|  |  |
| --- | --- |
| 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 |
| 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) | Ratio |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| 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 | Interval |

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

n(A)= {HHT, THH, HTH} =3

n(S)= {HHH, HHT, HTT, TTT, THH, TTH, HTH, THT} =8

Required Probability P(A) =3/8

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

1. Equal to 1

0

1. Less than or equal to 4

n(s)=36

n(A)= {(1,1) (1,2) (1,3) (3,1) (2,2) (2,1)

P(A)=6/36

1. Sum is divisible by 2 and 3

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

P(A)=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?

(7C2) = 7!/2!\*(7-2)! = 21

(5C2) = 5!/2!\*(5-2)! = 10

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-

Exp. Value =**∑xP(x)** =1\*0.015+4\*0.20+3\*0.65+5\*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.

> df <- read.csv(file.choose())

> #Points

> df[,2]

[1] 3.90 3.90 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 3.92 3.07 3.07 3.07 2.93 3.00

[17] 3.23 4.08 4.93 4.22 3.70 2.76 3.15 3.73 3.08 4.08 4.43 3.77 4.22 3.62 3.54 4.11

> #Mean

> Points.mean <- mean(df[,2])

> Points.mean

[1] 3.596563

> #Median

> Points.median <- median(df[,2])

> Points.median

[1] 3.695

> #Mode

> mode <- function(x) {

+ ux <- unique(x)

+ ux[which.max(tabulate(match(x,ux)))]

+ }

> Points.mode <- mode(df[,2])

> Points.mode

[1] 3.92

> #Variance

> var(df[,2])

[1] 0.2858814

> #SD

> Points.Standard\_Deviation <- sd(df[,2])

> Points.Standard\_Deviation

[1] 0.5346787

> #Range

> Points.Range <- (max(df[,2])-min(df[,2]))

> Points.Range

[1] 2.17

> df <- read.csv(file.choose())

> #Score

> df[,3]

[1] 2.620 2.875 2.320 3.215 3.440 3.460 3.570 3.190 3.150 3.440 3.440 4.070 3.730

[14] 3.780 5.250 5.424 5.345 2.200 1.615 1.835 2.465 3.520 3.435 3.840 3.845 1.935

[27] 2.140 1.513 3.170 2.770 3.570 2.780

> #Mean

> Score.mean <- mean(df[,3])

> Score.mean

[1] 3.21725

> #Median

> Score.median <- median(df[,3])

> Score.median

[1] 3.325

> #Mode

> mode <- function(x) {

+ ux <- unique(x)

+ ux[which.max(tabulate(match(x,ux)))]

+ }

> Score.mode <- mode(df[,3])

> Score.mode

[1] 3.44

> #Variance

> var(df[,3])

[1] 0.957379

> #SD

> Score.Standard\_Deviation <- sd(df[,3])

> Score.Standard\_Deviation

[1] 0.9784574

> #Range

> Score.Range <- (max(df[,3])-min(df[,3]))

> Score.Range

[1] 3.911

> df <- read.csv(file.choose())

> #weigh

> df[,4]

[1] 16.46 17.02 18.61 19.44 17.02 20.22 15.84 20.00 22.90 18.30 18.90 17.40 17.60

[14] 18.00 17.98 17.82 17.42 19.47 18.52 19.90 20.01 16.87 17.30 15.41 17.05 18.90

[27] 16.70 16.90 14.50 15.50 14.60 18.60

> #Mean

> weigh.mean <- mean(df[,4])

> weigh.mean

[1] 17.84875

> #Median

> weigh.median <- median(df[,4])

> weigh.median

[1] 17.71

> #Mode

> mode <- function(x) {

+ ux <- unique(x)

+ ux[which.max(tabulate(match(x,ux)))]

+ }

> weigh.mode <- mode(df[,4])

> weigh.mode

[1] 17.02

> #Variance

> var(df[,4])

[1] 3.193166

> #SD

> weigh.Standard\_Deviation <- sd(df[,4])

> weigh.Standard\_Deviation

[1] 1.786943

> #Range

> weigh.Range <- (max(df[,4])-min(df[,4]))

> weigh.Range

[1] 8.4

Inference:-

* The mean is useful for identifying trends in the data because we can compare means over a time period to spot trends. The mean is the most common measure of central tendency.
* The **median** divides a sample of data in half; it is the middle score. The median is a useful statistic if we think our data have some extreme cases. The median is not impacted by extreme cases, but the mean is.



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?

> X <- c(108, 110, 123, 134, 135, 145, 167, 187, 199)

> Expected\_weigh <- mean(X)

> Expected\_weigh

[1] 145.3333

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

**Cars speed and distance**

**Q9A**

> df <- read.csv(file.choose())

> names(df)

[1] "Index" "speed" "dist"

> print(df)

Index speed dist

1 1 4 2

2 2 4 10

3 3 7 4

4 4 7 22

5 5 8 16

6 6 9 10

7 7 10 18

8 8 10 26

9 9 10 34

10 10 11 17

11 11 11 28

12 12 12 14

13 13 12 20

14 14 12 24

15 15 12 28

16 16 13 26

17 17 13 34

18 18 13 34

19 19 13 46

20 20 14 26

21 21 14 36

22 22 14 60

23 23 14 80

24 24 15 20

25 25 15 26

26 26 15 54

27 27 16 32

28 28 16 40

29 29 17 32

30 30 17 40

31 31 17 50

32 32 18 42

33 33 18 56

34 34 18 76

35 35 18 84

36 36 19 36

37 37 19 46

38 38 19 68

39 39 20 32

40 40 20 48

41 41 20 52

42 42 20 56

43 43 20 64

44 44 22 66

45 45 23 54

46 46 24 70

47 47 24 92

48 48 24 93

49 49 24 120

50 50 25 85

> #For Speed

> sp=df['speed']

> library(moments)

> skewness(sp)

speed

-0.1139548

> kurtosis(sp)

speed

2.422853

> #For Distance

> dt=df['dist']

> skewness(dt)

dist

0.7824835

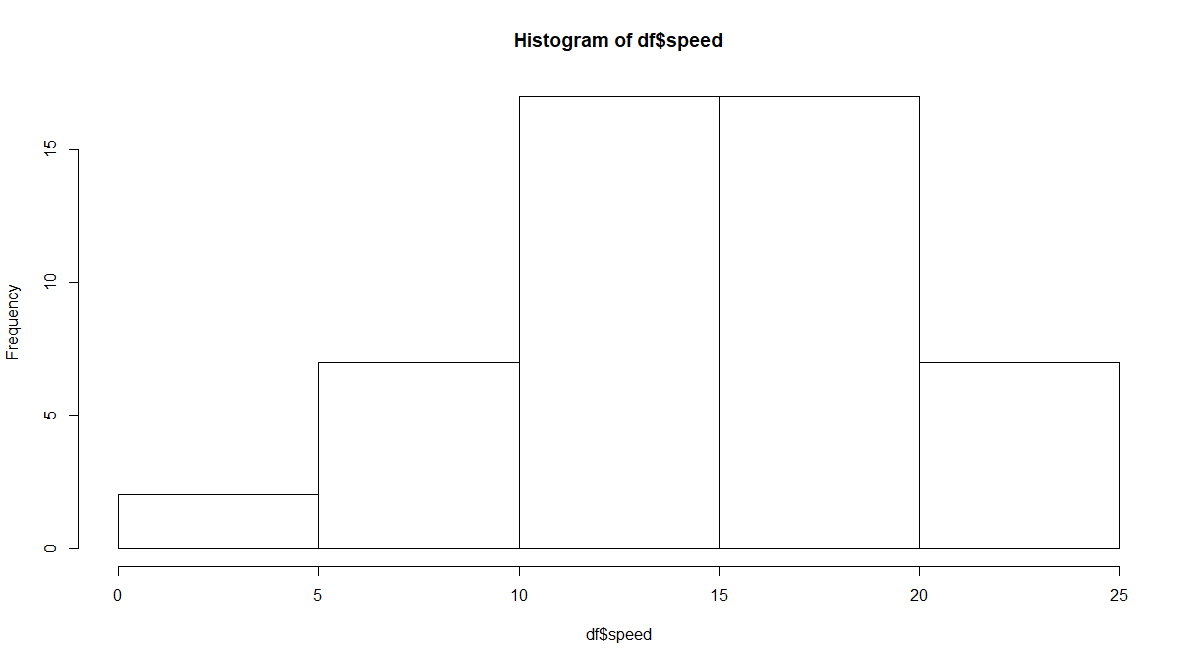
> kurtosis(dt)

dist

3.248019

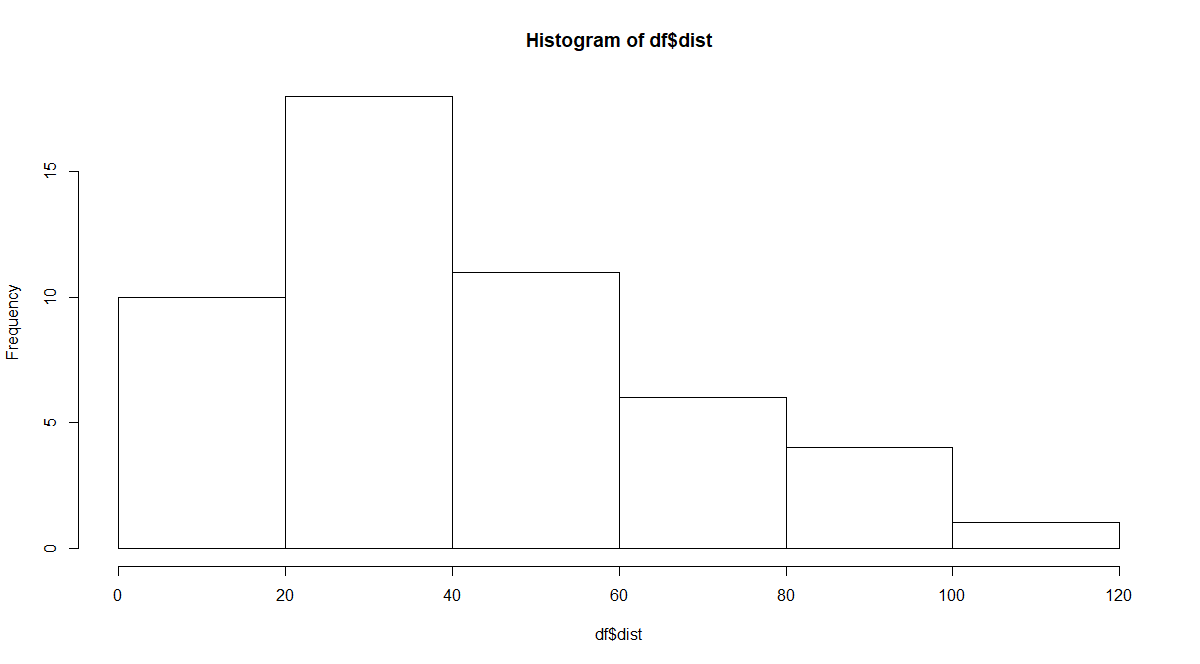
> #Draw Histogram

> hist(df$speed)



Negative Skewness as Distribution is skewed towards left. Kurtosis is Negative

> hist(df$dist)



Positive Skewness as Distribution is skewed towards Right. Positive Kurtosis



**SP and Weight(WT)**

> df <- read.csv(file.choose())

> names(df)

[1] "X" "SP" "WT"

> print(df)

X SP WT

1 1 104.18535 28.76206

2 2 105.46126 30.46683

3 3 105.46126 30.19360

4 4 113.46126 30.63211

5 5 104.46126 29.88915

6 6 113.18535 29.59177

7 7 105.46126 30.30848

8 8 102.59851 15.84776

9 9 102.59851 16.35948

10 10 115.64520 30.92015

11 11 111.18535 29.36334

12 12 117.59851 15.75353

13 13 122.10506 32.81359

14 14 111.18535 29.37844

15 15 108.18535 29.34728

16 16 111.18535 29.60453

17 17 114.36929 29.53578

18 18 117.59851 16.19412

19 19 114.36929 29.92939

20 20 118.47294 33.51697

21 21 119.10506 32.32465

22 22 110.84082 34.90821

23 23 120.28900 32.67583

24 24 113.82914 31.83712

25 25 119.18535 28.78173

26 26 114.59851 16.04317

27 27 120.76052 38.06282

28 28 119.10506 32.83507

29 29 99.56491 34.48321

30 30 121.84082 35.54936

31 31 113.48461 37.04235

32 32 112.28900 33.23436

33 33 119.92111 31.38004

34 34 121.39264 37.57329

35 35 111.28900 32.70164

36 36 115.01309 31.91122

37 37 114.09338 28.75400

38 38 116.90944 27.87992

39 39 116.90944 28.63050

40 40 128.46126 30.11543

41 41 116.39264 37.39252

42 42 115.74885 35.02718

43 43 117.46126 30.52743

44 44 114.09338 28.34398

45 45 114.38097 33.07863

46 46 117.10506 32.62192

47 47 118.20870 36.49862

48 48 116.47294 33.91006

49 49 127.90944 28.07060

50 50 118.28900 33.45847

51 51 118.28900 33.21395

52 52 118.28900 33.43671

53 53 120.40431 40.39816

54 54 143.39264 37.62069

55 55 135.39264 37.25439

56 56 126.40431 40.58907

57 57 110.46126 30.14754

58 58 118.28900 32.73452

59 59 112.64520 30.61528

60 60 115.57658 37.66287

61 61 130.20870 36.88815

62 62 117.66855 37.86041

63 63 126.04810 43.39099

64 64 125.31234 40.72283

65 65 128.12840 40.15948

66 66 126.59851 15.71286

67 67 132.48461 37.97996

68 68 133.68022 41.57397

69 69 133.31234 40.47204

70 70 158.30067 37.14173

71 71 164.59851 15.82306

72 72 133.41598 44.01314

73 73 133.14007 43.35312

74 74 124.71524 52.99775

75 75 121.86416 42.61870

76 76 132.86416 42.77822

77 77 169.59851 16.13295

78 78 150.57658 37.92311

79 79 151.59851 15.76963

80 80 167.94446 39.42310

81 81 139.84082 34.94861

> #For SP

> SP=df['SP']

> library(moments)

> skewness(SP)

SP

1.581454

> kurtosis(SP)

SP

5.723521

> #For Weight

> Weight=df['WT']

> skewness(Weight)

WT

-0.6033099

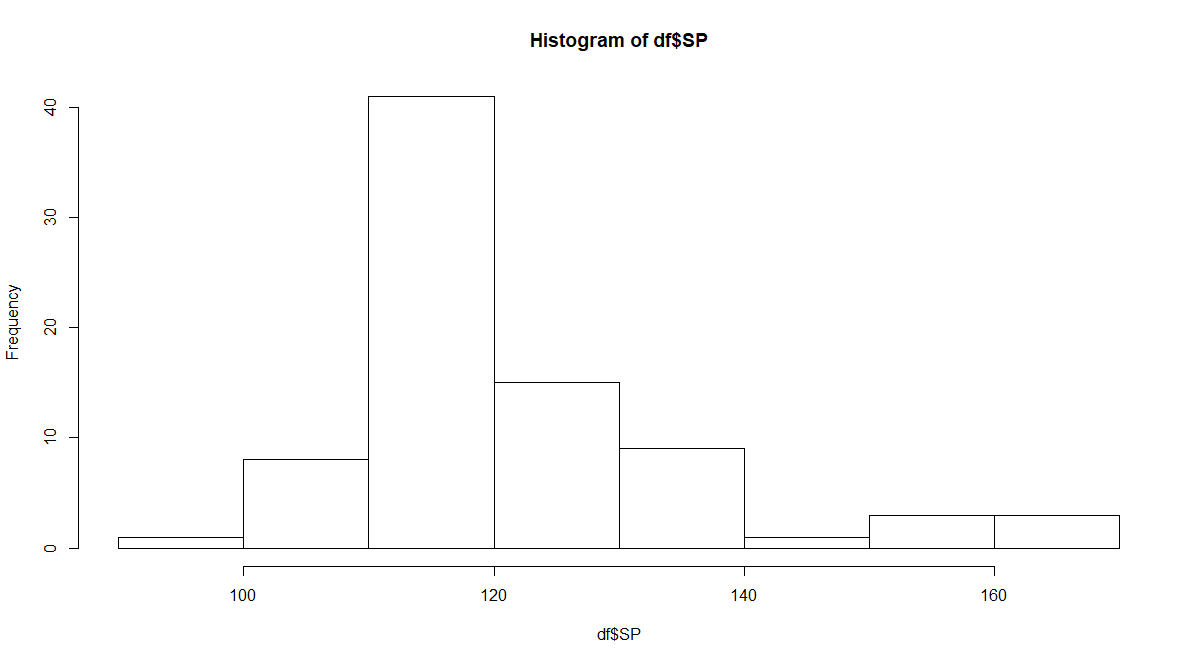
> kurtosis(Weight)

WT

3.819466

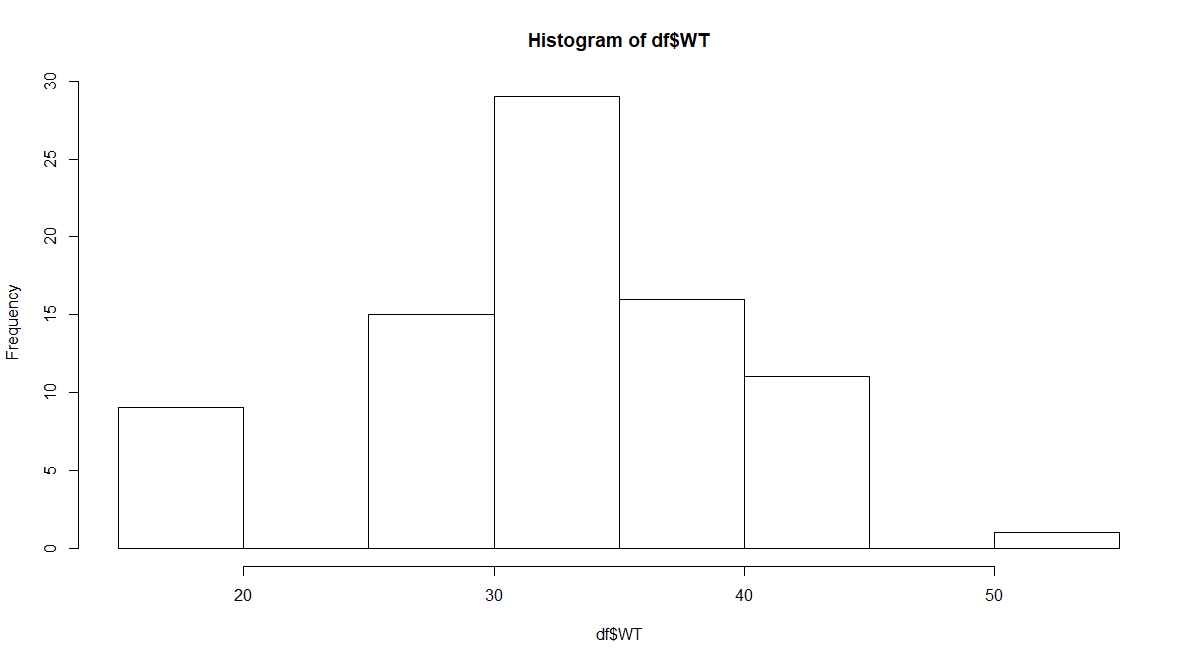
> #Histogram

> hist(df$SP)



**Positive Skewness. Positive Kurtosis**

> hist(df$WT)



**Negative Skewness. Positive Kurtosis**



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



=Distribution is Right Skewed

Mean>Median



= Distribution has 7 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 ?

> a <- 200 #Sample mean

> s <- 30

> n <- 2000 #sample size

> #alpha level

> (1-0.94)/2

[1] 0.03

> #step 2

> error <- qnorm(0.03)\*s/sqrt(n)

> error

[1] -1.261675

> left <- a-error

> right <- a+error

> left

[1] 201.2617

> right

[1] 198.7383

> #for 98%

> #at alpha level

> (1-0.98)/2

[1] 0.01

> error <- qnorm(0.01)\*s/sqrt(n)

> error

[1] -1.560562

> left <- a-error

> right <- a+error

> left

[1] 201.5606

> right

[1] 198.4394

> #for 96%

> #at alpha level

> (1-0.96)/2

[1] 0.02

> error <- qnorm(0.02)\*s/sqrt(n)

> error

[1] -1.377697

> left <- a-error

> right <- a+error

> left

[1] 201.3777

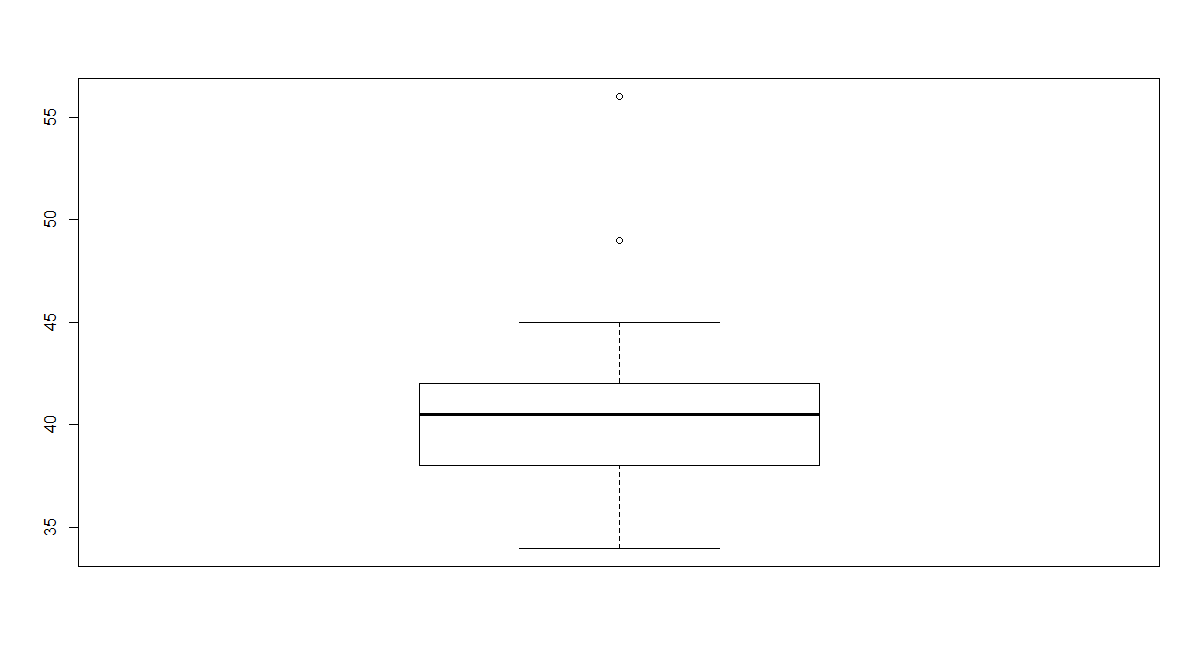
> right

[1] 198.6223

**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?
3. > x <- c(34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56)
4. > result.mean <- mean(x)
5. > result.mean
6. [1] 41
7. > result.median <- median(x)
8. > result.median
9. [1] 40.5
10. > result.variance <- var(x)
11. > result.variance
12. [1] 25.52941
13. > result.sd <- sd(x)
14. > result.sd
15. [1] 5.052664
16. > boxplot(x)

****

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

= Zero Skewed

Skewness=0,Symmetric

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

= Right Skewed Distribution

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

= Left Skewed Distribution

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

= A distribution with a positive kurtosis value indicates that the distribution has thick tails and a sharper peak than the normal distribution

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

= A distribution with a negative kurtosis value indicates that the distribution has lighter (Thin) tails and a flatter (Broad)peak than the normal distribution

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



What can we say about the distribution of the data?

= Asymmetrical Distribution

What is nature of skewness of the data?

= Left Skewed

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

= IQR = Q3-Q1=18-10=8

Q19) Comment on the below Boxplot visualizations?



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

= 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)

1. > df <- read.csv(file.choose())
2. > df
3. HP MPG VOL SP WT
4. 1 49 53.70068 89 104.18535 28.76206
5. 2 55 50.01340 92 105.46126 30.46683
6. 3 55 50.01340 92 105.46126 30.19360
7. 4 70 45.69632 92 113.46126 30.63211
8. 5 53 50.50423 92 104.46126 29.88915
9. 6 70 45.69632 89 113.18535 29.59177
10. 7 55 50.01340 92 105.46126 30.30848
11. 8 62 46.71655 50 102.59851 15.84776
12. 9 62 46.71655 50 102.59851 16.35948
13. 10 80 42.29908 94 115.64520 30.92015
14. 11 73 44.65283 89 111.18535 29.36334
15. 12 92 39.35409 50 117.59851 15.75353
16. 13 92 39.35409 99 122.10506 32.81359
17. 14 73 44.65283 89 111.18535 29.37844
18. 15 66 45.73489 89 108.18535 29.34728
19. 16 73 44.65283 89 111.18535 29.60453
20. 17 78 42.78991 91 114.36929 29.53578
21. 18 92 39.35409 50 117.59851 16.19412
22. 19 78 42.78991 91 114.36929 29.92939
23. 20 90 38.90183 103 118.47294 33.51697
24. 21 92 38.41100 99 119.10506 32.32465
25. 22 74 42.82848 107 110.84082 34.90821
26. 23 95 38.31061 101 120.28900 32.67583
27. 24 81 40.47472 96 113.82914 31.83712
28. 25 95 38.31061 89 119.18535 28.78173
29. 26 92 38.41100 50 114.59851 16.04317
30. 27 92 38.41100 117 120.76052 38.06282
31. 28 92 38.41100 99 119.10506 32.83507
32. 29 52 43.46943 104 99.56491 34.48321
33. 30 103 35.40419 107 121.84082 35.54936
34. 31 84 39.43124 114 113.48461 37.04235
35. 32 84 39.43124 101 112.28900 33.23436
36. 33 102 36.28546 97 119.92111 31.38004
37. 34 102 36.28546 113 121.39264 37.57329
38. 35 81 39.53163 101 111.28900 32.70164
39. 36 90 37.95874 98 115.01309 31.91122
40. 37 90 37.95874 88 114.09338 28.75400
41. 38 102 34.07067 86 116.90944 27.87992
42. 39 102 34.07067 86 116.90944 28.63050
43. 40 130 31.01413 92 128.46126 30.11543
44. 41 95 35.15273 113 116.39264 37.39252
45. 42 95 35.15273 106 115.74885 35.02718
46. 43 102 34.07067 92 117.46126 30.52743
47. 44 95 35.15273 88 114.09338 28.34398
48. 45 93 35.64356 102 114.38097 33.07863
49. 46 100 34.56150 99 117.10506 32.62192
50. 47 100 34.56150 111 118.20870 36.49862
51. 48 98 35.05233 103 116.47294 33.91006
52. 49 130 31.01413 86 127.90944 28.07060
53. 50 115 29.62994 101 118.28900 33.45847
54. 51 115 29.62994 101 118.28900 33.21395
55. 52 115 29.62994 101 118.28900 33.43671
56. 53 115 29.62994 124 120.40431 40.39816
57. 54 180 24.48737 113 143.39264 37.62069
58. 55 160 26.85228 113 135.39264 37.25439
59. 56 130 27.85625 124 126.40431 40.58907
60. 57 96 31.11358 92 110.46126 30.14754
61. 58 115 29.62994 101 118.28900 32.73452
62. 59 100 30.13192 94 112.64520 30.61528
63. 60 100 28.86023 115 115.57658 37.66287
64. 61 145 27.35427 111 130.20870 36.88815
65. 62 120 24.60913 116 117.66855 37.86041
66. 63 140 23.51592 131 126.04810 43.39099
67. 64 140 23.51592 123 125.31234 40.72283
68. 65 150 23.60516 121 128.12840 40.15948
69. 66 165 40.05000 50 126.59851 15.71286
70. 67 165 23.10317 114 132.48461 37.97996
71. 68 165 23.10317 127 133.68022 41.57397
72. 69 165 23.10317 123 133.31234 40.47204
73. 70 245 21.27371 112 158.30067 37.14173
74. 71 280 19.67851 50 164.59851 15.82306
75. 72 162 23.20357 135 133.41598 44.01314
76. 73 162 23.20357 132 133.14007 43.35312
77. 74 140 19.08634 160 124.71524 52.99775
78. 75 140 19.08634 129 121.86416 42.61870
79. 76 175 18.76284 129 132.86416 42.77822
80. 77 322 36.90000 50 169.59851 16.13295
81. 78 238 19.19789 115 150.57658 37.92311
82. 79 263 34.00000 50 151.59851 15.76963
83. 80 295 19.83373 119 167.94446 39.42310
84. 81 236 12.10126 107 139.84082 34.94861
85. > names(df)
86. [1] "HP" "MPG" "VOL" "SP" "WT"
87. > mpg <- df[,2]
88. > #p(mpg>38)
89. > pnorm(38,mean(mpg),sd(mpg),lower.tail = FALSE)
90. [1] 0.3475939
91. > #p(mpg>40)
92. > pnorm(40,mean(mpg),sd(mpg))
93. [1] 0.7293499
94. > #p(20<mpg<50)
95. > pnorm(20,mean(mpg),sd(mpg))-pnorm(50,mean(mpg),sd(mpg))
96. [1] -0.8988689

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

> df <- read.csv(file.choose())

> df

HP MPG VOL SP WT

1 49 53.70068 89 104.18535 28.76206

2 55 50.01340 92 105.46126 30.46683

3 55 50.01340 92 105.46126 30.19360

4 70 45.69632 92 113.46126 30.63211

5 53 50.50423 92 104.46126 29.88915

6 70 45.69632 89 113.18535 29.59177

7 55 50.01340 92 105.46126 30.30848

8 62 46.71655 50 102.59851 15.84776

9 62 46.71655 50 102.59851 16.35948

10 80 42.29908 94 115.64520 30.92015

11 73 44.65283 89 111.18535 29.36334

12 92 39.35409 50 117.59851 15.75353

13 92 39.35409 99 122.10506 32.81359

14 73 44.65283 89 111.18535 29.37844

15 66 45.73489 89 108.18535 29.34728

16 73 44.65283 89 111.18535 29.60453

17 78 42.78991 91 114.36929 29.53578

18 92 39.35409 50 117.59851 16.19412

19 78 42.78991 91 114.36929 29.92939

20 90 38.90183 103 118.47294 33.51697

21 92 38.41100 99 119.10506 32.32465

22 74 42.82848 107 110.84082 34.90821

23 95 38.31061 101 120.28900 32.67583

24 81 40.47472 96 113.82914 31.83712

25 95 38.31061 89 119.18535 28.78173

26 92 38.41100 50 114.59851 16.04317

27 92 38.41100 117 120.76052 38.06282

28 92 38.41100 99 119.10506 32.83507

29 52 43.46943 104 99.56491 34.48321

30 103 35.40419 107 121.84082 35.54936

31 84 39.43124 114 113.48461 37.04235

32 84 39.43124 101 112.28900 33.23436

33 102 36.28546 97 119.92111 31.38004

34 102 36.28546 113 121.39264 37.57329

35 81 39.53163 101 111.28900 32.70164

36 90 37.95874 98 115.01309 31.91122

37 90 37.95874 88 114.09338 28.75400

38 102 34.07067 86 116.90944 27.87992

39 102 34.07067 86 116.90944 28.63050

40 130 31.01413 92 128.46126 30.11543

41 95 35.15273 113 116.39264 37.39252

42 95 35.15273 106 115.74885 35.02718

43 102 34.07067 92 117.46126 30.52743

44 95 35.15273 88 114.09338 28.34398

45 93 35.64356 102 114.38097 33.07863

46 100 34.56150 99 117.10506 32.62192

47 100 34.56150 111 118.20870 36.49862

48 98 35.05233 103 116.47294 33.91006

49 130 31.01413 86 127.90944 28.07060

50 115 29.62994 101 118.28900 33.45847

51 115 29.62994 101 118.28900 33.21395

52 115 29.62994 101 118.28900 33.43671

53 115 29.62994 124 120.40431 40.39816

54 180 24.48737 113 143.39264 37.62069

55 160 26.85228 113 135.39264 37.25439

56 130 27.85625 124 126.40431 40.58907

57 96 31.11358 92 110.46126 30.14754

58 115 29.62994 101 118.28900 32.73452

59 100 30.13192 94 112.64520 30.61528

60 100 28.86023 115 115.57658 37.66287

61 145 27.35427 111 130.20870 36.88815

62 120 24.60913 116 117.66855 37.86041

63 140 23.51592 131 126.04810 43.39099

64 140 23.51592 123 125.31234 40.72283

65 150 23.60516 121 128.12840 40.15948

66 165 40.05000 50 126.59851 15.71286

67 165 23.10317 114 132.48461 37.97996

68 165 23.10317 127 133.68022 41.57397

69 165 23.10317 123 133.31234 40.47204

70 245 21.27371 112 158.30067 37.14173

71 280 19.67851 50 164.59851 15.82306

72 162 23.20357 135 133.41598 44.01314

73 162 23.20357 132 133.14007 43.35312

74 140 19.08634 160 124.71524 52.99775

75 140 19.08634 129 121.86416 42.61870

76 175 18.76284 129 132.86416 42.77822

77 322 36.90000 50 169.59851 16.13295

78 238 19.19789 115 150.57658 37.92311

79 263 34.00000 50 151.59851 15.76963

80 295 19.83373 119 167.94446 39.42310

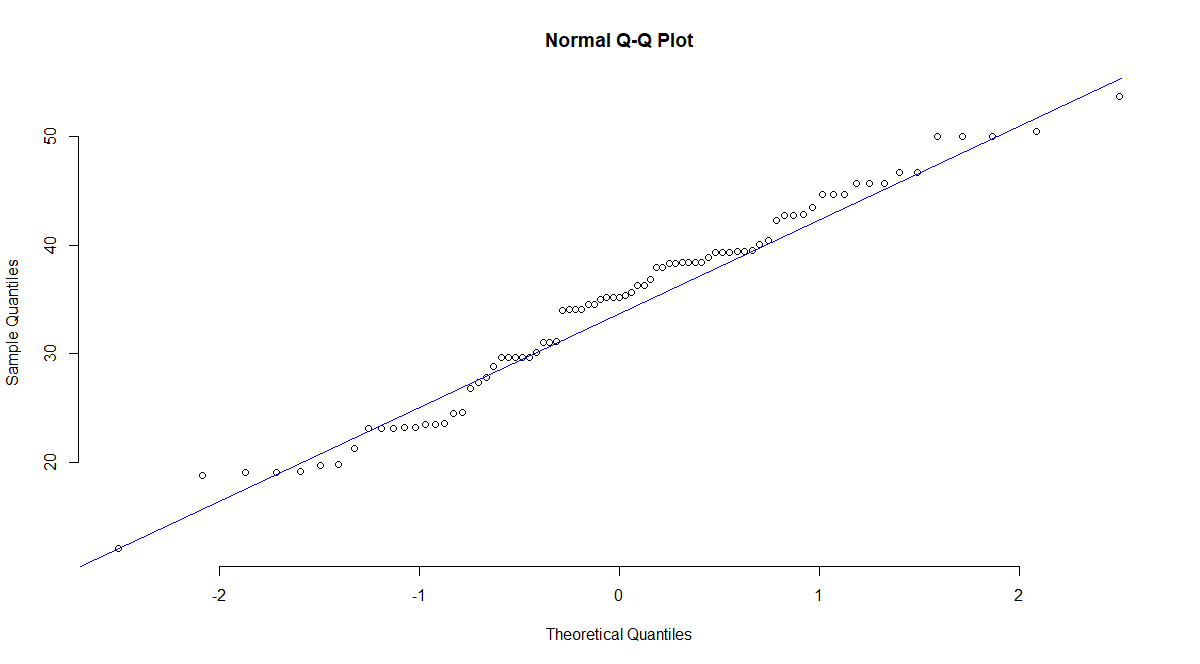
81 236 12.10126 107 139.84082 34.94861

> names(df)

[1] "HP" "MPG" "VOL" "SP" "WT"

> qqnorm(df$MPG, pch = 1 , frame = FALSE)

> qqline(df$MPG,col="Blue")



**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

> wc <- read.csv(file.choose())

> names(wc)

[1] "Waist" "AT"

> wc

Waist AT

1 74.75 25.72

2 72.60 25.89

3 81.80 42.60

4 83.95 42.80

5 74.65 29.84

6 71.85 21.68

7 80.90 29.08

8 83.40 32.98

9 63.50 11.44

10 73.20 32.22

11 71.90 28.32

12 75.00 43.86

13 73.10 38.21

14 79.00 42.48

15 77.00 30.96

16 68.85 55.78

17 75.95 43.78

18 74.15 33.41

19 73.80 43.35

20 75.90 29.31

21 76.85 36.60

22 80.90 40.25

23 79.90 35.43

24 89.20 60.09

25 82.00 45.84

26 92.00 70.40

27 86.60 83.45

28 80.50 84.30

29 86.00 78.89

30 82.50 64.75

31 83.50 72.56

32 88.10 89.31

33 90.80 78.94

34 89.40 83.55

35 102.00 127.00

36 94.50 121.00

37 91.00 107.00

38 103.00 129.00

39 80.00 74.02

40 79.00 55.48

41 83.50 73.13

42 76.00 50.50

43 80.50 50.88

44 86.50 140.00

45 83.00 96.54

46 107.10 118.00

47 94.30 107.00

48 94.50 123.00

49 79.70 65.92

50 79.30 81.29

51 89.80 111.00

52 83.80 90.73

53 85.20 133.00

54 75.50 41.90

55 78.40 41.71

56 78.60 58.16

57 87.80 88.85

58 86.30 155.00

59 85.50 70.77

60 83.70 75.08

61 77.60 57.05

62 84.90 99.73

63 79.80 27.96

64 108.30 123.00

65 119.60 90.41

66 119.90 106.00

67 96.50 144.00

68 105.50 121.00

69 105.00 97.13

70 107.00 166.00

71 107.00 87.99

72 101.00 154.00

73 97.00 100.00

74 100.00 123.00

75 108.00 217.00

76 100.00 140.00

77 103.00 109.00

78 104.00 127.00

79 106.00 112.00

80 109.00 192.00

81 103.50 132.00

82 110.00 126.00

83 110.00 153.00

84 112.00 158.00

85 108.50 183.00

86 104.00 184.00

87 111.00 121.00

88 108.50 159.00

89 121.00 245.00

90 109.00 137.00

91 97.50 165.00

92 105.50 152.00

93 98.00 181.00

94 94.50 80.95

95 97.00 137.00

96 105.00 125.00

97 106.00 241.00

98 99.00 134.00

99 91.00 150.00

100 102.50 198.00

101 106.00 151.00

102 109.10 229.00

103 115.00 253.00

104 101.00 188.00

105 100.10 124.00

106 93.30 62.20

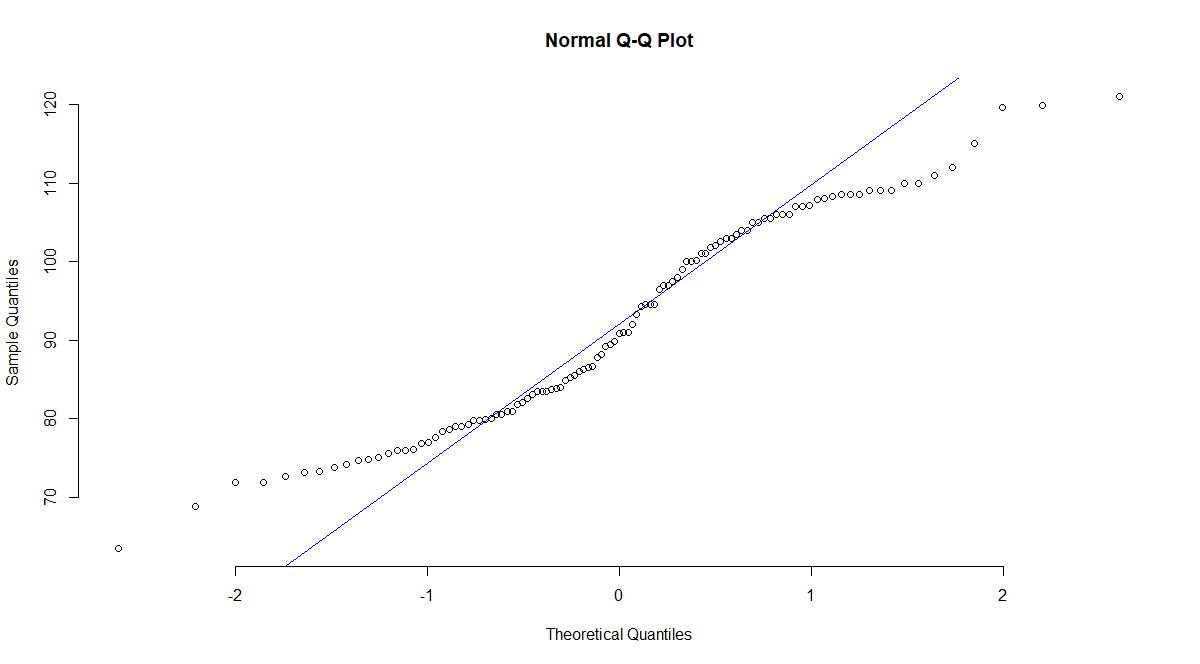
107 101.80 133.00

108 107.90 208.00

109 108.50 208.00

> qqnorm(wc$Waist, pch = 1, frame = FALSE)

> qqline(wc$Waist, col = "Blue")

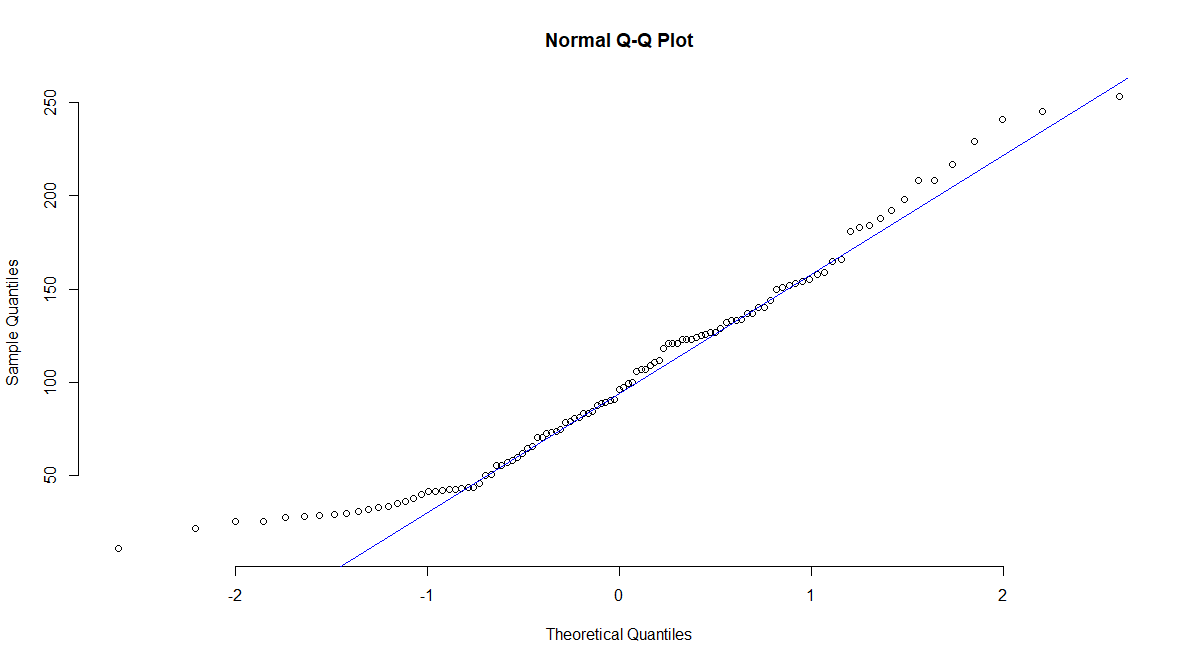


**Wc-at $waist follows Normal Distribution**

> #Q-Q Plot of AT

> qqnorm(wc$AT, pch = 1, frame = FALSE)

> qqline(wc$AT, col = "Blue")



**Wc-at $AT follows Normal Distribution**

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

> #For 90%

> p=0.90+0.05

> qnorm(0.95)

[1] 1.644854

> #For 94%

> p=0.94+0.03

> qnorm(0.97)

[1] 1.880794

> #For 60%

> p=0.60+0.2

> qnorm(0.8)

[1] 0.8416212

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

> #T scores

> #sample size = 25

> df=25-1

> #For 95%

> p=(1-0.95)/2 + 0.95

> qt(0.975,24)

[1] 2.063899

> #For 96%

> p=(1-0.96)/2 + 0.96

> qt(0.98,24)

[1] 2.171545

> #For 99%

> p=(1-0.99)/2 + 0.99

> 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

> x=260 #sample mean

> y=270 #Poplulation mean

> s=90 #sample standard deviation

> n=18

> df=18-1 #Degrees of freedom

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

> t= -(10)/21.2132 #tscore

> pt(-0.4714045,17)

[1] 0.3216725

**Probability is 32.16%**