EXCERISE-1

1.67+78-45

>67+78-45

[1] 100

2. 43+52 -3x81

>4^3+5^2-3\*81

[1] -154

3.

> sqrt(28)+547^(1/3)-47/53

[1] 12.583

4.+12% of 75

> exp(3)+0.12\*75

[1] 29.08554

5. ∛729+log(23/42)

> 729^(1/3)+log(23/42)

[1] 8.397825

6. (1.01)6+(2.67)3.4 – (3.2)(-2.1)

> (1.01)^6+(2.67)^3.4-(3.2)^(-2.1)

[1] 29.16739

7. 233 +4562 -56

> 23^3+456^2-56

[1] 220047

EXCERISE-2

1.Assign single values to X and Y as 3 and 4. Then find Z = X + Y; W = X\*Y; A = Z + W; B = A2 +Y; C= X3+Y3

> x=3;y=4

> z=x+y;z

[1] 7

> w=x\*y;w

[1] 12

> a=z+w;a

[1] 19

> b=a^2+sqrt(y);b

[1] 363

> c=x^3+y^3;c

[1] 91

2.Assign combination of values (equal length) to X and Y and do above calculations. for eg X= [2, 3, 5, 7] and Y= [11,13,17,19]> x=c(2,3,5,7);y=c(11,13,17,19)

> z=x+y;z

[1] 13 16 22 26

> w=x\*y;w

[1] 22 39 85 133

> a=z+w;a

[1] 35 55 107 159

> b=a^2+sqrt(y);b

[1] 1228.317 3028.606

[3] 11453.123 25285.359

> c=x^3+y^3;c

[1] 1339 2224 5038 7202

3. For problem 2 obtain the values for X/2, Y/3, X/Y

> x/2

[1] 1.0 1.5 2.5 3.5

> y/3

[1] 3.666667 4.333333 5.666667

[4] 6.333333

> x/y

[1] 0.1818182 0.2307692

[3] 0.2941176 0.3684211

4. Assign a vector of character strings (“Bob”, “Jack”, “Jill”) for names.

> names=c("bob","jack","jill");names

[1] "bob" "jack" "jill"

EXCERISE-3

1.Use sequence operator to get a sequence

i.from 1 to 20

ii.from 20 to 10

iii.From 2 to 30 of width 2

> 1:20

[1] 1 2 3 4 5 6 7 8

[9] 9 10 11 12 13 14 15 16

[17] 17 18 19 20

> 20:10

[1] 20 19 18 17 16 15 14 13

[9] 12 11 10

> 2\*1:15

[1] 2 4 6 8 10 12 14 16

[9] 18 20 22 24 26 28 30

2.Assign value 15 to n and find the difference between 1: n-1 and 1:(n-1)

> n=15

> 1:n-1

[1] 0 1 2 3 4 5 6 7

[9] 8 9 10 11 12 13 14

> 1:(n-1)

[1] 1 2 3 4 5 6 7 8

[9] 9 10 11 12 13 14

EXCERISE-4

1.Enter the following data using rep function

i)1,1,1,1,2,2,3,3,3,3,3,

ii)4,4,4,4,5,5,6,6,6,6,7,8,8,8

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

iv)10,10,10,10,11,11,11,11,12,12,12,12

> a=c(rep(1,4),rep(2,2),rep(3,5));a

[1] 1 1 1 1 2 2 3 3 3 3 3

> b=c(rep(4,4),rep(5,2),rep(6,4),7,rep(8,3));b

[1] 4 4 4 4 5 5 6 6 6 6 7 8 8 8

> c=rep(1:6,each=2);c

[1] 1 1 2 2 3 3 4 4 5 5 6 6

> d=rep(10:12,each=4);d

[1] 10 10 10 10 11 11 11 11 12 12 12 12

EXCERISE-5

1. Using data.frame function make the following frequency distribution.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Age | freq |  | Marks | freq |  | Variable | freq |
|  | 11 | 5 |  | 15 | 2 |  | 13 | 1 |
|  | 12 | 10 |  | 20 | 2 |  | 17 | 1 |
|  | 13 | 120 |  | 25 | 3 |  | 19 | 2 |
|  | 14 | 22 |  | 30 | 3 |  | 24 | 2 |
|  | 15 | 13 |  | 35 | 3 |  | 29 | 3 |
|  | 16 | 5 |  | 40 | 4 |  | 33 | 3 |

> age=11:16;freq=c(5,10,120,22,13,5);d1=data.frame(age,freq);d1

age freq

1 11 5

2 12 10

3 13 120

4 14 22

5 15 13

6 16 5

2.For the above data change the names of the columns

1. Mid age and No of cases.
2. Score and No of students
3. Income in ‘000 and No of families

> colnames(d1)=c("mid age","no of cases");d1

mid age no of cases

1 11 5

2 12 10

3 13 120

4 14 22

5 15 13

6 16 5

EXCERISE-6

1.Following is the data set: 5, 12, 21, 25, 25, 30, 25, 40, 42, 38, 50, 45, 60, 65, 50,70, 80, 50,13. Use the built-in functions discussed above, on the data set x.

> x=scan()

1: 5 12 21 25 30 25 40 42 38 50 45 60 65 50 70 80 50 13 20

20:

Read 19 items

> length(x)

[1] 19

> max(x)

[1] 80

> min(x)

[1] 5

> range(x)

[1] 5 80

EXCERISE-7

**For the given data sets;**

1. Enter the data set either using the scan function or c function .
2. Find the index for its maximum and minimum value
3. Find the summary.
4. Find all functions wrt this data set
5. Construct the discrete distribution.

**Data set I:** 13, 17, 24, 21, 28, 28, 13, 27, 17, 23, 17, 24, 21, 17, 23, 21

**Data set II:** 0, 1, 2, 3, 4, 5, 6, 6, 5, 4, 4, 5, 5, 4, 4, 3, 3, 3, 3, 2, 2, 2, 3, 2, 3, 2, 2, 2, 1, 1, 1, 0, 0, 1, 0, 3, 3, 2, 2, 2, 3, 2, 3, 2, 2, 2, 1, 1, 1, 0, 0, 1,0

**Data set I**

> x=scan()

1: 13 17 24 21 28 28 13 27 17 23 17 24 21 17 23 21

17:

Read 16 items

> max(x)

[1] 28

> min(x)

[1] 13

> summary(x)

Min. 1st Qu. Median Mean

13.00 17.00 21.00 20.88

3rd Qu. Max.

24.00 28.00

> quantile(x)

0% 25% 50% 75% 100%

13 17 21 24 28

> names(x)

NULL

> table(x)

x

13 17 21 23 24 27 28

2 4 3 2 2 1 2

**Data set II**

> x=scan()

1: 0 1 2 3 4 5 6 6 5 4 4 5 5 4 4 3 3 3 3 2 2 2 3 2 3 2 2 2 1 1 1 0 0 1 0 3 3 2 2 2 3 2 3 2 2 2 1 1 1 0 0 1 0

54:

Read 53 items

> max(x)

[1] 6

> min(x)

[1] 0

> summary(x)

Min. 1st Qu. Median Mean

0.00 1.00 2.00 2.34

3rd Qu. Max.

3.00 6.00

> quantile(x)

0% 25% 50% 75% 100%

0 1 2 3 6

> table(x)

x

0 1 2 3 4 5 6

7 9 15 11 5 4 2

EXCERISE-8

1.A psychologist estimates the I.Q. of 60 children. The values are as follows :103, 98, 87, 85, 67, 96, 115, 109, 127, 103, 95, 123, 94, 88, 102, 76, 73, 80, 84, 102, 115, 93, 76, 81, 132, 90, 119, 84, 97, 120, 114, 101, 153, 98, 99, 105, 110, 107, 110, 128, 89, 112, 118, 101, 122, 146, 96, 109, 72, 97, 94, 94, 79, 79, 100, 54, 102, 89, 43, 111.

> x=c(103, 98, 87, 85, 67, 96, 115, 109, 127, 103, 95, 123, 94, 88, 102, 76, 73, 80, 84, 102, 115, 93, 76, 81, 132, 90, 119, 84, 97, 120, 114, 101, 153, 98, 99, 105, 110, 107, 110, 128, 89, 112, 118, 101, 122, 146, 96, 109, 72, 97, 94, 94, 79, 79, 100, 54, 102, 89, 43, 111)

> summary(x)

Min. 1st Qu. Median Mean 3rd Qu. Max.

43.00 87.75 98.50 99.10 110.25 153.00

> (153-43)/5

[1] 22

> seq(43,160,by=22)

[1] 43 65 87 109 131 153

> ci=seq(43,160,by=22)

> length(x)

[1] 60

> range(x)

[1] 43 153

> y=cut(x,ci,right=F);y

[1] [87,109) [87,109) [87,109) [65,87)

[5] [65,87) [87,109) [109,131) [109,131)

[9] [109,131) [87,109) [87,109) [109,131)

[13] [87,109) [87,109) [87,109) [65,87)

[17] [65,87) [65,87) [65,87) [87,109)

[21] [109,131) [87,109) [65,87) [65,87)

[25] [131,153) [87,109) [109,131) [65,87)

[29] [87,109) [109,131) [109,131) [87,109)

[33] <NA> [87,109) [87,109) [87,109)

[37] [109,131) [87,109) [109,131) [109,131)

[41] [87,109) [109,131) [109,131) [87,109)

[45] [109,131) [131,153) [87,109) [109,131)

[49] [65,87) [87,109) [87,109) [87,109)

[53] [65,87) [65,87) [87,109) [43,65)

[57] [87,109) [87,109) [43,65) [109,131)

5 Levels: [43,65) [65,87) [87,109) ... [131,153)

> fd=cbind(table(y));fd

[,1]

[43,65) 2

[65,87) 12

[87,109) 27

[109,131) 16

[131,153) 2

2.The following data regarding weight of new born babies is obtained from the office records of a hospital. Weight (kgs.) 3.7, 3.4, 4.1, 4.0, 3.7, 4.7, 3.3, 2.4, 3.1, 4.2, 3.8, 3.6, 4.2, 4.3, 2.9, 3.6, 3.3, 4.8, 4.0, 3.9, 3.5, 3.5, 3.8, 3.8, 4.2, 3.9, 4.9, 3.2, 4.0, 3.8, 3.2, 2.7, 3.4., 3.3, 3.0, 3.1, 3.5, 3.7, 3.9, 4.3, 3.8, 3.7, 3.0, 4.4, 4.1, 3.6, 3.7, 3.4, 3.7, 3.3, 3.5, 3.7, 3.0, 2.9, 3.1, 3.3, 4.2.

> x=c(3.7, 3.4, 4.1, 4.0, 3.7, 4.7, 3.3, 2.4, 3.1, 4.2, 3.8, 3.6, 4.2, 4.3, 2.9, 3.6, 3.3, 4.8, 4.0, 3.9, 3.5, 3.5, 3.8, 3.8, 4.2, 3.9, 4.9, 3.2, 4.0, 3.8, 3.2, 2.7, 3.4, 3.3, 3.0, 3.1, 3.5, 3.7, 3.9, 4.3, 3.8,

3.7, 3.0, 4.4, 4.1, 3.6, 3.7, 3.4, 3.7, 3.3, 3.5, 3.7, 3.0, 2.9, 3.1, 3.3, 4.2)

> summary(x)

Min. 1st Qu. Median Mean 3rd Qu. Max.

2.400 3.300 3.700 3.651 4.000 4.900

> (4.9-2.4)/5

[1] 0.5

> ci=seq(2.4,5.5,by=0.5)

> y=cut(x,ci,right=F);y

[1] [3.4,3.9) [3.4,3.9) [3.9,4.4) [3.9,4.4) [3.4,3.9)

[6] [4.4,4.9) [2.9,3.4) [2.4,2.9) [2.9,3.4) [3.9,4.4)

[11] [3.4,3.9) [3.4,3.9) [3.9,4.4) [3.9,4.4) [2.9,3.4)

[16] [3.4,3.9) [2.9,3.4) [4.4,4.9) [3.9,4.4) [3.9,4.4)

[21] [3.4,3.9) [3.4,3.9) [3.4,3.9) [3.4,3.9) [3.9,4.4)

[26] [3.9,4.4) [4.9,5.4) [2.9,3.4) [3.9,4.4) [3.4,3.9)

[31] [2.9,3.4) [2.4,2.9) [3.4,3.9) [2.9,3.4) [2.9,3.4)

[36] [2.9,3.4) [3.4,3.9) [3.4,3.9) [3.9,4.4) [3.9,4.4)

[41] [3.4,3.9) [3.4,3.9) [2.9,3.4) [4.4,4.9) [3.9,4.4)

[46] [3.4,3.9) [3.4,3.9) [3.4,3.9) [3.4,3.9) [2.9,3.4)

[51] [3.4,3.9) [3.4,3.9) [2.9,3.4) [2.9,3.4) [2.9,3.4)

[56] [2.9,3.4) [3.9,4.4)

6 Levels: [2.4,2.9) [2.9,3.4) [3.4,3.9) ... [4.9,5.4)

> fd=cbind(table(y));fd

[,1]

[2.4,2.9) 2

[2.9,3.4) 15

[3.4,3.9) 22

[3.9,4.4) 14

[4.4,4.9) 3

[4.9,5.4) 1

EXCERISE-9

1. Access the data set *treering* containing tree-ring widths in dimensionless unit, from the base package of R. Use R-commands to answer the following

1. how many observations are in the data set?
2. What is the minimum and maximum observation?
3. List observation greater than the 1.8.
4. Find the quartiles of the data set.
5. Find the index for the maximum and minimum value of data set.
6. Construct appropriate frequency distribution table

> data(treering);d=treering;

> length(d)

[1] 7980

> summary(d)

Min. 1st Qu. Median Mean 3rd Qu. Max.

0.0000 0.8370 1.0340 0.9968 1.1970 1.9080

> d[d>1.8]

[1] 1.844 1.850 1.856 1.820 1.884 1.908 1.826 1.802

> length(d[d>1.8])

[1] 8

> d[1:5]

[1] 1.345 1.077 1.545 1.319 1.413

> d[7976:7980]

[1] 1.027 1.173 1.471 1.444 1.160

> which(d==.0000)

[1] 1395

> which(d==1.9080)

[1] 2185

> ci=seq(0,2,0.2);ci

[1] 0.0 0.2 0.4 0.6 0.8 1.0 1.2 1.4 1.6 1.8 2.0

> y=cut(d,ci,right=F);fd=cbind(table(y));fd

[,1]

[0,0.2) 121

[0.2,0.4) 254

[0.4,0.6) 473

[0.6,0.8) 914

[0.8,1) 1795

[1,1.2) 2457

[1.2,1.4) 1459

[1.4,1.6) 430

[1.6,1.8) 69

[1.8,2) 8

2. Access the data set *rivers*, from the base package of R. Use R-commands to answer the following

1. how many observations are in the data set?
2. What is the minimum and maximum observation?
3. List observation greater than the median.
4. Find the quartiles of the data set.
5. Find the index for the maximum and minimum value of data set.
6. Construct appropriate frequency distribution table

> data(rivers);d=rivers

> length(d)

[1] 141

> summary(d)

Min. 1st Qu. Median Mean 3rd Qu. Max.

135.0 310.0 425.0 591.2 680.0 3710.0

> d[d>425.0]

[1] 735 524 450 1459 465 600 870 906 1000 600

[11] 505 1450 840 1243 890 525 720 850 630 730

[21] 600 710 470 680 570 560 900 625 2348 1171

[31] 3710 2315 2533 780 460 431 760 618 981 1306

[41] 500 696 605 1054 735 435 490 460 1270 545

[51] 445 1885 800 538 1100 1205 610 540 1038 444

[61] 620 652 900 525 529 500 720 430 671 1770

> d[1:5]

[1] 735 320 325 392 524

> which(d==135.0)

[1] 8

> which(d==3710.0)

[1] 68

> (3710-135)/5

[1] 715

> ci=seq(135,3710,by=715)

> y=cut(d,ci,right=F);fd=cbind(table(y));fd

[,1]

[135,850) 117

[850,1.56e+03) 18

[1.56e+03,2.28e+03) 2

[2.28e+03,3e+03) 3

[3e+03,3.71e+03) 0

3.For the given data

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| X | 0 | 1 | 2 | 3 | 4 |
| F | 6 | 28 | 36 | 25 | 5 |

1. Add a column of cumulative frequency(cf)
2. Add a column of relative frequency(rf) (frequency/total frequency)
3. Add a column of relative cumulative frequency (cf/total frequency)

> x=0:4

> f=c(6,28,36,25,5)

> d1=data.frame(x,f);d1

x f

1 0 6

2 1 28

3 2 36

4 3 25

5 4 5

> cf=transform(d1,cfreq=cumsum(f));cf

x f cfreq

1 0 6 6

2 1 28 34

3 2 36 70

4 3 25 95

5 4 5 100

> cf1=transform(d1,rf=f/sum(f));cf1

x f rf

1 0 6 0.06

2 1 28 0.28

3 2 36 0.36

4 3 25 0.25

5 4 5 0.05

4. Access the data set *swiss*, from the base package of R. Use R-commands to answer the following Fertility Agriculture Examination Education Catholic

Find the mean and variance for Agriculture

* Construct a continuous frequency distribution for either Examination or Education
* Find the number of observation that has Catholic less than 60
* Get all the information with respect to 6th row
* Get all the information with respect to 6th column
* Get all the information with respect to the 5th,10th,….. & 45th observations.
* Get all the information with respect to the 1th,17th,29th,33rd,47th observations.

> data("swiss")

> mean(swiss$Fertility)

[1] 70.14255

> var(swiss$Agriculture)

[1] 515.7994

> summary(swiss$Examination)

Min. 1st Qu. Median Mean 3rd Qu. Max.

3.00 12.00 16.00 16.49 22.00 37.00

> (37-3)/5

[1] 6.8

> ci=seq(3,45,by=5)

> x=cut(swiss$Examination,ci,righth=F)

> fd=cbind(table(x));fd

[,1]

(3,8] 6

(8,13] 7

(13,18] 15

(18,23] 9

(23,28] 4

(28,33] 2

(33,38] 2

(38,43] 0

> sum(swiss$Catholic < 60)

[1] 31

> swiss[6, ]

Fertility Agriculture Examination Education

Porrentruy 76.1 35.3 9 7

Catholic Infant.Mortality

Porrentruy 90.57 26.6

>swiss[, 6]

[1] 22.2 22.2 20.2 20.3 20.6 26.6 23.6 24.9 21.0 24.4

[11] 24.5 16.5 19.1 22.7 18.7 21.2 20.0 20.2 10.8 20.0

[21] 18.0 22.4 16.7 15.3 21.0 23.8 18.0 16.3 20.9 22.5

[31] 15.1 19.8 18.3 19.4 20.2 17.8 16.3 18.1 20.3 20.5

[41] 18.9 23.0 20.0 19.5 18.0 18.2 19.3

> observations\_indices <- c(5, 10, seq(15, 45, by = 5))

> observations\_info <- swiss[observations\_indices, ]

> print("Information of specified observations:")

[1] "Information of specified observations:"

> print(observations\_info)

Fertility Agriculture Examination

Neuveville 76.9 43.5 17

Sarine 82.9 45.2 16

Cossonay 61.7 69.3 22

Lavaux 65.1 73.0 19

Oron 72.5 71.2 12

Yverdon 65.4 49.5 15

Monthey 79.4 64.9 7

La Chauxdfnd 65.7 7.7 29

V. De Geneve 35.0 1.2 37

Education Catholic Infant.Mortality

Neuveville 15 5.16 20.6

Sarine 13 91.38 24.4

Cossonay 5 2.82 18.7

Lavaux 9 2.84 20.0

Oron 1 2.40 21.0

Yverdon 8 6.10 22.5

Monthey 3 98.22 20.2

La Chauxdfnd 11 13.79 20.5

V. De Geneve 53 42.34 18.0

> specified\_indices <- c(1, 17, 29, 33, 47)

> specified\_observations\_info <- swiss[specified\_indices, ]

> print("Information of specified observations:")

[1] "Information of specified observations:"

> print(specified\_observations\_info)

Fertility Agriculture Examination Education

Courtelary 80.2 17.0 15 12

Grandson 71.7 34.0 17 8

Vevey 58.3 26.8 25 19

Herens 77.3 89.7 5 2

Rive Gauche 42.8 27.7 22 29

Catholic Infant.Mortality

Courtelary 9.96 22.2

Grandson 3.30 20.0

Vevey 18.46 20.9

Herens 100.00 18.3

Rive Gauche 58.33 19.3

EXCERISE-10

1. Access the data set cars from the base library of R

1. Construct Boxplot for the variables in it.
2. Obtain the summary of the variables

> data(cars);d1=cars;attach(d1)

> dim(d1)

[1] 50 2

> names(d1)

[1] "speed" "dist"

> s=speed

> boxplot(s,xlab="speed")

A diagram of a graph

Description automatically generated

>d=dist

> boxplot(d,xlab="distance")

> identify(rep(1,length(d)),d);

A diagram of a graph

Description automatically generated

> d1[49,]

speed dist

49 24 120

2. Access the data cats from the library MASS and plot sexwise boxplot for the variable Hwt(heart weight)

> library(MASS)

> data(cats)

> attach(cats);names(cats);

[1] "Sex" "Bwt" "Hwt"

> boxplot(Hwt~Sex);

> identify(as.numeric(Sex),Hwt)

integer(0)

> cats[c(47,144)]

> boxplot(Hwt~Sex, col=c("red","blue"), names=c("female","male"));

A graph with a red and blue box

Description automatically generated

3. Access the data set ***InsectSprays*** from the base package of R. Construct parallel boxplots for different sprays.

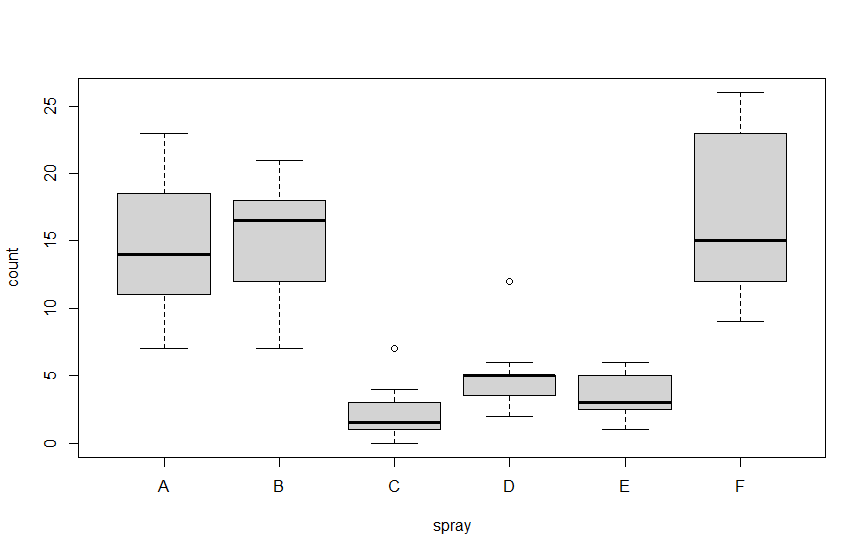
Hint: >boxplot(count~spray)

> data("InsectSprays")

> attach(InsectSprays);names(InsectSprays);

[1] "count" "spray"

> boxplot(count~spray);



4. Following are the body mass index values (kg/m2) for 14 subjects in sample

24.4, 3.04, 21.4, 25.4, 21.3, 23.8, 20.8, 22.9, 23.2, 21.1, 23.0, 20.6, 26.0, 20.9

i) compute mean, median, variance, standard deviation and coefficient of variation

ii) construct box and whisker plot. If outliers are found identify them.

iii)Compute Bowley’s measure of skewness

>x=c(24.4, 3.04, 21.4, 25.4, 21.3, 23.8, 20.8, 22.9, 23.2, 21.1, 23.0, 20.6, 26.0, 20.9)

> median(x)

[1] 22.15

> mean(x)

[1] 21.27429

> var(x)

[1] 30.62987

> sd(x)

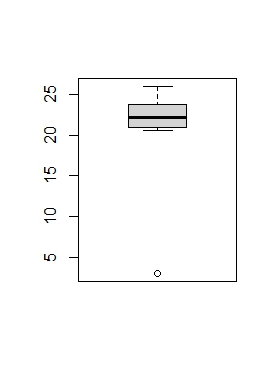
[1] 5.534426

> cv=sd(x) / mean(x)\*100

> cv

[1] 26.01463

> boxplot(x)



iii) > summary(x)

Min. 1st Qu. Median Mean 3rd Qu. Max.

3.04 20.95 22.15 21.27 23.65 26.00

> (23.65+20.95-2\*22.15)/(23.65-20.95)

[1] 0.1111111

EXCERISE-11

1.Following are the number of accidents that occurred at 60 major intersections in a certain city during a weekend: 0 1 0 2 4 2 5 0 3 0 2 0 1 4 4 4 1 2 1 2 5 0 4 1 0 2 1 1 4 2 5 3 2 0 5 1 1 0 6 3 1 5 0 3 0 0 6 3 2 2 3 1 4 0 3 0 0 1 2 4

Prepare a frequency distribution table and draw a bar chart. Comment on the nature of the distribution.

>x=c(0,1,0,2,4,2,5,0,3,0,2,0,1,4,4,4,1,2,1,2,5,0,4,1,0,2,1,1,4,2,5,3,2,0,5,1,1,0,6,3,1,5,0,3,0,0,6,3,2,2,3,1,4,0,3,0,0,1,2,4)

> t=table(x)

> t

x

0 1 2 3 4 5 6

15 12 11 7 8 5 2

> barplot(x)

A graph of a city

Description automatically generated with medium confidence

The distribution indicates a negative skew, with most major intersections experiencing few or no accidents during the weekend, while a minority encountered higher accident counts, up to 6 accidents.

2. From the information obtained in Q1 draw a pie diagram

> accidents = c(0, 1, 0, 2, 4, 2, 5, 0, 3, 0, 2, 0, 1, 4, 4, 4, 1, 2, 1, 2, 5, 0, 4, 1, 0, 2, 1, 1, 4, 2, 5, 3, 2, 0, 5, 1, 1, 0, 6, 3, 1, 5, 0, 3, 0, 0, 6, 3, 2, 2, 3, 1, 4, 0, 3, 0, 0, 1, 2, 4)

> accident\_counts <- table(accidents)

> pie(accident\_counts, main = "Number of Accidents at Major Intersections", labels = paste(names(accident\_counts), ": ", accident\_counts), col = rainbow(length(accident\_counts)))

A colorful pie chart with numbers

Description automatically generated

EXCERISE-12

1. Draw a histogram and and frequency polygon for the following data.

Height 0-7 7-14 14-21 21-28 20-35 35-42 42-49 49-50

No. of people: 26 31 35 42 82 71 54 19

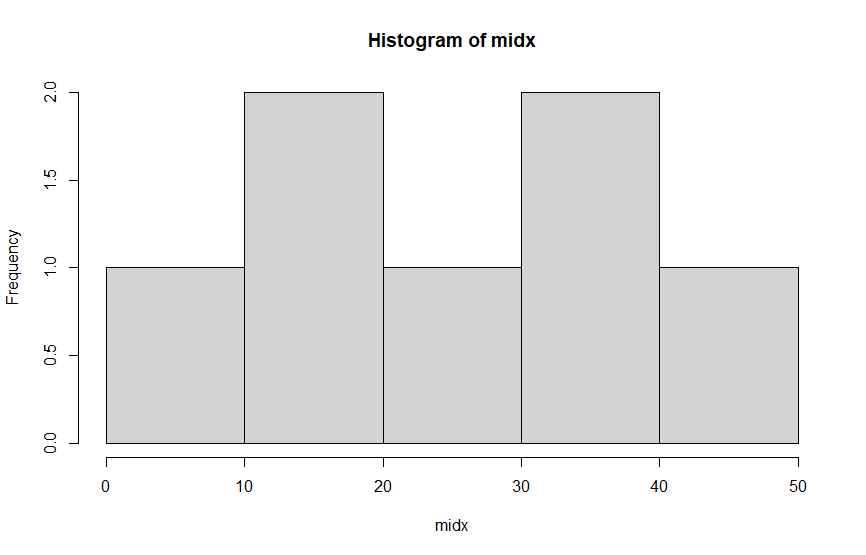
> midx=seq(3.5,49.5,7);freq=c(26,31,35,42,82,71,54,19)

> plot(midx,type = "o")

A graph of a number of numbers and a line

Description automatically generated with medium confidence

>hist(midx)



2. Plot the histogram and frequency polygon on the same graph for the given data

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Class interval | 20-30 | 30-40 | 40-50 | 50-60 | 60-70 | 70-80 | 80-90 |
| Frequency | 10 | 24 | 18 | 12 | 8 | 5 | 3 |

???

EXCERISE-13

1. Plot the scatter plot and compute the both the correlation coefficient for the following data

i)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| X | 0 | 4 | 8 | 12 |
| Y | 8.34 | 8.89 | 9.16 | 9.50 |

ii)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| A | 11.1 | 10.3 | 12.0 | 15.1 | 13.7 | 18.5 | 17.3 | 14.2 | 14.8 | 15.3 |
| B | 10.9 | 14.2 | 13.8 | 21.5 | 13.2 | 21.1 | 16.4 | 19.3 | 17.4 | 19.0 |

iii)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| C | 5.12 | 6.18 | 6.77 | 6.65 | 6.36 | 5.90 | 5.48 | 6.02 | 10.34 | 8.51 |
| D | 2.30 | 2.54 | 2.95 | 3.77 | 4.18 | 5.31 | 5.53 | 8.83 | 9.48 | 14.20 |

> x=c(0,4,8,12)

> y=c(8.34,8.89,9.16,9.50)

> p=plot(x,y)

> cor(x,y,method="spearman")

[1] 1

A graph with numbers and points

Description automatically generated

> a=c(11.1,10.3,12.0,15.1,13.7,18.5,17.3,14.2,14.8,15.3)

> b=c(10.9,14.2,13.8,21.5,13.2,21.1,16.4,19.3,17.4,19.0)

> p=plot(a,b)

> cor(a,b,method="spearman")

[1] 0.6969697

A graph with numbers and points

Description automatically generated

> d=c(2.30,2.54,2.95,3.77,4.18,5.31,5.53,8.83 ,9.48,14.20)

> p=plot(c,d)

> cor(c,d,method="spearman")

[1] 0.4181818

A graph with numbers and points

Description automatically generated

2.

|  |  |
| --- | --- |
| X1 | Y1 |
| 10 | 8.04 |
| 8 | 6.95 |
| 13 | 7.58 |
| 9 | 8.81 |
| 11 | 8.33 |
| 14 | 9.96 |
| 6 | 7.24 |
| 4 | 4.26 |
| 12 | 10.84 |
| 7 | 4.82 |
| 5 | 5.68 |

|  |  |
| --- | --- |
| X2 | Y2 |
| 10 | 9.14 |
| 8 | 8.14 |
| 13 | 8.74 |
| 9 | 8.77 |
| 11 | 9.26 |
| 14 | 8.10 |
| 6 | 6.13 |
| 4 | 3.10 |
| 12 | 9.13 |
| 7 | 7.26 |
| 5 | 4.78 |

For the above two data set verify the following

1. Mean of x1is same as mean of x2
2. Mean of y1 is same as mean of y2
3. Correlation coefficient between (x1,y1) is same as (x2,y2)
4. Draw the scatter plot and comment on the findings

> x1=c(10,8,13,9,11,14,6,4,12,7,5)

> x2=c(10,8,13,9,11,14,6,4,12,7,5)

> mean(x1)

[1] 9

> mean(x2)

[1] 9

y1=c(8.04,6.95,7.58,8.81,8.33,9.96,7.24,4.26,10.84,4.82,5.68)

> y2=c(9.14,8.14,8.74,8.77,9.26,8.10,3.10,9.13,7.26,4.78)

> mean(y1)

[1] 7.500909

> mean(y2)

[1] 7.642

> cor(x1,y1,method = "spearman")

[1] 0.8181818

>cor(x2,y2,method=”sperman”)

>

> p=plot(x1,y1)

A graph with numbers and points

Description automatically generated

> p=plot(x2,y2)

>

EXCERISE-14

1. The table shown the score of 10 students on maths(X) test and stats(Y) test. The maximum score in each test was 50.

1. Obtain the line of regression of X on Y.
2. Print this equation on the graph
3. if it is known that a student gets 28 in stats, what would be his/her score in maths?

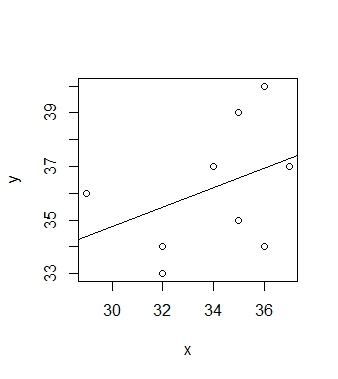
|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X | 34 | 37 | 36 | 32 | 32 | 36 | 35 | 34 | 29 | 35 |
| Y | 37 | 37 | 34 | 34 | 33 | 40 | 39 | 37 | 36 | 35 |

> x=c(34,37,36,32,32,36,35,34,29,35)

> y=c(37,37,34,34,33,40,39,37,36,35)

> plot(x,y)

> fit=lm(y~x);abline(fit);fit



2. Calculate person’s coefficient of correlation for the following data.

X : 45 55 56 58 60 65 68 70 75 80 85

Y : 56 50 48 60 62 64 65 70 74 82 90

Plot the line of best fit and Estimate Y when X = 78

> x=c(45,55,56,58,60,65,68,70,75,80,85)

> y=c(56,50,48,60,62,64,65,70,74,82,90)

> cor(x,y)

[1] 0.9188406

> fit=lm(y~x);abline(fit);fit

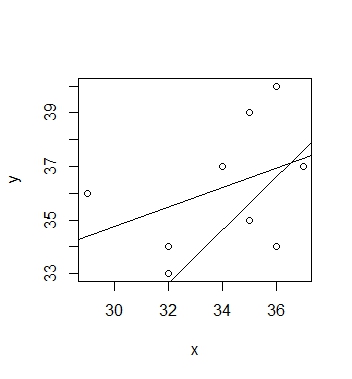
Call:

lm(formula = y ~ x)

Coefficients:

(Intercept) x

0.9044 0.9917



3.Calculate the coefficient of correlation by Karl Person’s method from the following data relating to overhead expenses and cost of production

Overhead expense (1000 Rs.) 80 90 100 110 120 130 140 150 160

Cost of (Rs. 1000) 15 15 16 19 17 18 16 18 19

Plot the line of best fit and estimate X when Y = 22

> x=10\*8:16

> y=c(15,15,16,19,17,18,16,18,19)

> mean\_overhead = mean(x)

> mean\_production = mean(y)

> deviation\_overhead = x - mean\_overhead

> deviation\_production = y - mean\_production

> product\_deviations = deviation\_overhead \* deviation\_production

> sum\_product\_deviations = sum(product\_deviations)

> sum\_squares\_overhead = sum(deviation\_overhead^2)

> sum\_squares\_production = sum(deviation\_production^2)

> correlation\_coefficient = sum\_product\_deviations / sqrt(sum\_squares\_overhead \* sum\_squares\_production)

> correlation\_coefficient

[1] 0.6928203

> plot(x,y, main = "Scatterplot of Overhead Expenses vs Cost of Production",

+ xlab = "Overhead Expenses (1000 Rs.)", ylab = "Cost of Production (Rs. 1000)")

> abline(lm(y ~ x), col = "red")

A graph with a red line

Description automatically generated

> X\_estimated = (22 - coef(lm(y ~ x))[1]) / coef(lm(y ~ x))[2]

> X\_estimated

(Intercept)

245

EXCERISE-15

1. The incident of occupational disease is such that the workers have 20% chance of catching it, what is the probability that out of 6 workers chosen (i) 4 or more are disease. (ii) atmost 2 catches the disease

> dbinom(4,6,0.2,log=FALSE)

[1] 0.01536

> pbinom(2,6,0.2,log=FALSE)

[1] 0.90112

2.The probability that a patient recovers from a sax blood disease 0.21. If 15 people are known to have contracted this disease what is the probability that: a) Atleast 10 survive? b) From 3 to 8 survive

>pbinom(10,15,0.21,log=FALSE)

[1] 0.9999796

> pbinom(8,15,0.21)-pbinom(2,15,0.21)

[1] 0.6373935

3. Find the probability that seven of ten persons will recover from a tropical disease, given that the probability is 0.8, that any one of these will recover from the disease.

> dbinom(7,10,0.8)

[1] 0.2013266

4. A basketball player hits on seventy-five percent of his shots from the free throw line. What is the probability that he makes exactly two of his next four free shots?

> n=4

> p=0.75

> pbinom(2,n,p)

[1] 0.2617188

5. In a certain city, incompatibility is given as the legal reason in 70% of all divorce cases. Find the probability that 5 of the next 6 divorce cases in this city will blame incompatible.

> n=6

> p=0.70

> pbinom(5,n,p)

[1] 0.882351

6. A automobile safety engineer claims that one in ten automobile accidents is due to driver fatigue. What is the probability that at least three of five automobile accidents are due to driver fatigue?

> 1-pbinom(2,5,0.1)

[1] 0.00856

7. Seven unbiassed and coins are tossed, and No. of heads are noted. The experiment is repeated 128 times and the following results are obtained. Fit a binomial distribution and obtain the expected frequencies.

No.of Heads (x) 0 1 2 3 4 5 6 7

Frequency 7 6 17 35 30 23 7 3

> x = 0:7

> observed\_freq = c(7, 6, 17, 35, 30, 23, 7, 3)

> total\_trials= 128

> prob\_heads=x / 7

> expected\_freq=dbinom(x, size = 7, prob = prob\_heads) \* total\_trials

> cat("Observed Frequencies:", observed\_freq, "\n")

Observed Frequencies: 7 6 17 35 30 23 7 3

> cat("Expected Frequencies:", round(expected\_freq), "\n")

Expected Frequencies: 128 51 41 38 38 41 51 128

8. A set of six similar coins are tossed 640 times and the following results are obtained

No. of Head(x) 0 1 2 3 4 5 6

Frequency 7 64 140 210 130 75 12

Fit a binomial distribution assuming that the nature of the coin is unknown

> x=0:6

> observed\_freq=c(7, 64, 140, 210, 130, 75, 12)

> total\_trials=640

> prob\_heads=sum(x \* observed\_freq) / (6 \* total\_trials)

> expected\_freq=dbinom(x, size = 6, prob = prob\_heads) \* total\_trials

> cat("Observed Frequencies:", observed\_freq, "\n")

Observed Frequencies: 7 64 140 210 130 75 12

> cat("Expected Frequencies:", round(expected\_freq), "\n")

Expected Frequencies: 9 57 147 200 153 63 11

EXCERISE-16

1. A hospital switch board receives an average of 4 emergency calls in a 10 minutes interval.

a) What is the probability that there are at the most 2 emergency calls in 10 minutes interval

b) There are exactly 3 emergency calls in 10 minutes

c) Atleast 4 calls in 10 minutes interval

> ppois(2,4,lower.tail = TRUE,log.p = FALSE)

[1] 0.2381033

> dpois(3,4,log=F)

[1] 0.1953668

> 1-ppois(3,4,lower.tail = TRUE,log.p = FALSE)

[1] 0.5665299

2. Assuming that the chance of a traffic accident in a City of Delhi is 0.001 on how many days out of 1000 days can we expect no accidents and more than 3 accidents.

> lambda = 0.001

> days = 1000

> prob\_no\_accidents = dpois(0, lambda)

> prob\_more\_than\_3\_accidents = 1 - ppois(3, lambda)

> expected\_days\_no\_accidents = days \* prob\_no\_accidents

> expected\_days\_no\_accidents

[1] 999.0005

> expected\_days\_more\_than\_3\_accidents <- days \* prob\_more\_than\_3\_accidents

> expected\_days\_more\_than\_3\_accidents

[1] 4.152234e-11

3. Fit a Poisson distribution to following data w.r.t. No.of. R.B.C.s per cell

No. of R.B.C. 0 1 2 3 4

No. of cells 142 156 69 27 5

4. If the number of mistakes made by a typist follows a Poisson distribution with mean 3, what is the chance that he/she

i) makes 2 mistakes, ii) makes atleast 2 mistakes

> dpois(2,3,log=F)

[1] 0.2240418

> 1-ppois(2,3,lower.tail = T,log.p = F)

[1] 0.5768099

5. The number of accidents occurring in a factory in a year is a Poission variate with mean 5. Find the probability that.

i) more than 2 accidents take place

ii) more than 4 accidents occur in 1 year

> 1-ppois(2,5,lower.tail = T,log.p=F)

[1] 0.875348

> 1-ppois(4,5,lower.tail = T,log.p=F)

[1] 0.5595067

6. A receptionist at an office receives on an average 3 telephone calls between 10 a.m. and 10.05 a.m. Find the probability that on a particular day

i) she does not receive any call

ii) she receives atleast 2 calls

> ppois(0,3,lower.tail = T,log.p = F)

[1] 0.04978707

> 1-ppois(2,3,lower.tail = T,log.p=F)

[1] 0.5768099

7. At 10.00 a.m. there is a city bus service. The number of passengers getting in at the 1st stop is a Poisson variate with parameter 6. What is the probability that on a particular day none of them gets in at the bus in the stop? On how many days of an year would you expect this to happen.

> lambda=6

> prob\_zero\_passengers <- dpois(0, lambda)

> cat("Probability of none of the passengers getting in at the bus stop on a particular day:", prob\_zero\_passengers, "\n")

Probability of none of the passengers getting in at the bus stop on a particular day: 0.002478752

> days\_in\_year=365

> expected\_days=days\_in\_year \* prob\_zero\_passengers

> cat("Expected number of days in a year where none of the passengers get in at the bus stop:", expected\_days)

Expected number of days in a year where none of the passengers get in at the bus stop: 0.9047445

8. On an average 3 street lights of a municipality fails every day. Find the standard deviation of number of failure per day and probability that atleast one light fails per day.

> sqrt(3)

[1] 1.732051

> 1-ppois(0,3,lower.tail = T,log.p = F)

[1] 0.9502129

9. On an average 1% of the pins are defective. If the box contains 300 pins, find the probability that the box has

i) atleast 1 defective pin

ii) more than 3 defective pins

> 1-ppois(0,3,lower.tail = T,log.p = F)

[1] 0.9502129

> 1-ppois(3,3,lower.tail = T,log.p = F)

[1] 0.3527681

10. On an average 1 in every 50 valves manufactured by a firm is substandard. If the valves are supplied in packers of 20 each

i) Find the probability that the packets will contain atleast 1 substandard valve

ii) In how many of a lot of 1000 packets would you expect substandard valves.

> p\_substandard = 1/50

> n\_valves\_per\_packet = 20

> total\_packets = 1000

> prob\_at\_least\_one\_substandard = 1 - dbinom(0, size = n\_valves\_per\_packet, prob = p\_substandard)

> cat("i) Probability of at least 1 substandard valve in a packet:", prob\_at\_least\_one\_substandard, "\n")

i) Probability of at least 1 substandard valve in a packet: 0.332392

> expected\_substandard\_valves = total\_packets \* p\_substandard

> cat("ii) Expected number of substandard valves in a lot of 1000 packets:", expected\_substandard\_valves)

ii) Expected number of substandard valves in a lot of 1000 packets: 20

11. Using the following data fit a Poisson distribution and find the expected frequencies

No.of Printing Mistakes 0 1 2 3 4 5

No.of days 42 33 14 6 4 1

> x = 0:5

> days = c(42, 33, 14, 6, 4, 1)

> total\_days = sum(days)

> lambda = sum(x \* days) / total\_days

> expected\_freq =dpois(x, lambda) \* total\_days

> cat("Observed Frequencies:", days, "\n")

Observed Frequencies: 42 33 14 6 4 1

> cat("Expected Frequencies:", round(expected\_freq), "\n")

Expected Frequencies: 37 37 18 6 2 0

12. The following is the distribution of daily sales of television sets in a shop, Fit a Poisson distribution and hence find the theoretical frequency.

No. of sets sold 0 1 2 3 4 5 6

No. of days 18 43 45 28 12 5 0

> x = 0:6

> days = c(18, 43, 45, 28, 12, 5, 0)

> total\_days = sum(days)

> lambda = sum(x \* days) / total\_days

> expected\_freq = dpois(x, lambda) \* total\_days

> cat("Observed Frequencies:", days, "\n")

Observed Frequencies: 18 43 45 28 12 5 0

> cat("Expected Frequencies:", round(expected\_freq), "\n")

Expected Frequencies: 22 42 41 26 13 5 2

EXCERISE-17

1. Given a normal distribution with mean = 50 and standard deviation = 8. Find the

probability that X assumes a value between 34 and 62.

> mean =50

> sd =8

> prob\_between\_34\_and\_62 = pnorm(62, mean, sd) - pnorm(34, mean, sd)

> prob\_between\_34\_and\_62

[1] 0.9104427

2. For a normal distribution with mean = 200 and S.D. = 25, find the probability

that X assumes a value between 200 and 260. Find the probability that X is greater

than 240.

> mean =200

> sd = 25

> prob\_between\_200\_and\_260 = pnorm(260, mean, sd) - pnorm(200, mean, sd)

> prob\_between\_200\_and\_260

[1] 0.4918025

> prob\_greater\_than\_240 = 1 - pnorm(240, mean, sd)

> prob\_greater\_than\_240

[1] 0.0547992

3. Given a Normal distribution with mean = 50 and S.D. = 13. Find the value of X

that has (a) 13% of the area to its left : b) 14% of the area to its right.

> mean = 50

> sd = 13

> X\_left = qnorm(0.13, mean, sd)

> X\_left

[1] 35.35692

> X\_right = qnorm(1 - 0.14, mean, sd)

> X\_right

[1] 64.04415

4. The accounts of a certain departmental store has an average balance of Rs. 120/-

and S.D. = Rs. 40/-. Assuming that the account balances are normally distributed.

a) what proportion of accounts is over Rs. 150/- b) what proportion is between 100

and 150; (c) between 60 and 90.

> mean = 120

> sd = 40

> prop\_over\_150 = 1 - pnorm(150, mean, sd)

> prop\_over\_150

[1] 0.2266274

> prop\_between\_100\_and\_150 = pnorm(150, mean, sd) - pnorm(100, mean, sd)

> prop\_between\_100\_and\_150

[1] 0.4648351

> prop\_between\_60\_and\_90 = pnorm(90, mean, sd) - pnorm(60, mean, sd)

> prop\_between\_60\_and\_90

[1] 0.1598202

5. The distribution of monthly income of 3000 workers of a factory follows normal

law with mean = 900 and S.D. = 100. Find

a) percentage of workers with income greater than Rs. 800

b) percentage of workers having on income less than Rs. 600.

> mean=900

> sd=100

> percentage\_greater\_than\_800 = 1 - pnorm(800, mean, sd)

> percentage\_greater\_than\_800

[1] 0.8413447

> percentage\_less\_than\_600 = pnorm(600, mean, sd)

> percentage\_less\_than\_600

[1] 0.001349898

6. 1200 students took an exam. The mean marks is 53% and S.D. = 15%. Assume

normal distribution of marks.

a)if 50% marks are required for passing, find how many students are expected to

score greater than 50%

b)if only 40% of students are required to be promoted what are the marks for

promotion.

>mean=53

> sd = 15

> prob\_passing = 0.5

> students\_greater\_than\_50 = (1 - pnorm(prob\_passing, mean, sd)) \* 1200

> students\_greater\_than\_50

[1] 1199.721

> prob\_promotion =0.4

> marks\_for\_promotion = qnorm(prob\_promotion, mean, sd)

> marks\_for\_promotion

[1] 49.19979