14.310x – Data Analysis for Social Scientists

MicroMasters en Statistics and Data Science (SDS) – MITx

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

Clase 4

Esperanza, Varianza y Momentos

Teoría

Módulos oficiales del MicroMaster:
M4 – Functions of Random Variables; Expectation, Variance & Intro to Regression (momentos, covarianza, correlación)

B.Sc. Cristian Lazo Quispe
clazoq@uni.pe
linkedin.com/in/cristian-lazo-quispe · github.com/CristianLazoQuispe

Material del *Advanced Program in Data Science & Global Skills*, diseñado y desarrollado por **BREIT (Brescia Institute of Technology)**, iniciativa de **Aporta** —plataforma de innovación e impacto social del **Grupo Breca**— en alianza con el **MIT IDSS** (Institute for Data, Systems, and Society).

11 de agosto de 2026

Índice

1. De qué trata esta clase

En la Clase 3 aprendimos a describir una variable aleatoria *por completo*: con su pmf, su pdf o su cdf. Pero una distribución contiene *muchísima* información, y a menudo no la necesitamos toda. Cuando un periodista pregunta "¿cuánto gana en promedio un hogar peruano?" o "¿qué tan desigual es el ingreso?", no quiere la densidad entera: quiere **un número** que resuma un rasgo de la distribución.

Esos números resumen son los **momentos (moments)**. El primero es la **esperanza (expectation)** $\mathbb{E}[X]$, que dice dónde está "centrada" la distribución. El segundo da pie a la **varianza (variance)** $\mathrm{Var}(X)$, que mide qué tan *dispersa* está. Con dos variables aparece la **covarianza (covariance)** y su versión estandarizada, la **correlación (correlation)**, que miden cómo se mueven juntas. Y todo esto desemboca, casi sin darnos cuenta, en la idea central de la econometría: la **regresión (regression)**.

Antes de los momentos veremos una herramienta de la Clase 3 llevada un paso más allá: cómo hallar la distribución (y luego la esperanza) de una **función de una variable aleatoria** $Y = g(X)$.

La distribución completa es como una fotografía en alta resolución. Los momentos son como decir "mide 1.70 m y pesa 68 kg": pierden detalle, pero resumen lo esencial y se comparan fácil. La esperanza dice *dónde* está el centro; la varianza, *qué tan ancha* es la nube de datos.

2. Funciones de variables aleatorias

Muchas cantidades de interés son *funciones* de una variable aleatoria: el logaritmo del ingreso $\log X$, el ingreso medido en dólares en vez de soles ($Y = X/3.7$), el cuadrado de una desviación, etc. Si X es aleatoria, $Y = g(X)$ también lo es, y queremos su distribución.

Método de la cdf (the cdf method)

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

1. Determina el **soporte inducido** de Y (qué valores puede tomar).
2. Escribe la cdf de Y como una probabilidad sobre X: $F_Y(y) = \mathbb{P}(Y \le y) = \mathbb{P}\Big(g(X) \le y\Big)$.
3. Resuelve la desigualdad $g(X) \le y$ para dejarla como un evento sobre X, y evalúalo con F_X.
4. Si Y es continua, deriva: $f_Y(y) = F_Y'(y)$.

Ejemplo 2.1 (Cambio de unidades: transformación lineal)**.** El ingreso mensual X (en soles) de cierta población tiene esperanza y densidad conocidas. Un organismo internacional quiere

el ingreso en dólares, $Y = aX$ con $a = 1/3.7$ (tipo de cambio). Como $a > 0$:

$$F_Y(y) = \mathbb{P}(aX \le y) = \mathbb{P}\Big(X \le \tfrac{y}{a}\Big) = F_X\Big(\tfrac{y}{a}\Big), \qquad f_Y(y) = \frac{1}{a}\, f_X\Big(\tfrac{y}{a}\Big).$$

Más adelante veremos que $\mathbb{E}[Y] = a\,\mathbb{E}[X]$: el ingreso *promedio* en dólares es, simplemente, el promedio en soles convertido al mismo tipo de cambio. (Caso general: $Y = aX + b$ da $f_Y(y) = \frac{1}{|a|} f_X\Big(\frac{y-b}{a}\Big)$.)

Ejemplo 2.2 (Función no monótona: $Y = X^2$)**.** Sea X una "desviación" centrada, uniforme en $[-1, 1]$, es decir $f_X(x) = \frac{1}{2}$ en $[-1, 1]$ (y $F_X(x) = \frac{x+1}{2}$ allí). Queremos la densidad de $Y = X^2$ (que vive en $[0, 1]$). El truco con una función *no monótona* es traducir bien el evento: $X^2 \le y \iff -\sqrt{y} \le X \le \sqrt{y}$. Así, para $0 \le y \le 1$,

$$F_Y(y) = \mathbb{P}\Big(-\sqrt{y} \le X \le \sqrt{y}\Big) = F_X(\sqrt{y}) - F_X(-\sqrt{y}) = \frac{\sqrt{y}+1}{2} - \frac{-\sqrt{y}+1}{2} = \sqrt{y}.$$

Derivando, $f_Y(y) = \dfrac{1}{2\sqrt{y}}$ para $0 < y \le 1$. Verificación: $\int_0^1 \frac{1}{2\sqrt{y}}\, dy = \Big[\sqrt{y}\Big]_0^1 = 1$ (✓).

Cuando g no es monótona (como x^2), dos valores distintos de X pueden dar el mismo Y. El error típico es escribir $\mathbb{P}(X^2 \le y) = \mathbb{P}(X \le \sqrt{y})$, olvidando la rama negativa. Traduce siempre la desigualdad *con cuidado* y dibújala en la recta.

3. Esperanza (expectation)

Definición 3.1 (Esperanza / valor esperado)**.** La **esperanza (expectation, expected value, mean)** de una variable aleatoria X, denotada $\mathbb{E}[X]$ o μ, es

$$\textbf{Discreta:}\quad \mathbb{E}[X] = \sum_x x\, f_X(x), \qquad\qquad \textbf{Continua:}\quad \mathbb{E}[X] = \int_{-\infty}^{\infty} x\, f_X(x)\, dx.$$

(Siempre que la suma/integral converja absolutamente.)

Si pensamos la pdf como una densidad física de masa repartida sobre la recta, la esperanza es el **punto de equilibrio (balancing point)**: el lugar donde la distribución "hace balanza". Por eso, en una distribución muy asimétrica (como el ingreso), la *media* se va hacia la cola larga, mientras que la *mediana* se queda en el centro de los datos.

Ejemplo 3.1 (Esperanza de una v.a. discreta: años de educación)**.** En una encuesta, el máximo nivel educativo X de un adulto (en años, agrupado) tiene pmf

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

La educación *promedio* es

$$\mathbb{E}[X] = 6(0.30) + 11(0.40) + 14(0.20) + 17(0.10) = 1.8 + 4.4 + 2.8 + 1.7 = 10.7 \text{ años}.$$

Ejemplo 3.2 (Esperanza de una v.a. continua: la exponencial)**.** El tiempo X (en meses) hasta que un desempleado encuentra trabajo se modela con $f_X(x) = \lambda e^{-\lambda x}$, $x \geq 0$. Integrando por partes,

$$\mathbb{E}[X] = \int_0^\infty x\, \lambda e^{-\lambda x}\, dx = \Big[-xe^{-\lambda x} \Big]_0^\infty + \int_0^\infty e^{-\lambda x}\, dx = 0 + \frac{1}{\lambda} = \frac{1}{\lambda}.$$

Si $\lambda = \frac{1}{4}$ por mes, el tiempo de espera promedio es $\mathbb{E}[X] = 1/\lambda = 4$ meses.

3.1. Propiedades de la esperanza

Linealidad de la esperanza (linearity of expectation)

Para constantes a, b y variables aleatorias X, Y (¡*no* hace falta independencia!):

$$\textbf{(1)}\ \mathbb{E}[a] = a, \qquad \textbf{(2)}\ \mathbb{E}[aX + b] = a\,\mathbb{E}[X] + b, \qquad \textbf{(3)}\ \mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y].$$

Más en general, $\mathbb{E}[a_1X_1 + \cdots + a_nX_n + b] = a_1\mathbb{E}[X_1] + \cdots + a_n\mathbb{E}[X_n] + b$. Y si $X \perp\!\!\!\perp Y$, además $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$ (esto último *sí* requiere independencia).

$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$ vale *siempre*, estén o no relacionadas X e Y. En cambio $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ **solo** vale bajo independencia. Confundir ambas es uno de los errores más comunes del curso.

Ley del estadístico inconsciente (LOTUS)

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

$$\textbf{Discreta:}\ \mathbb{E}[g(X)] = \sum_x g(x)\, f_X(x), \qquad \textbf{Continua:}\ \mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x)\, f_X(x)\, dx.$$

El nombre "ley del estadístico inconsciente" (*law of the unconscious statistician*, LOTUS) viene de que mucha gente la usa sin notar que es un teorema.

Ejemplo 3.3 (LOTUS y la paradoja de San Petersburgo)**.** Lanzo una moneda justa hasta que sale cara; si hacen falta X lanzamientos, te pago $g(X) = 2^X$ soles. Aquí X es geométrica con $\mathbb{P}(X = x) = (1/2)^x$. ¿Cuánto deberías pagar por jugar? Tu pago esperado, por LOTUS, es

$$\mathbb{E}[2^X] = \sum_{x=1}^{\infty} 2^x \left(\tfrac{1}{2}\right)^x = \sum_{x=1}^{\infty} 1 = \infty.$$

¡El pago esperado es infinito! Y sin embargo nadie pagaría ni 20 soles por jugar: esa es la *paradoja de San Petersburgo*. La salen los economistas con la *utilidad marginal decreciente del dinero*: si valoras tus ganancias como $\log(2^X)$, entonces $\mathbb{E}[\log(2^X)] = \log 2 \cdot \mathbb{E}[X] = 2\log 2 < \infty$, un número finito y razonable.

4. Varianza, desviación estándar y momentos

La esperanza dice dónde está el centro, pero no qué tan dispersa está la distribución. Dos países pueden tener el mismo ingreso promedio y, aun así, uno ser mucho más desigual. Para eso está la varianza.

Definición 4.1 (Varianza y desviación estándar)**.** La **varianza (variance)** de X, denotada $\mathrm{Var}(X)$ o σ^2, es la esperanza de la desviación al cuadrado respecto de la media $\mu = \mathbb{E}[X]$:

$$\mathrm{Var}(X) = \mathbb{E}\Big[(X-\mu)^2\Big].$$

La **desviación estándar (standard deviation)** es $\sigma = \mathrm{SD}(X) = \sqrt{\mathrm{Var}(X)}$; tiene las *mismas unidades* que X.

Fórmula de atajo para la varianza

Casi siempre conviene calcular la varianza así:

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

Es decir: "el promedio de los cuadrados menos el cuadrado del promedio". Como $\mathrm{Var}(X) \geq 0$, esto implica $\mathbb{E}[X^2] \geq (\mathbb{E}[X])^2$ siempre.

Propiedades de la varianza

(1) $\mathrm{Var}(X) \geq 0$, **(2)** $\mathrm{Var}(a) = 0$, **(3)** $\mathrm{Var}(aX+b) = a^2\,\mathrm{Var}(X)$.

La propiedad (3) dice: *sumar* una constante b desplaza la distribución pero **no** cambia su dispersión; *multiplicar* por a la estira, y la varianza cambia por el **cuadrado** del factor (la desviación estándar, por $|a|$).

Ejemplo 4.1 (Varianza con la fórmula de atajo)**.** Retomando la educación con pmf $f_X(6) = 0.30$, $f_X(11) = 0.40$, $f_X(14) = 0.20$, $f_X(17) = 0.10$ (ya vimos $\mathbb{E}[X] = 10.7$). El segundo momento es

$$\mathbb{E}[X^2] = 36(0.30) + 121(0.40) + 196(0.20) + 289(0.10) = 10.8 + 48.4 + 39.2 + 28.9 = 127.3.$$

Entonces

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

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

Definición 4.2 (Momentos (moments) de una distribución)**.** El **k-ésimo momento** de X (respecto del origen) es $\mathbb{E}[X^k]$. El **k-ésimo momento central** es $\mathbb{E}[(X-\mu)^k]$. Así:

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

Cada momento resume *un* rasgo de la forma. Con suficientes momentos podríamos casi reconstruir la distribución; pero en la práctica casi siempre nos basta con los dos primeros: media y varianza.

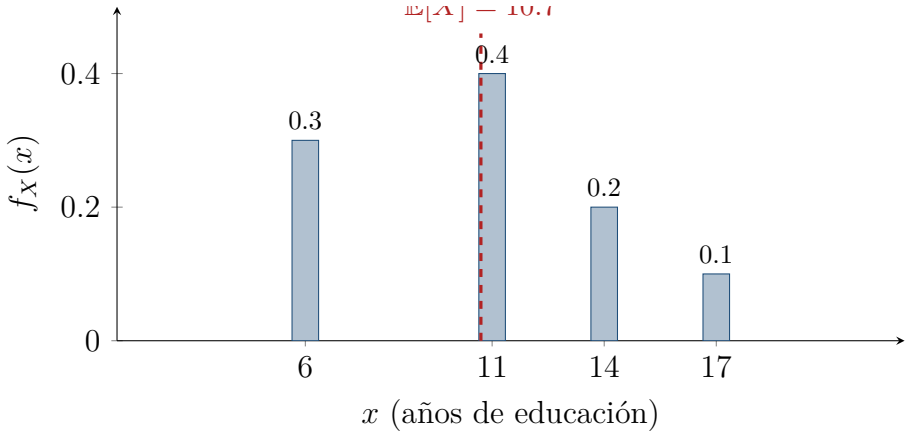

Figura 1: pmf de los años de educación X. La línea roja marca la *esperanza* $\mathbb{E}[X] = 10.7$, el "punto de equilibrio" de la distribución.

5. Esperanza y varianza de combinaciones

En economía rara vez miramos una sola variable: nos interesa el ingreso *total* de un hogar (suma de los ingresos de sus miembros), el gasto *conjunto*, una cartera de inversiones, etc. Necesitamos reglas para $\mathbb{E}$ y Var de *combinaciones*.

Esperanza de una combinación lineal

$$\mathbb{E}[a_1X_1 + a_2X_2 + \cdots + a_nX_n + b] = a_1\mathbb{E}[X_1] + a_2\mathbb{E}[X_2] + \cdots + a_n\mathbb{E}[X_n] + b.$$

Vale *siempre* (con o sin independencia). En particular $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$.

Varianza de una suma (caso independiente)

Si $X_1, \ldots, X_n$ son **independientes**,

$$\mathrm{Var}(X_1 + \cdots + X_n) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n),$$

$$\mathrm{Var}(a_1X_1 + \cdots + a_nX_n + b) = a_1^2\,\mathrm{Var}(X_1) + \cdots + a_n^2\,\mathrm{Var}(X_n).$$

Atención: a diferencia de la esperanza, la varianza de la suma *sí* necesita independencia. El caso general (con dependencia) lo veremos con la covarianza.

Ejemplo 5.1 (Ingreso total de una pareja)**.** Dos personas tienen ingresos X e Y con $\mathbb{E}[X] = 1\,800$, $\mathbb{E}[Y] = 1\,200$ soles, $\mathrm{Var}(X) = 400^2$, $\mathrm{Var}(Y) = 300^2$. Si sus ingresos fueran

independientes, el ingreso total $T = X + Y$ cumple

$$\mathbb{E}[T] = 1\,800 + 1\,200 = 3\,000, \qquad \mathrm{Var}(T) = 400^2 + 300^2 = 160\,000 + 90\,000 = 250\,000,$$

así que $\mathrm{SD}(T) = \sqrt{250\,000} = 500$ soles. (Si trabajan en el mismo sector y sus ingresos suben y bajan juntos, la varianza será *mayor*: ver la sección siguiente.)

6. Covarianza y correlación

La esperanza y la varianza describen *una* variable. Para dos variables a la vez, el momento que describe *cómo se mueven juntas* es la covarianza.

Definición 6.1 (Covarianza)**.** La **covarianza (covariance)** de X e Y, denotada $\mathrm{Cov}(X, Y)$ o σ_{XY}, es

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

La covarianza es positiva si, típicamente, cuando X está *por encima* de su media, Y también lo está (suben y bajan juntas: educación e ingreso). Es negativa si se mueven en sentidos opuestos (precio y cantidad demandada). Es cero si no hay tendencia lineal conjunta.

Propiedades de la covarianza

(1) $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$, **(2)** $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$,
(3) $\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$, **(4)** $X \perp\!\!\!\perp Y \;\Rightarrow\; \mathrm{Cov}(X, Y) = 0$,
(5) $\mathrm{Cov}(aX + b,\, cY + d) = ac\,\mathrm{Cov}(X, Y)$,
(6) $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X, Y)$.

La implicación (4) va en **un solo sentido**. Si X e Y son independientes, su covarianza es 0. Pero $\mathrm{Cov}(X, Y) = 0$ *no* garantiza independencia: la covarianza solo ve la relación **lineal**, y dos variables pueden estar fuertemente ligadas de forma no lineal y tener covarianza cero. (En la hoja de retos verás un contraejemplo explícito.)

Definición 6.2 (Correlación)**.** La **correlación (correlation)** es la covarianza *estandarizada*:

$$\mathrm{Corr}(X, Y) = \rho_{XY} = \frac{\mathrm{Cov}(X, Y)}{\sigma_X\,\sigma_Y}.$$

Cumple $-1 \le \rho_{XY} \le 1$ (sin unidades). Decimos que X e Y están *positivamente correlacionadas* si $\rho > 0$, *negativamente* si $\rho < 0$, y *no correlacionadas* si $\rho = 0$. Además $|\rho_{XY}| = 1$ *si y solo si* $Y = aX + b$ (relación lineal exacta).

Ejemplo 6.1 (Var de una suma con covarianza)**.** Volvamos a la pareja con $\mathrm{Var}(X) = 400^2$, $\mathrm{Var}(Y) = 300^2$, pero ahora sus ingresos están *positivamente* correlacionados: $\rho_{XY} = 0.5$. Entonces $\mathrm{Cov}(X, Y) = \rho\,\sigma_X\sigma_Y = 0.5 \cdot 400 \cdot 300 = 60\,000$, y

$$\mathrm{Var}(X + Y) = 400^2 + 300^2 + 2(60\,000) = 250\,000 + 120\,000 = 370\,000,$$

así que $\mathrm{SD}(X + Y) = \sqrt{370\,000} \approx 608$ soles, mayor que los 500 del caso independiente. **Cuando los riesgos se mueven juntos, el riesgo total crece**: esa es la idea detrás de *diversificar* una cartera.

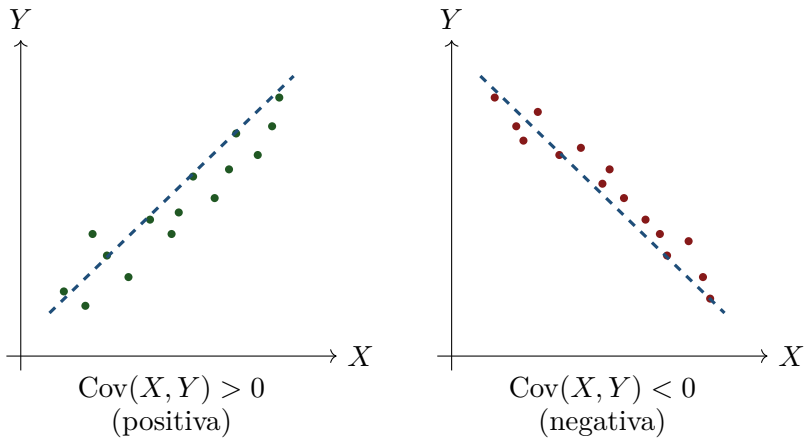

Figura 2: Nubes de dispersión (scatterplots). A la izquierda, X e Y tienden a subir juntas (covarianza/correlación *positiva*); a la derecha, cuando una sube la otra baja (covarianza/correlación *negativa*). La recta punteada sugiere la tendencia lineal.

7. Intro a la regresión

Aquí empieza, de verdad, la econometría. Tenemos dos variables, X (p.ej. años de educación) e Y (p.ej. ingreso), con medias μ_X, μ_Y, varianzas σ_X^2, σ_Y^2 y correlación ρ_{XY}. Queremos **predecir Y a partir de X** con una recta.

La mejor predicción lineal (best linear predictor)

Buscamos la recta $\alpha + \beta X$ que mejor predice Y, en el sentido de minimizar el **error cuadrático medio (mean squared error, ECM)** $\mathbb{E}\big[(Y - \alpha - \beta X)^2\big]$. La solución es

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

Con esa elección, el residuo $U = Y - \alpha - \beta X$ cumple $\mathbb{E}[U] = 0$ y $\mathrm{Cov}(X, U) = 0$. Llamamos a α, β *coeficientes de regresión*: $\alpha + \beta X$ es la parte de Y "explicada por" X, y U la parte "no explicada".

La pendiente β es, esencialmente, la covarianza reescalada por cuánto varía X. Si $\rho = \pm 1$, la recta predice Y a la perfección ($U = 0$). Si $\rho = 0$, la mejor predicción lineal es simplemente la media $\mathbb{E}[Y] = \mu_Y$: X no ayuda en nada (linealmente). La *esperanza condicional* $\mathbb{E}[Y \mid X]$ es la mejor predicción *de cualquier forma*; la regresión lineal es su mejor aproximación con una recta.

Ejemplo 7.1 (Pendiente de educación → ingreso)**.** Supongamos $\sigma_X = 3$ años, $\sigma_Y = 900$ soles y $\rho_{XY} = 0.6$. La pendiente de la mejor predicción lineal es

$$\beta = \rho_{XY}\frac{\sigma_Y}{\sigma_X} = 0.6 \cdot \frac{900}{3} = 180.$$

Interpretación: cada año adicional de educación se asocia, en promedio, con unos 180 soles más de ingreso mensual. (Asociación, *no* necesariamente causalidad: eso es tema de las clases de inferencia causal.)

Nota. También existen dos resultados muy útiles que conectan momentos con probabilidades de cola: la *desigualdad de Markov* ($\mathbb{P}(X \geq t) \leq \mathbb{E}[X]/t$ para $X \geq 0$, $t > 0$) y la *desigualdad de Chebyshev* ($\mathbb{P}(|X - \mu| \geq t) \leq \mathrm{Var}(X)/t^2$). Permiten acotar probabilidades sabiendo solo la media o la varianza.

8. En R: estimar esperanza y varianza por simulación

Monte Carlo para E(X) y Var(X)

```
set.seed(1)
x <- sample(1:6, 1e6, replace = TRUE)  # un dado justo
mean(x)     # E[X] ~ 3.5
var(x)      # Var(X) ~ 2.917  (teorico 35/12)
mean(x^2)   # E[X^2] por LOTUS ~ 15.17
```

La media y varianza muestrales se acercan a $\mathbb{E}[X] = 3.5$ y $\mathrm{Var}(X) = \frac{35}{12} \approx 2.917$.

Resumen de la clase

Lo esencial de la Clase 4

- **Función de una v.a.:** $Y = g(X)$ por el *método de la cdf*; $F_Y(y) = \mathbb{P}(g(X) \leq y)$, luego derivar. Cuidado con las g no monótonas.
- **Esperanza:** $\mathbb{E}[X] = \sum x f_X(x)$ (discreta) $= \int x f_X(x)\,dx$ (continua). Es el "punto de equilibrio".
- **Linealidad:** $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$ y $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$ *siempre*. **LOTUS:**

$\mathbb{E}[g(X)] = \int g(x) f_X(x)\, dx$.

- **Varianza:** $\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$; $\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X)$; $\sigma = \sqrt{\mathrm{Var}(X)}$ (mismas unidades).
- **Momentos:** $\mathbb{E}[X^k]$ (k-ésimo momento). El 1.^{o} localiza, el 2.^{o} (central) dispersa.
- **Covarianza:** $\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$; $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X, Y)$. **Correlación:** $\rho = \mathrm{Cov}\,/(\sigma_X \sigma_Y) \in [-1, 1]$.
- **Independencia** $\Rightarrow \mathrm{Cov} = 0$ (¡no al revés!).
- **Regresión:** la mejor predicción lineal tiene $\beta = \mathrm{Cov}(X, Y)/\,\mathrm{Var}(X)$, $\alpha = \mu_Y - \beta\mu_X$.

Fuentes y atribución

Bibliografía.

- Duflo, E. & Ellison, S. F. *14.310x: Data Analysis for Social Scientists*. MIT OpenCourseWare (slides, lecture notes y problem sets). ocw.mit.edu.
- Larsen, R. J. & Marx, M. L. *An Introduction to Mathematical Statistics and Its Applications*, 6.ª ed., Pearson, 2018.
- DeGroot, M. H. & Schervish, M. J. *Probability and Statistics*, 4.ª ed., Pearson, 2012.
- Blitzstein, J. & Hwang, J. *Introduction to Probability* (Harvard Stat 110); W. Chen, *Probability Cheatsheet*.
- Diez, D., Çetinkaya-Rundel, M. & Barr, C. *OpenIntro Statistics*. openintro.org.
- Materiales abiertos de la comunidad del curso en GitHub: `gevorg-hunanyan/probability`; `hanmochen/Probability-Theory-2020`; `mynameisjanus/14310xDataAnalysis`.