Variables Aleatorias I. Variables aleatorias discretas.

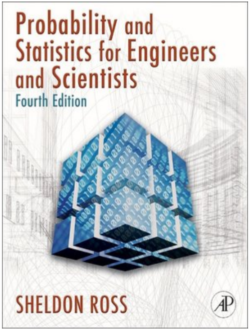
Estadística, Grado en Sistemas de Información

Constantino Antonio García Martínez

Universidad San Pablo Ceu

Contenidos

References



S. Ross. Introduction to Probability and Statistics for Engineers and Scientists. Chapter 4.

C.D. Barr, D.M. Diez, M. Çetinkaya-Rundel. OpenIntro Statistics. Chapters 2-3.

Introducción

Cuando realizamos un experimento, generalmente no nos interesan los detalles del mismo, sino solo un valor numérico determinado por el resultado.

Ejemplo: Variables aleatorias
Después de 20 lanzamientos de una moneda, podemos haber obtenido el resultado:

$$HHHHHHHTHHTTHHHHHHTT$$

Sin embargo, la información puede sintetizarse con números:

- Hay 5 cruces.
- La frecuencia de las caras es 0.75.
- La primera cruz ocurre en la posición 8.

En resumen, generalmente nos interesa **asociar a cada resultado de un experimento aleatorio, rasgos cuantitativamente medibles**.

Variable aleatoria
Una **variable aleatoria** X es una función que asocia un resultado de un experimento a un número:

$$X : \Omega \to \mathbb{R}.$$

Ejemplo: Variable aleatoria
Denotemos con X la variable aleatoria número de caras en 3 lanzamientos.

Variable aleatoria
Una **variable aleatoria** X es una función que asocia un resultado de un experimento a un número:

$$X : \Omega \to \mathbb{R}.$$

Ejemplo: Variable aleatoria
Denotemos con X la variable aleatoria número de caras en 3 lanzamientos.

$$\begin{aligned}
&X(HHH) = 3\\
&X(HHT) = X(HTH) = X(THH) = 2\\
&X(TTH) = X(THT) = X(HTT) = 1\\
&X(TTT) = 0
\end{aligned}$$

Variable aleatoria
Una **variable aleatoria** X es una función que asocia un resultado de un experimento a un número:

$$X : \Omega \to \mathbb{R}.$$

Ejemplo: Variable aleatoria
Denotemos con X la variable aleatoria número de caras en 3 lanzamientos.

$$\begin{aligned}
&X(HHH) = 3 \\
&X(HHT) = X(HTH) = X(THH) = 2 \\
&X(TTH) = X(THT) = X(HTT) = 1 \\
&X(TTT) = 0
\end{aligned}$$

Ejemplo: Variable aleatoria
Denotemos con X la variable aleatoria número de caras en 3 lanzamientos.
¿Cuál es la probabilidad de obtener dos caras?

Variable aleatoria

Una **variable aleatoria** X es una función que asocia un resultado de un experimento a un número:

$$X : \Omega \to \mathbb{R}.$$

Ejemplo: Variable aleatoria

Denotemos con X la variable aleatoria número de caras en 3 lanzamientos.

$$\begin{aligned}
&X(HHH) = 3 \\
&X(HHT) = X(HTH) = X(THH) = 2 \\
&X(TTH) = X(THT) = X(HTT) = 1 \\
&X(TTT) = 0
\end{aligned}$$

Ejemplo: Variable aleatoria

Denotemos con X la variable aleatoria número de caras en 3 lanzamientos.

¿Cuál es la probabilidad de obtener dos caras?

$$P(X = 2) = P(\{HHT, HTH, THH\}) = \frac{3}{8}$$

Además de permitirnos ocultar el espacio muestral, algunas de las ventajas de emplear **variables aleatorias** son...

1. Permiten caracterizar la **distribución de probabilidades** del evento de interés.
2. Nos permitirán movernos a **espacios muestrales no-numerables** (p.ej.: estudiar la altura de los europeos). (Parte 2.)
3. Nos permitirán resolver problemas **re-usando conocimiento**. (Parte 3.)

Distribuciones de variables aleatorias discretas

Distribución de una VA
LLamamos **función de probabilidad**, o **función de masa de probabilidad** a

$$P(X = x) = p(x) = f(x).$$

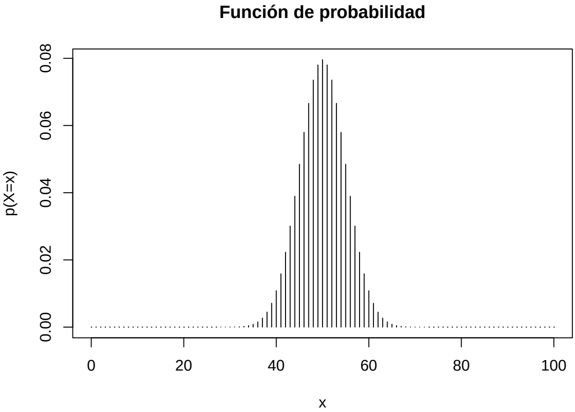
Siempre se cumple que

- $P(X = x) \geq 0$.
- $\sum_x P(X = x) = 1$.

Su cálculo permite caracterizar la **distribución de una VA discreta**.

Ejercicio: Función de probabilidad
Sea la VA X: "nº de caras en n lanzamientos de una moneda cuya probabilidad de cara es p". Estudia su distribución para el caso $p = 1/2$, $n = 100$ mediante la función de probabilidad.

Ejercicio: Función de probabilidad (continuación)



Una forma alternativa de caracterizar la **distribución** de una VA es emplear

distribución acumulada de probabilidad (o función de distribución)

$$F(x) = P(X \leq x).$$

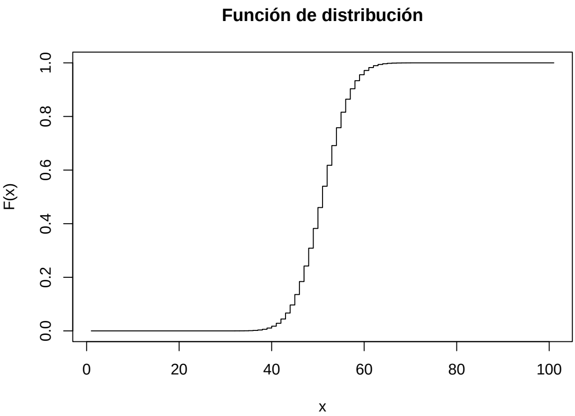
$F(x)$ es particularmente útil para responder a preguntas del tipo $P(a < X < b)$ (o similares).

Ejercicio: Función de distribución

Escribe una función de R para la función de distribución de la VA aleatoria X: "nº de caras en 100 lanzamientos de una moneda cuya probabilidad de cara es 0.5". Dibújala y úsala para responder a las siguientes preguntas:

a) $P(40 \leq X \leq 60)$
b) $P(40 < X < 60)$
c) $P(40 \leq X < 60)$

Ejercicio: Función de distribución (continuación)

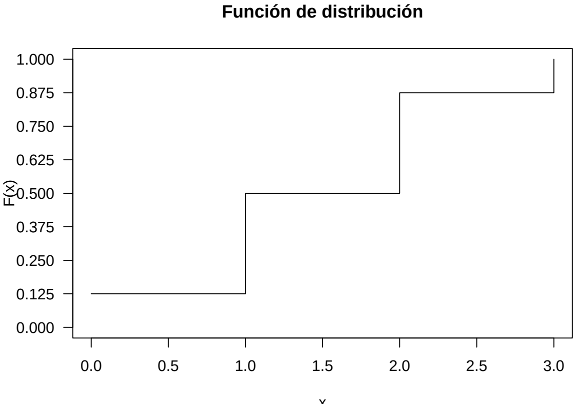


Es posible

- Hallar $F(x)$ a partir de $p(x)$,
- Hallar $p(x)$ a partir de $F(x)$.

Ejercicio: Función de distribución

Halla la función de probabilidad de X: "nº de caras en 3 lanzamientos de una moneda cuya probabilidad de cara es 0.5" a partir de la siguiente función de distribución:



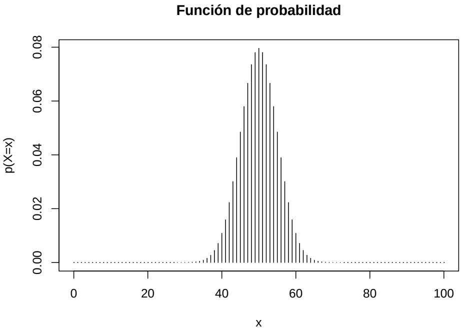
Esperanza, varianza y otros estadísticos resumen

Esperanza, varianza y otros estadísticos resumen

Medidas de tendencia central

Medidas de tendencia central

La distribución de una VA nos da información detallada acerca de todos los posibles resultados de un experimento. Sin embargo, a veces nos gustaría **resumir** la información.



Uno de los "resúmenes" más útiles es saber en torno a qué punto se concentran los resultados de un experimento aleatorio. Existen varias formas de medir la **tendencia central** de un experimento: 1) **Esperanza o media**, 2) **mediana**, 3) **moda**.

Moda, mediana y cuantiles

Moda
Es cualquier x tal que $p(x)$ tenga un **máximo local**. Una distribución puede tener 0, 1, 2, 3, etc. Hablamos de **distribuciones sin moda, unimodales, bimodales, trimodales, ...**

Mediana
Es el valor x que separa la mitad superior de la mitad inferior de todos los posibles resultados de X. De manera formal, la mediana es el valor más pequeño que verifica que

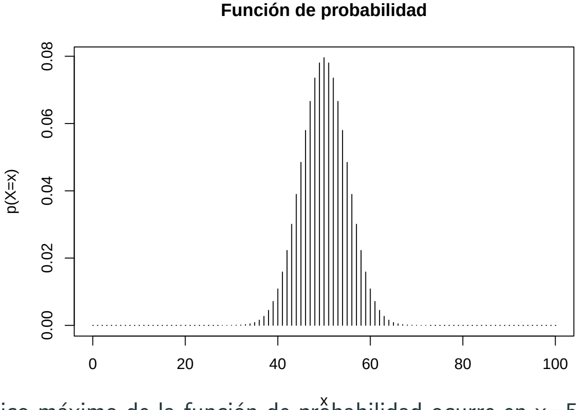
$$F(x) \geq 1/2.$$

Cuantiles
Una generalización de la mediana es el cuantil. El $\alpha-$cuantil es el valor x más pequeño que deja una probabilidad de α a su izquierda:

$$F(x) \geq \alpha. \qquad \text{(La mediana es el 50 \%-cuantil.)}$$

Ejemplo: Mediana y Moda

La mediana y la moda de X ("nº de caras en 100 lanzamientos de una moneda sin trucar") son fáciles de calcular por razonamiento:



- El único máximo de la función de probabilidad ocurre en x=50. Por tanto, 50 es la moda de X.
- La distribución es simétrica respecto a este máximo. Como deja igual probabilidad a izquierda y derecha, esta debe ser 0.5 en ambos casos. Por tanto, x=50 también es la mediana de X.

Ejercicio: Mediana y Moda
Realiza los cálculos anteriores explícitamente usando R.

Esperanza matemática
La **esperanza matemática** de una VA representa el **valor que esperamos obtener en media si el experimento aleatorio se repitiese indefinidamente**.

Definimos la esperanza para una variable aleatoria discreta X como

$$E[X] = \mu_x = \mu = \sum_i x_i P(X = x_i).$$

De cara a las simulaciones, el siguiente resultado será útil:

Ley de los grandes números
La media aritmética de n observaciones

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

converge a la esperanza $E[X]$ si $n \to \infty$.

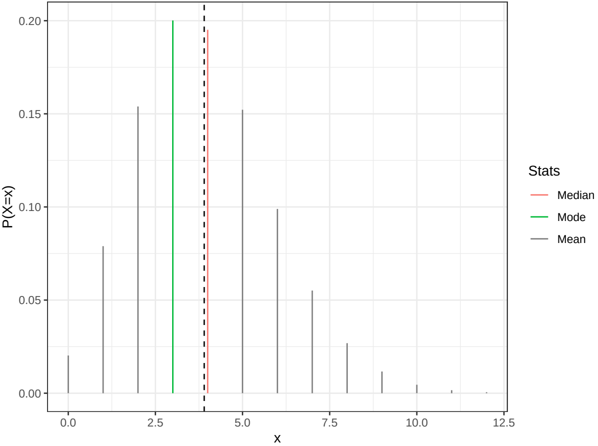
Generalmente llamamos **media muestral** a $\bar{X}$ y **media poblacional** a $E[X]$.

Ejercicio: Esperanza

Calcula la esperanza de la variable aleatoria X: "nº de caras en 100 lanzamientos de una moneda sin trucar" usando 1) la definición y 2) simulaciones. ¿Cuadra con tu intuición?

Esperanza, mediana y moda

¡Ojo! Aunque en el ejemplo de la moneda, media, moda y mediana tomen el mismo valor, en general esto no es cierto. Solo ocurre con distribuciones unimodales y simétricas.



Muchas veces, la esperanza permite tomar decisiones rápidas...

Ejercicio: Esperanza en la toma de decisiones
Un jugador gana 1 euro si al tirar un dado obtiene un 1 o un 3; pierde 2 euros si sale un 2, 4, 6; y gana 4 euros si sale un 5. ¿Cuál es la ganancia esperada? ¿Jugarías a este juego?

Propiedades de la esperanza

- $\mathbb{E}[cX] = c \cdot \mathbb{E}[X]$.
- $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$.
- Si X e Y son VAs independientes

$$\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y].$$

Ejercicio: Propiedades de la esperanza

Sea X la VA: “nº de caras en n lanzamientos si la probabilidad de cara es p”.

Calcula la esperanza de X usando las propiedades de la esperanza.

Ejemplo: Funciones de VAs

Si X se distribuye como

$$p(-2) = 0.15, p(0) = 0.2, p(1) = 0.5, p(2) = 0.15,$$

¿cuál es la esperanza de $Y = X^2$?

Ejemplo: Funciones de VAs
Si X se distribuye como

$$p(-2) = 0.15, p(0) = 0.2, p(1) = 0.5, p(2) = 0.15,$$

¿cuál es la esperanza de $Y = X^2$?

$$\begin{aligned}\mathbb{E}[Y] &= 0p(0) + 1p(1) + 4\left(p(-2) + p(2)\right) \\ &= 4p(-2) + 0p(0) + 1p(1) + 4p(2) \\ &= \sum_{x_i \in \{-2,0,1,2\}} x_i^2 P(X = x_i).\end{aligned}$$

El ejemplo anterior sugiere que podemos calcular la esperanza de Y usando la distribución de X. Esta propiedad es, de hecho, cierta (y a veces se llama la **ley del estadísta inconsciente**):

$$\mathbb{E}[g(X)] = \sum g(x)p(x)$$

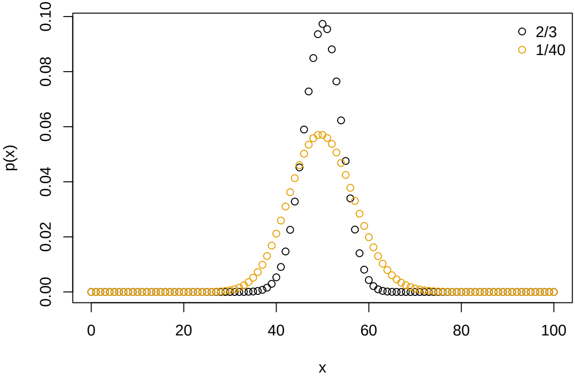
Esperanza, varianza y otros estadísticos resumen

Varianza

La esperanza es muy útil para "resumir" una VA, ya que $\mathbb{E}[X]$ nos permite saber cuál es la tendencia central de X. Pero a veces esto no es suficiente...

Ejemplo: Dispersiones distintas con igual media

Compara las distribuciones de las VAs X: "nº de caras en 100 lanzamientos si la probabilidad de cara es 2/3" e Y: "nº de caras en 2000 lanzamientos si la probabilidad de cara es 1/40".

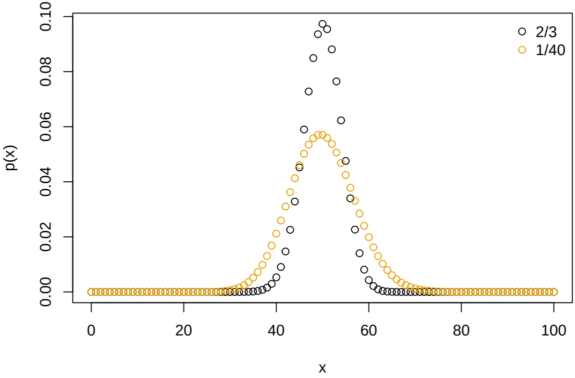


Varianza

La esperanza es muy útil para "resumir" una VA, ya que $\mathbb{E}[X]$ nos permite saber cuál es la tendencia central de X. Pero a veces esto no es suficiente...

Ejemplo: Dispersiones distintas con igual media

Compara las distribuciones de las VAs X: "nº de caras en 100 lanzamientos si la probabilidad de cara es $2/3$" e Y: "nº de caras en 2000 lanzamientos si la probabilidad de cara es $1/40$".



Lo que cambia de una distribución a otra es la **dispersión** de los valores en torno a su punto medio.

Medidas de dispersión

Para caracterizar la **dispersión** usamos la **varianza**:

$$\mathrm{Var}[X] = \sigma_x^2 = \sigma^2 = \mathbb{E}[(X-\mu)^2],$$

o bien la **desviación estándar** o **desviación típica**:

$$\sigma_x = \sigma = \sqrt{\mathrm{Var}[X]} = \sqrt{\mathbb{E}[(X-\mu)^2]},$$

Ejemplo: Varianza Vs. Desviación estándar

La desviación típica es generalmente más fácil de interpretar que la varianza. Para entenderlo, considera la variable aleatoria X: " altura de una persona en cm". La varianza de X se mediría en [cm^2] mientras que la desviación típica se mediría en [cm].

Ejercicio: Varianza y desviación típica

Calcula la varianza y desviación típica de la variable aleatoria X: "nº de caras en 100 lanzamientos de una moneda sin trucar" usando 1) la definición y 2) simulaciones. Visualiza la desviación típica sobre la función de probabilidad.

Propiedades de la varianza

- $\mathsf{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}[X^2] - \mu^2$.
- $\mathsf{Var}[cX] = c^2\mathsf{Var}[X]$.
- Si X e Y son independientes:

$$\mathsf{Var}[X + Y] = \mathsf{Var}[X] + \mathsf{Var}[Y], \tag{1}$$

$$\mathsf{Var}[X - Y] = \mathsf{Var}[X] + \mathsf{Var}[Y]. \tag{2}$$

Distribuciones conjuntas de variables discretas

Distribuciones conjuntas de variables discretas

Distribuciones conjuntas de variables discretas

Distribuciones conjuntas de variables discretas

Para especificar la relación entre dos VAs discretas X e Y usamos la **distribución de probabilidad conjunta** $P(X = x, Y = y) = p(x, y) = f(x, y)$. Se verifica:

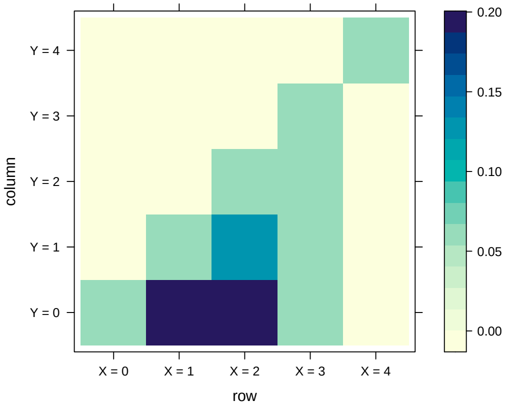
1. $p(x, y) \geq 0$.
2. $\sum_x \sum_y p(x, y) = 1$.

Ejercicio: Distribuciones conjuntas

Se lanza una moneda n veces (prob. de cara es p). Considera las VAs X: "nº de caras" e Y: "nº de caras iniciales (antes de la primera cruz o del fin del experimento)". Halla la distribución conjunta para cualquier n y p y luego particulariza para $n = 4, p = 0.5$.

```
##        Y = 0  Y = 1  Y = 2  Y = 3  Y = 4
## X = 0 0.0625 0.0000 0.0000 0.0000 0.0000
## X = 1 0.1875 0.0625 0.0000 0.0000 0.0000
## X = 2 0.1875 0.1250 0.0625 0.0000 0.0000
## X = 3 0.0625 0.0625 0.0625 0.0625 0.0000
## X = 4 0.0000 0.0000 0.0000 0.0000 0.0625
```

Ejercicio: Distribuciones conjuntas (continuación)



Distribuciones conjuntos: marginalización

El conocimiento de la probabilidad conjunta nos permite calcular cualquiera de las **distribuciones marginales**. Por ejemplo, para calcular la distribución marginal de X:

$$P(X = x) = p(x) = \sum_y p(x, y).$$

Ejercicio: Marginalización

Partiendo de la función de probabilidad conjunta del ejemplo anterior, ($n = 4$ $p = 0.5$), calcula las funciones de probabilidad marginales para X (nº de caras) e Y (nº de caras iniciales).

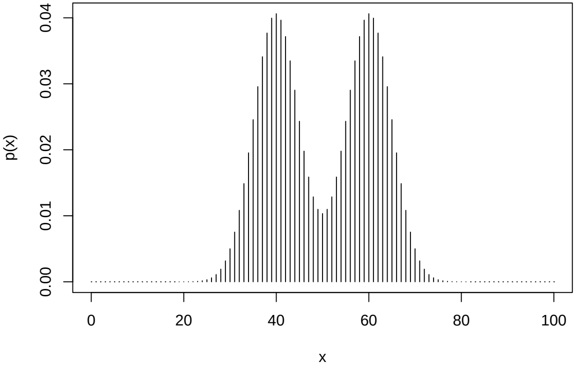
	Y = 0	Y = 1	Y = 2	Y = 3	Y = 4
X = 0	0.0625	0.0000	0.0000	0.0000	0.0000
X = 1	0.1875	0.0625	0.0000	0.0000	0.0000
X = 2	0.1875	0.1250	0.0625	0.0000	0.0000
X = 3	0.0625	0.0625	0.0625	0.0625	0.0000
X = 4	0.0000	0.0000	0.0000	0.0000	0.0625

Distribuciones conjuntas de variables discretas

Distribuciones condicionales de variables discretas

Ejercicio: Distribuciones condicionales

En una urna hay dos monedas trucadas con probabilidad de cara $p_0 = 0.4$ y $p_1 = 0.6$. Se elige una al azar y se tira 100 veces. Sea X: "nº de caras obtenidas" e Y: "moneda elegida". Obtener la función de probabilidad de X.



Ejemplo: Media, moda y distribuciones bimodales

De nuevo, podemos obtener la media, moda y mediana del problema anterior por razonamiento. Por simetría, la media y la mediana son 50, mientras que hay dos modas: 40 y 60. ¿Qué estadístico es más adecuado para estudiar **distribuciones bimodales**?

Ejercicio: Distribuciones condicionales

Si se han obtenido X=48 caras, ¿cuál es la probabilidad de que la moneda usada sea la correspondiente a p_0?

Distribuciones conjuntas de variables discretas

Estadísticos de variables aleatorias conjuntas

El **teorema del estadística inconsciente** se puede generalizar a más de una VA, P.ej.:

$$\mathbb{E}[g(X,Y)] = \sum_{x,y} g(x,y)p(x,y)$$

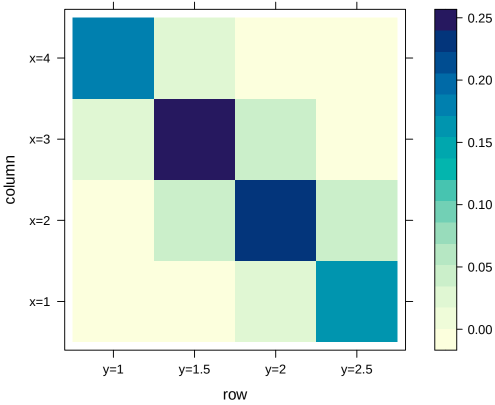
Ejercicio:
Sea X: "Cantidad mensual de lotes comprados por una empresa a su proveedor" e Y: "Precio por lote ofertado por el proveedor (en miles de euros)". La distribución conjunta de ambas variables se recoge en la siguiente tabla:

	x=1	x=2	x=3	x=4
y=1	0.00	0.00	0.03	0.18
y=1.5	0.00	0.04	0.24	0.02
y=2	0.02	0.23	0.04	0.00
y=2.5	0.16	0.04	0.00	0.00

¿Cuál es el coste esperado para la empresa en el siguiente mes de actividad?

Ejemplo: Covarianza de dos variables aleatorias

En el anterior ejemplo, el proveedor hace mejores ofertas cuanto más producto compre la empresa:



Es decir, parece que las variables aleatorias varían siguiendo alguna regla (**co-varían**). ¿Cómo medir este tipo de covariaciones?

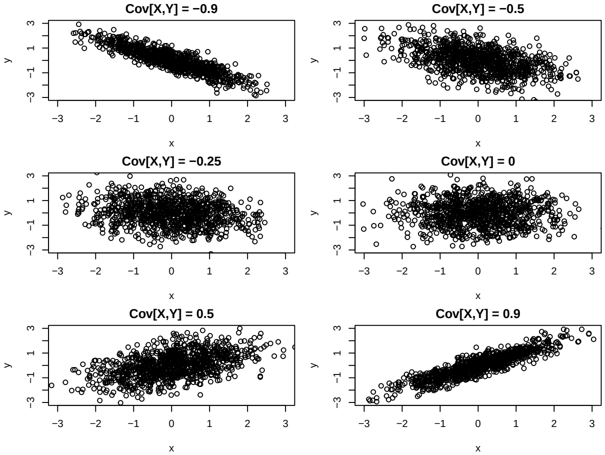
Covarianza
Definimos **la covarianza** para estudiar como varía X con Y:

$$\sigma_{xy} = \mathrm{Cov}[X, Y] = \mathbb{E}[(X - \mu_x)(Y - \mu_y)].$$

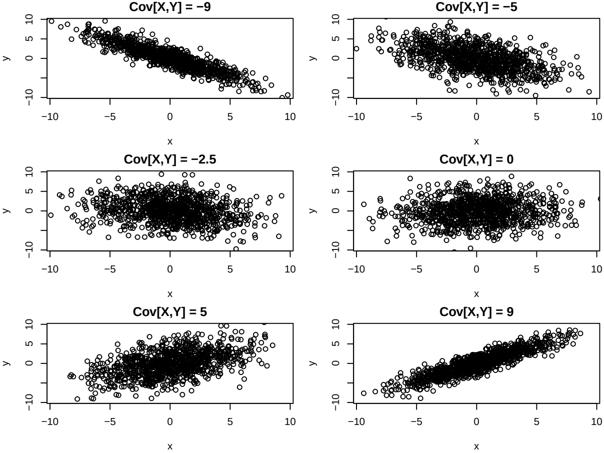
La covarianza tiene dos pequeños problemas:

- **Aumente con la varianza de X e Y**.
- Sólo captura **relaciones lineales**.

Covarianza: dependencia con la varianza



Covarianza: dependencia con la varianza



Correlación
La **correlación (o correlación de Pearson)** es una pequeña modificación de la covarianza para hacerla independiente de las escalas de X e Y:

$$\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y}.$$

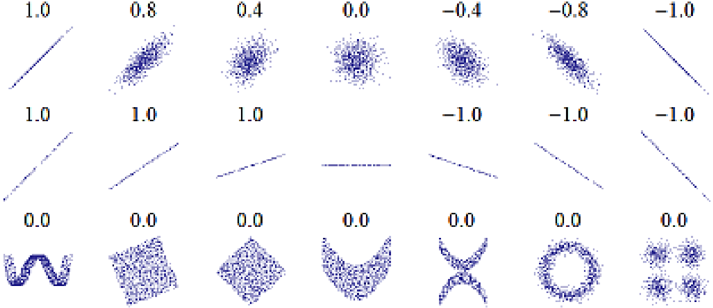
De esta forma,

$$-1 < \rho < 1.$$

Si $\rho = 0$, decimos que las variables están **incorreladas**.

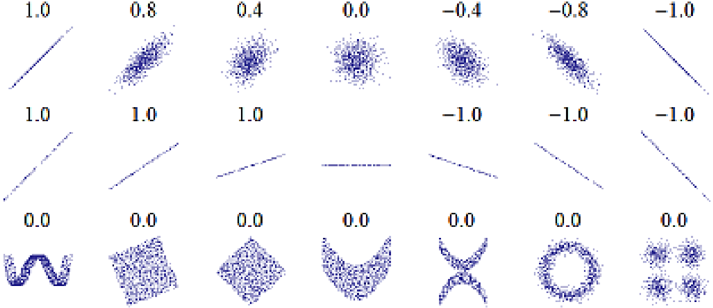
En general, **la correlación es una medida de la dependencia lineal entre X e Y**.

La correlación **solo captura relaciones lineales**.



Correlación: solo para dependencias lineales

La correlación **solo captura relaciones lineales**.



Por tanto, $\rho = 0$ **NO implica independencia**.

Propiedades de la covarianza/correlación

1. $\sigma_{xy} = \mathrm{Cov}[X, Y] = \mathbb{E}[XY] - \mu_x \mu_y$.
2. Si X e Y son independientes entonces

$$\sigma_{xy} = 0 \qquad \text{y} \qquad \rho = 0.$$

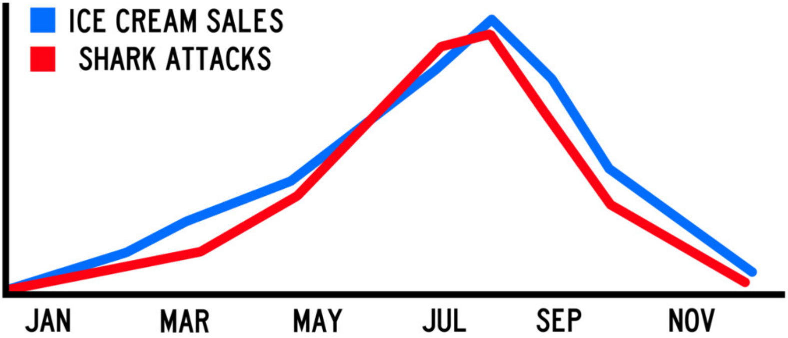
3. $\mathrm{Var}[X \pm Y] = \sigma_x^2 + \sigma_y^2 \pm 2\sigma_{xy}$.
4. $|\sigma_{xy}| \leq \sigma_x \sigma_y$.
5. $-1 < \rho < 1$.

Ejercicio:
Calcula la correlación entre X e Y en el problema de la empresa y el proveedor.

Ejercicio:
Sean D_1 y D_2 los resultados de tirar el dado 1 y el dado 2. Sean $X = D_1 + D_2$ e $Y = D_1 - D_2$. Demuestra que X e Y están incorreladas pero que no son independientes.

¡Correlación no implica causalidad!

CORRELATION IS NOT CAUSATION!



Both ice cream sales and shark attacks increase when the weather is hot and sunny, but they are not caused by each other (they are caused by good weather, with lots of people at the beach, both eating ice cream

Correlación no implica causalidad

¡Correlación no implica causalidad!