MATH102–Statistics

Solutions to the Exercises (W7)

Tuesday 5/9/06

1. $\text{Cov}(X, Y) = E(X - \mu_x)(Y - \mu_y) = E(XY) - E(X)E(Y)$

   and

   $E(X) = \mu_x$

   with

   $E(Y) = \mu_y = E(a + bX) = a + b\mu_x$

   which gives

   $E(XY) = EX(a + bX) = a\mu_x + bE(X^2)$

   Therefore

   $$\text{Cov}(X, Y) = a\mu_x + bEX^2 - \mu_x(a + b\mu_x) = bEX^2 - b(EX)^2 = bV(X)$$

2. $$\rho_{xy} = \frac{\text{Cov}(X, Y)}{\sigma_x \sigma_y} = \frac{bV(X)}{\sqrt{V(X)b^2V(X)}} = 1$$

   as expected since we have a perfect linear relation between X and Y.

3. $\text{Cov}(X, Y) = -bV(X)$ and $\rho_{xy} = -1$ since $Y = a - bX$, and the solution to 1. and 2. is for general b.

4. In general,

   $V(X_1 + X_2) = V(X_1) + V(X_2) + 2Cov(X_1, X_2)$

   (i) If $X_1 \propto X_2$ then $Cov(X_1, X_2)$ is positive, and so the independence form for the variance of the sum $V(X_1) + V(X_2)$ will under–estimate the true value.

   (ii) If $X_1 \propto -X_2$ then $Cov(X_1, X_2)$ is negative, and so the sum of the individual variances will over–estimate the true variance.

5. In general,

$V(X_1 - X_2) = V(X_1) + V(X_2) - 2Cov(X_1, X_2)$

(i) If $X_1 \propto X_2$ then $Cov(X_1, X_2)$ is positive, and so the independence form for the variance of the difference $V(X_1)+V(X_2)$ will over–estimate the true value.

(ii) If $X_1 \propto -X_2$ then $Cov(X_1, X_2)$ is negative, and so the sum of the individual variances will under–estimate the true variance.

6.

X	Y	$(X-\bar{X})$	$(Y-\bar{Y})$	$(X-\bar{X})(Y-\bar{Y})$	$(X-\bar{X})^2$	$(Y-\bar{Y})^2$
1	3	-2	-3	6	4	9
2	3	-1	-3	3	1	9
3	6	0	0	0	0	0
4	7	1	1	1	1	1
5	11	2	5	10	4	25
				20	10	44

$\widehat{\mathrm{Cov}}(X, Y) = 20/4 = 5$ and

$$\hat{\rho} = \frac{5}{\sqrt{10/4 \times 44/4}} = 0.95346$$

The covariance is positive, as expected since Y is effectively $2X$, and the correlation is close to 1 for the same reason.

```
> x <- c(1,2,3,4,5)
> y <- c(3,3,6,7,11)
> cov(x,y)
[1] 5
> cor(x,y)
[1] 0.9534626
```