0.97)

[1] 1.880794

> qnorm(0.80)

[1] 0.8416212

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

ANS: qt(0.975,24)

[1] 2.063899

> qt(0.98,24)

[1] 2.171545

> qt(0.995,24)

[1] 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

Hint:

rcode 🡪 pt(tscore,df)

df 🡪 degrees of freedom

ANS: I am going to assume this as a normal distribution

You want to compare the same mean to the reported mean



In this case you are comparing =260 , s=90, =270



(n=18) , we use the T distribution , to calculate the t-statistics use



Therefore the probability is 0.3215076

Code written in r

p <- pt((260-270)/(90/sqrt(18)),17)

n<- 18

pt((260-270)/(90/sqrt(18)),17)

output received :

> p <- pt((260-270)/(90/sqrt(18)),17)

> n<- 18

> pt((260-270)/(90/sqrt(18)),17)

[1] 0.3216725