# Math 644, Fall 2012

# Solution for Homework 2

## Problem 1.

(a) the 99% CI (confidence interval) is

$$b_1 \pm t(1-0.01/2, 118) * s(b_1) = 0.03883 \pm 2.62 * 0.01277 = [0.0053726, 0.0722874]$$

It does not include 0 (or we have 99% confidence to believe it is not 0). If it is not 0, a student's PCA can be predicted by his ACT. In other word, ACT can be an indicator for universities to evaluate a students and to decide whether a student should be admitted.

(b)

$$H_0 : \beta_1 = 0 \quad H_1 : \beta_1 \neq 0$$

From the calculation

$$|t| = |0.03883/0.01277| = |3.040| > t(0.995, 118) = 2.62$$

Thus, we reject $H_0$, i.e. there is significant linear association between ACT and GPA when $\alpha = 0.01$

(c) Because

$$p - value = 0.00292 < 0.01$$

we also reject the $H_0$ above

(a) [3.061384, 3.341033] on average, with 95% confidence, the mean freshman GPA is between 3.061384 and 3.341033 when their ACT test scores are 28

(b) [1.959355, 4.443063], with 95% confidence her GPA will be between 1.959355 and 4.443063

(c) Yes

## Problem 2.

(a) Set up the ANOVA table

Response: y

| source | Df | SS | MS | F-value | p-value |
|---|---|---|---|---|---|
| regression(x) | 1 | 3.588 | 3.588 | 9.2402 | 0.002917 |
| Residuals | 118 | 45.818 | 0.388 | | |
| Total | 119 | 49.406 | | | |

(b) conduct an F-test for $H_0 : \beta_1 = 0$ with $\alpha = 0.01$

Since p-value is smaller than $\alpha = 0.01$, we reject $H_0$, i.e. $\beta_1$ is significantly different from 0.

(c) what is the absolute magnitude of the reduction in the variation of $Y$ when $X$ is introduced into the model? what is the relative reduction? what is the name for the later measure?

the absolute magnitude of the reduction in the variation of $Y$ when $X$ is 3.588

the relative reduction 3.588/SST = 3.588/(3.588+45.818) = 7.262276%

called $R^2$

(d) obtain $r_{XY}$ and attach the appropriate sign

$$r_{XY} = +\sqrt{R^2} = 0.2694818$$

(e) which measure $R^2$ or $r$ has more clear-cut operational interpretation, explain.

$r$, because it give clear relationship between $x$ and $y$.

## Problem 3.

The absolute value of the coefficient $-1.4$ looks different from 0, but its p-value is very big, indicating that the coefficient is not significantly different from 0. Thus, the correlation between $X$ and $Y$ is not strong and the conclusion is not statistically solid.

## Problem 4.

the $\alpha$ level used by the analyst was greater than 0.033, If the $\alpha$ level had been 0.01, he shod accept $H_0$

## Problem 5.

(a)

$$\begin{array}{ccccc} \hat{Y} & = & 10.2 & + & 4.00X \\ (SE) & & (0.6633) & & (0.4690) \end{array}$$

$$MSE = 2.199289, \quad R^2 = 0.9009, \quad F = 72.73$$

Yes, the linear regression function fits the data well.

(b)

$$\hat{Y} = 10.2 + 4.00 * 1 = 14.2$$

(c)

$$b_1 \pm t(0.975, 8) * s(b_1) = 4 \pm 2.306 * 0.469 = [2.918486, 5.081514]$$

The interval does not include 0, indicating that we have (big) confidence that $\beta_1$ is different from 0, i.e. $\beta_1$ is not zero, and thus the linear association is significant.

(d) We need to test

$$H_0 : \beta_1 = 0 \qquad v.s. H_1 : \beta_1 \neq 0$$

Note that

$$|t| = |\frac{b_1 - 0}{s(b_1)}| = 8.528 > t(1 - \alpha/2, n - 2) = 2.306$$

We reject $H_0$. In other words, there is significant linear association.

(e)

$$b_0 \pm t(0.975, 8) * s(b_0) = 10.2 \pm 2.306 * 0.6633 = [8.67043, 11.72957]$$

The interval does not include 0, indicating we have (big) confidence that $\beta_0$ is different from 0. Even if no shipment, there are still broken ampules (due to the other reasons)

## Problem 6.

(a) set up the ANOVA table. which elements are additive?

Response: y

| source | Df | SS | MS | F-value | p-value |
|---|---|---|---|---|---|
| regression(x) | 1 | 160.0 | 160.0 | 72.727 | 2.749e-05 |
| Residuals | 8 | 17.6 | 2.2 | | |
| Total | 9 | 177.6 | | | |

(b)

$$H_0 : \beta_1 = 0$$

$$F^* = 72.727 > F(1 - 0.05, 1, 8) = 5.32$$

reject $H_0$, there is significant linear association between $X$ and $Y$

(c)

$$H_0 : \beta_1 = 0$$

$$|t^*| = 8.528 > t(1 - 0.05/2, 8) = 2.262$$

reject $H_0$, there is significant linear association between $X$ and $Y$

(d)

$$R^2 = 90.09\%, \qquad r = 0.9491575$$

90.09% of the variation in $Y$ is accounted for by introducing $X$ into the model