|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Integer |
| Results of rolling a dice | Integer |
| Weight of a person | Float |
| Weight of Gold | Float |
| Distance between two places | Float |
| Length of a leaf | Float |
| Dog's weight | Float |
| Blue Color | Character |
| Number of kids | Integer |
| Number of tickets in Indian railways | Integer |
| Number of times married | Integer |
| Gender (Male or Female) | Character |

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

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

ANS=

Occurance=(HHH,HHT,HTH,THH,TTH,TTH,THT,TTT)=8

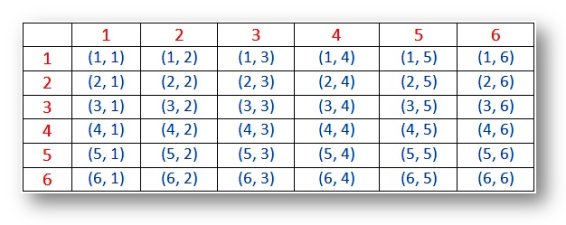
Occurances of 2H & 1T=3

Probability=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



1. Ans= 0
2. Ans=P(Sum<=4)=6/36
3. Ans=P(Sum/2&4)=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?

Total number of balls  
= (2 + 3 + 2)  
= 7  
Let S be the sample space  
Then, n(S) = Number of ways of drawing 2 balls out of 7

n(S)=7C2n(S)=(7×6)(2×1)n(S)=21n(S)=7C2n(S)=(7×6)(2×1)n(S)=21

Let E = Event of 2 balls, none of which is blue  
∴ n(E) = Number of ways of drawing 2 balls out of (2 + 3) balls

n(E)=5C2

n(E)=(5×4)(2×1)

n(E)=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 candies for a randomly selected child  = 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-**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Points** | **Score** | **Weigh** |
| **count** | 32 | 32 | 32 |
| **mean** | 3.59656 | 3.21725 | 17.8488 |
| **std** | 0.53468 | 0.97846 | 1.78694 |
| **min** | 2.76 | 1.513 | 14.5 |
| **25%** | 3.08 | 2.58125 | 16.8925 |
| **50%** | 3.695 | 3.325 | 17.71 |
| **75%** | 3.92 | 3.61 | 18.9 |
| **max** | 4.93 | 5.424 | 22.9 |

Q8) Calculate Expected Value for the problem below

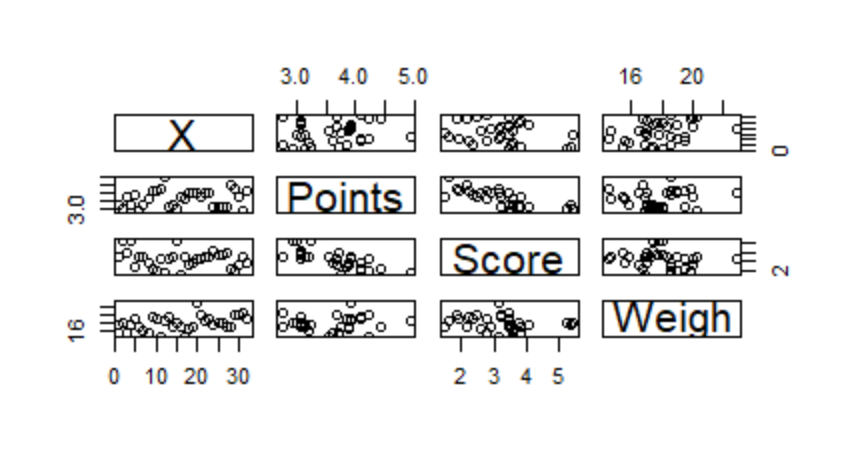
1. The weights (X) of patients at a clinic (in pounds), are

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-

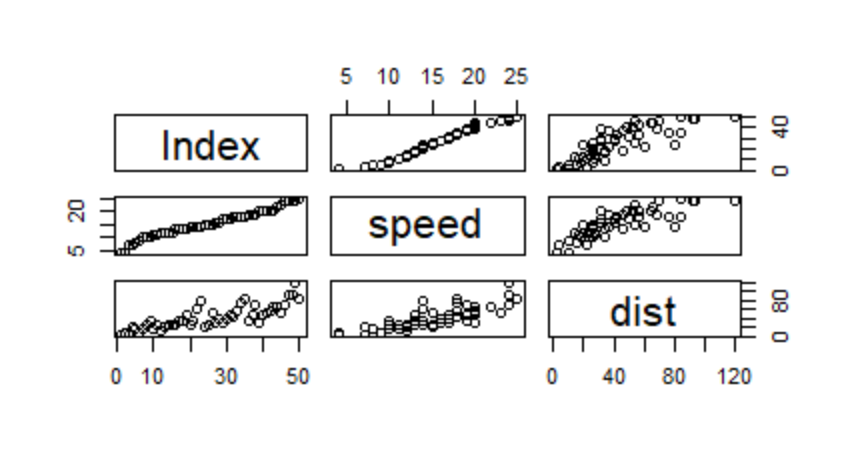
The mean is:Expected Value=(108 + 110 + 123 + 134 + 135 + 145 + 167 + 187 + 199)/9 = 145.333.It is not mandatory for the expected value to be present in sample space. As calculated above Expected Value of 145.33 is not presentin our sample space



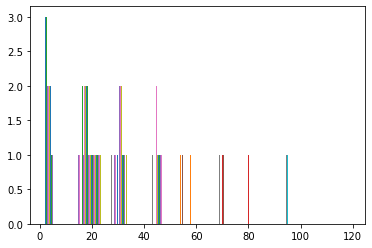
**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

**Cars speed and distance**

**Use Q9\_a.csv**



**SP and Weight(WT)**

****

**Use Q9\_b.csv**

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



Ans- from the above histogram one can predict that this is positively skewed.



ANS- from above box plot we can say that, it is having outliers.

**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?

1.Average weight of an adult male in Mexico: at 95% confidence interval[141.2011 258.7989]

2. Average weight of an adult male in Mexico: at 94% confidence interval [143.5762 256.4238]

3. Average weight of an adult male in Mexico: at 98% confidence interval [130.2096 269.7904]

4. Average weight of an adult male in Mexico: at 96% confidence interval [138.3875 261.6125]

**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.

ASN- Mean=41,median=40.5,Variance=25.5294,SD=4.9103

1. What can we say about the student marks?

kurtosis value indicates

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

Ans- If the distribution is symmetric, then the mean is equal to the median, and the distribution has zero skewness.

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

Ans-  If the mean is greater than the mode, the distribution is positively skewed.

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

Ans- If the mean is less than the median, the distribution is negatively skewed.

Q16) What does positive for a data ?

Ans- Kurtosis is a measure of whether the data are heavy-tailed or light-tailed relative to a normal distribution. That is, data sets with high kurtosis tend to have heavy tails, or outliers. Data sets with low kurtosis tend to have light tails, or lack of outliers. A uniform distribution would be the extreme case.

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

Ans- A negative kurtosis means that your distribution is flatter than a normal curve with the same mean and standard deviation. The easiest way to visualise this is to plot a histogram with a fitted normal curve.

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



What can we say about the distribution of the data?

Ans- Mean is in between 14to 16.

What is nature of skewness of the data?

Ans- Data is negatively skewed.

What will be the IQR of the data (approximately)?   
  
Ans- The IQR tells how spread out the "middle" values are; it can also be used to tell when some of the other values are "too far" from the central value. These "too far away" points are called "outliers", because they "lie outside" the range in which we expect them.

Q19) Comment on the below Boxplot visualizations?

Ans- Median of both the boxplots are inline with no outliers present.



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

Ans- The box plot shape will show if a statistical data set is normally distributed or skewed. When the median is in the middle of the box, and the whiskers are about the same on both sides of the box, then the distribution is symmetric.

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)

Ans= 0.6524098

* 1. P(MPG<40)

Ans= 0.7293536

* 1. P (20<MPG<50)

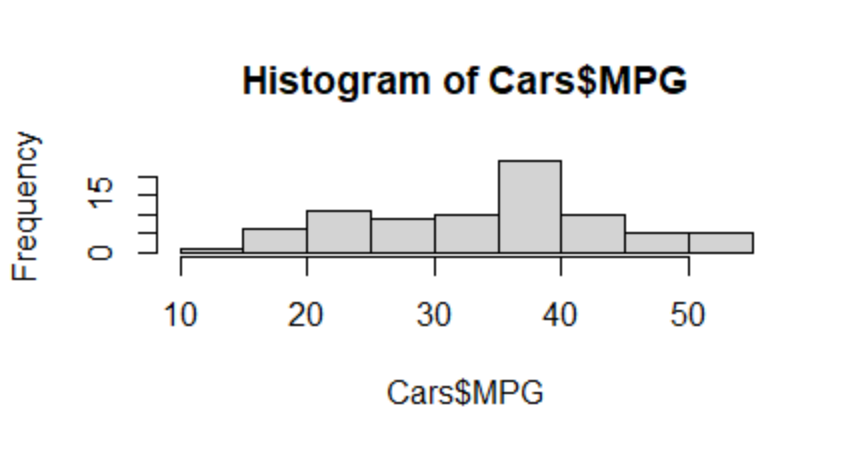
Ans=

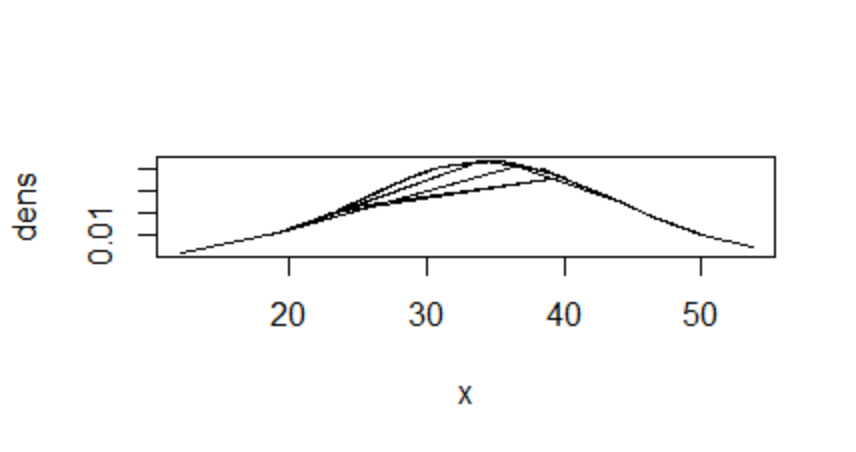
Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans=

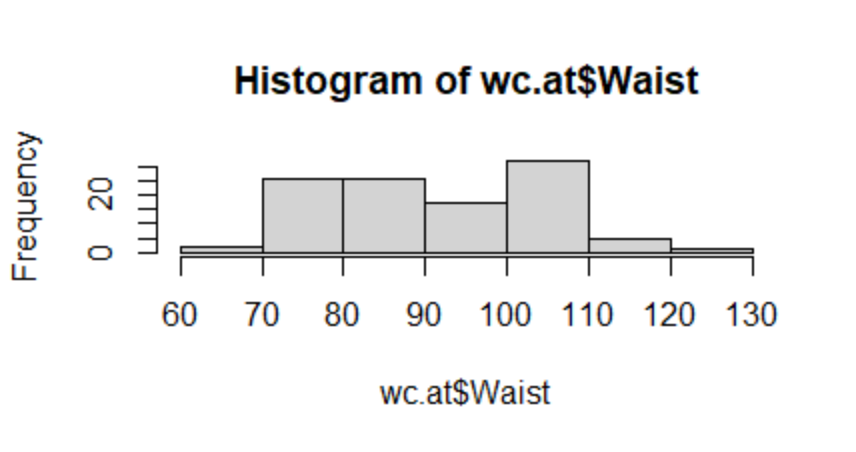




From the above histogram and density curve it is seen that the data is normally distributed.

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

Dataset: wc-at.csv



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

Z-score :

[[-1.27080988 -1.3356523 ]

[-1.43010719 -1.33267148]

[-0.74846287 -1.03967467]

[-0.58916555 -1.03616783]

[-1.27821906 -1.26341132]

[-1.48567602 -1.40649055]

[-0.81514547 -1.27673732]

[-0.62991603 -1.20835387]

[-2.10434234 -1.58604096]

[-1.38565213 -1.22167987]

[-1.48197144 -1.29006333]

[-1.25228694 -1.01758155]

[-1.39306131 -1.1166499 ]

[-0.95591984 -1.04177878]

[-1.10410339 -1.24377299]

[-1.70795135 -0.80857365]

[-1.18189975 -1.01898429]

[-1.31526494 -1.20081415]

[-1.34119706 -1.02652401]

[-1.18560434 -1.27270445]

[-1.11521715 -1.14487999]

[-0.81514547 -1.08088009]

[-0.88923724 -1.16539503]

[-0.20018374 -0.73300117]

[-0.73364451 -0.9828638 ]

[ 0.00727323 -0.55222336]

[-0.39282235 -0.32340179]

[-0.84478218 -0.30849771]

[-0.43727742 -0.40335784]

[-0.69659863 -0.6512917 ]

[-0.62250685 -0.51434945]

[-0.28168469 -0.22065127]

[-0.0816369 -0.40248112]

[-0.18536538 -0.32164837]

[ 0.74819097 0.44021348]

[ 0.19250267 0.33500816]

[-0.06681855 0.08952908]

[ 0.82228275 0.47528192]

[-0.88182806 -0.48874949]

[-0.95591984 -0.81383392]

[-0.62250685 -0.50435494]

[-1.17819516 -0.90115434]

[-0.84478218 -0.89449133]

[-0.40023153 0.66815834]

[-0.65955274 -0.09387886]

[ 1.12605902 0.2824055 ]

[ 0.17768431 0.08952908]

[ 0.19250267 0.3700766 ]

[-0.9040556 -0.63077667]

[-0.93369231 -0.36127571]

[-0.15572867 0.15966596]

[-0.60027932 -0.19575267]

[-0.49655084 0.5454188 ]

[-1.21524105 -1.05194863]

[-1.0003749 -1.05528013]

[-0.98555655 -0.76684221]

[-0.30391222 -0.22871701]

[-0.41504988 0.93117163]

[-0.4743233 -0.5457357 ]

[-0.6076885 -0.47016321]

[-1.05964832 -0.7863052 ]

[-0.51877837 -0.0379447 ]

[-0.89664642 -1.29637565]

[ 1.21496915 0.3700766 ]

[ 2.0522062 -0.20136362]

[ 2.07443373 0.07199486]

[ 0.34068621 0.73829521]

[ 1.00751218 0.33500816]

[ 0.9704663 -0.08353367]

[ 1.11864984 1.12404805]

[ 1.11864984 -0.24379644]

[ 0.6740992 0.91363741]

[ 0.3777321 -0.03321046]

[ 0.60000742 0.3700766 ]

[ 1.19274162 2.01829326]

[ 0.60000742 0.66815834]

[ 0.82228275 0.12459752]

[ 0.89637452 0.44021348]

[ 1.04455807 0.17720018]

[ 1.26683339 1.57993776]

[ 0.85932863 0.52788458]

[ 1.34092517 0.42267926]

[ 1.34092517 0.89610319]

[ 1.48910872 0.98377429]

[ 1.22978751 1.42212979]

[ 0.89637452 1.43966401]

[ 1.41501694 0.33500816]

[ 1.22978751 1.00130851]

[ 2.15593469 2.50925141]

[ 1.26683339 0.61555568]

[ 0.41477799 1.10651383]

[ 1.00751218 0.87856897]

[ 0.45182388 1.38706135]

[ 0.19250267 -0.36723734]

[ 0.3777321 0.61555568]

[ 0.9704663 0.40514504]

[ 1.04455807 2.43911453]

[ 0.52591565 0.56295302]

[-0.06681855 0.84350053]

[ 0.78523686 1.68514308]

[ 1.04455807 0.86103475]

[ 1.27424257 2.2287039 ]

[ 1.71138404 2.64952517]

[ 0.6740992 1.50980089]

[ 0.6074166 0.38761082]

[ 0.10359254 -0.69600396]

[ 0.73337262 0.5454188 ]

[ 1.18533244 1.86048528]

[ 1.22978751 1.86048528]]

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

99% Confidence interval t= 2.7969

96% Confidence interval t= 2.1715

95% Confidence interval t= 2.0638

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-t = - 0.471

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 than the t-value obtained With 17 degrees of freedom and a 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.