Theoretical exercises 1: Regression

7. (Weighted Least Squares) Suppose that in the model $y_i = \beta_0 + \beta_1 x_i + e_i$, the errors have mean zero and are independent, but $\mathrm{Var}(e_i) = \rho_i^2\sigma^2$, where the ρ_i are known constants, so the errors do not have equal variance. This situation arises when the y_i are averages of several observations at x_i; in this case, if y_i is an average of n_i independent observations, $\rho_i^2 = 1/n_i$ (why?). Because the variances are not equal, the theory developed in this chapter does not apply; intuitively, it seems that the observations with large variability should influence the estimates of β_0 and β_1 less than the observations with small variability.

The problem may be transformed as follows:

$$\rho_i^{-1} y_i = \rho_i^{-1}\beta_0 + \rho_i^{-1}\beta_1 x_i + \rho_i^{-1} e_i$$

or

$$z_i = u_i\beta_0 + v_i\beta_1 + \delta_i$$

where

$$u_i = \rho_i^{-1} \qquad v_i = \rho_i^{-1} x_i \qquad \delta_i = \rho_i^{-1} e_i$$

a. Show that the new model satisfies the assumptions of the standard statistical model.

b. Find the least squares estimates of β_0 and β_1.

c. Show that performing a least squares analysis on the new model, as was done in part (b), is equivalent to minimizing

$$\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 \rho_i^{-2}$$

This is a weighted least squares criterion; the observations with large variances are weighted less.

d. Find the variances of the estimates of part (b).

15. Find the least squares estimate of β for fitting the line $y = \beta x$ to points (x_i, y_i), where $i = 1, \ldots, n$.

16. Consider fitting the curve $y = \beta_0 x + \beta_1 x^2$ to points (x_i, y_i), where $i = 1, \ldots, n$.

a. Use the matrix formalism to find expressions for the least squares estimates of β_0 and β_1.

b. Find an expression for the covariance matrix of the estimates.

24. Suppose that the independent variables in a least squares problem are replaced by rescaled variables $u_{ij} = k_j x_{ij}$ (for example, centimeters are converted to meters.) Show that $\hat{Y}$ does not change. Does $\hat{\beta}$ change? (*Hint:* Express the new design matrix in terms of the old one.)

25. Suppose that each setting x_i of the independent variables in a simple least squares problem is duplicated, yielding two independent observations Y_{i_1}, Y_{i_2}. Is it true that the least squares estimates of the intercept and slope can be found by doing a regression of the mean responses of each pair of duplicates, $\overline{Y}_i = (Y_{i_1} + Y_{i_2})/2$ on the x_i? Why or why not?