

**project for stat 332**

**عهد سليمان المهيلب**

**434200796**

**Project for stat 332**

**For Muscle mass data:**

SSE= = 3874.448

The Variance of the residuals & Standard deviation:

At X=20

When X (predictor) increase by one point, Y (response) increase by 1,19.

**By using minitab**

Stat → Regression → Regression

Response → Y

Predictor → X

Choose the option storage to apply the residuals → ok →ok

**Regression Analysis: Y versus X**

The regression equation is

**Y = 156 - 1.19 X**

Predictor Coef SE Coef T P

Constant 156.347 5.512 28.36 0.000

**X -1.19000** 0.09020 -13.19 0.000

S = 8.17318 R-Sq = 75.0% R-Sq(adj) = 74.6%

Analysis of Variance

Source DF SS MS F P

Regression 1 11627 11627 174.06 0.000

Residual Error 58 **3874 67**

Total 59 15502

Unusual Observations

Obs X Y Fit SE Fit Residual St Resid

53 78.0 87.00 63.53 1.94 23.47 2.96R

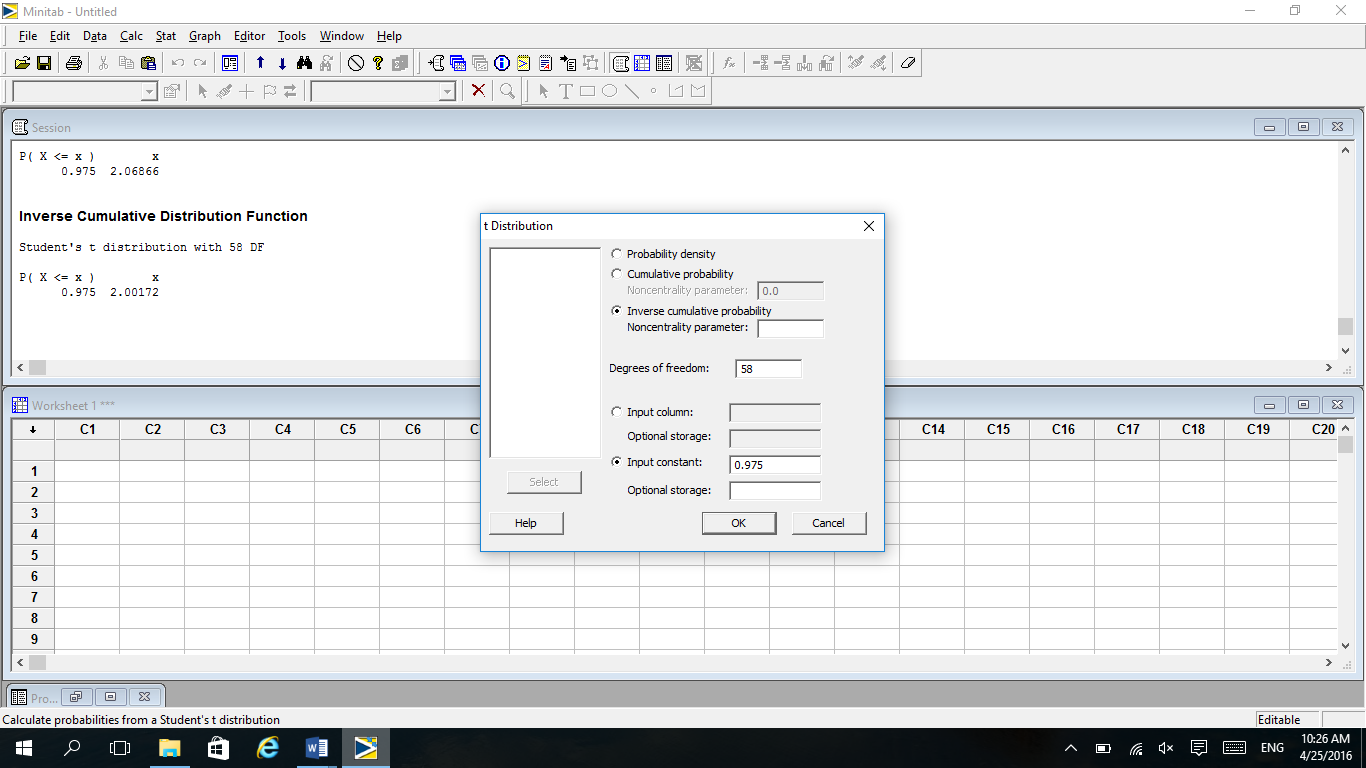
R denotes an observation with a large standardized residual.

To calculate the variance & S.D of :

Confidence Interval for :

95% C.I for :

Calculating by using minitab:

Calc → probability distribution → t…→ 

(-1.371, -1.0094 )

This mean that when y increases by one unite x increases from -1.371 to -1.0094

**Sampling Distribution of**

In Muscle mass data Test if

1. Hypothesis

2. Test statistic

3. Decision: Reject if , 2.002

Then reject

To calculate the variance & S.D of :

Confidence Interval for :

95% C.I for :

(145.311,167.382)

This mean that from 145.311 to 167.382 of X doesn’t depend on y

**Sampling Distribution of**

In Muscle mass data Test if

1. Hypothesis

2. Test statistic

3. Decision: Reject if , 2.002

Then reject

**General Regression Analysis: Y versus X**

Regression Equation

Y = 156.347 - 1.19 X

Coefficients

Term Coef SE Coef T P 95% CI

Constant 156.347 5.51226 **28.3634** 0.000 **(145.313, 167.381)**

X -1.190 0.09020 **-13.1933** 0.000 **( -1.371, -1.009)**

Summary of Model

S = 8.17318 R-Sq = 75.01% R-Sq(adj) = 74.58%

PRESS = 4170.23 R-Sq(pred) = 73.10%

Analysis of Variance

Source DF Seq SS Adj SS Adj MS F P

Regression 1 11627.5 11627.5 11627.5 174.062 0.000000

X 1 11627.5 11627.5 11627.5 174.062 0.000000

Error 58 3874.4 3874.4 66.8

Lack-of-Fit 30 2024.8 2024.8 67.5 1.022 0.478882

Pure Error 28 1849.7 1849.7 66.1

Total 59 15501.9

Fits and Diagnostics for Unusual Observations

Obs Y Fit SE Fit Residual St Resid

53 87 63.5269 1.93756 23.4731 2.95624 R

R denotes an observation with a large standardized residual.

ANOVA test for linearty:

Refare to the data:

SSE= = 3874.448

ANOVA TABLE

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Source of Variation | d.f | SS | MS | F |
| Regression | 1 | SSR=11627.57 |  | 174.063 |
| Error | 58 | SSE= |  |
| Total | 59 | SSTo= 15502.02 |  |  |

1. Hypothesis

(Non liner)

1. Test statistic

= 174.063

1. Decision: Reject if

*4.00687*

Then Reject

In the data find

This mean that 75.01% of the reduction in the total error of Y due to X.

**Regression Analysis: Y versus X**

The regression equation is

Y = 156 - 1.19 X

Predictor Coef SE Coef T P

Constant 156.347 5.512 28.36 0.000

X -1.19000 0.09020 -13.19 0.000

S = 8.17318 **R-Sq = 75.0%** R-Sq(adj) = 74.6%

Prediction for the expected response:

At X=24

(120.958 , 134.616)

The expected of Y when X=24 is from 120.958 to 134.616.

Prediction for new response:

By using the data calculate 95% PI for Y at X=45.

(86.086 , 119.507)

At X=45 Y can be predicted from 86.086 to 119.507

**Regression Analysis: Y versus X**

The regression equation is

Y = 156 - 1.19 X

Predictor Coef SE Coef T P

Constant 156.347 5.512 28.36 0.000

X -1.19000 0.09020 -13.19 0.000

S = 8.17318 R-Sq = 75.0% R-Sq(adj) = 74.6%

Analysis of Variance

Source DF SS MS F P

Regression 1 11627 11627 174.06 0.000

Residual Error 58 3874 67

Total 59 15502

Unusual Observations

Obs X Y Fit SE Fit Residual St Resid

53 78.0 87.00 63.53 1.94 23.47 2.96R

R denotes an observation with a large standardized residual.

Predicted Values for New Observations

New

Obs Fit SE Fit 95% CI 95% PI

1 102.80 1.71 (99.36, 106.23) **(86.08, 119.51)**

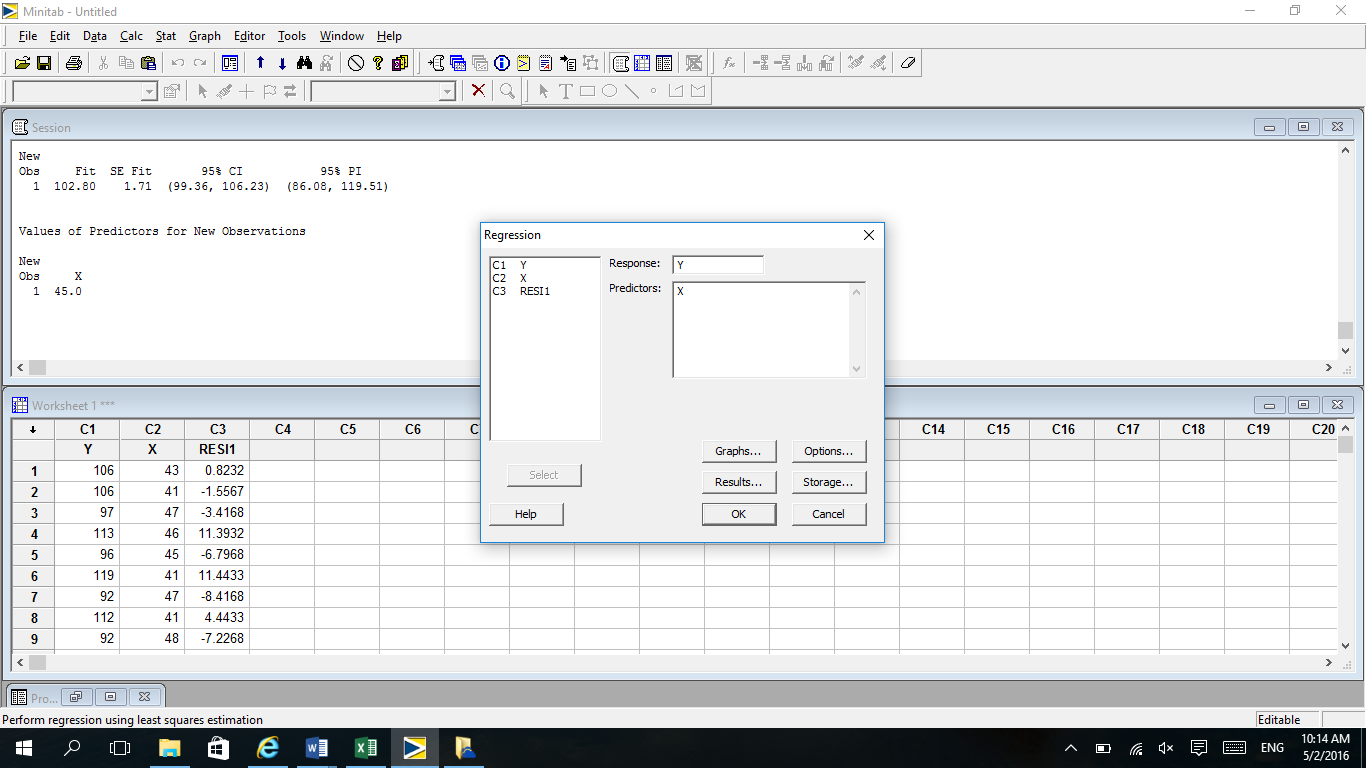
Values of Predictors for New Observations

New

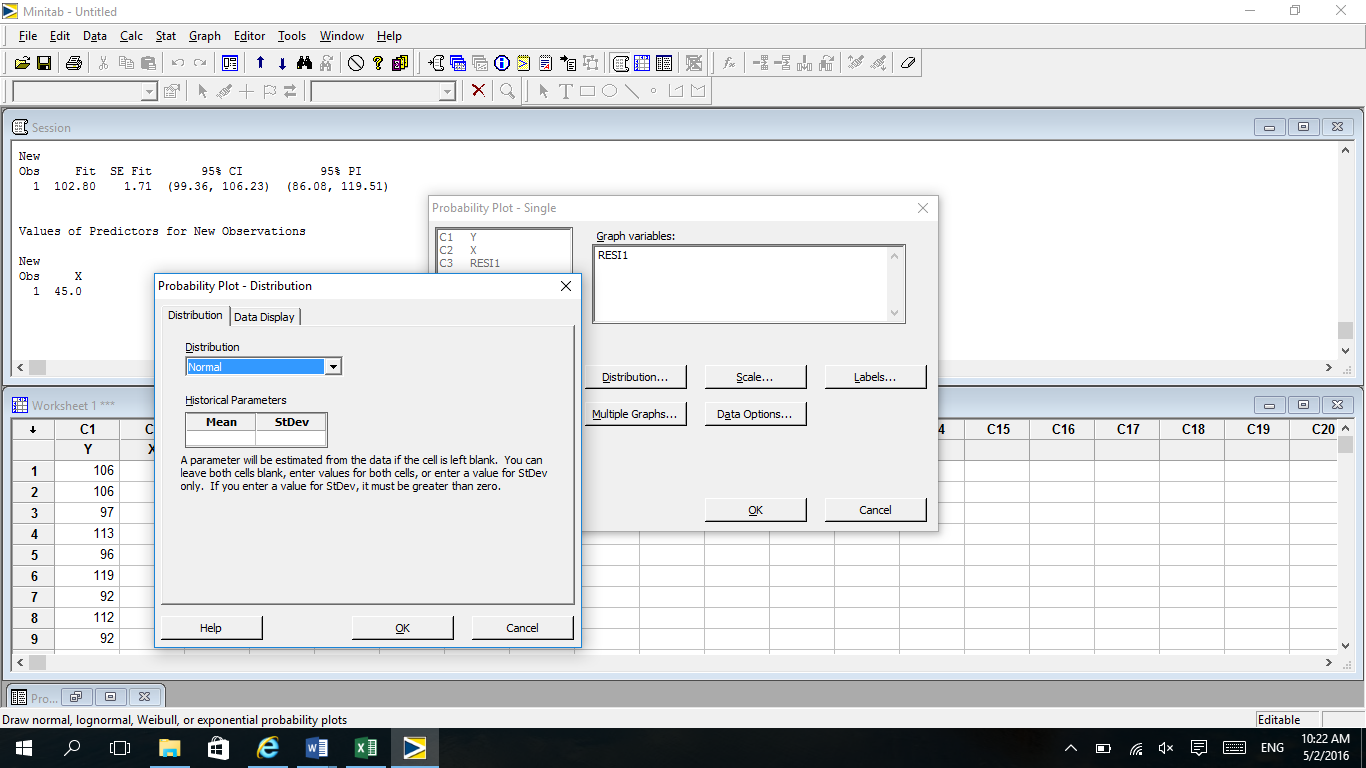
Obs X

1 45.0

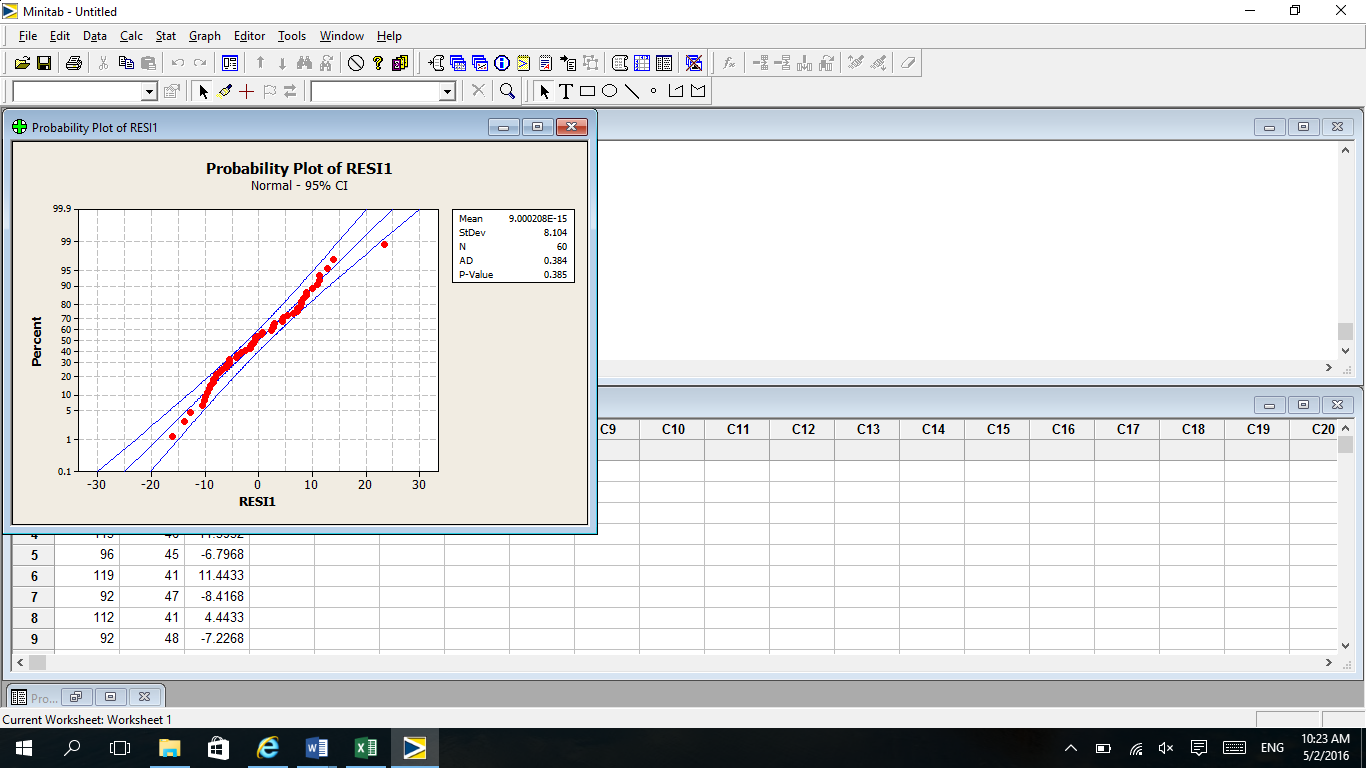
Residual analysis:

stat→ regression → regression → 

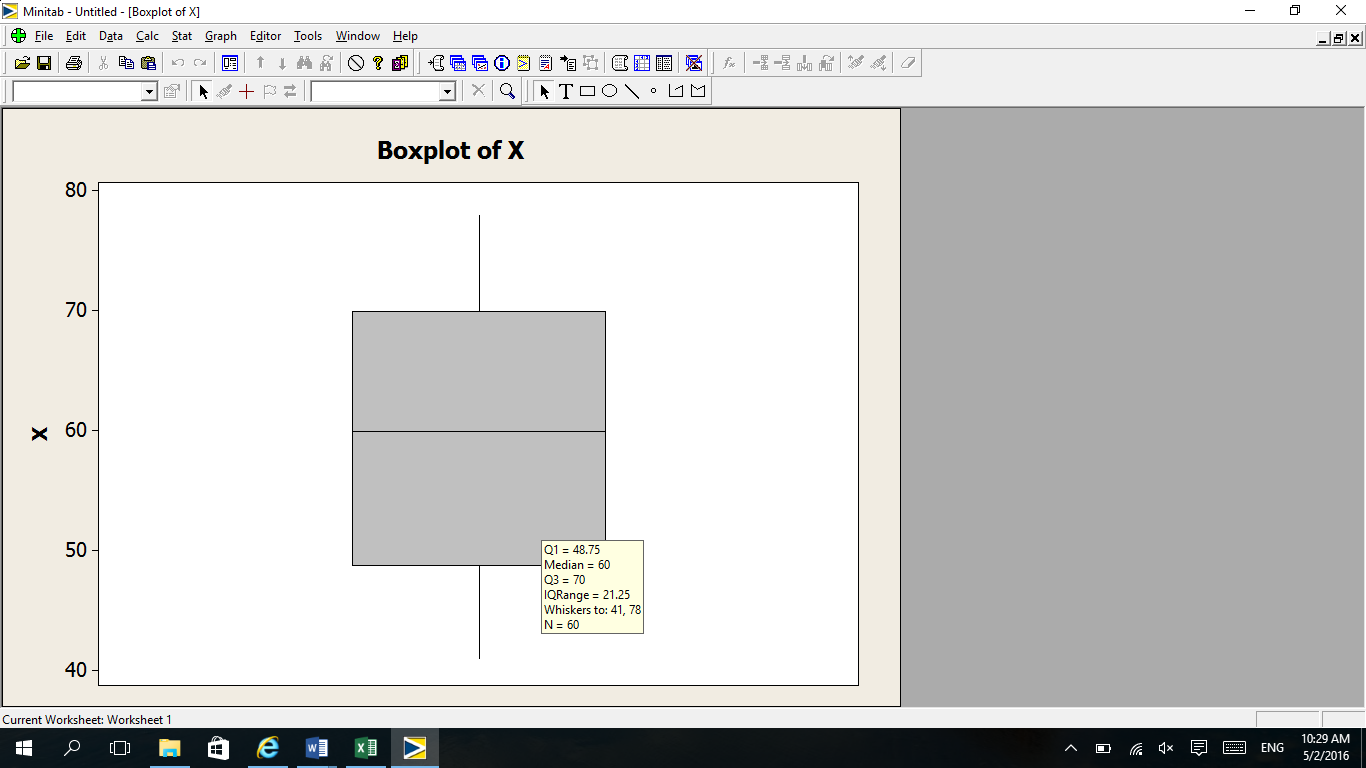
to apply the residuals from storage →●residuals →ok →ok

graph → probability plot → single →

→ ok



Diagnostics and remedial measures:



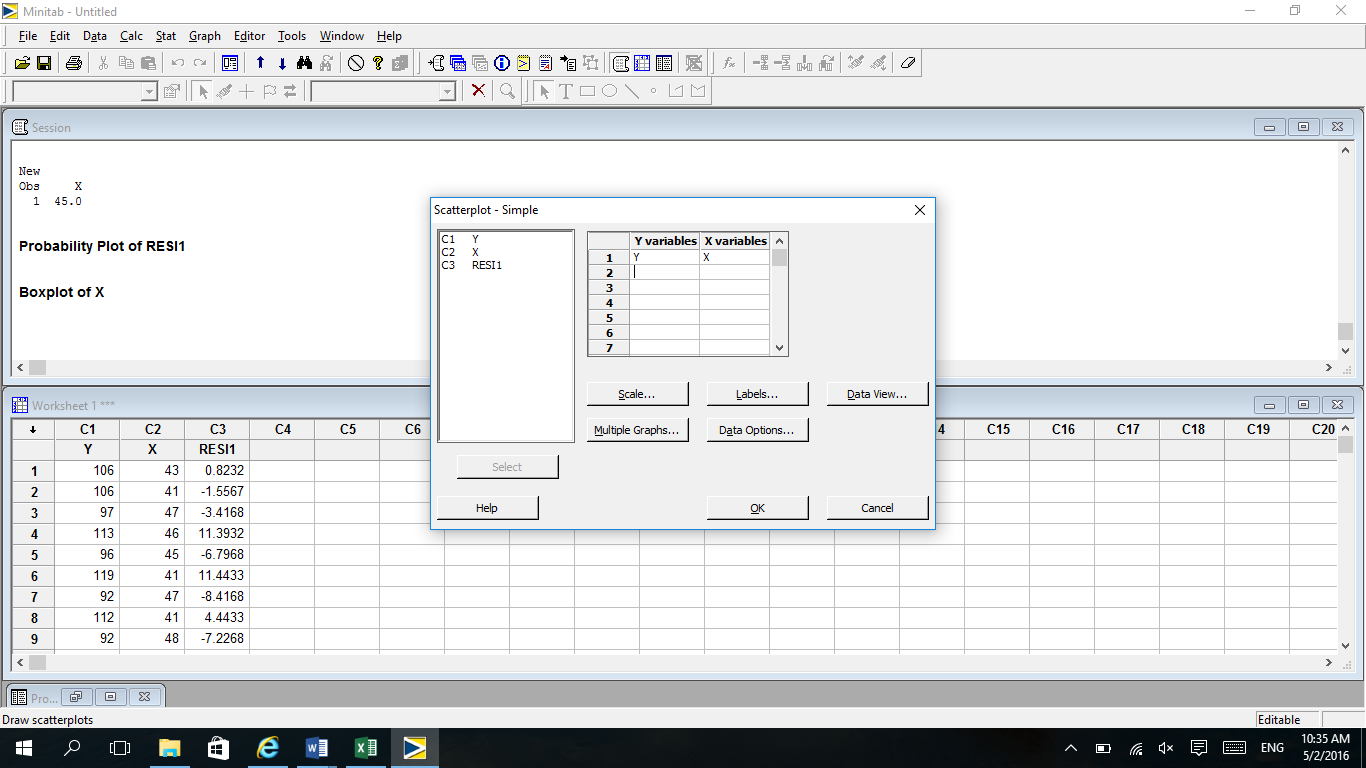
Q1=48.75

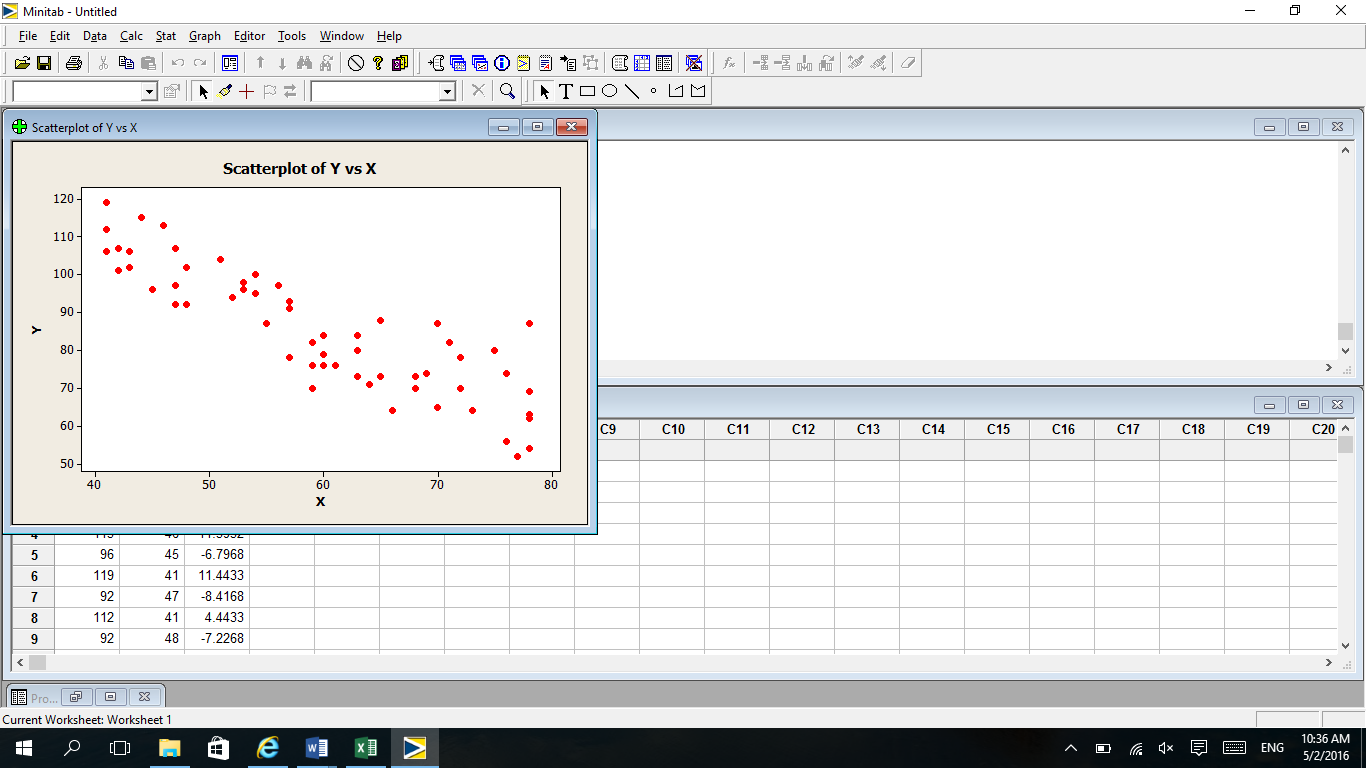
Q2 or median =60

Q3=70

1/ linearity problem:

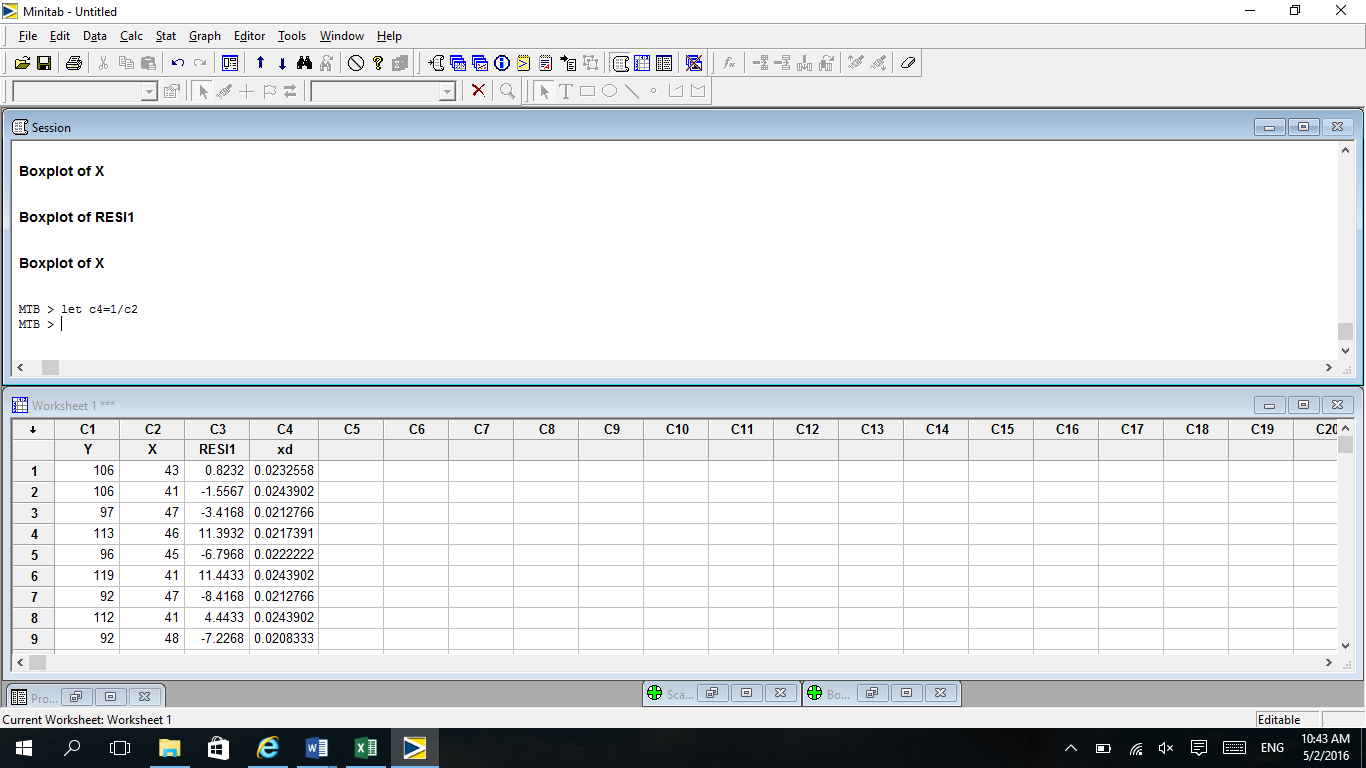
Graph → scatter plot →simple →ok

 → ok

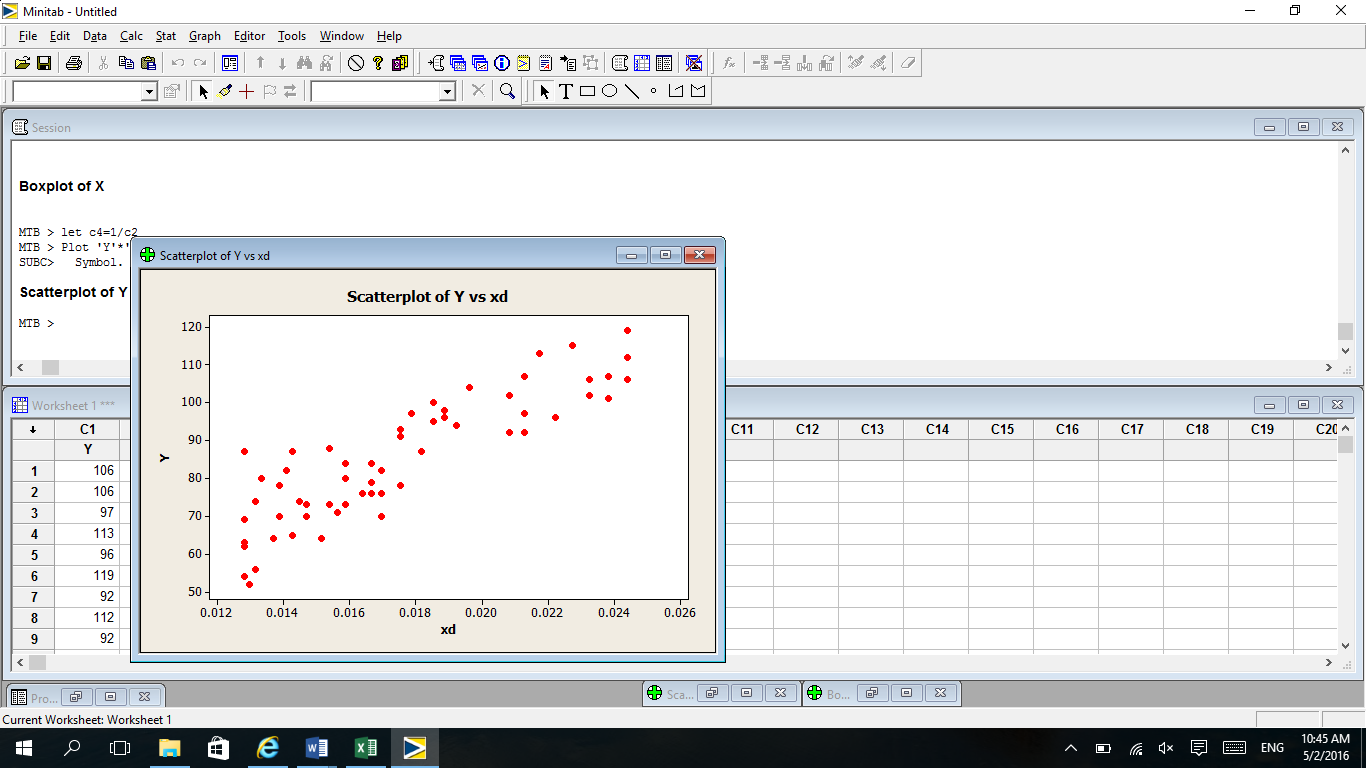


Non linear

The transformation x’=



MTB > let c4=1/c2



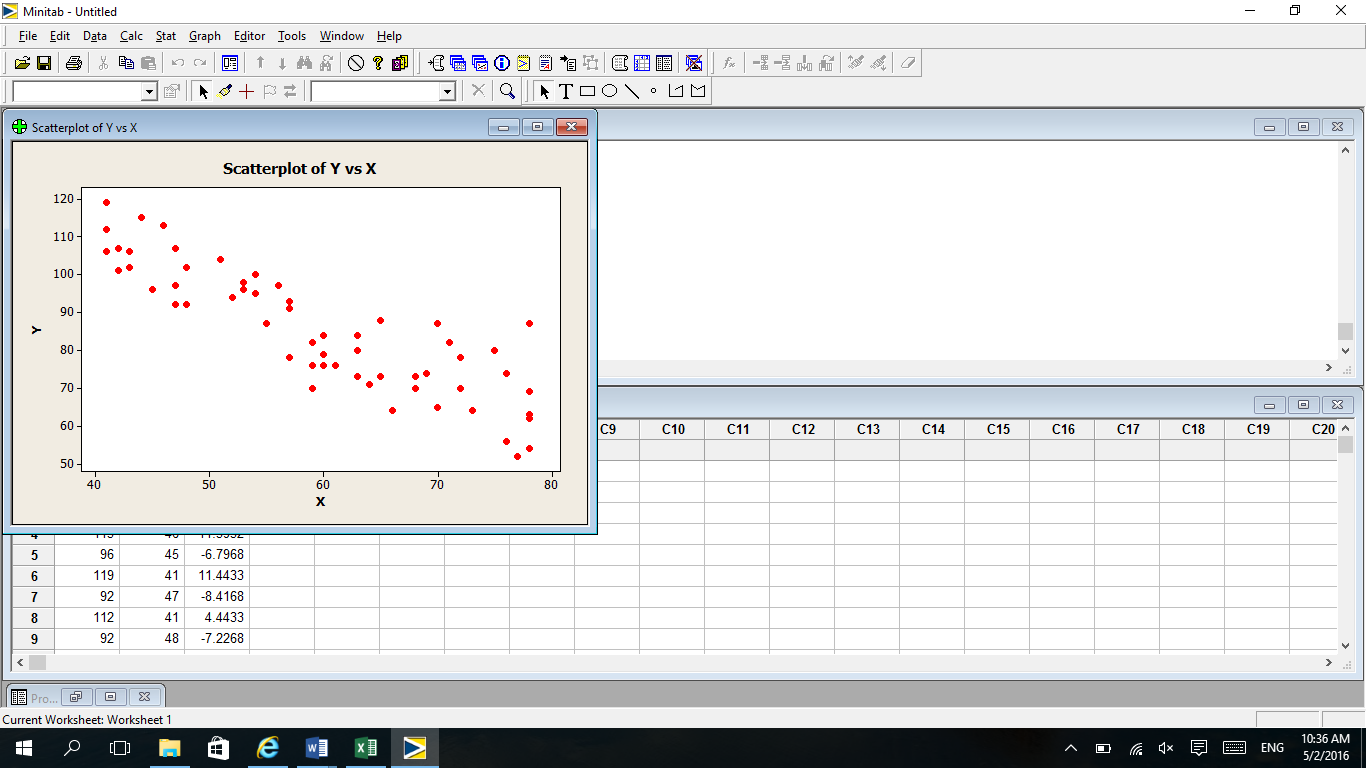
Model is linear

The estimated model is:

Calculating Y when X= 3

2/ variance is not fixed for each

Graph → scatter plot →simple →ok



yd=y’=logy

**Regression Analysis: yd versus X**

The regression equation is

yd = 2.29 - 0.00617 X

y’=2.29-0.00617X

log(y)= 2.29-0.00617X

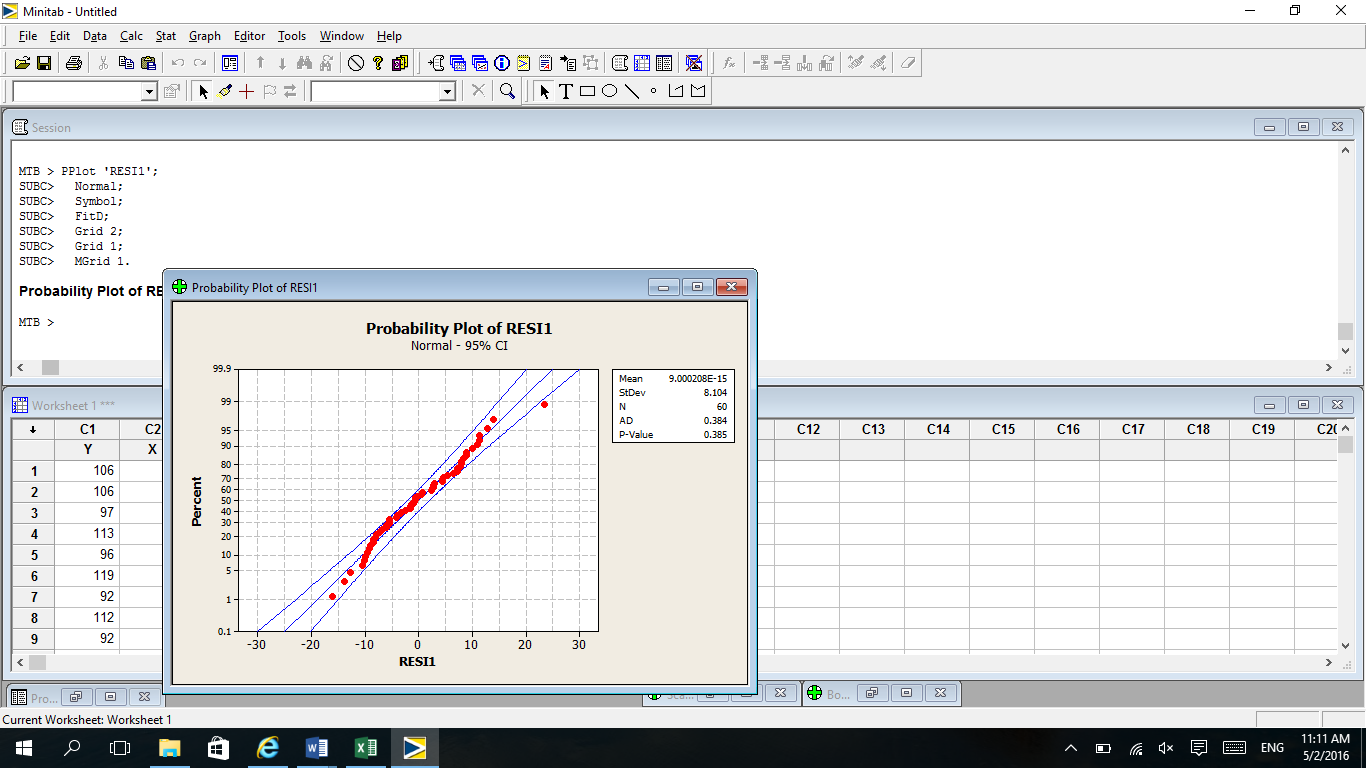
calculating y when x=8

log(y)= 2.29-0.00617(8)=2.24064

y=174.0364

Normality problem:

Test for normality “Anderson test”:



1. P-value =0.385
2. Accept

Simple linear regression in matrix form:

MTB > copy c3 c2 m1

MTB > Transpose M1 m2.

MTB > Multiply M2 M1 m3.

MTB > print m3

**Data Display**

Matrix M3

60 3599

3599 224091

MTB > Invert M3 m4.

MTB > print m4

**Data Display**

Matrix M4

0.454860 -0.0073053

-0.007305 0.0001218

MTB > copy c1 m5

MTB > Multiply M2 M5 m6.

MTB > print m6

**Data Display**

Matrix M6

5098

296024

MTB > Multiply M4 M6 m7.

MTB > print m7

**Data Display**

Matrix M7

156.347

-1.190

MTB > Multiply 66.8008 m4 m8.

MTB > print m8

**Data Display**

Matrix M8

30.3850 -0.487997

-0.4880 0.008136

Confidence Interval for :

95% C.I for :

(145.311,167.382)

Confidence Interval for :

95% C.I for :

(-1.371, -1.0094 )

**For Brand preference data:**

Multiple linear regression model:

MTB > copy c4 c2 c3 m1

MTB > Transpose M1 m2.

MTB > Multiply M2 M1 m3.

MTB > print m3

**Data Display**

Matrix M3

16 112 48

112 864 336

48 336 160

MTB > Invert M3 m4.

MTB > print m4

**Data Display**

Matrix M4

1.2375 -0.0875 -0.1875

-0.0875 0.0125 -0.0000

-0.1875 0.0000 0.0625

MTB > copy c1 m5

MTB > Multiply M2 M5 m6.

MTB > print m6

**Data Display**

Matrix M6

1308

9510

3994

MTB > Multiply M4 M6 m7.

MTB > print m7

**Data Display**

Matrix M7

37.650

4.425

4.375

There is 37.65 from y doesn’t depend on X1 and X2.

If X1 increase by 1 unite and X2 fixed, y will increase by 4.425.

If X2 increase by 1 unite and X1 fixed, y will increase by 4.357

MTB > copy c5 m8

MTB > Transpose M8 m9.

MTB > Multiply M9 M8 m10.

Answer = 94.3000

MTB > Multiply 7.254 M4 m11.

MTB > print m11

**Data Display**

Matrix M11

8.97682 -0.634725 -1.36013

-0.63472 0.090675 -0.00000

-1.36013 0.000000 0.45338

Confidence Interval for

95% C.I:

(31.1772 , 44.1228)

The expected of y doesn’t depend on X1 and X2.

(3.775 , 5.0755)

If X1 increase by 1 unite y increase from 3.775 to 5.0775 with fixed X2.

(2.9024 , 5.8116)

If X2 increase by 1 unite y increase from 2.9024 to 5.8116with fixed X1.

**Regression Analysis: Y versus X1, X2**

The regression equation is

Y = 37.7 + 4.42 X1 + 4.37 X2

Predictor Coef SE Coef T P

Constant 37.650 2.996 12.57 0.000

X1 4.4250 0.3011 14.70 0.000

X2 4.3750 0.6733 6.50 0.000

S = 2.69330 R-Sq = 95.2% R-Sq(adj) = 94.5%

Analysis of Variance

Source DF SS MS F P

Regression 2 1872.70 936.35 129.08 0.000

Residual Error 13 94.30 7.25

Total 15 1967.00

Source DF Seq SS

X1 1 1566.45

X2 1 306.25

1. Hypothesis

2. Test statistic

3. Decision: Reject if

p-value=

Reject if

In Brand preference data Test if

1. Hypothesis

2. Test statistic

3. Decision: Reject if

reject

In Brand preference data Test if

1. Hypothesis

2. Test statistic

3. Decision: Reject if

reject

In Brand preference data Test if

1. Hypothesis

2. Test statistic

3. Decision: Reject if

reject

**Regression Analysis: Y versus X1, X2**

The regression equation is

Y = 37.7 + 4.42 X1 + 4.37 X2

Predictor Coef SE Coef T P

Constant 37.650 2.996 **12.57** 0.000

X1 4.4250 0.3011 **14.70** 0.000

X2 4.3750 0.6733 **6.50** 0.000

S = 2.69330 R-Sq = 95.2% R-Sq(adj) = 94.5%

Analysis of variance in the multiple linear regression model:

ANOVA

MTB > Transpose M5 m12.

MTB > Multiply M12 M5 m13.

Answer = 108896.0000

MTB > copy c8 c9 c10 c11 c12 c13 c14 c15 c16 c17 c18 c19 c20 c21 c22 c23 m14

MTB > Multiply M12 M14 m15.

MTB > Multiply M15 M5 m16.

Answer = 1710864.0000

MTB > Transpose M7 m17.

MTB > Multiply M17 M2 m18.

MTB > Multiply M18 M5 m19.

Answer = 108801.7000

ANOVA TABLE

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Source of Variation | d.f | SS | MS | F |
| Regression | 2 | SSR=1872.7 |  | 129.0805 |
| Error | 13 | SSE= |  |
| Total | 15 | SSToT=1967 |  |  |

1. Hypothesis

2. Test statistic

3.Decision: Reject if

Then reject

Analysis of Variance

Source DF SS MS F P

Regression 2 1872.70 936.35 129.08 0.000

Residual Error 13 94.30 7.25

Total 15 1967.00

p-value approach:

1. Hypothesis

2. p-value = 0.000 < 0,05

3. Reject

Confidence interval for the expected of response:

Prediction of new observation

Find 95%C.I for and when x1=12 and x2=15

Predicted Values for New Observations

New

Obs Fit SE Fit 95% CI 95% PI

1 156.375 8.247 (**138.560, 174.190**) (**137.633, 175.117**)XX

XX denotes a point that is an extreme outlier in the predictors

MTB > copy c7 m20

MTB > print m20

**Data Display**

Matrix M20

1

12

15

MTB > Transpose m20 m21.

MTB > Multiply m21 m11 m22.

MTB > Multiply m22 m20 m23.

Answer = 68.0063

(138.2299 , 173.98)

(137.363 , 174.847)