LINEAR REGRESSION MODELS W4315

HOMEWORK 1 ANSWERS

October 4, 2009

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

1. (25 points) Let $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$ be a linear regression model with distribution of error terms unspecified (but with mean $E(\epsilon) = 0$ and variance $V(\epsilon_i) = \sigma^2$ (σ^2 finite) known). Show that $s^2 = MSE = \frac{\sum(Y_i - \hat{Y}_i)^2}{n-2}$ is an unbiased estimator for σ^2. $\hat{Y}_i = b_0 + b_1 X_i$ where $b_0 = \bar{Y} - b_1 \bar{X}$ and $b_1 = \frac{\sum_i((X_i - \bar{X})(Y_i - \bar{Y}))}{\sum_i(X_i - \bar{X})}$

Answer:

First, let's denote the followings:

$\hat{e}_i = y_i - \hat{y}_i$

$SXX = \sum_{i=1}^{n}(x_i - \bar{x})^2$

$SYY = \sum_{i=1}^{n}(y_i - \bar{y})^2$

$SXY = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$

Now we set out to prove the following equation which accomplishes essentially the final result:

$Var\hat{e}_i = E\hat{e}_i{}^2 = (\frac{n-2}{n} + \frac{1}{SXX}(\frac{1}{n}\sum_{j=1}^{n} x_j^2 + x_i^2 - 2(x_i - \bar{x})^2 - 2x_i\bar{x}))\sigma^2$

To prove the above, realize that:

$$\begin{aligned}
Var(\hat{e}_i) &= Var(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) \\
&= Var((y_i - \beta_0 - \beta_1 x_i) - (\hat{\beta}_0 - \beta_0) - x_i(\hat{\beta}_1 - \beta_1)) \\
&= Var(y_i) + Var(\hat{\beta}_0) + x_i^2 Var(\hat{\beta}_1) - 2Cov(y_i, \hat{\beta}_0) - 2x_i Cov(y_i, \hat{\beta}_1) + 2x_i Cov(\hat{\beta}_0, \hat{\beta}_1)
\end{aligned}$$

The last equation holds because the covariance between any random variable and a constant is zero, and all the y_i's are independent entailing that the $Cov(y_i, y_j) = 0, i \neq j$

$Var(y_i) = \sigma^2$

Notice that(some algebras needed here, and the following tricks are crucial in reducing the amount of calculation):
$\sum (x_i - \bar{x}) = 0$
$\beta_1 = \frac{\sum x_i - \bar{x} y_i}{SXX}$
So now we have:

$$
\begin{aligned}
Var(\beta_1) &= Var(\frac{SXY}{SXX}) \\
&= Var(\frac{\sum (x_i - \bar{x}) y_i}{SXX}) \\
&= \frac{1}{SXX^2} \sum x_i - \bar{x}^2 Var(y_i) \\
&= \frac{\sigma^2}{SXX}
\end{aligned}
$$

And:

$$
\begin{aligned}
Var(\beta_0) &= Var(\bar{y} - \hat{\beta}_1 \bar{x}) \\
&= Var(\sum (\frac{1}{n} - \frac{(x_i - \bar{x})\bar{x}}{SXX}) y_i) \\
&= \sum (\frac{1}{n} - \frac{x_i - \bar{x}}{SXX} \bar{x})^2 \sigma^2 \\
&= \sum [\frac{1}{n^2} + \frac{SXX * \bar{x}^2}{SXX^2} - \frac{2}{n} \frac{\bar{x}(x_i - \bar{x})}{XSS}] \sigma^2 \\
&= [\frac{1}{n} + \frac{n\bar{x}^2}{SXX}] \sigma^2 \\
&= \frac{\sum {x_i}^2}{n * SXX} \sigma^2
\end{aligned}
$$

For the other terms in the decomposition of $Var(\hat{e_i})$, we have:

$$
\begin{aligned}
Cov(y_i, \hat{\beta}_1) &= Cov(y_i, \frac{sumx_i - \bar{x} y_i}{SXX}) \\
&= \frac{x_i - \bar{x}}{SXX} Var(y_i) \\
&= \frac{x_i - \bar{x}}{SXX} \sigma^2
\end{aligned}
$$

and:

$$
\begin{aligned}
Cov(y_i, \hat{\beta}_0) &= Cov(y_i, \bar{y} - \hat{\beta}_1 \bar{x}) \\
&= Cov(y_i, \frac{\sum y_i}{n} - \sum (x_i - \bar{x}) y_i SXX \bar{x}) \\
&= \frac{\sigma^2}{n} + \bar{x}\frac{x_i - \bar{x}}{SXX}\sigma^2
\end{aligned}
$$

At last, we have:

$$
\begin{aligned}
Cov(\hat{\beta}_0, \hat{\beta}_1) &= Cov(\bar{y} - \hat{\beta}_1 \bar{x}, \hat{\beta}_1) \\
&= Cov(\frac{\sum y_i}{n} - \sum \frac{(x_i - \bar{x})\bar{x}}{SXX} y_i, \sum \frac{(x_i - \bar{x}) y_i}{SXX}) \\
&= \sum_{i=i}^{n} (\frac{1}{n} - \frac{x_i - \bar{x}}{SXX}\bar{x}) \frac{x_i - \bar{x}}{SXX}\sigma^2 \\
&= -\frac{\bar{x}}{SXX}\sigma^2
\end{aligned}
$$

Then plug in all the parts back to the decomposition of $Var(\hat{e_i})$, we have:

$Var(\hat{e_i}) = (\frac{n-1}{n} + \frac{1}{SXX}(\frac{1}{n}\sum_{j=1}^{n} {x_j}^2 + x_i^2 - 2(x_i - \bar{x})^2 - 2x_i\bar{x}))\sigma^2$

Thus,

$$
\begin{aligned}
E\hat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n} E{\hat{e_i}}^2 \\
&= \frac{1}{n}\sum_{i=1}^{n} [\frac{n-2}{n} + \frac{1}{SXX}(\frac{1}{n}\sum_{j=1}^{n} x_j^2 + x_i^2 - 2(x_i - \bar{x})^2 - 2x_i\bar{x})]\sigma^2 \\
&= [\frac{n-2}{n} + \frac{1}{nS_{xx}}\{\sum_{j=1}^{n} x_j^2 + \sum_{i=1}^{n} x_i^2 - 2SXX - 2\frac{1}{n}(\sum_{i=1}^{n} x_i)^2\}]\sigma^2 \\
&= (\frac{n-2}{n} + 0)\sigma^2 \\
&= \frac{n-2}{n}\sigma^2
\end{aligned}
$$

where the third equation holds because: $\sum x_i \bar{x} = \frac{1}{n}(\sum x_i)^2$

and the second to last equation holds since $\sum x_i^2 - \frac{1}{n}(\sum x_i)^2 = SXX$

From the above equation, the result flows.

2. (25 points) Derive the maximimum likelihood estimators $\hat{\beta}_0, \hat{\beta}_1$, and $\hat{\sigma}^2$ for parameters β_0, β_1, and σ^2 for the normal linear regression model (i.e. $\epsilon_i \sim_{iid} N(0, \sigma^2)$).

Answer:

To figure the MLE of the parameters, we need to first write down the likelihood function of the data, so under normal assumption, we have the log-likelihood function as follows:

$logL(\beta_0, \beta_1, \sigma^2|x, y) = -\frac{n}{2}log(2\pi) - \frac{n}{2}log\sigma^2 - \frac{\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}$.

For any fixed value of σ^2, $logL$ is maximized as a function of β_0 and β_1, that minimize

$$\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2 \tag{1}$$

But to minimize this function is just to principle behind LSE, so it's apparent that the MLE of β_0 and β_1 are the same as their LSE's. Now, substituting in the log-likelihood, to find the MLE of σ^2 we need to maximize

$$-\frac{n}{2}log(2\pi) - \frac{n}{2}log\sigma^2 - \frac{\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}{2\sigma^2}$$

This maximization problem is nothing but MLE of σ^2 in ordinary normal sampling problems, which is easily given as

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

If you are not familiar with the MLE in normal sampling setting, you can take derivative with respect to σ^2 (N.B. not σ), and then set the derivative to be zero. The solution of the equation is just the MLE of σ^2.

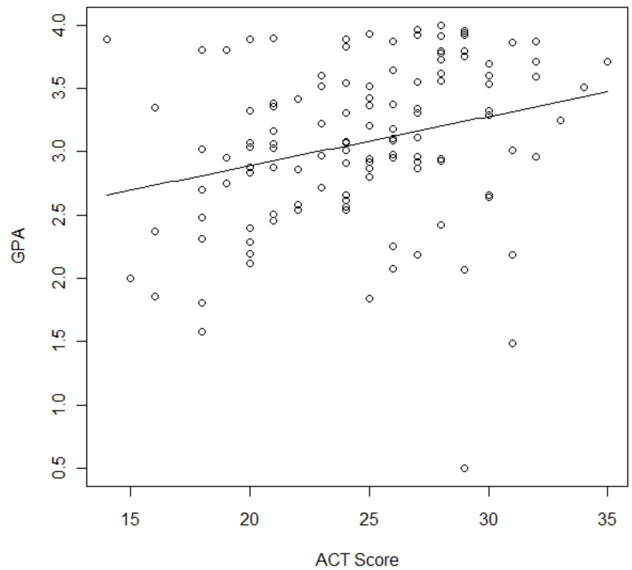
3. (50 points) Do problem 1.19 in the book.

Answer:

(a)

$data <- read.table("f:/TA1.txt")$

$attach(data)$



$x <- data[,2]$
$y <- data[,1]$
$SXX <- sum((x - mean(x))^2)$
$SYY <- sum((y - mean(y))^2)$
$SXY <- sum((x - mean(x)) * (y - mean(y)))$
$beta.hat <- SXY/SXX$
$alpha.hat <- mean(y) - beta.hat * mean(x)$
We get the result the the LSE of the intercept and the slope are 2.11 and .038.
The corresponding regression line is thus

$$Y = .038 + 2.11X \tag{2}$$

(b) From the graph, we can see that the regression passes through the center of the major part of the data, but does not capture all the features of the data.
(c)
Plug $x = 30$ into (2), and the result is the point estimation of mean GPA with ACT score being 30.

$$\hat{y} = \hat{\alpha} + \hat{\beta} * x = 2.11 + .038 * 30 = 3.28$$

(d)

It is nothing but the slope of the estimated regression line, since the slope can be interpreted as the average GPA will increase by .038 when the ACT score is enhanced by one point.