APUNTES DE ESTADÍSTICA
PARA UN CURSO INTRODUCTORIO DE ECONOMETRÍA

Julio César Alonso C.

No. 12
Marzo 2007

UNIVERSIDAD ICESI
DEPARTAMENTO DE ECONOMÍA

APUNTES DE ECONOMÍA

ISSN 1794-029X

No. 12, Marzo de 2007

Editor

Julio César Alonso C.

jcalonso@icesi.edu.co

Vanessa Ospina López

Asistente de Edición

Gestión Editorial

Departamento de Economía - Universidad Icesi

www.icesi.edu.co

Tel: 5552334 ext: 8398. Fax: 5551441

Calle 18 # 122-135 Cali, Valle del Cauca, Colombia

APUNTES DE ESTADÍSTICA
PARA UN CURSO DE PREPARACIÓN DE ECONOMETRÍA

Julio Cesar Alonso C[1].

Marzo de 2007

Resumen

Este documento presenta una breve introducción a los conceptos básicos de estadística que forman la base de un curso introductorio de Econometría de pregrado. Se discuten conceptos como variables, vectores y matrices aleatorias, distribución de probabilidad, valor esperado, independencia estadística, distribución conjunta y marginal, el teorema del límite central, el sesgo de un estimador y aspectos generales de la construcción de intervalos de confianza y evaluación de hipótesis. Este documento está dirigido principalmente a estudiantes de pregrado de economía o primer año de maestría en finanzas o economía, pero por la sencillez del lenguaje, puede ser de utilidad para cualquier estudiante o profesional interesado en repasar los concetos básicos de álgebra matricial.

Palabras Clave: Econometría, Principios de Estadística, variables aleatorias, vectores y matrices aleatorias, distribución de probabilidad, valor esperado, independencia estadística, distribución conjunta y marginal, el teorema del límite central, el sesgo de un estimador.

Apuntes de Economía es una publicación del Departamento de Economía de la Universidad Icesi, cuya finalidad es divulgar las notas de clase de los docentes y brindar material didáctico para la instrucción en el área económica a diferentes niveles. El contenido de esta publicación es responsabilidad absoluta del autor.

DO NOT COPY

---

[1] Profesor del Departamento de Economía y Director del Centro de Investigación en Economía y Finanzas (CIENFI) de la Universidad Icesi, jcalonso@icesi.edu.co.

1 Elementos de Estadística.

Algunos autores describen la ciencia estadística como la tecnología del método científico, pues es la estadística y sus métodos los que proveen a los investigadores con las herramientas para probar sus hipótesis de trabajo (Por ejemplo Mood (1950)). En general, la Estadística es definida como "la ciencia de estimar la distribución de probabilidad de una variable aleatoria basada en repetidas observaciones de variables aleatorias de la misma variable aleatoria" (Amemiya (1994)).

Así, la estadística es una ciencia que emplea conjuntos de datos para obtener a partir de ellos inferencias (proyección, adivinanza) sobre una población (valor real).De manera que el problema estadístico consiste en encontrar la mejor predicción para un valor real desconocido para el investigador, a partir de datos recolectados (muestra) de una población.

En este documento repasaremos los conceptos básicos de estadística y probabilidad que son las bases de un curso introductorio de econometría.

1.1 Variables, vectores y matrices aleatorias.

Una variable se define como una magnitud que puede tener un valor cualquiera de los comprendidos en un conjunto. En otras palabras, es una "letra" que puede tomar uno o diferentes valores. Por ejemplo, si la variable x cumple la siguiente condición $3x = 2$, entonces la variable necesariamente tomará el valor de $2/3$ $(x = 2/3)$. Otro ejemplo, si la variable x cumple la condición $x^2 = 1$, entonces x puede tomar los valores de 1 o -1.

Ahora, consideremos la definición de una ***variable aleatoria***, también conocida como variable estocástica. Una variable aleatoria es una "letra" que toma diferentes valores, cada uno con una probabilidad previamente definida. Por ejemplo, tiremos una moneda

DO NOT COPY

justa[2] al aire, y sea X la variable aleatoria que toma el valor de uno si la cara superior de la moneda es sello, en caso contrario la variable toma el valor de cero. Es decir

$$X = \begin{cases} 1 \;\; si \; sello \\ 0 \;\; si \; cara \end{cases}$$

Entonces, en este caso, diremos que la variable aleatoria X tiene dos posibles realizaciones. Ahora bien, si la moneda es una moneda normal, existirá igual probabilidad que la variable aleatoria tome el valor de uno o cero, en otras palabras tendremos que la probabilidad de que la variable aleatoria sea igual a uno es 0.5, al igual que la probabilidad que la variable aleatoria sea cero. Esto se puede abreviar de la siguiente forma: $P(X=1)=0.5$ y $P(X=0)=0.5$.

Si el conjunto de valores que toma la variable aleatoria es un conjunto finito o infinito contable, entonces la variable estocástica se denomina una ***variable aleatoria discreta***. Por otro lado, si las posibles realizaciones de la variable aleatoria son un conjunto de realizaciones infinitamente divisible y, por tanto, imposible de contar, entonces la variable estocástica se conoce como una ***variable aleatoria continua***. En general, si las posibles realizaciones toman valores discretos entonces estamos hablando de una variable estocástica discreta; por el contrario, si los posibles valores son parte de un rango continuo de valores, entonces estamos hablando de una variable estocástica continua.

Un ***vector aleatorio*** $\underset{\sim}{X}$ es un vector cuyos elementos son variables aleatorias ya sean continuas o discretas, es decir,

$$\underset{\sim}{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}, \qquad (1)$$

---

[2] Por una moneda justa, se entiende una moneda que tiene una probabilidad igual de obtener cualquiera de las dos caras.

DO NOT COPY

donde X_i para $i=1,2,...,n$ representan diferentes variables aleatorias. Análogamente, una matriz aleatoria es una matriz cuyos elementos son variables aleatorias.

Es importante anotar que los economistas interpretamos algunos aspectos de la economía como resultados de un proceso estocástico. En la práctica observamos un único valor de una variable como el PIB o los rendimientos de un activo. Los valores observados en la realidad para esas variables aleatorias, se interpretan como las realizaciones de una variable aleatoria después de que los "dados" de la economía ya han sido tirados, es decir, lo que observa el investigador es la realización de un evento aleatorio.

1.2 Distribución de probabilidad

Una ***distribución de probabilidad de una variable aleatoria discreta***, también conocida como la función de densidad discreta, $f(x)$, es una lista de las probabilidades asociadas a las diferentes realizaciones x que puede tomar una variable aleatoria discreta X. Para una variable aleatoria discreta X tenemos que

$$f(x) = \mathrm{P}(X = x) \tag{2}$$

donde $f(x)$ debe cumplir que:

- $0 \leq f(x) \leq 1$
- $\sum_{todo\ x_i} f(x_i) = 1$

Dado que en el caso de una variable aleatoria continua, ésta puede tomar cualquier valor dentro de un número infinito de valores, será imposible asignar una probabilidad para cada uno de los valores que puede tomar la variable aleatoria continua. Por tanto, en el caso de variables aleatorias continuas es necesario un enfoque diferente al

DO NOT COPY

seguido con las variables aleatorias discretas. En este caso definiremos una función que nos permita conocer la probabilidad de ocurrencia de un intervalo (conjunto continuo de puntos) y no un punto como lo hicimos para las variables aleatorias discretas.

Ejemplo 1

Suponga que lanzamos un dado y definimos la variable aleatoria W como el valor de la cara superior del dado. Encuentre la distribución de probabilidad de W.

Respuesta: Suponiendo que se trata de un dado sin truco, cada una de las 5 caras tiene la misma probabilidad de realización, $1/6$. Así, la distribución de probabilidad de W es:

$$P(W=w)=\begin{cases}\frac{1}{6} & \text{Si } w=1\\ \frac{1}{6} & \text{Si } w=2\\ \frac{1}{6} & \text{Si } w=3\\ \frac{1}{6} & \text{Si } w=4\\ \frac{1}{6} & \text{Si } w=5\\ \frac{1}{6} & \text{Si } w=6\\ 0 & o.w.\end{cases}$$

donde $o.w.$ significa en otros casos (del inglés otherwise). Una forma corta de reescribir esto es $P(W=w)=\frac{1}{6}I_{(w=1,2,3,4,5,6)}$, donde $I_{(A)}$ se conoce como la ***función indicador*** que toma el valor de uno cuando A es cierto y cero en otros casos.

Una ***distribución de probabilidad de una variable aleatoria continua***, también conocida como la función de densidad continua, $f(x)$, es una función asociada a la variable aleatoria continua X, tal que

$$\int_a^b f(x)\,dx = \mathrm{P}(a \le x \le b) \tag{3}$$

donde $f(x)$ debe cumplir que:

- $f(x) \ge 0$
- $\int_{-\infty}^{\infty} f(x)\,dx = 1$

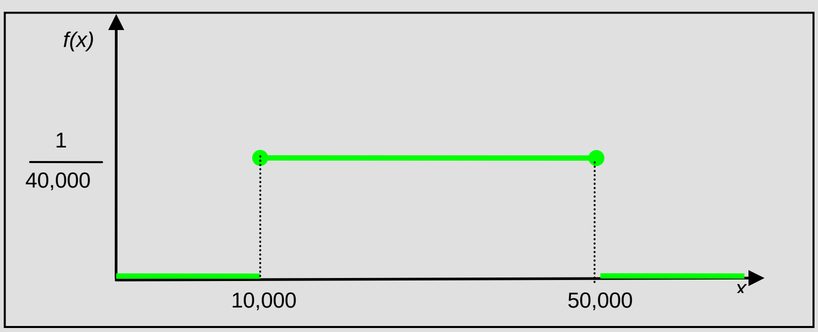
COPY

Ejemplo 2

Un concesionario de automotores, después de contratar un estudio, está seguro que la función de distribución de su consumo de carburante mensual está dada por

$$f(x) = \begin{cases} \dfrac{1}{40,000} & \text{si} \quad 10,000 \le x \le 50,000 \\ 0 & o.w. \end{cases}.$$

donde x corresponde a la cantidad de galones de carburante efectivamente empleados en el mes. Esta función de distribución se puede graficar de la siguiente forma:



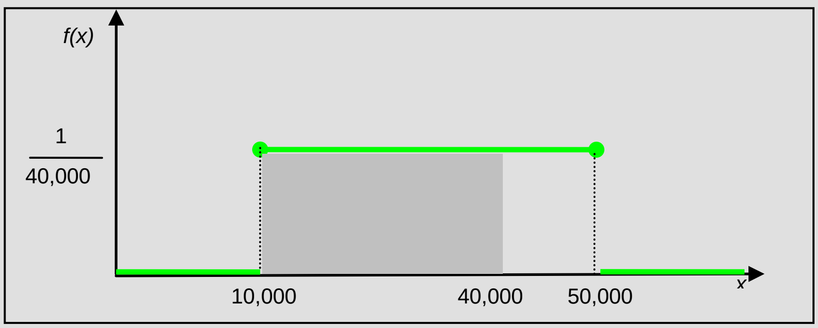
Encuentre la probabilidad de que este mes sean empleados: a) exactamente 30,000 litros, b) no más de 40,000, y c) entre 20,000 y 30,000.

Respuesta: Esta función de distribución se conoce con el nombre de distribución uniforme. A continuación se responden las preguntas:

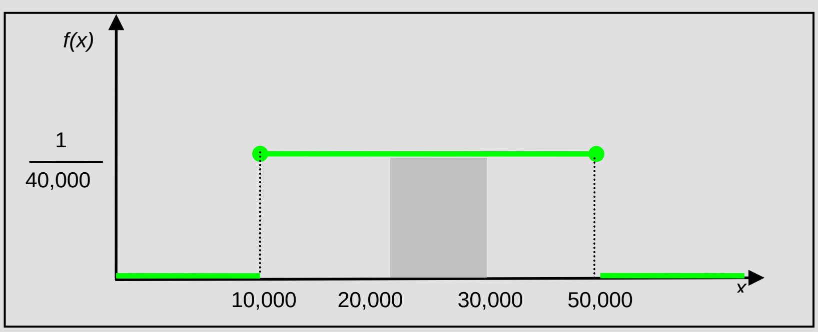
a) Debe ser claro que $P(X = 30,000) = P(30,000 \leq X \leq 30,000) = \int_{30,000}^{30,000} f(x)dx = 0$.

Así, la probabilidad de que el consumo sea exactamente 30,000 galones es cero.

b) $P(X \leq 40,000) = \int_{0}^{40,000} f(x)dx = \frac{3}{4} = .75$. Por tanto, la probabilidad de que el consumo de carburante mensual sea no mayor de 40,000 galones es de 0.75. Gráficamente esto se puede representar por el área sombreada del siguiente gráfico.



c) $P(20,000 \leq X \leq 30,000) = \int_{20,000}^{30,000} f(x)dx = .25$, gráficamente:



1.3 Valor esperado de una variable aleatoria y otros momentos de las variables aleatorias.

Eventualmente, las distribuciones de probabilidad se pueden describir con sus momentos[3]. El primer momento de una distribución se conoce como el valor esperado o esperanza matemática.

El valor esperado de una variable aleatoria corresponde a su media poblacional y se interpreta como el valor promedio que se espera de la variable aleatoria cuando se obtiene cualquier muestra de ésta.

El valor esperado de una variable aleatoria discreta denotado por $E[X]$ se define como

$$E[X] = \sum_{todo \ x_i} x_i P(X = x_i) = \sum_{todo \ x_i} x_i f(x_i). \quad (4)$$

Ejemplo 3

Continuando con el Ejemplo 1

Encuentre el valor esperado de W.

Respuesta: De acuerdo a la definición, $E[W] = \sum_{i=1}^{6} w_i P(W = w_i) = \sum_{i=1}^{6} w_i \frac{1}{6} = \frac{1}{6}\sum_{i=1}^{6} w_i$ $= \frac{21}{6} = \frac{7}{2} = 3.5$. Así, en promedio se espera que la variable aleatoria tome un valor de 3.5.

El valor esperado de una variable aleatoria continua, también denotado por $E[X]$ se define como:

[3] Los momentos de una distribución son representados por parámetros poblacionales que se representarán de aquí en adelante con letras griegas. Los momentos de una distribución describen las características de la distribución poblacional de la variable aleatoria.

DRAFT COPY

$$E[X]=\int_{-\infty}^{\infty} xf(x)\,dx \qquad (5)$$

Ejemplo 4

Continuando con el Ejemplo 2

Encuentre el valor esperado del consumo de carburante mensual.

Respuesta: De acuerdo a la definición, tenemos que

$$E[X]=\int_{-\infty}^{\infty} xf(x)\,dx=\int_{10,000}^{50,000} x\frac{1}{40,000}dx$$

$$=\frac{1}{40,000}\left[\frac{x^2}{2}\right]_{10,000}^{50,000}=\frac{1}{8\cdot 10^4}\left[50,000^2-10,000^2\right]=\frac{10^8}{8\cdot 10^4}\left[5^2-1^2\right]=\frac{10^4}{8}24=30,000.$$

Es decir, el consumo mensual esperado de carburante es de 30,000 galones.

El valor esperado también es conocido como el operador de esperanza matemática, particularmente es un operador lineal cuyas principales características son:

- $E[c]=c$, donde c es una constante, o una variable no estocástica.
- $E[aX+b]=aE[X]+b$, donde a y b son constantes y X es una variable aleatoria.
- En general $E[g(X)]\neq g(E(X))$, donde $g(\bullet)$ es cualquier función. La única excepción de esto es cuando $g(\bullet)$ es una función lineal.
- $E[a_1X_1+a_2X_2+\ldots+a_nX_n]=a_1E[X_1]+a_2E[X_2]+\ldots+a_nE[X_n]$ donde cada uno de los a_i y X_i $(i=1,2,\ldots,n)$ son constantes y variables aleatorias, respectivamente.
- $E[g(X)]=\begin{cases}\sum\limits_{todo\ x_i} g(x_i)\,f(x_i) & \text{si } X \text{ es discreta}\\ \int_{-\infty}^{\infty} g(x)\,f(x)\,dx & \text{si } X \text{ es continua}\end{cases}$

Como se mencionó anteriormente, $E\left[X\right]$ se conoce como el primer momento de una variable aleatoria y también se denota como μ_X, es decir, la media poblacional de X. El ***i-ésimo momento*** (alrededor del origen) de una variable aleatoria, μ_i', está definido por $\mu_i' = E\left[X^i\right]$

1.4 Independencia lineal

Dos variables aleatorias, X y Y, se consideran estadísticamente independientes, u ortogonales, si y solamente si

$$E\left[XY\right] = E\left[X\right]E\left[Y\right] \tag{6}$$

Es importante notar que independencia estadística entre dos variables no implica que no exista relación alguna entre las variables, como se verá más adelante, independencia estadística sólo implica que no existe una relación lineal entre las dos variables.

1.5 Varianza y momentos alrededor de la media de una variable aleatoria.

La varianza de una variable aleatoria, denotada por σ^2 o $Var\left[X\right]$, se define como

$$Var\left[X\right] = E\left[\left(x-\mu\right)^2\right] \tag{7}$$

Así, en el caso de una variable aleatoria discreta tendremos que $Var\left[X\right] = \sum_{todo\ \ x_i} \left(x_i-\mu\right)^2 f\left(x_i\right)$, y para una variable estocástica continua la varianza será calculada de la siguiente manera: $Var\left[X\right] = \int_{-\infty}^{\infty} \left(x-\mu\right)^2 f\left(x\right)dx$. La varianza es una medida de la dispersión de una distribución. Generalmente se emplea la raíz cuadrada de la varianza, la desviación estándar ($\sigma = \sqrt[+]{Var\left(X\right)}$), para describir una distribución. La ventaja de la desviación estándar es que ésta está medida en las mismas unidades de X y μ.

DO NOT COPY

Un ejemplo de cómo la desviación estándar puede ser empleada para describir la dispersión de una distribución está dado por la desigualdad de ***Chebychev***; para cualquier variable aleatoria X y para cualquier constante k se tiene que:

$$\mathrm{P}\left(\mu - k\sigma \leq x \leq \mu + k\sigma\right) \geq 1 - \frac{1}{k^2}. \tag{8}$$

Antes de continuar, es importante anotar que el cálculo directo de la varianza es relativamente engorroso, afortunadamente es fácil mostrar (vea el Ejercicio 4) que

$$Var\left[X\right] = E\left[X^2\right] - \left(E\left[X\right]\right)^2 \tag{9}$$

este resultado permite en la práctica agilizar el cálculo de la varianza de cualquier variable aleatoria.

Ejemplo 5

Continuando con el Ejemplo 1

Calcule la varianza de W.

Respuesta: Empleando la expresión **(9)** tenemos que $Var\left[W\right] = E\left[W^2\right] - \left(E\left[W\right]\right)^2$

En el Ejemplo 3

encontramos que $E\left[W\right] = \frac{7}{2}$.

Recuerde que $E\left[W^2\right] \neq \left(\frac{7}{2}\right)^2 = \left(E\left[W\right]\right)^2$. Así, es necesario calcular $E\left[W^2\right]$, aplicando nuevamente la definición de valor esperado de una función, tenemos que

$$E\left[W^2\right] = \sum_{i=1}^{6} \left(w_i\right)^2 P\left(W = w_i\right) = \sum_{i=1}^{6} \left(w_i\right)^2 \frac{1}{6} = \frac{1}{6} \sum_{i=1}^{6} \left(w_i\right)^2 = \frac{91}{6}.$$

Por tanto, $Var\left[W\right] = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$.

Las principales propiedades de la varianza son:

- $Var[c]=0$, donde c es una constante, o una variable no estocástica.
- $Var[aX+b]=a^2Var[X]$, donde a y b son constantes y X es una variable aleatoria.
- $Var[aX+bY]=a^2Var[X]+b^2Var[Y]+2abCov[X,Y]$, donde a y b son constantes y X y Y son variables aleatorias (en la próxima sección repasaremos el concepto de Cov).

La varianza de una variable aleatoria también es conocida como el segundo momento alrededor de la media. En general, el ***i-ésimo momento alrededor de la media de una variable aleatoria***, denotada por $\mu_i = E\left[(X-\mu)^i\right]$.

Ejemplo 6

Continuando con el Ejemplo 2

Encuentre la varianza del consumo de carburante mensual.

Respuesta: Similarmente al caso de una variable estocástica discreta, tenemos que $Var[X]=E[X^2]-(E[X])^2$, antes encontramos que $E[X]=3\cdot 10^4$. Ahora necesitamos encontrar $E[X^2]$; este valor esperado es

$$E[X^2]=\int_{-\infty}^{\infty} x^2 f(x)\,dx=\int_{10,000}^{50,000} x^2\frac{1}{40,000}dx$$

$$=\frac{1}{40,000}\left[\frac{x^3}{3}\right]_{10,000}^{50,000}=\frac{1}{12\cdot 10^4}\left[50,000^3-10,000^3\right]=\frac{10^{12}}{12\cdot 10^4}\left[5^3-1^3\right]=\frac{10^8}{12}124=\frac{31}{3}\cdot 10^8$$

. Por tanto tenemos que $Var[X]=E[X^2]-(E[X])^2=\frac{31}{3}\cdot 10^8-9\cdot 10^8=\frac{4}{3}\cdot 10^8$. Es decir la varianza del consumo mensual es de 133,333,333.33.

El tercero y cuarto momento alrededor de la media se conocen como la asimetría (skewness en inglés) y curtosis, respectivamente. Una medida de asimetría comúnmente empleada es el coeficiente de asimetría definido como:

$$A = \frac{\mu_3}{\sigma^3}. \tag{10}$$

Otro estadístico comúnmente empleado para describir que tan aplanada o "picuda" es una distribución, es el coeficiente de curtosis que se define como:

$$C = \frac{\mu_4}{\sigma^4}. \tag{11}$$

Tabla 1. Interpretación del Coeficiente de Simetría

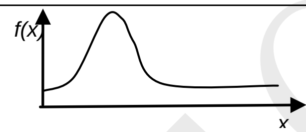
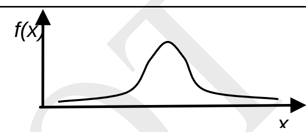
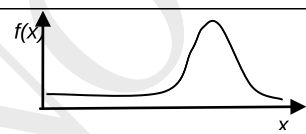
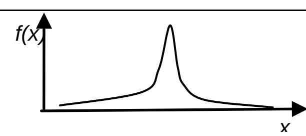
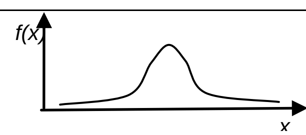
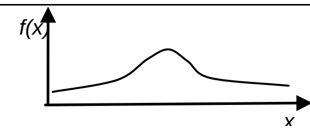
A ? 0	Interpretación	Gráfico
>	Asimetría a la derecha[4]	
=	Simetría	
<	Asimetría a la izquierda	

Tabla 2. Interpretación del Coeficiente de Curtosis

C ? 3	Interpretación	Gráfico
>	Distribución platicúrticas (picuda, ancha o de colas cortas)	
=	Distribución mesocúrtica (por Ej.: distribución normal)	

[4] En otras palabras, "la cola grande" de la distribución está hacia la derecha.

DO NOT COPY

<	Distribución Leptocúrtica (plana, delgada o de colas largas)	

1.6 Covarianza y Correlación entre dos variables aleatorias

Ahora consideremos la covarianza entre dos variables aleatorias X y Y denotada por $Cov(X,Y)$ ó $\sigma_{X,Y}$ y definida como

$$Cov(X,Y) = E\left[\left(X - E[X]\right)\left(Y - E[Y]\right)\right]. \qquad (12)$$

Al igual que lo que ocurre con la varianza de una variable aleatoria, el cálculo directo de una covarianza es muy engorroso. Afortunadamente, es fácil mostrar que

$$Cov[X,Y] = E[XY] - E[X]E[Y]. \qquad (13)$$

La expresión (13) ayuda a entender la utilidad de la covarianza entre dos variables aleatorias. Noten que en caso de que las variables estocásticas X y Y sean independientes, se tendrá que $E[XY] = E[X]E[Y]$. Así, $Cov[X,Y] = 0$.

Por tanto, la covarianza entre dos variables aleatorias será cero si no existe relación lineal (hay independencia) entre ellas; y será diferente de cero si no hay independencia estadística entre ellas. Por otro lado, en el caso de que al mismo tiempo que una realización de la variable aleatoria X está por encima de su media, la realización de la variable estocástica Y también está por encima de su media, entonces la covarianza de estas dos variables será positiva. Si cuando la realización de una variable aleatoria está por encima de su media la realización de la otra variable está por debajo de la media, entonces la covarianza será negativa.

Una importante propiedad de la covarianza es:

- $Cov[a + bX, c + dY] = bdCov(X,Y)$, donde a, b, c, y d son constantes, y X y Y son variables aleatorias.

DO NOT COPY

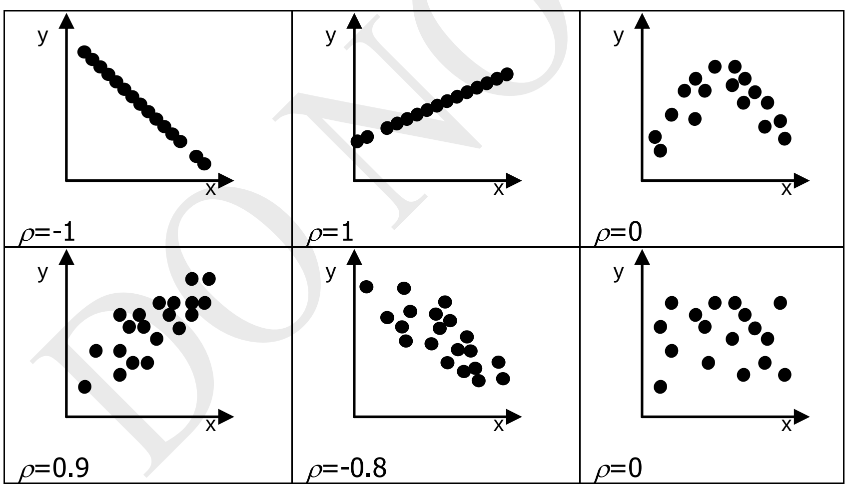
Como se mencionó anteriormente, la covarianza entre dos variables estocásticas mide la relación lineal entre las variables, pero ésta depende de las unidades en que están medidas X y Y. Para tener una medida del grado de dependencia lineal entre dos variables aleatorias, que no dependa de las unidades, se emplea el coeficiente de correlación.

La correlación entre dos variables aleatorias, denotado por ρ, está definida por:

$$\rho = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}. \qquad (14)$$

Es muy fácil mostrar que $-1 \leq \rho \leq 1$ (ver Ejercicio 7). La correlación entre dos variables aleatorias tiene una interpretación muy sencilla; por ejemplo, una correlación de 1/-1 entre las variables aleatorias X y Y implica una relación lineal positiva/negativa y perfecta entre ellas. Mientras que una correlación de cero implica que no existe relación lineal entre las variables.

Gráfico 1. Diagramas de dispersión y sus correspondientes correlaciones.



DO NOT COPY

1.7 Esperanza y Varianza de vectores aleatorios.

Como se mencionó anteriormente, un vector aleatorio es un vector cuyos elementos son todos variables aleatorias. Así, el valor esperado de un vector aleatorio corresponde a un vector cuyos elementos son los valores esperados de los correspondientes elementos del vector estocástico. En otras palabras, sea $\underset{\sim}{X}$ un vector aleatorio, entonces

$$E\left[\underset{\sim}{X}\right]=E\left[\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}\right]=\begin{bmatrix} E\left[X_1\right] \\ E\left[X_2\right] \\ \vdots \\ E\left[X_n\right] \end{bmatrix}=\begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix}=\mu\,. \tag{15}$$

Es muy fácil extender esta idea para encontrar el valor esperado de una matriz aleatoria. Sea $X_{n\times m}$ una matriz aleatoria de dimensiones $n\times m$, entonces

$$E\left[X_{n\times m}\right]=\begin{bmatrix} E\left[X_{11}\right] & E\left[X_{12}\right] & \cdots & E\left[X_{1n}\right] \\ E\left[X_{21}\right] & E\left[X_{22}\right] & \cdots & E\left[X_{2n}\right] \\ \vdots & \vdots & \vdots & \vdots \\ E\left[X_{m1}\right] & E\left[X_{m2}\right] & \cdots & E\left[X_{mn}\right] \end{bmatrix}$$

Análogamente al caso de una variable aleatoria, la varianza de un vector aleatorio $\underset{\sim}{X}$ se define como:

$$Var\left[\underset{\sim}{X}\right]=E\left[\left(\underset{\sim}{X}-\mu\right)\left(\underset{\sim}{X}-\mu\right)^T\right]=E\left[\underset{\sim}{X}\underset{\sim}{X}^T\right]-\mu\mu^T \tag{16}$$

en este caso tenemos que

$$Var\left[\underset{\sim}{X}\right]=\begin{bmatrix} E\left[\left(X_1-\mu_1\right)\left(X_1-\mu_1\right)\right] & E\left[\left(X_1-\mu_1\right)\left(X_2-\mu_2\right)\right] & \cdots & E\left[\left(X_1-\mu_1\right)\left(X_n-\mu_n\right)\right] \\ E\left[\left(X_2-\mu_2\right)\left(X_1-\mu_1\right)\right] & E\left[\left(X_2-\mu_2\right)\left(X_2-\mu_2\right)\right] & \cdots & E\left[\left(X_1-\mu_1\right)\left(X_2-\mu_2\right)\right] \\ \vdots & \vdots & \vdots & \vdots \\ E\left[\left(X_n-\mu_n\right)\left(X_1-\mu_1\right)\right] & E\left[\left(X_n-\mu_n\right)\left(X_2-\mu_2\right)\right] & \cdots & E\left[\left(X_n-\mu_n\right)\left(X_n-\mu_n\right)\right] \end{bmatrix}$$

DO NOT COPY

por tanto, la matriz de varianzas de un vector aleatorio $\underset{\sim}{X}$, conocida como la ***matriz de covarianzas*** o la ***matriz de varianzas y covarianzas***, está dada por:

$$\begin{aligned} Var\left[\underset{\sim}{X}\right] &= \begin{bmatrix} Var[X_1] & Cov(X_1,X_2) & \cdots & Cov(X_1,X_n) \\ Cov(X_2,X_1) & Var[X_2] & \cdots & Cov(X_2,X_n) \\ \vdots & \vdots & \vdots & \vdots \\ Cov(X_n,X_1) & Cov(X_n,X_2) & \cdots & Var[X_n] \end{bmatrix} \\ &= \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{bmatrix} = \Sigma \end{aligned} \tag{17}$$

Dividiendo cada uno de los σ_{ij} por las respectivas σ_i y σ_j obtendremos la ***matriz de correlaciones***:

$$R = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{21} & 1 & \cdots & \rho_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} & \rho_{n2} & \cdots & 1 \end{bmatrix} \tag{18}$$

Antes de continuar, consideremos las siguientes propiedades. Sean $\underset{\sim}{a}$ un vector de constantes, A una matriz de constantes y $\underset{\sim}{X}$ un vector aleatorio, entonces:

- $E\left[\underset{\sim}{a}^T \underset{\sim}{X}\right] = \underset{\sim}{a}^T \mu$
- $Var\left[\underset{\sim}{a}^T \underset{\sim}{X}\right] = \underset{\sim}{a}^T Var\left[\underset{\sim}{X}\right]\underset{\sim}{a} = \underset{\sim}{a}^T \Sigma \underset{\sim}{a}$ donde
- $E\left[A\underset{\sim}{X}\right] = A\mu$
- $Var\left[A\underset{\sim}{X}\right] = A\Sigma A^T$
- $E\left[tr\left(A_{n\times n}\right)\right] = tr\left(E\left[A_{n\times n}\right]\right)$

DO NOT COPY

1.8 Resultados importantes relacionados con Distribuciones de Probabilidad

Sin duda, la distribución de probabilidad más empleada y común es la distribución normal. Una variable aleatoria X sigue una distribución normal con media μ y varianza σ^2 si y solamente si

$$f\left(x \middle| \mu,\sigma^2\right) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{(x-\mu)^2}{\sigma^2}\right]}. \tag{19}$$

En este caso se escribe $X \sim N\left(\mu,\sigma^2\right)$ (se lee X está distribuida normalmente con media μ y varianza σ^2). En particular, la distribución $N(0,1)$ se conoce como la distribución estándar normal, una variable que sigue esta distribución comúnmente se denota por z.

A continuación consideraremos varios resultados importantes que involucran variables aleatorias que se distribuyen normalmente:

- Si $X \sim N\left(\mu,\sigma^2\right)$, entonces $Y = \dfrac{X-\mu}{\sigma} \sim N(0,1)$.
- Si $Y \sim N(0,1)$, entonces Y^2 sigue una distribución Chi-cuadrado con un grado de libertad, denotado por $Y^2 \sim \chi^2_{(1)}$.
- Si $Y_i \sim N(0,1)$ para $i = 1,2,...,n$, entonces $\sum_{i=1}^{n}(Y_i)^2 \sim \chi^2_{(n)}$
- Si $X_i \sim N\left(0,\sigma^2\right)$ para $i = 1,2,...,n$, entonces $\sum_{i=1}^{n}\left(\dfrac{Y_i}{\sigma}\right)^2 \sim \chi^2_{(n)}$
- Si $X_1 \sim \chi^2_{(n_1)}$, $X_2 \sim \chi^2_{(n_2)}$ y adicionalmente estas dos variables aleatorias son independientes, entonces $X_1 + X_2 \sim \chi^2_{(n_1+n_2)}$.

DO NOT COPY

- Si $X_1 \sim \chi^2_{(n_1)}$, $X_2 \sim \chi^2_{(n_2)}$ y adicionalmente estas dos variables aleatorias son independientes, entonces $\frac{X_1/n_1}{X_2/n_2}$ sigue una ***distribución F*** con n_1 grados de libertad en el numerador y n_2 grados de libertad en el denominador[5] (una abreviatura para esto es $F_{(n_1,n_2)}$)
- Si $X \sim N\left(0,\sigma^2\right)$, $Y \sim \chi^2_{(n)}$ y X y Y son independientes, entonces $\frac{X}{\sqrt{Y/n}}$ sigue ***una distribución t*** con n grados de libertad, denotado por $t_{(n)}$.
- Si $X \sim N_n\left(0,\Sigma\right)$, $ran\left(A\right)=k$ y $A\Sigma$ es idempotente, entonces $X^T AX \sim \chi^2_k$.
- Si $W \sim t_{(n)}$, entonces $W^2 \sim F_{(1,n)}$.

1.9 Distribución conjunta de probabilidad

La ***función conjunta de distribución*** para dos variables aleatorias X y Y, denotada por $f\left(x,y\right)$, se define como:

$$f\left(x,y\right)=P\left(a\le x\le b, c\le y\le d\right)=\begin{cases}\int_a^b\int_c^d f\left(x,y\right)dydx & \text{si } X \text{ y } Y \text{ continuas}\\ \sum_{a\le x\le b}\sum_{c\le y\le d} \mathrm{f}\left(x,y\right) & \text{si } X \text{ y } Y \text{ discretas}\end{cases} \quad (20)$$

Toda función conjunta de distribución debe cumplir:

- $f\left(x,y\right)\ge 0$

[5] La distribución F también es conocida como la distribución F de Snedecor en Honor a Gerge W. Snedecor unos de los padres de la estadística moderna.

DO NOT COPY

- $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dydx = 1$ si X y Y son variables aleatorias continuas, o $\sum_{\text{Todo x}}\sum_{\text{Todo y}} \mathrm{f}(\mathrm{x,y}) = 1$ si X y Y son variables aleatorias discretas

Es importante anotar que, en este caso, el valor esperado de cualquier función $g(x,y)$ está dado por:

$$E\left[g(x,y)\right] = \begin{cases} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\, f(x,y)\,dydx & \text{si } X \text{ y } Y \text{ continuas} \\ \sum_{\text{Todo x}}\sum_{\text{Todo y}} g(x,y)\,\mathrm{f}(\mathrm{x,y}) & \text{si } X \text{ y } Y \text{ discretas} \end{cases} \tag{21}$$

Noten que esta idea es fácilmente extensible a más de dos variables aleatorias, en ese caso tendremos que la función de distribución para las variables aleatorias $X_1, X_2, ..., X_n$ será $f(x_1, x_2, ..., x_n)$. Y todos los resultados de las distribuciones bivariadas se pueden extender al caso de n variables aleatorias.

Ejemplo 7. Distribución Normal Bivariada.

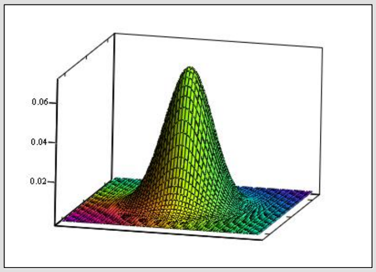
La distribución bivariada normal está definida como

$$f\left(x, y \middle| \mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho\right) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}\, e^{\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_X)^2}{\sigma_X^2} - 2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right]\right\}}$$

donde μ_X y σ_X^2 representan la media y varianza de X, respectivamente. Similarmente, μ_Y y σ_Y^2 corresponde a la media y varianza de Y, respectivamente. Finalmente, ρ corresponde a la correlación entre las variables aleatorias X y Y.

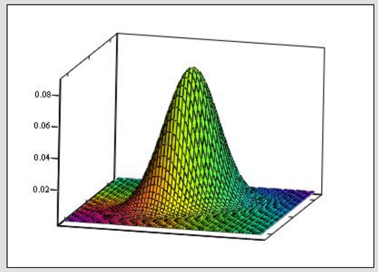
Gráficos de la Distribución Normal Bivariada
con diferentes valores de Correlación

$f\left(x, y | 0, 0, 9/4, 9/4, 0\right)$



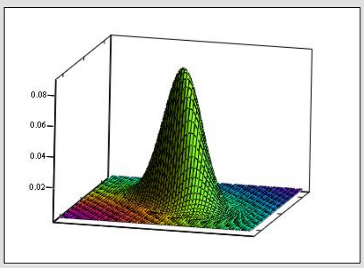
s

$f\left(x, y | 0, 0, 9/4, 9/4, 3/5\right)$



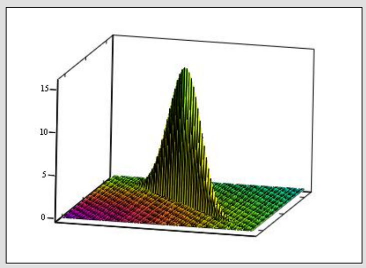
s

$f\left(x, y | 0, 0, 9/4, 9/4, -3/5\right)$



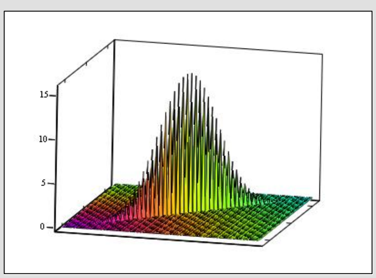
s

$f\left(x, y | 0, 0, 9/4, 9/4, -1\right)$



s

$f\left(x, y | 0, 0, 9/4, 9/4, -1\right)$



s

Ejemplo 8. Distribución Normal Multivariada.

Una generalización de la distribución bivariada normal presentada en el Ejemplo 7 es la distribución Normal Multivariada. Si consideramos un vector aleatorio $\underset{\sim}{X}$ de tamaño n, entonces $\underset{\sim}{X}$ estará distribuido normalmente si y solamente si:

$$f\left(\underset{\sim}{X}\right)=\left(2\pi\right)^{-\frac{n}{2}}\cdot\left|\Sigma\right|^{-1}e^{\left\{-\frac{1}{2}(x-\mu)^{T}\Sigma^{-1}(x-\mu)\right\}}$$

donde Σ es la matriz de varianzas y covarianzas y μ es el vector de medias, respectivamente.

1.10 Distribución marginal y condicional de probabilidad

Sean X y Y dos variables aleatorias con su respectiva función conjunta de distribución $f\left(x,y\right)$. La ***distribución marginal de probabilidad***, o función de distribución marginal, de la variable aleatoria X se define como

$$f_X\left(x\right)=\begin{cases}\int_{-\infty}^{\infty} f\left(x,t\right)dt & \text{si } X \text{ y } Y \text{ continuas}\\ \sum_{\text{Todo y}} \mathrm{f}\left(\mathrm{x,y}\right) & \text{si } X \text{ y } Y \text{ discretas}\end{cases}. \quad (22)$$

Intuitivamente, la distribución marginal se puede entender como la función de distribución de una sola variable aleatoria cuando ignoramos las otras variables aleatorias que hacen parte de una función de distribución conjunta.

Ahora supongamos que queremos saber cuál es la probabilidad de que X sea igual a un valor o esté en un intervalo predeterminado dado que la variable Y es exactamente igual a un valor determinado. En otras palabras queremos conocer la distribución de probabilidad de la variable X condicionada a un valor determinado de la variable Y.

Si consideramos la distribución bivariada normal para las variables estocásticas X y Y ilustrada en el Ejemplo 7, entonces la distribución condicional de X dado que $Y=y$ es equivalente a cortar una "tajada" fina con un cuchillo ubicado paralelo al eje de las x y

DO NOT COPY

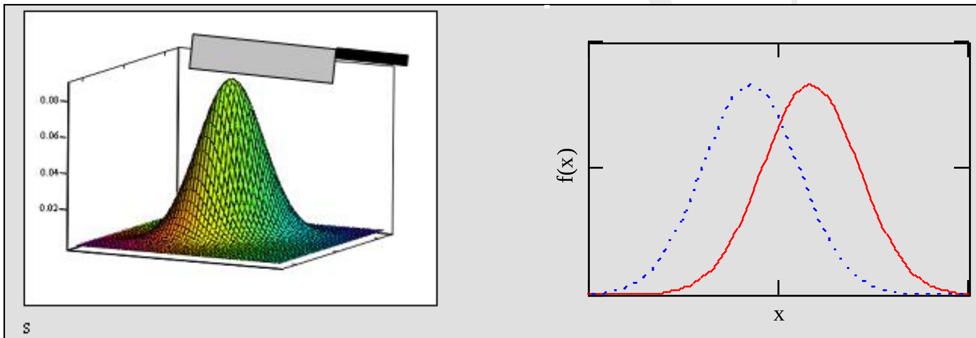
la tajada fina será cortada exactamente en el punto que $Y = y$ (Ver Gráfico 2). Si colocamos esta "tajada" sobre una mesa, encontraremos un gráfico en dos dimensiones de una distribución normal, la media de esta distribución "tajada" dependerá de cuál es la altura a la que se corta la distribución bivariada, es decir el valor $Y = y$.

Formalmente, la ***distribución condicional de probabilidad*** , o función de distribución condicional, de la variable aleatoria X dado $Y = y$ se define como:

$$f\left(x \middle| Y = y\right) = f\left(x \middle| y\right) = \frac{f\left(x, y\right)}{f_Y\left(y\right)} \tag{23}$$

COPY

Gráfico 2. Función de distribución condicional Normal.



Panel a. **Panel b.**

En el **Panel a**, se observa una distribución bivariada normal $f\left(x, y \middle| 0, 0, 9/4, 9/4, 3/5\right)$. Imagine que corta con un cuchillo una "fina tajada" de esta "montaña" colocando el cuchillo paralelo al eje de las x y exactamente a la altura de $Y = 2$. Esta "fina tajada" corresponderá a la línea sólida del **Panel b** (en este caso $X \middle| y = 2 \sim N\left(1.2, 1.44\right)$). Si se realiza un corte similar, pero a la altura de $Y = -1$, obtendremos la función de densidad punteada mostrada en el **Panel b** (en este caso $X \middle| y = -1 \sim N\left(-.6, 1.44\right)$).

1.11 Valor esperado y varianza condicional.

El ***valor esperado condicional***, o media condicional, de una variable aleatoria X es el valor esperado de una distribución condicional y se define como:

$$\mu_{X|y} = E\left[X \middle| Y = y\right] = E\left[X \middle| y\right] = \begin{cases} \int_{-\infty}^{\infty} xf\left(x|y\right)dx & \text{si } X \text{ y } Y \text{ continuas} \\ \sum_{\text{Todo } x_i} x_i f\left(x_i|y\right) & \text{si } X \text{ y } Y \text{ discretas} \end{cases}. \qquad (24)$$

Similarmente, la ***varianza condicional*** de una variable aleatoria X es la varianza de una distribución condicional y se define como:

$$\begin{aligned} \sigma^2_{X|y} &= Var\left[X \middle| Y = y\right] = Var\left[X \middle| y\right] = E\left[\left(X - \mu_{X|y}\right)^2 \middle| y\right] \\ &= E\left[X^2 \middle| y\right] - \left(E\left[X \middle| y\right]\right)^2 \end{aligned} \qquad (25)$$

1.12 Estimadores puntuales y sus propiedades deseadas.

Intuitivamente, un estimador se puede entender como una “fórmula” que permite pronosticar un valor poblacional (parámetro) desconocido a partir de una muestra. Por ejemplo, supongamos que deseamos conocer la media de una población. Regularmente no conocemos este valor y por tanto se recolectan observaciones de parte de la población total (muestra), y a partir de estas observaciones evaluamos una fórmula para conocer nuestro pronóstico del valor poblacional real.

Formalmente un ***estimador***, también conocido como estimador puntual, de un parámetro poblacional es una función que indica cómo calcular una matriz, vector o escalar a partir de una muestra. El valor arrojado por esta función una vez los valores muestrales son reemplazados en el estimador se denomina ***estimación***.

Así, un estimador $\hat{\theta}$ para pronosticar un parámetro θ a partir de una muestra aleatoria de tamaño n se define como:

DO NOT COPY

$$\hat{\theta} = h\left(X_1, X_2, ..., X_n\right) \tag{26}$$

donde $h(\bullet)$ es una función cualquiera y $X_1, X_2, ..., X_n$ corresponden a cada uno de los n puntos muestrales. Es importante notar que los estimadores son variables aleatorias, pues son función de variables aleatorias.

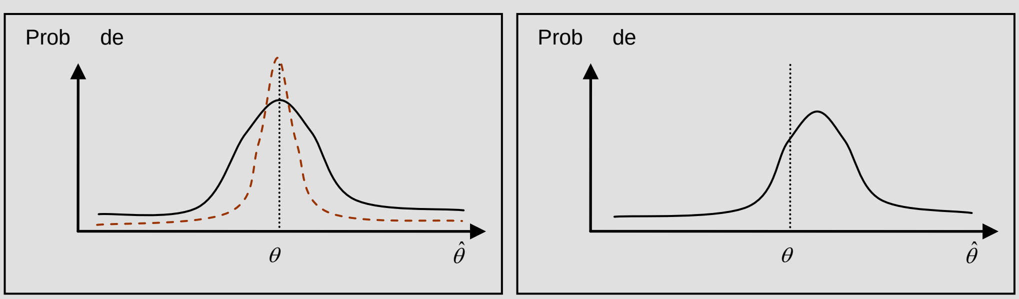
Claramente cualquier función de los puntos muestrales por definición es un estimador. Pero, ¿cómo escoger cuál función de la muestra es un buen estimador para el parámetro deseado? Existen varias propiedades deseadas en los estimadores que discutiremos a continuación.

Una propiedad muy deseable es que el valor esperado de la distribución del estimador esté en promedio lo más cercano o coincida con el valor población del parámetro. De esta forma, cada vez que se analice información nueva se estará seguro que en promedio el estimador estará correcto. En general, diremos que un estimador es ***insesgado*** si $E\left[\hat{\theta}\right] = \theta$. Así definiremos el sesgo de un estimador como:

$$Sesgo\left[\hat{\theta}\right] = E\left[\hat{\theta}\right] - \theta \tag{27}$$

La insesgadez es una propiedad deseable en un estimador, pero como lo ilustra el panel a) del Gráfico 3 la ausencia de sesgo no dice nada sobre la dispersión que tiene el estimador alrededor de su media.

Gráfico 3. Sesgo de un Estimador.



a) Estimadores Insesgados **b)**Estimador Sesgado

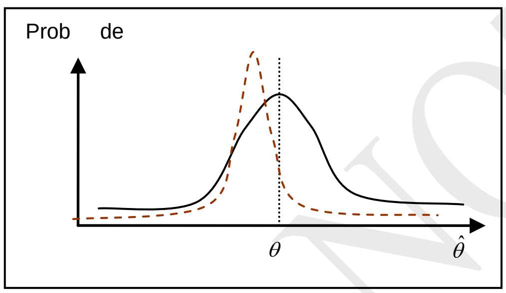
DO NOT COPY

Como es de esperarse, se preferirá un estimador que tenga una menor dispersión alrededor de la media (varianza) a uno con mayor dispersión. Un estimador $\hat{\theta}_1$ es considerado un ***estimador insesgado más eficiente*** que $\hat{\theta}_2$ si

$$\operatorname{Var}\left(\hat{\theta}_1\right) < \operatorname{Var}\left(\hat{\theta}_2\right) \tag{28}$$

Ahora consideremos el caso en que estamos comparando un estimador sesgado con una varianza relativamente pequeña con un estimador insesgado con una varianza relativamente grande (Ver Gráfico 4). La pregunta es: ¿cuál de los dos estimadores deberá ser preferido?

Gráfico 4. Sesgo versus Mínima Varianza.



Un criterio para escoger un estimador entre otros, es considerar el estimador con el mínimo Error Medio al Cuadrado, denotado MSE por su nombre en inglés ("Mean Square Error"), éste se define como:

$$MSE\left(\hat{\theta}\right) = E\left(\hat{\theta} - \theta\right)^2. \tag{29}$$

Es fácil mostrar que:

$$MSE\left[\hat{\theta}\right] = \left[Sesgo\left[\hat{\theta}\right]\right]^2 + Var\left[\hat{\theta}\right] \tag{30}$$

Así al minimizar el MSE, se está teniendo en cuenta tanto el sesgo como la dispersión del estimador.

DO NOT COPY

Ejemplo 9

Tal vez el ejemplo más familiar es el estimador de la media. Suponga que queremos estimar la media poblacional μ; el estimador más empleado para la media poblacional es $h(X_1, X_2, ..., X_n) = \frac{\sum_{i=1}^{n} X_i}{n}$, este estimador comúnmente se conoce como $\bar{x}$. Noten que $E[\bar{x}] = E\left[\frac{\sum_{i=1}^{n} X_i}{n}\right] = \frac{1}{n}E\left[\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i]$, dado que todas las X_i provienen de la misma población, entonces $E[X_i] = \mu$ para todo i. Así $E[\bar{x}] = \frac{1}{n}\sum_{i=1}^{n} \mu = \mu$. Por tanto $\bar{x}$ es insesgado.

Finalmente, otra propiedad deseada en un estimador es la consistencia. Intuitivamente, un estimador es consistente si cuando la muestra se hace grande y más cercana a la población total, entonces la probabilidad de que el estimador sea diferente del valor poblacional θ es cero. Formalmente, $\hat{\theta}$ es un estimador ***consistente*** si

$$\lim_{n \to \infty} P\left(\left|\theta - \hat{\theta}\right| < \delta\right) = 1 \qquad (31)$$

donde δ es una constante positiva arbitrariamente pequeña.

1.13 Teorema del límite Central

El teorema del límite central puede ser uno de los resultados más poderosos y asombrosos de la ciencia estadística. La intuición detrás de este teorema es muy sencilla; si se suman un número suficientemente grande de variables aleatorias, entonces sin importar la distribución de las variables aleatorias sumadas, la sumatoria seguirá una distribución normal.

Existen numerosas versiones del teorema del límite central, uno con restricciones más fuertes que otras, pero la versión más sencilla de este teorema es:

$$\sum_{i=1}^{n} X_i \sim^a N\left(n\mu, n\sigma^2\right) \tag{32}$$

donde μ y σ^2 son la media y la varianza poblacional, respectivamente. Una versión más conocida de este teorema es:

$$\bar{x} = \frac{\sum_{i=1}^{n} X_i}{n} \sim^a N\left(\mu, \frac{\sigma^2}{n}\right) \tag{33}$$

1.14 Intervalos de confianza y pruebas de hipótesis.

La probabilidad de que una estimación puntual sea exactamente igual al valor real del parámetro es cero (¿Por qué?). Para aumentar la certidumbre entorno a nuestra estimación, podemos emplear la distribución de probabilidad del estimador para ampliar nuestra estimación a un intervalo o probar diferentes hipótesis.

Para aquellos estimadores que siguen una distribución simétrica como la t o la normal, la estructura de un intervalo de confianza es muy sencilla. La idea es crear un intervalo con una confianza del $100(1-\alpha)\%$ que tenga como centro la estimación y cuyos límites inferior y superior sigan la estructura:

$$Estimacion \mp \left(\text{Valor de una distribucion}\right)_{\frac{\alpha}{2}} \sqrt{Var\left[Estimador\right]} \tag{34}$$

donde el valor de una distribución depende de la distribución del estimador y del nivel de confianza α deseado. En general un intervalo de confianza de $100(1-\alpha)\%$ no significa que con un $100(1-\alpha)\%$ de certeza el valor real del parámetro estará en el intervalo. La interpretación de un intervalo de confianza del $100(1-\alpha)\%$ es que de 100

DO NOT COPY

muestras que se generen de la misma población en $100(1-\alpha)$ veces las estimaciones estarán dentro de dicho intervalo.

En cuanto a las pruebas de hipótesis en torno a parámetros, éstas presentan una estructura similar (Ver Esquema 1.). Ahora consideremos rápidamente uno de los aspectos más importantes de una prueba de hipótesis, los tipos de errores involucrados en cada decisión. En especial, cuando escogemos un nivel de significancia α, y, por ejemplo, rechazamos la hipótesis nula, en ese caso es posible que incorrectamente rechacemos la hipótesis nula cuando ésta es verdadera; a este error lo llamamos ***error tipo I*** y la probabilidad que este ocurra será del α%. Supongamos ahora que no es posible rechazar la hipótesis nula, en este caso es posible que no estemos rechazando la hipótesis nula cuando ésta en verdad es falsa; este error recibe el nombre de ***error tipo II*** y la probabilidad de que este ocurra es de β. Por tanto, lo ideal al diseñar una prueba de hipótesis es que tanto α como β sean lo más pequeño posible. Lastimosamente existe un compromiso entre la probabilidad de cometer el error tipo I y el error tipo II, pues cuando se disminuye el error tipo I, el error tipo II se aumenta.

Esquema 1. Estructura de una prueba de Hipótesis

1. H_0: Hipótesis Nula (hipótesis que se quiere refutar)
2. H_A: Hipótesis Alterna (hipótesis que se quiere aceptar)
3. Cálculo de un estadístico (depende de la distribución del estimador)
4. Decisión (Comparar el estadístico calculado con un valor crítico de la correspondiente función de distribución)

Es por esta razón, que siempre que se plantea una prueba de hipótesis, se trata de construir las hipótesis nula y alterna de tal forma que lo que se desea comprobar sea planteado en la hipótesis alterna y no en la hipótesis nula. Así, se pretende cometer un error tipo I que es más fácilmente controlable por el investigador.

DO NOT COPY

1.15 Referencias.

- Amemiya, Takeshi. 1994. *Introduction to Statistics and Econometrics*. London: Harvard University Press.

- Mood, Alexander McFarlane. 1950. *Introduction to the theory of statistics*: McGraw-Hill.

1.16 Ejercicios.

1. Dos dados son lanzados al mismo tiempo sobre una mesa con superficie nivelada; un dado es de color azul y el otro de color rojo. Sean X el número en la cara superior del dado azul después de lanzado multiplicado por 3, Y el número en la cara superior del dado rojo después de lanzado dividido por 3, Z la suma del número en la cara superior de cada uno de los dos dados después de ser lanzados, y finalmente sea $W = XZ$. Encuentre:

   $E(X)$, $E(Y)$, $E(Z)$, y $E(W)$.

   $Var(X)$, $Var(Y)$, $Var(Z)$, y $Var(W)$

   ¿Son W y E independientes?

2. Una distribución discreta muy usada es la distribución Poisson. Por ejemplo, suponga que X es el número de personas que se presentan en una ventanilla de atención de un banco en un período de tiempo dado escogido aleatoriamente. Entonces un modelo frecuentemente usado es

   a. $f(x) = \dfrac{e^{-\lambda}\lambda^{x}}{x!}, \qquad x = 0,1,2,\ldots$

3. Demuestre que la distribución Poisson cumple las condiciones $0 \le f(x) \le$ y $\sum_{todo\ x_i} f(x_i) = 1$

4. Muestre que $Var[X] = E\left[X^2\right] - \left(E[X]\right)^2$.

5. Muestre que $Var[aX + b] = a^2 Var[X]$.

6. Muestre que $Var[aX + bY] = a^2 Var[X] + b^2 Var[Y] + 2abCov[X,Y]$.

DO NOT COPY

7. Muestre que $Cov[X,Y]=E[XY]-E[X]E[Y]$.
8. Muestre que $-1\leq\rho\leq 1 \quad 0\leq\rho\leq 1$.
9. Muestre que $Var[\underset{\sim}{X}]=E[\underset{\sim}{X}\underset{\sim}{X}^T]-\mu\mu^T$.
10. Muestre que si $Y\sim\chi^2_{(1)}$, Entonces $E[Y]=1$ y $Var[Y]=2$.
11. Muestre que $MSE[\hat{\theta}]=[Sesgo[\hat{\theta}]]^2+Var[\hat{\theta}]$.
12. Encuentre $Var[\bar{x}]$.

DO NOT COPY