|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Discrete |
| Weight of a person | Continues |
| Weight of Gold | Continues |
| Distance between two places | Continues |
| Length of a leaf | Continues |
| Dog's weight | Continues |
| Blue Color(wave length) | Continues |
| 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: data which is represented in the form of whole number is discrete .(can not represent in form of fractions )

Data which can be represented in the form of fractions is continues

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

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

number of events output E = 2\*2\*2= 8

HHH

HHT

HTH

HTT

THH

THT

TTH

TTT

So , {HHT,HTH,THH}

The probablility of getting 2 heads and 1 tail = 3

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 :** b

Possibilities when 2 dies are rolled =6\*6= 36

Possibilities of getting 4 = (1,3)(2,2)(3,1) = (4)(4)(4)

**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 no of balls =2+3+2 = 7

Then , s be sample

n(S)=7C2

n(S)= (7×6)/ (2×1)

n(S)=21

let E = event of 2 balls , non of which is blue

n(E) = number of ways to draw 2 balls from (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 : 3.09

1 \* 0.015  + 4\*0.20  + 3 \*0.65  + 5\*0.005  + 6 \*0.01  + 2 \* 0.12

= 0.015 + 0.8  + 1.95 + 0.025 + 0.06 + 0.24

=       3.090

**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:**

view(cars)

mean(cars$points) = 3.36563

median(cars$points)=3.695

mfv(cars$points)=3.07 3.92v

mean(cars$Score)=3.21725

median(cars$Score)=3.325

mfv(cars$Score)=3.44

mean(cars$weight)=17.84875

median(cars$weight)=17.71

mfv(cars$weight)=17.02 18.90

#mearures of dispersion

var(cars$Points)=0.2858814

sd(cars$Points)=0.5346787

range(cars$Points)=02.76 4.93

var(cars$Score)= 0.957379

sd(cars$Score)= 0.9784574

range(cars$Score)= 1.513 5.424

var(cars$Weigh)= 3.193166

sd(cars$Weigh)= 1.786943

range(cars$Weigh)= 14.5 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:**

Expected Value  =  ∑ ( probability  \* Value )

 ∑ P(x).E(x)

there are 9 patients

Probability of selecting each patient = 1/9

Ex   : 108, 110, 123, 134, 135, 145, 167, 187, 199

P(x)  : 1/9  1/9   1/9  1/9   1/9   1/9   1/9   1/9  1/9

Expected Value  =  (1/9)(108) + (1/9)110  + (1/9)123 + (1/9)134 + (1/9)135 + (1/9)145 + (1/9(167) + (1/9)187 + (1/9)199

= (1/9) ( 108 + 110 + 123 + 134 + 135 + 145 + 167 + 187 + 199)

= (1/9)  (  1308)

= 145.33

Expected Value of the Weight of that patient = 145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans :**

**Skewness(velocity$speed)=-0.1139548**

**Skewness(velocity$dist)= 0.7824835**

speed data is negatively skewed and dist data is positive

**kurtosis(velocity$speed)**

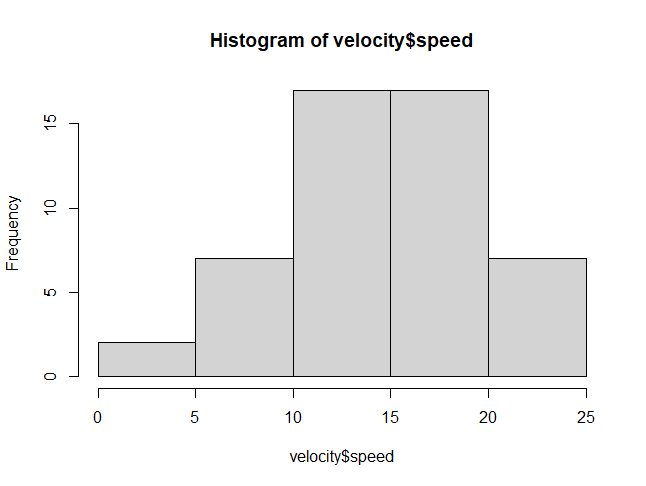
**[1] 2.422853**

**> kurtosis(velocity$dist)**

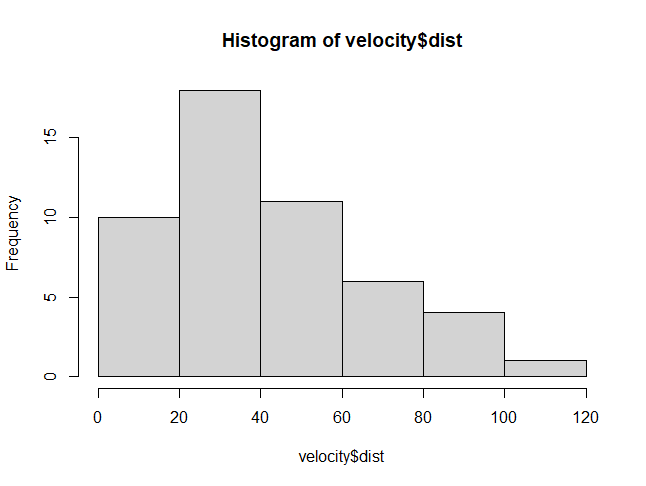
**[1] 3.248019**

Kurtosis Inference: 1. Speed distribution is platykurtic (negative kurtosis i.e. flatter than normal distribution) 2. Distance distributin is leptokurtic (positive kurtosis i.e. peaked than noramal distribution)

**Hist(velocity$speed)**



Hist(Velocity$dist)



**SP and Weight(WT)**

**Use Q9\_b.csv**

> skewness(data1$SP)

[1] 1.581454

> skewness(data1$WT)

[1] -0.6033099

> skewness(data1)

X SP WT

1. 1.5814537 -0.6033099

SP distribution is Right skewed (Positive skewness) 2. WT distribution is Left skewed (Negative skewness)

> #mesures of kurtosis

> kurtosis(data1)

X SP WT

1.799634 5.723521 3.819466

> kurtosis(data1$SP)

[1] 5.723521

> kurtosis(data1$WT)

[1] 3.819466

Kurtosis Inference: Both the SP and WT distributions are leptokurtic (have positive kurtosis i.e. Peaked than normal distribution)

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



Ans : by seeing histogram graph we can it is right skewed,because histogram tells the shape of plot. And the main purpose of Box-plot is finding the outilers.

**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:

sample weight=200 pounds

sample(s) = 30

n=2000

population (x)=3,000,000

standard error= s/sqrt(n)= 30/sqrt(2000) = 30/44.7213= 0.67082

1. 94%: Alpha=1-(94/100)=1-0.94=0.06

Critical probability = 1-alpha/2=1-(0.06/2)=1-0.03=0.97

Digrees of freedom = n-1=2000-1=1999

Find t-distribution df= 1999 and cumulative probability= 0.97

For 94% z and t distribution = 2003.98748 & 1996.01252

for 98% z and t distribution = 2004.933446 & 1995.006554

for 96% z and t distribution = 2004.354413 & 1995.645587

**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**: a <- c(34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56)

> median(a)

[1] 40.5

> mean(a)

[1] 41

> var(a)

[1] 25.52941

> sd(a)

[1] 5.052664

> range(a)

[1] 34 56

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

Ans :the distribution is [symmetric](https://en.wikipedia.org/wiki/Symmetric_probability_distribution), the mean is equal to the median, and the distribution has zero skewness

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

Ans: It is positive/right skew, means the mean is greater than the median,

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

Ans : negative/left skew means the mean is less than (to the left of) the median.

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

Ans : Positive values of kurtosis indicate that a distribution is peaked and possess thick tails.

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

Ans: Positive values of kurtosis indicate that a distribution is peaked and possess thick tails.

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



**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:**

A) Let’s assume above box plot is about ages of the employees in a company . 50% of the people are above 30 yrs. old and remaining are less.And students whose age is above 35 are approx. 40%

.What is nature of skewness of the data :

Left skewed, median is greater than mean.

the IQR of the data Approximately= -8

Q19) Comment on the below Boxplot visualizations?



**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 of Cars for the below cases.

MPG <- Cars$MPG

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

c. P (20<MPG<50)

Ans :

Rcode:MPG <-c (Cars$MPG)

MPGsample(MPG)

a=subset(MPG,MPG>38)

b=subset(MPG,MPG<40)

c=subset(MPG,MPG>20 & MPG <50)

**Q 21) Check whether the data follows normal distribution**

1. **Check whether the MPG of Cars follows Normal Distribution**

**Dataset: Cars.csv**

**Ans :**

We can interpret that the data of MPG of Cars follows the normal distribution by:

1.Conducting shapiro test (w=0,97797; p value =0,1764)

2.Evaluating kurtosis value which is -0,7054604

3.Finding of mean value (34,42208) which is not so far difference from median value (35, 15273).

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

**Dataset: wc-at.csv**

**Ans :**

We can interpret that the data of Weight of WC\_AT follows the normal distribution by:

1.Conducting shapiro test (w=0,95586; p value =0,00117)

2.Evaluating kurtosis value which is -1,141846.

3.Finding of mean value (91.902) which is not so far difference from median value (90.8).

We can interpret that the data of AT of WC\_AT follows the non-normal distribution by:

1.Conducting shapiro test (w=0,95234; p value =0,000654) which is significant lower than 0,05 .

2.Evaluating kurtosis value which is -0,37600593.Finding of mean value (101,894) which is quite far difference from median value (96.54)

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

Ans : Z score of 90% confidence interval is 1.65

Z score of 94% confidence interval is 1.55

Z score of 60% confidence interval is 0.85

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

Ans : A) For 95%= 1.96

For 96%= 2.5

For 99% = 2.47

**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 : mu=270

n=18

xbar=260

sigma=90

z=x-mu/sigma=260-270/90=-0.11

pnorm(-0.11)=0.4562p=45%

T=x-mu/s/sqrt(n) =260-270/90/sqrt(18)=-0.4714Pt-(0.4714, 17)=0.3216P=32%