LINEAR REGRESSION MODELS W4315

HOMEWORK 2 ANSWERS

October 4, 2009

Instructor: Frank Wood (10:35-11:50)

1. (25 points) Problem 2.4 in the textbook on page 90

Answer:

(a)

Following the code from problem 3 in homework 1, below is the R code for this problem:
$y.hat < -alpha.hat + beta.hat * x$
$RSS < -sum((y - y.hat)^2)$
$MSE < -RSS/(120 - 2)$
$var.b1 < -MSE/SXX$
$sd.b1 < -sqrt(var.b1)$
From the code, we have standard error of β_1 is 0.0128.
Since t(.995,118)=2.618137, we have then the β_1's .99 confidence interval is $.0.0388 + / -$ $2.618137 * .0128$=(0.005287846,0.07231215).
It doesn't include zero. The reason that the director cares about the coverage of zero of CI is that he wants to be very much sure of if there is a positive relation between ACT score and GPA score.
(b)
Use the formula (2.20) to calculate the corresponding t-value. All the components are already known from the code above, so after plugging in the values, we have $t^* = \frac{b_1 - 0}{sd(b_1)} = 3.04$, and this value is greater than 2.618, so we reject the null hypothesis.
(c)
The P-value is $2P(t(118) > 3.04) = .003$, which is smaller than .01, so we reject the null hypothesis. It's in sync with the result concluded from (b).

2. (25 points) Do problem 2.51 in the book.

Answer:

From (2.21), we have the explicit formula of b_0, so plugging in every term's formula we have the followings:

$$
\begin{aligned}
Eb_0 &= E(\bar{Y}) - Eb_1 * \bar{X} \\
&= \frac{1}{n}\sum_{i=1}^{n} EY_i - \frac{\bar{X}}{SXX} * E(SXY)
\end{aligned}
$$

where $EY_i = \beta_0 + \beta_1 * X_i$, and

$$
\begin{aligned}
E(SXY) &= E(\sum_{i=1}^{n}(X_i - \bar{X})Y_i) \\
&= \sum_{i=1}^{n}(X_i - \bar{X})(\beta_0 + \beta_1 * X_i)
\end{aligned}
$$

Then we have:

$$
\begin{aligned}
Eb_0 &= \frac{1}{n}\sum_{i=1}^{n} EY_i - \frac{\bar{X}}{SXX} * E(SXY) \\
&= \frac{1}{n}\sum_{i=1}^{n} \beta_0 + \beta_1 * X_i - \sum_{i=1}^{n}(X_i - \bar{X})(\beta_0 + \beta_1 * X_i) \\
&= \beta_0 + \sum \frac{1}{n} - \frac{\bar{X}(X_i - \bar{X})}{SXX}(\beta_0 - \beta_1 X_i) \\
&= \beta_0 + \frac{nSXX - n\sum(X_i - \bar{X})X_i}{nSXX} \\
&= \beta_0
\end{aligned}
$$

Thus, we proved that b_0 is an unbiased estimator of β_0.

3. (50 points) Problem 2.52 in the textbook on page 97

Answer:

(2.31) tells us that $\bar{Y}$ is independent of b_1. N.B. if 2 random variables X and Y and independent, then $Var(X+Y) = Var(X) + Var(Y)$.
So given the above result, we have the followings:

$$
\begin{aligned}
Var(b_0) &= Var(\bar{Y} - b_1\bar{X}) \\
&= Var(\bar{Y}) + Var(b_1) * \bar{X}^2
\end{aligned}
$$

Since:

$$\begin{aligned} Var(\bar{Y}) &= Var(\frac{1}{n}\sum Y_i) \\ &= \frac{1}{n^2}\sum Var(Y_i) \\ &= \frac{\sigma^2}{n} \end{aligned}$$

and $Var(b_1) = \frac{\sigma^2}{\sum(X_i - \bar{X})^2}$ by "Variance" on Page 43 of the textbook, so we have:

$$Var(b_0) = \sigma^2(\tfrac{1}{n} + \tfrac{\bar{X}^2}{\sum(X_i - \bar{X})^2})$$

The above equation is a special case of (2.29b) in the sense that in (2.29b), if X_h equals 0 or $2\bar{X}$ then it becomes (2.22b).