Chapter 11

- Sums of squares:

Total sum of squares

$$SST = \sum_{i=1}^{n} (Y_i - \bar{Y})^2,$$

regression sum of squares

$$SSR = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2,$$

error sum of squares

$$SSE = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2.$$

Properties: a) SSR and SSE are independent, b) $SST = SSR + SSE$.

- Coefficient of determination (a measure of quality of fit)

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}.$$

$0 \leq R^2 \leq 1$. The greater R^2 the better fit.

- ANOVA approach. Testing

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

Test statistic

$$F = \frac{SSR}{SSE/(n-2)}.$$

Under H_0, F has F-distribution with 1 and $n-2$ degrees of freedom. Test: H_0 is rejected if $F \geq f_{\alpha,1,n-2}$.

- Prediction. Observe $Y_1, Y_2, ..., Y_n$. Predict $Y_0 = \beta_0 + \beta_1 x_0 + \epsilon_0$.

$(1-\alpha)$ confidence interval for EY_0:

$$\left[\hat{Y}_0 - t_{\alpha/2,n-2} S \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}, \hat{Y}_0 + t_{\alpha/2,n-2} S \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}} \right]$$

where $\hat{Y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$.

- $(1-\alpha)$ prediction interval for Y_0:

$$\left[\hat{Y}_0 - t_{\alpha/2,n-2} S \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}, \hat{Y}_0 + t_{\alpha/2,n-2} S \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}} \right].$$