Expectations and covariances

Tiago M. Magalhães

Department of Statistics - ICE-UFJF

Juiz de Fora, September 04, 2026

Outline

Outline

Introduction

Let $(X, Y)^\top$ be a two-dimensional random variable (RV) and g be a function of these two variables.

Expectation

If $(X, Y)^\top$ is a two-dimensional discrete RV, with joint probability mass function $p(x, y)$, then

$$\mathbb{E}\left[g(X, Y)\right] = \sum_{y=-\infty}^{+\infty} \sum_{x=-\infty}^{+\infty} g(x, y)\, p(x, y).$$

Expectation

If $(X, Y)^\top$ is a two-dimensional discrete RV, with joint probability mass function $p(x, y)$, then

$$\mathbb{E}\left[g(X, Y)\right] = \sum_{y=-\infty}^{+\infty} \sum_{x=-\infty}^{+\infty} g(x, y)\, p(x, y).$$

Expectation

If $(X, Y)^\top$ is a two-dimensional continuous RV, with joint probability density function (PDF) $f(x, y)$, then

$$\mathbb{E}\left[g(X, Y)\right] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) f(x, y) dx dy.$$

Expectation

If $(X, Y)^\top$ is a two-dimensional continuous RV, with joint probability density function (PDF) $f(x, y)$, then

$$\mathbb{E}\left[g(X, Y)\right] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x, y) f(x, y) dx dy.$$

Covariance

The covariance between X and Y, denoted by $\mathrm{Cov}(X, Y)$,

Covariance

The covariance between X and Y, denoted by $\mathrm{Cov}(X, Y)$, is defined by

$$\mathrm{Cov}(X, Y) = \mathbb{E}\left\{[X - \mathbb{E}(X)]\,[Y - \mathbb{E}(Y)]\right\}.$$

Covariance

The covariance between X and Y, denoted by $\text{Cov}(X, Y)$, is defined by

$$\text{Cov}(X, Y) = \mathbb{E}\left\{[X - \mathbb{E}(X)]\,[Y - \mathbb{E}(Y)]\right\}.$$

Alternatively, the covariance between X and Y can be written as

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

Covariance

The covariance between X and Y, denoted by $\text{Cov}(X, Y)$, is defined by

$$\text{Cov}(X, Y) = \mathbb{E}\left\{[X - \mathbb{E}(X)]\,[Y - \mathbb{E}(Y)]\right\}.$$

Alternatively, the covariance between X and Y can be written as

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

Covariance

Let a, b, c, d be constants, from the definition of covariance, we have the following properties:

Covariance

Let a, b, c, d be constants, from the definition of covariance, we have the following properties:

- $\text{Cov}(X, Y) = \text{Cov}(Y, X)$;

Covariance

Let a, b, c, d be constants, from the definition of covariance, we have the following properties:

- $\text{Cov}(X, Y) = \text{Cov}(Y, X)$;
- $\text{Cov}(X, X) = \text{Var}(X)$;

Covariance

Let a, b, c, d be constants, from the definition of covariance, we have the following properties:

- $\text{Cov}(X, Y) = \text{Cov}(Y, X)$;
- $\text{Cov}(X, X) = \text{Var}(X)$;
- $\text{Cov}(aX + b, cY + d) = ac\text{Cov}(X, Y)$;

Covariance

Let a, b, c, d be constants, from the definition of covariance, we have the following properties:

- $\text{Cov}(X, Y) = \text{Cov}(Y, X)$;
- $\text{Cov}(X, X) = \text{Var}(X)$;
- $\text{Cov}(aX + b, cY + d) = ac\text{Cov}(X, Y)$;
- $\text{Cov}\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right) = \sum_{i=1}^{n} \sum_{j=1}^{m} \text{Cov}(X_i, Y_j)$.

Covariance

Let a, b, c, d be constants, from the definition of covariance, we have the following properties:

- $\text{Cov}(X, Y) = \text{Cov}(Y, X)$;
- $\text{Cov}(X, X) = \text{Var}(X)$;
- $\text{Cov}(aX + b, cY + d) = ac\text{Cov}(X, Y)$;
- $\text{Cov}\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} \text{Cov}(X_i, Y_j)$.

Covariance

From the definition and properties of covariance, we have some results. The first is that, if X and Y are independent,

Covariance

From the definition and properties of covariance, we have some results. The first is that, if X and Y are independent, then

$$\mathbb{E}(XY) = \mathbb{E}(X) \times \mathbb{E}(Y),$$

Covariance

From the definition and properties of covariance, we have some results. The first is that, if X and Y are independent, then

$$\mathbb{E}(XY) = \mathbb{E}(X) \times \mathbb{E}(Y),$$

that is, if X and Y are independent, $\text{Cov}(X, Y) = 0$.

Covariance

From the definition and properties of covariance, we have some results. The first is that, if X and Y are independent, then

$$\mathbb{E}(XY) = \mathbb{E}(X) \times \mathbb{E}(Y),$$

that is, if X and Y are independent, $\text{Cov}(X, Y) = 0$.

Remark: $\text{Cov}(X, Y) = 0$ does not imply that X and Y are independent.

Covariance

From the definition and properties of covariance, we have some results. The first is that, if X and Y are independent, then

$$\mathbb{E}(XY) = \mathbb{E}(X) \times \mathbb{E}(Y),$$

that is, if X and Y are independent, $\text{Cov}(X, Y) = 0$.

Remark: $\text{Cov}(X, Y) = 0$ does not imply that X and Y are independent.

Covariance

For the second result, let $X_1, X_2, \ldots, X_n$ be a sequence of RVs, then

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}\left(X_i\right) + 2 \sum_{1 \leq i < j \leq n} \mathrm{Cov}(X_i, X_j). \tag{1}$$

Covariance

For the second result, let $X_1, X_2, \ldots, X_n$ be a sequence of RVs, then

$$\operatorname{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \operatorname{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} \operatorname{Cov}(X_i, X_j). \tag{1}$$

For the two-dimensional case $(X, Y)^\top$,

Covariance

For the second result, let $X_1, X_2, \ldots, X_n$ be a sequence of RVs, then

$$\text{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \text{Var}\left(X_i\right) + 2 \sum_{1 \le i < j \le n} \text{Cov}(X_i, X_j). \tag{1}$$

For the two-dimensional case $(X, Y)^\top$, we have that (1) results in

$$\text{Var}\left(X + Y\right) = \text{Var}\left(X\right) + \text{Var}\left(Y\right) + 2\text{Cov}(X, Y).$$

Covariance

For the second result, let $X_1, X_2, \ldots, X_n$ be a sequence of RVs, then

$$\text{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \text{Var}\,(X_i) + 2 \sum_{1 \leq i < j \leq n} \text{Cov}(X_i, X_j). \tag{1}$$

For the two-dimensional case $(X, Y)^\top$, we have that (1) results in

$$\text{Var}\,(X + Y) = \text{Var}\,(X) + \text{Var}\,(Y) + 2\text{Cov}(X, Y).$$

Covariance

Now, if $X_1, X_2, \ldots, X_n$ form a sequence of pairwise independent RVs, (1) results in

$$\text{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \text{Var}\left(X_i\right).$$

Covariance

Now, if $X_1, X_2, \ldots, X_n$ form a sequence of pairwise independent RVs, (1) results in

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}\left(X_i\right).$$

Correlation

Let $(X, Y)^\top$ be a two-dimensional RV. We define ρ_{xy}, the correlation coefficient between X and Y, as follows:

Correlation

Let $(X, Y)^\top$ be a two-dimensional RV. We define ρ_{xy}, the correlation coefficient between X and Y, as follows:

$$\rho_{xy} = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}},$$

with $\text{Var}(X) > 0$ and $\text{Var}(Y) > 0$.

Correlation

Let $(X, Y)^\top$ be a two-dimensional RV. We define ρ_{xy}, the correlation coefficient between X and Y, as follows:

$$\rho_{xy} = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}},$$

with $\text{Var}(X) > 0$ and $\text{Var}(Y) > 0$.

Correlation

- ρ_{xy} is a dimensionless quantity;

Correlation

- ρ_{xy} is a dimensionless quantity;
- $-1 \leq \rho_{xy} \leq 1$;

Correlation

- ρ_{xy} is a dimensionless quantity;
- $-1 \leq \rho_{xy} \leq 1$;
- If X and Y are independent, then $\rho_{xy} = 0$;

Correlation

- ρ_{xy} is a dimensionless quantity;
- $-1 \leq \rho_{xy} \leq 1$;
- If X and Y are independent, then $\rho_{xy} = 0$;
- If $\rho_{xy}^2 = 1$, then there exist constants a and b such that $Y = aX + b$, almost surely;

Correlation

- ρ_{xy} is a dimensionless quantity;
- $-1 \leq \rho_{xy} \leq 1$;
- If X and Y are independent, then $\rho_{xy} = 0$;
- If $\rho_{xy}^2 = 1$, then there exist constants a and b such that $Y = aX + b$, almost surely;
- If $Y = aX + b$ and $\rho_{xy}^2 = 1$, then, if $a > 0$, $\rho_{xy} = 1$, while, if $a < 0$, $\rho_{xy} = -1$.

Correlation

- ρ_{xy} is a dimensionless quantity;
- $-1 \leq \rho_{xy} \leq 1$;
- If X and Y are independent, then $\rho_{xy} = 0$;
- If $\rho_{xy}^2 = 1$, then there exist constants a and b such that $Y = aX + b$, almost surely;
- If $Y = aX + b$ and $\rho_{xy}^2 = 1$, then, if $a > 0$, $\rho_{xy} = 1$, while, if $a < 0$, $\rho_{xy} = -1$.

Correlation

Result

If ρ_{xy} is the correlation coefficient between X and Y,

Correlation

Result

If ρ_{xy} is the correlation coefficient between X and Y, and if $V = aX + b$ and $W = cY + d$,

Correlation

Result

If ρ_{xy} is the correlation coefficient between X and Y, and if $V = aX + b$ and $W = cY + d$, where a, b, c, and d, with $a \neq 0$ and $c \neq 0$, are constants,

Correlation

Result

If ρ_{xy} is the correlation coefficient between X and Y, and if $V = aX + b$ and $W = cY + d$, where a, b, c, and d, with $a \neq 0$ and $c \neq 0$, are constants, then

$$\rho_{vw} = \frac{ac}{|ac|}\rho_{xy}.$$

Correlation

Result

If ρ_{xy} is the correlation coefficient between X and Y, and if $V = aX + b$ and $W = cY + d$, where a, b, c, and d, with $a \neq 0$ and $c \neq 0$, are constants, then

$$\rho_{vw} = \frac{ac}{|ac|}\rho_{xy}.$$

Example

Let $(X, Y)^\top$ be a two-dimensional RV, with joint PDF given by:

$$
\begin{aligned}
f(x, y) &= 2\,\mathbb{I}_{(0,1)}(x)\,\mathbb{I}_{(x,1)}(y) \\
&= 2\,\mathbb{I}_{(0,1)}(y)\,\mathbb{I}_{(0,y)}(x).
\end{aligned}
$$

Find ρ_{xy}.

Example

Let $(X, Y)^\top$ be a two-dimensional RV, with joint PDF given by:

$$f(x,y) = 2\,\mathbb{I}_{(0,1)}(x)\,\mathbb{I}_{(x,1)}(y)$$
$$= 2\,\mathbb{I}_{(0,1)}(y)\,\mathbb{I}_{(0,y)}(x).$$

Find ρ_{xy}.

Outline

Conditional expectation

Just as we defined the expectation of a random variable X, we can also define the **conditional expectation** of an RV.

Conditional expectation

Just as we defined the expectation of a random variable X, we can also define the **conditional expectation** of an RV.

Conditional expectation

If $(X, Y)^\top$ is a two-dimensional discrete RV, we define the conditional expectation of X, for a given $Y = y$,

Conditional expectation

If $(X, Y)^\top$ is a two-dimensional discrete RV, we define the conditional expectation of X, for a given $Y = y$, as

$$\mathbb{E}(X|Y=y) = \sum_{x=-\infty}^{+\infty} x\, p_{X|Y}(x|y).$$

Conditional expectation

If $(X, Y)^\top$ is a two-dimensional discrete RV, we define the conditional expectation of X, for a given $Y = y$, as

$$\mathbb{E}(X|Y = y) = \sum_{x=-\infty}^{+\infty} x\, p_{X|Y}(x|y).$$

Conditional expectation

If $(X, Y)^\top$ is a two-dimensional continuous RV, we define the conditional expectation of X, for a given $Y = y$,

Conditional expectation

If $(X, Y)^\top$ is a two-dimensional continuous RV, we define the conditional expectation of X, for a given $Y = y$, as

$$\mathbb{E}(X|Y = y) = \int_{-\infty}^{+\infty} x\, f_{X|Y}(x|y)\, dx.$$

Conditional expectation

If $(X, Y)^\top$ is a two-dimensional continuous RV, we define the conditional expectation of X, for a given $Y = y$, as

$$\mathbb{E}(X|Y = y) = \int_{-\infty}^{+\infty} x\, f_{X|Y}(x|y)\, dx.$$

Conditional expectation

In general, if $(X, Y)^\top$ is a two-dimensional RV and g is a function of X, we have that

$$\begin{cases} \mathbb{E}\left[g(X)|Y=y\right] = \sum_{x=-\infty}^{+\infty} g(x)\, p_{X|Y}(x|y), & \text{discrete case,} \\ \mathbb{E}\left[g(X)|Y=y\right] = \int_{-\infty}^{+\infty} g(x)\, f_{X|Y}(x|y)dx, & \text{continuous case.} \end{cases}$$

Conditional expectation

In general, if $(X, Y)^\top$ is a two-dimensional RV and g is a function of X, we have that

$$\begin{cases} \mathbb{E}\left[g(X)|Y=y\right] = \sum_{x=-\infty}^{+\infty} g(x)\, p_{X|Y}(x|y), & \text{discrete case,} \\ \mathbb{E}\left[g(X)|Y=y\right] = \int_{-\infty}^{+\infty} g(x)\, f_{X|Y}(x|y)dx, & \text{continuous case.} \end{cases}$$

Remark

In general, $\mathbb{E}(X|Y = y)$ is a function of y. Thus, $\mathbb{E}(X|Y)$ is a random variable.

Remark

In general, $\mathbb{E}(X|Y = y)$ is a function of y. Thus, $\mathbb{E}(X|Y)$ is a random variable. Strictly speaking, $\mathbb{E}(X|Y = y)$ is the value of the RV $\mathbb{E}(X|Y)$.

Remark

In general, $\mathbb{E}(X|Y = y)$ is a function of y. Thus, $\mathbb{E}(X|Y)$ is a random variable. Strictly speaking, $\mathbb{E}(X|Y = y)$ is the value of the RV $\mathbb{E}(X|Y)$.

Conditional expectation

An important property of the conditional expectation is

$$\mathbb{E}(X) = \mathbb{E}\left[\mathbb{E}(X|Y)\right]. \tag{2}$$

Similarly, $\mathbb{E}(Y) = \mathbb{E}\left[\mathbb{E}(Y|X)\right]$.

Conditional expectation

An important property of the conditional expectation is

$$\mathbb{E}(X) = \mathbb{E}\left[\mathbb{E}(X|Y)\right]. \tag{2}$$

Similarly, $\mathbb{E}(Y) = \mathbb{E}\left[\mathbb{E}(Y|X)\right]$.

Conditional expectation

The result in (2) also remains valid for situations such as

- $\mathbb{E}[g(X)] = \mathbb{E}\left\{\mathbb{E}[g(X)|Y]\right\}$;

Conditional expectation

The result in (2) also remains valid for situations such as

- $\mathbb{E}[g(X)] = \mathbb{E}\left\{\mathbb{E}[g(X)|Y]\right\}$;
- $\mathbb{E}(XY) = \mathbb{E}\left[\mathbb{E}(XY|Y)\right]$.

Conditional expectation

The result in (2) also remains valid for situations such as

- $\mathbb{E}[g(X)] = \mathbb{E}\{\mathbb{E}[g(X)|Y]\}$;
- $\mathbb{E}(XY) = \mathbb{E}[\mathbb{E}(XY|Y)]$.

Conditional expectation

Another important property is

$$\mathsf{Var}(X) = \mathbb{E}\left[\mathsf{Var}(X|Y)\right] + \mathsf{Var}\left[\mathbb{E}(X|Y)\right].$$

Example

Suppose that shipments containing a variable number of parts arrive each day. If N is the number of parts in the shipment,

Example

Suppose that shipments containing a variable number of parts arrive each day. If N is the number of parts in the shipment, the probability distribution of the random variable N,

Example

Suppose that shipments containing a variable number of parts arrive each day. If N is the number of parts in the shipment, the probability distribution of the random variable N, will be given as follows:

N	10	11	12	13	14	15
$\mathbb{P}(N)$	0.05	0.10	0.10	0.20	0.35	0.20

Example

Suppose that shipments containing a variable number of parts arrive each day. If N is the number of parts in the shipment, the probability distribution of the random variable N, will be given as follows:

N	10	11	12	13	14	15
$\mathbb{P}(N)$	0.05	0.10	0.10	0.20	0.35	0.20

Thus, $\mathbb{E}(N) = 13.3$.

Example

Suppose that shipments containing a variable number of parts arrive each day. If N is the number of parts in the shipment, the probability distribution of the random variable N, will be given as follows:

N	10	11	12	13	14	15
$\mathbb{P}(N)$	0.05	0.10	0.10	0.20	0.35	0.20

Thus, $\mathbb{E}(N) = 13.3$.

Example

The probability that any particular part is defective is the same for all parts

and equal to 0,10.

Example

The probability that any particular part is defective is the same for all parts and equal to 0,10. If X is the number of defective parts that arrive each day, what is the expectation of X?

Example

The probability that any particular part is defective is the same for all parts and equal to 0,10. If X is the number of defective parts that arrive each day, what is the expectation of X?

Outline

Regression function

As we have already pointed out, $\mathbb{E}(X|Y = y)$ is the value of the RV $\mathbb{E}(X|Y)$ and is a function of y. The graph of this function of y is known as the **regression curve** of X on Y.

Regression function

As we have already pointed out, $\mathbb{E}(X|Y = y)$ is the value of the RV $\mathbb{E}(X|Y)$ and is a function of y. The graph of this function of y is known as the **regression curve** of X on Y. For each fixed value y, $\mathbb{E}(X|Y = y)$ is the expected value of the (one-dimensional) RV $X|Y = y$.

Regression function

As we have already pointed out, $\mathbb{E}(X|Y=y)$ is the value of the RV $\mathbb{E}(X|Y)$ and is a function of y. The graph of this function of y is known as the **regression curve** of X on Y. For each fixed value y, $\mathbb{E}(X|Y=y)$ is the expected value of the (one-dimensional) RV $X|Y=y$.

Theorem

Let $(X, Y)^\top$ be a two-dimensional RV and suppose that

$$\mathbb{E}(X) = \mu_X,\ \mathbb{E}(Y) = \mu_Y,\ \text{Var}(X) = \sigma_X^2 \text{ and } \text{Var}(Y) = \sigma_Y^2.$$

Theorem

Let $(X, Y)^\top$ be a two-dimensional RV and suppose that

$$\mathbb{E}(X) = \mu_X,\ \mathbb{E}(Y) = \mu_Y,\ \text{Var}(X) = \sigma_X^2 \text{ and } \text{Var}(Y) = \sigma_Y^2.$$

Theorem

Let ρ be the correlation coefficient between X and Y. If the regression of Y on X is linear,

Theorem

Let ρ be the correlation coefficient between X and Y. If the regression of Y on X is linear, we have

$$\mathbb{E}(Y|X=x) = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X).$$

Theorem

Let ρ be the correlation coefficient between X and Y. If the regression of Y on X is linear, we have

$$\mathbb{E}(Y|X=x) = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X).$$

If the regression of X on Y is linear,

Theorem

Let ρ be the correlation coefficient between X and Y. If the regression of Y on X is linear, we have

$$\mathbb{E}(Y|X = x) = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X).$$

If the regression of X on Y is linear, we have

$$\mathbb{E}(X|Y = y) = \mu_X + \rho\frac{\sigma_X}{\sigma_Y}(y - \mu_Y).$$

Theorem

Let ρ be the correlation coefficient between X and Y. If the regression of Y on X is linear, we have

$$\mathbb{E}(Y|X=x) = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X).$$

If the regression of X on Y is linear, we have

$$\mathbb{E}(X|Y=y) = \mu_X + \rho\frac{\sigma_X}{\sigma_Y}(y-\mu_Y).$$

Remarks

- One of the regression functions may be linear, while the other is not;

Remarks

- One of the regression functions may be linear, while the other is not;
- If the regression of X on Y, for example, is linear and if $\rho = 0$,

Remarks

- One of the regression functions may be linear, while the other is not;
- If the regression of X on Y, for example, is linear and if $\rho = 0$, then we see that $\mathbb{E}(X|Y = y)$ does not depend on y.

Remarks

- One of the regression functions may be linear, while the other is not;
- If the regression of X on Y, for example, is linear and if $\rho = 0$, then we see that $\mathbb{E}(X|Y = y)$ does not depend on y. If $\rho \neq 0$, the sign of ρ determines the sign of the slope of the regression line;

Remarks

- One of the regression functions may be linear, while the other is not;
- If the regression of X on Y, for example, is linear and if $\rho = 0$, then we see that $\mathbb{E}(X|Y = y)$ does not depend on y. If $\rho \neq 0$, the sign of ρ determines the sign of the slope of the regression line;
- If both regression functions are linear,

Remarks

- One of the regression functions may be linear, while the other is not;
- If the regression of X on Y, for example, is linear and if $\rho = 0$, then we see that $\mathbb{E}(X|Y = y)$ does not depend on y. If $\rho \neq 0$, the sign of ρ determines the sign of the slope of the regression line;
- If both regression functions are linear, we see that the regression lines intersect at the "center" of the distribution, (μ_X, μ_Y).

Remarks

- One of the regression functions may be linear, while the other is not;
- If the regression of X on Y, for example, is linear and if $\rho = 0$, then we see that $\mathbb{E}(X|Y = y)$ does not depend on y. If $\rho \neq 0$, the sign of ρ determines the sign of the slope of the regression line;
- If both regression functions are linear, we see that the regression lines intersect at the "center" of the distribution, (μ_X, μ_Y).

Least squares principle

Although the regression functions do not need to be linear, we may be interested in approximating the regression curve with a linear function. This is usually done using the **least squares principle**.

Least squares principle

Although the regression functions do not need to be linear, we may be interested in approximating the regression curve with a linear function. This is usually done using the **least squares principle**.

Least squares principle

Under the least squares principle, two constants a and b are chosen so that

$$\mathbb{E}\left\{[\mathbb{E}(Y|X) - (aX + b)]^2\right\}$$

becomes minimum.

Least squares principle

Under the least squares principle, two constants a and b are chosen so that

$$\mathbb{E}\left\{[\mathbb{E}(Y|X) - (aX + b)]^2\right\}$$

becomes minimum. The line $y = ax + b$ is called the least-squares approximation to the regression curve $\mathbb{E}(Y|X = x)$.

Least squares principle

Under the least squares principle, two constants a and b are chosen so that

$$\mathbb{E}\left\{[\mathbb{E}(Y|X) - (aX + b)]^2\right\}$$

becomes minimum. The line $y = ax + b$ is called the least-squares approximation to the regression curve $\mathbb{E}(Y|X = x)$.

Theorem

If $y = ax + b$ is the least-squares approximation to the regression function $\mathbb{E}(Y|X = x)$ and if $\mathbb{E}(Y|X = x)$ is in fact a linear function of x,

Theorem

If $y = ax + b$ is the least-squares approximation to the regression function $\mathbb{E}(Y|X = x)$ and if $\mathbb{E}(Y|X = x)$ is in fact a linear function of x, that is,

$$\mathbb{E}(Y|X = x) = a'x + b',$$

then $a = a'$ and $b = b'$.

Theorem

If $y = ax + b$ is the least-squares approximation to the regression function $\mathbb{E}(Y|X = x)$ and if $\mathbb{E}(Y|X = x)$ is in fact a linear function of x, that is,

$$\mathbb{E}(Y|X = x) = a'x + b',$$

then $a = a'$ and $b = b'$.

Remark: similarly, the results hold for $\mathbb{E}(X|Y = y)$.

Theorem

If $y = ax + b$ is the least-squares approximation to the regression function $\mathbb{E}(Y|X = x)$ and if $\mathbb{E}(Y|X = x)$ is in fact a linear function of x, that is,

$$\mathbb{E}(Y|X = x) = a'x + b',$$

then $a = a'$ and $b = b'$.

Remark: similarly, the results hold for $\mathbb{E}(X|Y = y)$.

Outline

Bibliography

Meyer, P. L. (2017). *Probabilidade: aplicações à Estatística*. Livros Técnicos e Científicos Editora Ltda, Rio de Janeiro, 2 edition.

Ross, S. (2019). *A first course in probability*. Pearson Education, Malaysia, 10 edn. global edition.

Thank you!

✉ tiago.magalhaes@ufjf.br

ufjf.br/tiago_magalhaes

Department of Statistics, Room 319