

Stat 324: Lecture 20

Inference on Regression Parameters

Moo K. Chung
mchung@stat.wisc.edu

April 19, 2005

1. *Coefficient of determination* measures the proportion of the total variations in y that can be explained by the linear model and it is defined as

$$r^2 = 1 - \frac{SSE}{SST}$$

Since $SST = S_{yy}$ and $SSE = S_{yy} - S_{xy}^2/S_{xx}$, we can see that $r^2 = S_{xy}^2/(S_{xx}S_{yy})$. r is called the *sample correlation coefficient*.

2. *Example*. Hooke's law states that the length change y in spring is proportional to applied force x . The following measurements are available.

x (kg)	29.4	39.2	49.0	58.8	68.6	78.4
y (mm)	4.25	5.25	6.50	7.85	8.75	10.10

We fit the model $Y = \beta_0 + \beta_1 x + \epsilon$.

```
>x<-c(29.4, 39.2, 49.0, 58.8, 68.6, 78.4)
>y<-c(4.25, 5.25, 6.50, 7.85, 8.75, 10.10)
>summary(lm(y~x))
```

```
Call: lm(formula = y ~ x)
```

Residuals:

```
      1      2      3      4      5      6
0.06905 -0.10524 -0.02952  0.14619 -0.12810  0.04762
```

Coefficients:

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.658095    0.163995   4.013   0.0160 *
x            0.119825    0.002906  41.238 2.07e-06 ***
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.1191 on 4 degrees of freedom Multiple
R-Squared:  0.9977,    Adjusted R-squared:  0.9971 F-statistic: 1701
on 1 and 4 DF,  p-value: 2.067e-06
```

3. The least squares estimation for β_1 are given by $\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$ where $S_{xy} = n(\bar{x}\bar{y} - \bar{x}\bar{y})$. Since $\sum_{j=1}^n (x_j - \bar{x})\bar{Y} = 0$, $\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \sum_{j=1}^n c_j Y_j$, where $c_j = (x_j - \bar{x})/S_{xx}$. From $\sum_{j=1}^n c_j = 0$, $\sum_{j=1}^n c_j x_j = 1$, we can show that $\mathbb{E}\hat{\beta}_1 = \beta_1$ showing unbiasedness. From $\sum_{j=1}^n c_j^2 = S_{xx}^{-1}$, $\mathbb{V}\hat{\beta}_1 = \sigma^2/S_{xx}$. Since $\hat{\beta}_1$ is a linear combination of normal distributions, $\hat{\beta}_1 \sim N(\beta_1, \sigma^2/S_{xx})$. The test statistic

$$T = \frac{\hat{\beta}_1 - \beta_1}{S_{\hat{\beta}_1}} \sim t_{n-2},$$

where $S_{\hat{\beta}_1} = \hat{\sigma}/\sqrt{S_{xx}}$ and $\hat{\sigma} = \sqrt{SSE/(n-2)}$. It can be shown that $SSE = S_{yy} - S_{xy}^2/S_{xx}$ and $S_{\hat{\beta}_1} = \frac{1}{\sqrt{n-2}} \sqrt{\frac{S_{yy}}{S_{xx}} - \left(\frac{S_{xy}}{S_{xx}}\right)^2}$. In a slightly different way, $SSE = \sum_{j=1}^n y_j^2 - \hat{\beta}_0 \sum_{j=1}^n y_j - \hat{\beta}_1 \sum_{j=1}^n x_j y_j$.

4. Based on the test statistic, a $100(1 - \alpha)\%$ CI for β_1 is $\hat{\beta}_1 \pm t_{\alpha/2, n-2} S_{\hat{\beta}_1}$. For hypothesis testing

$$H_0 : \beta_1 = 0 \text{ vs. } H_1 : \beta_1 \neq 0$$

we reject H_0 if $|t| > t_{\alpha/2, n-2}$ at $100(1 - \alpha)\%$ significance. A hypothesis testing on β_0 and a CI construction are similar. $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x}$ is distributed as normal with mean β_0 and variance $\sigma^2 \bar{x}^2 / S_{xx}$. Testing $H_0 : \beta_0 = 0$ is based on $T = \frac{\hat{\beta}_0 - \beta_0}{S_{\hat{\beta}_0}} \sim t_{n-2}$, where $S_{\hat{\beta}_0} = \sigma^2 \bar{x}^2 / S_{xx}$.

Assigned problems. Exercise 12.30., 12. 36. HW 6 due April 28th. Solve the assigned problems in Lecture 19-22.