COVARIANCE AND REGRESSION

1. Let x and y be random variables, and imagine that the conditional expectation of y given x is determined by the linear function $E(y|x) = \alpha + \beta x$. Find expressions for α and β that are in terms of the moments of the joint distribution of x and y.

In a skiing competition, a competitor's overall time is found by adding his time in the downhill section to his time in the slalom section. Downhill times have an expected value of 1 minute and 15 seconds with a standard deviation of 5 seconds. Slalom times have an expected value of 1 minute and 45 seconds with a standard deviation of 6 seconds. Overall times have a standard deviation of 11 seconds. What is the expected overall time of a competitor who has recorded 1 minute and 10 seconds in the downhill section?

Answer. The downhill time is x, the slalom time is y. We have

$$\begin{aligned}
&E(x) = 75, \quad V(x) = 25,\\
&E(y) = 105, \quad V(y) = 36, \quad \text{and}\\
&V(x+y) = V(x) + V(y) + 2C(x,y) = 121.
\end{aligned}$$

Hence

$$C(x,y) = \frac{121 - 36 - 25}{2} = 30, \quad \text{and} \quad \beta = \frac{C(x,y)}{V(x)} = \frac{30}{25}.$$

If $x = 70$, then

$$\begin{aligned}
E(y|x) &= E(y) + \beta\{x - E(x)\}\\
&= 105 + \frac{30}{25}\{70 - 75\} = 99.
\end{aligned}$$

Also

$$E(x + y|x) = x + E(y|x) = 70 + 99 \quad \text{i.e.} \quad \text{2mins. 49secs.}$$

2. Let x and y be jointly distributed random variables such that $E(y|x) = \alpha + \beta x$. Prove that $\beta = C(x,y)/V(x)$ and that $\alpha = E(y) - \beta E(x)$

The average height of each generation of adult males in Britain is 69 ins with a standard deviation of 5 ins. The correlation coefficient for the heights of fathers and the heights of their sons is 0.7. Given that I am 75 ins tall, what is the expected height of my male offspring, and what is the expected height of my male grandchildren.

3. Let x, y be linearly related random variables such that $E(y|x) = \alpha + \beta x$. Show that $\beta = C(x, y)/V(x)$ and that $\alpha = E(y) - \beta E(x)$

Let the expected yields in pounds of three investments be $E(x, y, z) =$ (200, 150, 350) and let

$$\begin{bmatrix} V(x) & C(x,y) & C(x,z) \\ * & V(y) & C(y,z) \\ * & * & V(z) \end{bmatrix} = \begin{bmatrix} 14 & -2 & 6 \\ * & 11 & -9 \\ * & * & 9 \end{bmatrix}$$

What is the variance of the total earnings? Assuming that $E(y|x) = \alpha + \beta x$ and that $E(z|x) = \gamma + \delta x$, what are the expected earnings given that $x = 120$ has already been received?

4. Let x and y be jointly distributed random variables with conditional expectations which can be written as $E(y|x) = \alpha + \beta x$ and $E(x|y) = \gamma + \delta y$. Express β and δ in terms of the moments of the joint distributions. Use the fact that $C^2(x, y) \leq V(x)V(y)$ to show that $\beta < 1/\delta$.

A man runs on Hampstead Heath twice a week. The average duration of one of his outings is 40 minutes with a standard deviation of 5 minutes. His average running time per week is 1 hour 20 minutes with a standard deviation of 4 minutes. Given that he ran for only 15 minutes on Monday, what is the expected duration of his Friday outing?

Answer. The duration of the Monday outing is x, the duration of the Friday outing is is y. We have

$$E(x) = E(y) = 40, \quad V(x) = V(y) = 25, \quad \text{and}$$

$$V(x + y) = V(x) + V(y) + 2C(x, y) = 16.$$

Hence

$$C(x, y) = \frac{16 - 25 - 25}{2} = -17, \quad \text{and} \quad \beta = \frac{C(x, y)}{V(x)} = -\frac{17}{25}.$$

If $x = 15$, then

$$\begin{aligned} E(y|x) &= E(y) + \beta\big\{x - E(x)\big\} \\ &= 40 - \frac{17}{25}\{15 - 40\} = 57. \end{aligned}$$

5. Let x and y be jointly distributed random variables such that $E(y|x) = \alpha + \beta x$. Show that $\beta = C(x, y)/V(x)$ and that $\alpha = E(y) - \beta E(x)$.

The centigrade temperatures recorded at 20 minute intervals in an air-conditioned room constitute a sequence of random variables. The expected value of the readings is 19° with a standard deviation of 2°. The correlation between successive temperature readings is 0.9. If a temperature of 20° is recorded at one reading, what is the expected value at the next reading?

Answer. We have

$$Corr\{y(t), y(t-1)\} = \frac{C\{y(t), y(t-1)\}}{V\{y(t)\}} = \beta = 0.9.$$

Therefore

$$\begin{aligned} E\{y(t+1)|y(t)\} &= \alpha + \beta y(t) \\ &= (1-\beta) + \beta\Big[y(t) - E\{y(t)\}\Big] \\ &= 19 + 0.9\big[20 - 19\big] = 19.9. \end{aligned}$$

6. Let x, y be linearly related random variables such that $E(y|x) = \alpha + \beta x$. Show that $\beta = C(x,y)/V(x)$ and that $\alpha = E(y) - \beta E(x)$

Let the expected yields in pounds of three investments be $E(x, y, z) =$ (200, 150, 350) and let

$$\begin{bmatrix} V(x) & C(x,y) & C(x,z) \\ C(y,x) & V(y) & C(y,z) \\ C(z,x) & C(z,y) & V(z) \end{bmatrix} = \begin{bmatrix} 14 & -2 & 6 \\ -2 & 11 & -9 \\ 6 & -9 & 9 \end{bmatrix}$$

What is the variance of the total earnings? Assuming that $E(y|x) = \alpha+\beta x$ and that $E(z|x) = \gamma + \delta x$, what are the expected earnings given that $x = 120$ has already been received?

7. The expected rainfalls in September, October and November are 10 ins, 8 ins 6 ins respectively, with a variance–covariance matrix of

$$\begin{bmatrix} 6 & -3 & 1.5 \\ -3 & 6 & -3 \\ 1.5 & -3 & 6 \end{bmatrix}.$$

Calculate the expected rainfall throughout these three months and find its variance.

If the September rain was unusually high, in what direction would ones estimates of rainfall be revised

(a) in the two months following, and

(b) for the three month period?

Answer. The Expected rainfall is

$$E(y) = E\left(\sum x_i\right) = \sum E(x_i)$$
$$= 10 + 8 + 6 = 24.$$

Its variance is

$$\begin{aligned} V(x_1 + x_2 + x_3) &= V(x_1) + V(x_2) + V(x_3) \\ &\qquad + 2\big\{C(x_1, x_2) + C(x_1, x_3) + C(x_2, x_3)\big\} \\ &= (6 + 6 + 6) + 2\{1.5 - 3 - 3\} = 9. \end{aligned}$$

Let x, y, z be the rainfall in September, October and November, and let us write $x = E(x) + \Delta x$. Then we have

$$E(x + y + z|z) = x + E(y|x) + E(z|x),$$

with

$$\begin{aligned} E(y|x) &= E(y) + \beta\{x - E(x)\}; \quad \beta = C(x, y)/V(x); \\ E(z|x) &= E(z) + \delta\{x - E(x)\}; \quad \delta = C(x, z)/V(x), \end{aligned}$$

which may be combined to give

$$\begin{aligned} E(x + y + z|z) &= x + E(y) + E(z) + \{\beta + \delta\}\{x - E(x)\} \\ &= E(x) + E(y) + E(z) + \{1 + \beta + \delta\}\Delta x. \end{aligned}$$

Given that

$$\beta = \frac{C(x, y)}{V(x)} = -\frac{3}{6} = -\frac{1}{2} \quad \text{and} \quad \delta = \frac{C(x, z)}{V(z)} = \frac{1.5}{6} = \frac{1}{4},$$

It follows that

$$E(x + y + z|z) = E(x) + E(y) + E(z) + \left\{1 - \frac{1}{2} + \frac{1}{4}\right\}\Delta x.$$

8. An investor has a choice of three financial assets. The expected yields of these assets are given in the vector $[\,0.04 \quad 0.03 \quad 0.05\,]$ and the variances and covariances of the yields are given in the matrix

$$10^{-4}\begin{bmatrix} 3.29 & -0.83 & 0 \\ -0.83 & 3.41 & 0 \\ 0 & 0 & 2.01 \end{bmatrix}.$$

Derive expressions for the expected yield and variance of a portfolio containing λQ of the first asset and $(1-\lambda)Q$ of the second asset, and ascertain whether there is any value of λ such that the variance of this portfolio is less than the variance of an investment Q in the third asset.

Answer. The investor has two objectives (a) to obtain high returns and (b) to minimise risk. A trade-off must be established between the two objectives. However, some portfolios dominate others in the sense that they fulfil one of the objectives better that the alternative portfolio whilst fulfilling the other objective at least as well. The question we must answer is whether or not there exists a mixed portfolio containing the first and the second assets which dominates a portfolio containing only the third asset.

First we find the expected yield and variance of a generic portfolio containing λQ of the first asset and $(1-\lambda)Q$ of the second asset. Let x_1, x_2 and x_3 be the yields of the three assets. We find that

$$\begin{aligned}E\{\lambda Qx_1 + (1-\lambda)Qx_2\} &= Q\{\lambda E(x_1) + (1-\lambda)E(x_2)\}\\ &= Q\Big\{E(x_2) + \lambda\big[E(x_1) - E(x_2)\big]\Big\}\\ &= Q\{0.03 + \lambda 0.01\}.\end{aligned}$$

Also

$$\begin{aligned}&V\{\lambda Qx_1 + (1-\lambda)Qx_2\}\\ &\quad= Q^2\Big\{\lambda^2 V(x_1) + (1-\lambda)^2 V(x_2) + 2\lambda(1-\lambda)C(x_1,x_2)\Big\}\\ &\quad= Q^2\Big[\lambda^2\big\{V(x_1) + V(x_2) - 2C(x_1,x_2)\big\}\\ &\qquad\qquad + \lambda\big\{2C(x_1,x_2) - 2V(x_2)\big\} + V(x_2)\Big]\\ &\quad= Q^2\left[\lambda^2\left\{\begin{matrix}3.29\\ +3.41\\ +1.66\end{matrix}\right\} - \lambda\left\{\begin{matrix}1.66\\ +6.82\end{matrix}\right\} + 3.41\right]\\ &\quad= Q^2(8.36\lambda^2 - 8.48\lambda + 3.41).\end{aligned}$$

Now we ask whether there is a value $\lambda \in (0,1)$ such that this variance is less that $V(Qx_3) = Q^2 2.01$. The answer can be found by investigating the roots of

$$V\big\{\lambda x_1 + (1-\lambda)x_2\big\} - V(x_3) = 8.36\lambda^2 - 8.48\lambda + 1.40 = 0.$$

The roots $\lambda_1 = 0.2075$ and $\lambda_2 = 0.8009$ are the points at which the alternative portfolios have the same variance. The optimum value of λ which minimises the variance of the mixed portfolio, and which can be obtained by ordinary calculus, lies somewhere in the interval $[\lambda_1, \lambda_2]$