Chapter 11

• Unbiased estimator of σ^2 is

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

S^2 is independent on $\bar{Y}$, $\hat{\beta}_0$ and $\hat{\beta}_1$. Distribution of $(n-2)S^2/\sigma^2$ is χ^2-distribution with $n-2$ degrees of freedom.

• Inference on the slope and intercept.

$(1-\alpha)$ confidence interval for β_1:

$$\left[\hat{\beta}_1 - t_{\alpha/2,n-2}\frac{S}{\sqrt{\sum(x_i-\bar{x})^2}}, \hat{\beta}_1 + t_{\alpha/2,n-2}\frac{S}{\sqrt{\sum(x_i-\bar{x})^2}}\right].$$

$(1-\alpha)$ confidence interval for β_0:

$$\left[\hat{\beta}_0 - t_{\alpha/2,n-2}\frac{S\sqrt{\sum x_i^2}}{\sqrt{n\sum(x_i-\bar{x})^2}}, \hat{\beta}_0 + t_{\alpha/2,n-2}\frac{S\sqrt{\sum x_i^2}}{\sqrt{n\sum(x_i-\bar{x})^2}}\right].$$

Testing $H_0 : \beta_1 = \beta_{10}$. Alternatives a) $H_1 : \beta_1 > \beta_{10}$, b) $H_1 : \beta_1 < \beta_{10}$, c) $H_1 : \beta_1 \neq \beta_{10}$. Significance level α.

Test statistic

$$T = \frac{\hat{\beta}_1 - \beta_{10}}{S/\sqrt{S_{xx}}}$$

where $S_{xx} = \sum(x_i - \bar{x})^2$. Under H_0, T has t-distribution with $n-2$ degrees of freedom. Critical region: a) $T \geq t_{\alpha,n-2}$, b) $T \leq -t_{\alpha,n-2}$, c) $|T| \geq t_{\alpha/2,n-2}$.

Testing $H_0 : \beta_0 = \beta_{00}$. Alternatives a) $H_1 : \beta_0 > \beta_{00}$, b) $H_1 : \beta_0 < \beta_{00}$, c) $H_1 : \beta_0 \neq \beta_{00}$. Significance level α.

Test statistic

$$T = \frac{\hat{\beta}_0 - \beta_{00}}{S\sqrt{\sum x_i^2/nS_{xx}}}$$

where $S_{xx} = \sum(x_i - \bar{x})^2$. Under H_0, T has t-distribution with $n-2$ degrees of freedom. Critical region: a) $T \geq t_{\alpha,n-2}$, b) $T \leq -t_{\alpha,n-2}$, c) $|T| \geq t_{\alpha/2,n-2}$.

• 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)

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

$0 \le R^2 \le 1$. 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 $F \ge f_{\alpha,1,n-2}$.
• 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 $\hat{Y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$.