|  |  |
| --- | --- |
| %Activity | Data Type |
| Number of beatings from Wife | Discrete data |
| Results of rolling a dice | Discrete |
| Weight of a person | Continuous |
| Weight of Gold | Continous |
| Distance between two places | Continous |
| Length of a leaf | Continous |
| Dog's weight | Continous |
| 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 |
| Celsius Temperature | Interval |
| Weight | Interval |
| Hair Color | Nominal |
| Socioeconomic Status | Ordinal data |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation | Ordinal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ratio |
| Sales Figures | Ratio |
| Blood Group | Nominal data |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Nominal |
| 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**) total no of events={hhh,hht,htt,ttt,tth,thh,hth,tht}

Interested events=8

n=3,x=2,p(x)=0.5 p(t)=0.5

P(x)=n!/x!(n-x)!\*p^x(1-p)n-x

P(2+1)=3!/2!\*1!\*(0.5)^2\*(o.5)^3-2

P(3)=3!/2!\*1!\*(1/4)\*(1/2)

**Ans=3/8**

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

Equal to 1=0%probaility

Less than or equal to 4=6/36=1/6

Sum is divisible by 2 and 3

**Ans=total no of outcomes when 2 dice are rolled=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)**

**(c)sum is divisible by 2 and 3**

{2,3,4,5,6,7} {3,4,5,6,7,8} {4,5,6,7,8 9,} {5,6,7,8,9,10} {6,7,8,9,10,11}, {7,8,9,10,11,12}

=6/36=1/36s

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?

a)7c2=7\*6/2\*1=42/2=21

5c2=5\*4/2\*1=20/2=10

P€=n(e)/n(s)=5c2/7c2

**Ans=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)**

|  |  |  |  |
| --- | --- | --- | --- |
| CHILD | Candies count | Probability | Expected Value |
| A | 1 | 0.015 | 0.015 |
| B | 4 | 0.2 | 0.8 |
| C | 3 | 0.65 | 1.95 |
| D | 5 | 0.005 | 0.025 |
| E | 6 | 0.01 | 0.06 |
| F | 2 | 0.12 | 0.24 |
|  |  | 1 | **3.09** |

**Child A – probability of having 1 candy = 0.015.**

**Child B – probability of having 4 candies = 0.20**

**Expected value=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.

View(asg7)

> **mean(asg7 $Points)**

**[1] 3.596563**

> mean(asg7 $Score)

[1] 3.21725

> mean(asg7 $Weigh)

[1] 17.84875

> median(asg7 $Points)

**[1] 3.695**

> median(asg7 $Score)

[1] 3.325

> median(asg7 $Weigh)

[1] 17.71

**> mode(asg7 $Points)**

**[1] "numeric"**

**> var(asg7 $Points)**

**[1] 0.2858814**

**> var(asg7 $Score)**

**[1] 0.957379**

> var(asg7 $Weigh)

[1] 3.193166

**> range(asg7 $Points)**

**[1] 2.76 4.93**

**> range(asg7 $Score)**

**[1] 1.513 5.424**

> range(asg7 $Weigh)

[1] 14.5 22.9

**> sqrt(3.193166)**

**[1] 1.786943**

> sqrt(0.957379)

[1] 0.9784575

> sqrt(2858814)

[1] 1690.803

points

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Type | Points | Score | Weigh |  |
| Mean | 3.516 | 3.217 | 17.8487 |  |
| median | 3.625 | 3.325 | 17.71 |  |
| Variance | 0.2858814 | 0.957379 | 3.193166 |  |
| Standard deviation | 1690.803 | 0.9784575 | 1.786943 |  |
| Range | 2.76 4.93 | 1.513 5.424 | 14.5 22.9 |  |

Q8) Calculate Expected Value for the problem below

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)sigma(x/n)=

**108+110+123+134+135+145+167+187+199/9=145.33**

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

**Cars speed and distance**

**Cc**

**Ans)**

> library("moments")

> skewness(asg9 $speed)

[1] -0.1139548

> skewness(asg 9 $dist)

[1] 0.7824835

> kurtosis(asg9 $speed)

[1] 2.422853

> kurtosis(asg9 $dist)

[1] 3.248019

**SP and Weight(WT)**

> library("moments")

> skewness(asg9b$speed)

[1] 1.581454

> skewness(asg9b$weight)

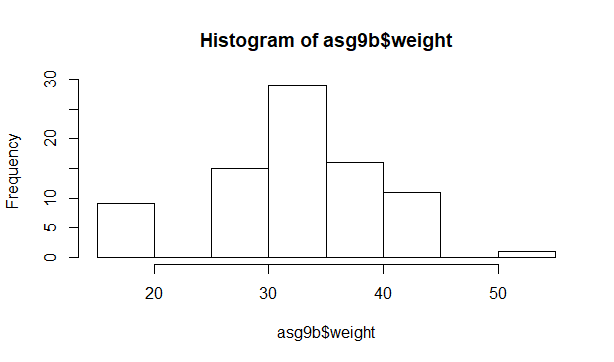
[1] -0.6033099

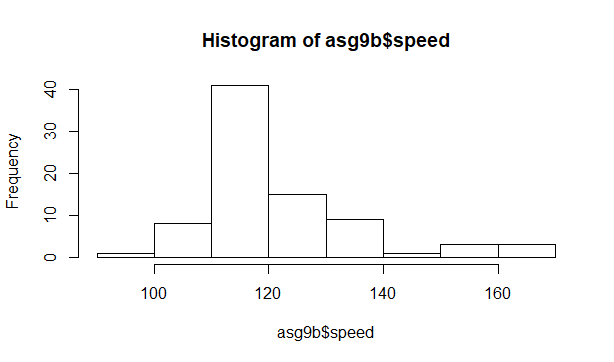
> kurtosis(asg9b$speed)

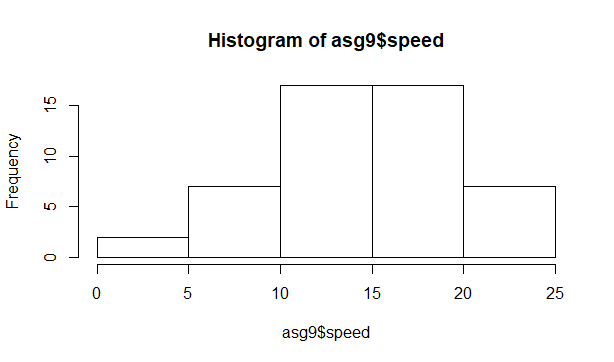
[1] 5.723521

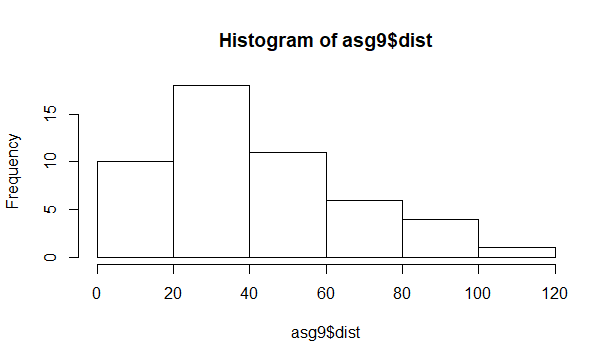
> kurtosis(asg9b$weight)

[1] 3.819466

****

****

****

****

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



**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%=**2000+1.89(30/sqrt(3,000,000)=**198.74-201.26**

**-value of -94%=**2000-1.89(30/sqrt(3,000000)=1999.9672

**98%=**2000=0.99(30)/sqrt(3000000)=20000.040

**-98%=**2000-2.33(0.01732.05)=**198.44-201.56**

**96%=**2000+2.06(0.01732.05)=200003568

**-96%=**2000-2.06(0.01732.05)=**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**

Find mean,median,variance,standard deviation.

What can we say about the student marks?

Ans)

library(readxl)

**> mean(asg12$scores)**

**[1] 41**

**> median(asg12$scores)**

**[1] 40.5**

**> var(asg12$scores)**

**[1] 25.52941**

**> sd(asg12$scores)**

**[1] 5.052664**

**Average marks scored by students=41**

**Mean is greater than median it is right skewed**

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

**Ans) no skewness,symmetric,uniform**

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 peakness and less variation**

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

**Ans) less peakness and broad peak,more varition**

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



What can we say about the distribution of the data?

It is not normal distribution

What is nature of skewness of the data?

Left skewed

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

> library(moments)

> hist(asg12&value)

Error in hist(asg12 & value) : object 'value' not found

> hist(asg12$value)

> skewness(asg12$value)

[1] 0

> boxplot(asg12$value)

> IQR(asg12$value)

[1] 8

Q19) Comment on the below Boxplot visualizations?

Ans) median**value lies in between at 260**

**2)there are no outliers in both of the boxplots**

**3)boxplots are not skewed either in positive or negative direction**

**Scale is starting from less than 200 to greater than 325**



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

Q 20) Calculate probability from the given dataset for the below cases

Data \_set: Cars.csv

Calculate the probability of MPG ofCars for the below cases.

MPG<- Cars$MPG

P(MPG>38)

P(MPG<40)

P (20<MPG<50)

Ans) > library("readr")

> mean(Cars$MPG)

[1] 34.42208

> sd(Cars$MPG)

[1] 9.131445

> pnorm(38,34.4,9.131445)

[1] 0.6532988

> pnorm(40,34.4,9.131)

[1] 0.7301595

> pnorm(50,34.4,9.131)

[1] 0.956225

> pnorm(20,34.4,9.131)

[1] 0.05739256

> pnorm(0.956225)-pnorm(0.05739256)

[1] 0.307637

> pnorm(50,34.4,9.131)-pnorm(20,34.4,9.131)

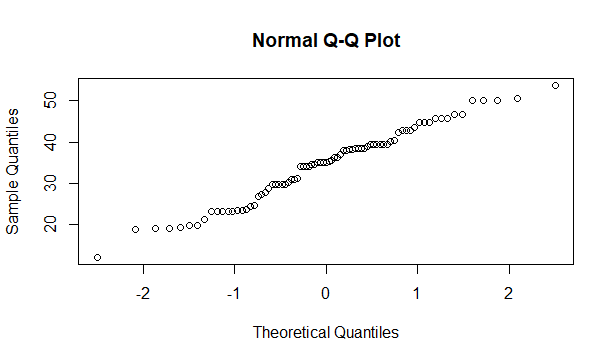
[1] **0.8988325**

Q 21) Check whether the data follows normal distribution

Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

shapiro.test(Cars$MP G)



Shapiro-Wilk normality test

data: Cars$MPG

W = 0.97797, p-value = 0.1764

P value is greater than 0.05 data is normal.

Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) fromwc-at data set follows Normal Distribution

Dataset: wc-at.csv

shapiro.test(wc\_at\_3\_$Waist)

Shapiro-Wilk normality test

data: wc\_at\_3\_$Waist

W = 0.95586, p-value = 0.00117

P value less than 0.05 data is non-normal

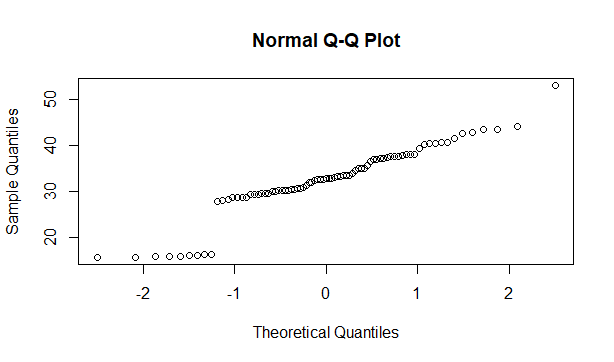
shapiro.test(wc\_at\_3\_$AT)

Shapiro-Wilk normality test

data: wc\_at\_3\_$AT

W = 0.95234, p-value = 0.000654

P value less than 0.05 data is non –normal



Q 22) Calculate the Z scoresof 90% confidence interval,94% confidence interval,24 l, 60% confidence interval

**Ans)90%=qnorm(0.95)=1.644854**

**94%=qnorm(0.97)=1.880794**

**60%=qnorm(0.8)=0.8416212**

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

Ans)

95%=> **qt(0.975,24)**

**[1] 2.063899**

96%> qt(96,24)

> qt(0.98,24)

**[1] 2.171545**

99%> **qt(0.995,24)**

**[1] 2.79694**

**95%=2.063899**

**96%=2.171545**

**99%=2.79694**

Q 24**)**A Government companyclaims 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) mew=270,x=260,sd=90,n=18,df=n-1=18-1=17**

**t=260-270/90/sqrt(18)**

**-10/90sqrt94.2426)**

**=-10/21.23**

**Ans=-0.471**

**Pt(-0.47,17)=0.32=32%**

=