SMF PROBLEMS 4. 30.10.2017

Q1. Show that the regression (= least-squares) line for the data $(x_1, y_1), \ldots, (x_n, y_n)$ is

$$y - \overline{y} = r_{xy} \frac{s_y}{s_x} (x - \overline{x}),$$

with $r = r_{xy}$ the sample correlation coefficient, s_x, s_y the sample standard deviations.

Q2. With data y and two predictor variables (regressors) u and v, show that the regression (= least-squares) plane is $y - \overline{y} = a(u - \overline{u}) + b(v - \overline{v})$, where a, b satisfy

$$as_{uu} + bs_{uv} = s_{yu},$$
$$as_{uv} + bs_{vv} = s_{yv}.$$

Q3. Grain crops (wheat, barley etc.) are harvested in the summer. Yields are much affected by the weather at harvest time, which is not predictable much in advance. The two main predictors that are known months in advance are amounts of spring rainfall and spring sunshine, during the growing season. Discuss the use of this in the markets for grain options and futures.

Q4. In the regression model

$$y = A\beta + \epsilon$$

(data y an n-vector, the design matrix A an $n \times p$ matrix of constants, β a p-vector of parameters, ϵ an n-vector of errors with independent $N(0, \sigma^2)$ components), show that the maximum-likelihood estimators, and also the least-squares estimators, are

$$\hat{\beta} = (A^T A)^{-1} A^T y.$$

Show also that (in the notation of lectures)

$$Py = A\hat{\beta}.$$

NHB