**Lecture 1 overview**

Introduction: Discussion of syllabus, course website, online questions. We start in the text in Chapter 5.

**Review of simple linear regression**

In the text's notation the equation is:

$$Y_i = A + BX_i + E_i \ .$$

The least-squares estimates minimize the quantity:

$$S(A, B) = \sum E_i^2 = \sum (Y_i - A - BX_i)^2,$$

and are given by:

$$A = \overline{Y} - B\overline{X}, \ B = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sum (X_i - \overline{X})^2} = \frac{S_{XY}}{S_X^2}.$$

The sum of squares decomposition is

$$\sum (Y_i - \overline{Y})^2 = \sum (\overline{Y} - \widehat{Y}_i)^2 + \sum (Y_i - \widehat{Y}_i)^2 \text{ or TSS} = \text{RegSS} + \text{RSS},$$

which leads to:

$$r^2 = \frac{\text{RegSS}}{\text{TSS}},$$

and the correlation coefficient $r$ is the square root of $r^2$ multiplied by the sign of $B$ . Also $r = S_{XY}/(S_X S_Y)$. The residual standard error is

$$S_E = \sqrt{\frac{\sum E_i^2}{n-2}}.$$

**Review of multiple regression**

For two explanatory variables, note Figures 5.6 and 5.7 which give a geometric perspective and illustrate the partial slope interpretation of the coefficients. For several explanatory variables, the equation is:

$$Y_i = A + B_1 X_{i1} + B_2 X_{i2} + ... + B_k X_{ik} + E_i = \widehat{Y}_i + E_i \ .$$

For this equation the least-squares estimates minimize the quantity:

$$S(A, B_1, B_2, ..., B_k) = \sum [Y_i - (A + B_1 X_{i1} + B_2 X_{i2} + ... + B_k X_{ik})]^2.$$

Taking derivatives of $S(A, B_1, B_2, ..., B_k)$ yields the set of normal equations shown on page 90 of the text, which are solved to yield the coefficients $A, B_1, B_2, ..., B_k$. We again have the decomposition of TSS = RegSS + RSS and

$$R^2 = \frac{\text{RegSS}}{\text{TSS}},$$

is the squared multiple correlation (also called the coefficient of determination) which is the percent of variation that is explained by the regression. The residual standard error is

$$S_E = \sqrt{\frac{\sum E_i^2}{n - k - 1}}.$$

The text also discusses the use of standardized regression coefficients and when it is appropriate to use them.