

Esperanza, Varianza y Momentos

Clase 4 — 14.310x: Data Analysis for Social Scientists

B.Sc. Cristian Lazo Quispe

BREIT (Brescia Institute of Technology) · Aporta — Grupo Breca

Programa SDS (Statistics and Data Science, MITx) · **BREIT** (Brescia Institute of Technology), iniciativa de Aporta —Grupo Breca— en alianza con MIT IDSS.
Basado en MIT 14.310x (CC BY-NC-SA 4.0).

✉ clazoq@uni.pe ·  linkedin.com/in/cristian-lazo-quispe ·  github.com/CristianLazoQuispe

¿Qué veremos hoy?

- 1 Funciones de variables aleatorias
- 2 Esperanza $\mathbb{E}[X]$
- 3 Varianza
- 4 Momentos
- 5 Covarianza y correlación
- 6 Hacia la regresión
- 7 Manos a la obra

Contenido

1 Funciones de variables aleatorias

2 Esperanza $\mathbb{E}[X]$

3 Varianza

4 Momentos

5 Covarianza y correlación

6 Hacia la regresión

7 Manos a la obra

De una v.a. a una función de ella

La idea

Si X es aleatoria, cualquier transformación $Y = g(X)$ *también* lo es.

Muchas cantidades de interés son funciones de una v.a.:

- el logaritmo del ingreso $\log X$,
- el ingreso convertido de soles a dólares $Y = X/3.7$,
- una desviación al cuadrado $(X - \mu)^2$.

La pregunta: **¿cuál es la distribución de Y ?**

El método de la cdf

Definición 1 (Método de la cdf (the cdf method))

Para hallar la distribución de $Y = g(X)$:

- 1 Determina el **soporte inducido** de Y .
- 2 Escribe la cdf de Y como probabilidad sobre X :
 $F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(g(X) \leq y)$.
- 3 Resuelve $g(X) \leq y$ como un evento sobre X y evalúalo con F_X .
- 4 Si Y es continua, deriva: $f_Y(y) = F'_Y(y)$.

Ejemplo: la función no monótona $Y = X^2$

Sea $X \sim \text{Unif}[-1, 1]$, con $f_X(x) = \frac{1}{2}$ y $F_X(x) = \frac{x+1}{2}$ en $[-1, 1]$. Queremos f_Y con $Y = X^2$ (que vive en $[0, 1]$).

El truco: traducir bien el evento

$$X^2 \leq y \iff -\sqrt{y} \leq X \leq \sqrt{y}.$$

$$F_Y(y) = F_X(\sqrt{y}) - F_X(-\sqrt{y}) = \frac{\sqrt{y} + 1}{2} - \frac{-\sqrt{y} + 1}{2} = \sqrt{y}, \quad f_Y(y) = \frac{1}{2\sqrt{y}}.$$

¡Cuidado!

Si g no es monótona, no olvides la rama negativa: $\mathbb{P}(X^2 \leq y) \neq \mathbb{P}(X \leq \sqrt{y})$.

Demo resuelto: $Y = X^2$ con $X \sim \text{Unif}[0, 1]$

Problema

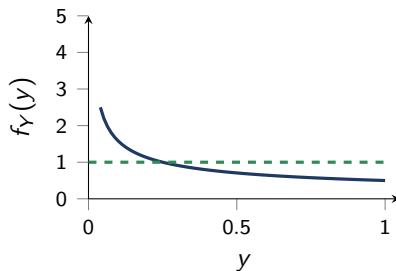
Sea $X \sim \text{Unif}[0, 1]$, con $f_X(x) = 1$ y $F_X(x) = x$ en $[0, 1]$. Se pide: f_Y de $Y = X^2$ (vive en $[0, 1]$).

Solución (método de la cdf)

Para $0 \leq y \leq 1$, como $X \geq 0$:

$$F_Y(y) = \mathbb{P}(X \leq \sqrt{y}) = F_X(\sqrt{y}) = \sqrt{y}.$$

Derivando, $f_Y(y) = \frac{1}{2\sqrt{y}}, 0 < y \leq 1.$



Línea azul: $f_Y(y) = \frac{1}{2\sqrt{y}}.$

Verde punteada: la $f_X = 1$ de partida.

Contenido

- 1 Funciones de variables aleatorias
- 2 Esperanza $\mathbb{E}[X]$**
- 3 Varianza
- 4 Momentos
- 5 Covarianza y correlación
- 6 Hacia la regresión
- 7 Manos a la obra

¿Dónde está el centro? La esperanza

Definición 2 (Esperanza / valor esperado)

La *esperanza (expectation, mean)* de X , denotada $\mathbb{E}[X]$ o μ , es

$$\underbrace{\mathbb{E}[X] = \sum_x x f_X(x)}_{\text{discreta}}$$

$$\underbrace{\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x) dx}_{\text{continua}}$$

Intuición: el punto de equilibrio (balancing point)

Si la pdf es masa repartida sobre la recta, $\mathbb{E}[X]$ es el lugar donde la distribución “hace balanza”.

Demo: esperanza discreta (años de educación)

Problema

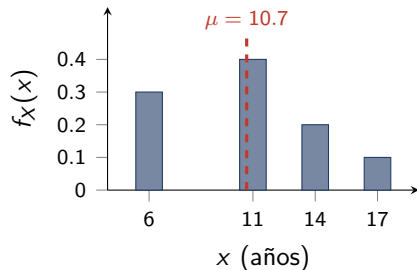
Nivel educativo X (años) de un adulto, con pmf:

x	6	11	14	17
$f_X(x)$	0.30	0.40	0.20	0.10

Se pide: la educación promedio $\mathbb{E}[X]$.

Solución

$$\begin{aligned}\mathbb{E}[X] &= 6(0.30) + 11(0.40) + 14(0.20) + 17(0.10) \\ &= 10.7 \text{ años.}\end{aligned}$$



Barras: pmf. Línea roja: la media $\mu = \mathbb{E}[X]$ (punto de equilibrio).

Demo: esperanza continua (la exponencial)

Problema

El tiempo X (en meses) hasta conseguir empleo es exponencial con $f_X(x) = \lambda e^{-\lambda x}$ para $x \geq 0$, con $\lambda = \frac{1}{4}$ por mes.

Se pide: la espera media $\mathbb{E}[X]$.

Solución

Integrando por partes,

$$\mathbb{E}[X] = \int_0^{\infty} x \lambda e^{-\lambda x} dx = \frac{1}{\lambda} = \frac{1}{1/4} = 4 \text{ meses.}$$

Propiedades: linealidad de la esperanza

Linealidad (linearity of expectation)

Para constantes a, b y v.a. X, Y (¡no hace falta independencia!):

$$\mathbb{E}[a] = a, \quad \mathbb{E}[aX + b] = a\mathbb{E}[X] + b, \quad \mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y].$$

Error clásico del curso

$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$ vale **siempre**.

$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ **solo** si $X \perp\!\!\!\perp Y$.

LOTUS: la ley del estadístico inconsciente

Teorema 1 (LOTUS (law of the unconscious statistician))

Para calcular $\mathbb{E}[g(X)]$ no hace falta hallar primero la distribución de $Y = g(X)$. Basta:

$$\mathbb{E}[g(X)] = \sum_x g(x) f_X(x) \quad \text{o} \quad \mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx.$$

Ejemplo express

Con un dado justo y $g(X) = 2X + 1$:

por linealidad $\mathbb{E}[2X + 1] = 2(3.5) + 1 = 8$; por LOTUS

$\frac{1}{6}(3 + 5 + 7 + 9 + 11 + 13) = \frac{48}{6} = 8$. Coinciden.

Contenido

- 1 Funciones de variables aleatorias
- 2 Esperanza $\mathbb{E}[X]$
- 3 Varianza**
- 4 Momentos
- 5 Covarianza y correlación
- 6 Hacia la regresión
- 7 Manos a la obra

¿Qué tan dispersa está? La varianza

Dos países pueden tener el mismo ingreso *promedio* y, aun así, uno ser mucho más desigual. La media no captura eso; la varianza sí.

Definición 3 (Varianza y desviación estándar)

$$\text{Var}(X) = \mathbb{E}[(X - \mu)^2], \quad \sigma = \text{SD}(X) = \sqrt{\text{Var}(X)}.$$

La desviación estándar σ tiene las mismas unidades que X .

Fórmula de atajo (la que usarás siempre)

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

“El promedio de los cuadrados menos el cuadrado del promedio”.

Propiedades de la varianza

Reglas

$$\text{Var}(X) \geq 0, \quad \text{Var}(a) = 0, \quad \boxed{\text{Var}(aX + b) = a^2 \text{Var}(X)}$$

- **Sumar** una constante b desplaza la distribución pero **no** cambia su dispersión.
- **Multiplicar** por a la estira: la varianza cambia por a^2 (la desviación estándar, por $|a|$).

Estandarizar

$$Z = \frac{X - \mu}{\sigma} \text{ tiene siempre } \mathbb{E}[Z] = 0 \text{ y } \text{Var}(Z) = 1.$$

Demo resuelto: $\mathbb{E}[X]$ y $\text{Var}(X)$ de una pmf

Problema

Misma educación X con pmf

$f_X(6) = 0.30$, $f_X(11) = 0.40$, $f_X(14) = 0.20$, $f_X(17) = 0.10$ (ya calculamos $\mathbb{E}[X] = 10.7$). **Se pide:** la varianza $\text{Var}(X)$ y la desviación estándar σ .

Paso 1: segundo momento $\mathbb{E}[X^2]$

$$\mathbb{E}[X^2] = 36(0.30) + 121(0.40) + 196(0.20) + 289(0.10) = 127.3.$$

Paso 2: atajo

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = 127.3 - (10.7)^2 = 127.3 - 114.49 = 12.81,$$

de donde $\sigma = \sqrt{12.81} \approx 3.58$ años.

Contenido

- 1 Funciones de variables aleatorias
- 2 Esperanza $\mathbb{E}[X]$
- 3 Varianza
- 4 Momentos**
- 5 Covarianza y correlación
- 6 Hacia la regresión
- 7 Manos a la obra

Momentos: resumir la forma con números

Definición 4 (Momentos (moments))

El k -ésimo momento de X es $\mathbb{E}[X^k]$. El k -ésimo momento central es $\mathbb{E}[(X - \mu)^k]$.

- 1.^{er} momento $\mathbb{E}[X] = \mu$: la *localización* (centro).
- 2.^{do} momento central $\mathbb{E}[(X - \mu)^2] = \text{Var}(X)$: la *dispersión*.
- 3.^{er} momento central: la *asimetría* (skewness).
- 4.^{to} momento central: la *curtosis* (kurtosis, grosor de colas).

Idea

Cada momento resume *un* rasgo de la forma. En la práctica casi siempre nos basta con los dos primeros: **media** y **varianza**.

Demo resuelto: asimetría (skewness) de una pmf

Problema

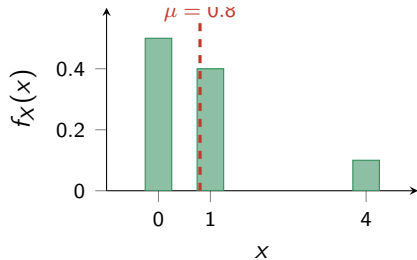
Una v.a. X toma valores $\{0, 1, 4\}$ con pmf $f_X(0) = 0.5$, $f_X(1) = 0.4$, $f_X(4) = 0.1$. Se pide: el signo del 3.^{er} momento central $\mathbb{E}[(X - \mu)^3]$ (¿asimétrica a la derecha?).

Solución

Media $\mu = 1(0.4) + 4(0.1) = 0.8$.

$$\begin{aligned}\mathbb{E}[(X - \mu)^3] &= \sum_x (x - 0.8)^3 f_X(x) \\ &= -0.256 + 0.0032 + 3.2768 = +3.024.\end{aligned}$$

Positivo $\Rightarrow$ cola a la **derecha** (sesgo +).



La masa larga hacia la derecha (en $x = 4$) genera el sesgo positivo.

Contenido

1 Funciones de variables aleatorias

2 Esperanza $\mathbb{E}[X]$

3 Varianza

4 Momentos

5 Covarianza y correlación

6 Hacia la regresión

7 Manos a la obra

Dos variables a la vez: covarianza

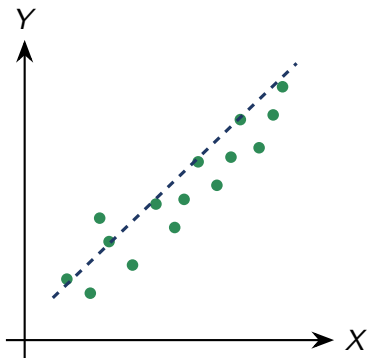
Definición 5 (Covarianza)

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)] \stackrel{\text{atajo}}{=} \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

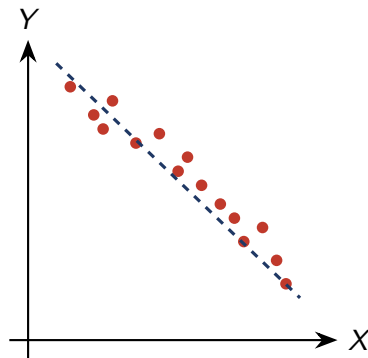
Intuición

- $\text{Cov} > 0$: cuando X está por encima de su media, Y también (suben juntas: educación e ingreso).
- $\text{Cov} < 0$: se mueven en sentidos opuestos (precio y cantidad demandada).
- $\text{Cov} = 0$: no hay tendencia *lineal* conjunta.

Dos nubes de dispersión (scatterplots)



$\text{Cov}(X, Y) > 0$
(positiva)



$\text{Cov}(X, Y) < 0$
(negativa)

Izquierda: X e Y suben juntas. Derecha: una sube y la otra baja.

Correlación: la covarianza estandarizada

Definición 6 (Correlación)

$$\text{Corr}(X, Y) = \rho_{XY} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}, \quad -1 \leq \rho_{XY} \leq 1 \text{ (sin unidades).}$$

Además $|\rho_{XY}| = 1$ si y solo si $Y = aX + b$ (relación lineal exacta).

Varianza de una suma (caso general)

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

Si $X \perp\!\!\!\perp Y$ entonces $\text{Cov} = 0$ y se reduce a $\text{Var}(X) + \text{Var}(Y)$.

Independencia y covarianza: ¡un solo sentido!

Teorema 2 (Independencia $\Rightarrow$ Cov = 0)

Si $X \perp\!\!\!\perp Y$, entonces $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ y por tanto $\text{Cov}(X, Y) = 0$.

¡Pero NO al revés!

$\text{Cov}(X, Y) = 0$ **no** implica independencia. La covarianza solo ve la relación **lineal**; dos variables pueden estar fuertemente ligadas de forma no lineal y tener covarianza cero.

Contraejemplo

$X \sim \text{Unif}[-1, 1]$, $Y = X^2$ (¡total dependencia!). Sin embargo
 $\text{Cov}(X, Y) = \mathbb{E}[X^3] - \mathbb{E}[X]\mathbb{E}[Y] = 0 - 0 = 0$.

Demo resuelto: Cov y Corr con datos

Problema

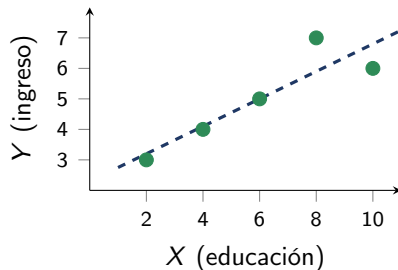
Cinco hogares, X = años de educación, Y = ingreso (miles):

X	2	4	6	8	10
Y	3	4	5	7	6

Se pide: $\text{Cov}(X, Y)$ y ρ_{XY} (poblacional, $\div n$).

Solución

$\bar{X} = 6$, $\bar{Y} = 5$; $\text{Cov} = \frac{1}{5}(8+2+0+4+4) = 3.6$. Con $\sigma_X^2 = 8$, $\sigma_Y^2 = 2$: $\rho_{XY} = \frac{3.6}{\sqrt{8}\sqrt{2}} = \frac{3.6}{4} = 0.9$.



Nube ascendente: X e Y suben juntas ($\rho = 0.9$).

Punteada: tendencia lineal.

Contenido

- 1 Funciones de variables aleatorias
- 2 Esperanza $\mathbb{E}[X]$
- 3 Varianza
- 4 Momentos
- 5 Covarianza y correlación
- 6 Hacia la regresión**
- 7 Manos a la obra

La mejor predicción lineal (best linear predictor)

Queremos **predecir** Y a **partir de** X con una recta $\alpha + \beta X$, minimizando el error cuadrático medio $\mathbb{E}[(Y - \alpha - \beta X)^2]$.

La solución

$$\beta = \frac{\text{Cov}(X, Y)}{\text{Var}(X)} = \rho_{XY} \frac{\sigma_Y}{\sigma_X}, \quad \alpha = \mu_Y - \beta \mu_X.$$

- La pendiente β es la covarianza reescalada por cuánto varía X .
- Si $\rho = 0$: la mejor predicción lineal es simplemente μ_Y (X no ayuda).
- Esto es, literalmente, el comienzo de la **econometría**.

Demo: educación $\rightarrow$ ingreso

Problema

Para X (años de educación) e Y (ingreso mensual, en soles) se estima $\mu_X = 10$, $\mu_Y = 1500$, $\sigma_X = 3$, $\sigma_Y = 900$ y $\rho_{XY} = 0.6$.

Se pide: la pendiente β y el intercepto α de la mejor predicción lineal $\hat{Y} = \alpha + \beta X$.

Solución: pendiente e intercepto

$$\beta = \rho_{XY} \frac{\sigma_Y}{\sigma_X} = 0.6 \cdot \frac{900}{3} = 180, \quad \alpha = \mu_Y - \beta \mu_X = 1500 - 180(10) = -300.$$

Interpretación

Cada año adicional de educación se *asocia* con unos 180 soles más de ingreso mensual.

Asociación, no necesariamente causalidad.

Contenido

- 1 Funciones de variables aleatorias
- 2 Esperanza $\mathbb{E}[X]$
- 3 Varianza
- 4 Momentos
- 5 Covarianza y correlación
- 6 Hacia la regresión
- 7 Manos a la obra**

¿Y ahora?

- Revisa el **PDF de teoría** para el detalle (Markov, Chebyshev y más ejemplos).
- Practica con el **PDF de ejercicios** (3 niveles: Basic · Core · Challenge).
- Verifica con el **PDF de soluciones** paso a paso.
- En **R**: `mean(x)`, `var(x)`, `sd(x)`, `cov(x,y)` y `cor(x,y)` hacen el trabajo numérico.

¡A practicar!