Another Least Squares Problem

(a) Show for any $n \geq 2$, $x_1, \ldots, x_n$, and $y_1, \ldots, y_n$, that

$$\sum (y_i - x_i)^2/n \quad \geq \quad (\sigma_y - \sigma_x)^2 + (\bar{y} - \bar{x})^2,$$

where $\bar{x} = \sum x_i/n$, $\sigma_x = \sqrt{\sum (x_i - \bar{x})^2/n}$, $\bar{y} = \sum y_i/n$, and $\sigma_y = \sqrt{\sum (y_i - \bar{y})^2/n}$.

(b) Show that (x_i, y_i) are colinear if and only if

$$\sum (y_i - x_i)^2/n \quad = \quad (\sigma_y - \sigma_x)^2 + (\bar{y} - \bar{x})^2.$$

(c) Express y_i as a function of x_i, $\bar{x}$, σ_x, $\bar{y}$, and σ_y when (x_i, y_i) are colinear.

Jeffrey A. Barnett (`jbb@notatt.com`) and Pablo Gonzalez
Albuquerque, NM 87106
USA

Solutions

Given the integer $n \geq 2$ and the real numbers $x_1, \ldots, x_n$, v_y, and $\bar{y}$,

Find the $y_1^*, \ldots, y_n^*$ that minimizes $G = \sum(y_i - x_i)^2/n$ subject to the constraints that $\sum y_i/n = \bar{y}$ and $\sum(y_i - \bar{y})^2/n = v_y$.

Define $\bar{x} = \sum x_i/n$ and $v_x = \sum(x_i - \bar{x})^2/n$.

Lagrangean multipliers are used below to show that the minimizing y_i are

$$y_i^* = (x_i - \bar{x})\sqrt{\frac{v_y}{v_x}} + \bar{y}. \tag{1}$$

Two definitions make this result slightly more interesting:

Define $\sigma_x = \sqrt{v_x}$ and $\sigma_y = \sqrt{v_y}$.

Now simple algebra shows that

$$y_i^* = (x_i - \bar{x})\frac{\sigma_y}{\sigma_x} + \bar{y} \tag{2}$$

and that

$$\min G = (\sigma_y - \sigma_x)^2 + (\bar{y} - \bar{x})^2. \tag{3}$$

Answers for parts a, b, and c

(a) Given $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ and, hence, $\bar{x}$, σ_x, $\bar{y}$, and σ_y, Equ (3) establishes the minimum G.

(b) Equ (2) shows that the minimizing y_i are *unique*, given the x_i, and that achieving equality entails that the (x_i, y_i) are colinear. It remains to show that if the (x_i, y_i) are colinear then equality 3 follows. Let $y_i = ax_i + b$, then $\bar{y} = a\bar{x} + b$ and $\sigma_y = a\sigma_x$. Therefore,

$$\begin{aligned} y_i &= ax_i + b \\ &= a(x_i - \bar{x}) + a\bar{x} + b \\ &= (x_i - \bar{x})\frac{\sigma_y}{\sigma_x} + \bar{y} \end{aligned}$$

and Equ (3) will follow as before.

(c) Equ (2) or $(y_i - \bar{y})\sigma_x = (x_i - \bar{x})\sigma_y$ are such simple relations.

Proof: The use of Lagrangean multipliers is straightforward but tedious. As usual, an equation is manufactured consisting of the actual objective function plus weighted zero-valued expressions formed from the problem constraints. The weights, λ_1 and λ_2 here, provide an extra parameter for each constraint. Thus, the two constraints will be expressed in terms of two parameters so we can solve for them. The manufactured function for minimization is

$$\sum (y_i - x_i)^2 + \lambda_1 \left(\sum y_i - n\bar{y} \right) + \lambda_2 \left(\sum (y_i - \bar{y})^2 - nv_y \right).$$

Now take the derivative with respect to y_i and set it equal to 0. Note, we actually obtain n equations in this process.

$$2(y_i - x_i) + \lambda_1 + 2\lambda_2(y_i - \bar{y}) = 0$$

Therefore,

$$y_i = \frac{2\lambda_2\bar{y} + 2x_i - \lambda_1}{2(1 + \lambda_2)}. \tag{4}$$

Next, substitute the above into the constraint $\sum y_i/n = \bar{y}$ and simplify to show that

$$\lambda_1 = 2(\bar{x} - \bar{y}).$$

Substitute this value for λ_1 into Equ (4) to show that

$$y_i = \frac{x_i - \bar{x}}{1 + \lambda_2} + \bar{y}. \tag{5}$$

Now substitute this value for y_i into the constraint $\sum (y_i - \bar{y})^2/n = v_y$ and simplify to show that

$$\lambda_2 = -1 + \sqrt{\frac{v_x}{v_y}}.$$

Finally, substitute this value for λ_2 in Equ (5) and it follows that the minimizing values of y_i are as shown by Equ (1) above. ▮