Mathematical Statistics 1

Chapter 2. Multivariate Distributions

2.5. The Correlation Coefficient—Proofs of Theorems

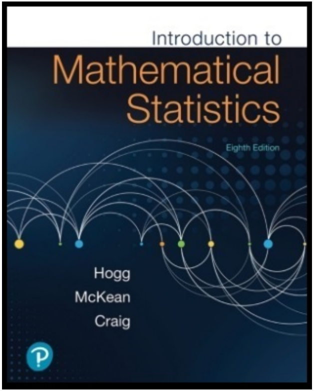

Table of contents

Theorem 2.5.1

Theorem 2.5.1. For all jointly distributed random variables (X, Y) whose correlation coefficient ρ exists (so that $\sigma_1 > 0$ and $\sigma_2 > 0$ by the definition of ρ), we have $-1 \leq \rho \leq 1$.

Proof. With $E[X] = \mu_1$ and $E[Y] = \mu_2$, consider the polynomial in v given by

$$h(v) = E[((X - \mu_1) + v(Y - \mu_2))^2].$$

Then $h(v) \geq 0$ for all v. Hence the discriminant of $h(v)$ is less than or equal to 0 (for if the discriminant is positive then h has two roots and h is negative between those two roots).

Theorem 2.5.1

Theorem 2.5.1. For all jointly distributed random variables (X, Y) whose correlation coefficient ρ exists (so that $\sigma_1 > 0$ and $\sigma_2 > 0$ by the definition of ρ), we have $-1 \leq \rho \leq 1$.

Proof. With $E[X] = \mu_1$ and $E[Y] = \mu_2$, consider the polynomial in v given by

$$h(v) = E[((X - \mu_1) + v(Y - \mu_2))^2].$$

Then $h(v) \geq 0$ for all v. Hence the discriminant of $h(v)$ is less than or equal to 0 (for if the discriminant is positive then h has two roots and h is negative between those two roots).

Theorem 2.5.1 (continued)

Proof (continued). Since E is linear (by Theorem 1.8.2) we have

$$\begin{aligned} h(v) &= E[((X-\mu_1)+v(Y-\mu_2))^2] \\ &= E[(X-\mu_1)^2+2v(X-\mu_1)(Y-\mu_2)+v^2(Y-\mu_2)^2] \\ &= E[(x-\mu_1)^2]+2vE[(X-\mu_1)(Y-\mu_2)]+v^2E[(Y-\mu_2)^2] \\ &= \sigma_1^2+2v\text{cov}(X,Y)+v^2\sigma_2^2 \text{ by Definition 2.5.1} \\ &= \sigma_1^2+2v\rho\sigma_1\sigma_2+v^2\sigma_2^2 \text{ by Definition 2.5.2.} \end{aligned}$$

So the discriminant is nonpositive then $4(\rho^2-1)\sigma_1^2\sigma_2^2 \le 0$ or $\rho^2-1\le 0$ or $|\rho|\le 1$, as claimed. □

Theorem 2.5.2

Theorem 2.5.2. If X and Y are independent random variables then $\text{cov}(X, Y) = 0$ and hence $\rho = 0$.

Proof. If X and Y are independent then by Theorem 2.4.2 we have $E[XY] = E[X]E[Y]$. So by Note 2.5.A,

$$\text{cov}(X, Y) = E[XY] - \mu_1\mu_2 = \mu_1\mu_2 - \mu_1\mu_2 = 0,$$

as claimed. □

Theorem 2.5.2

Theorem 2.5.2. If X and Y are independent random variables then $\operatorname{cov}(X, Y) = 0$ and hence $\rho = 0$.

Proof. If X and Y are independent then by Theorem 2.4.2 we have $E[XY] = E[X]E[Y]$. So by Note 2.5.A,

$$\operatorname{cov}(X, Y) = E[XY] - \mu_1\mu_2 = \mu_1\mu_2 - \mu_1\mu_2 = 0,$$

as claimed. □

Theorem 2.5.3

Theorem 2.5.3. Suppose (X, Y) have a joint distribution with the variances of X and Y finite and positive. Denote the means and variances of X and Y by μ_1, μ_2 and σ_1^2, σ_2^2, respectively, and let ρ be the correlation coefficient between X and Y. If $E[Y \mid X]$ is linear in X then

$$E[Y \mid X]\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(X - \mu_1) \text{ and } E[\mathrm{Var}(Y \mid X)] = \sigma_2^2(1 - \rho^2).$$

Proof. We give a proof for the continuous case and leave the discrete case as an exercise. Suppose $E[Y \mid X]$ is linear in X, say $E[Y \mid x] = a + bx$. By the definition of conditional probability function (see Section 2.3) we have

$$f_{Y|X}(y \mid x) = f_{2|1}(y \mid x) = \frac{f_{X,Y}}{(x,y)} f_X(x) = \frac{f(x,y)}{f_1(x)}$$

and so by the definition of expected value (Definition 1.8.1) we have ...

Theorem 2.5.3

Theorem 2.5.3. Suppose (X, Y) have a joint distribution with the variances of X and Y finite and positive. Denote the means and variances of X and Y by μ_1, μ_2 and σ_1^2, σ_2^2, respectively, and let ρ be the correlation coefficient between X and Y. If $E[Y \mid X]$ is linear in X then

$$E[Y \mid X]\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(X - \mu_1) \text{ and } E[\mathrm{Var}(Y \mid X)] = \sigma_2^2(1 - \rho^2).$$

Proof. We give a proof for the continuous case and leave the discrete case as an exercise. Suppose $E[Y \mid X]$ is linear in X, say $E[Y \mid x] = a + bx$. By the definition of conditional probability function (see Section 2.3) we have

$$f_{Y|X}(y \mid x) = f_{2|1}(y \mid x) = \frac{f_{X,Y}}{(x, y)} f_X(x) = \frac{f(x, y)}{f_1(x)}$$

and so by the definition of expected value (Definition 1.8.1) we have ...

Theorem 2.5.3 (continued 1)

Proof (continued). ... we have the conditional expectation

$$a + bx = E[Y \mid x] = \int_{-\infty}^{\infty} y f_{Y|X}(y \mid x)\, dy = \int_{-\infty}^{\infty} \frac{yf(x,y)}{f_1(x)}\, dy$$
$$= \frac{1}{f_1(x)} \int_{-\infty}^{\infty} yf(x,y)\, dy.$$

Hence
$$\int_{-\infty}^{\infty} yf(x,y)\, dy = (a + bx) f_1(x). \tag{2.56}$$

By the definition of $E[Y]$ (see Section 2.1),
$E[Y] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} yf(x,y)\, dx\, dy$ so from (2.56) we have

$$E[Y] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} yf(x,y)\, dy = \int_{-\infty}^{\infty} (a + bx) f_1(x)\, dx$$
$$= a \int_{-\infty}^{\infty} f_1(x)\, dx + b \int_{-\infty}^{\infty} x f_1(x)\, dx = a + bE[X] \ldots$$

Theorem 2.5.3 (continued 1)

Proof (continued). ...we have the conditional expectation

$$a + bx = E[Y \mid x] = \int_{-\infty}^{\infty} y f_{Y|X}(y \mid x)\, dy = \int_{-\infty}^{\infty} \frac{yf(x,y)}{f_1(x)}\, dy$$

$$= \frac{1}{f_1(x)} \int_{-\infty}^{\infty} yf(x,y)\, dy.$$

Hence

$$\int_{-\infty}^{\infty} yf(x,y)\, dy = (a+bx)f_1(x). \tag{2.56}$$

By the definition of $E[Y]$ (see Section 2.1),
$E[Y] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(x,y)\, dx\, dy$ so from (2.56) we have

$$E[Y] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(x,y)\, dy = \int_{-\infty}^{\infty} (a+bx)f_1(x)\, dx$$

$$= a\int_{-\infty}^{\infty} f_1(x)\, dx + b\int_{-\infty}^{\infty} xf_1(x)\, dx = a + bE[X] \ldots$$

Theorem 2.5.3 (continued 2)

Proof (continued). or $\mu_2 = a + b\mu_1$ where $\mu_1 = E[X]$ and $\mu_2 = E[Y]$. Also from (2.56) (multiplying both sides by x and then integrating with respect to x):

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)\,dy\,dx = \int_{-\infty}^{\infty}(ax + bx^2)f_1(x)\,dx$$

$$= a\int_{-\infty}^{\infty} xf_1(x)\,dx + b\int_{-\infty}^{\infty} x^2 f_1(x)\,dx$$

or $E[XY] = aE[X] + bE[X^2]$. Now $E[XY] = \mu_1\mu_2 + \rho\sigma_1\sigma_2$ by Note 2.5.B and $E[X^2] = \sigma_1^2 + \mu_1^2$ by Note 1.9.A so that we now have $\mu_1\mu_2 + \rho\sigma_1\sigma_2 = a\mu_1 + b(\sigma_1^2 + \mu_1^2)$. We can now solve for a and b using the two linear equations $a + \mu_1 b = \mu_2$ and $\mu_1 a + (\sigma_1^2 + \mu_1^2)b = \mu_1\mu_2 + \rho\sigma_1\sigma_2$ to get $a = \mu_2 - \rho\mu_1\sigma_2/\sigma_1$ and $b = \rho\sigma_2/\sigma_1$. Therefore,

$$E[Y \mid x] = a + bx = \mu_2 = \rho\mu_1\sigma_2/\sigma_1 + \rho\sigma_2 x/\sigma_1 = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1),$$

or $E[Y \mid X] = \mu_2 + \rho\dfrac{\sigma_2}{\sigma_1}(X - \mu_1)$, as claimed.

Theorem 2.5.3 (continued 2)

Proof (continued). or $\mu_2 = a + b\mu_1$ where $\mu_1 = E[X]$ and $\mu_2 = E[Y]$. Also from (2.56) (multiplying both sides by x and then integrating with respect to x):

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)\,dy\,dx = \int_{-\infty}^{\infty}(ax + bx^2)f_1(x)\,dx$$
$$= a\int_{-\infty}^{\infty} xf_1(x)\,dx + b\int_{-\infty}^{\infty} x^2 f_1(x)\,dx$$

or $E[XY] = aE[X] + bE[X^2]$. Now $E[XY] = \mu_1\mu_2 + \rho\sigma_1\sigma_2$ by Note 2.5.B and $E[X^2] = \sigma_1^2 + \mu_1^2$ by Note 1.9.A so that we now have $\mu_1\mu_2 + \rho\sigma_1\sigma_2 = a\mu_1 + b(\sigma_1^2 + \mu_1^2)$. We can now solve for a and b using the two linear equations $a + \mu_1 b = \mu_2$ and $\mu_1 a + (\sigma_1^2 + \mu_1^2)b = \mu_1\mu_2 + \rho\sigma_1\sigma_2$ to get $a = \mu_2 - \rho\mu_1\sigma_2/\sigma_1$ and $b = \rho\sigma_2/\sigma_1$. Therefore,

$$E[Y \mid x] = a + bx = \mu_2 = \rho\mu_1\sigma_2/\sigma_1 + \rho\sigma_2 x/\sigma_1 = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1),$$

or $E[Y \mid X] = \mu_2 + \rho\dfrac{\sigma_2}{\sigma_1}(X - \mu_1)$, as claimed.

Theorem 2.5.3 (continued 3)

Proof (continued). Next, we consider $\mathrm{var}(Y \mid x)$. To do so, we need the conditional mean of Y given x which we see from above is $E[Y \mid x] = \mu_2 + \rho \frac{\sigma_2}{\sigma_1}(x - \mu_1)$. We have

$$\mathrm{var}(Y \mid x) = \int_{-\infty}^{\infty} \left(y - \left(\mu_2 + \rho \frac{\sigma_2}{\sigma_1}(x - \mu_1) \right) \right)^2 f_{2|1}(y \mid x)\, dy$$

$$= \frac{1}{f_1(x)} \int_{-\infty}^{\infty} \left(y - \mu_2 - \rho \frac{\sigma_2}{\sigma_1}(x - \mu_1) \right)^2 f(x, y)\, dy.$$

So

$$E[\mathrm{var}(Y \mid X)] = \int_{-\infty}^{\infty} \mathrm{var}(Y \mid x) f_1(x)\, dx$$

$$= \int_{-\infty}^{\infty} \left(\frac{1}{f_1(x)} \int_{-\infty}^{\infty} \left(y - \mu_2 - \rho \frac{\sigma_2}{\sigma_1}(x - \mu_1) \right)^2 f(x, y)\, dy \right) f_1(x)\, dx$$

Theorem 2.5.3 (continued 4)

Proof (continued). ...

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(y-\mu_2-\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)\right)^2 f(x,y)\,dy\,dx$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left((y=\mu_2)^2-2\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)(y-\mu_2)+\rho^2\frac{\sigma_2^2}{\sigma_1^2}(x-\mu_1)^2\right)\,dy\,dx$$

$$= E[(Y-\mu_2)^2]-2\rho\frac{\sigma_2}{\sigma_1}E[(X-\mu_1)(Y-\mu_2)]+\rho^2\frac{\sigma_2^1}{\sigma_1^2}E[(X-\mu_1)^2]$$

$$= \sigma_2^2-2\rho\frac{\sigma_2}{\sigma_1}\text{cov}(X,Y)+\rho^2\frac{\sigma_2^2}{\sigma_1^2}\sigma_1^2$$

$$= \sigma_2^2-2\rho\frac{\sigma_2}{\sigma_1}(\rho\sigma_1\sigma_2)+\rho^2\sigma_2^2 \text{ since, by Definition 2.5.2, } \rho=\frac{\text{cov}(X,Y)}{\sigma_1\sigma_2}$$

$$= \sigma_2-2\rho^2\sigma_2^2+\rho^2\sigma_2^2=\sigma_2^2(1-2\rho^2+\rho^2)=\sigma_2^2(1-\rho^2),$$

as claimed. □