9.8 Exercises: Some Solutions (November 15, 2011)

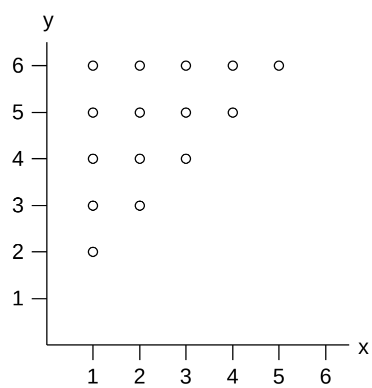

Figure 9.4 (See Exercise 9.1.) A two-dimensional discrete distribution, each of the 15 pairs appearing with the same probability.

9.1. Consider the Figure 9.4. Find the best predictor for y on the basis of x. What is $\operatorname{cor}(x, y)$? (Do your calculations without pen and paper.)

• Solution to Ex. 9.1:

The conditional means of y lie in the line $y = 0.5x + 3.5$ and hence the best predictor for y on the basis of x is simply $\mathrm{BP}(y; x) = 0.5x + 3.5$, which is trivially also the best *linear* predictor

$$\mathrm{BLP}(y; x) = \varrho_{xy} \frac{\sigma_y}{\sigma_x} x + (\mu_y - 0.5\mu_x)\,.$$

Because $\sigma_x = \sigma_y$, we have $\varrho_{xy} = 0.5$. □

9.2. Prove that $\mathrm{E}(y \mid \mathbf{x})$ is the best predictor for y on the basis of $\mathbf{x}$, i.e., $\mathrm{E}(y \mid \mathbf{x}) = \mathrm{BP}(y; \mathbf{x})$.

Christensen (2002, p. 132), Rao (1973a, p. 264), Searle, Casella & McCulloch (1992, §7.2).

9.3. Confirm that $\mathrm{E}[y - (\mathbf{a}'\mathbf{x} + b)]^2 = \mathrm{var}(y - \mathbf{a}'\mathbf{x}) + (\mu_y - \mathbf{a}'\boldsymbol{\mu}_{\mathbf{x}} - b)^2$.

• SOLUTION TO EX. 9.1:

$$\begin{aligned}
\mathrm{E}[y - (\mathbf{a}'\mathbf{x} + b)]^2 &= \mathrm{E}[y - \mathbf{a}'\mathbf{x} - (\mu_y - \mathbf{a}'\boldsymbol{\mu}_{\mathbf{x}}) + (\mu_y - \mathbf{a}'\boldsymbol{\mu}_{\mathbf{x}} - b)]^2 \\
&= \mathrm{E}[y - \mathbf{a}'\mathbf{x} - (\mu_y - \mathbf{a}'\boldsymbol{\mu}_{\mathbf{x}})]^2 + \mathrm{E}[\mu_y - \mathbf{a}'\boldsymbol{\mu}_{\mathbf{x}} - b]^2 \\
&\quad + 2\,\mathrm{E}[y - \mathbf{a}'\mathbf{x} - (\mu_y - \mathbf{a}'\boldsymbol{\mu}_{\mathbf{x}})][\mu_y - \mathbf{a}'\boldsymbol{\mu}_{\mathbf{x}} - b] \\
&= \mathrm{var}(y - \mathbf{a}'\mathbf{x}) + (\mu_y - \mathbf{a}'\boldsymbol{\mu}_{\mathbf{x}} - b)^2 \\
&= \mathrm{var}(y - \mathbf{a}'\mathbf{x}) + [\mathrm{E}(y - \mathbf{a}'\mathbf{x}) - b]^2 \\
&= \text{“variance”} + \text{“bias}^2\text{”}.
\end{aligned}$$

□