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 $\beta_0$and β_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.

Q2.4. Refer to **Grade point average** Problem 1.19.
a. Obtain a 99 percent confidence interval for β_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]

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

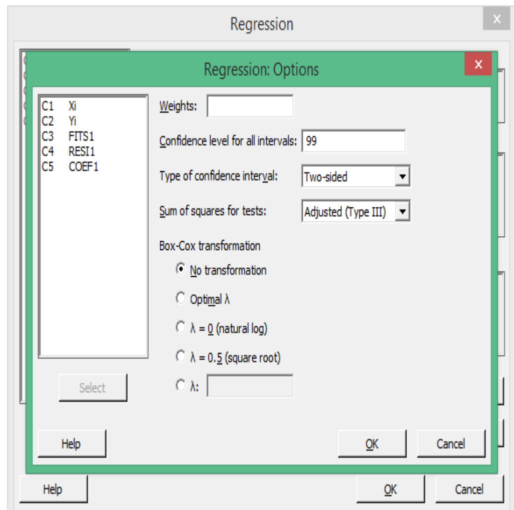
Yi = 2.114 + 0.0388 Xi

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



Current Worksheet: Worksheet 2

Minitab - Grade point Average.MPJ - [Session]

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

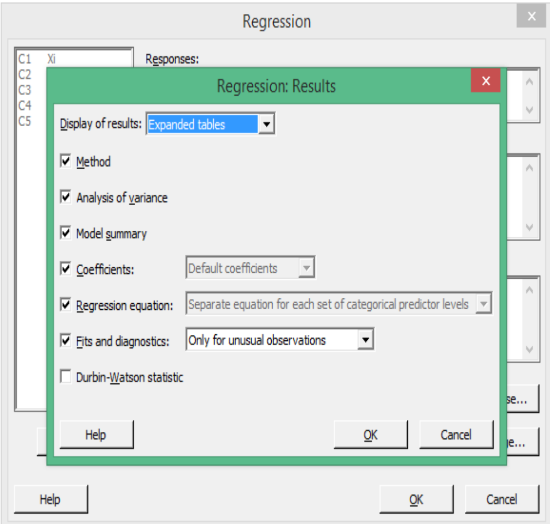
Yi = 2.114 + 0.0388 Xi

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



Current Worksheet: Worksheet 2

Editable

22:14
15/02/2016

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   SECoef        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_0if $|T_0| > t_{\left(1-\frac{\alpha}{2},n-2\right)}$, $3.04 > t_{(0.995,118)} = 1.70943$

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.

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

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

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

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

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

At $X_h = 28$

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

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

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

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

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

$3.22012 \pm 1.9807(0.0706)$

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

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

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 هوعددالناسخ اتالخدمات

Y هوالعددالإجماليللدقائقالتييقضيهاالشخصالخدمة.

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

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

Q2.5. Refer to Copier maintenance 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: $b_1 - t_{\left(1-\frac{\alpha}{2}, n-2\right)} s(b_1) \leq \beta_1 \leq b_1 + t_{(1-\alpha/2, n-2)} s(b_1)$

$\alpha = 1 - 0.9 = 0.1$

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

$$= \frac{22660 - 45 * 5.1111 * 76.2667}{1516 - 45 * 5.1111^2} = 15.035$$

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

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

$$t_{\left(1-\frac{\alpha}{2},n-2\right)} = t_{(0.95,43)} = 1.68107$$

$$b_1 - t_{\left(1-\frac{\alpha}{2},n-2\right)}s(b_1) = 15.035 - 1.68107 * 0.48308 = 14.222$$

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

$$14.222 \le \beta_1 \le 15.847$$

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_0if $|T_0| > t_{\left(1-\frac{\alpha}{2},n-2\right)}$, $31.123 > t_{(0.995,43)} = 2.695$

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_0if $T_0 > t_{(1-\alpha,n-2)}$, $2.143 > t_{(0.95,43)} = 1.861$
Then reject H_0
p-value=$P(t_{(n-2)} > T_0) = \left(1 - P\left(t_{(n-2)} < 2.143\right)\right) = (1 - 0.981) = 0.019 < 0.05$
, then we reject H_0.