Probabilidad y Estadística

Ing. Remigio Hurtado Ortiz, PhD.

2023

Probabilidad y Estadística

1. Estadística
2. Teoría de la Probabilidad

Estadística Descriptiva

Las **estadísticas** son la recopilación y el análisis de datos mediante técnicas matemáticas.
Las **medidas descriptivas** nos permiten obtener información sobre la distribución de variables

Media: es el valor medio que una variable puede tomar

Varianza: es una medida de lo dispersa que es la distribución de x, o cuan es x, o cuanto varía x

Desviación estándar: es la raíz cuadrada de la varianza

	Varianza	Desviación Estándar	Media
Población	$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}$	$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}}$	$\mu = \frac{\sum_{i=1}^{N}x_i}{N}$
Muestra	$s^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$	$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}}$	$\bar{x} = \frac{\sum_{i=1}^{n}x_i}{n}$

Covarianza: determina la relación entre dos variables

$$Cov(X,Y) = \frac{\sum_1^n (x_i-\bar{x})(y_i-\bar{y})}{n}$$

Varianza, covarianza y correlación

La <u>**Correlación (coeficiente de correlación)** es una medida del grado de existencia de una relación lineal entre X e Y.</u>

$$corr(X,Y) = \frac{cov(X,Y)}{Sd(X)Sd(Y)} = \frac{cov(X,Y)}{\sqrt{var(X)var(Y)}}$$

$$-1 \leq corr(X,Y) \leq 1$$

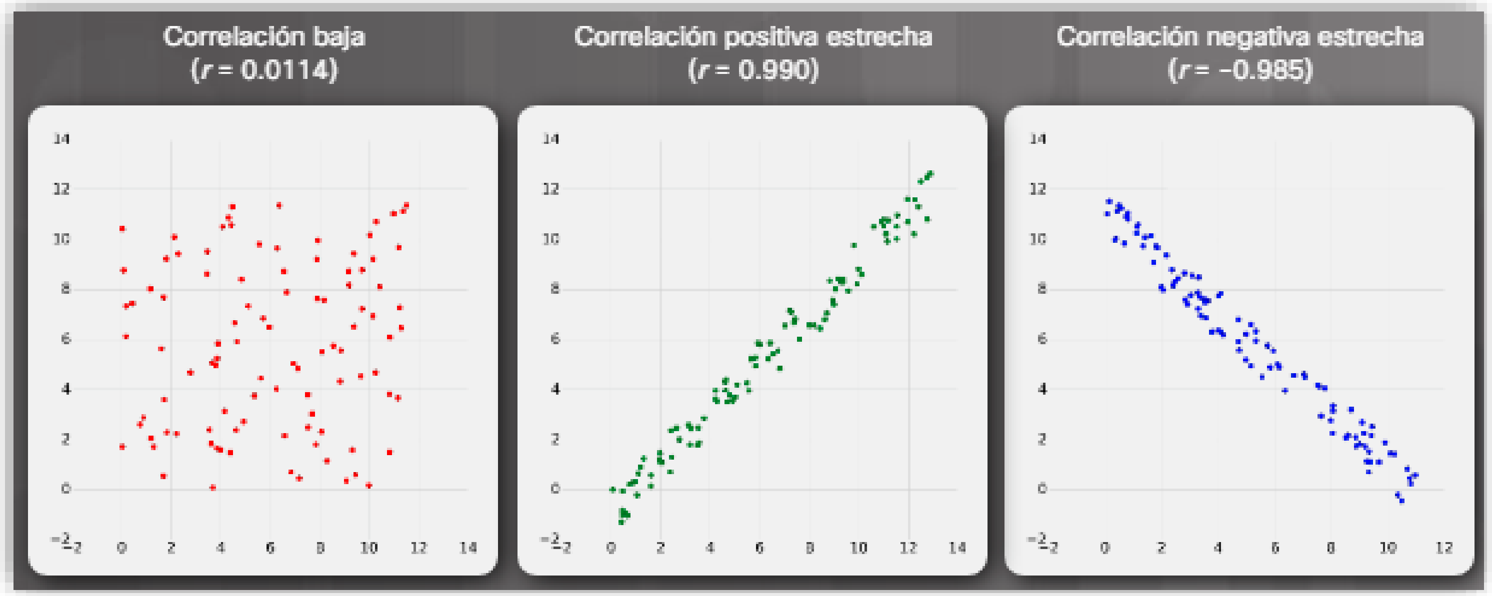
$corr(X,Y)$ =1 → Y aumenta X aumenta

$corr(X,Y) = -1$ → Z disminuye X aumenta

Recursos para análisis de correlación y varianza: https://rpsychologist.com/d3/correlation/

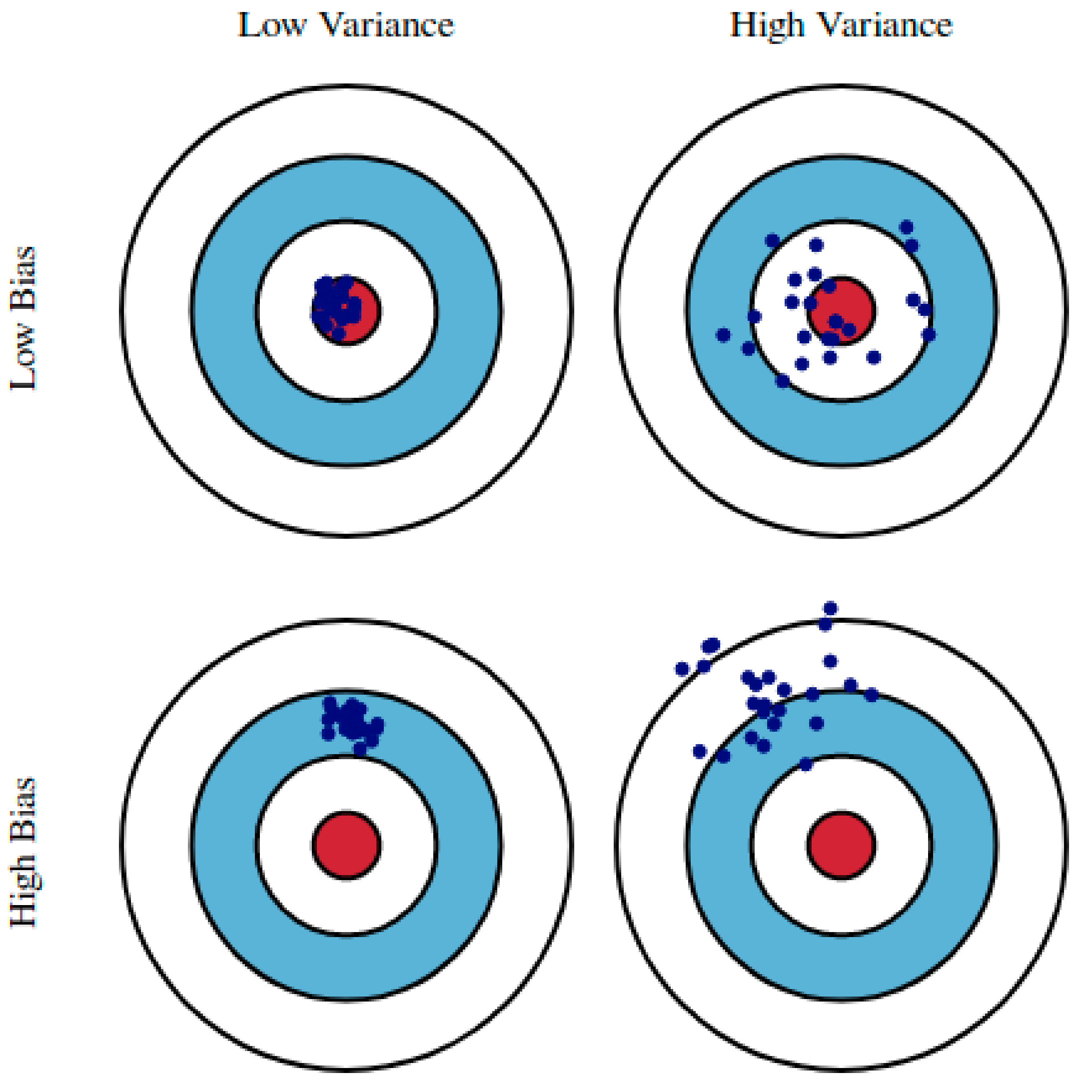
Mapa de Calor

Var 1	Var 1									
Var 2	0.385347	Var 2								
Var 3	0.980488	-0.041384	Var 3							
Var 4	0.334169	0.773025	0.252985	Var 4						
Var 5	0.148042	0.704236	0.631272	0.935709	Var 5					
Var 6	0.738391	-0.352484	0.891105	-0.697668	0.060915	Var 6				
Var 7	0.323975	0.611637	0.071302	0.358152	0.355059	0.113838	Var 7			
Var 8	0.056357	0.107547	0.739079	0.718235	0.301726	0.119437	0.018915	Var 8		
Var 9	0.211220	0.846530	0.284073	0.312393	0.542448	-0.434659	-0.428654	-0.343555	Var 9	
Var 10	-0.305630	0.678628	-0.685378	0.677347	0.418519	0.291790	0.289261	0.731888	0.179380	Var 10



Análisis de medidas

Sesgo (bias) y varianza



Teoría de la Probabilidad

La probabilidad es la ciencia de la incertidumbre. Da reglas matemáticas precisas que permiten comprender y analizar nuestra propia ignorancia. Ese conocimiento nos ayuda a hacer predicciones, tomar decisiones, valorar los riesgos e incluso ganar dinero.

Modelo de probabilidad

1. **S**: es el espacio muestral (conjunto no vacío) S= {lluvia, nieve, despejado}

2. Conjunto de **sucesos**: subconjuntos de S. Se les puede asignar una probabilidad.

Subconjuntos: {lluvia, nieve} {lluvia} Ø={} es vacío

3. **P**: medida de probabilidad. Probabilidad a cada suceso A. P(A).

P(A)= número real no negativo entre 0 y 1

P(Ø)=0

P(S)=1

P es aditiva -> $P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \ldots$

P([a-b])=b-a siempre que $0 \leq a \leq b \leq 1$. La probabilidad del intervalo es simplemente su longitud -> Distribución Uniforme en [0,1]

Teoría de la Probabilidad

Diagramas de Venn y subconjuntos: proporcionan un método gráfico muy útil para representar el espacio muestral y subconjuntos del mismo.

Probabilidad Uniforme

S si es finito, una posible medida de probabilidad es la probabilidad uniforme.

La probabilidad uniforme asigna una probabilidad $1/|S|$ a cada resultado. $|S|$ es el número de elementos de S o cardinalidad de S.

Por aditividad: $P(A)= |A|/|S|$ Se requiere determinar los tamaños de los conjuntos A y S (mediante principios de combinatoria: principio del producto, enumeración de permutaciones, enumeración de subconjuntos)

Ejemplos:

1 Moneda: S = {cara, cruz} $|S|=2$ P({cara})=P({cruz})=1/2

3 Monedas: S = {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT} $|S|=8$

P({HHH})=1/8 P({HHH,TTT})=2/8=1/4

Exactamente 2 caras= P({HHT,HTH,THH})=1/8 + 1/8 + 1/8 = 3/8

Probabilidad condicional e independencia

S = {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT} P(s)=1/8 para todo s ϵ S

¿Probabilidad de que la primera moneda sea cara?

P ({HHH,HHT,HTH,HTT})=4/8 =1/2 =0.5

Información extra = Condición = sabemos que 2 de 3 monedas han salido cara. Esto cambia nuestra información disponible, es decir nuestro grado de ignorancia. Cambian las probabilidades.

P({HHT,HTH,THH})=1 P(s)=1/3

P(cara en la primera moneda | cara en dos monedas) = P({HHT,HTH})=1/3+1/3=2/3

P(suceso | condición) -> | significa "bajo la condición" o "dado que"

P(A|B) = representa la fracción de veces que ocurre A sabiendo que ha ocurrido B.

P(A|B) = $\frac{P(A \cap B)}{P(B)}$ El cociente entre la probabilidad que ocurra A y B (ambos) y la prob. que ocurra B.

P(A|B) = P({HHT,HTH}) / P({HHT,HTH,THH}) = $\frac{2/8}{3/8}$ = 2/3 = 0.666 (AUMENTA LA PROBABILIDAD)

¿Cruz en la primera moneda (A) dado que 2 monedas han salido cara (B)?

P(A|B) = P({THH}) / P({HHT,HTH,THH}) = $\frac{1/8}{3/8}$ = 1/3 = 0.333 (DISMINUYE LA PROBABILIDAD)

Probabilidad condicional e independencia

Fórmula del producto P(A|B) = $\frac{P(A \cap B)}{P(B)}$ -> $P(A \cap B)$ = P(B) P(A|B) = P(A) P(B|A)

Ley de la probabilidad total

P(B) = $P(A_1)\, P(B|A_1) + P(A_2)\, P(B|A_2) + \ldots$

P(B) = $P(A_1 \cap B) + P(A_2 \cap B) + \cdots$

Ejemplo: A1=60% mujeres, A2=40% hombres, B=PELO LARGO: 30% mujeres, 20% hombres

P(B)= $P(A_1)\, P(B|A_1) + P(A_2)\, P(B|A_2)$ = (0.6)(0.3) + (0.4)(0.2) = 0.26 = 26%

Teorema de Bayes: cuando se conoce P(A), P(B) y P(B|A), se desea conocer P(A|B)

$$P(A|B) = \frac{P(A)}{P(B)}\, P(B|A)$$

Sistemas de 2 etapas (ver ejemplo 1.5.2 Pag 42 [1]):

Fórmula del producto: para calcular probabilidades conjuntas de ambas etapas

Ley de la probabilidad total: para calcular probabilidades de la segunda etapa.

Teorema de Bayes: para calcular probabilidades de la primera etapa. El teorema de Bayes se utiliza para revisar probabilidades previamente calculadas cuando se posee nueva información. Desarrollado por Thomas Bayes en el siglo XVII, el teorema de Bayes es una extensión de la probabilidad condicional.

¿Cual es la probabilidad que la bolita del segundo buzón sea roja, si la bolita del primer buzón salió negra?

Sucesos independientes: si $P(A \cap B) = P(A)P(B)$.

Sucesos cuya ocurrencia no afecta la probabilidad de los otros. Es decir, P(A|B)=P(A) y P(B|A)=P(B)

Teorema de Bayes: utilidades

Este teorema es de mucha utilidad, ya que gracias a él podemos relacionar la probabilidad de que un evento A ocurra sabiendo que ocurrió B, con la probabilidad de que ocurra lo contrario, es decir, que ocurra B dado A.

Por ejemplo en el estudio de enfermedades, el teorema de Bayes puede ayudar a discernir la probabilidad de que una enfermedad sea encontrada en un grupo de personas con una característica dada, tomando como datos las tasas globales de la enfermedad y el predominio de dicha característica en personas tanto sanas como enfermas.

Por otro lado, en el mundo de las altas tecnologías ha influenciado a grandes compañías ha desarrollar software "Basados en el Conocimiento". Como ejemplo cotidiano tenemos el asistente de Microsoft Office. El teorema de Bayes ayuda al software a evaluar los problemas que presenta el usuario y determinar qué consejo proporcionarle y así poder ofrecer un mejor servicio según los hábitos del usuario.

Teorema de Bayes

Sean $A_1, A_2, A_3, \ldots, A_n$ eventos mutuamente excluyentes y exhaustivos, con $P(A_i) \neq 0$ para cada A_i. Sea B cualquier evento con $P(B) \neq 0$, entonces:

$$P(A_i|B) = \frac{P(A_i) \cdot P(B|A_i)}{\sum_{j=1}^{n} P(A_j) \cdot P(B|A_j)}$$

Variables aleatorias

Existen formas más sencillas de representar una asignación particular de probabilidad que P. Formas más convenientes para trabajar con P.

Una variable aleatoria asigna un valor numérico a cada posible resultado s de S.

S={lluvia, nieve, despejado} Variable aleatoria X

X=3 si llueve, X=6 si nieva, X=-2,7 si esta despejado

Una variable aleatoria es una función definida sobre el espacio muestral S que asigna valores del conjunto R^1 de todos los números reales.

Dado que las variables aleatorias se definen como funciones de un resultado s y dado también que el resultado s se supone aleatorio (es decir, toma distintos valores con diferentes probabilidades), se sigue que el valor de una variable aleatorio será a su vez aleatorio (como su nombre implica).

Distribuciones de variables aleatorias

Si X es una variable aleatoria, ¿cuál es la probabilidad de que X tome un valor concreto x?

X=x cuando el resultado se escoge de modo que X(s)=x

Ejemplo: X=3, X=6, X=-2.7

- P(X=3)=P(lluvia)=**0.4**, P(X=6)=P(nieve)=**0.15**, P(X=-2.7)=P(despejado)=**0.45**
- P(X ϵ {3,6})= P(X=3) + P(X=6) = 0.4+0.15 = 0.55
- P(X<5)= P(X=3) + P(-2.7) = 0.4+0.45 = 0.85

La distribución de una variable aleatoria X es el conjunto de probabilidades P(X ϵ B) de X pertenecientes a diversos conjuntos.

Ejemplo: ¿Cuál es la distribución de X?

P(x ϵ B) 0.4 si 3 ϵ B, 0.15 si 6 ϵ B, 0.45 si -2.7 ϵ B

P(x ϵ B)= 0.4 I_B (3) += 0.15 I_B(6) += 0.45 I_B (-2.7) ← La Distribución $I_B(x)=1$ si $x \in B$, $I_B(x)=0$ si $x \notin B$

→Resulta engorrosa la forma de obtener la distribución para todos los subconjuntos B. Existen funciones más sencillas para obtener una distribución de probabilidad: funciones de distribución acumulada, funciones de probabilidad y funciones de densidad de probabilidad.

Distribuciones discretas

Una variable aleatoria es discreta si $\sum_x \boldsymbol{P(X = x) = 1}$, es decir si sus probabilidades son iguales a determinados valores.

Principales distribuciones: Uniforme discreta, Degenerada, **Bernoulli, Binomial, Geométrica**, Binomial negativa, **Poisson**, **Hipergeométrica, Multinomial**, Polya, De dos puntos.

Toda la información sobre la distribución de X esta contenida en su función de probabilidad, siempre y cuando sepamos que X es una variable aleatoria discreta.

Distribuciones continuas

Una variable aleatoria es continua si **P(X=x)=0**. Ninguna de las probabilidades puede tener un valor predeterminado para un valor de la variable.

Distribución Uniforme [0,1] P(a $\leq$ X $\leq$ b)=b-a siempre que a=0 y b=1 P(X<0)=P(X>1)=0

$f(x) = \begin{cases} 1, & 0 \leq x \leq 1 \\ 0, & otro\ caso \end{cases}$ $\mathrm{P(a \leq X \leq b)} = \int_a^b f(x)dx$ → **Ventaja:** Modificando la función se obtienen otras distribuciones continuas.

$f(x)$ es una función densidad si $f(x) \geq 0$ para todo x $\underline{\epsilon}$ R1 y $\int_{-\infty}^{\infty} f(x)dx = 1$

Principales distribuciones: Uniforme continua, **Exponencial**, **Gamma**, Normal, T de Student, Chi-cuadrada, F de Fisher, Erlang, Cauchy, Beta, Laplace, Log-normal, Rayleigh, Weibull, Maxwell, Valor extremo.

Teoría de la Probabilidad

Distribuciones discretas

Bernoulli: éxito o fracaso

Binomial: cantidad de éxitos (dos categorías o resultados)

Geométrica: cantidad de eventos para éxito. Numero de fracasos para el éxito.

Binomial negativa: cantidad de pruebas para conseguir un número determinado de resultados favorables (por primera vez)

Poisson: Probabilidad de que ocurra un determinado número de eventos durante cierto período de tiempo.

Hipergeométrica: el número de eventos en una muestra de tamaño fijo cuando usted conoce el número total de elementos en la población de la cual proviene la muestra

Multinomial: generalización de binomial (tres o más categorías o resultados). Casos de manera simultánea.

Distribuciones continuas

Uniforme continua: todos los intervalos de igual longitud en la distribución en su rango son igualmente probables.

Exponencial: tiempos de espera para la ocurrencia de un cierto evento. Tiempo hasta fallar (tiempo de vida). Probabilidad de funcionar hasta tal tiempo.

Gamma: distribución exponencial y distribución de Erlang son casos particulares de gamma. Gamma modela el comportamiento de variables aleatorias continuas con asimetría positiva y/o los experimentos en donde está involucrado el tiempo.

Erlang: para describir el tiempo de espera hasta el suceso número n en un proceso de Poisson. cuando es usada para medir el tiempo entre llamadas entrantes, se puede utilizar junto con la duración esperada de las mismas para producir información sobre la carga de tráfico.

Valor esperado

El valor esperado de una variable aleatoria es el <u>valor medio</u> que esta variable puede tomar.

Ejemplo:

- si la mitad de veces X=0, y la otra mitad X=10, E(X)=5
- si 1/3 Y=6, 2/3 Y=15, entonces (1/3)(6) + (2/3)(15)=2+10=12   E(Y)=12

Caso Discreto: E(X)= μ_x = $\sum_{x \epsilon R^1} x\, P(X = x)$ = $\sum_{x \epsilon R^1} x\, Px(x)$ = $\sum_i x_i\, Pi$

Ejemplo:

P(W=-3)=0.2, P(W=-11)=0.7, P(W=31)=0.1

E(W)=(-3)(0.2)+(-11)(0.7)+(31)(0.1)= -0.6-7.7+3.1= -5.2

Caso totalmente continuo: E(X)=$\int_{-\infty}^{\infty} x f_x\,(x) dx$

El <u>**valor medio $\boldsymbol{\mu_x}$ de X es una medida de la posición**</u> de la distribución de probabilidad de X. La media se desplaza con la distribución de probabilidad.

Los valores de una o dos variables aleatorias también pueden calcularse como la suma de los productos de los valores de la función por sus probabilidades. E(XY)= $\sum_z z\, P(XY = z)$ → Pag 157 [1]

Varianza, covarianza y correlación de variables aleatorias

Estos valores nos permiten obtener información sobre la distribución de variables aleatorias.

El valor medio de X será E(X), esto no nos dice nada sobre el modo en que X tiende a ser E(X).

La varianza de una variable aleatoria: σ_x^2 = var(X) = E((x- μ_x)^2). Desviación estándar: σ_x= $\sqrt{var(X)}$ → evita efecto escala varianza

La varianza es una medida de lo dispersa que es la distribución de X, o cuan aleatoria es X, o cuanto varía X.

Ejemplo:

$$P_X(x) = \begin{cases} 1, & x = 10 \\ 0, & otro\ caso \end{cases}$$

$$P_Y(y) = \begin{cases} \frac{1}{2}, & Y = 5 \\ \frac{1}{2}, & Y = 15 \\ 0, & otro\ caso \end{cases}$$

E(X)=E(Y)=10

Var(X) = $(10-10)^2(1)=0$

Var(Y) = $(5-10)^2\,(1/2) + (15-10)^2\,(1/2)=25$

X e Y tienen el mismo valor esperado. La varianza de Y es mucho mayor que la de X. Y es más aleatoria (cambiante) que X, varía más de lo que hace X.

Varianza, covarianza y correlación de variables aleatorias

Propiedades de la varianza:

a) Var (X)≥ 0

b) a y b dos números reales var(aX+b) = a^2var(X)

c) Var(X) = E(X^2) - $(\mu_x)^2$ = E(X^2) - E(X)^2

d) Var (X) $\leq$ E(X^2)

El <u>efecto de la varianza</u> es la dispersión de cada distribución respecto a su valor medio.

Si la varianza aumenta, aumenta la dispersión.

Si la varianza disminuye, la distribución resulta más “concentrada” sobre el valor de la media.

Varianza, covarianza y correlación de variables aleatorias

La <u>**Covarianza**</u> determina la relación entre dos variables aleatorias.

$$Cov(X,Y)=E((X-\mu_x)(Y-\mu_y))$$

Ejemplo:

$$P_{X,Y}(x,y)=\begin{cases}\frac{1}{2} & x=3, y=4\\ \frac{1}{3} & x=3, y=6\\ \frac{1}{6} & x=5, y=6\\ 0 & otro\ caso\end{cases}$$

E(X)= (3)(1/2) + (3)(1/3) + (5)(1/6) = 10/3

E(Y)= (4)(1/2) + (6)(1/3) + (6)(1/6) = 5

Cov (X,Y)=E((X-10/3)(Y-5))= (3-10/3)(4-5)/2 + (3-10/3)(6-5)/3 + (5-10/3)(6-5)/6 = 1/3

Linealidad: Cov(X,Y)= E(XY) - E(X)E(Y)

Si X e Y son independientes: Cov(X,Y)=0. Cov(X,Y)=0 no implica independencia en todos los casos.

Varianza, covarianza y correlación de variables aleatorias

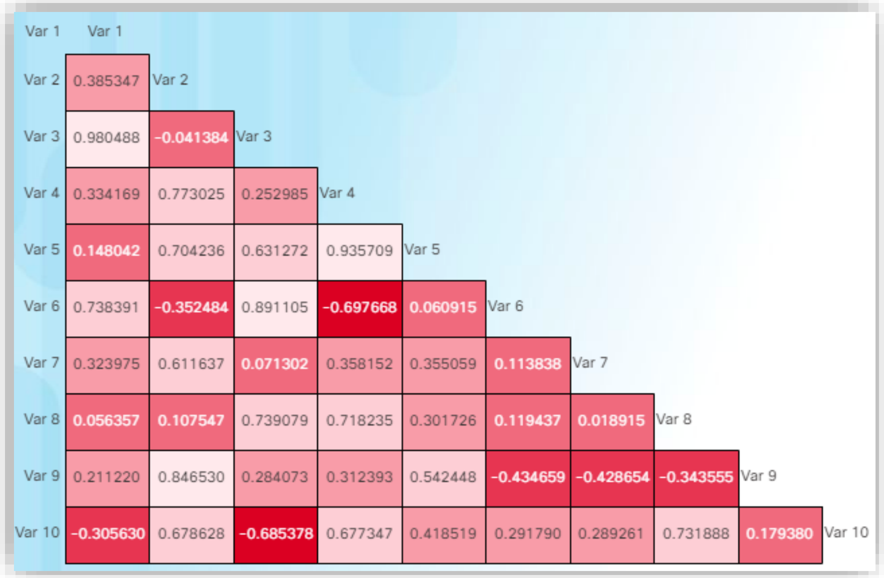
La **Correlación (coeficiente de correlación)** es una medida del grado de existencia de una relación lineal entre **dos variables aleatorias** X e Y.

$$corr(X,Y) = \frac{cov(X,Y)}{Sd(X)Sd(Y)} = \frac{cov(X,Y)}{\sqrt{var(X)var(Y)}}$$

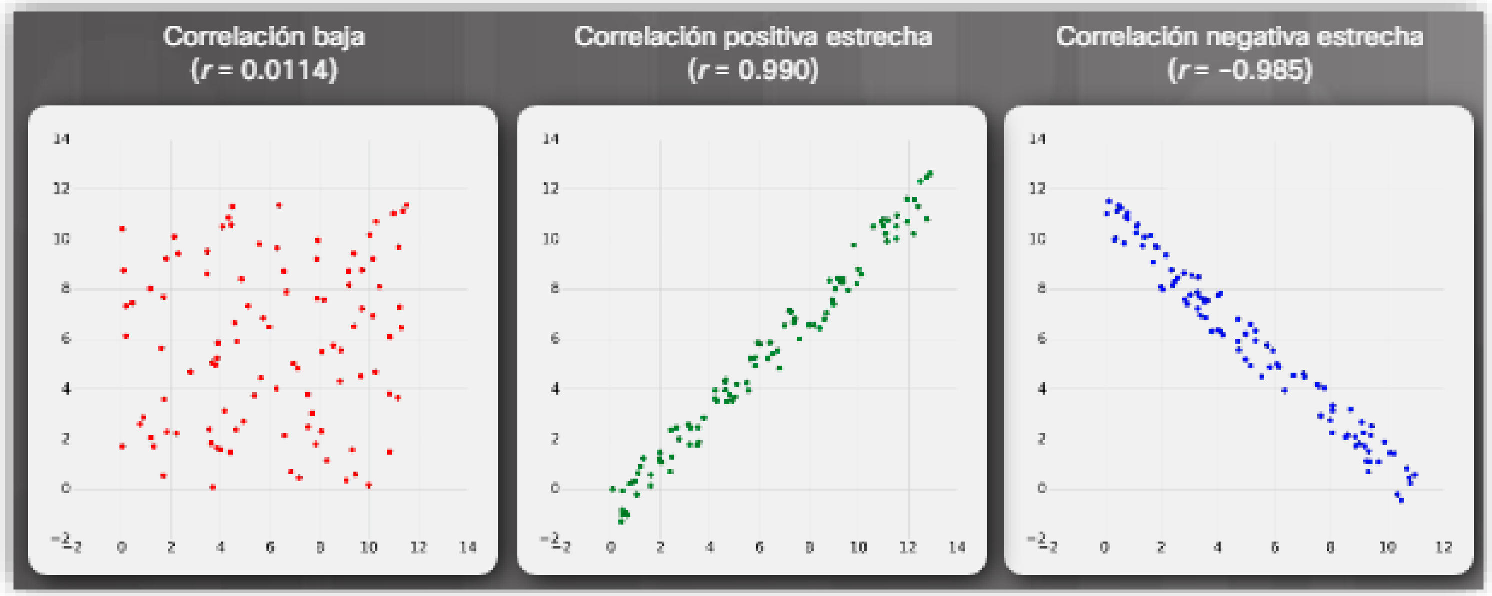
$$-1 \leq corr(X,Y) \leq 1$$

$corr(X,Y)$ =1 → Y aumenta X aumenta $corr(X,Y) = -1$ → Z disminuye X aumenta

Recursos para análisis de correlación y varianza:
https://rpsychologist.com/d3/correlation/
Mapa de Calor



	Var 1	Var 2	Var 3	Var 4	Var 5	Var 6	Var 7	Var 8	Var 9
Var 2	0.385347								
Var 3	0.980488	-0.041384							
Var 4	0.334169	0.773025	0.252985						
Var 5	0.148042	0.704236	0.631272	0.935709					
Var 6	0.738391	-0.352484	0.891105	-0.697668	0.060915				
Var 7	0.323975	0.611637	0.071302	0.358152	0.355059	0.113838			
Var 8	0.056357	0.107547	0.739079	0.718235	0.301726	0.119437	0.018915		
Var 9	0.211220	0.846530	0.284073	0.312393	0.542448	-0.434659	-0.428654	-0.343555	
Var 10	-0.305630	0.678628	-0.685378	0.677347	0.418519	0.291790	0.289261	0.731888	0.179380



Teoría de la Probabilidad: Momentos

Los primeros cuatro momentos estadísticos son: media, varianza, asimetría, curtosis.

Los Momentos estadísticos son una serie de medidas que se utilizan para describir la distribución de una variable aleatoria.

La distribución de probabilidad describe el comportamiento de una variable aleatoria

Una función que toma distintos valores con diferentes probabilidades

Funciones Generatrices

Función de distribución acumulada de X $F_X(x)=P(X\leq x)$, contiene toda la información de la distribución de probabilidad de X. En muchas situaciones, no es tan fácil calcular la media, varianza y otros momentos de una variable directamente. A menudo se pueden usar funciones generadoras para derivar los momentos indirectamente.

Otras funciones también contienen información:

- Función generatriz de probabilidad
- Función generatriz de momentos

Función generatriz de probabilidad

$$r_x(t)= E(t^x) \qquad t \in R^1$$

$$r_X(0)= P(X=0)$$

$$r'_X(0)= P(X=1)$$

$$r''_X(0)= 2P(X=2)$$

…

$$r_X^{(k)}(0)=k!\ P(X=k)$$

Una vez que conocemos la función generatriz de probabilidad $r_x(t)$, podemos calcular todos los valores de probabilidad P(X=0), P(X=1), P(X=2), etc.

Ejemplo: en el caso de la distribución binomial, estas probabilidades dan información relevante. La probabilidad de que X=1, es decir que haya 1 éxito de n ensayos. Así se pueden obtener todas las probabilidades y otras medidas.

Función generatriz de probabilidad

Ejemplo:

Distribución Binomial (n,$\boldsymbol{\theta}$): i éxitos de n ensayos. θ es la probabilidad de éxito.

$$r_x(t)= E(t^x) = \sum_{i=0}^{n} P(X=i)\, t^i = \sum_{i=0}^{n} \binom{n}{i} \theta^i (1-\theta)^{n-i} t^i = (t\theta + 1 - \theta)^n$$

$$r_X(0)= P(X=0) = (1-\theta)^n$$

$$r'_X(0)= P(X=1)= n(1-\theta)^{n-1}\theta$$

$$r''_X(0)= 2P(X=2) = n(n-1)(1-\theta)^{n-2}\theta^2$$

Podemos calcular todas las probabilidades de X a partir de r_X (al menos en el caso discreto).

Si X e Y toman valores {0,1,2,…} y $r_X = r_Y$ -> X=Y tienen la misma distribución.

Sólo se utiliza la función generadora de probabilidades para variables discretas y no negativas. Una ***función más general***, que se puede utilizar para cualquiera variable es la ***función generadora de momentos***.

Funciones Generatrices

Función generatriz de momentos

$m_x(s)= r_x(e^{s})= E(e^{sX})$ $s \in R^1$

Una vez conocida la función generatriz de momentos $m_x(s)$, podemos calcular todos los momentos $E(X)$, $E(X^2)$, $E(X^3)$, etc.

$m_x(0)= 1$

$m'_x(0)= E(X)$

$m''_x(0)= E(X^2)$

…

$m^{(k)}_X= E(X^k)$

Ejemplo:

Distribución Binomial (n,$\boldsymbol{\theta}$): $r_Y(t)= (t\theta + 1 - \theta)^n$ → $m_Y(s)=r_y(e^s)= (e^{s}\,\theta + 1 - \theta)^n$

$m'_Y(s)= n\theta e^s(\theta e^s - \theta + 1)^{(n-1)}$ → $m'_y(0)=$ $E(Y) = n\,\theta$

$m''_Y(s)= n\theta e^s(\theta e^s - \theta + 1)^{(n-2)}(n\theta e^s - \theta + 1)$ → $m''_Y(0)= E(Y^2)=$ $n^2\theta^2 - n\theta^2 + n\theta$

$Var(Y)= E(Y^2) - (E(Y))^2=$ $n^2\theta^2 - n\theta^2 + n\theta - n^2\theta^2 = -n\theta^2 + n\theta = n\theta\,(1 - \theta)$

Funciones Generatrices - Procesos

Proceso 1. Cálculo de la función generatriz de momentos para una distribución continua o discreta

Continua: $m_x(s) = E(e^{sX}) = \int_{-\infty}^{\infty} e^{sX} fx(x)dx$ Discreta: $m_x(s) = E(e^{sX}) = \sum_{i=0}^{n} P(X = i)\, e^{si}$

- Reemplazar la función de distribución en cuestión en $fx(x)$ o en $P(X = i)$, dependiendo el tipo de distribución
- Resolver la ecuación para obtener la función generatriz de la distribución en cuestión

Proceso 2. Cálculo de momentos y medidas estadísticas

- Calcular la primera derivada de la función generatriz de momentos y reemplazar s=0. Esto da el primer momento E(X). El primer momento es la media.
- Calcular la segunda derivada de la función generatriz de momentos y reemplazar s=0. Esto da el segundo momento $E(X^2)$. La varianza se calcula con el primer y segundo momento. $Var(X) = E(X^2) - (E(X))^2$
- …
- Calcular la k derivada de la función generatriz de momentos y reemplazar s=0. Esto da el k-ésimo momento $E(X^k)$.

Funciones Generatrices

Función generatriz de momentos

$$m_x(s) = r_x(e^s) = E(e^{sX}) \qquad s \in R^1$$

Una vez conocida la función generatriz de momentos $m_x(s)$, podemos calcular todos los momentos $E(X)$, $E(X^2)$, $E(X^3)$, etc.

$$m_x(0) = 1$$

$$m'_x(0) = E(X)$$

$$m''_x(0) = E(X^2)$$

$$\ldots$$

$$m^{(k)}_X = E(X^k)$$

Ejemplo:

Distribución Exponencial ($\boldsymbol{\lambda}$): $m_x(s) = E(e^{sX}) = \int_{-\infty}^{\infty} e^{sX} fx(x)dx = \int_0^{\infty} e^{sX} \lambda e^{-\lambda x} \mathrm{dx} = \lambda(\lambda - s)^{-1}$

$m'_x(0) = E(X) = 1/\lambda$

$m''_x(0) = E(X^2) = 2/(\lambda^2)$

$Var(X) = E(X^2) - (E(X))^2 = 2/(\lambda^2) - 1/\lambda^2 = 1/\lambda^2$

Funciones Generatrices

Momentos:

- **Primer momento E(X):** la media de la variable aleatoria. Medida de la posición sobre la recta real en que se localiza el centro de probabilidad de X, al menos cuando la distribución es unimodal (un solo máximo) y no sea demasiado sesgada.
- **Segundo momento $E(X^2)$:** junto con el primer momento, nos da la varianza a través de la relación $Var(X)= E(X^2) – (E(X))^2$
- Los dos primeros momentos nos informan de la posición y dispersión (o grado de concentración) de la distribución respecto a su media.
- Lo momentos de orden superior también proporcionan información sobre la distribución. Con el tercer momento se obtiene el Sesgo (Oblicuidad/Asimetría/Skewness). Con el cuarto momento se obtiene la Curtosis (Apuntamiento/Kurtosis). La mayoría de distribuciones poseen función generatriz de momentos.
- Las funciones generatrices pueden ayudarnos con las distribuciones compuestas.

REFERENCIAS

- Evans, M. Rosenthal, J. (2013) Probabilidad y estadística la ciencia de la incertidumbre. Reverté.
- Forsyth, D. (2018). Probability and Statistics for Computer Science. Springer.
- Triola, M. F. (2017). Estadística: conceptos y aplicaciones. Pearson Educación.
- Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. (2011). Probabilidad y estadística para ingeniería y ciencias. Pearson Educación.
- Johnson, R. A., & Wichern, D. W. (2018). Applied multivariate statistical analysis. Pearson Educación.
- Rencher, A. C., & Schaalje, G. B. (2008). Linear models in statistics. John Wiley & Sons.
- Joshi, P. (2017). Artificial intelligence with python. Packt Publishing Ltd. Enlace de repositorio de ejercicios: https://github.com/PacktPublishing/Artificial-Intelligence-with-Python
- S. Raschka, V. Mirjalili. Python Machine Learning, Packt Publishing Ltd, 2017. Enlace de repositorio de ejercicios: https://github.com/PacktPublishing/Python-Machine-Learning-Second-Edition