

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS  
DEPARTMENT OF MATHEMATICAL SCIENCES  
DHAHRAN, SAUDI ARABIA

**STAT 212: BUSINESS STATISTICS II**

Third Major Exam  
Sunday May 6, 2007  
6:00 pm – 7:15 pm

Please **circle** your instructor's name & section #:

<u>Instructor's name</u>	<u>Section number</u>		
Rahimov I.	1	5	
Anabosi R.	2	3	4
Al Sawi E.		6	

Name:

*Solution Key*

ID#:

Serial:

Question No	Full Points	Points Obtained
1	18	
2	16	
3	26	
Total	60	

Q1. (6+2+2+2+3+3 points)

The general manager of a chain of furniture stores believes that experience is the most important factor in determining the level of success of a salesperson. To examine this belief he records last month's sales (in SR 1,000s) and the years of experience of 10 randomly selected salespeople. These data are listed below.

Salesperson #	1	2	3	4	5	6	7	8	9	10
Years of Experience	0	2	10	3	8	5	12	7	20	15
Sales	7	9	20	15	18	14	20	17	30	25

Given that

$$\sum x = 82$$

$$\sum y = 175$$

$$\sum xy = 1811$$

$$\sum x^2 = 1020$$

$$\sum y^2 = 3489$$

$$MSE = 2.472454 \Rightarrow s_e = 1.5724$$

Answer the following (Do NOT do any rounding to anything):

- a. What is the estimated regression equation for predicting the Sales using the Years of Experience?

$$b_1 = \frac{\sum xy - \frac{(\sum x)(\sum y)}{n}}{\sqrt{\sum x^2 - \frac{(\sum x)^2}{n}}} = \frac{1811 - \frac{(82)(175)}{10}}{\sqrt{1020 - \frac{(82)^2}{10}}} = 1.081703$$

$$b_0 = \bar{y} - b_1 \bar{x} = 17.5 - (1.081703)(8.2) = 8.630035$$

$$\hat{y} = 8.630035 + (1.081703107)X$$

- b. How much sales do you expect a salesperson with 10 years of experience to make?

$$\hat{y} = 8.630035 + (1.081703)10 = 19.44707$$

- c. What is the amount of error in your expectation in part (b) above?

$$e = y - \hat{y} = 20 - 19.44707 = 0.55293$$

d. What is the standard error of the slope of the regression model?

$$s_{b_1} = \frac{s_e}{\sqrt{\sum x^2 - \frac{(\sum x)^2}{n}}} = \sqrt{\frac{MSE}{\sum x^2 - \frac{(\sum x)^2}{n}}} = \sqrt{\frac{2.472454}{347.6}} \quad (1)$$

$$= \boxed{0.084338} \quad (1)$$

e. What is the 90% confidence interval for the estimated sales for a salesperson with 10 years of experience?

A 90% C.I. for  $\mu_{y|x}$  is  $\hat{y} \pm t_{\alpha/2, n-2} s_e \sqrt{1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{\sum(x-\bar{x})^2}}$

$$19.44707 \pm t_{0.05, 8} \sqrt{2.472454 \left(1 + \frac{1}{10} + \frac{(10-8.2)^2}{347.6}\right)} \quad (1)$$

$$19.44707 \pm (1.8595)(1.656123) = 19.44707 \pm 3.07956 \quad (1)$$

$$= (16.36751, 22.52663) \quad (1)$$

f. Do you support the claim that the estimated sales for a salesperson with 10 years of experience is SR 23,000, at the 10% level of significance?

- (1) NO, this claim is rejected because the
- (2) P.I. in (e) before does NOT include the claimed value which is SR 23,000.



$$\Sigma x = 99$$

$$\Sigma y = 373$$

$$\Sigma x^2 = 1275$$

$$\Sigma xy = 4761$$

$$\Sigma y^2 = 17849$$

Q2. (4+8+2+2 points)

A professor of economics wants to study the relationship between income ( $y$  in SR 1,000s) and education ( $x$  in years). A random sample eight individuals is taken and the results are shown below.

Education	16	11	15	8	12	10	13	14
Income	58	40	55	35	43	41	52	49

Answer the following (Do NOT do any rounding to anything):

- a. How much is the Pearson Correlation Coefficient between the Income and the years of Education? And interpret its value.

$$r = \frac{\Sigma xy - \frac{(\Sigma x)(\Sigma y)}{n}}{\sqrt{(\Sigma x^2 - \frac{(\Sigma x)^2}{n}) (\Sigma y^2 - \frac{(\Sigma y)^2}{n})}} = \frac{4761 - \frac{(99)(373)}{8}}{\sqrt{(1275 - \frac{(99)^2}{8})(17849 - \frac{(373)^2}{8})}}$$

$$= 0.96$$

There is a STRONG DIRECT (Positive) relationship between Education & Income

- b. Do you think that the Income and the Education years are significantly linearly related?

① ①  $H_0: \rho = 0$  vs.  $H_A: \rho \neq 0$      $\alpha = 0.05$

② ② Assumptions: ① Bivariate Normal variable ② Quantitative variables

① ③ TS:  $t_0 = r \sqrt{\frac{n-2}{1-r^2}} = (0.96) \sqrt{\frac{6}{0.0777}} = 8.436$

① ④ CV:  $t_{\frac{\alpha}{2}, n-2} = t_{0.025, 6} = 2.4469$

① ⑤ DR: If  $|t_0| > t_{\frac{\alpha}{2}, n-2} \Rightarrow$  Reject  $H_0$

① ⑥ since  $8.436 > 2.4469 \Rightarrow$  Reject  $H_0$ .

① ⑦ Yes, the two variables are sig. linearly related.

- c. How much is the Coefficient of Determination?

$$R^2 = r^2 = (0.96)^2 = 0.9216 \approx 92.16\%$$

- d. What can you say about the adequacy, or the goodness, of a regression model if it is to be fitted to predict the income using the number of years of education?

The model is ADEQUATE.

**Q3.** (6+6+6+4+2+2 points)

An actuary wanted to develop a model to predict how long individuals will live. After consulting a number of physicians, she collected the age at death ( $Y$ ), the average number of hours of exercise per week ( $X_1$ ), the cholesterol level ( $X_2$ ), and the number of points that the individual's blood pressure exceeded the recommended value ( $X_3$ ). A random sample of 40 individuals was selected. The Minitab output of the regression model is shown below.

The regression equation is

$$Y = 55.8 + 1.79 X_1 - 0.021 X_2 - 0.016 X_3$$

Predictor	Coef	SE Coef	T
Constant	55.8	11.8	4.729
X1	1.79	0.44	???
X2	-0.021	0.011	-1.909
X3	-0.016	0.014	-1.143

S = ???      R-Sq = ???

Analysis of Variance

Source	DF	SS	MS	F
Regression	3	936	312	3.4774
Residual Error	36	3230	89.722	
Total	39	4166		

Answer the following (Do NOT do any rounding to anything):

- Complete the ANOVA table.
- Is there enough evidence at the 5% significance level to infer that the model is useful in predicting length of life?

① ①  $H_0: \beta_1 = \beta_2 = \beta_3 = 0$  vs.  $H_A: \text{at least one } \beta_i \neq 0.$

- ① ② Assumptions:
- ⊖ Assumed relation is linear between  $X$ 's &  $Y$ .
  - ⊖ Residuals are Indep.
  - = have constant variance
  - = are normally distributed.

① ③  $F_0 = \boxed{3.4774}.$

① ④  $F_{\alpha, k, n-k-1} = F_{0.05, 3, 36} = \boxed{2.839}$

- ① ⑤ If  $F_0 > F_\alpha \Rightarrow \text{Reject } H_0.$
- ① ⑥ Since  $3.4774 > 2.839 \Rightarrow \text{Reject } H_0.$
- ① ⑦ The model is USEFUL.



c. Is there enough evidence at the 1% significance level to infer that the average number of hours of exercise per week and the age at death are linearly related?

① ①  $H_0: \beta_1 = 0$  vs.  $H_A: \beta_1 \neq 0$ .

① ② Assumptions: see part (b) before

① ③ TS:  $t_0 = \frac{1.79}{0.44} = \boxed{4.475}$

① ④ CV:  $t_{\frac{\alpha}{2}, n-k-1} = t_{\frac{0.01}{2}, 36} = \boxed{2.7045}$

① ⑤ DR: If  $|t_0| > t_{\frac{\alpha}{2}, n-k-1} \Rightarrow$  Reject  $H_0$

① ⑥ Since  $4.475 > 2.7045 \Rightarrow$  Reject  $H_0$ .

① ⑦ Yes, Age at death & average # of hours/week are sig. linearly related.

d. How much is the coefficient of determination? What does this statistic tell you?

②  $R^2 = \frac{SSR}{SST} \times 100 = \frac{936}{4166} \times 100 = \boxed{22.46\%}$

② 22.46% of the variation in  $Y$  is explained by the common variation of the  $X$ 's.

e. Interpret the coefficient  $b_2$ .

① If the cholesterol level DECREASES by 1 unit the

① the average age at death INCREASES by 0.021 years given that all the other variables are fixed (kept constant).

f. Calculate  $S_e$  (The standard deviation of the regression model)

$S_e = \sqrt{MSE} = \sqrt{89.722} = \boxed{9.4722}$