Chapter 1

Q1.19)

Grade point average. The director of admissions of a small college selected 120 students at random from the new freshman class in a study to determine whether a student's grade point average (OPA) at the end of the freshman year (Y) can be predicted from the ACT test score *(X)*. The results of the study follow. Assume that first-order regression model (1.1) is appropriate.

a. Obtain the least squares estimates of $\boldsymbol{\beta_0}$ and $\boldsymbol{\beta_1}$, and state the estimated regression function.
b. Plot the estimated regression function and the data."Does the estimated regression function appear to fit the data well?
c. Obtain a point estimate of the mean freshman OPA for students with ACT test score *X* = 30.
d. What is the point estimate of the change in the mean response when the entrance test score increases by one point?

Solution:

$\bar{X} = 24.725, \bar{Y} = 3.07405$

$$\sum_{i=1}^{n=120} (X_i - \bar{X})\,(Y_i - \bar{Y}) = 92.40565$$

$$\sum_{i=1}^{n=120} (X_i - \bar{X})^2 = 2379.925$$

$$\sum_{i=1}^{n=120} (Y_i - \bar{Y})^2 = 49.40545$$

$$b_1 = \widehat{\beta_1} = \sum_{i=1}^{n=120} \frac{(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n=120}(X_i - \bar{X})^2} = \frac{92.40565}{2379.925} = 0.038827$$

$$b_0 = \widehat{\beta_0} = \bar{Y} - b_1\bar{X} = 3.07405 - 0.038827 * 24.725 = 2.114049$$

$$\hat{Y} = 2.114 + 0.0388\,X$$

At X=30

$\widehat{Y_h} = 2.114 + 0.0388\ (30) = 3.278863$

when the entrance test score increases by one point, the mean response increase by 0.038827.

Q1.20)

Copier maintenance. The Tri-City Office Equipment Corporation sells an imported copier on a franchise basis and performs preventive maintenance and repair service on this copier. The data below have been collected from 45 recent calls on users to perform routine preventive maintenance service; for each call, X is the number of copiers serviced and *Y* is the total number of minutes spent by the service person. Assume that first-order regression model (1.1) is appropriate.
(مصنع يعمل على الصنعة الوقائية)
X هو عدد الناسخات الخدمات
Y هو العدد الإجمالي للدقائق التي يقضيها الشخص الخدمة.

a. Obtain the estimated regression function.
b. Plot the estimated regression function and the data. How well does the estimated regression function fit the data?
c. Interpret b_0 in your estimated regression function. Does b_0 provide any relevant information here? Explain.
d. Obtiun a poim estimate of the mean service time when X = 5 copiers are serviced.

Solution:

$\bar{X} = 5.11111, \bar{Y} = 76.26667$

$$\sum_{i=1}^{n=120} (X_i - \bar{X})(Y_i - \bar{Y}) = 5118.667$$

$$\sum_{i=1}^{n=120} (X_i - \bar{X})^2 = 340.4444$$

$$\sum_{i=1}^{n=120} (Y_i - \bar{Y})^2 = 80376.8$$

$$b_1 = \widehat{\beta_1} = \sum_{i=1}^{n=120} \frac{(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n=120}(X_i - \bar{X})^2} = 15.03525$$

$$b_0 = \widehat{\beta_0} = \bar{Y} - b_1\bar{X} = -0.58016$$

$$\hat{Y} = -0.58016 + 15.03525\, X$$

At X=5

$$\widehat{Y_h} = -0.58016 + 15.03525\,(5) = 74.59608$$

Q1.21) (H.W)

Airfreight breakage. A substance used in biological and medical research is shipped by airfreight to users in cartons of 1,000 ampules. The data below, involving 10 shipments, were collected on the number of times the carton was transferred from one aircraft to another over the shipment route (X) and the number of ampules found to be broken upon arrival (Y). Assume that first-order regression model (1.1) is appropriate.

a. Obtain the estimated regression function. Plot the estimated regression function and the data. Does a linear regression function appear to give a good fit here?
b. Obtain a point estimate of the expected number of broken ampules when X = 1 transfer is made.
c. Estimate the increase in the expected number of ampules broken when there are 2 transfers as compared to 1 transfer.
d. Verify that your fitted regression line goes through the point $(\bar{X}, \bar{Y})$.

Q1.22)

Plastic hardness. Refer to Problems 1.3 and 1.14. Sixteen batches of the plastic were made, and from each batch one test item was molded. Each test item was randomly assigned to one of the four predetermined time levels, and the hardness was measured after the assigned elapsed time. The results are shown below; X is the elapsed time in hours? and *Y* is hardness in Brinell units. Assume that first-order regression model (1.1) is appropria'te.

a. Obtain the estimated regression function. Plot the estimated regression function and the data. Does a linear regression function appear to give a good fit here?
b. Obtain a point estimate of the mean hardness when X = 40 hours.
c. Obtain a point estimate of the change in mean hardness when X increases by 1 hour.

Solution:

$\bar{X} = 28, \bar{Y} = 225.5625$

$$\sum_{i=1}^{n=120} (X_i - \bar{X})(Y_i - \bar{Y}) = 2604$$

$$\sum_{i=1}^{n=120} (X_i - \bar{X})^2 = 1280$$

$$\sum_{i=1}^{n=120} (Y_i - \bar{Y})^2 = 5443.938$$

$$b_1 = \widehat{\beta_1} = \sum_{i=1}^{n=120} \frac{(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n=120}(X_i - \bar{X})^2} = 2.034375$$

$$b_0 = \widehat{\beta_0} = \bar{Y} - b_1\bar{X} = 168.6$$

$$\hat{Y} = 168.6 + 2.034375\ X$$

At X=40

$$\widehat{Y_h} = 168.6 + 2.034375\ (40) = 249.975$$

Q1.24) Refer to Copier maintenance Problem 1.20.
a Obtain the residuals e_i and the sum of the squared residuals $\sum e_i^2$. What is the relation between the sum of the squared residuals here and the quantity Q in (1.8)?
b. Obtain point estimates of σ^2 and . In what units is σ expressed?

$$\sum e_i^2 = 3416.377$$
$$\sum e_i^2 = Q$$
$$\widehat{\sigma^2} = \frac{\sum e_i^2}{n-2} = \frac{3416.377}{43} = 79.45063 = MSE$$
$$\sigma = \sqrt{MSE} = \sqrt{79.45063}$$

Q1.25) (H.W) Refer to Airfreight breakage Problem 1.21.
a. Obtain the residual for the first case. What is its relation to e_1?
b. Compute $\sum e_i^2$ and MSE. What is estimated by MSE?

Q1.26) (H.W) Refer to Plastic hardness Problem 1.22.
a. Obtain the residuals *ej*. Do they sum to zero in accord with (1.17)?
b. Estimate σ^2 and . In what units is σ expressed?

Q1.21) **Solution** (H.W)

Airfreight breakage. A substance used in biological and medical research is shipped by airfreight to users in cartons of 1,000 ampules. The data below, involving 10 shipments, were collected on the number of times the carton was transferred from one aircraft to another over the shipment route (X) and the number of ampules found to be broken upon arrival (Y). Assume that first-order regression model (1.1) is appropriate.

a. Obtain the estimated regression function. Plot the estimated regression function and the data. Does a linear regression function appear to give a good fit here?

$$\hat{Y} = 10.2 + 4.0X$$

b. Obtain a point estimate of the expected number of broken ampules when X = 1 transfer is made.
If X=1
Then

$$\widehat{Y_h} = 10.2 + 4.0(1) = 14.20$$

c. Estimate the increase in the expected number of ampules broken when there are 2 transfers as compared to 1 transfer.

$$\widehat{Y_{h2}} = 10.2 + 4.0(2) = 18.20$$
$$\widehat{Y_{h1}} = 10.2 + 4.0(1) = 14.20$$
$$\widehat{Y_{h2}} - \widehat{Y_{h1}} = b_1 = 4.0$$

d. Verify that your fitted regression line goes through the point $(\bar{X}, \bar{Y})$.

$$\bar{X} = 1, \bar{Y} = 14.2$$

$$(\bar{X}, \bar{Y}) = (1, 14.2)$$

If X=1
Then

$$\widehat{Y_h} = 10.2 + 4.0(1) = 14.20$$

Then we can say the regression line goes through the point $(\bar{X}, \bar{Y}) = (1, 14.2)$

Chapter 2

We assume that the normal error regression model is applicable. This model is:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

where:
β_0 and β_1, are parameters
X_i are known constants
ε_i are independent $N\ (0, \sigma^2)$

$$E(Y_i) = \beta_0 + \beta_1 X_i$$

Sampling Distribution of $\widehat{\boldsymbol{\beta_1}}$

$$\widehat{\beta_1} = b_1 = \sum_{i=1}^{n} \frac{(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

$$E\left(\widehat{\beta_1}\right) = \beta_1$$

$$\sigma^2\left(\widehat{\beta_1}\right) = \frac{\sigma^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

$$s^2\left(\widehat{\beta_1}\right) = \frac{MSE}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

$$\frac{b_1 - \beta_1}{s(b_1)} \sim t_{(n-2)}$$

Confidence Interval for $\boldsymbol{\beta_1}$

$$P\left[b_1 - t_{\left(1-\frac{\alpha}{2}, n-2\right)} s(b_1) \le \beta_1 \le b_1 + t_{(1-\alpha/2, n-2)} s(b_1)\right] = 1 - \alpha$$

C.I $(1-\alpha)\%$ for β_1

$b_1 - t_{\left(1-\frac{\alpha}{2},n-2\right)} s(b_1) \le \beta_1 \le b_1 + t_{(1-\alpha/2,n-2)} s(b_1)$

Tests Concerning $\boldsymbol{\beta_1}$

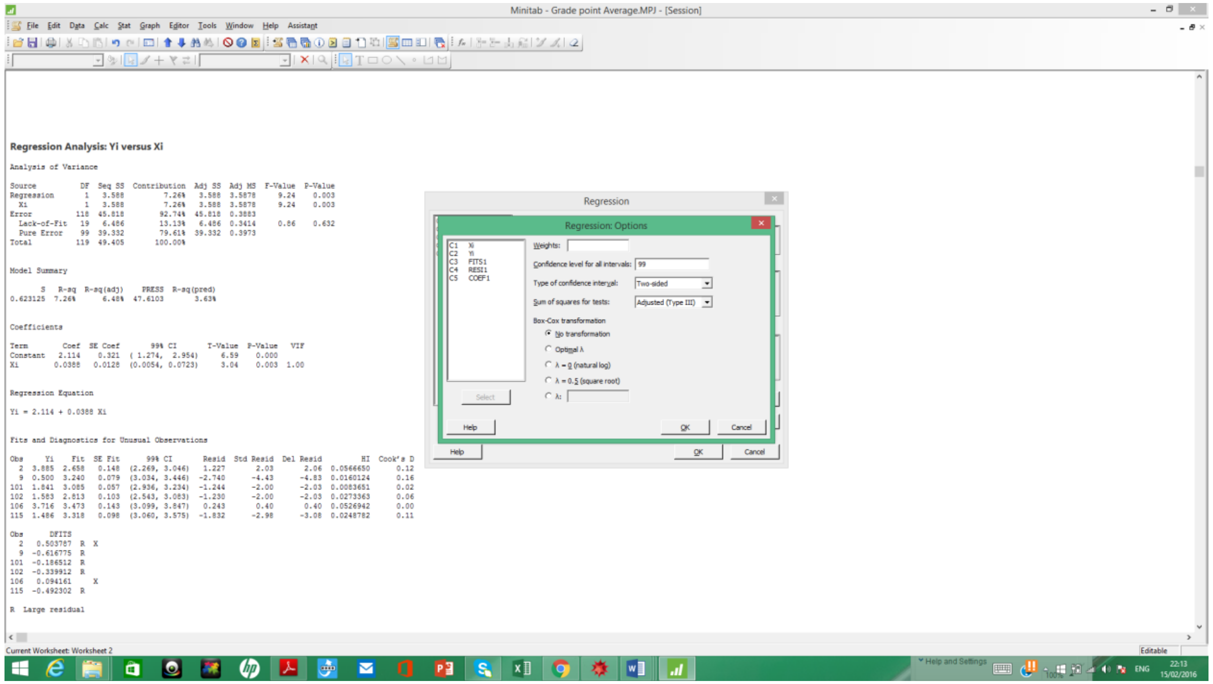
1. Hypothesis		
$H_0: \beta_1 = \beta_{10}$ $H_1: \beta_1 \ne \beta_{10}$	$H_0: \beta_1 = \beta_{10}$ $H_1: \beta_1 > \beta_{10}$	$H_0: \beta_1 = \beta_{10}$ $H_1: \beta_1 < \beta_{10}$
2. Test statistic		
$T_0 = \dfrac{b_1 - \beta_{10}}{s(b_1)}$		
3. Decision: Reject H_0 if		
$\lvert T_0 \rvert > t_{\left(1-\frac{\alpha}{2},n-2\right)}$	$T_0 > t_{(1-\alpha,n-2)}$	$T_0 < t_{(\alpha,n-2)}$
P-value: Reject H_0 if $p-value < \alpha$		
p-value=$2P\left(t_{(n-2)} > \lvert T_0 \rvert\right)$	$p-value = P\left(t_{(n-2)} > T_0\right)$	$p-value = P\left(t_{(n-2)} < T_0\right)$

Page 114

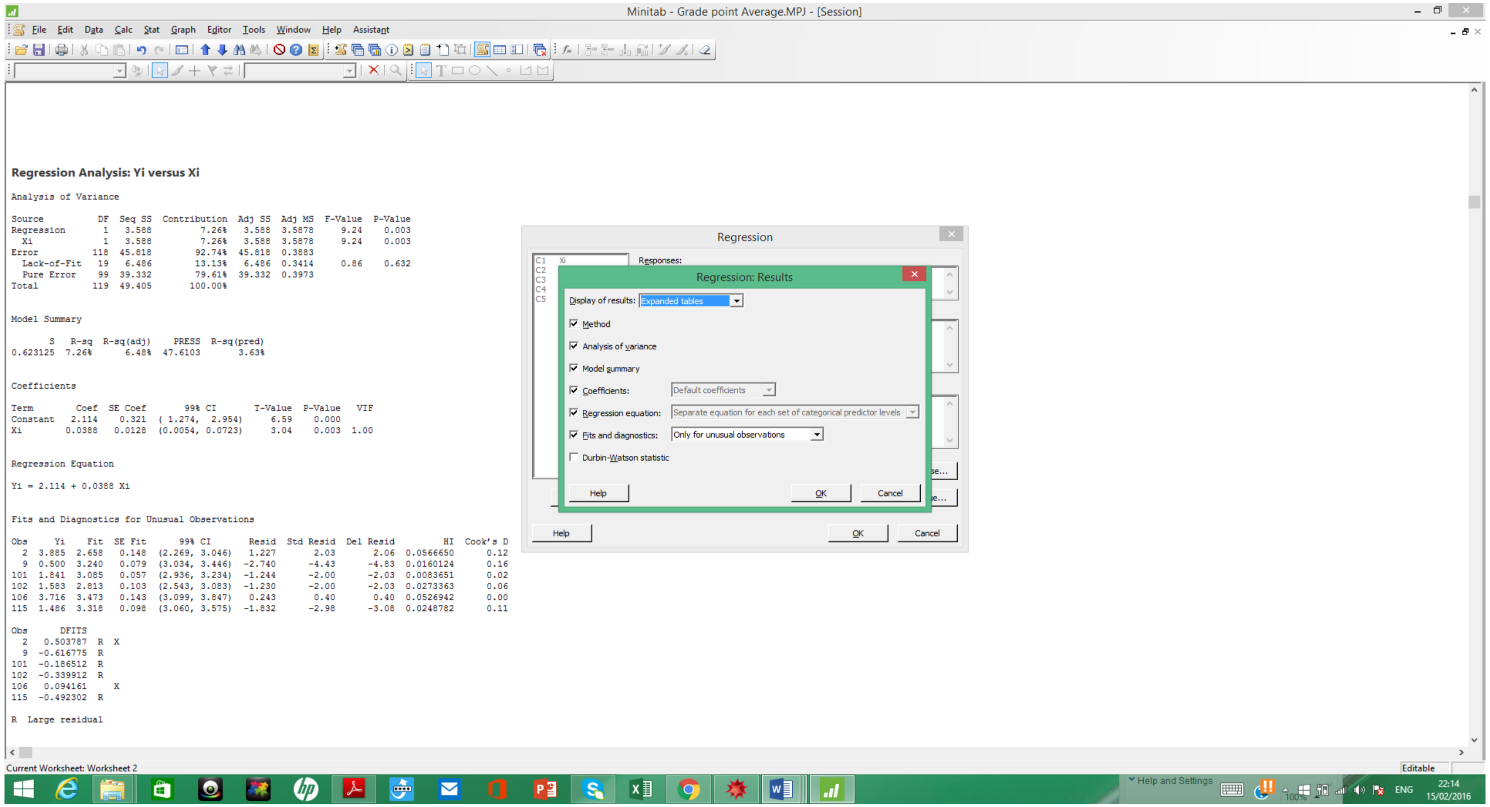
Q2.4. Refer to **Grade point average** Problem 1.19.
a. Obtain a 99 percent confidence interval for $\boldsymbol{\beta_1}$. Interpret your confidence interval. Does it include zero? Why might the director of admissions be interested in whether the confidence interval includes zero?

Solution:
By using Minitab:

$$Stat \rightarrow Regression \rightarrow Regression \rightarrow Fit\ Regression\ Mode$$



Minitab - Grade point Average.MPJ - [Session]
File Edit Data Calc Stat Graph Editor Tools Window Help Assistant
Regression Analysis: Yi versus Xi
Analysis of Variance
Source DF Seq SS Contribution Adj SS Adj MS F-Value P-Value
Regression 1 3.588 7.26% 3.588 3.5878 9.24 0.003
Xi 1 3.588 7.26% 3.588 3.5878 9.24 0.003
Error 118 45.818 92.74% 45.818 0.3883
Lack-of-Fit 19 6.486 13.13% 6.486 0.3414 0.86 0.632
Pure Error 99 39.332 79.61% 39.332 0.3973
Total 119 49.405 100.00%
Model Summary
S R-sq R-sq(adj) PRESS R-sq(pred)
0.623125 7.26% 6.48% 47.6103 3.63%
Coefficients
Term Coef SE Coef 99% CI T-Value P-Value VIF
Constant 2.114 0.321 (1.274, 2.954) 6.59 0.000
Xi 0.0388 0.0128 (0.0054, 0.0723) 3.04 0.003 1.00
Regression Equation
Yi = 2.114 + 0.0388 Xi
Fits and Diagnostics for Unusual Observations
Obs Yi Fit SE Fit 99% CI Resid Std Resid Del Resid HI Cook's D
2 3.885 2.658 0.148 (2.269, 3.046) 1.227 2.03 2.06 0.0566650 0.12
9 0.500 3.240 0.079 (3.034, 3.446) -2.740 -4.43 -4.83 0.0160124 0.16
101 1.841 3.085 0.057 (2.936, 3.234) -1.244 -2.00 -2.03 0.0083651 0.02
102 1.583 2.813 0.103 (2.543, 3.083) -1.230 -2.00 -2.03 0.0273363 0.06
106 3.716 3.473 0.143 (3.099, 3.847) 0.243 0.40 0.40 0.0526942 0.00
115 1.486 3.318 0.098 (3.060, 3.575) -1.832 -2.98 -3.08 0.0248782 0.11
Obs DFITS
2 0.503787 R X
9 -0.616775 R
101 -0.186512 R
102 -0.339912 R
106 0.094161 X
115 -0.492302 R
R Large residual
Regression
Regression: Options
C1 Xi
C2 Yi
C3 FITS1
C4 RESI1
C5 COEF1
Weights:
Confidence level for all intervals: 99
Type of confidence interval: Two-sided
Sum of squares for tests: Adjusted (Type III)
Box-Cox transformation
No transformation
Optimal λ
λ = 0 (natural log)
λ = 0.5 (square root)
λ:
Select
Help
OK
Cancel
Help
OK
Cancel
Current Worksheet: Worksheet 2
Editable



Regression Analysis: Yi versus Xi

Analysis of Variance

Source	DF	Seq SS	Contribution	Adj SS	Adj MS	F-Value	P-Value
Regression	1	3.588	7.26%	3.588	3.5878	9.24	0.003
Xi	1	3.588	7.26%	3.588	3.5878	9.24	0.003
Error	n-2=118	45.818	92.74%	SSE=45.818	MSE=0.3883		
Lack-of-Fit	19	6.486	13.13%	6.486	0.3414	0.86	0.632

```
  Pure Error     99   39.332         79.61%   39.332   0.3973
Total           119   49.405        100.00%
```

Model Summary

```
       S    R-sq  R-sq(adj)    PRESS  R-sq(pred)
0.623125   7.26%      6.48%  47.6103       3.63%
```

Coefficients

Term	Coef	SE Coef	99% CI	T-Value	P-Value	VIF
Constant	2.114	0.321	(1.274, 2.954)	6.59	0.000	
Xi	0.0388	0.0128	**(0.0054, 0.0723)**	3.04	**0.003**	1.00

Regression Equation

Yi = 2.114 + 0.0388 Xi

99% C.I for β_1: $b_1 - t_{\left(1-\frac{\alpha}{2},n-2\right)} s(b_1) \le \beta_1 \le b_1 + t_{(1-\alpha/2,n-2)} s(b_1)$

$$0.0054 \le \beta_1 \le 0.0723$$

Interpret your confidence interval. Does it include zero? No

Why might the director of admissions be interested in whether the confidence interval includes zero?

If the C.I of β_1 include zero, then β_1 can tack zero and $\beta_1 = 0$

b. Test, using the test statistic t*, whether or not a linear association exists between student's ACT score (X) and GPA at the end of the freshman year (Y). Use a level of significance of 0.01 State the alternatives, decision rule, and conclusion.

$\alpha = 0.01$

1. Hypothesis
$H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$
2. Test statistic
$T_0 = \frac{b_1 - \beta_{10}}{s(b_1)} = \frac{b_1}{s(b_1)} = \frac{0.0388}{0.0128} = 3.04$
3. Decision: Reject H_0 if $|T_0| > t_{\left(1-\frac{\alpha}{2}, n-2\right)}$, $3.04 > t_{(0.995,118)} = 1.70943$

Then reject H_0

c. What is the P-value of your test in part (b)? How does it support the conclusion reached in part (b)?
p-value= 0.003<0.01, then we reject H_0.

Q2.5. Refer to <u>Copier maintenance</u> Problem 1.20.

$$n = 45, \sum_{i=1}^{n=45} X_i = 230, \sum_{i=1}^{45} Y_i = 3432, \sum_{i=1}^{45} X_i^2 = 1516, \quad \sum_{i=1}^{45} X_i Y_i = 22660$$
$$SSE = 3416.377$$

a. Estimate the change in the mean service time when the number of copiers serviced increases by one. Use a 90 percent confidence interval. Interpret your confidence interval.

90% C.I for β_1: $b_1 - t_{\left(1-\frac{\alpha}{2},n-2\right)} s(b_1) \leq \beta_1 \leq b_1 + t_{(1-\alpha/2,n-2)} s(b_1)$

$\alpha = 1 - 0.9 = 0.1$

$$b_1 = \sum_{i=1}^{n} \frac{(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \sum_{i=1}^{n} \frac{(X_i Y_i - \bar{X} Y_i - X_i \bar{Y} + \bar{X}\bar{Y})}{\sum_{i=1}^{n}\left({X_i}^2 - 2\bar{X}X_i + \bar{X}^2\right)} = \frac{\sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y}}{\sum_{i=1}^{n} {X_i}^2 - n\bar{X}^2} = \frac{22660 - 45 * 5.1111 * 76.2667}{1516 - 45 * 5.1111^2} = 15.035$$

$$s^2(b_1) = \frac{MSE}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{3416.377/(45-2)}{1516 - 45 * 5.1111^2} = 0.23337$$

$s(b_1) = \sqrt{0.2337} = 0.48308$

$t_{\left(1-\frac{\alpha}{2},n-2\right)} = t_{(0.95,43)} = 1.68107$
$b_1 - t_{\left(1-\frac{\alpha}{2},n-2\right)} s(b_1) = 15.035 - 1.68107 * 0.48308 = 14.222$
$b_1 + t_{\left(1-\frac{\alpha}{2},n-2\right)} s(b_1) = 15.035 + 1.68107 * 0.48308 = 15.84709$
$14.222 \leq \beta_1 \leq 15.847$

b. Conduct a t test to determine whether or not there is a linear association between X and Y here; control the α a risk at 0.01. State the alternatives, decision rule, and conclusion. What is the P-value of your test?
$\alpha = 0.01$

1. Hypothesis

$H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$
2. Test statistic
$$T_0 = \frac{b_1 - \beta_{10}}{s(b_1)} = \frac{b_1}{s(b_1)} = \frac{15.035}{0.48308} = 31.123$$
3. Decision: Reject H_0 if $|T_0| > t_{\left(1-\frac{\alpha}{2}, n-2\right)}$, $31.123 > t_{(0.995,43)} = 2.695$

Then reject H_0

p-value=$2P\left(t_{(n-2)} > |T_0|\right) = 2\left(1 - P\left(t_{(n-2)} < 31.123\right)\right) = 2(1-1)$

$0.00 < 0.01$, then we reject H_0.

c. Are your results in parts (a) and (b) consistent? Explain.
Yes, the C.I of β_1 does not include zero, and we reject H_0.

d. The manufacturer has suggested that the mean required time should not increase by more than 14 minutes for each additional copier that is serviced on a service call. Conduct a test to decide whether this standard is being satisfied by Tri-City. Control the risk of a Type I error at 0.05. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

$\alpha = 0.05$

1. Hypothesis
$H_0: \beta_1 \leq 14$

$H_1: \beta_1 > 14$
2. Test statistic
$$T_0 = \frac{b_1 - \beta_{10}}{s(b_1)} = \frac{b_1 - 14}{s(b_1)} = \frac{15.035 - 14}{0.48308} = 2.143$$
3. Decision: Reject H_0 if $T_0 > t_{(1-\alpha, n-2)}$, $2.143 > t_{(0.95,43)} = 1.861$
Then reject H_0

p-value=$P\left(t_{(n-2)} > T_0\right) = \left(1 - P\left(t_{(n-2)} < 2.143\right)\right) = (1 - 0.981) = 0.019 < 0.05$

, then we reject H_0.

Q2.6. Refer to Airfreight breakage Problem 1.21.
$\bar{X} = 1, \bar{Y} = 14.2, \sum_{i=1}^{n=10}(X_i - \bar{X})(Y_i - \bar{Y}) = 40$

$\sum_{i=1}^{10}(X_i - \bar{X})^2 = 10, MSE = 2.2$

a. Estimate $\boldsymbol{\beta_1}$ with a 95 percent confidence interval. Interpret your interval estimate.
95% C.I for β_1: $b_1 - t_{\left(1-\frac{\alpha}{2},n-2\right)}s(b_1) \leq \beta_1 \leq b_1 + t_{(1-\alpha/2,n-2)}s(b_1)$

$\alpha = 1 - 0.95 = 0.05$

$$b_1 = \widehat{\beta_1} = \sum_{i=1}^{n=120} \frac{(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n=120}(X_i - \bar{X})^2} = 4$$

$$s^2(b_1) = \frac{MSE}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{2.2}{10} = 0.22$$

$s(b_1) = \sqrt{0.22} = 0.469$

$t_{\left(1-\frac{\alpha}{2},n-2\right)} = t_{(0.975,8)} = 2.306$
$b_1 - t_{\left(1-\frac{\alpha}{2},n-2\right)}s(b_1) = 4 - 2.306 * 0.469 = 2.918$
$b_1 + t_{\left(1-\frac{\alpha}{2},n-2\right)}s(b_1) = 4 + 2.306 * 0.469 = 5.081$
$2.918 \leq \beta_1 \leq 5.081$

b. Conduct a t test to decide whether or not there is a linear association between number of times a carton is transferred (X) and number of broken ampules (Y). Use a level of significance of 0.05. State the alternatives, decision rule, and conclusion. What is the P-value of the test?
$\alpha = 0.05$

1. Hypothesis

$H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$

2. Test statistic

$$T_0 = \frac{b_1 - \beta_{10}}{s(b_1)} = \frac{b_1}{s(b_1)} = \frac{4}{0.469} = 8.528$$

3. Decision: Reject H_0 if $|T_0| > t_{\left(1-\frac{\alpha}{2}, n-2\right)}$, $8.528 > t_{(0.975,8)} = 2.308$

Then reject H_0

p-value=$2P\left(t_{(n-2)} > |T_0|\right) = 2\left(1 - P\left(t_{(8)} < 8.528\right)\right) = 2(1 - 0.9999)$

$0.0002 < 0.05$, then we reject H_0.

```
Analysis of Variance
```

Source	DF	Seq SS	Contribution	Adj SS	Adj MS	F-Value	P-Value
Regression	1	160.000	90.09%	160.000	160.000	72.73	0.000
Xi	1	160.000	90.09%	160.000	160.000	72.73	0.000
Error	8	17.600	9.91%	17.600	2.200		
Lack-of-Fit	2	0.933	0.53%	0.933	0.467	0.17	0.849
Pure Error	6	16.667	9.38%	16.667	2.778		
Total	9	177.600	100.00%				

```
Model Summary
```

S	R-sq	R-sq(adj)	PRESS	R-sq(pred)
1.48324	90.09%	88.85%	25.8529	85.44%

```
Coefficients
```

Term	Coef	SE Coef	95% CI	T-Value	P-Value	VIF
Constant	10.200	0.663	(8.670, 11.730)	15.38	0.000	
Xi	4.000	0.469	(2.918, 5.082)	8.53	0.000	1.00

Regression Equation

```
Yi = 10.200 + 4.000 Xi
```

H.W:

Q2.7 Refer to Plastic hardness Problem 1.22.

a. Estimate the change in the mean hardness when the elapsed time increases by one hour. Use a 99 percent confidence interval. Interpret your interval estimate.

b. The plastic manufacturer has stated that the mean hardness should increase by 2 Brinell units per hour. Conduct a two-sided test to decide whether this standard is being satisfied; use $\alpha = 0.01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

Chapter 2

We assume that the normal error regression model is applicable. This model is:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

where:
β_0 and β_1, are parameters
X_i are known constants
ε_i are independent $N\ (0, \sigma^2)$

$$E(Y_i) = \beta_0 + \beta_1 X_i$$

Sampling Distribution of $\widehat{\boldsymbol{\beta_1}}$

$$\widehat{\beta_1} = b_1 = \sum_{i=1}^{n} \frac{(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

$$E\left(\widehat{\beta_1}\right) = \beta_1$$

$$\sigma^2\left(\widehat{\beta_1}\right) = \frac{\sigma^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

$$s^2\left(\widehat{\beta_1}\right) = \frac{MSE}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

$$\frac{b_1-\beta_1}{s(b_1)} \sim t_{(n-2)}$$

Confidence Interval for $\boldsymbol{\beta_1}$

$$P\left[b_1 - t_{\left(1-\frac{\alpha}{2},n-2\right)} s(b_1) \le \beta_1 \le b_1 + t_{(1-\alpha/2,n-2)} s(b_1)\right] = 1-\alpha$$

C.I $(1-\alpha)\%$ for β_1

$$b_1 - t_{\left(1-\frac{\alpha}{2},n-2\right)} s(b_1) \le \beta_1 \le b_1 + t_{(1-\alpha/2,n-2)} s(b_1)$$

Tests Concerning $\boldsymbol{\beta_1}$

1. Hypothesis		
$H_0: \beta_1 = \beta_{10}$ $H_1: \beta_1 \neq \beta_{10}$	$H_0: \beta_1 = \beta_{10}$ $H_1: \beta_1 > \beta_{10}$	$H_0: \beta_1 = \beta_{10}$ $H_1: \beta_1 < \beta_{10}$
2. Test statistic		
$T_0 = \frac{b_1 - \beta_{10}}{s(b_1)}$		
3. Decision: Reject H_0 if		
$\lvert T_0 \rvert > t_{\left(1-\frac{\alpha}{2},n-2\right)}$	$T_0 > t_{(1-\alpha,n-2)}$	$T_0 < t_{(\alpha,n-2)}$
P-value: Reject H_0 if $p-value < \alpha$		
p-value=$2P\left(t_{(n-2)} > \lvert T_0 \rvert\right)$	p-value= $P\left(t_{(n-2)} > T_0\right)$	$p-value = P\left(t_{(n-2)} < T_0\right)$

Sampling Distribution of $\widehat{\boldsymbol{\beta_0}}$

$$\widehat{\beta_0} = b_0 = \bar{Y} - b_1\bar{X}$$

$$E\left(\widehat{\beta_0}\right) = \beta_0$$

$$\sigma^2\left(\widehat{\beta_0}\right) = \sigma^2\left(\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right)$$

$$s^2\left(\widehat{\beta_0}\right) = MSE\left(\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right)$$

$$\frac{b_0 - \beta_0}{s(b_0)} \sim t_{(n-2)}$$

Confidence Interval for $\boldsymbol{\beta_1}$

$$P\left[b_0 - t_{\left(1-\frac{\alpha}{2},n-2\right)}s(b_0) \leq \beta_0 \leq b_0 + t_{(1-\alpha/2,n-2)}s(b_0)\right] = 1 - \alpha$$

C.I $(1-\alpha)\%$ for β_0

$$b_0 - t_{\left(1-\frac{\alpha}{2},n-2\right)}s(b_0) \leq \beta_0 \leq b_0 + t_{(1-\alpha/2,n-2)}s(b_0)$$

Tests Concerning $\boldsymbol{\beta_1}$

1. Hypothesis		
$H_0: \beta_0 = \beta_{00}$ $H_1: \beta_0 \neq \beta_{00}$	$H_0: \beta_0 = \beta_{00}$ $H_1: \beta_0 > \beta_{00}$	$H_0: \beta_1 = \beta_{00}$ $H_1: \beta_1 < \beta_{00}$
2. Test statistic		
$T_0 = \frac{b_0 - \beta_{00}}{s(b_0)}$		
3. Decision: Reject H_0 if		
$\lvert T_0 \rvert > t_{\left(1-\frac{\alpha}{2}, n-2\right)}$	$T_0 > t_{(1-\alpha, n-2)}$	$T_0 < t_{(\alpha, n-2)}$
P-value: Reject H_0 if $p - value < \alpha$		
p-value=$2P\left(t_{(n-2)} > \lvert T_0 \rvert\right)$	p-value= $P\left(t_{(n-2)} > T_0\right)$	$p - value = P\left(t_{(n-2)} < T_0\right)$

$$Y_h = b_0 + b_1 X_h$$

ANOVA TABLE

Source of Variation	d.f	SS	MS	F	p-value
Regression	1	SSR=$\sum\left(\hat{Y}_i - \bar{Y}\right)^2$	$MSR = \frac{SSR}{1}$	$\frac{MSR}{MSE}$	
Error	n-2	SSE=$\sum\left(Y_i - \hat{Y}_i\right)^2$	$MSE = \frac{SSE}{n-2}$		
Total	n-1	SSTo= $\sum(Y_i - \bar{Y})^2$			

1. Hypothesis

$H_0: \beta_1 = 0$ (Non liner)

$H_1: \beta_1 \neq 0$

2. Test statistic

$$F^* = \frac{MSR}{MSE}$$

3. Decision: Reject H_0 if

$F > F_{(1-\alpha,1,n-2)}$

P-value: Reject H_0 if $p - value < \alpha$

$p - value = P\left(F_{(1,n-2)} > F^*\right)$

Q2.6**. Refer to Airfreight breakage Problem 1.21.**
$\bar{X} = 1, \bar{Y} = 14.2,$

$$\sum_{i=1}^{n=10} (X_i - \bar{X})\,(Y_i - \bar{Y}) = 40, \sum_{i=1}^{10} (X_i - \bar{X})^2 = 10$$

$$\sum_{i=1}^{n=10} (Y_i - \bar{Y})^2 = 177.6, MSE = 2.2$$

$b_0 = 10.2, b_1 = 4$

d) A consultant has suggested, on the basis of previous experience, that the mean number of broken ampules should not exceed 9.0 when no transfers are made. Conduct an appropriate test, using $\alpha = 0.025$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?
$\alpha = 0.025$

1. Hypothesis

$H_0: \beta_0 \leq 9$

$H_1: \beta_0 > 9$

2. Test statistic

$$T_0 = \frac{b_0 - \beta_{00}}{s(b_0)} = \frac{10.2 - 9}{0.6633} = 1.809$$

$$s^2(\widehat{\beta_0}) = MSE\left(\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right) = 2.2\left(\frac{1}{10} + \frac{1^2}{10}\right) = 0.44$$

$s(b_0) = 0.6633$

3. Decision: Reject H_0 if $T_0 > t_{(1-\alpha, n-2)}$,

$1.809 \ngtr t_{(0.975,8)} = 2.306$

Then not reject H_0

p-value=$P\left(t_{(n-2)} > T_0\right) = \left(1 - P\left(t_{(n-2)} < 1.809\right)\right) = (1 - 0.945) = 0.055 \nless 0.025$

, then we not reject H_0.

at $\alpha = 0.05$

$$b_0 - t_{\left(1-\frac{\alpha}{2}, n-2\right)} s(b_0) \leq \beta_0 \leq b_0 + t_{(1-\alpha/2, n-2)} s(b_0)$$

$$t_{\left(1-\frac{\alpha}{2}, n-2\right)} = t_{(0.975,8)} = 2.306$$

$$10.2 - 2.306 * 0.6633 \leq \beta_0 \leq 10.2 + 2.306 * 0.6633$$

$$8.76 \leq \beta_0 \leq 11.728$$

Analysis of Variance

Source	DF	Seq SS	Contribution	Adj SS	Adj MS	F-Value	P-Value
Regression	1	160.000	90.09%	160.000	160.000	72.73	0.000
Xi	1	160.000	90.09%	160.000	160.000		
Error	8	17.600	9.91%	17.600	2.200		
Total	9	177.600	100.00%				

Coefficients

Term	Coef	SE Coef	95% CI	T-Value	P-Value	VIF
Constant	10.200	0.663	(8.670, 11.730)	15.38	0.000	
Xi	4.000	0.469	(2.918, 5.082)	8.53	0.000	1.00

Regression Equation

Yi = 10.200 + 4.000 Xi

Q2.25. Refer to Airfreight breakage Problem 1.21.
a. Set up the ANOVA table. Which elements are additive?
b. Conduct an F test to decide whether or not there is a linear association between the number of times a carton is transferred and the number of broken ampules; control the α risk at 0.05. State the alternatives, decision rule, and conclusion.
c. Obtain the t^* statistic for the test in part (b) and demonstrate numerically its equivalence to the F^* statistic obtained in part (b).

$\bar{X} = 1, \bar{Y} = 14.2, \sum_{i=1}^{n=10}(X_i - \bar{X})(Y_i - \bar{Y}) = 40, \quad \sum_{i=1}^{10}(X_i - \bar{X})^2 = 10$

$$\sum_{i=1}^{n=10}(Y_i - \bar{Y})^2 = 177.6, \qquad MSE = 2.2, \qquad b_0 = 10.2, \qquad b_1 = 4$$

X_i	Y_i	$(X_i - \bar{X})$	$(Y_i - \bar{Y})$	$(X_i - \bar{X})^2$	$(X_i - \bar{X}) * (Y_i - \bar{Y})$	$(Y_i - \bar{Y})^2$	$\hat{Y}_i$	$(Y_i - \hat{Y}_i)^2$
1	16	0	1.8	0	0	3.24	14.2	3.24
0	9	-1	-5.2	5.2	1	27.04	10.2	1.44
2	17	1	2.8	2.8	1	7.84	18.2	1.44
0	12	-1	-2.2	2.2	1	4.84	10.2	3.24
3	22	2	7.8	15.6	4	60.84	22.2	0.04
1	13	0	-1.2	0	0	1.44	14.2	1.44
0	8	-1	-6.2	6.2	1	38.44	10.2	4.84
1	15	0	0.8	0	0	0.64	14.2	0.64
2	19	1	4.8	4.8	1	23.04	18.2	0.64
0	11	-1	-3.2	3.2	1	10.24	10.2	0.64
10	142	0	0	40	10	177.6	142	17.6

$$\sum\left(Y_i - \widehat{Y}_i\right)^2 = 17.6$$

ANOVA TABLE

Source of Variation	d.f	SS	MS	F	p-value
Regression	1	SSR=$177.6 - 17.6 = 160$	$MSR = 160$	$\frac{160}{2.2} = 72.72$	0.00
Error	8	SSE=17.6	$MSE = \frac{17.6}{8} = 2.2$		
Total	9	SSTo= 177.6			

$\alpha = 0.05$

1. Hypothesis
$H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$
2. Test statistic
$F^* = 72.72$
3. Decision: Reject H_0 if $F^* > F_{(1-\alpha,1,n-2)}$, $72.72 > F_{(0.95,1,8)} = 5.31$
Then reject H_0
p-value=$P\left(F_{(1,n-2)} > F^*\right) = \left(1 - P\left(F_{(1,8)} < 72.72\right)\right) = (1 - 0.9999) = 0.0001 < 0.05$
, then we reject H_0.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	160.000	160.000	72.73	0.000
Xi	1	160.000	160.000	72.73	0.000
Error	8	17.600	2.200		
Lack-of-Fit	2	0.933	0.467	0.17	0.849
Pure Error	6	16.667	2.778		
Total	9	177.600			

$t^* = 8.528$, $(t^*)^2 = (8.528)^2 = 72.72 = F^*$

Q2.26. Refer to Plastic hardness Problem 1.22.
a. Set up the ANOVA table.
b. Test by means of an F test whether or not there is a linear association between the hardness of the plastic and the elapsed time. Use $a = .01$. State the alternatives, decision rule, and conclusion.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	5297.51	5297.51	506.51	0.000
Xi	1	5297.51	5297.51	506.51	0.000
Error	14	146.43	10.46		
Lack-of-Fit	2	17.67	8.84	0.82	0.462
Pure Error	12	128.75	10.73		
Total	15	5443.94			

$\alpha = 0.01$

1. Hypothesis
$H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$
2. Test statistic
$F^* = 506.51$
3. Decision:
p-value=0.000<0.01
, then we reject H_0.

Prove that

$$Q = \sum_{i=1}^{n} {\epsilon_i}^2 = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)^2$$

$$\frac{\partial Q}{\partial \beta_0} = 0, \frac{\partial Q}{\partial \beta_1} = 0$$

$$\frac{\partial Q}{\partial \beta_0} = 2\sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i)\,(-1) = 0$$

$$\sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i) = 0$$

$$\sum_{i=1}^{n} (Y_i) - \sum_{i=1}^{n} (b_0) - \sum_{i=1}^{n} (b_1 X_i) = 0$$

$$\sum_{i=1}^{n} (Y_i) - n b_0 - b_1 \sum_{i=1}^{n} (X_i) = 0$$

$$\sum_{i=1}^{n} (Y_i) = n b_0 + b_1 \sum_{i=1}^{n} (X_i) \to (1)$$

$$\frac{\partial Q}{\partial \beta_0} = 2\sum_{i=1}^{n} [(Y_i - b_0 - b_1 X_i)(-X_i)] = 0$$

$$\sum_{i=1}^{n} \left(Y_i X_i - b_0 X_i - b_1 {X_i}^2\right) = 0$$

$$\sum_{i=1}^{n}(Y_iX_i) - b_0\sum_{i=1}^{n}(X_i) - b_1\sum_{i=1}^{n}\left({X_i}^2\right) = 0$$

$$\sum_{i=1}^{n}(Y_iX_i) = b_0\sum_{i=1}^{n}(X_i) + b_1\sum_{i=1}^{n}\left({X_i}^2\right) \rightarrow (2)$$

By solving 1 and 2 together

$\sum_{i=1}^{n}(Y_i) = nb_0 + b_1\sum_{i=1}^{n}(X_i)$

$\sum_{i=1}^{n}(Y_iX_i) = b_0\sum_{i=1}^{n}(X_i) + b_1\sum_{i=1}^{n}\left({X_i}^2\right)$

From 1

$\bar{Y} = b_0 + b_1\bar{X}$

$b_0 = \bar{Y} - b_1\bar{X}$

$$\sum_{i=1}^{n}(Y_iX_i) = (\bar{Y} - \beta_1\bar{X})\sum_{i=1}^{n}(X_i) + b_1\sum_{i=1}^{n}\left({X_i}^2\right)$$

$$\sum_{i=1}^{n}(Y_iX_i) = \bar{Y}\sum_{i=1}^{n}(X_i) + b_1\left[\sum_{i=1}^{n}\left({X_i}^2\right) - \bar{X}\sum_{i=1}^{n}(X_i)\right]$$

$$\sum_{i=1}^{n}(Y_iX_i) - \bar{Y}\sum_{i=1}^{n}(X_i) = b_1\left[\sum_{i=1}^{n}\left({X_i}^2\right) - \bar{X}\sum_{i=1}^{n}(X_i)\right]$$

$$\sum_{i=1}^{n}(Y_i X_i) - n\bar{Y}\bar{X} = b_1\left[\sum_{i=1}^{n}\left({X_i}^2\right) - n\bar{X}^2\right]$$

$$b_1 = \frac{\sum_{i=1}^{n}(Y_i X_i) - n\bar{Y}\bar{X}}{\left[\sum_{i=1}^{n}\left({X_i}^2\right) - n\bar{X}^2\right]} = \sum_{i=1}^{n}\frac{(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

$$\left(\sum_{i=1}^{n} X_i\right)^2 \neq \sum_{i=1}^{n}\left({X_i}^2\right)$$

$$\sum_{i=1}^{n}\frac{(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

$$\left(\sum_{i=1}^{n}(X_i - \bar{X})^2\right)^2 \neq \sum_{i=1}^{n}(X_i - \bar{X})^4$$

$$\sum_{i=1}^{n} e_i = 0$$

$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n}\left(Y_i - \widehat{Y_\iota}\right) = \sum_{i=1}^{n}(Y_i - b_0 - b_1 X_i)$$

$$= \sum_{i=1}^{n}(Y_i - \bar{Y} + b_1\bar{X} - b_1 X_i) = \sum_{i=1}^{n}(Y_i - \bar{Y}) - b_1\sum_{i=1}^{n}(X_i - \bar{X}) = 0 - 0 = 0$$

$b_0 = \bar{Y} - b_1\bar{X}$

$$\sum_{i=1}^{n} e_i X_i = 0, \sum_{i=1}^{n} e_i X_i = \sum_{i=1}^{n} [(Y_i - b_0 - b_1 X_i) X_i] = \sum_{i=1}^{n} \left(Y_i X_i - b_0 X_i - b_1 {X_i}^2\right) = \sum_{i=1}^{n} (Y_i X_i) - b_0 \sum_{i=1}^{n} (X_i) - b_1 \sum_{i=1}^{n} \left({X_i}^2\right)$$

$$= \sum_{i=1}^{n} (Y_i X_i) - (\bar{Y} - b_1 \bar{X}) \sum_{i=1}^{n} (X_i) - b_1 \sum_{i=1}^{n} \left({X_i}^2\right) = \sum_{i=1}^{n} (Y_i X_i) - \bar{Y} \sum_{i=1}^{n} (X_i) + b_1 \bar{X} \sum_{i=1}^{n} (X_i) - b_1 \sum_{i=1}^{n} \left({X_i}^2\right)$$

$$= \sum_{i=1}^{n} (Y_i X_i) - n\bar{Y}\bar{X} + n b_1 \bar{X}^2 - b_1 \sum_{i=1}^{n} \left({X_i}^2\right) = \left[\sum_{i=1}^{n} (Y_i X_i) - n\bar{Y}\bar{X}\right] - b_1 \left[\sum_{i=1}^{n} \left({X_i}^2\right) - n\bar{X}^2\right]$$

$$= \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y}) - b_1 \sum_{i=1}^{n} (X_i - \bar{X})^2 = \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y}) - \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n} (X_i - \bar{X})^2} \sum_{i=1}^{n} (X_i - \bar{X})^2 = 0$$

$$\sum_{i=1}^{n} \hat{Y}_i = \sum_{i=1}^{n} (b_0 + b_1 X_i) = \sum_{i=1}^{n} (\bar{Y} - b_1 \bar{X} + b_1 X_i) = n\bar{Y} + b_1 \sum_{i=1}^{n} (X_i - \bar{X}) = \sum_{i=1}^{n} Y_i$$

$$Var\left(\frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n} (X_i - \bar{X})^2}\right)$$

$$\frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n} (X_i - \bar{X})^2} = \frac{\sum_{i=1}^{n} (X_i - \bar{X}) Y_i - \bar{Y} \sum_{i=1}^{n} (X_i - \bar{X})}{\sum_{i=1}^{n} (X_i - \bar{X})^2} = \frac{\sum_{i=1}^{n} (X_i - \bar{X}) Y_i}{\sum_{i=1}^{n} (X_i - \bar{X})^2}$$

$$Var\left(\frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sum_{i=1}^{n}(X_i-\bar{X})^2}\right)=Var\left(\frac{\sum_{i=1}^{n}(X_i-\bar{X})Y_i}{\sum_{i=1}^{n}(X_i-\bar{X})^2}\right)=Var\left(\sum_{i=1}^{n}\frac{(X_i-\bar{X})}{\sum_{i=1}^{n}(X_i-\bar{X})^2}Y_i\right)=\sum_{i=1}^{n}\left(\frac{(X_i-\bar{X})}{\sum_{i=1}^{n}(X_i-\bar{X})^2}\right)^2 Var(Y_i)$$

$$=\sum_{i=1}^{n}\frac{(X_i-\bar{X})^2}{(\sum_{i=1}^{n}(X_i-\bar{X})^2)^2}\sigma^2=\sigma^2\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{(\sum_{i=1}^{n}(X_i-\bar{X})^2)^2}=\frac{\sigma^2}{\sum_{i=1}^{n}(X_i-\bar{X})^2}$$

Prove that SSTo=SSR+SSE.

L.H.S= $SSTo=\sum_{i=1}^{n}(Y_i-\bar{Y})^2=\sum_{i=1}^{n}\left(Y_i-\widehat{Y}_i+\widehat{Y}_i-\bar{Y}\right)^2=\sum_{i=1}^{n}\left(\left(Y_i-\widehat{Y}_i\right)+\left(\widehat{Y}_i-\bar{Y}\right)\right)^2=\sum_{i=1}^{n}\left[\left(Y_i-\widehat{Y}_i\right)^2+\left(\widehat{Y}_i-\bar{Y}\right)^2+2\left(Y_i-\widehat{Y}_i\right)\left(\widehat{Y}_i-\bar{Y}\right)\right]$

$$=\sum_{i=1}^{n}\left(Y_i-\widehat{Y}_i\right)^2+\sum_{i=1}^{n}\left(\widehat{Y}_i-\bar{Y}\right)^2+2\sum_{i=1}^{n}\left(Y_i-\widehat{Y}_i\right)\left(\widehat{Y}_i-\bar{Y}\right)$$

$\because \left(Y_i-\widehat{Y}_i\right)=e_i$

$$\sum_{i=1}^{n}\left(Y_i-\widehat{Y}_i\right)\left(\widehat{Y}_i-\bar{Y}\right)=\sum_{i=1}^{n}e_i\left(\widehat{Y}_i-\bar{Y}\right)=\sum_{i=1}^{n}e_i\widehat{Y}_i-\bar{Y}\sum_{i=1}^{n}e_i$$

$$\because \sum_{i=1}^{n}e_i\widehat{Y}_i=\sum_{i=1}^{n}e_i=0$$

Then

$$\sum_{i=1}^{n}\left(Y_i-\widehat{Y}_i\right)\left(\widehat{Y}_i-\bar{Y}\right)=0$$

Then

$$SSTo = \sum_{i=1}^{n} \left(Y_i - \widehat{Y}_i\right)^2 + \sum_{i=1}^{n} \left(\widehat{Y}_i - \bar{Y}\right)^2$$

$$\sum_{i=1}^{n} \left(Y_i - \widehat{Y}_i\right)^2 = SSE \quad \& \quad \sum_{i=1}^{n} \left(\widehat{Y}_i - \bar{Y}\right)^2 = SSR$$

$$SSTo = SSR + SSE = L.H.S$$

Chapter 2

2.13 Refer to Grade point average.

Calculate R^2. What proportion of the variation in Y is accounted for by introducing X into the regression model? From page 98

$\bar{X} = 24.725, \sum_{i=1}^{n=120}(X_i - \bar{X})^2 = 2379.925$

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	SSR=3.588	3.5878	9.24	0.003
Xi	1	3.588	3.5878	9.24	0.003
Error	118	SSE=45.818	MSE=0.3883		
Lack-of-Fit	19	6.486	0.3414	0.86	0.632
Pure Error	99	39.332	0.3973		
Total	119	SSTo=49.405			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.623125	7.26%	6.48%	3.63%

$$R^2 = \frac{SSR}{SSTo} = \frac{3.588}{49.405} = 0.0726$$

$$R^2 = 1 - \frac{SSE}{SSTo} = 1 - \frac{45.818}{49.405} = 1 - 0.9274 = 0.0726$$

This means that 7.26% of change in the mean freshman OPA for students is by ACT test score

a. Obtain a 95 percent interval estimate of the mean freshman OPA for students whose ACT test score is 28. Interpret your confidence interval. From page 76- to 79

$$\widehat{Y_h} = b_0 + b_1\, X_h$$

$$s^2(\widehat{Y_h}) = MSE\left(\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right)$$

$$\widehat{Y_h} \pm t\left(1 - \frac{\alpha}{2}; n-2\right) s(\widehat{Y_h})$$

$$\alpha = 0.05, \frac{\alpha}{2} = 0.025$$

At $X_h = 28$

$$\widehat{Y_h} = 2.114 + 0.0388\ (28) = 3.2012$$

$$s^2(\widehat{Y_h}) = MSE\left(\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right)$$

$$s^2(\widehat{Y_h}) = MSE\left(\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right) = \mathbf{0.3883}\left(\frac{1}{120} + \frac{(28 - 24.725)^2}{2379.925}\right) = 0.004986$$

$$s(\widehat{Y_h}) = \sqrt{0.007776} = 0.0706$$

$$t\left(1 - \frac{\alpha}{2}; n-2\right) = t(0.975; 118) = 1.9807$$

$$3.22012 \pm 1.9807(0.0706)$$

$$3.0614 < E(Y_h) < 3.3410$$

b. Mary Jones obtained a score of 28 on the entrance test. Predict her freshman OPA-using a 95 percent prediction interval. Interpret your prediction interval.

$$s^2(\widehat{Y_{new}}) = MSE\left(1 + \frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right)$$

$$\widehat{Y_h} \pm t\left(1 - \frac{\alpha}{2}; n - 2\right) s(\widehat{Y_{new}})$$

$$s^2(\widehat{Y_{new}}) = MSE\left(1 + \frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right) = \mathbf{0.3883}\left(1 + \frac{1}{120} + \frac{(28 - 24.725)^2}{2379.925}\right) = 0.39328$$

$$s(\widehat{Y_{new}}) = 0.6271$$

$$3.22012 \pm 1.9807(0.6271)$$

$$1.9594 < Y_{h(new)} < 4.4430$$

c. Is the prediction interval in part (b) wider than the confidence interval in part (a)? Should it be?

هل فترة الثقة للتنبؤ في الجزء (ب) أوسع من فترة الثقة في الجزء (أ)؟ هل يجب أن تكون؟

Yes, Yes

2.15. Refer to Airfreight breakage Problem 1.21.

$$\bar{X} = 1, \quad \sum_{i=1}^{10} (X_i - \bar{X})^2 = 10$$

ANOVA TABLE

Source of Variation	d.f	SS	MS	F	p-value
Regression	1	SSR=160	$MSR = 160$	72.72	0.00
Error	8	SSE=17.6	$MSE = 2.2$		
Total	9	SSTo= 177.6			

a. Because of changes in airline routes, shipments may have to be transferred more frequently than in the past. Estimate the mean breakage for the following numbers of transfers: X = 2, 4. Use separate 99 percent confidence intervals. Interpret your results.

At $X_h = 2$

$$\hat{Y}_h = 10.2 + 4\,(2) = 18.2$$

$$s^2(\hat{Y}_h) = MSE\left(\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right) = \mathbf{2.2}\left(\frac{1}{10} + \frac{(2-1)^2}{10}\right) = 0.44$$

$$s(\hat{Y}_h) = \sqrt{0.44} = 0.6633$$

$$t\left(1 - \frac{\alpha}{2}; n-2\right) = t(0.995; 8) = 3.355$$

$$18.2 \pm 3.355(0.6633)$$

$$15.976 < E(Y_h) < 20.424$$

At $X_h = 4$

$\widehat{Y}_h = 10.2 + 4\,(4) = 26.2$

$$s^2(\widehat{Y}_h) = MSE\left(\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right) = \mathbf{2.2}\left(\frac{1}{10} + \frac{(4-1)^2}{10}\right) = 2.2$$

$s(\widehat{Y}_h) = \sqrt{2.2} = 1.483$

$t\left(1 - \frac{\alpha}{2}; n - 2\right) = t(0.995; 8) = 3.355$

$26.2 \pm 3.355(1.483)$

$$12.748 < E(Y_h) < 23.652$$

We conclude that the mean number of ampules found to be broken upon arrival when 2 transfers from one aircraft to another over the shipment route of 2 are produced is somewhere between 15.976 and 20.424 ampules

أن متوسط عدد أمبولات وجدت منكسره عند وصولهم عندما تم نقله عبر 2 مرات من طائرة واحدة إلى آخر عبر مسار الشحنة , بين 15.976 و 20.424 أمبوله.

We conclude that the mean number of ampules found to be broken upon arrival when 4 transfers from one aircraft to another over the shipment route are produced is somewhere between 12.748 and 23.652 ampules.

أن متوسط عدد أمبولات وجدت منكسره عند وصولهم عندما تم نقله عبر 4 مرات من طائرة واحدة إلى آخر عبر مسار الشحنة , بين 12.748 و 23.652 أمبوله.

b. The next shipment will entail two transfers. Obtain a 99 percent <u>prediction</u> interval for the number of broken ampules for this shipment. Interpret your prediction interval.

$$s^2(\widehat{Y_{new}}) = MSE\left(1 + \frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right) = \mathbf{2.2}\left(1 + \frac{1}{10} + \frac{(2-1)^2}{10}\right) = 2.64$$

$s(\widehat{Y}_h) = \sqrt{2.64} = 1.6248$

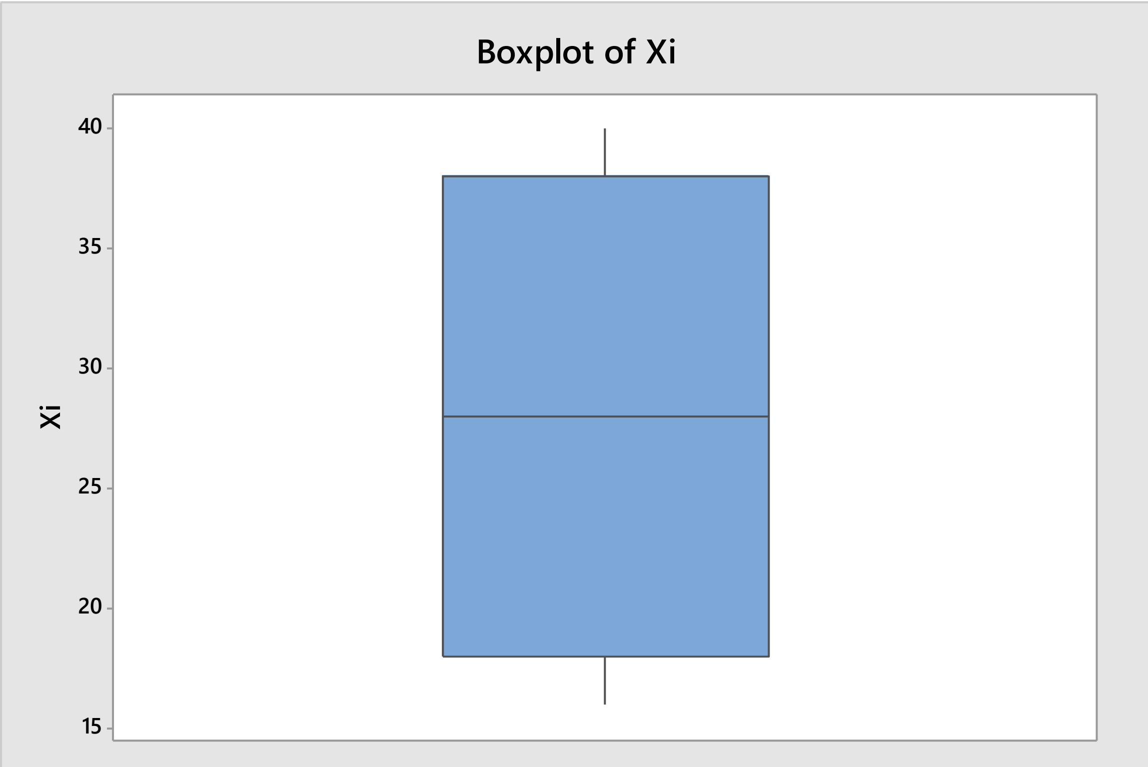
$18.2 \pm 3.355(1.6248)$

$$12.748 < Y_{h(new)} < 23.652$$

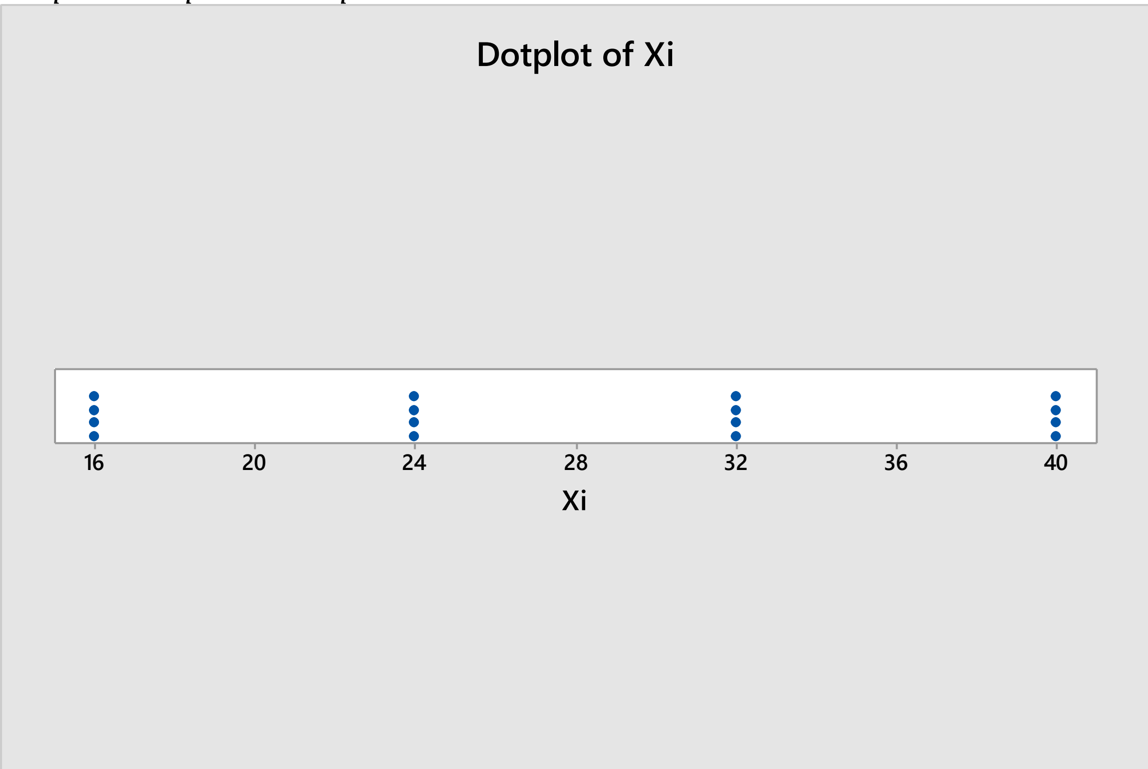
With confidence coefficient 0.99, we predict that the mean number of ampules found to be broken upon arrival when 2 transfers from one aircraft to another over the shipment route of 2 are produced is somewhere between 12.748 and 23.652 ampules.

Refer to Plastic hardness.

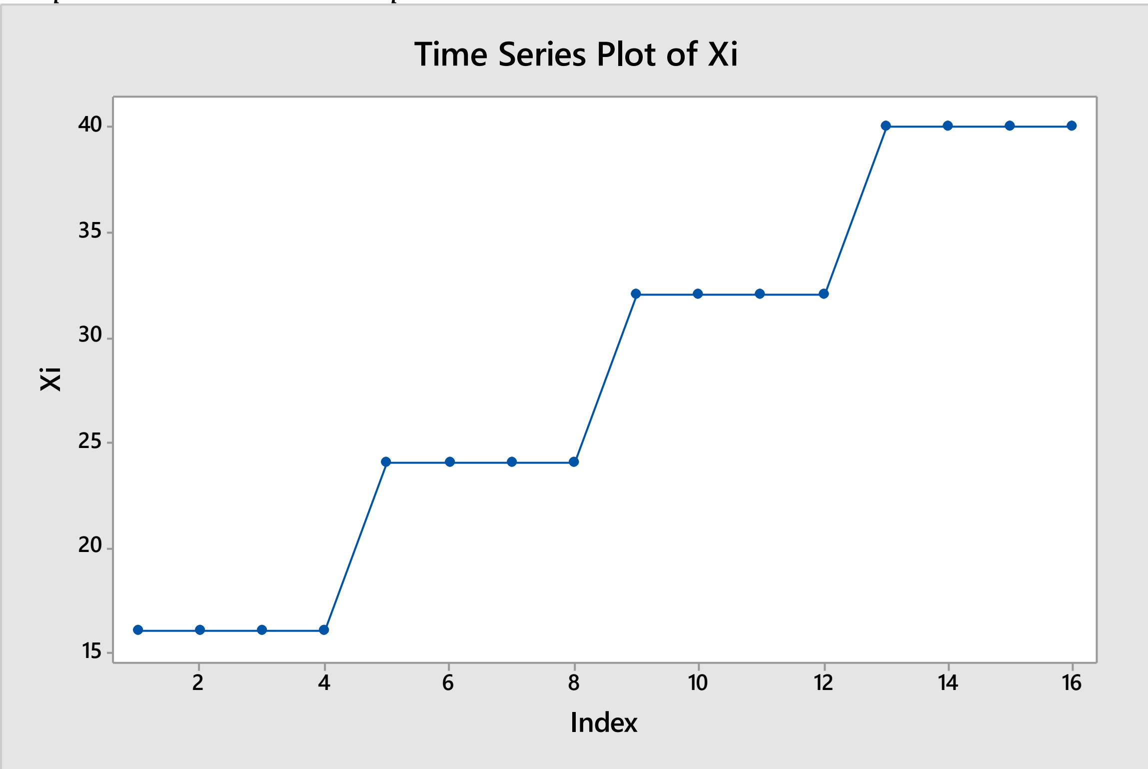
$Graph \rightarrow Boxplot \rightarrow simple \rightarrow X \rightarrow ok$



$Graph \rightarrow Dotplot \rightarrow simple \rightarrow X \rightarrow ok$

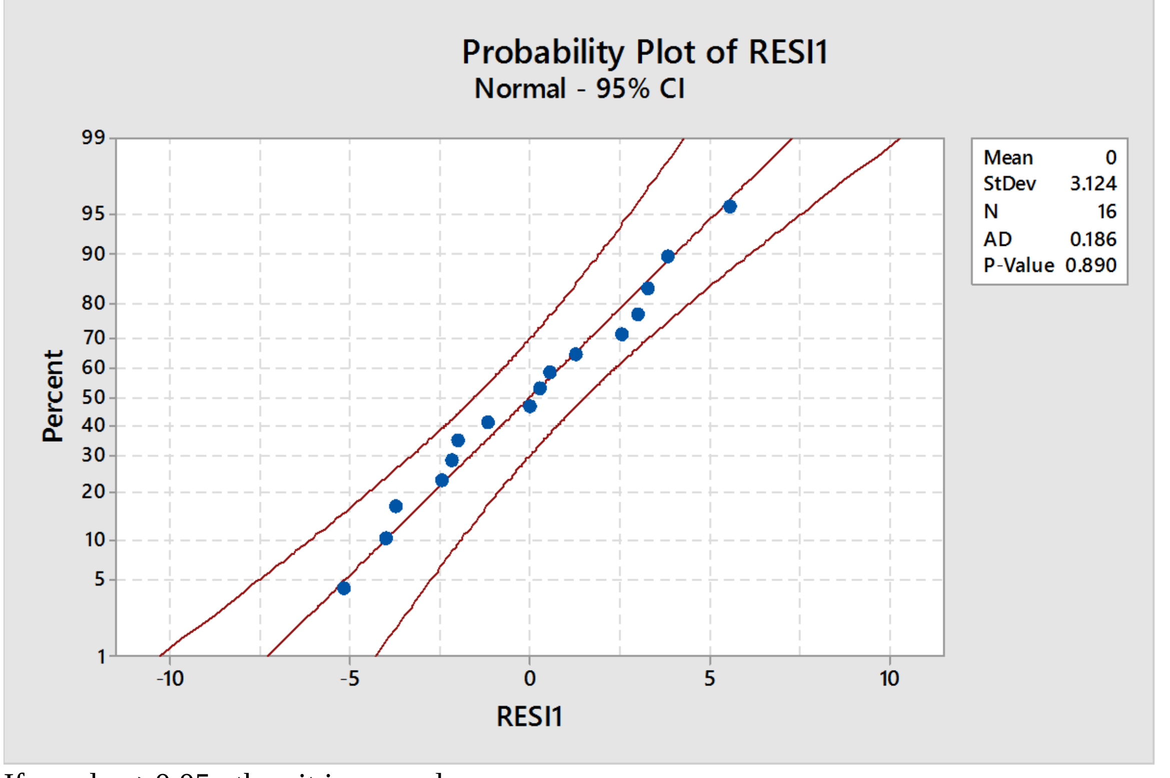


$Graph \rightarrow time\ series \rightarrow simple \rightarrow X \rightarrow ok$



For test normality of residuals

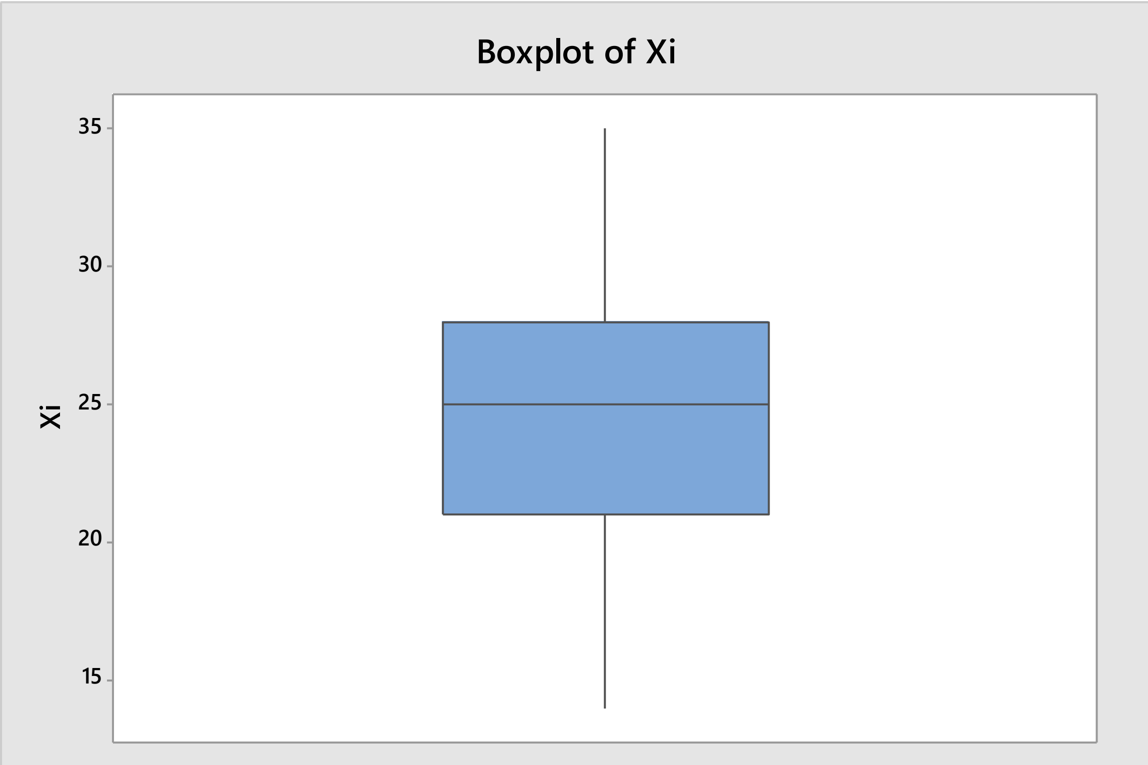
$Graph \rightarrow probability\ plot \rightarrow single \rightarrow (distribution\ Normal) \rightarrow X \rightarrow ok$



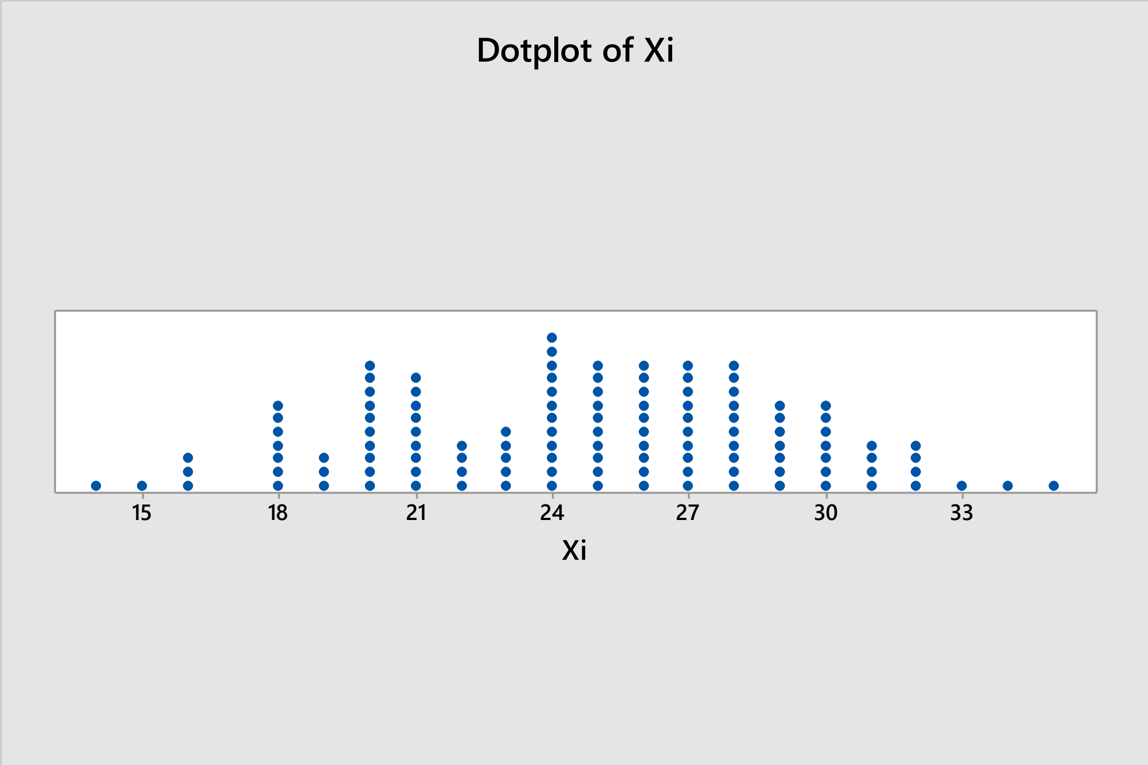
If p-value >0.05 , then it is normal

Refer to Grade point average.

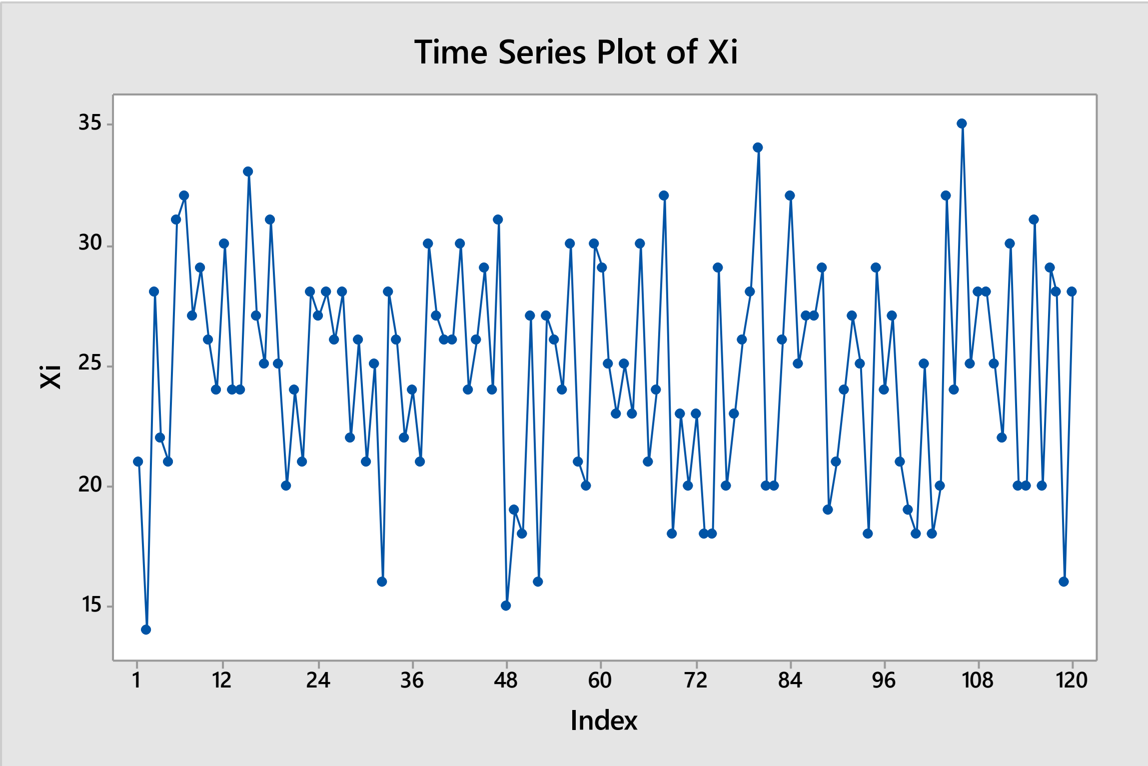
$Graph \rightarrow Boxplot \rightarrow simple \rightarrow X \rightarrow ok$



$Graph \rightarrow Dotplot \rightarrow simple \rightarrow X \rightarrow ok$

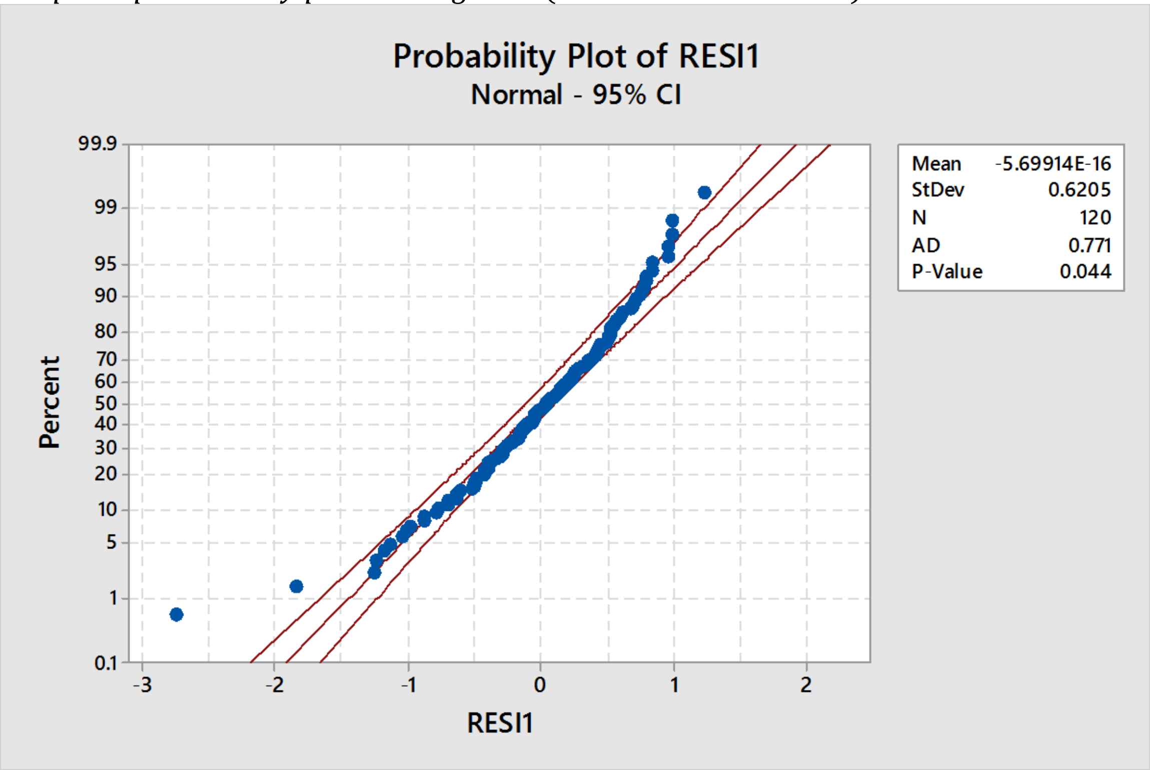


$Graph \rightarrow time\ series \rightarrow simple \rightarrow X \rightarrow ok$



For test normality of residuals

$Graph \rightarrow probability\ plot \rightarrow single \rightarrow (distribution\ Normal) \rightarrow X \rightarrow ok$



At 0.01 it is normal but at 0.05 , it is not normal

Q5.1. For the matrices below, obtain (1) A +B, (2) A - B, (3) AC, (4) AB', (5) B'A.

$$A = \begin{bmatrix} 1 & 4 \\ 2 & 6 \\ 3 & 8 \end{bmatrix} \quad B = \begin{bmatrix} 1 & 3 \\ 1 & 4 \\ 2 & 5 \end{bmatrix} \quad C = \begin{bmatrix} 3 & 8 & 1 \\ 5 & 4 & 0 \end{bmatrix}$$

Solution:

$$A + B = \begin{bmatrix} 1 & 4 \\ 2 & 6 \\ 3 & 8 \end{bmatrix} + \begin{bmatrix} 1 & 3 \\ 1 & 4 \\ 2 & 5 \end{bmatrix} = \begin{bmatrix} 1+1 & 4+3 \\ 2+1 & 6+4 \\ 3+2 & 8+5 \end{bmatrix} = \begin{bmatrix} 2 & 7 \\ 3 & 10 \\ 5 & 13 \end{bmatrix}$$

$$A - B = \begin{bmatrix} 1 & 4 \\ 2 & 6 \\ 3 & 8 \end{bmatrix} - \begin{bmatrix} 1 & 3 \\ 1 & 4 \\ 2 & 5 \end{bmatrix} = \begin{bmatrix} 1-1 & 4-3 \\ 2-1 & 6-4 \\ 3-2 & 8-5 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}$$

$$AC = \begin{bmatrix} 1 & 4 \\ 2 & 6 \\ 3 & 8 \end{bmatrix} \begin{bmatrix} 3 & 8 & 1 \\ 5 & 4 & 0 \end{bmatrix} = \begin{bmatrix} 1*3+4*5 & 1*8+4*4 & 1*1+4*0 \\ 2*3+6*5 & 2*8+6*4 & 2*1+6*0 \\ 3*3+8*5 & 3*8+8*4 & 3*1+8*0 \end{bmatrix} = \begin{bmatrix} 23 & 24 & 1 \\ 36 & 40 & 2 \\ 49 & 56 & 3 \end{bmatrix}$$

$$B' = \begin{bmatrix} 1 & 1 & 2 \\ 3 & 4 & 5 \end{bmatrix}$$

$$(A_{3\times2}B'_{2\times3})_{3\times3} = \begin{bmatrix} 1 & 4 \\ 2 & 6 \\ 3 & 8 \end{bmatrix} \begin{bmatrix} 1 & 1 & 2 \\ 3 & 4 & 5 \end{bmatrix} = \begin{bmatrix} 1*1+4*3 & 1*1+4*4 & 1*2+4*5 \\ 2*1+6*3 & 2*1+6*4 & 2*2+6*5 \\ 3*1+8*3 & 3*1+8*4 & 3*2+8*5 \end{bmatrix} = \begin{bmatrix} 13 & 17 & 22 \\ 20 & 26 & 34 \\ 27 & 35 & 46 \end{bmatrix}$$

$$(B'_{2\times3}A_{3\times2})_{2\times2} = \begin{bmatrix} 1 & 1 & 2 \\ 3 & 4 & 5 \end{bmatrix} \begin{bmatrix} 1 & 4 \\ 2 & 6 \\ 3 & 8 \end{bmatrix} = \begin{bmatrix} 1*1+1*2+2*3 & 1*4+1*6+2*8 \\ 3*1+4*2+5*3 & 3*4+4*6+5*8 \end{bmatrix} = \begin{bmatrix} 9 & 26 \\ 26 & 76 \end{bmatrix}$$

Q5.4. <u>Flavor deterioration</u>. The results shown below were obtained in a small-scale experiment to study the relation between ^{o}F of storage temperature (X) and number of weeks before flavour deterioration of a food product begins to occur (Y).

i	**1**	**2**	**3**	**4**	**5**
X_i	**8**	**4**	**0**	**-4**	**-8**
Y_i	**7.8**	**9.0**	**10.2**	**11.0**	**11.7**

Assume that first-order regression model (2.1) is applicable. Using matrix methods, find (1) $Y'Y$, (2) $X'X$, (3) $X'Y$.

$$\boldsymbol{Y}_{n\times 1} = \boldsymbol{X}_{n\times 2}\mathbf{B}_{2\times 1} + \boldsymbol{\varepsilon}_{n\times 1}$$
$$\boldsymbol{E}(\boldsymbol{Y}_{n\times 1}) = \boldsymbol{X}_{n\times 2}\mathbf{B}_{2\times 1}$$
$$\mathbf{B}_{2\times 1} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$$

$$\boldsymbol{V}(\mathbf{B}) = MSE(\boldsymbol{X}'\boldsymbol{X})^{-1}$$
$$MSE = \frac{\boldsymbol{e}'\boldsymbol{e}}{n-2}$$

	C1	C2	C3
1	8	7.8	1
2	4	9.0	1
3	0	10.2	1
4	-4	11.0	1
5	-8	11.7	1

$$\boldsymbol{X} = \begin{bmatrix} 1 & 8 \\ 1 & 4 \\ 1 & 0 \\ 1 & -4 \\ 1 & -8 \end{bmatrix}, \quad \boldsymbol{Y} = \begin{bmatrix} 7.8 \\ 9.0 \\ 10.2 \\ 11.0 \\ 11.7 \end{bmatrix}, \quad \boldsymbol{X}' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 8 & 4 & 0 & -4 & -8 \end{bmatrix}$$

$$(\boldsymbol{X'X})_{\mathbf{2\times2}} = \begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 8 & 4 & 0 & -4 & -8 \end{bmatrix} \begin{bmatrix} 1 & 8 \\ 1 & 4 \\ 1 & 0 \\ 1 & -4 \\ 1 & -8 \end{bmatrix} = \begin{bmatrix} 1+1+1+1+1 & 8+4+0-4-8 \\ 8+4+0-4-8 & 8*8+4*4+0*0+-4*-4+-8*-8 \end{bmatrix}$$

$$= \begin{bmatrix} 5 & 0 \\ 0 & 160 \end{bmatrix}$$

$$(\boldsymbol{X'X})^{-\mathbf{1}} = \frac{\mathbf{1}}{\Delta} \begin{bmatrix} \sum X_i^2 & -\sum X_i \\ -\sum X_i & \boldsymbol{n} \end{bmatrix}$$

$$\Delta = \boldsymbol{n} \sum X_i^2 - \sum X_i \sum X_i = 5 * 160 - 0 = 800$$

$$(\boldsymbol{X'X})^{-\mathbf{1}} = \frac{1}{800} \begin{bmatrix} 160 & 0 \\ 0 & 5 \end{bmatrix} = \begin{bmatrix} 0.2 & 0 \\ 0 & 0.00625 \end{bmatrix}$$

$$(\boldsymbol{X'}_{\mathbf{2\times n}} \boldsymbol{Y}_{\boldsymbol{n}\times\mathbf{1}})_{\mathbf{2\times1}} = \begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 8 & 4 & 0 & -4 & -8 \end{bmatrix} \begin{bmatrix} 7.8 \\ 9.0 \\ 10.2 \\ 11.0 \\ 11.7 \end{bmatrix} = \begin{bmatrix} 7.8+9+10.2+11+11.7 \\ 8*7.8+4*9+0*10.2-4*11-8*11.7 \end{bmatrix} = \begin{bmatrix} 49.7 \\ -39.2 \end{bmatrix}$$

$$\mathbf{B}_{\mathbf{2\times1}} = (\boldsymbol{X'X})^{-\mathbf{1}} \boldsymbol{X'Y} = \begin{bmatrix} 0.2 & 0 \\ 0 & 0.00625 \end{bmatrix} \begin{bmatrix} 49.7 \\ -39.2 \end{bmatrix} = \begin{bmatrix} 0.2*49.7+0*-39.2 \\ 0*49.7+0.00625*-39.2 \end{bmatrix} = \begin{bmatrix} 9.94 \\ -0.245 \end{bmatrix}$$

$$\hat{Y} = 9.940 - 0.245X$$

$$\widehat{\boldsymbol{Y}}_{\boldsymbol{n}\times\mathbf{1}} = \boldsymbol{X}_{\boldsymbol{n}\times\mathbf{2}} \mathbf{B}_{\mathbf{2\times1}} = \begin{bmatrix} 1 & 8 \\ 1 & 4 \\ 1 & 0 \\ 1 & -4 \\ 1 & -8 \end{bmatrix} \begin{bmatrix} 9.94 \\ -0.245 \end{bmatrix} = \begin{bmatrix} 9.94-0.245*8 \\ 9.94-0.245*4 \\ 9.94-0.245*0 \\ 9.94+0.245*4 \\ 9.94+0.245*8 \end{bmatrix} = \begin{bmatrix} 7.98 \\ 8.96 \\ 9.94 \\ 10.92 \\ 11.9 \end{bmatrix}$$

$$\boldsymbol{e}_{\boldsymbol{n}\times\mathbf{1}} = \boldsymbol{Y}_{\boldsymbol{n}\times\mathbf{1}} - \widehat{\boldsymbol{Y}}_{\boldsymbol{n}\times\mathbf{1}} = \begin{bmatrix} 7.8 \\ 9.0 \\ 10.2 \\ 11.0 \\ 11.7 \end{bmatrix} - \begin{bmatrix} 7.98 \\ 8.96 \\ 9.94 \\ 10.92 \\ 11.9 \end{bmatrix} = \begin{bmatrix} 7.8-7.98 \\ 9-8.96 \\ 10.2-9.94 \\ 11-10.92 \\ 11.7-11.9 \end{bmatrix} = \begin{bmatrix} -0.18 \\ 0.04 \\ 0.26 \\ 0.08 \\ -0.2 \end{bmatrix}$$

$$\boldsymbol{e'}_{\boldsymbol{1\times n}}\boldsymbol{e}_{\boldsymbol{n\times 1}} = \left[\sum e_i^2\right] = [-0.18 \quad 0.04 \quad 0.26 \quad 0.08 \quad -0.2]\begin{bmatrix}-0.18\\0.04\\0.26\\0.08\\-0.2\end{bmatrix} = [0.148]$$

$$MSE = \frac{0.148}{3} = 0.049333$$

$$\boldsymbol{V}(\mathbf{B}) = MSE(\boldsymbol{X'X})^{-\mathbf{1}} = 0.049333\begin{bmatrix}0.2 & 0\\0 & 0.00625\end{bmatrix} = \begin{bmatrix}0.009867 & 0\\0 & 0.000308\end{bmatrix}$$

$$\boldsymbol{V}\begin{bmatrix}\widehat{\beta_0}\\\widehat{\beta_1}\end{bmatrix} = \begin{bmatrix}Var(\widehat{\beta_0}) & cov(\widehat{\beta_0},\widehat{\beta_1})\\cov(\widehat{\beta_0},\widehat{\beta_1}) & Var(\widehat{\beta_1})\end{bmatrix}$$

$$\boldsymbol{Y'}_{\boldsymbol{1\times n}}\boldsymbol{Y}_{\boldsymbol{n\times 1}} = \left[\sum y_i^2\right] = [7.8 \quad 9.0 \quad 10.2 \quad 11.0 \quad 11.7]\begin{bmatrix}7.8\\9.0\\10.2\\11.0\\11.7\end{bmatrix} = [503.77]$$

```
MTB > Copy C3 C1 m1
MTB > Print m1
```
$\boldsymbol{X}_{\boldsymbol{n\times 2}}$

```
Data Display

 Matrix M1

1   8
1   4
1   0
1  -4
1  -8
```
$\boldsymbol{X}_{\boldsymbol{n\times 2}}$

```
MTB > tran m1 m2
MTB > print m2
```
$\boldsymbol{X'}_{\boldsymbol{2\times n}}$

```
Data Display

 Matrix M2

1  1  1   1   1
8  4  0  -4  -8
```
$\boldsymbol{X'}_{\boldsymbol{2\times n}}$

```
MTB > mult m2 m1 m3
MTB > print m3
```
$(\boldsymbol{X'X})_{\boldsymbol{2\times 2}}$

```
Data Display

 Matrix M3
```
$$(\boldsymbol{X'X})_{\boldsymbol{2\times 2}} = \begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix}$$

```
5    0
0  160

MTB > inver m3 m4
MTB > print m4
```
$(\boldsymbol{X'X})^{-\boldsymbol{1}}_{\boldsymbol{2\times 2}}$

```
Data Display

 Matrix M4

0.2  0.00000
0.0  0.00625

MTB > copy c2 m5
MTB > Print m5
```
$\boldsymbol{Y}_{\boldsymbol{n\times 1}}$

```
Data Display

 Matrix M5

 7.8
 9.0
10.2
11.0
11.7
```

$\boldsymbol{Y}_{\boldsymbol{n}\times\mathbf{1}}$

```
MTB > mult m2 m5 m6
MTB > print m6
```

$\boldsymbol{X'}_{\mathbf{2}\times\boldsymbol{n}}\boldsymbol{Y}_{\boldsymbol{n}\times\mathbf{1}} = \boldsymbol{X'Y}_{\mathbf{2}\times\mathbf{1}}$

```
Data Display

 Matrix M6

 49.7
-39.2

MTB > mult m4 m6 m7
MTB > print m7
```

$(\boldsymbol{X'X})^{-\mathbf{1}}_{\mathbf{2}\times\mathbf{2}}(\boldsymbol{X'Y})_{\mathbf{2}\times\mathbf{1}} = \mathbf{B}_{\mathbf{2}\times\mathbf{1}}$

```
Data Display

 Matrix M7

 9.940
-0.245
```

$\mathbf{B}_{\mathbf{2}\times\mathbf{1}}$

$$\hat{Y} = 9.940 - 0.245X$$

```
MTB > tran m5 m13
MTB > print m13
```

Data Display

```
 Matrix M13

7.8  9  10.2  11  11.7

MTB > mult m13 m5 m14

Answer = 503.7700
```

```
MTB > mult m1 m7 m8
```

$\hat{Y}_{n\times 1} = X_{n\times 2}\mathbf{B}_{2\times 1}$

```
MTB > print m8

Data Display

 Matrix M8

 7.98
 8.96
 9.94
10.92
11.90
```

```
MTB > copy m8 c4
```

```
MTB > Let c5 = 'y'-C4
```

```
MTB > copy c5 m9
```
$\boldsymbol{e}$

```
MTB > tran m9 m10
MTB > print m10
```
$\boldsymbol{e'}$

Data Display

```
 Matrix M10

-0.18  0.04  0.26  0.08  -0.2
```

```
MTB > mult m10 m9 m11
```
$\boldsymbol{e'e}$

```
Answer = 0.1480
```

MSE=0.1480/3=0.049333 $\boldsymbol{MSE = \frac{e'e}{n-2}}$

```
MTB > mult 0.049333 m4 m12
MTB > print m12
```

$$\boldsymbol{V(\mathbf{B}) = MSE(X'X)^{-1}}$$

Data Display

```
 Matrix M12

0.0098666  0.0000000
0.0000000  0.0003083
```

$$\boldsymbol{V}\begin{bmatrix}\widehat{\beta_0}\\ \widehat{\beta_1}\end{bmatrix} = \begin{bmatrix}\boldsymbol{Var}(\widehat{\beta_0}) & \boldsymbol{cov}(\widehat{\beta_0},\widehat{\beta_1})\\ \boldsymbol{cov}(\widehat{\beta_0},\widehat{\beta_1}) & \boldsymbol{Var}(\widehat{\beta_1})\end{bmatrix}$$

Regression Analysis: y versus x

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	9.6040	9.60400	194.68	0.001
x	1	9.6040	9.60400	194.68	0.001
Error	3	0.1480	0.04933		
Total	4	9.7520			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.222111	98.48%	97.98%	94.11%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	9.9400	0.0993	100.07	0.000	
x	-0.2450	0.0176	-13.95	0.001	1.00

Regression Equation

y = 9.9400 - 0.2450 x

H.W

Q5.2 For the matrices below, obtain (1) $A + C$, (2) $A - C$, (3) $B'A$, (4) AC', (5) $C'A$.

$$A = \begin{bmatrix} 2 & 1 \\ 3 & 5 \\ 5 & 7 \\ 4 & 8 \end{bmatrix} \quad B = \begin{bmatrix} 6 \\ 9 \\ 3 \\ 1 \end{bmatrix} \quad C = \begin{bmatrix} 3 & 8 \\ 8 & 6 \\ 5 & 1 \\ 2 & 4 \end{bmatrix}$$

Q5.5 Consumer finance. The data below show, for a consumer finance company operating in six cities, the number of competing loan companies operating in the city *(X)* and the number per thousand of the company's loans made in that city that are currently delinquent *(Y);*

i	1	2	3	4	5	6
X_i	4	1	2	3	3	4
Y_i	16	5	10	15	13	22

Assume that first-order regression model (2.1) is applicable. Using matrix methods, find (1) $Y'Y$, (2) $X'X$, (3) $X'Y$.

Chapter 6

Q6.10. Refer to **Grocery retailer**
a. Fit regression model (6.5) to the data for three predictor variables. State the estimated regression function. How are bl , *b2* , and *b3* interpreted here?

e. Estimate β_1, β_2 and β_3 jointly confidence interval, using a 95 percent confidence coefficient.

	C1	C2	C3	C4	C5
1	4264	305657	7.17	0	1
2	4496	328476	6.20	0	1
3	4317	317164	4.61	0	1
⋮	⋮	⋮	⋮	⋮	⋮
52	4342	292087	7.77	0	1

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$$
$$\boldsymbol{Y_{n\times 1}} = \boldsymbol{X_{n\times 4}} \mathbf{B_{4\times 1}} + \boldsymbol{\varepsilon_{n\times 1}}$$
$$\boldsymbol{E}(\boldsymbol{Y_{n\times 1}}) = \boldsymbol{X_{n\times 4}} \mathbf{B_{4\times 1}}$$
$$\mathbf{B_{4\times 1}} = (\boldsymbol{X'X})^{-1}\boldsymbol{X'Y}$$

$$\boldsymbol{V}(\mathbf{B}) = MSE(\boldsymbol{X'X})^{-1}$$
$$MSE = \frac{\boldsymbol{e'e}}{n-(p+1)}$$

$$\hat{\beta}_k - t_{1-\frac{\alpha}{2}, n-(p+1)} S.E(\hat{\beta}_k) \leq \beta_k \leq \hat{\beta}_k + t_{1-\frac{\alpha}{2}, n-(p+1)} S.E(\hat{\beta}_k)$$

```
MTB > copy c5 c2 c3 c4 m1     X
MTB > tran m1 m2              X'
MTB > mult m2 m1 m3           X'X
MTB > print m3
```

Data Display

```
Matrix M3

      52  1.57400E+07        383        6
15740042  4.92022E+12  116223168  1857680
     383  1.16223E+08       2864       46
       6  1.85768E+06         46        6
```

$$(\boldsymbol{X'X})_{\boldsymbol{4\times 4}} = \begin{bmatrix} \boldsymbol{n} & \sum X_{1i} & \sum X_{2i} & \sum X_{3i} \\ \sum X_{1i} & \sum {X_{1i}}^2 & \sum X_{1i}X_{2i} & \sum X_{1i}X_{3i} \\ \sum X_{2i} & \sum X_{1i}X_{2i} & \sum {X_{2i}}^2 & \sum X_{2i}X_{3i} \\ \sum X_{3i} & \sum X_{1i}X_{3i} & \sum X_{2i}X_{3i} & \sum {X_{3i}}^2 \end{bmatrix} = \begin{bmatrix} \mathbf{52} & 1.57400\text{E}+07 & 383 & 6 \\ 15740042 & 4.92022\text{E}+12 & 116223168 & 1857680 \\ 383 & 1.16223\text{E}+08 & 2864 & 46 \\ 6 & 1.85768\text{E}+06 & 46 & 6 \end{bmatrix}$$

```
MTB > inver m3 m4
MTB > print m4
```

$(\boldsymbol{X'X})^{-\mathbf{1}}$

Data Display

```
Matrix M4

 1.86275  -0.0000017  -0.180557   0.047324
-0.00000   0.0000000  -0.000000  -0.000000
-0.18056  -0.0000000   0.025971  -0.007750
 0.04732  -0.0000000  -0.007750   0.191112
```

```
MTB > copy c1 m5
```

Y

```
MTB > mult m2 m5 m6
```

$$\boldsymbol{X'}_{\boldsymbol{4\times n}}\boldsymbol{Y}_{\boldsymbol{n\times 1}} = \boldsymbol{X'Y}_{\boldsymbol{4\times 1}}$$

```
MTB > print m6
```

Data Display

```
Matrix M6
```

```
2.26878E+05
6.88202E+10
1.67289E+06
2.94990E+04
```

$$(\boldsymbol{X'Y})_{\mathbf{4\times1}} = \begin{bmatrix} \sum Y_i \\ \sum Y_i X_{1i} \\ \sum Y_i X_{2i} \\ \sum Y_i X_{3i} \end{bmatrix} = \begin{bmatrix} 226878 \\ 68820200000 \\ 1672890 \\ 29499 \end{bmatrix}$$

```
MTB > mult m4 m6 m7
MTB > print m7
```

$$(\boldsymbol{X'X})^{-\mathbf{1}}_{\mathbf{4\times4}}(\boldsymbol{X'Y})_{\mathbf{4\times1}} = \mathbf{B}_{\mathbf{4\times1}}$$

Data Display

```
 Matrix M7

4149.89
   0.00
 -13.17
 623.55
```

$$\mathbf{B}_{\mathbf{4\times1}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \begin{bmatrix} 4149.89 \\ 0.00 \approx 0.000787 \\ -13.17 \\ 623.55 \end{bmatrix}$$

$$\hat{Y} = 4149.89 + 0.000787\ X_1 - 13.17\ X_2 + 623.55\ X_3$$

```
MTB > tran m5 m8
```
$\boldsymbol{Y'}$

```
MTB > mult m8 m5 m9
```
$\boldsymbol{Y'Y} = \sum {Y_i}^2$

```
Answer = 993039576.0000

MTB > mult m1 m7 m10
MTB > copy m10 c6
```
$\boldsymbol{\hat{Y}_{n\times 1}} = \boldsymbol{X_{n\times 4}}\mathbf{B_{4\times 1}}$

```
MTB > Let C7=C1-C6
```
$\boldsymbol{e} = \boldsymbol{Y} - \boldsymbol{\hat{Y}}$

```
MTB > copy c7 m11
```
$\boldsymbol{e}$

```
MTB > tran m11 m12
```
$\boldsymbol{e'}$

```
MTB > mult m12 m11 m13
```
$\boldsymbol{e'e} = \boldsymbol{SSE}$

```
Answer = 985529.7464
```

MSE=985529.7464/(52-4)= 20531.87 $\quad MSE = \frac{e'e}{n-(p+1)}$

```
MTB > mult 20531.87 m4 m14
```
$$\boldsymbol{V}(\mathbf{B}) = \boldsymbol{MSE}(\boldsymbol{X'X})^{-1}$$

```
MTB > print m14
```

Data Display

```
 Matrix M14

38245.8  -0.0351482  -3707.17   971.66
   -0.0   0.0000001     -0.00    -0.00
-3707.2  -0.0006762    533.23  -159.12
```

971.7 -0.0008312   -159.12 3923.89

$$\boldsymbol{V}\begin{bmatrix}\hat{\beta}_0\\ \hat{\beta}_1\\ \hat{\beta}_2\\ \hat{\beta}_3\end{bmatrix}=\begin{bmatrix}\boldsymbol{Var}(\widehat{\beta_0}) & \boldsymbol{cov}(\widehat{\beta_0},\widehat{\beta_1}) & \boldsymbol{cov}(\widehat{\beta_0},\widehat{\beta_2}) & \boldsymbol{cov}(\widehat{\beta_0},\widehat{\beta_3})\\ \boldsymbol{cov}(\widehat{\beta_0},\widehat{\beta_1}) & \boldsymbol{Var}(\widehat{\beta_1}) & \boldsymbol{cov}(\widehat{\beta_1},\widehat{\beta_2}) & \boldsymbol{cov}(\widehat{\beta_1},\widehat{\beta_3})\\ \boldsymbol{cov}(\widehat{\beta_0},\widehat{\beta_2}) & \boldsymbol{cov}(\widehat{\beta_2},\widehat{\beta_1}) & \boldsymbol{Var}(\widehat{\beta_2}) & \boldsymbol{cov}(\widehat{\beta_2},\widehat{\beta_3})\\ \boldsymbol{cov}(\widehat{\beta_0},\widehat{\beta_3}) & \boldsymbol{cov}(\widehat{\beta_1},\widehat{\beta_3}) & \boldsymbol{cov}(\widehat{\beta_2},\widehat{\beta_3}) & \boldsymbol{Var}(\widehat{\beta_3})\end{bmatrix}=\begin{bmatrix}38245.8 & -0.0351482 & -3707.17 & 971.66\\ -0.0351482 & 0.0000001 & -0.0006762 & -0.0008312\\ -3707.2 & -0.0006762 & 533.23 & -159.12\\ 971.7 & -0.0008312 & -159.12 & 3923.89\end{bmatrix}$$

$$\hat{\beta}_k - t_{1-\frac{\alpha}{2},n-(p+1)} S.E(\hat{\beta}_k) \le \beta_k \le \hat{\beta}_k + t_{1-\frac{\alpha}{2},n-(p+1)} S.E(\hat{\beta}_k)$$

$t_{1-\frac{\alpha}{2},n-(p+1)} = t_{0.975,52-4} = 2.0106$

$S.E(\hat{\beta}_0) = \sqrt{38245.8} = 195.5653$

$S.E(\hat{\beta}_1) = \sqrt{0.0000001} = 0.000316$

$S.E(\hat{\beta}_2) = \sqrt{533.23} = 23.09177$

$S.E(\hat{\beta}_3) = \sqrt{3923.89} = 62.64096$

$\hat{\beta}_0 - t_{0.975,52-4} S.E(\hat{\beta}_0) \le \beta_0 \le \hat{\beta}_0 + t_{0.975,52-4} S.E(\hat{\beta}_0)$

$4149.89 - 2.0106 * 195.5653 \le \beta_0 \le 4149.89 + 2.0106 * 195.5653$

$3756.686 \le \beta_0 \le 4543.094$

$\hat{\beta}_1 - t_{0.975,52-4} S.E(\hat{\beta}_1) \le \beta_1 \le \hat{\beta}_1 + t_{0.975,52-4} S.E(\hat{\beta}_1)$

$0.000787 - 2.0106 * 0.000316 \le \beta_1 \le 0.000787 + 2.0106 * 0.000316$

$0.000151 \le \beta_1 \le 0.001423$

$\hat{\beta}_2 - t_{0.975,52-4} S.E(\hat{\beta}_2) \le \beta_2 \le \hat{\beta}_2 + t_{0.975,52-4} S.E(\hat{\beta}_2)$

$-13.17 - 2.0106 * 23.09177 \le \beta_2 \le -13.17 + 2.0106 * 23.09177$

$-59.5983 \le \beta_2 \le 33.25832$

$\hat{\beta}_3 - t_{0.975,52-4} S.E(\hat{\beta}_3) \leq \beta_3 \leq \hat{\beta}_3 + t_{0.975,52-4} S.E(\hat{\beta}_3)$
$623.55 - 2.0106 * 62.64096 \leq \beta_3 \leq 623.55 + 2.0106 * 62.64096$
$497.6041 \leq \beta_3 \leq 749.4959$

1. Hypothesis
$H_0: \beta_k = 0$

$H_1: \beta_k \neq 0$
2. Test statistic
$$T_0 = \frac{b_k - \beta_{k0}}{s(b_k)} = \frac{b_k}{s(b_k)}$$
3. Decision: Reject H_0 if $|T_0| > t_{\left(1-\frac{\alpha}{2}, n-(p+1)\right)}$
p-value=$2P\left(t_{(n-(p+1))} > |T_0|\right)$
Reject H_0 if $\text{p} - \text{value} \leq \alpha$

1. Hypothesis
$H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$
2. Test statistic
$$T_0 = \frac{b_1 - \beta_{10}}{s(b_1)} = \frac{b_1}{s(b_1)} = \frac{0.000787}{0.000316} = 2.4887$$
3. Decision: Reject H_0 if $|T_0| > t_{\left(1-\frac{\alpha}{2}, n-(p+1)\right)} = 2.0106$
p-value=$2P\left(t_{(n-(p+1))} > |T_0|\right) = 2P\left(t_{(48)} > |2.4887|\right) = 2(1 - 0.9918) = 0.0164$
reject H_0

1. Hypothesis

$H_0: \beta_2 = 0$

$H_1: \beta_2 \neq 0$

2. Test statistic

$$T_0 = \frac{b_2}{s(b_2)} = \frac{-13.17}{23.091} = -0.5703$$

3. Decision: Reject H_0 if $|T_0| > t_{\left(1-\frac{\alpha}{2}, n-(p+1)\right)} = 2.0106$

p-value=$2P\left(t_{(n-(p+1))} > |T_0|\right) = 2P\left(t_{(48)} > |-0.5703|\right) = 2(1 - 0.7144) = 0.5712$

not reject H_0

1. Hypothesis

$H_0: \beta_3 = 0$

$H_1: \beta_3 \neq 0$

2. Test statistic

$$T_0 = \frac{b_3}{s(b_3)} = \frac{623.55}{62.6409} = 9.9543$$

3. Decision: Reject H_0 if $|T_0| > t_{\left(1-\frac{\alpha}{2}, n-(p+1)\right)} = 2.0106$

p-value=$2P\left(t_{(n-(p+1))} > |T_0|\right) = 2P\left(t_{(48)} > |9.9543|\right) = 2(1 - 1) = 0.00$

reject H_0

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	2176606	725535	35.34	0.000
X1	1	95707	95707	4.66	0.036
X2	1	6675	6675	0.33	0.571
X3	1	2034514	2034514	99.09	0.000
Error	48	985530	20532		
Total	51	3162136			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
143.289	68.83%	66.89%	64.78%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	4150	196	21.22	0.000	
X1	0.000787	0.000365	2.16	0.036	1.01
X2	-13.2	23.1	-0.57	0.571	1.02
X3	623.6	62.6	9.95	0.000	1.01

Regression Equation

Y = 4150 + 0.000787 X1 - 13.2 X2 + 623.6 X3

Q8.6. Refer to **Steroid level**

a. Fit regression model (8.2) to the data for three predictor variables. State the estimated regression function.

e. Estimate β_1 and β_2 jointly confidence interval, using a 95 percent confidence coefficient.

	C1	C2	C3	C4
1	27.1	23	529	1
2	22.1	19	361	1
3	21.9	25	625	1
⋮	⋮	⋮	⋮	⋮
27	20.6	18	324	1

$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \epsilon$
Let $X_1 = X$, and $X_2 = X^2$, then
$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$
$\boldsymbol{Y_{n\times 1}} = \boldsymbol{X_{n\times 3}}\mathbf{B_{3\times 1}} + \boldsymbol{\varepsilon_{n\times 1}}$
$\boldsymbol{E}(\boldsymbol{Y_{n\times 1}}) = \boldsymbol{X_{n\times 3}}\mathbf{B_{3\times 1}}$
$\mathbf{B_{3\times 1}} = (\boldsymbol{X'X})^{-1}\boldsymbol{X'Y}$

$\boldsymbol{V}(\mathbf{B}) = MSE(\boldsymbol{X'X})^{-1}$
$$MSE = \frac{\boldsymbol{e'e}}{n-(p+1)}$$

$$\hat{\beta}_k - t_{1-\frac{\alpha}{2},n-(p+1)} S.E(\hat{\beta}_k) \le \beta_k \le \hat{\beta}_k + t_{1-\frac{\alpha}{2},n-(p+1)} S.E(\hat{\beta}_k)$$

```
MTB > copy c4 c2 c3 m1
MTB > tran m1 m2
MTB > mult m2 m1 m3
MTB > print m3
```

Data Display

```
 Matrix M3

  27     426     7508
 426    7508   144168
7508  144168  2945984

MTB > inver m3 m4
MTB > copy c1 m5
MTB > mult m2 m5 m6
MTB > print m6
```

Data Display

```
 Matrix M6
```

```
   476
  8307
156154

MTB > mult m4 m6 m7
MTB > print m7
```

Data Display

```
 Matrix M7

-26.3254
  4.8736
 -0.1184
```

$$\mathbf{B_{4\times 1}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} -26.3254 \\ 4.8736 \\ -0.1184 \end{bmatrix}$$

$$\hat{Y} = -26.3254 + 4.8736X - 0.1184X^2$$

```
MTB > tran m5 m8
MTB > mult m8 m5 m9

Answer = 9690.6200

MTB > mult m1 m7 m10
MTB > copy m10 c5
MTB > let c6=c1-c5

MTB > copy c6 m11
MTB > tran m11 m12
MTB > mult m12 m11 m13

Answer = 238.5408

MTB > mult 9.9392 m4 m14
MTB > print m14
```

Data Display

```
Matrix M14

34.5925  -4.47501   0.130833
-4.4750   0.60085  -0.017999
 0.1308  -0.01800   0.000551
```

$$\boldsymbol{V}\begin{bmatrix}\hat{\beta}_0\\ \hat{\beta}_1\\ \hat{\beta}_2\end{bmatrix}=\begin{bmatrix}\boldsymbol{Var}(\widehat{\beta_0}) & \boldsymbol{cov}(\widehat{\beta_0},\widehat{\beta_1}) & \boldsymbol{cov}(\widehat{\beta_0},\widehat{\beta_2})\\ \boldsymbol{cov}(\widehat{\beta_0},\widehat{\beta_1}) & \boldsymbol{Var}(\widehat{\beta_1}) & \boldsymbol{cov}(\widehat{\beta_1},\widehat{\beta_2})\\ \boldsymbol{cov}(\widehat{\beta_0},\widehat{\beta_2}) & \boldsymbol{cov}(\widehat{\beta_2},\widehat{\beta_1}) & \boldsymbol{Var}(\widehat{\beta_2})\end{bmatrix}=\begin{bmatrix}34.5925 & -4.4750 & 0.130833\\ -4.4750 & 0.60085 & -0.017999\\ 0.1308 & -0.017999 & 0.000551\end{bmatrix}$$

$$\hat{\beta}_k - t_{1-\frac{\alpha}{2},n-(p+1)}S.E(\hat{\beta}_k) \le \beta_k \le \hat{\beta}_k + t_{1-\frac{\alpha}{2},n-(p+1)}S.E(\hat{\beta}_k)$$

$t_{1-\frac{\alpha}{2},n-(p+1)} = t_{0.975,27-3} = 2.0639$

$S.E(\hat{\beta}_0) = \sqrt{34.5925} = 5.8815$

$S.E(\hat{\beta}_1) = \sqrt{0.60085} = 0.7746$

$S.E(\hat{\beta}_2) = \sqrt{0.000551} = 0.02347$

$\hat{\beta}_1 - t_{0.975,27-3}S.E(\hat{\beta}_1) \le \beta_1 \le \hat{\beta}_1 + t_{0.975,27-3}S.E(\hat{\beta}_1)$

$3.2747 \le \beta_1 \le 6.4724$

$\hat{\beta}_2 - t_{0.975,27-3}S.E(\hat{\beta}_2) \le \beta_2 \le \hat{\beta}_2 + t_{0.975,27-3}S.E(\hat{\beta}_2)$

$-0.16685 \le \beta_2 \le -0.06995$

1. Hypothesis

$H_0: \beta_k = 0$

$H_1: \beta_k \neq 0$
2. Test statistic
$$T_0 = \frac{b_k - \beta_{k0}}{s(b_k)} = \frac{b_k}{s(b_k)}$$
3. Decision: Reject H_0 if $|T_0| > t_{\left(1-\frac{\alpha}{2}, n-(p+1)\right)}$
p-value=$2P\left(t_{(n-(p+1))} > |T_0|\right)$
Reject H_0 if $\mathrm{p-value} \leq \alpha$

1. Hypothesis
$H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$
2. Test statistic
$$T_0 = \frac{b_1}{s(b_1)} = 6.2913$$
3. Decision: Reject H_0 if $|T_0| > t_{\left(1-\frac{\alpha}{2}, n-(p+1)\right)} = 2.0639$
p-value=$2P\left(t_{(n-(p+1))} > |T_0|\right) = 2P\left(t_{(24)} > |6.2913|\right) = 2(1-1) = 0.00$
reject H_0

1. Hypothesis
$H_0: \beta_2 = 0$

$H_1: \beta_2 \neq 0$
2. Test statistic
$$T_0 = \frac{b_2}{s(b_2)} = -5.044$$

3. Decision: Reject H_0 if $|T_0| > t_{\left(1-\frac{\alpha}{2}, n-(p+1)\right)} = 2.0639$

p-value=$2P\left(t_{(n-(p+1))} > |T_0|\right) = 2P\left(t_{(24)} > |-5.044|\right) = 2(1 - 0.999981) = 0.000038$

reject H_0

Regression Analysis: Y versus X, X^2

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	1046.27	523.133	52.63	0.000
X	1	392.90	392.897	39.53	0.000
X^2	1	252.99	252.985	25.45	0.000
Error	24	238.54	9.939		
Lack-of-Fit	12	79.85	6.654	0.50	0.876
Pure Error	12	158.69	13.224		
Total	26	1284.81			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
3.15265	81.43%	79.89%	77.01%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-26.33	5.88	-4.48	0.000	
X	4.874	0.775	6.29	0.000	47.56
X^2	-0.1184	0.0235	-5.05	0.000	47.56

Regression Equation

Y = -26.33 + 4.874 X - 0.1184 X^2

Project:
Use Data in Brand preference
And Data in Muscle mass

Find all what we learned in Stat 332 by calculate and MINITAB.

Chapter 6 (2)

Q6.11. Refer to **Grocery retailer**. Assume that regression model (6.5) for three predictor variables with independent normal error terms is appropriate.

a. Test whether there is a regression relation, using level of significance *0.05*. State the alternatives, decision rule, and conclusion. What does your test result imply about β_1 β_2, and β_3? What is the P-value of the test?

c. Calculate the coefficient of multiple determination R^2. How is this measure interpreted here?

1- Hypotheses:

$H_0: \beta_1 = \beta_2 = \beta_3 = 0$

$H_1: at\ least\ one\ of\ \beta_k \neq 0.\ k = 1,2,3$

2- Test Statistic:

$$F = \frac{MSR}{MSE}$$

3- Disjoin:

Reject H_0 if $F > F_{p-1,n-p,\alpha}$

or $p - value \leq \alpha$

Source of Variation	d.f	SS	MS	F
Regression	p-1	SSR=$\sum\left(\hat{Y}_i - \bar{Y}\right)^2$	$MSR = \frac{SSR}{p-1}$	$\frac{MSR}{MSE}$
Error	n-p	SSE=$\sum\left(Y_i - \hat{Y}_i\right)^2$	$MSE = \frac{SSE}{n-p}$	
Total	n-1	SSTo= $\sum(Y_i - \bar{Y})^2$		

$$SSTo = Y'Y - \left(\frac{1}{n}\right)Y'JY$$

$$SSE = Y'Y - b'X'Y$$

$$SSR = b'X'Y - \left(\frac{1}{n}\right)Y'JY$$

$$J_{n\times n} = \begin{bmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{bmatrix}$$

```
MTB > copy c5 c2 c3 c4 m1
```
X
```
MTB > tran m1 m2
```
$\boldsymbol{X'}$
```
MTB > mult m2 m1 m3
```
$\boldsymbol{X'X}$
```
MTB > inver m3 m4
```
$(\boldsymbol{X'X})^{-1}$
```
MTB > copy c1 m5
```
Y
```
MTB > mult m2 m5 m6
```
$\boldsymbol{X'Y}$
```
MTB > mult m4 m6 m7
```
$(\boldsymbol{X'X})^{-1}_{4\times4}(\boldsymbol{X'Y})_{4\times1} = \mathbf{B}_{4\times1}$
```
MTB > print m7
```

Data Display

```
 Matrix M7

4149.89
   0.00
 -13.17
 623.55
```

```
MTB > tran m5 m8
```
$\boldsymbol{Y'}$
```
MTB > mult m8 m5 m9
```
$\boldsymbol{Y'Y} = \sum {Y_i}^2$

```
Answer = 993039576.0000
```

```
MTB > copy c5 m10
```
I
```
MTB > tran m10 m11
```
I'
```
MTB > mult m10 m11 m12
```
$II' = J$
```
MTB > mult m8 m12 m13
```
$Y'J$
```
MTB > mult m13 m5 m14
```
$Y'JY$

```
Answer = 51473626884.0000
```

```
MTB > tran m7 m15         b'
MTB > mult m15 m2 m16     b'X'
MTB > mult m16 m5 m17     b'X'Y
```

```
Answer = 992054046.2536
```

$$SSTo = Y'Y - \left(\frac{1}{n}\right)Y'JY = 993039576 - \left(\frac{1}{52}\right)51473626884 = 3162136$$
$$SSE = Y'Y - b'X'Y = 993039576 - 992054046.2536 = 985529.7464$$
$$SSR = SSTo - SSE = 3162136 - 985529.7464 = 2176606.177$$

Source of Variation	d.f	SS	MS	F
Regression	3	2176606.177	725535.4	35.33703
Error	48	985529.7464	20531.87	
Total	51	3162136		

1- Hypotheses:
$H_0: \beta_1 = \beta_2 = \beta_3 = 0$
$H_1: at\ least\ one\ of\ \beta_k \neq 0.\ \ k = 1,2,3$
2- Test Statistic:
$F = 35.337$
3- Disjoin:
Reject H_0 if $F > F_{p-1,n-p,\alpha} = F_{3,48,0.05} = 2.79806$
Then we reject H_0

or $p - value \leq \alpha = 0.05$
$p - value = P\left(F_{p-1,n-p} > 35.337\right) = 1 - 1 = 0.00 < 0.05$

```
MTB > CDF 35.337;
SUBC>   F 3 48.
```

Then we reject H_0

$$R^2 = \frac{SSR}{SSTo} = \frac{2176606.177}{3162136} = 0.68833 = 68.83\%$$

Thus, when the three predictor variables, the number of cases shipped, the indirect cost of the total labor hours as percentage and the qualitative predictor which call holiday, are considered, the variation in the total labor hours is reduced by 68.83 percent.

$Stat \rightarrow Regression \rightarrow Regression \rightarrow Fit\ Regression\ Model$

Regression Analysis: Y versus X1, X2, X3

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	2176606	725535	35.34	0.000
X1	1	95707	95707	4.66	0.036
X2	1	6675	6675	0.33	0.571
X3	1	2034514	2034514	99.09	0.000
Error	48	985530	20532		
Total	51	3162136			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
143.289	68.83%	66.89%	64.78%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	4150	196	21.22	0.000	
X1	0.000787	0.000365	2.16	0.036	1.01
X2	-13.2	23.1	-0.57	0.571	1.02
X3	623.6	62.6	9.95	0.000	1.01

Regression Equation

Y = 4150 + 0.000787 X1 - 13.2 X2 + 623.6 X3

Q6.12. Refer to **Grocery retailer**. Assume that regression model (6.5) for three predictor variables with independent normal error terms is appropriate.

a. Management desires simultaneous interval estimates of the total labor hours for the following typical weekly shipments:

$$X_{h1} = 302000,\ \ X_{h2} = 7.20,\ \ X_{h3} = 0$$

Obtain the family of estimates using a 95 percent confidence coefficient.

$\hat{y}_h \pm t_{1-\frac{\alpha}{2},n-p} S.E(\hat{y}_h)$

$S.E(\hat{y}_h) = \sqrt{Var(\hat{y}_h)}$

$Var(\hat{y}_h) = X_h{}'\boldsymbol{V}(\widehat{\boldsymbol{\beta}})X_h$

```
MTB > copy c6 m24
```
X_h

C7
1
302000
7.20
0

```
MTB > tran m24 m25
```
$X_h{}'$

```
MTB > mult m25 m7 m26
```
$\hat{y}_h$

```
Answer = 4292.7901
```

```
MTB > mult 20531.87 m4 m19
```
$\boldsymbol{V}(\mathbf{B}) = \boldsymbol{MSE}(\boldsymbol{X}'\boldsymbol{X})^{-1}$

```
MTB > mult m25 m19 m27
```
$X_h{}'\boldsymbol{V}(\mathbf{B})$

```
MTB > mult m27 m24 m28
```
$X_h{}'\boldsymbol{V}(\mathbf{B})X_h$

```
Answer = 456.1072
```

$Var(\hat{y}_h) = X_h'\boldsymbol{V}(\widehat{\boldsymbol{\beta}})X_h = 456.1072$

$S.E(\hat{y}_h) = \sqrt{456.1072} = 21.3567$

$$t_{1-\frac{\alpha}{2},n-p} = t_{0.975,48} = 2.01063$$

$4292.7901 \pm 2.01063(21.3567)$

$$4249.8497 < E(Y_h) < 4335.7305$$

$Stat \rightarrow Regression \rightarrow Regression \rightarrow Predict$

Prediction for Y

```
Regression Equation

Y = 4150 + 0.000787 X1 - 13.2 X2 + 623.6 X3


Variable  Setting
X1         302000
X2            7.2
X3              0


    Fit   SE Fit        95% CI               95% PI
4292.79  21.3567  (4249.85, 4335.73)  (4001.50, 4584.08)
```

Q6.14. Refer to **Grocery retailer**. Assume that regression model (6.5) for three predictor variables with independent normal error terms is appropriate. Three new shipments are to be received, each with $X_{h1} = 282000,\ X_{h2} = 7.10$, and $X_{h3} = 0$.

a. Obtain a 95 percent prediction interval for the mean handling time for these shipments.

$\hat{y}_h \pm t_{1-\frac{\alpha}{2}, n-p} S.E(\hat{y}_{new})$

$S.E(\hat{y}_{new}) = \sqrt{Var(\hat{y}_{new})}$

$Var(\hat{y}_{new}) = MSE + X_h{}' \boldsymbol{V}(\hat{\boldsymbol{\beta}}) X_h$

MTB > copy c7 m18 X_h

C7
1
282000
7.10
0

```
MTB > mult 20531.87 m4 m19
```
$\boldsymbol{V}(\mathbf{B}) = \boldsymbol{MSE}(\boldsymbol{X}'\boldsymbol{X})^{-1}$

```
MTB > tran m18 m20
```
$X_h{}'$

```
MTB > mult m20 m19 m21
```
$X_h{}'\boldsymbol{V}(\mathbf{B})$

```
MTB > mult m21 m18 m22
```
$X_h{}'\boldsymbol{V}(\mathbf{B})X_h$

```
Answer = 521.5551
MTB > mult m20 m7 m23
```
$\hat{y}_h$

```
Answer = 4278.3651
```

$$Var(\hat{y}_{new}) = MSE + X_h'\boldsymbol{V}(\hat{\boldsymbol{\beta}})X_h = 20531.87 + 521.5551 = 21053.42482$$

$$S.E(\hat{y}_{new}) = \sqrt{21053.42482} = 145.098$$

$$t_{1-\frac{\alpha}{2},n-p} = t_{0.975,48} = 2.01063$$

$$4278.3651 \pm 2.01063(145.098)$$

$$3986.626748 < Y_{h(new)} < 4570.103452$$

$Stat \rightarrow Regression \rightarrow Regression \rightarrow Predict$

```
Variable  Setting
X1         282000
X2            7.1
X3              0


    Fit   SE Fit         95% CI                 95% PI
4278.37  22.8376  (4232.45, 4324.28)  (3986.63, 4570.10)
```