Stat 312: Lecture 19
Inference on Slope

Moo K. Chung
mchung@stat.wisc.edu

April 3, 2003

Concepts

1. Given a model
$$Y_j = \beta_0 + \beta_1 x_j + \epsilon_j,$$
we wish to test if $H_0 : \beta_1 = 0$. Assume that $\epsilon \sim N(0, \sigma^2)$. The least squares estimator of β_1 is $\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \sum_{j=1}^n c_j Y_j$, where $c_j = (x_j - \bar{x})/S_{xx}$.

2. From $\sum_{j=1}^n c_j = 0$, $\sum_{j=1}^n c_j x_j = 1$ and $\sum_{j=1}^n c_j^2 = S_{xx}^{-1}$ we can show that $\hat{\beta}_1 \sim N(\beta_1, \sigma^2/S_{xx})$.

3. Inference on the slope parameter β_1 is based on
$$T = \frac{\hat{\beta}_1 - \beta_1}{S_{\hat{\beta}_1}} \sim t_{n-2},$$
where $S_{\hat{\beta}_1} = \hat{\sigma}/\sqrt{S_{xx}} = \sqrt{\frac{SSE}{(n-2)S_{xx}}}$. From Lecture 18 formula, $SSE = S_{yy} - S_{xy}^2/S_{xx}$, we get $S_{\hat{\beta}_1} = \frac{1}{\sqrt{n-2}}\sqrt{\frac{S_{yy}}{S_{xx}} - (\frac{S_{xy}}{S_{xx}})^2}$. There is a reason $S_{\hat{\beta}_1}$ is written in this way (see example 1).

   We reject H_0 if $|t| > t_{\alpha/2,n-2}$ at $100(1-\alpha)$ significance.

In-class problems

Example 1. 10 students took tow midterm exams.

```
> x<-c(80, 75, 60, 90, 99, 60, 55,
85, 65, 70)
> y<-c(70, 60, 70, 72, 95, 66, 60,
80, 70, 60)
> summary(lm(y~x))
Call: lm(formula = y ~ x)
 Residuals:
    Min      1Q  Median   3Q     Max
-10.908  -6.312   1.758 4.354 10.836
 Coefficients:
 Estimate Std. Error t value Pr(>|t|)
(Intercept) 29.48  13.23  2.23  0.056
 x           0.55   0.18  3.14  0.014
```

The P-value here is based on two-sided test. Let's see if we are getting the same P-value using Concept 3. Under H_0, $\beta_1 = 0$. So $t = \hat{\beta}_1/S_{\hat{\beta}_1}$. $\hat{\beta}_1 = 0.55$. In R, we first compute $\frac{S_{yy}}{S_{xx}} - (\frac{S_{xy}}{S_{xx}})^2$:

```
>cov(y,y)/cov(x,x)-
(cov(x,y)/cov(x,x))^2
[1] 0.2476849
> Sbeta<-sqrt(0.2477)/sqrt(8)
> Sbeta
[1] 0.1759616
> 0.5523/Sbeta
[1] 3.138752
> 2*pt(-3.138,8)
[1] 0.01384698
```

P-value$= P(|T| > 3.138) = 0.014$.

Example 2. There was a mistake in the solution to example 3, lecture 18. Let S_x^2 be the sample variance of x_j. Since $S_{xx} = \sum_{j=1}^n (x_j - \bar{x})^2$, $S_{xx} = (n-1)S_x^2$, $S_{yy} = (n-1)S_y^2$. So $S_{xy}/(n-1)$ is the sample covariance between x_j and y_j. In R, `cov` and `var` are sample covariance and sample variance respectively. The correct way to estimate $\sigma^2 = \mathbf{Var} Y_j$ is to use the following R computation.

```
> 9*(cov(y,y)-beta1*cov(x,y))/8
[1] 58.48568
```

Alternately, we can use

```
> 9*(cov(y,y)-cov(x,y)^2/cov(x,y))/8
[1] 58.4815
```

Due to the rounding-off error, you are getting the different values for the third decimal place.

Self-study problems

Example 12.10., 12.11.