Mods Statistics, Sheet 4, HT 2012

1. Suppose $x_1, \ldots, x_n$ are known constants and that $Y_1, \ldots, Y_n$ satisfy the 'regression through the origin' model $Y_i = \beta x_i + \varepsilon_i$, where the ε_i are independent $N(0, \sigma^2)$ random variables. Show that the maximum likelihood estimator of β is $\widehat{\beta} = \sum_i x_i Y_i / \sum_i x_i^2$.

What is the distribution of $\widehat{\beta}$?

The following data give the distance, in miles, by road (y) and in a straight line (x) for several different journeys. Why might we prefer to consider the model above to the model $Y_i = \alpha + \beta x_i + \varepsilon_i$?

y	10.7	11.7	6.5	25.6	29.4	$\cdots$	25.7	40.5	26.5	14.2	33.1
x	9.5	9.8	5.0	19.0	23.0	$\cdots$	21.7	28.2	18.0	12.1	28.0

If the straight-line distance between two locations is 12 miles, how would you use the model to predict the expected distance by road?

2. In the model $Y_i = \alpha + \beta x_i + \varepsilon_i$, $i = 1, \ldots, n$, where $E(\varepsilon_i) = 0$, show that the least squares estimator of β is

$$\widehat{\beta} = \frac{n \sum x_i Y_i - (\sum x_i)(\sum Y_i)}{n \sum x_i^2 - (\sum x_i)^2}.$$

Show that $\widehat{\beta}$ is unbiased for β. Under what additional assumptions is $\widehat{\beta}$ the maximum likelihood estimator of β?

3. Suppose $Y_1, \ldots, Y_n$ satisfy $Y_i = \alpha + \beta(x_i - \overline{x}) + \varepsilon_i$, where the ε_i are independent $N(0, \sigma^2)$ and the constants x_i are not all equal.

Find the maximum likelihood estimators $\widehat{\alpha}$ and $\widehat{\beta}$. Show that $\widehat{\alpha}$ and $\widehat{\beta}$ are unbiased, for α and β respectively, and find their variances.

Assuming σ^2 is known, show how the distribution of $\widehat{\beta}$ can be used to construct a $100(1 - \gamma)\%$ confidence interval for β.

4. Suppose that in the model $Y_i = \alpha + \beta x_i + \varepsilon_i$, the errors ε_i are independent and normally distributed with mean 0, but that $\mathrm{var}(\varepsilon_i) = \rho_i^2 \sigma^2$ where the ρ_i are known constants.

Show that the maximum likelihood estimates of α and β can be found by minimizing

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

and find these estimates of α and β.

Can you think of a situation in which this model might arise?