



GOBIERNO
DE ESPAÑA

MINISTERIO
DE HACIENDA
Y ADMINISTRACIONES PÚBLICAS

INSTITUTO DE ESTUDIOS FISCALES

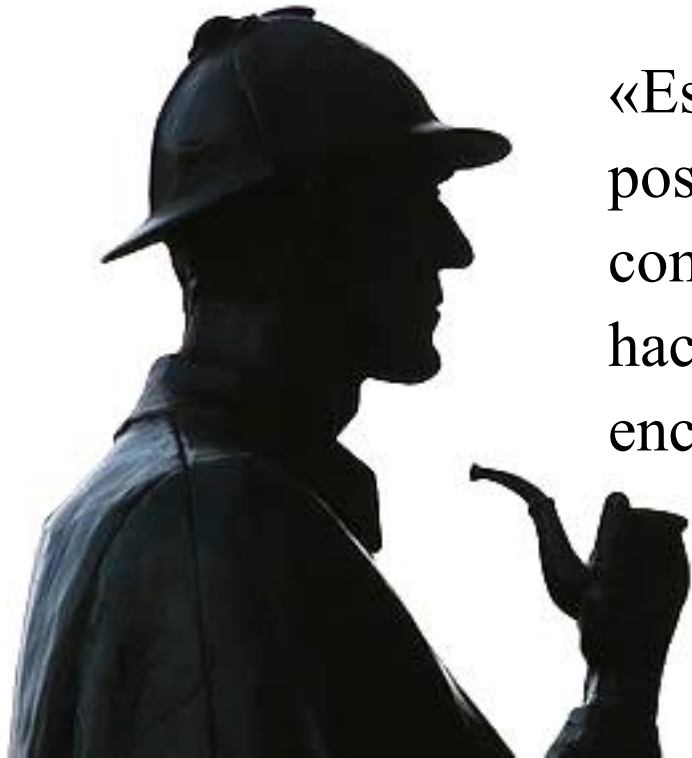


Evaluación de Políticas Públicas

EUROsociAI II

Técnicas cuantitativas de evaluación de políticas públicas

Lima, 9 de septiembre de 2013



«Es un error capital el teorizar antes de poseer datos. Insensiblemente, uno comienza a deformar los hechos para hacerlos encajar en las teorías en lugar de encajar las teorías en los hechos».

La evaluación

Definición de evaluación

Noción de lógica de la intervención

- ▶ Reconstrucción de la lógica de la intervención
- ▶ Utilidad de la lógica de la intervención

Noción de efecto

Noción de impacto

Definición de evaluación

La evaluación es el “proceso sistemático de observación, medida, análisis e interpretación encaminado al conocimiento de una intervención pública, sea esta una norma, programa, plan o política, para alcanzar un juicio valorativo basado en evidencias y criterios referenciales, respecto de su diseño, puesta en práctica, efectos, resultados e impactos”, de acuerdo con la definición que nos aporta la Agencia de Evaluación y Calidad española.

Atribución: imputación de un vínculo causal entre cambios observados (o que se espera observar) y una intervención específica

Noción de lógica de la intervención

Se trata del conjunto de actividades puestas en práctica, de los efectos esperados (productos, resultados e impactos) así como de los supuestos que explican cómo las actividades van a conducir a los efectos en el contexto de la intervención.

Con frecuencia, la lógica de una intervención evoluciona a través del tiempo. Llegado el caso, ello justifica la reconstrucción de varias lógicas correspondientes a períodos sucesivos.

Reconstrucción de la lógica de la intervención

Recabar los documentos oficiales de la intervención.

Identificar las principales actividades.

Identificar los objetivos.

Traducir los objetivos para que adopten la forma de resultados e impactos esperados.

Conectar las actividades con los impactos esperados mediante relaciones causa-efecto lógicas.

Debatir la lógica con expertos externos y con gestores internos.

Utilidad de la lógica de la intervención

Ayudar a aclarar los objetivos y a traducirlos en una jerarquía de efectos esperados de tal manera que ello permita su evaluación.

Sugerir preguntas de evaluación sobre dichos efectos.

Ayudar a evaluar la coherencia interna de la intervención.

Noción de efecto

Efecto es todo comportamiento o acontecimiento del que puede razonablemente decirse que ha sido influido por algún aspecto del programa o proyecto.

Efecto \neq Objetivo. Los objetivos constituyen el estado deseado que se pretende alcanzar con la realización del proyecto, se ubican temporalmente antes de su realización y son fijados por sus diseñadores. En cambio, los efectos constituyen resultados de las acciones llevadas a cabo por el proyecto y se verifican durante o después del mismo.

Noción de impacto

El impacto se define como un resultado de los efectos de un proyecto. La determinación del impacto exige el establecimiento de objetivos operacionales y de un modelo causal que permita vincular el proyecto con los efectos resultantes de su implementación.

El impacto expresa el grado de cumplimiento de los objetivos respecto a la población-meta del proyecto.



«Watson, no hay que suponer si no tenemos evidencia».

Diseño de experimentos

¿Qué es el diseño de experimentos?

Preguntas que formulan los experimentos

Causas, efectos y relaciones causales

Cuasiexperimentos

Validez de las conclusiones

Diseño de experimentos

Diseñar estadísticamente un experimento es realizar una prueba para poder caracterizar las variables explicativas de mayor influencia a través de los efectos en varias variables respuesta, de forma que se minimice el efecto de las variables no controlables (covariables) y se estabilice y minimice la variabilidad de las respuestas.

Preguntas que formulan los experimentos

Descriptivas: recuento de los valores concretos de ciertas variables en un conjunto de observaciones.

Relacionales: ¿existen relaciones entre fenómenos, y en qué sentido y magnitud se da la covariación?

Causales: ¿una o más variables independientes tienen efectos sobre una o más variables dependientes? ¿Cómo se obran estos efectos?

Causas, efectos y relaciones causales

En los experimentos:

- ▶ Las supuestas causas se manipulan para observar los efectos.
- ▶ La variabilidad en la causa se relaciona con la variación en el efecto.
- ▶ El diseño y el conocimiento acumulado se usan para identificar y descartar explicaciones alternativas plausibles.

Causas manipulables y no manipulables

Los experimentos implican agentes casuales que pueden ser manipulados por el investigador.

Las causas no manipulables (origen étnico, desastres naturales) no pueden ser causas en los experimentos ya que no están sujetos a la manipulación del investigador.

Causas: descripción y explicación

El primer paso consiste en identificar que existe una relación causal entre A y B (*molar causation*): relación general entre un tratamiento y sus efectos.

El segundo paso es explicar cómo A causa B (*molecular causation*): llegar a saber con detalle qué aspectos del tratamiento son responsables de qué aspectos de un efecto.

Cuasiexperimentos

El término cuasiexperimento se refiere a diseños de investigación experimentales en los cuales los sujetos o grupos de sujetos de estudio no están asignados aleatoriamente.

Los cuasiexperimentos son más vulnerables a las amenazas a la validez que las pruebas aleatorias.

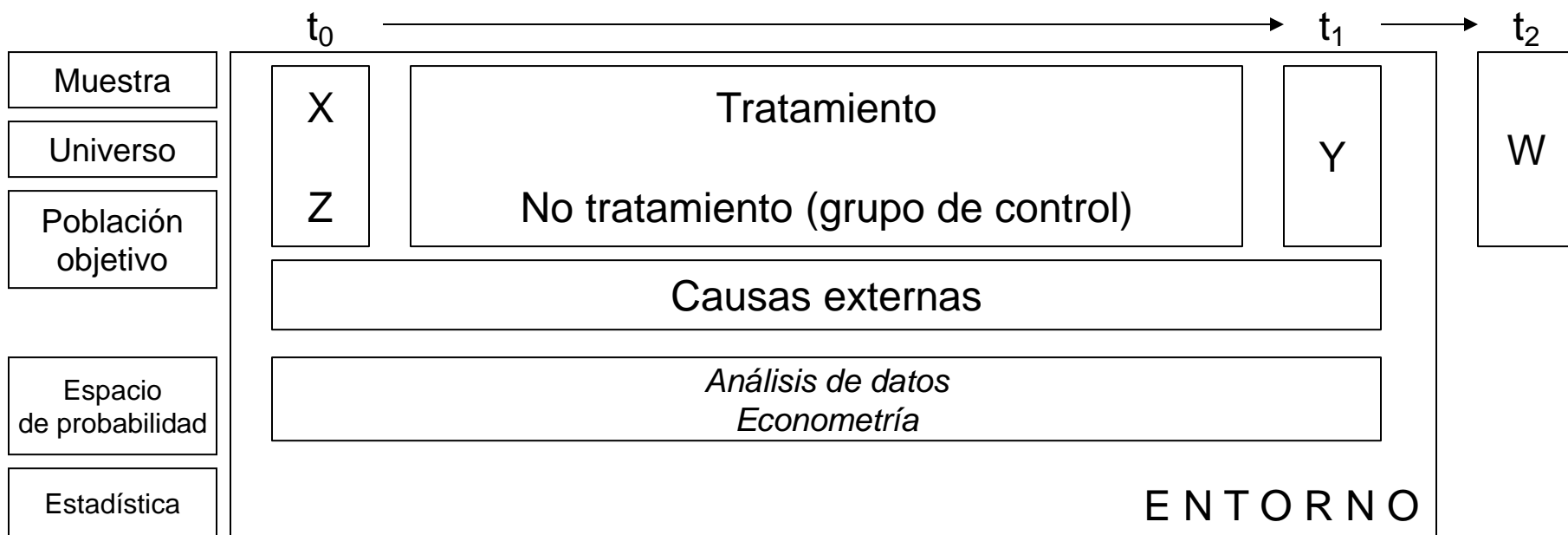
Cuasiexperimentos

Cuando se diseñan, controlan y analizan apropiadamente, los cuasiexperimentos pueden ofrecer una evidencia casi tan fuerte del impacto del programa como la de las pruebas aleatorias.

El cuasiexperimento es más vulnerable a los sesgos de selección: el grupo de tratamiento puede diferir del grupo control en características no observadas correlacionadas con los resultados, lo que distorsiona el impacto: problema de la heterogeneidad no observada.

Cuasiexperimentos

Técnicas de evaluación de impacto: el problema del contrafactual



X: variables exógenas
Z: variables instrumentales

Y, W: variables endógenas
Y: resultados
W: impactos

t_0 : inicio del tratamiento
 t_1 : corto plazo
 t_2 : medio o largo plazo

Ejercicio: el programa PIPE del ICEX

Lea el folleto informativo sobre el programa PIPE.

Reconstruya la lógica de la intervención del programa.

Esboce el diseño experimental para evaluar el impacto del programa PIPE.

En su opinión, ¿se trata de un experimento o de un cuasiexperimento?

- ▶ Fichero: *Folleto_PIPE.pdf*
- ▶ Tiempo: 15 minutos
- ▶ Ejercicio en grupo

Validez de las conclusiones

Un experimento o cuasiexperimento genera conclusiones que, bien validadas e interpretadas, pueden transformarse en conocimiento.

La validez de una conclusión es su grado de fidelidad y de corrección. Hay varios tipos de validez: estadística, interna, estructural y externa. Incrementar la validez de un tipo implica a menudo reducir la validez de otros tipos.

Validez estadística

La validez estadística de las conclusiones se centra en la covariación entre el tratamiento (causa) y el resultado (efecto).

- ▶ ¿Ocurren A y B de verdad a la vez?
- ▶ ¿Cuál es la magnitud de la covarianza?
- ▶ ¿Es significativa la conclusión?

La importancia real se halla en estimar el tamaño del efecto en relación a los intervalos de confianza más que en demostrar la significación estadística de la hipótesis nula (p -valor).

Validez interna

La validez interna de las conclusiones se centra en si la covarianza observada entre A y B refleja una relación causal.

- ▶ ¿Ha sido de verdad A la causa de B o ha sido otro factor C el responsable?
- ▶ Contrafactual: asignación aleatoria al grupo de tratamiento y al grupo de control.

La importancia reside en probar que la relación causal se debe en exclusiva a la manipulación o la variación de las variables medidas. El mayor riesgo es el sesgo de selección, presente por lo general en cuasiexperimentos.

Validez estructural

La validez estructural de las conclusiones se centra en las estructuras de orden superior que están representados por aspectos o elementos particulares presentes en la muestra, es decir: entre el concepto teórico y la medida empírica que se usa.

- ▶ ¿Hemos medido A en realidad u otro A' en su lugar?
- ▶ ¿Cómo se mide el carácter emprendedor de una empresa?
- ▶ Conceptualización clara: se define un concepto o variable de forma que pueda ser medido o que pueda dejar constancia de haber sufrido un cambio (efecto).

Validez externa

La validez externa de las conclusiones se centra en si una relación causa-efecto sigue siendo cierta aunque varíen los individuos, el entorno, los tratamientos y los resultados.

- ▶ ¿Hay otras cosas parecidas a A que provoquen cosas como B?
- ▶ Diseño de muestreo: factor de elevación a población
- ▶ Generalización por extrapolación

Bibliografía

HM Treasury Magenta book

<https://www.gov.uk/government/publications/the-magenta-book>

Impact Evaluation in Practice

<http://issuu.com/world.bank.publications/docs/9780821385418>

Glosario de evaluación de la OECD

http://www.oecd.org/findDocument/0,2350,en_2649_34435_1_119678_1_1_1_00.html

Páginas de evaluación de las Relaciones Exteriores de la Unión Europea

http://ec.europa.eu/europeaid/how/evaluation/index_es.htm

Técnicas cuantitativas de evaluación de políticas públicas

Lima, 9 de septiembre de 2013



«Cuando se excluye lo imposible, lo que queda, por improbable que sea, debe ser la verdad».

Nociones de probabilidad

Población y muestra

Espacio de probabilidad

Variable aleatoria discreta y continua

Esperanza matemática y momentos de una variable aleatoria

Vectores aleatorios

Población y muestra

La población objetivo es el grupo de todos los individuos del que el investigador está interesado en extraer conclusiones válidas mediante extrapolación

La población o universo es la colección completa, finita o infinita, de individuos distintos y perfectamente identificables sin ambigüedad que podemos estudiar y de la que podemos obtener datos.

Una muestra es un grupo de individuos seleccionados de la población y representativo de esta.

Población y muestra

Ejemplo 1: Estudio de salud infantil

Universo: Todos los niños nacidos en el Reino Unido en 1980.

Muestra: Todos los niños nacidos el 7 de mayo en cualquiera de los años del período.

Población objetivo: Todos los niños del mundo desarrollado.

Población y muestra

Ejemplo 2: Programa de becas para pymes

Universo: las empresas beneficiarias del programa.

Muestra: una muestra aleatoria simple del universo.

Población objetivo: todas las pymes del país.

Espacio de probabilidad

Un fenómeno o experimento aleatorio es un fenómeno que cumple las siguientes propiedades:

- ▶ Las mismas condiciones iniciales pueden dar lugar a diferentes resultados.
- ▶ Todos los resultados posibles se conocen de antemano.
- ▶ No se puede predecir el resultado de cada experiencia particular.
- ▶ Puede repetirse en las mismas condiciones de forma indefinida.
- ▶ Si se repite bajo condiciones idénticas un número determinado de veces n , y anotamos el número de veces que aparecerá cada resultado n_i , el cociente n_i/n tiende a estabilizarse a un valor fijo conforme aumenta n .

$$\forall i, \quad \exists \lim_{n \rightarrow \infty} \frac{n_i}{n} = p_i$$

Espacio de probabilidad

Dado un experimento aleatorio, llamaremos espacio muestral Ω al conjunto de todos los resultados posibles de dicho experimento aleatorio. Los elementos de Ω se denominan sucesos elementales.

Asociados a un mismo experimento aleatorio pueden considerarse diferentes espacios muestrales.

Espacio de probabilidad

Sea \mathcal{A} una colección no vacía de subconjuntos del espacio muestral Ω . Llamaremos suceso a un conjunto A de \mathcal{A} . Diremos que se ha presentado el suceso A al realizar el experimento si el resultado de dicho experimento es algún punto $\omega \in A$.

Espacio de probabilidad

Dado el espacio muestral Ω , una clase $\mathcal{A} \subset \mathcal{P}(\Omega)$ tiene estructura de σ -álgebra si y solo si

- ▶ $\Omega \in \mathcal{A}$
- ▶ $\forall A \in \mathcal{A}$ es $A^* \in \mathcal{A}$, donde A^* es el complementario de A .
- ▶ \forall toda sucesión $\{A_n\}_{n \in \mathbb{N}} \subset \mathcal{A}$ es $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$

La σ -álgebra es cerrada por uniones numerables, intersecciones numerables, diferencias, complementarios y paso al límite para sucesiones monótonas.

Espacio de probabilidad

Al par $\{\Omega, \mathcal{A}\}$, donde $\mathcal{A} \subset \mathcal{P}(\Omega)$ es una σ -álgebra de subconjuntos de Ω , se le denomina espacio medible o probabilizable. A los elementos de \mathcal{A} se les denomina conjuntos medibles.

Espacio de probabilidad

Axiomática de Kolmogorov

Sea $\{\Omega, \mathcal{A}\}$ un espacio probabilizable. Definimos una función de conjunto P , que va ser una medida normada sobre \mathcal{A} , mediante una aplicación de \mathcal{A} en \mathbf{R} que cumple los siguientes axiomas:

- ▶ $\forall A \in \mathcal{A}$ es $P(A) \geq 0$
- ▶ $P(\Omega)=1$
- ▶ \forall toda sucesión $\{A_n\}_{n \in \mathbf{N}} \subset \mathcal{A}$ tal que $A_i \cap A_j = \Phi$, $\forall i \neq j$, es

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$$

Espacio de probabilidad

El trío $\{ \Omega, \mathcal{A}, P \}$, donde Ω es un espacio muestral, \mathcal{A} una σ -álgebra de sucesos sobre Ω y P una medida de probabilidad sobre \mathcal{A} recibe el nombre de espacio probabilístico o espacio de probabilidades. A los elementos de \mathcal{A} se les llama sucesos.

Variable aleatoria

Sea $\{\Omega, \mathcal{A}, P\}$ un espacio probabilístico y sea $\{\mathbf{R}, \mathbf{B}\}$ el espacio probabilizable formado por la recta real \mathbf{R} y la σ -álgebra de conjuntos de Borel $\mathbf{B} = \mathbf{B}(\mathbf{R})$.

Una aplicación $X: \Omega \rightarrow \mathbf{R}$ es una variable aleatoria (v.a.) si y solo si $X^{-1}(B) \in \mathcal{A}$, $\forall B \in \mathbf{B}$.

Variable aleatoria

La variable aleatoria X induce una medida de probabilidad P_X sobre $\{\mathbf{R}, \mathbf{B}\}$ que lo convierte en el espacio probabilístico $\{\mathbf{R}, \mathbf{B}, P_X\}$, donde:

$$P_X(B) = P\{X^{-1}(B)\} = P(A), \quad \forall B \in \mathbf{B}, \text{ donde } X(A) = B$$

Variable aleatoria

De forma intuitiva, una variable aleatoria es una función real definida sobre el espacio muestral Ω que conserva la probabilidad asociada al experimento aleatorio.

Diremos que se ha definido una variable aleatoria o que se ha construido un modelo de distribución de probabilidad cuando se especifican los posibles valores de la variable con sus probabilidades respectivas.

Variable aleatoria discreta

Diremos que una variable aleatoria X es discreta cuando toma un número de valores finito o infinito numerable, es decir: X puede tomar cualquiera de los valores $x = 1, 2, 3, 4, \dots$

Ejemplos: el número de piezas defectuosas que aparecen en un proceso de fabricación, el número de contactos que ha hecho una empresa en una feria, etc.

Variable aleatoria discreta

La función dada por $f(x) = P(X = x)$ (la probabilidad de que la variable aleatoria X tome el valor x), se llama distribución de probabilidad, función de probabilidad o función de cuantía de una variable aleatoria discreta. $f(x)$ debe satisfacer las siguientes condiciones:

- ▶ $f(x) \geq 0, \quad \forall x \in \text{dom}(X)$

- ▶ $\sum_{x \in \text{dom}(X)} f(x) = 1$

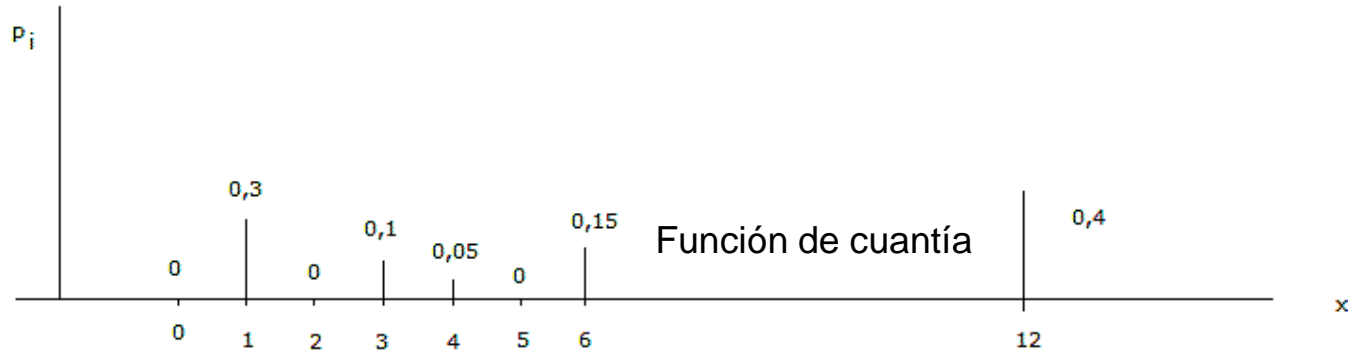
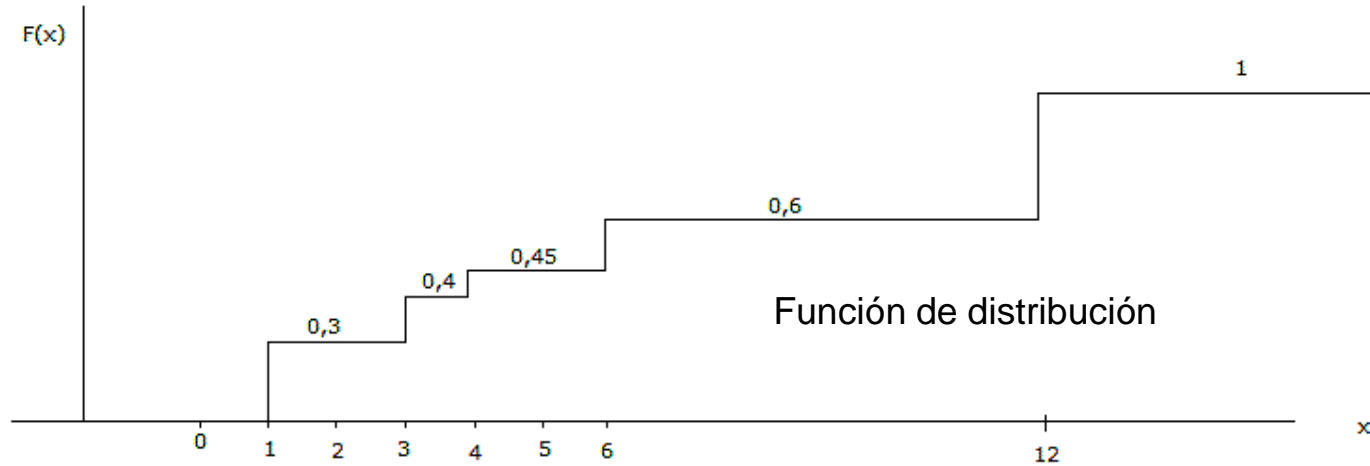
Variable aleatoria discreta

Si X es una variable aleatoria discreta, la función dada por:

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t), \quad -\infty < x < +\infty$$

donde $f(t)$ es el valor de la distribución de probabilidad de X en t , es llamada función de distribución acumulativa de la variable aleatoria X .

Variable aleatoria discreta



Variable aleatoria continua

Una función $f: \mathbf{R} \rightarrow \mathbf{R}$ se llama función de densidad sobre \mathbf{R} si cumple las siguientes condiciones:

- ▶ $f(x) \geq 0, \quad \forall x \in \mathbf{R}$
- ▶ f admite a lo sumo un número finito de discontinuidades sobre cada intervalo finito de \mathbf{R} .
- ▶ $\int_{-\infty}^{\infty} f(x) dx = 1$

Variable aleatoria continua

Una variable aleatoria se dice continua si su función de distribución $F(x)$ puede representarse para cada $x \in \mathbf{R}$ por:

$$F(x) = \int_{-\infty}^x f(t) dt, \quad -\infty < x < +\infty$$

donde $f(t)$ es una función de densidad sobre \mathbf{R} . A esta función se la llama función de densidad de la variable aleatoria continua X .

Variable aleatoria continua

Una variable aleatoria es continua cuando puede tomar cualquier valor en un intervalo. Por ello, no es posible asignar una probabilidad a cada uno de los infinitos posibles valores de la variable aleatoria, sino que se asigna a intervalos o rangos de valores.

Ejemplos: longitud de las piezas defectuosas que aparecen en un proceso de fabricación, la cifra de exportación de una empresa, la renta anual de las personas de una población, etc.

Variable aleatoria continua

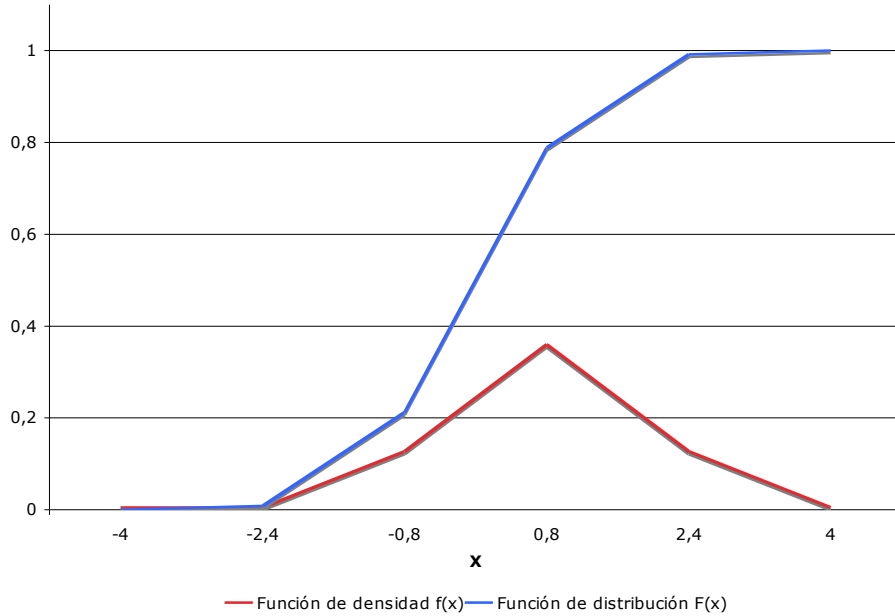
El histograma de una variable aleatoria continua depende del número de observaciones que tomemos. Conforme aumentan las observaciones y se hacen más finos los intervalos, el histograma tenderá a una curva suave y diferenciable.

La derivada en cada punto a esta función límite nos dará la función de densidad de la variable aleatoria.

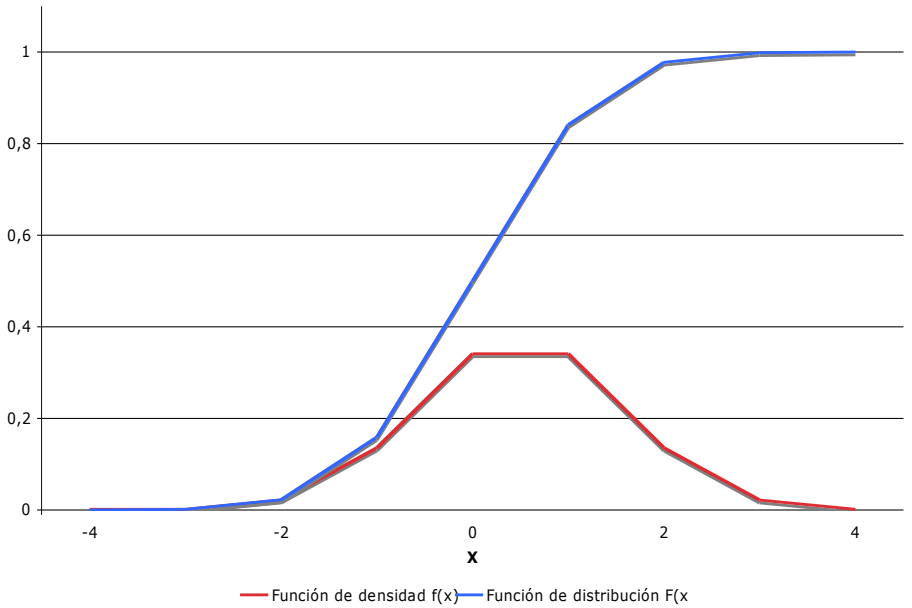
Variable aleatoria continua

La distribución de probabilidad de la variable aleatoria X se describe mediante una curva de densidad. La probabilidad de cualquier suceso es el área por debajo de la curva de densidad y a lo largo de los valores de X que definen el suceso.

Aproximación a la distribución normal

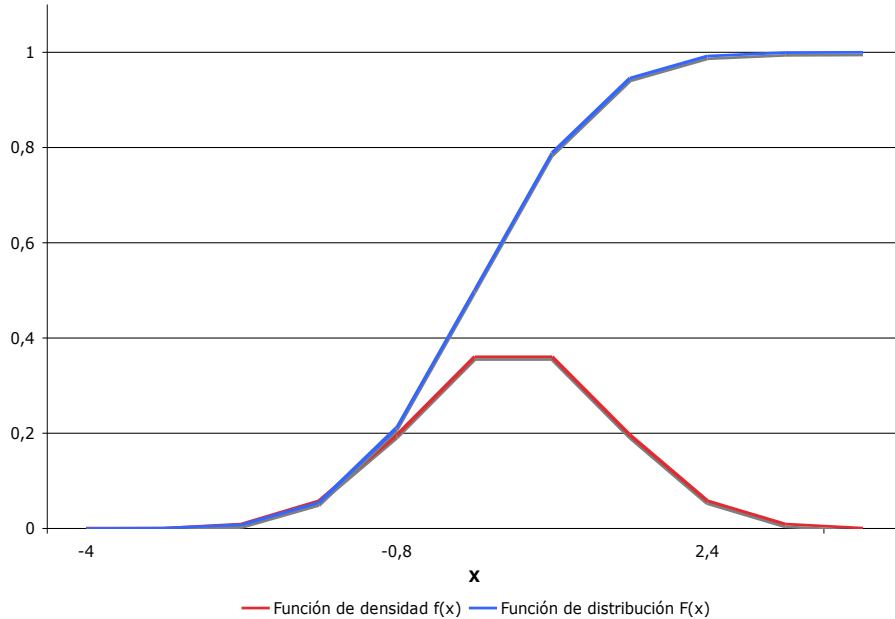


$h=1,6$

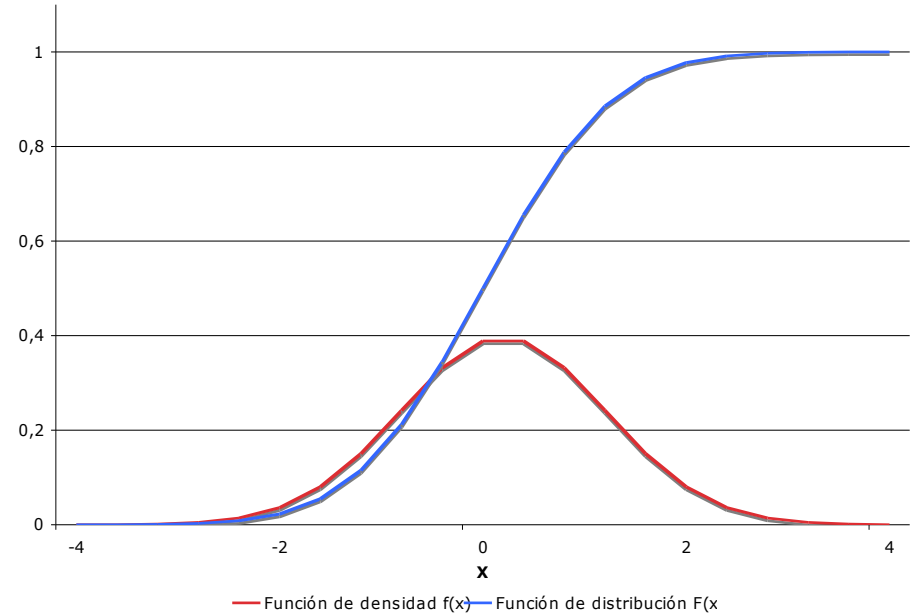


$h=1,0$

Aproximación a la distribución normal

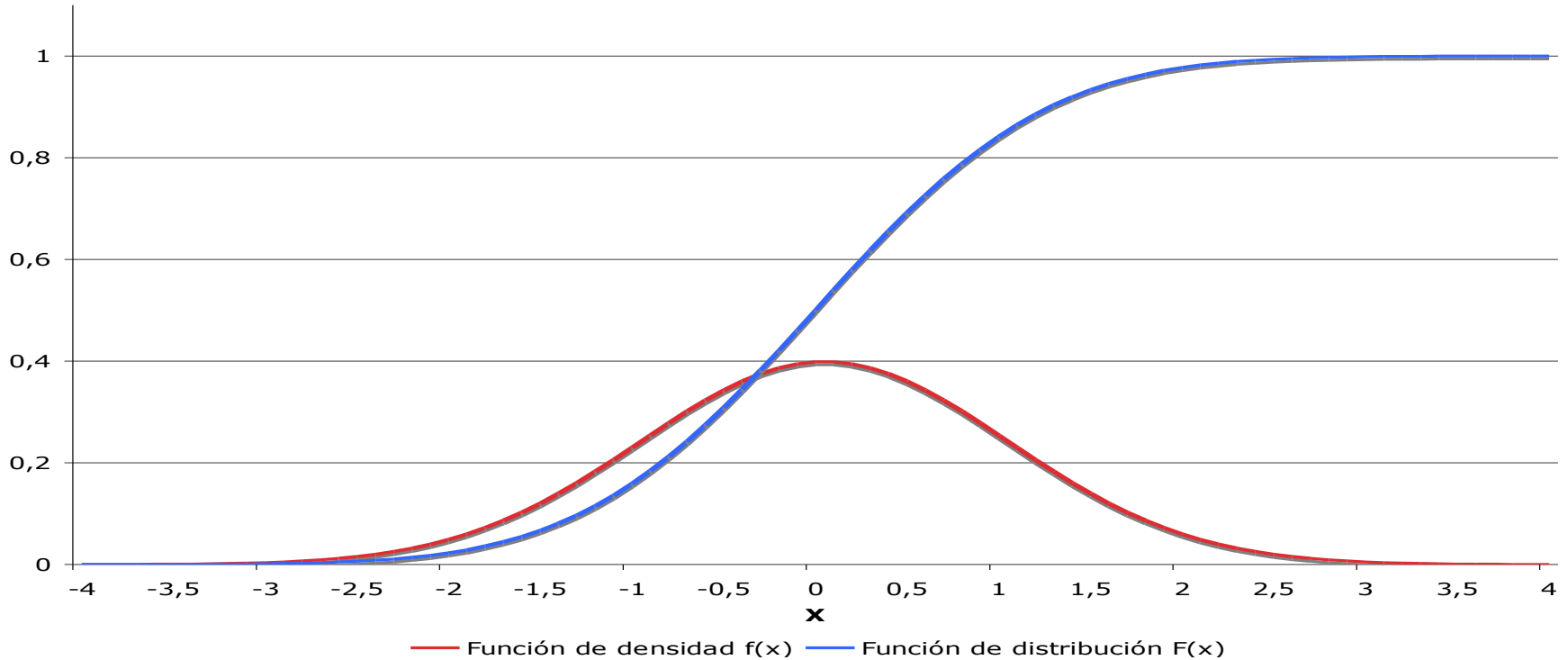


$h=0,8$



$h=0,4$

Aproximación a la distribución normal



$h=0,1$

Diferencia entre v.a. y distribución de probabilidad

Una variable aleatoria representa el resultado particular de un experimento y una distribución de probabilidad representa todos los posibles resultados que pueden suceder en el experimento así como la correspondiente probabilidad.

El valor obtenido en una variable aleatoria describe el pasado mientras que la distribución de probabilidad describe cuán probable es que suceda dicho resultado en el futuro.

Esperanza matemática

Sea X una v.a. Llamaremos esperanza matemática de la v.a. a:

$$E[X] = \int_{\Omega} X dP(\omega) = \int_{\mathbf{R}} \xi(x) dP_X(x) = \int_{\mathbf{R}} \xi(x) dF_X(x)$$

siendo ξ la función medible identidad en $\{\mathbf{R}, \mathbf{B}\}$.

Esperanza matemática

Si X es una v.a. discreta con función de cuantía p_X y que puede tomar valores en un conjunto numerable $\{x_j\} \subset \mathbf{R}$:

$$E[X] = \sum_{j=1}^{\infty} x_j p_X(x_j)$$

Si X es una v.a. continua con función de densidad f :

$$E[X] = \int_{\mathbf{R}} x f_X(x) dx$$

Momentos

Sea $g : \mathbf{R} \rightarrow \mathbf{R}$ una función real medible y X una v.a. Entonces $Y=g(X)$ es también una variable aleatoria.

Se llama momento de orden k de la v.a. X a la esperanza de $g(X) = X^k$. Se representa por α_k .

$$\blacktriangleright \alpha_k = E[X^k]$$

Se llama momento respecto a la media de orden k de la v.a. X a la esperanza de $g(X) = (X - \alpha_1)^k$. Se representa por μ_k .

$$\blacktriangleright \mu_k = E[(X - \alpha_1)^k] = E[(X - \mu)^k]$$

Momentos destacados

El momento de orden 1 recibe el nombre de media y se suele representar normalmente como μ .

- ▶ $\mu = \alpha_1 = E[X]$

El momento respecto a la media de orden 2 recibe el nombre de varianza y se representa por σ^2 o por $V(X)$

- ▶ $\sigma^2 = \mu_2 = E[(X - \alpha_1)^2] = E[(X - \mu)^2]$

La raíz cuadrada positiva de la varianza recibe el nombre de desviación típica y se representa por σ o por $D[X]$.

Vectores aleatorios

Sea $\{\Omega, \mathcal{A}, P\}$ un espacio probabilístico y sea $\{\mathbf{R}^n, \mathbf{B}_n\}$ el espacio probabilizable formado por el espacio vectorial real \mathbf{R}^n y la σ -álgebra de conjuntos de Borel $\mathbf{B}_n = \mathbf{B}(\mathbf{R}^n)$.

Una aplicación vectorial $\mathbf{X} : \Omega \rightarrow \mathbf{R}^n$ es un vector aleatorio o una variable aleatoria n -dimensional si y solo si $\mathbf{X}^{-1}(B) \in \mathcal{A}$, $\forall B \in \mathbf{B}_n$.

Vectores aleatorios

El vector aleatorio \mathbf{X} induce una medida de probabilidad $P_{\mathbf{X}}$ sobre $\{\mathbf{R}^n, \mathbf{B}_n\}$ que lo convierte en el espacio probabilístico $\{\mathbf{R}^n, \mathbf{B}_n, P_{\mathbf{X}}\}$, donde:

$$P_{\mathbf{X}}(B) = P\{\mathbf{X}^{-1}(B)\} = P(A), \forall B \in \mathbf{B}_n, \text{ donde } \mathbf{X}(A) = B$$

$\mathbf{X}=(X_1, X_2, \dots, X_n)$ es una variable aleatoria n -dimensional si y solo si las $X_i : \Omega \rightarrow \mathbf{R}$ son variables aleatorias unidimensionales.

Ejercicio: los resultados del programa PIPE del ICEX como variables aleatorias

Considere los resultados establecidos del programa PIPE.

Describa dichos resultados como variable aleatoria. En particular, indique:

- ▶ El espacio muestral
 - ▶ ¿Es la variable aleatoria unidimensional o multidimensional?
 - ▶ ¿Es la variable aleatoria discreta o continua?
 - ▶ ¿Qué modelo de distribución de probabilidad cree que define dicha variable aleatoria?
-
- ▶ Tiempo: 15 minutos
 - ▶ Ejercicio en grupo

Bibliografía

Lecciones de cálculo de probabilidades

Autores: Vicente Quesada y Alfonso García

Ediciones Díaz de Santos, 1988

Consultable en Google Books

Técnicas cuantitativas de evaluación de políticas públicas

Lima, 10 de septiembre de 2013



«Nada resulta más engañoso que un hecho evidente».

Nociones de estadística

Teorema central del límite

Distribuciones más comunes en estadística

Inferencia estadística

- ▶ Muestra aleatoria
- ▶ Estadísticos
- ▶ Momentos muestrales
- ▶ Intervalos y regiones de confianza
- ▶ Contrastes de hipótesis
 - ▶ Tests de diferencias de medias

Teorema central del límite

No existe un único Teorema Central del Límite, sino un conjunto de teoremas. Todos ellos dan condiciones para que una sucesión de variables aleatorias tienda a distribuirse según una distribución normal.

Teorema central del límite

Sea $\{X_n\}_{n \in \mathbf{N}}$ una sucesión de variables aleatorias definidas sobre el mismo espacio probabilístico $\{\Omega, \mathcal{A}, P\}$, con medias y varianzas finitas.

Diremos que la sucesión $\{X_n\}_{n \in \mathbf{N}}$ obedece al teorema central del límite y lo denotaremos $X_n \in LN$, si y solo si la sucesión $\{S_n\}_{n \in \mathbf{N}}$, donde

$$S_n = \sum_{i=1}^n X_i, \quad \forall n \in \mathbf{N}$$

converge en ley hacia una distribución normal. Tipificando:

$$X_n \in LN \Leftrightarrow Z_n = \frac{S_n - E[S_n]}{D[S_n]} \xrightarrow{\ell} N(0,1)$$

Teorema central del límite

Teorema de Levy-Lindeberg

Sea $\{X_n\}_{n \in \mathbb{N}}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas, con media y varianza finitas. Entonces $X_n \in LN$.

Generalización vectorial del Teorema de Levy-Lindeberg

Sea $\{\mathbf{X}_n\}_{n \in \mathbb{N}}$ una sucesión de vectores aleatorios independientes e idénticamente distribuidos, con vector de medias finitas $\boldsymbol{\mu}$ y matriz de varianzas-covarianzas finitas Σ (definida positiva). Entonces:

$$Y_n = \frac{\sum_{i=1}^n \mathbf{X}_i - n\boldsymbol{\mu}}{\sqrt{n}} \xrightarrow{\ell} N(0, \Sigma)$$

La distribución normal

Una variable aleatoria X se dice que tiene una distribución normal o de Gauss-Laplace si su función de densidad es de la forma:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, -\infty < x < \infty$$

Considerada por primera vez en 1753, se le dio el nombre de normal porque se creyó que, en la práctica, la mayoría de las distribuciones eran de este tipo y el resto, anormales. La distribución normal se denota como $N(0,1)$.

La distribución normal

Una variable aleatoria X se dice que tiene una distribución normal con parámetros μ y σ , y se denota como $N(\mu, \sigma)$, si su función de densidad es de la forma:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty < x < \infty, -\infty < \mu < \infty, \sigma > 0$$

Si X tiene distribución $N(\mu, \sigma)$, la variable

$$Y = \frac{X - \mu}{\sigma} \equiv N(0, 1)$$

La distribución normal

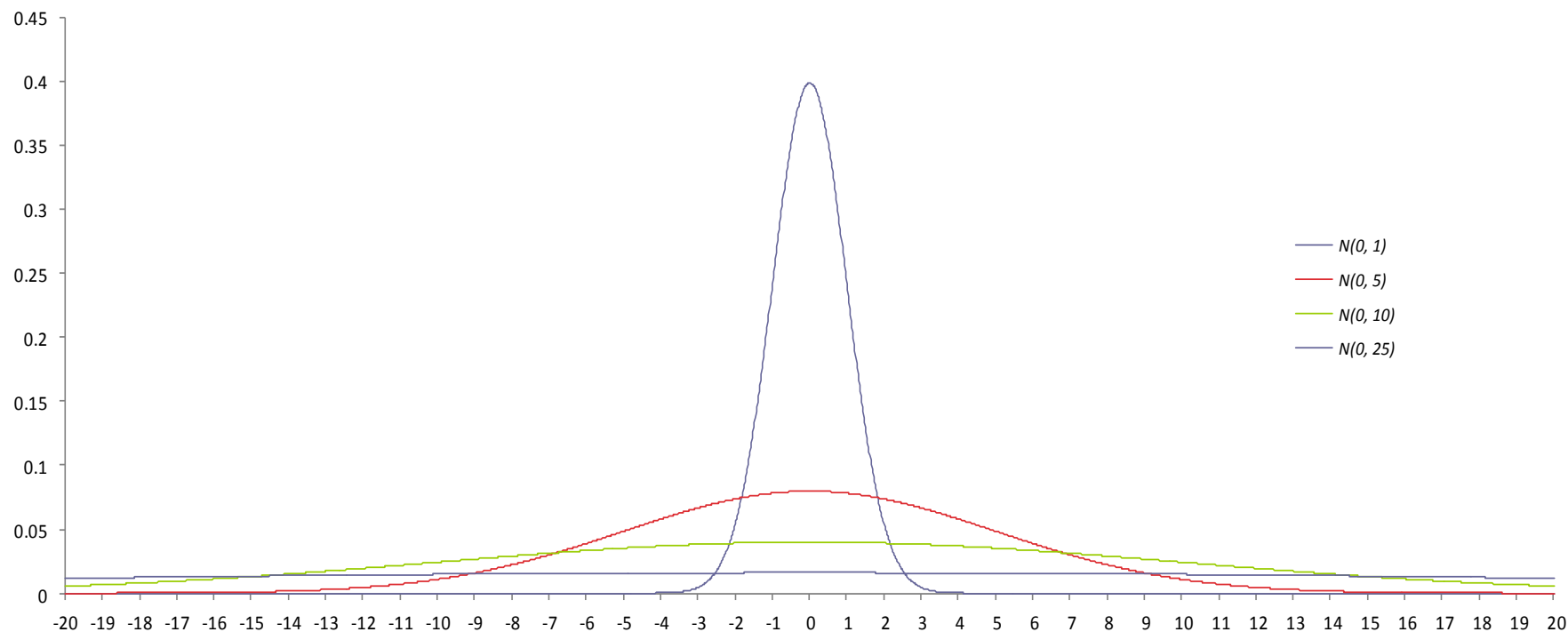
Sus momentos son:

$$E[N(\mu, \sigma)] = \mu, \quad V(N(\mu, \sigma)) = \sigma^2$$

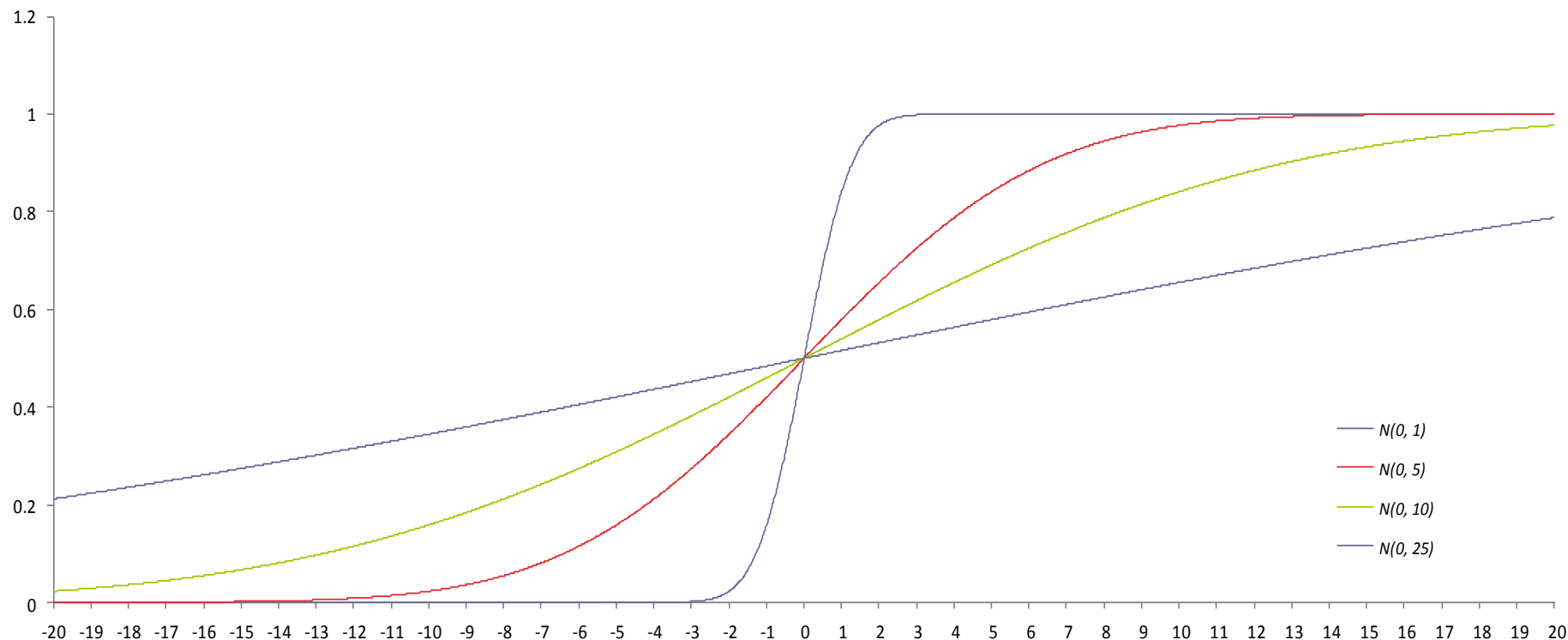
En particular, los momentos de la distribución $N(0,1)$ son:

$$E[N(0,1)] = 0, \quad V(N(0,1)) = 1$$

Función de densidad de la distribución $N(0, \sigma)$ para diferentes valores de σ

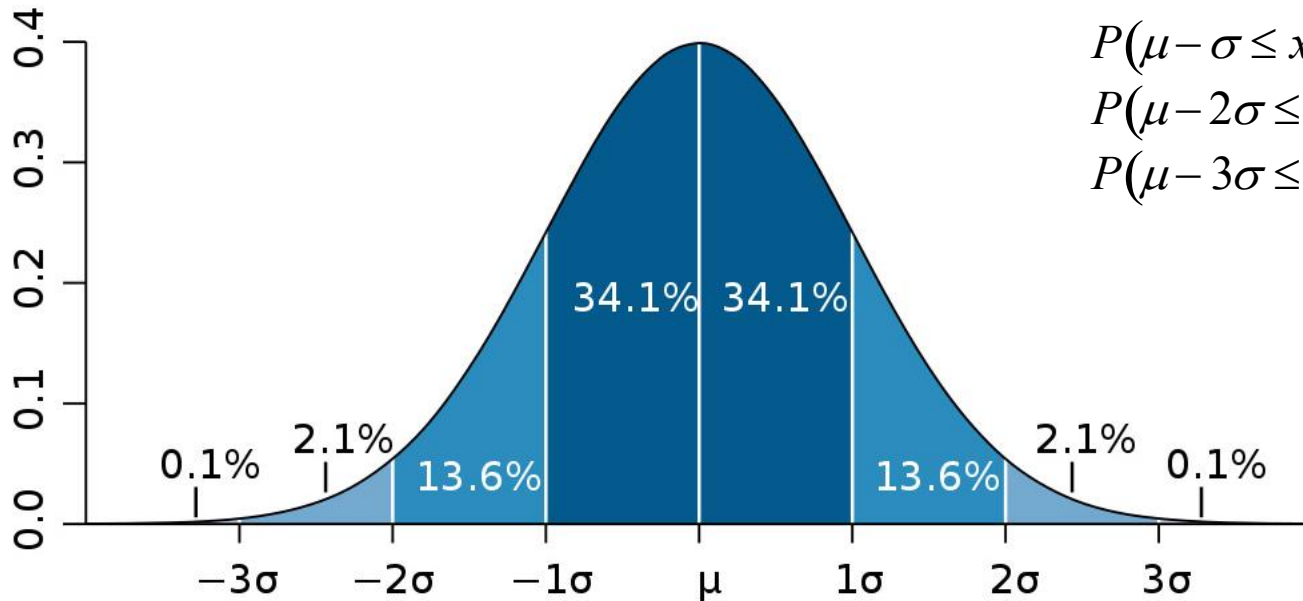


Función de distribución de la distribución $N(0, \sigma)$ para diferentes valores de σ



La distribución normal

En la distribución $N(\mu, \sigma)$ se cumple:



$$P(\mu - \sigma \leq x \leq \mu + \sigma) \approx 0.6827$$
$$P(\mu - 2\sigma \leq x \leq \mu + 2\sigma) \approx 0.9545$$
$$P(\mu - 3\sigma \leq x \leq \mu + 3\sigma) \approx 0.9973$$

Distribución χ^2 (ji-cuadrado) de Pearson

Si X_1, X_2, \dots, X_n son variables aleatorias independientes y con distribución $N(0,1)$, la distribución de la variable

$$\sum_{i=1}^n X_i^2 \equiv \chi_n^2$$

se denomina χ^2 (ji-cuadrado) de Pearson con n grados de libertad.

Distribución χ^2 (ji-cuadrado) de Pearson

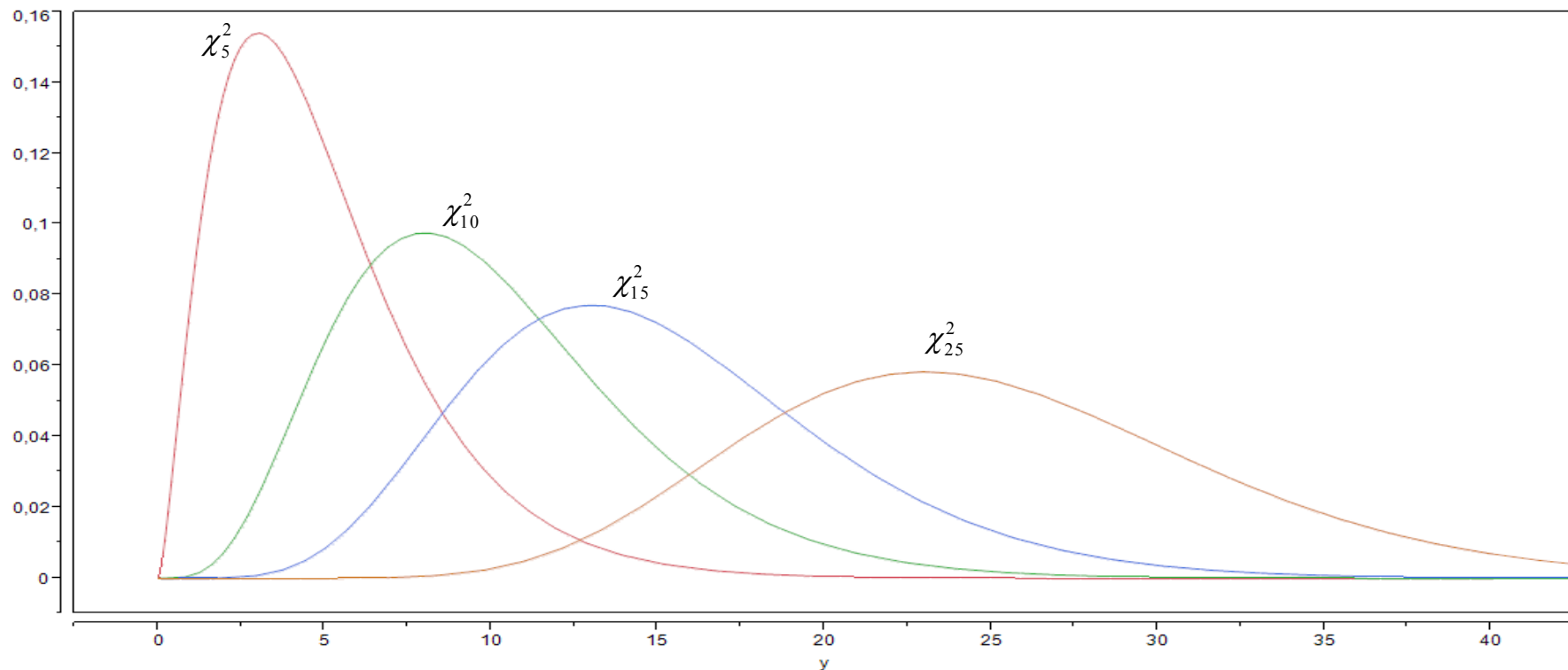
Sus momentos son:

$$E[\chi_n^2] = n, V(\chi_n^2) = 2n$$

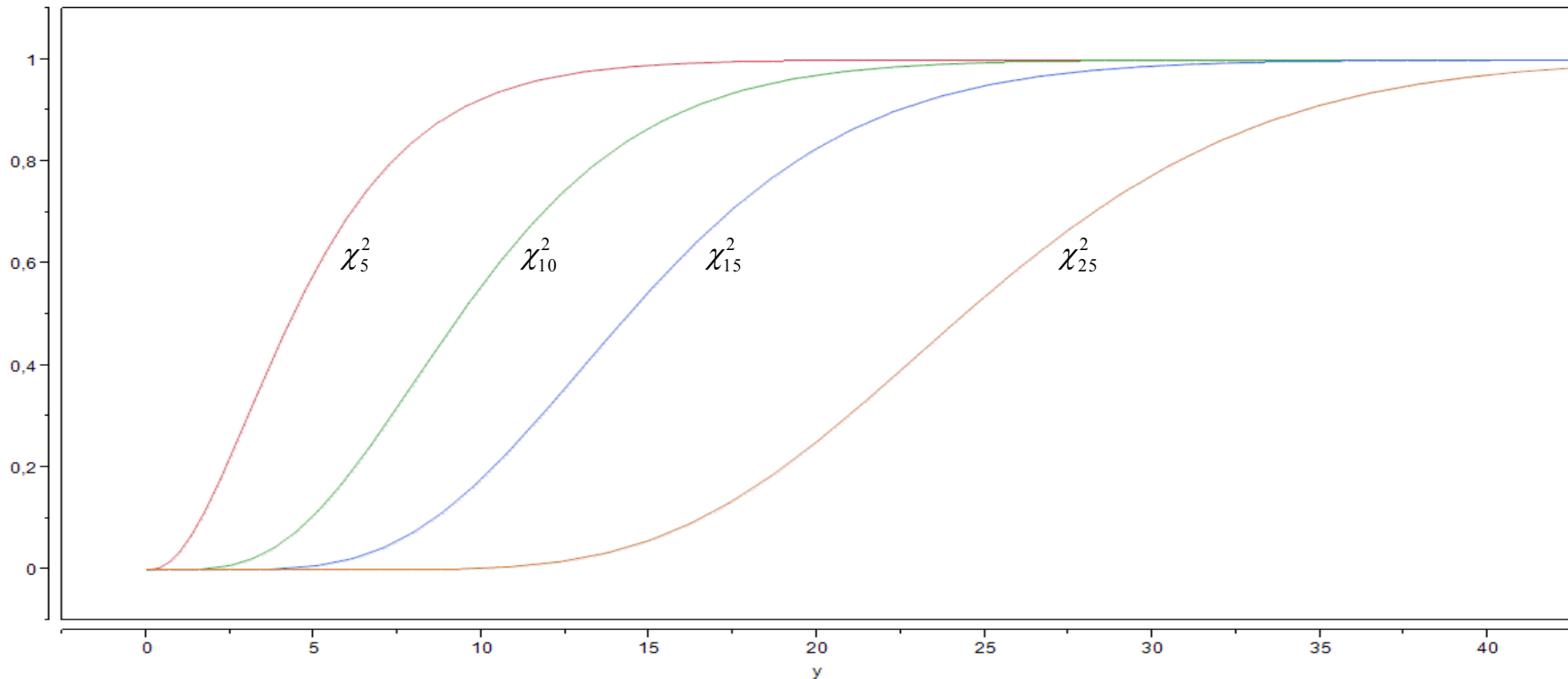
La función de densidad de la distribución χ_n^2 es:

$$f(y) = \frac{1}{2^{n/2} \Gamma(n/2)} y^{\frac{n}{2}-1} e^{-y/2}, \forall y > 0$$

Función de densidad de la distribución χ^2 para diferentes grados de libertad



Función de distribución de la distribución χ^2 para diferentes grados de libertad



Distribución t de Student

Si X, X_1, X_2, \dots, X_n son variables aleatorias independientes y con distribución $N(0, \sigma)$, la distribución de la variable

$$\frac{X}{\sqrt{\frac{1}{n} \sum_{i=1}^n X_i^2}} \equiv t_n$$

se denomina distribución t de Student n grados de libertad.

Distribución t de Student

Sus momentos son:

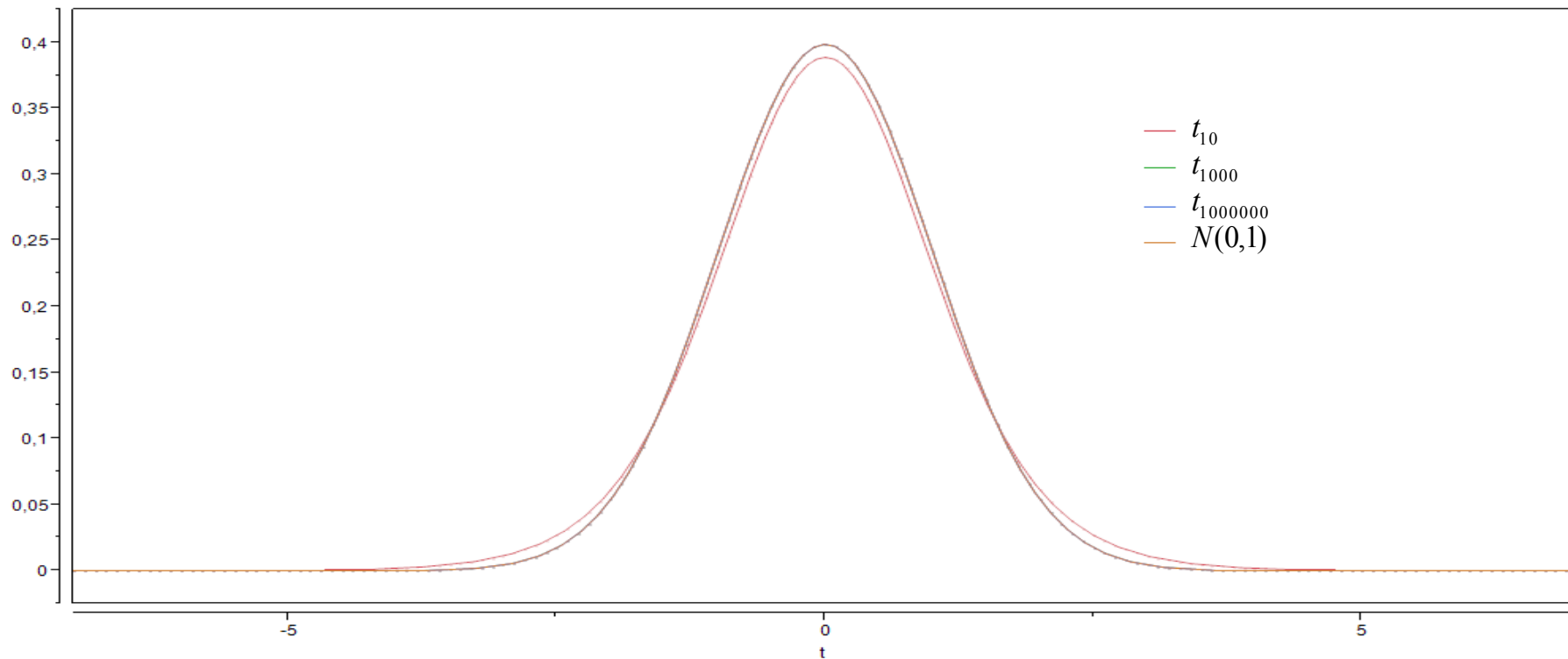
$$E[t_n] = 0, V(t_n) = \frac{n}{n-2}, n > 2$$

Si $\sigma = 1$, $t_n \xrightarrow{d} N(0,1)$

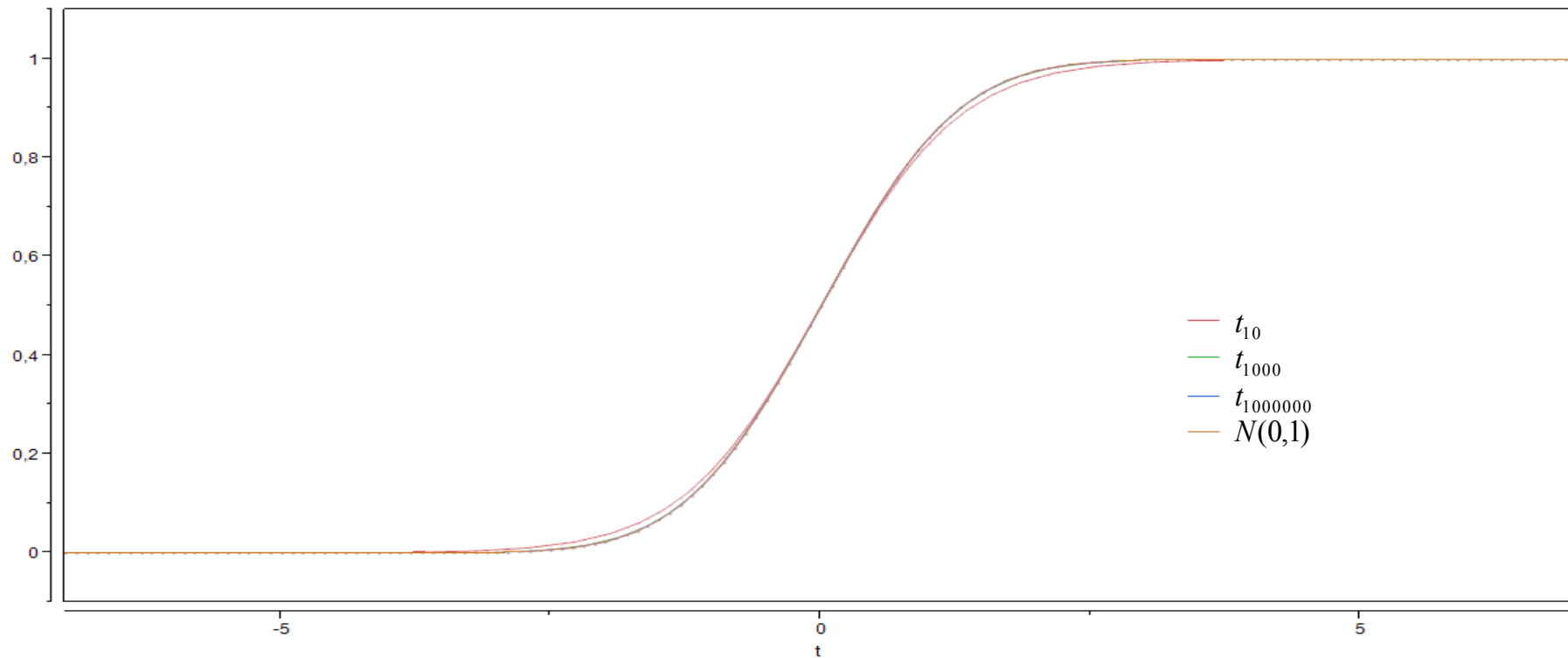
La función de densidad de la distribución t_n es

$$f(t) = \frac{1}{\sqrt{n\pi}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}, \forall t \in \mathbf{R}$$

Función de densidad de la distribución t para diferentes grados de libertad



Función de distribución de la distribución t para diferentes grados de libertad



Distribución F de Snedecor

Si $X_1, X_2, \dots, X_n, Y_1, Y_2, \dots, Y_m$ son variables aleatorias independientes y con distribución $N(0, \sigma)$, la distribución de la variable

$$\frac{\frac{1}{n} \sum_{i=1}^n X_i^2}{\frac{1}{m} \sum_{i=1}^m Y_i^2} \equiv F_{n,m}$$

se denomina distribución F de Snedecor con n y m grados de libertad.

Distribución F de Snedecor

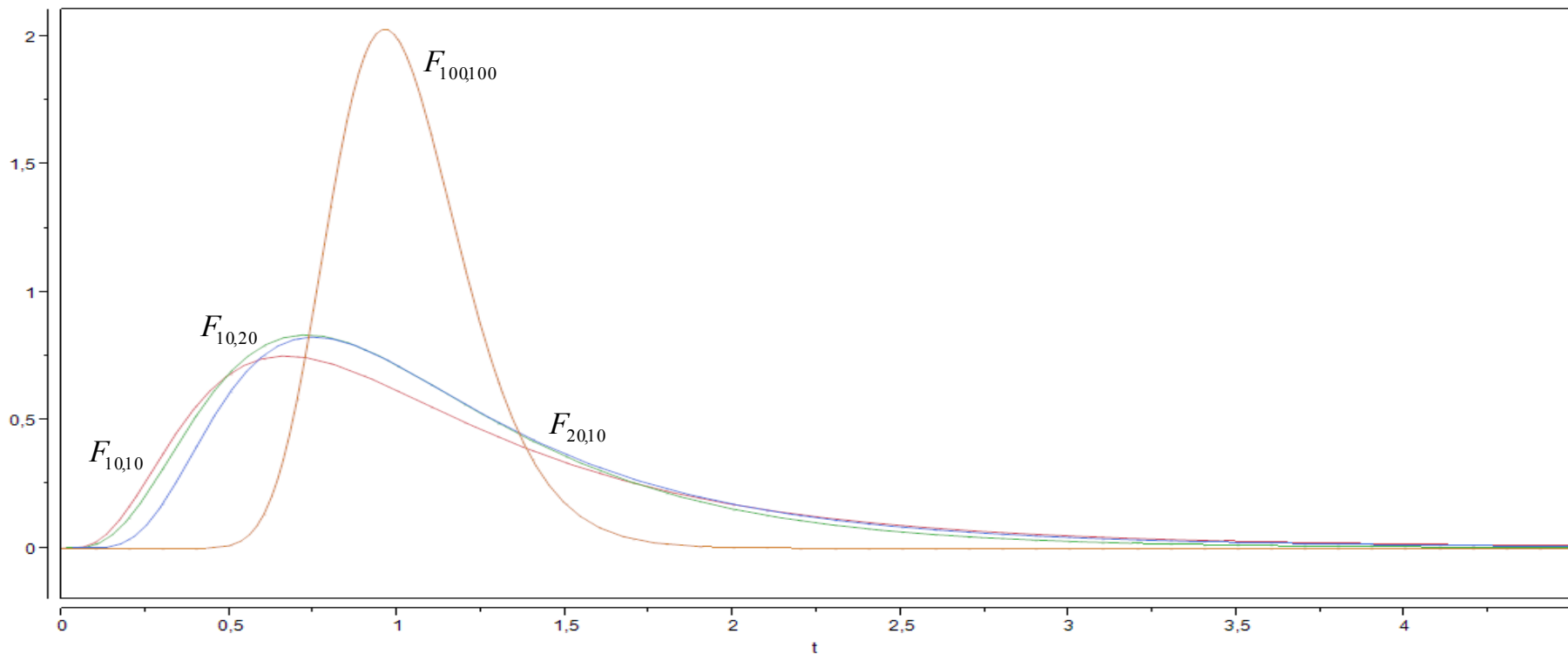
Sus momentos son:

$$E[F_{n,m}] = \frac{m}{m-2}, m > 2; V(F_{n,m}) = \frac{2m^2(n+m+2)}{n(m-2)^2(m-4)}, m > 4$$

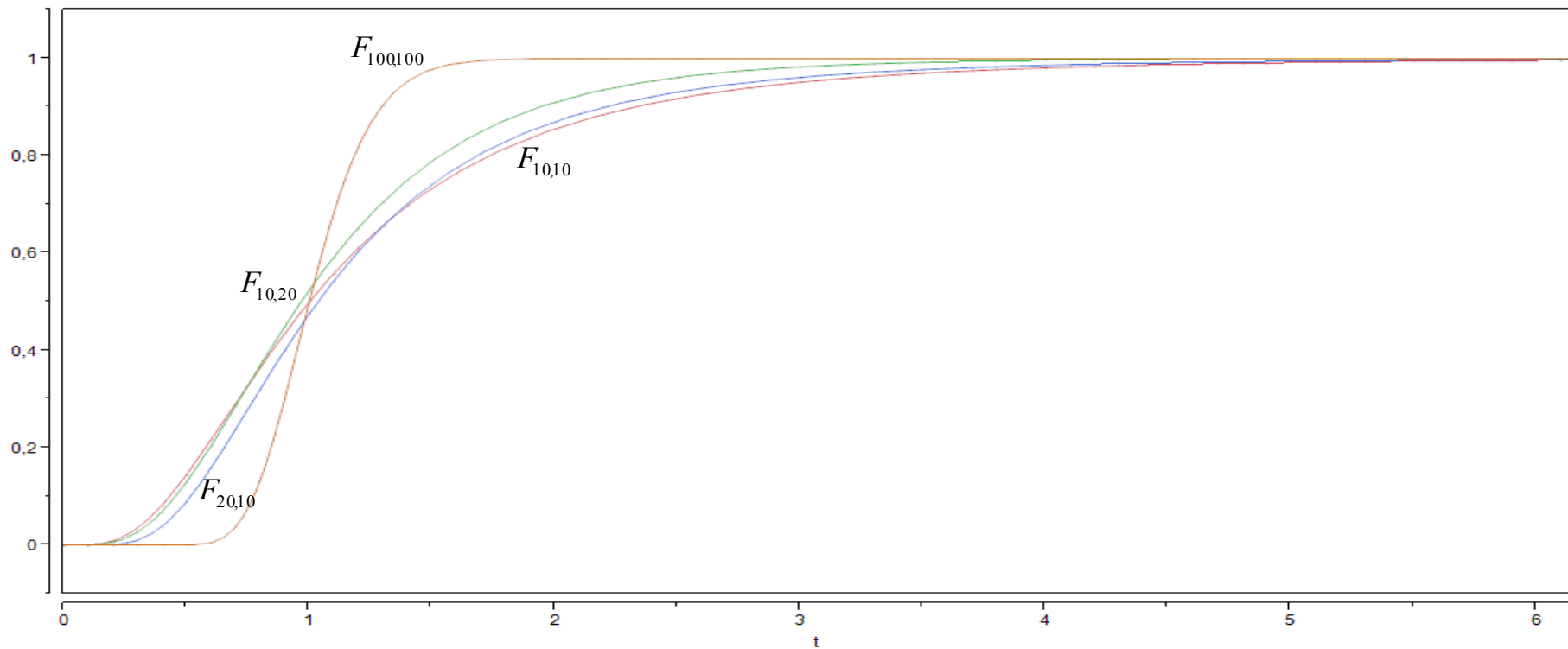
La función de densidad de la distribución $F_{n,m}$ es

$$f(t) = \frac{\Gamma\left(\frac{n+m}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)} \left(\frac{n}{m}\right)^{\frac{n}{2}} t^{(n-2)/2} \left(1 + \frac{n}{m}t\right)^{-\frac{n+m}{2}}, \forall t \geq 0$$

Función de densidad de la distribución F para diferentes grados de libertad



Función de distribución de la distribución F para diferentes grados de libertad



La distribución normal n -dimensional

Un vector aleatorio \mathbf{X} se dice que tiene una distribución normal n -dimensional con vector de medias $\boldsymbol{\mu}$ y matriz de varianzas-covarianzas Σ , y se denota como $N(\boldsymbol{\mu}, \Sigma)$, si su función de densidad es de la forma:

$$f(\mathbf{x}) = f(x_1, x_2, \dots, x_n) = \frac{|\Sigma|^{-1/2}}{(2\pi)^{n/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})' \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})}, \forall \mathbf{x} \in \mathbf{R}^n, \boldsymbol{\mu} \in \mathbf{R}^n, |\Sigma| > 0$$

Los momentos del vector aleatorio son, efectivamente:

$$\mathbf{E}[\mathbf{X}] = \boldsymbol{\mu}, \mathbf{V}[\mathbf{X}] = \Sigma$$

La distribución normal n -dimensional

Si el vector aleatorio $\mathbf{X} = (X_1, X_2, \dots, X_n)$ tiene una distribución normal n -dimensional $N(\boldsymbol{\mu}, \Sigma)$, siendo

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_i \\ \vdots \\ \mu_n \end{pmatrix}, \Sigma = \begin{bmatrix} \sigma_{11} & \dots & \sigma_{1i} & \dots & \sigma_{1n} \\ \vdots & \ddots & \vdots & & \vdots \\ \sigma_{i1} & \dots & \sigma_{ii} & \dots & \sigma_{in} \\ \vdots & & \vdots & \ddots & \vdots \\ \sigma_{n1} & \dots & \sigma_{i1} & \dots & \sigma_{nn} \end{bmatrix}$$

entonces cada X_i sigue una distribución normal unidimensional $N(\mu_i, \sigma_{ii})$. Solo serán independientes entre sí si Σ es diagonal.

Inferencia estadística

La estadística matemática o inferencia estadística tiene como finalidad obtener información acerca de la ley de probabilidad de un fenómeno aleatorio a través de una observación no exhaustiva del mismo.

Para ello, es necesario solo disponer de un conjunto finito de observaciones del fenómeno aleatorio considerado.

Inferencia estadística

Si el objetivo es obtener un pronóstico numérico único de un determinado parámetro de la distribución poblacional, hablamos de problemas de estimación puntual.

Si el objetivo es obtener un margen de variación para un determinado parámetro de la distribución poblacional, hablamos de problemas de estimación por intervalos.

Si el objetivo es corroborar o invalidar una determinada afirmación acerca de la distribución poblacional, hablamos de problemas de contraste de hipótesis.

Inferencia estadística

La distribución desconocida F de la variable aleatoria involucrada recibe el nombre de distribución teórica o distribución de la población.

El mayor o menor grado de desconocimiento acerca de la distribución teórica F se refleja mediante la familia \mathcal{F} de distribuciones candidatas a ser la distribución de la población.

Cuando \mathcal{F} está compuesta por distribuciones de forma funcional fija y conocida dependientes de un parámetro θ que varía dentro de un subconjunto $\Theta \subset \mathbf{R}^k$, denominado espacio paramétrico, hablamos de inferencia estadística paramétrica.

$$\mathcal{F} = \{F_\theta \mid \theta \in \Theta \subset \mathbf{R}^k\}$$

Muestra aleatoria

Llevando a cabo repeticiones del experimento aleatorio que da lugar a la variable aleatoria X obtenemos un conjunto de valores numéricos (x_1, x_2, \dots, x_n) que constituyen una muestra aleatoria de X . El número n se denomina tamaño de la muestra.

Todo problema de inferencia estadística debe incluir la definición precisa del procedimiento de muestreo con el que se obtienen las observaciones. Dicha definición recibe el nombre de diseño del muestreo y permitirá elevar los resultados del muestro a la población de origen.

Muestra aleatoria

Una muestra aleatoria simple (m.a.s) de tamaño n de una variable aleatoria X con distribución teórica F son n variables aleatorias (X_1, X_2, \dots, X_N) independientes e idénticamente distribuidas con distribución común F .

La función de distribución conjunta de una muestra aleatoria simple correspondiente a una distribución de la población F es:

$$F(x_1, x_2, \dots, x_n) = F(x_1)F(x_2) \cdots F(x_n)$$

Muestra aleatoria

Además se puede asociar a cada muestra concreta (x_1, x_2, \dots, x_n) una distribución muestral F^* con la pretensión de emular la distribución teórica F . F^* se define como:

$$F_n^*(x) = \frac{\text{número de elementos muestrales} \leq x}{n}, \quad \forall x \in \mathbf{R}$$

La distribución muestral es siempre discreta, concentrada en los valores que aparecen en la muestra y su función de probabilidad es:

$$p_n^*(x) = \frac{j}{n}, \quad \forall x \in \mathbf{R}$$

siendo j el número de valores muestrales que coinciden con x .

Espacio muestral

Se llama espacio muestral \mathcal{X} al conjunto de muestras posibles que pueden obtenerse al seleccionar una muestra aleatoria de tamaño determinado de una cierta población.

Si la variable aleatoria de la que se extrae la muestra aleatoria de tamaño n es m -dimensional, \mathcal{X} es un subconjunto del espacio euclídeo \mathbf{R}^{mn} , sobre el que podemos considerar la σ -álgebra restringida de la σ -álgebra de Borel \mathbf{B}_{mn} , que representamos por \mathcal{B} . La distribución de la muestra determina una medida de probabilidad \mathcal{P} en el espacio muestral $(\mathcal{X}, \mathcal{B})$. Si el problema es paramétrico suele incluirse en la notación el parámetro y escribimos \mathcal{P}_θ .

Comportamiento asintótico de la distribución muestral

Si elegimos la muestra mediante muestreo aleatorio simple, la función de distribución muestral F_n^* , donde n hace referencia al tamaño de la muestra, debe considerarse también una variable aleatoria.

El siguiente teorema prueba que F_n^* converge a F_n uniformemente en x , con probabilidad uno.

Comportamiento asintótico de la distribución muestral

Teorema de Glivenko-Cantelli:

Sea $\{X_i\}_{i=1}^{\infty}$ una sucesión de variables aleatorias independientes y con distribución común F . Si F_n^* es la función de distribución muestral asociada a la muestra aleatoria simple (X_1, X_2, \dots, X_n) y

$$\Delta_n = \sup_{x \in \mathbf{R}} |F_n^*(x) - F(x)|$$

entonces

$$\lim_{n \rightarrow \infty} \Delta_n = 0, \mathcal{P} - \text{casi seguro}$$

Estadísticos

Se denomina estadístico a cualquier función T del espacio muestral $(\mathcal{X}, \mathcal{B})$ en un espacio euclídeo $(\mathbf{R}^k, \mathbf{B}_k)$ que sea medible respecto a las σ -álgebras \mathcal{B} y \mathbf{B}_k . La dimensión k del espacio euclídeo imagen se denomina dimensión del estadístico.

Estadísticos

Cuando se considera la muestra (X_1, X_2, \dots, X_n) como una variable aleatoria, el estadístico T es a su vez una variable aleatoria $T(X_1, X_2, \dots, X_n)$.

Se denomina distribución en el muestreo de un estadístico T a la distribución de la variable aleatoria $T(X_1, X_2, \dots, X_n)$: es decir, la distribución en el muestreo del estadístico T es la medida de probabilidad que induce la distribución de la muestra, \mathcal{P} , mediante la función $T: (\mathcal{X}, \mathcal{B}) \rightarrow (\mathbf{R}^k, \mathbf{B}_k)$

Estadísticos

Parámetro y estadístico son conceptos muy diferentes. Un parámetro es una constante que describe la población y cuando se conoce queda determinado el modelo probabilístico, mientras que un estadístico es una variable aleatoria cuyo valor no contiene ningún valor o parámetro desconocido pues depende solamente de las observaciones muestrales, por lo que sí está sometido a una distribución de probabilidad.

Momentos muestrales

Los momentos muestrales son estadísticos definidos para cada muestra posible de tamaño n (x_1, x_2, \dots, x_n) según la expresión siguiente:

$$a_k(x) = \frac{1}{n} \sum_{i=1}^n x_i^k, \quad b_k(x) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^k$$

Tienen particular importancia la media y la varianza muestrales:

$$\bar{x} = a_1(x) = \frac{1}{n} \sum_{i=1}^n x_i, \quad s^2 = b_2(x) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

Momentos muestrales

La distribución en el muestreo de los momentos muestrales considerados como variables aleatorias no admiten una expresión general, pues dependen de la forma funcional de la distribución poblacional.

Sin embargo, los momentos de los momentos muestrales sí admiten expresiones generales.

Momentos muestrales

Suponiendo que μ y σ^2 (la media y la varianza poblacionales) existen, los momentos de la media y la varianza muestrales son:

$$E[\bar{X}] = \mu, V(\bar{X}) = \frac{\sigma^2}{n}, E[s^2] = \frac{n-1}{n} \sigma^2$$

Es frecuente considerar también la cuasivarianza muestral:

$$S^2 = \frac{n}{n-1} s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

que cumple:

$$E[S^2] = \sigma^2$$

Distribución de la media y la varianza muestral

Teorema de Fisher

Si (X_1, X_2, \dots, X_n) es una muestra aleatoria simple de una población $N(\mu, \sigma)$, entonces s^2 y \bar{X} son variables aleatorias independientes y la distribución en el muestreo de ambas es:

$$n \frac{s^2}{\sigma^2} \equiv \chi_{n-1}^2$$

$$\bar{X} \equiv N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Estadístico de Student

Teorema de Student

Si (X_1, X_2, \dots, X_n) es una muestra aleatoria simple de una población $N(\mu, \sigma)$, el estadístico de Student:

$$\sqrt{n-1} \frac{\bar{X} - \mu}{s} = \sqrt{n} \frac{\bar{X} - \mu}{S} \equiv t_{n-1}$$

tiene distribución t de Student con $n-1$ grados de libertad.

Distribución del cociente de cuasivarianzas muestrales

Teorema

Si S_1^2 y S_2^2 son las cuasivarianzas de sendas muestras aleatorias simples de tamaño n y m , respectivamente, de dos poblaciones normales de varianzas σ_1^2 y σ_2^2 , respectivamente, el estadístico

$$U = \frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2} \equiv F_{n-1, m-1}$$

tiene distribución F de Snedecor con $n-1$ y $m-1$ grados de libertad.

Distribución del coeficiente de correlación muestral

Sea una muestra aleatoria simple de tamaño n de una población normal bidimensional (X, Y) con matriz de varianzas y covarianzas Σ . Llamamos coeficiente de correlación muestral a

$$R = \frac{S_{11}}{S_1 S_2}, \text{ donde } S_{11} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$$

es la cuasicovarianza muestral.

Distribución del coeficiente de correlación muestral

Si R es el coeficiente de correlación muestral de una distribución normal bidimensional con coeficiente de correlación poblacional $\rho = 0$, el estadístico

$$R^* = \sqrt{n-2} \frac{R}{\sqrt{1-R^2}} \equiv t_{n-2}$$

tiene distribución t de Student con $n-2$ grados de libertad.

Intervalos de confianza

Sea \mathcal{F} una familia de distribuciones de forma funcional fija y conocida dependientes de un parámetro θ que varía dentro de un subconjunto $\Theta \subset \mathbf{R}^k$,

$$\mathcal{F} = \left\{ F_{\theta} \mid \theta \in \Theta \subset \mathbf{R}^k \right\}$$

Si disponemos de un estadístico T para estimar θ , en muchas ocasiones la distribución en el muestreo de dicho estadístico está muy concentrada en torno al verdadero valor del parámetro. Esto nos permite conocer los márgenes de variación previsibles para el parámetro con una probabilidad fijada de antemano.

Intervalos de confianza

De una población descrita por una variable aleatoria X cuya distribución teórica $F \in \mathcal{F}$, se considera una muestra aleatoria (X_1, X_2, \dots, X_n) con distribución \mathcal{P}_θ . Sea $g(\theta) : \Theta \rightarrow \mathbf{R}$ una función real y $T_1 \leq T_2$ dos estadísticos unidimensionales tales que

$$\mathcal{P}_\theta\{T_1(X_1, \dots, X_n) \leq g(\theta) \leq T_2(X_1, \dots, X_n)\} \geq 1 - \alpha, \forall \theta \in \Theta$$

Entonces, para cualquier muestra concreta (x_1, x_2, \dots, x_n) , el intervalo

$$[T_1(X_1, \dots, X_n), T_2(X_1, \dots, X_n)]$$

se denomina intervalo de confianza para $g(\theta)$ de nivel de confianza $1 - \alpha$.

Intervalos de confianza

Ejemplo:

Consideramos una muestra aleatoria de tamaño n (X_1, X_2, \dots, X_n) proveniente de una población con distribución teórica $N(\mu, \sigma)$ con μ desconocida pero σ conocida. Sabemos que la media muestral

$$\bar{X} \equiv N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Calculamos entonces la media muestral y sabemos que

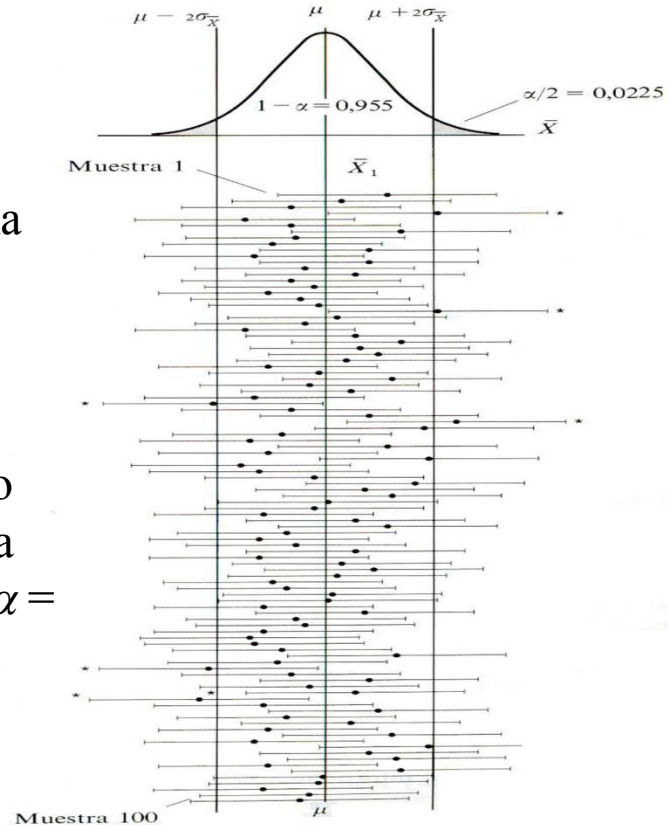
$$P\left\{\bar{x} - 2\sigma_{\bar{x}} \leq \mu \leq \bar{x} + 2\sigma_{\bar{x}}\right\} = P\left\{\bar{x} - 2\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + 2\frac{\sigma}{\sqrt{n}}\right\} \approx 0.955$$

Intervalos de confianza

Veámoslo de otro modo. Suponemos conocida la media poblacional μ . Como asintóticamente la media muestral converge hacia μ , si construimos un intervalo de la forma

$$\left[\bar{x} - 2 \frac{\sigma}{\sqrt{n}}, \bar{x} + 2 \frac{\sigma}{\sqrt{n}} \right]$$

para el 95.5% de las muestras posibles de tamaño n que obtengamos, la media muestral estará contenida en dicho intervalo, es decir: de cada 100 muestras, solo para 4,5 la media muestral estará fuera del intervalo. En este caso, $\alpha = 0.045$ y el nivel de confianza es del 95.5%.



Regiones de confianza

Si $\theta \in \Theta \subset \mathbf{R}^k$ es un parámetro k-dimensional y se cumplen las mismas condiciones anteriores, si $S(x_1, x_2, \dots, x_n)$ es un subconjunto de Θ , dependiente de la muestra (x_1, x_2, \dots, x_n) , que verifica

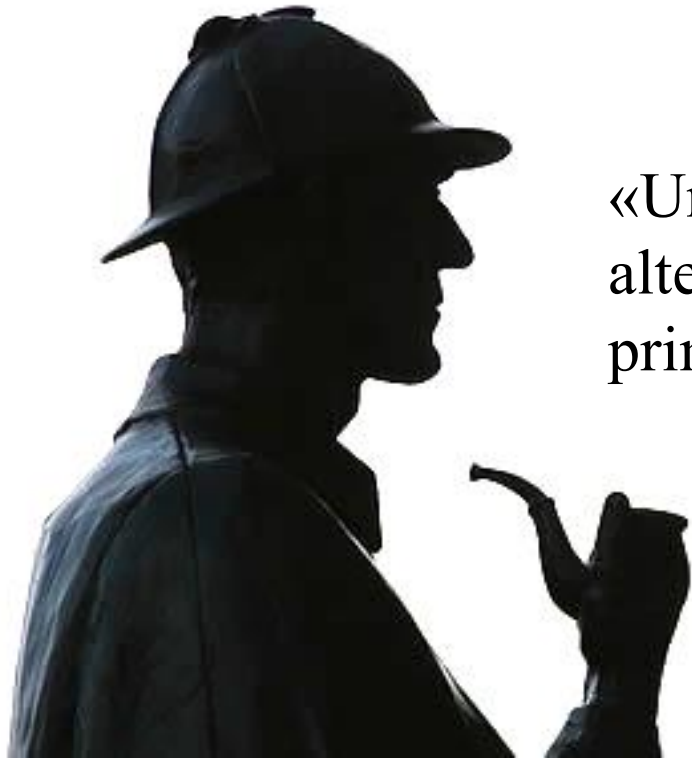
$$\mathcal{P}_\theta \{ \theta \in S(X_1, \dots, X_n) \} \geq 1 - \alpha, \forall \theta \in \Theta$$

el subconjunto $S(x_1, x_2, \dots, x_n)$, asociado a cualquier muestra concreta, se denomina región de confianza para el parámetro θ de nivel de confianza $1 - \alpha$.

Importancia de los intervalos y las regiones de confianza

Hay dos maneras de construir intervalos y regiones de confianza: el método de la cantidad pivotal y el método de Neymann.

Poder construir estas regiones de confianza permite diseñar contrastes de hipótesis formales, en los que la región de confianza para aceptar la hipótesis se denomina región de aceptación y los puntos del espacio muestral que quedan fuera de la región de confianza constituyen la denominada región crítica.



«Uno debe buscar siempre una posible alternativa y demostrar que es falsa. Es la primera regla de la investigación».

Contrastes de hipótesis

La finalidad del contraste de hipótesis es decidir si una determinada hipótesis sobre la distribución poblacional estudiada es confirmada o invalidada estadísticamente a partir de las observaciones contenidas en una muestra.

La hipótesis que se quiere confirmar o refutar se denomina hipótesis nula y se representa por H_0 . Esta hipótesis nula se contrasta frente a una segunda hipótesis, llamada hipótesis alternativa y designada por H_1 , que agrupa todas aquellas distribuciones poblacionales posibles en las que H_0 no es cierta.

Contrastes de hipótesis

Un test de hipótesis es cualquier partición del espacio muestral \mathcal{X} en una región crítica C , en la cual se rechaza la hipótesis nula, y una región de aceptación C^c , en la cual la hipótesis nula no es rechazada. Tales tests reciben el calificativo de no aleatorizados.

Un test aleatorizado es cualquier función medible

$$\varphi: \mathcal{X} \rightarrow [0,1]$$

llamada función crítica del test, tal que $\varphi(x_1, x_2, \dots, x_n)$ expresa la probabilidad de rechazar la hipótesis nula cuando se observa la muestra $(x_1, x_2, \dots, x_n) \in \mathcal{X}$.

Contrastes de hipótesis

Rechazar la hipótesis nula cuando es cierta recibe el nombre de error de tipo I, mientras que no descartarla siendo falsa se denomina error de tipo II.

	H_0 es cierta	H_0 es falsa
Rechazar H_0	Error de tipo I	Decisión correcta
Aceptar H_0	Decisión correcta	Error de tipo II

Lo idóneo sería conseguir un test que hiciese mínimas las probabilidades de ambos errores.

Contrastes de hipótesis

Si la distribución teórica depende de un parámetro $\theta \in \Theta$, acerca del cual hay que contrastar la hipótesis nula $H_0: \theta \in \Theta_0$ frente a la alternativa $H_1: \theta \in \Theta_1$, se denomina función de potencia $\beta(\theta)$ a la probabilidad de rechazar H_0 cuando el valor del parámetro es θ .

Contrastes de hipótesis

Si se trata de un test no aleatorizado de región crítica C , la potencia del test es:

$$\beta(\theta) = \mathcal{P}_{\theta}(C)$$

mientras que para un test de función crítica φ , la función de potencia es:

$$\beta(\theta) = E_{\theta}[\varphi(X_1, X_2, \dots, X_n)]$$

Contrastes de hipótesis

Un test tiene nivel de significación α si

$$\beta(\theta) \leq \alpha, \forall \theta \in \Theta_0$$

Se denomina tamaño del test al número

$$\sup_{\theta \in \Theta_0} \beta(\theta)$$

Contrastes de hipótesis

Para seleccionar el test más adecuado para el contraste de una hipótesis H_0 , consideramos tests de nivel de significación α (es decir, de tamaño menor o igual que α). A continuación intentamos maximizar la potencia para todos los valores de $\theta \in \Theta_1$ (que constituyen la hipótesis H_1) simultáneamente.

Contrastes de hipótesis

En la práctica, dada la dificultad para calcular la función de potencia, se diseña el test en función de α y, una vez obtenida la muestra concreta, se observa el nivel de significación más pequeño para el que tal muestra obliga a rechazar H_0 . Este número $\alpha(x_1, x_2, \dots, x_n) \in [0, 1]$ se denomina nivel crítico asociado a la muestra obtenida e indica el apoyo que la hipótesis nula recibe de las observaciones; cuando más grande sea su valor, más confirmada queda la hipótesis H_0 . En la terminología, este valor recibe el nombre de p -valor.

Fases en la construcción de un contraste de hipótesis

Fase 1. Formular la hipótesis nula H_0 y la hipótesis alternativa H_1 en términos de probabilidad, de forma que sean mutuamente excluyentes y complementarias.

Fase 2. Determinar el estadístico apropiado que se utilizará para rechazar o no rechazar la hipótesis nula H_0 .

Fases en la construcción de un contraste de hipótesis

Fase 3. Seleccionar el nivel de significación, valor α . Este valor indica la importancia o significado que el investigador atribuye a las consecuencias asociadas de rechazar incorrectamente la hipótesis nula H_0 . Valores habituales son $\alpha = 0.1$, $\alpha = 0.05$ y $\alpha = 0.01$.

Fase 4. Determinar la región crítica o región de rechazo y la región de aceptación en la gráfica de la distribución del estadístico. La región de rechazo es el conjunto de realizaciones muestrales para las que rechazamos H_0 . Por el contrario la región de aceptación es el conjunto de realizaciones muestrales para las que aceptamos H_0 .

Fases en la construcción de un contraste de hipótesis

Fase 5. Seleccionar aleatoriamente la muestra y calcular el valor del estadístico.

Fase 6. Interpretación. Si el valor calculado para el estadístico cae dentro de la región crítica definida por el valor o valores críticos, entonces la hipótesis nula H_0 se rechaza. Y si el valor calculado cae dentro de la región de aceptación, entonces la hipótesis nula H_0 no se rechaza.

Ejemplo de construcción de un contraste de hipótesis

Vamos a contrastar si la tensión ocular, X , de los consumidores de un fármaco tiene distribución $N(15,1)$ (H_0) o $N(18,1)$ (H_1).

Fase 1: $H_0: \mu = 15$; $H_1: \mu > 15$

Fase 2: Usaremos como estadístico la media muestral

$$\bar{X} \equiv N(\theta, 1/\sqrt{n})$$

Fase 3: Fijamos el nivel de significación en $\alpha = 0.05$.

Contrastes de hipótesis

Fase 4: La región crítica será de la forma $C = \{\bar{X} > c\}$ donde c es el punto crítico que cumple

$$\mathcal{P}_{\theta=15}(\bar{X} > c) = \alpha = 0.05$$

Despejando, tenemos:

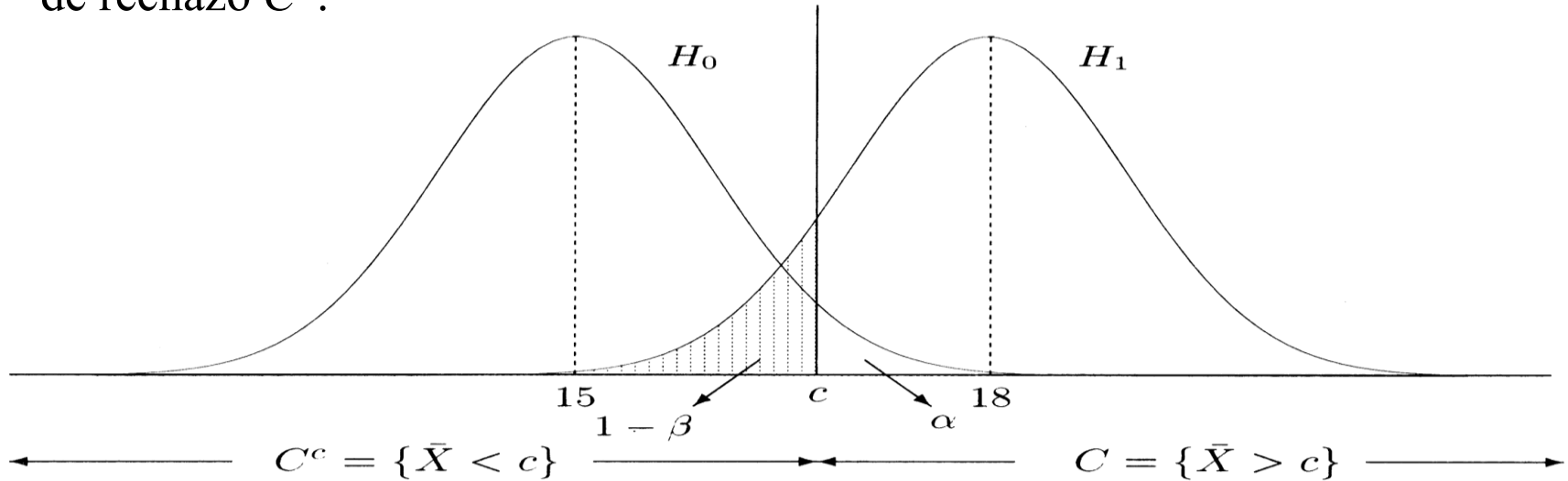
$$c = 15 + \frac{z_{\alpha}}{\sqrt{n}}$$

donde z_{α} es el valor tal que

$$P\{Z > z_{\alpha}\} = \alpha = 0.05 \text{ si } Z \equiv N(0,1).$$

Contrastes de hipótesis

Fases 5 y 6: Extraemos la muestra de tamaño n , calculamos el valor de la media muestral y vemos si cae en la región de aceptación C o en la región de rechazo C^c .



Tests de diferencias de medias

En evaluación de impacto, interesa comprobar los efectos, es decir, si hay diferencia significativa entre las medias de las variables de resultado para el grupo de tratamiento y el grupo de control.

Por ello, cobran especial relevancia los tests de diferencias de medias.

Tests de diferencias de medias

Sea (X, Y) una variable aleatoria bidimensional cuyas componentes son independientes y siguen, respectivamente, distribuciones $N(\mu_1, \sigma_1)$ y $N(\mu_2, \sigma_2)$. Para comprobar diferencias de medias entre X e Y , se utiliza como estadístico la variable aleatoria:

$$\bar{X} - \bar{Y} : \mathcal{X} = \mathbf{R}^{n+m} \rightarrow \mathbf{R}$$

En el espacio muestral \mathcal{X} se considera la distribución de la muestra $(X_1, X_2, \dots, X_n, Y_1, Y_2, \dots, Y_m)$. Por ser X e Y independientes, podemos considerarlas como variables aleatorias separadas y por ello tener muestras de tamaño diferente.

Tests de diferencias de medias

Dependiendo de las suposiciones que hagamos en las hipótesis nula y alternativa, el estadístico $\bar{X} - \bar{Y}$ habrá que modificarlo o transformarlo para conseguir que su distribución en el muestreo sea conocida y, por tanto, fácil de trabajar.

En todos los casos, consideramos $H_0 : \mu_1 = \mu_2$ y las tres posibles hipótesis alternativas siguientes:

- ▶ $H_1 : \mu_1 \neq \mu_2$. Test bilateral o de dos colas.
- ▶ $H_2 : \mu_1 > \mu_2$. Test unilateral o de una cola (cola superior).
- ▶ $H_3 : \mu_1 < \mu_2$. Test unilateral o de una cola (cola inferior).

Tests de diferencias de medias

En primer lugar, supongamos que tanto σ_1 como σ_2 son conocidas.

En este caso, el estadístico para el contraste y las regiones críticas son:

$$Z_0 = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \equiv N(0,1)$$

$H_1 : \mu_1 \neq \mu_2$	$H_2 : \mu_1 > \mu_2$	$H_3 : \mu_1 < \mu_2$
$ z_0 > z_{\alpha/2}$	$z_0 > z_{\alpha}$	$z_0 < -z_{\alpha}$

Tests de diferencias de medias

En segundo lugar, supongamos que tanto σ_1 como σ_2 son desconocidas y $\sigma_1 \neq \sigma_2$. En este caso, el estadístico para el contraste y las regiones críticas, debidas a la aproximación de Welch, son:

$$T_0 = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{m}}} \equiv t_f \xrightarrow{\ell} N(0,1), f = \frac{\left(\frac{S_1^2}{n} + \frac{S_2^2}{m} \right)^2}{\frac{(S_1^2/n)^2}{n+1} + \frac{(S_2^2/n)^2}{m+1}}$$

$H_1 : \mu_1 \neq \mu_2$	$H_2 : \mu_1 > \mu_2$	$H_3 : \mu_1 < \mu_2$
$ t_0 > t_{\alpha/2, f}$	$t_0 > t_{\alpha, f}$	$t_0 < -t_{\alpha, f}$

Tests de diferencias de medias

En tercer lugar, supongamos que tanto σ_1 como σ_2 son desconocidas pero $\sigma_1 = \sigma_2$.

En este caso, el estadístico para el contraste y las regiones críticas son:

$$T_0 = \frac{\bar{X} - \bar{Y}}{S_p \sqrt{\frac{1}{n} + \frac{1}{m}}} \equiv t_{n+m-2}, S_p = \frac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2}$$

$H_1 : \mu_1 \neq \mu_2$	$H_2 : \mu_1 > \mu_2$	$H_3 : \mu_1 < \mu_2$
$ t_0 > t_{\alpha/2, n+m-2}$	$t_0 > t_{\alpha, n+m-2}$	$t_0 < -t_{\alpha, n+m-2}$

Ejercicio: diferencias de medias en una población con tres grupos

Disponemos de una muestra de tamaño 100 de tres poblaciones, identificadas como grupos A, B y C, y queremos contrastar diferencias entre el grupo de control B y los otros dos grupos, A y C.

Realice con el *software* Stata sendos tests de diferencias de medias entre los grupos A-B y B-C e interprete los resultados.

- ▶ ¿Existe diferencia entre A y B?
- ▶ ¿Existe diferencia entre B y C?

- ▶ Ficheros: *Diferencias de medias AB.dta* y *Diferencias de medias BC.dta*
- ▶ Tiempo: 10 minutos
- ▶ Ejercicio individual

Ejercicio: diferencias de medias en una población con tres grupos

Considere ahora la salida del *software* JMP de SAS de un test de diferencias de medias entre los tres grupos.

- ▶ ¿Hay evidencia de que los tres grupos son distintos en la característica medida?
- ▶ ¿Refrenda esta salida sus conclusiones al comparar los grupos dos a dos?
- ▶ ¿Cuántos grupos considera que hay en realidad en la muestra: dos o tres?

- ▶ Fichero: *Diferencia de medias A B C informe.TIFF*
- ▶ Tiempo: 10 minutos
- ▶ Ejercicio en grupo

Bibliografía

Principios de inferencia estadística

Autores: Ricardo Vélez Ibarrola y Alfonso García Pérez
UNED, 1993

Técnicas cuantitativas de evaluación de políticas públicas

Lima, 10 de septiembre de 2013



«Datos, datos, datos. No puedo fabricar ladrillos sin arcilla».

Probabilidad condicionada

Concepto de probabilidad condicionada

Teorema de la probabilidad total

Teorema de Bayes o de la probabilidad inversa

Probabilidad condicionada

Sea $\{ \Omega , \mathcal{A}, P \}$ un espacio probabilístico y $A \in \mathcal{A}$ un suceso tal que $P(A) > 0$. Llamaremos probabilidad condicionada del suceso $B \in \mathcal{A}$ respecto al suceso A a

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

$\mathcal{A}_A = \mathcal{A} \cap A$ es una σ -álgebra y la probabilidad $P(\cdot | A)$ es una probabilidad sobre \mathcal{A}_A . Por ello, $\{A, \mathcal{A}_A, P(\cdot | A)\}$ es un espacio probabilístico si $P(A) > 0$.

Teorema de la probabilidad total

Sea $\{\Omega, \mathcal{A}, P\}$ un espacio probabilístico y sea $\{A_n\}_{n \in \mathbb{N}} \subset \mathcal{A}$ un sistema completo de sucesos, es decir, una sucesión de sucesos disjuntos tal que

$$A_i \cap A_j = \emptyset, \forall i \neq j; \bigcup_{n=1}^{\infty} A_n = \Omega$$

Sea $B \in \mathcal{A}$ un suceso para el que se conocen las probabilidades condicionadas $P(B|A_i)$ y supongamos que se conocen también las probabilidades $P(A_i) > 0, \forall i \in \mathbb{N}$. Entonces

$$P(B) = \sum_{i=1}^{\infty} P(B|A_i)P(A_i)$$

Teorema de Bayes o de la probabilidad inversa (1764)

Sea $\{\Omega, \mathcal{A}, P\}$ un espacio probabilístico y sea $\{A_n\}_{n \in \mathbb{N}} \subset \mathcal{A}$ un sistema completo de sucesos tal que $P(A_i) > 0, \forall i \in \mathbb{N}$. Sea $B \in \mathcal{A}$ un suceso con $P(B) > 0$, para el que se conocen las probabilidades condicionadas $P(B|A_i)$. Entonces,

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{i=1}^{\infty} P(B|A_i)P(A_i)}, \forall i \in \mathbb{N}$$

A las probabilidades $P(A_i)$ se les suele llamar probabilidades *a priori*; a las $P(A_i|B)$, probabilidades *a posteriori*; y a las $P(B|A_i)$, verosimilitudes.

Teorema de Bayes o de la probabilidad inversa (1764)

$\{A_n\}_{n \in \mathbb{N}}$ se consideran hipótesis sobre una cuestión en estudio y tenemos una cierta información, subjetiva o no, acerca de dichas hipótesis, reflejadas en $P(A_i)$. Supongamos que se produce el suceso B asociado a un experimento aleatorio, usualmente provocado por el investigador. El teorema de Bayes nos proporciona el medio para obtener la alteración provocada por B en nuestras hipótesis mediante las probabilidades *a posteriori* $P(A_i | B)$. Tras esto, el investigador provocará otro suceso C del experimento aleatorio y utilizará las $P(A_i | B)$ antes calculadas como probabilidades *a priori* para obtener nuevas probabilidades *a posteriori*.

Teorema de Bayes o de la probabilidad inversa (1764)

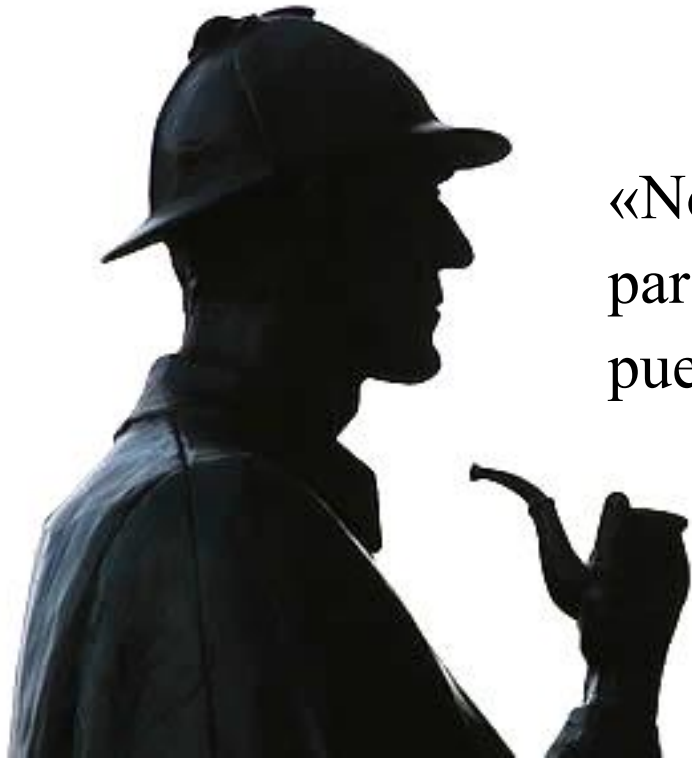
Ejemplo: Tenemos una urna cerrada con cinco bolas cuya composición es desconocida. Hay dos posibilidades: la urna contiene tres bolas rojas y dos bolas blancas (hipótesis que llamaremos Urna I); o la urna contiene cuatro bolas blancas y una roja (Urna II).

Sea A_1 el suceso “la urna es la Urna I” y A_2 el suceso “la urna es la Urna II”. Al principio: $P(A_1) = P(A_2) = 1/2$. Extraemos una bola de la urna y provocamos el suceso B : la bola extraída es blanca. Por el teorema de Bayes, calculamos $P(A_1|B) = 1/3$ y $P(A_2|B) = 2/3$. Utilizamos estas probabilidades condicionadas como nuevas probabilidades *a priori* y extraemos una segunda bola para provocar otro suceso C que nos dé más información sobre la composición de la urna.

Ejercicio: teorema de Bayes

En el ejemplo anterior, considere que el suceso B es: “la primera bola extraída es roja” y el suceso C : “la segunda bola extraída es también roja”. Calcule las probabilidades *a posteriori* tras conocer ambos sucesos. ¿Qué conclusión obtenemos de dichas probabilidades *a posteriori*?

- ▶ Tiempo: 10 minutos
- ▶ Ejercicio individual



«No cabe combinación de acontecimientos
para los que la inteligencia del hombre no
pueda concebir una explicación».

Nociones de econometría

Distribución condicionada

Modelos de regresión lineal: la recta de regresión

Estimación por mínimos cuadrados ordinarios (MCO)

Diagnos de modelos econométricos

Problema de endogeneidad

Método de las variables instrumentales

Estimación por mínimos cuadrados en dos etapas: 2SLS

Distribución condicionada

Sea un vector aleatorio n -dimensional $\xi = (X_1, X_2, \dots, X_n) : \Omega \rightarrow \mathbf{R}^n$ que sigue una distribución $N(\mu, \Sigma)$. Consideramos una partición del vector ξ en dos vectores: un vector p -dimensional ξ_1 , y un segundo vector $(n-p)$ -dimensional ξ_2 . Podemos suponer, sin pérdida de generalidad, que ξ_1 está compuesto por las p primeras componentes de ξ , mientras que ξ_2 está compuesto por las $(n-p)$ últimas primeras componentes de ξ .

Distribución condicionada

Con esta división podemos escribir:

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \boldsymbol{\Sigma} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

Si llamamos

$$\boldsymbol{\beta} = \Sigma_{12} \Sigma_{22}^{-1} \quad y \quad \Sigma_{11.2} = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$$

La variable aleatoria $\xi_1 \mid \xi_2 = \mathbf{x}_2$ tiene una distribución $N(\mu_1 + \boldsymbol{\beta} (\mathbf{x}_2 - \mu_2), \Sigma_{11.2})$. El punto hace referencia a que la variable indicada ha sido fijada.

Recta de regresión

La ecuación

$$\mathbf{x}_1 - \boldsymbol{\mu}_1 = \Sigma_{12} \Sigma_{22}^{-1} (\mathbf{x}_2 - \boldsymbol{\mu}_2)$$

representa la línea general de regresión de ξ_1 sobre ξ_2 . La matriz $\Sigma_{12} \Sigma_{22}^{-1}$ se denomina matriz de regresión. El elemento que ocupa el lugar (i,j) en la matriz de regresión lo denominamos $\beta_{ij.p+1,p+2,\dots,n}$.

Recta de regresión

A los elementos de la matriz de covarianzas de la distribución condicionada $\Sigma_{11.2}$ los denotamos por $\sigma_{ij.p+1,p+2,\dots,n}$.

El coeficiente de correlación entre X_i y X_j calculado a partir de la matriz de covarianzas de la distribución condicionada se denomina coeficiente de correlación parcial entre X_i y X_j cuando $\xi_2 = (X_{p+1}, X_{p+2}, \dots, X_n)$ ha sido fijado. Lo representamos por $\rho_{ij.p+1,p+2,\dots,n}$ y viene dado por:

$$\rho_{ij.p+1,p+2,\dots,n} = \frac{\sigma_{ij.p+1,p+2,\dots,n}}{\sqrt{\sigma_{ii.p+1,p+2,\dots,n} \sigma_{jj.p+1,p+2,\dots,n}}}$$

Recta de regresión

En resumen, la variable aleatoria

$$\xi_{1.2} = \xi_1 - \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2 - \mu_2)$$

denominada residual, representa las discrepancias entre los verdaderos valores de ξ_1 y el valor esperado. $\xi_{1.2}$ cumple las siguientes condiciones:

$$\begin{aligned}\xi_{1.2} &\equiv N(0, \Sigma_{12} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}) \\ E[\xi_{1.2}(\xi_2 - \mu_2)'] &= \Sigma_{12} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{22} = 0\end{aligned}$$

Recta de regresión

Podemos por tanto considerar el siguiente modelo de previsión para la variable ξ_1 :

$$\xi_1 = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2 - \mu_2) + \xi_{1.2}$$

Esta es la base matemática de la regresión lineal por mínimos cuadrados.

Recta de regresión

Ahora, llamando δ_i al vector fila i -ésimo de la matriz Σ_{12} , suponemos $\mu=0$ y consideramos una combinación lineal cualquiera del vector ξ_2 , obtenida multiplicado ξ_2 por un vector fila cualquiera ζ . Entonces la combinación lineal de ξ_2 que minimiza la varianza de $X_i - \zeta \xi_2$ y maximiza la correlación entre X_i y $\zeta \xi_2$ se obtiene para el vector $\zeta = \beta_i$ donde $\beta_i = \delta_i \Sigma_{22}^{-1}$.

Recta de regresión

Al coeficiente de correlación entre X_i y $\beta_i \xi_2$ se le denomina coeficiente de correlación múltiple. Se representa por $\rho_{i.p+1,p+2,\dots,n}$ y se calcula de la siguiente forma:

$$\rho_{i.p+1,p+2,\dots,n} = \frac{\sqrt{\delta_i \Sigma_{22}^{-1} \delta_i'}}{\sqrt{\sigma_{ii}}}$$

Recta de regresión

Siendo $\sigma_{ii.p+1,p+2,\dots,n}$ la varianza de $X_i \mid \xi_2 = \mathbf{x}_2$, se cumple que:

$$\sigma_{ii.p+1,p+2,\dots,n} = \sigma_{ii} \left(1 - \rho_{i.p+1,p+2,\dots,n}^2 \right)$$

por lo que el coeficiente de correlación múltiple mide la proporción de varianza de X_i atribuible a la variación de $\xi_2 = (X_{p+1}, X_{p+2}, \dots, X_n)$

Recta de regresión en dos dimensiones

Consideremos dos variables aleatorias X e Y cuyo coeficiente de correlación lineal $\rho \neq 0$. Queremos hallar la recta que mejor ajuste la relación lineal entre ambas variables a través de unas pocas observaciones obtenidas a través de una muestra aleatoria de tamaño n $\{(x_i, y_i)\}_{i=1, 2, \dots, n}$. Dicha relación vendrá dada por la siguiente expresión:

$$Y = \alpha + \beta X + u \quad , \quad u \equiv N(0, \sigma)$$

Recta de regresión en dos dimensiones

El problema se reduce a determinar los coeficientes α y β de la ecuación que mejor se ajuste a los n pares (x_i, y_i) observados, en el sentido de que se minimice la distancia entre lo encontrado (y_i) y lo pronosticado (\hat{y}_i). Estas distancias o diferencias entre los valores se denominan residuos y se denotan como u_i .

Recta de regresión en dos dimensiones

Los estimadores de los parámetros vienen dados por:

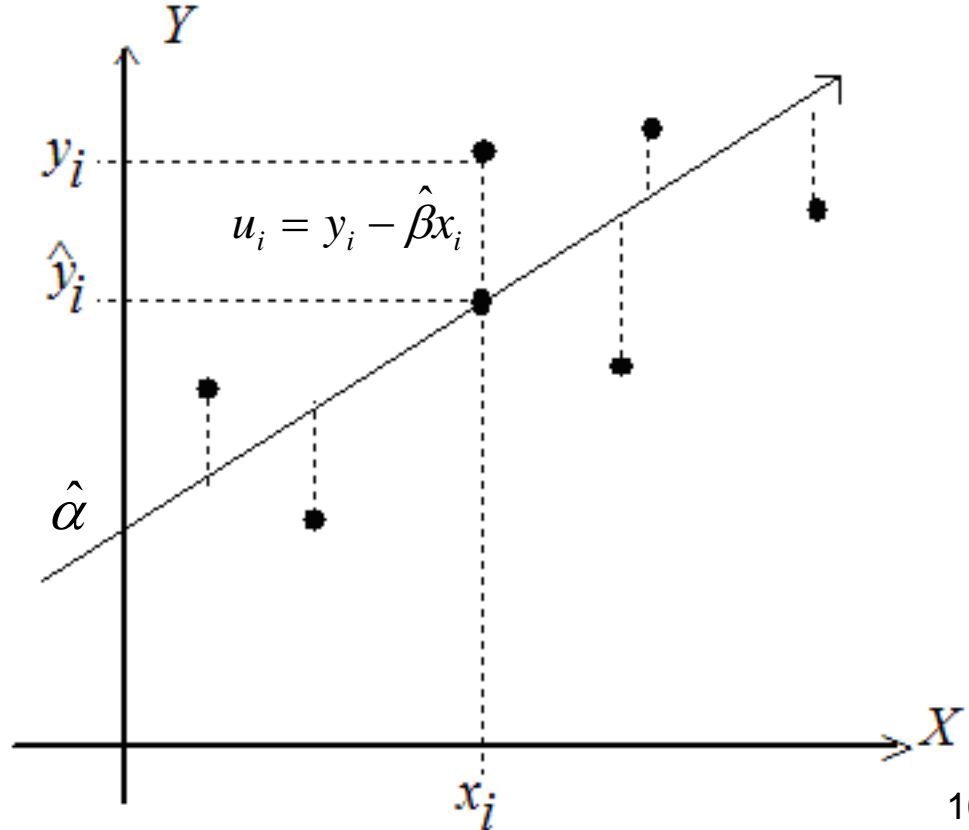
$$\hat{\beta} = \frac{s_{XY}}{s_X^2}; \hat{\alpha} = \bar{y} - \hat{\beta} \bar{x}$$

El parámetro α representa la ordenada en el origen, esto es, el valor que toma la variable dependiente cuando la variable independiente toma el valor 0. El parámetro β es la pendiente de la recta de regresión.

Recta de regresión en dos dimensiones

Gráficamente:

$$Y = \hat{\alpha} + \hat{\beta}X$$



Recta de regresión en dos dimensiones

Geométricamente, la recta de regresión lineal de Y con respecto a X minimiza la suma de las distancias verticales de los puntos a la recta. La recta de regresión lineal de X con respecto a Y minimiza las distancias horizontales. Las dos rectas se cortan en el centro de gravedad (\bar{x}, \bar{y}) de la nube de puntos. La separación entre las dos rectas es mayor cuando la correlación es más débil.

Estimación por mínimos cuadrados ordinarios (MCO)

El modelo poblacional a estudiar es:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + u \quad , \quad u \equiv N(0, \sigma)$$

donde Y es una variable dependiente unidimensional y X_i , $i=1, \dots, k$, son variables aleatorias unidimensionales con distribución conocida, llamadas regresores.

Estimar un modelo consiste en encontrar valores para β_i , que denotamos $\hat{\beta}_i$, $i = 1, 2, \dots, k$, a partir de una muestra aleatoria.

Estimación por mínimos cuadrados ordinarios (MCO)

De forma equivalente, podemos escribir el modelo poblacional en forma matricial:

$$Y = \boldsymbol{\beta}\mathbf{X} + u \quad , \quad u \equiv N(0, \sigma)$$

donde

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}, \quad \mathbf{X} = (1 \quad X_1 \quad X_2 \quad \cdots \quad X_k)$$

Estimación por mínimos cuadrados ordinarios (MCO)

Si se cumplen los dos siguientes supuestos:

$$E[\mathbf{X}'u] = 0 \Rightarrow E[u] = 0, E[u|\mathbf{X}] = 0$$
$$rango(\mathbf{X}'\mathbf{X}) = k$$

el modelo está identificado, es decir: los parámetros β_i , $i=0, \dots, k$, pueden escribirse en términos de momentos de las variables observables X_i . Operando, obtenemos:

$$\boldsymbol{\beta} = E[\mathbf{X}'\mathbf{X}]^{-1} E[\mathbf{X}'Y]$$

Estimación por mínimos cuadrados ordinarios (MCO)

Tomando una muestra aleatoria de tamaño N de la población, $\{y_i, \mathbf{x}_i=(1, x_{i1}, x_{i2}, \dots, x_{ik})\}_{i=1, 2, \dots, N}$, el estimador para los parámetros se obtiene sustituyendo los momentos poblacionales por los momentos muestrales:

$$\hat{\boldsymbol{\beta}} = \left(N^{-1} \sum_{i=1}^N \mathbf{x}_i' \mathbf{x}_i \right)^{-1} \left(N^{-1} \sum_{i=1}^N \mathbf{x}_i' y_i \right) = (\mathbf{x}' \mathbf{x})^{-1} \mathbf{x}' \mathbf{y}$$

donde

$$\mathbf{x} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1j} & \cdots & x_{1k} \\ \vdots & \vdots & & \vdots & & \vdots \\ 1 & x_{i1} & \cdots & x_{ij} & \cdots & x_{ik} \\ \vdots & \vdots & & \vdots & & \vdots \\ 1 & x_{N1} & \cdots & x_{Nj} & \cdots & x_{Nk} \end{pmatrix} \quad \mathbf{e} \mathbf{y} = \begin{pmatrix} y_1 \\ y_i \\ y_N \end{pmatrix}$$

Estimación por mínimos cuadrados ordinarios (MCO)

Si además se cumple el supuesto débil de homocedasticidad

$$E[u^2 \mathbf{X}' \mathbf{X}] = \sigma^2 E[\mathbf{X}' \mathbf{X}],$$

que implica que el error cuadrático está incorrelado con cada x_i , x_i^2 o cualquier producto cruzado de la forma $x_i x_j$, $\forall 1 \leq i < j \leq k$, podemos asegurar que, asintóticamente,

$$\sqrt{N}(\hat{\beta} - \beta) \xrightarrow{\ell, N \rightarrow \infty} N(0, \sigma^2 A^{-1})$$

donde $A = E[\mathbf{X}' \mathbf{X}]$.

Ejercicio 1: estimación por MCO

Tenemos una muestra aleatoria de los valores de antigüedad (en años) y de precio (en centenas de dólares) de 10 coches Corvettes.

Ajuste un modelo lineal por mínimos cuadrados ordinarios con los datos.

¿Cree que el modelo lineal es suficiente para explicar la relación existente entre la antigüedad del vehículo y su precio de venta?

- ▶ Fichero: *Corvettes.dta*
- ▶ Tiempo: 5 minutos
- ▶ Ejercicio individual

Ejercicio 1: estimación por MCO

Esta es la salida de Stata.

```
. regress price age
```

Source	SS	df	MS
Model	24057.8913	1	24057.8913
Residual	1623.70874	8	202.963592
Total	25681.6	9	2853.51111

Number of obs = **10**
F(1, 8) = **118.53**
Prob > F = **0.0000**
R-squared = **0.9368**
Adj R-squared = **0.9289**
Root MSE = **14.247**

price	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
age	-27.90291	2.562889	-10.89	0.000	-33.81295	-21.99288
_cons	371.6019	11.4329	32.50	0.000	345.2376	397.9663

Ejercicio 2: estimación por MCO

Tenemos una muestra aleatoria de tamaño 100 con una variable dependiente Y y 5 variables independientes X_1, \dots, X_5 .

Ajuste un modelo lineal por mínimos cuadrados ordinarios con los datos.

¿Son todas las variables relevantes?

- ▶ Fichero: *MCO.dta*
- ▶ Tiempo: 5 minutos
- ▶ Ejercicio individual

Ejercicio 2: estimación por MCO

Esta es la salida de Stata con las variables X_1 , X_2 , X_3 , X_4 y X_5 .

```
. regress y x1 x2 x3 x4 x5
```

Source	SS	df	MS	Number of obs = 100		
Model	45984.9296	5	9196.98592	F(5, 94) =13126.70		
Residual	65.8593881	94	.700631788	Prob > F = 0.0000		
Total	46050.789	99	465.159485	R-squared = 0.9986		
				Adj R-squared = 0.9985		
				Root MSE = .83704		

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
x1	4.887785	.1742717	28.05	0.000	4.541765	5.233806
x2	2.878389	.0141319	203.68	0.000	2.850329	2.906448
x3	-10.59501	.0739852	-143.20	0.000	-10.74191	-10.44812
x4	-.0019691	.1729999	-0.01	0.991	-.3454644	.3415262
x5	-.0102901	.0283126	-0.36	0.717	-.0665054	.0459252
_cons	1.610365	1.077654	1.49	0.138	-.529342	3.750072

Ejercicio 2: estimación por MCO

Esta es la salida de Stata con las variables X_1 , X_2 y X_3 .

```
. regress y x1 x2 x3
```

Source	SS	df	MS	Number of obs = 100		
Model	45984.8365	3	15328.2788	F(3, 96) =22311.75		
Residual	65.9524569	96	.687004759	Prob > F = 0.0000		
Total	46050.789	99	465.159485	R-squared = 0.9986		
				Adj R-squared = 0.9985		
				Root MSE = .82886		

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
x1	4.882145	.1683644	29.00	0.000	4.547944	5.216346
x2	2.878164	.0138633	207.61	0.000	2.850645	2.905682
x3	-10.59433	.0728564	-145.41	0.000	-10.73895	-10.44971
_cons	1.452581	.9772035	1.49	0.140	-.4871528	3.392315

Econometría

La econometría es la rama de la economía que hace un uso extensivo de modelos matemáticos y estadísticos así como de la programación lineal y la teoría de juegos para analizar, interpretar y hacer predicciones sobre sistemas económicos.

La econometría, en tanto que “medida de los procesos económicos”, provee modelos y técnicas útiles a la hora de evaluar el óptimo uso de los recursos.

Diagnosis de modelos econométricos

En econometría la diagnosis de un modelo es el punto esencial.

Tras construir un modelo (por ejemplo, mediante mínimos cuadrados ordinarios) hay que realizar su correcta diagnosis.

Para ello, existen contrastes de hipótesis matemáticos bien formulados.

Solo cuando un modelo pasa la diagnosis, se pueden hacer predicciones.

Diagnosis de modelos econométricos

Para diagnosticar un modelo, realizamos en este orden los siguientes contrastes. Los marcados con asterisco (*) pueden obviarse para la diagnosis de un modelo provisional, pero deben incluirse en el modelo definitivo.

- ▶ Significatividad individual de los parámetros estimados: los parámetros no pueden ser nulos.
- ▶ Significatividad conjunta de los parámetros estimados: no todos pueden ser nulos.
- ▶ R^2 : indica el buen ajuste del modelo. Es suficiente pero no necesario.

Diagnosis de modelos econométricos

- ▶ R^2 : indica el buen ajuste del modelo. Es suficiente pero no necesario.
- ▶ Ausencia de autocorrelación. Si en el modelo hay varios términos de error u_t , $t = 1, \dots, T$, (por ejemplo cuando se trabaja con series temporales o con varias variable dependientes), se debe cumplir que $\text{Cov}(u_r, u_s) = 0$, $\forall r \neq s$.
- ▶ Ausencia de heterocedasticidad. $V(u | X_1, X_2, \dots, X_k) = \sigma^2$.
- ▶ *Ausencia de endogeneidad. Las variables X_1, X_2, \dots, X_k no están correlacionadas con el término de error u : $E(u | X_1, X_2, \dots, X_k) = 0$.

Diagnosis de modelos econométricos

- ▶ Ausencia de multicolinealidad. Las variables X_1, X_2, \dots, X_k son linealmente independientes, es decir, no existe relación lineal exacta entre ellas. Esta hipótesis se denomina hipótesis de independencia, y cuando no se cumple, decimos que el modelo presenta multicolinealidad.
- ▶ *El modelo debe presentar linealidad. La variable Y es aleatoria, ya que depende de la variable aleatoria u . Además, la relación entre Y y X_1, X_2, \dots, X_k es efectivamente lineal.

Diagnosis de modelos econométricos

- ▶ *El modelo debe estar bien especificado: forma funcional, variables omitidas y redundantes. Ausencia de errores de especificación: suponemos que todas las variables X_i que son relevantes para la explicación de la variable Y están incluidas en la definición del modelo lineal.
- ▶ Ausencia de regresores estocásticos. En el modelo, las variables X_1, X_2, \dots, X_k se consideran deterministas (no variables aleatorias), ya que su valor es constante al provenir de una muestra.
- ▶ Normalidad de los residuos. Todas las variables $u_t \equiv N(0, \sigma_t)$.

Problema de endogeneidad

En los diseños cuasiexperimentales es habitual la presencia de sesgo de selección. Eso implica que no se verifica una de las condiciones para estimar el modelo lineal por MCO, a saber:

$$E[\mathbf{X}'u] \neq 0 \Rightarrow E[u|\mathbf{X}] \neq 0$$

Es decir: una o varias de las X_i no son exógenas.

Problema de endogeneidad

Podemos suponer, sin pérdida de generalidad, que solo X_k podría ser endógena mientras que son exógenas el resto de las X_i , $i=1, \dots, k-1$. Una posible razón es que u contiene una variable omitida (no observada) q incorrelada con todos los regresores menos con X_k .

Método de las variables instrumentales

El método de las variables instrumentales resuelve el problema de variables exógenas entre los regresores. Requerimos contar con una variable observable, Z_1 , distinta a todas las X_i , $i=1, \dots, k$, que satisfaga las siguientes condiciones:

$$E[Z_1 u] = 0 \Rightarrow E[u|Z_1] = 0$$

$$\exists \theta \neq 0 : X_k = \delta_0 + \delta_1 X_1 + \delta_2 X_2 + \dots + \delta_{k-1} X_{k-1} + \theta Z_1 + r_k, \quad r_k \equiv N(0, \sigma_k)$$

$$E[(X_1, \dots, X_{k-1}, Z_1) r_k] = 0$$

Entonces podemos decir que Z_1 es una posible variable instrumental o un posible instrumento para X_k y la ecuación se llama forma reducida de X_k .

Método de las variables instrumentales

En el modelo a estimar, sustituimos X_k por la expresión de su forma reducida para obtener:

$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \cdots + \alpha_{k-1} X_{k-1} + \lambda_1 Z_1 + v, v = u + \beta_k r_k \equiv N(0, \sigma_{Z_1})$$

$$\alpha_j = \beta_j + \beta_k \delta_j, j = 1, \cdots, k-1$$

$$\lambda_1 = \beta_k \theta_1$$

donde v es el error de la forma reducida y α_j y λ_1 los parámetros de la forma reducida. En esta ecuación, gracias a nuestros supuestos, se cumple que:

$$E[(X_1, \cdots, X_{k-1}, Z_1)v] = 0$$

y por ello dichos parámetros se pueden estimar por MCO.

Mínimos cuadrados en dos etapas: 2SLS

Supongamos ahora que tenemos más de un instrumento para X_k , es decir tenemos $m > 1$ variables Z_h , $h=1, \dots, m$, distintas todas ellas de las X_i , $i=1, \dots, k$, que satisfacen las siguientes condiciones:

$$E[Z_h u] = 0 \Rightarrow E[u|Z_h] = 0, \forall h = 1, \dots, m$$

$$\forall h = 1, \dots, m, \exists \theta_h \neq 0: X_k = \delta_0 + \delta_1 X_1 + \delta_2 X_2 + \dots + \delta_{k-1} X_{k-1} + \theta_1 Z_1 + \dots + \theta_m Z_m + r_k$$

$$r_k \equiv N(0, \sigma_k)$$

$$E[(X_1, \dots, X_{k-1}, Z_1, \dots, Z_m) r_k] = 0$$

Mínimos cuadrados en dos etapas: 2SLS

En este caso, hacemos la regresión de X_k sobre X_i y Z_h , $i=1,\dots,k-1$, $h=1,\dots,m$, y obtenemos una estimación de X_k , que denotamos como X_k^* . En la ecuación original utilizamos X_k^* en lugar de la X_k y resolvemos por MCO.

De hecho, X_k^* se interpreta a menudo como la parte de X_k incorrelada con u . Si X_k es endógena, es porque r_k está correlado con u :

$$E[ur_k] \neq 0$$

Mínimos cuadrados en dos etapas: 2SLS

Vamos a ver ahora el caso general. Supongamos el modelo:

$$Y = \beta \mathbf{X} + u, \quad u \equiv N(0, \sigma)$$

donde

$$\mathbf{X} = (1 \quad X_1 \quad X_2 \quad \dots \quad X_k), \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}$$

pero algunos elementos de \mathbf{X} están correlados con u :

$$E[\mathbf{X}'u] = 0 \Rightarrow E[u|\mathbf{X}] = 0$$

Mínimos cuadrados en dos etapas: 2SLS

Supongamos que existe un vector \mathbf{Z} de dimensiones $1 \times L$ que incluye constante que cumple:

$$E[\mathbf{Z}'u] = 0 \Rightarrow E[u|\mathbf{Z}] = 0$$

y satisface la condición de rango

$$\text{rango } E[\mathbf{Z}'\mathbf{Z}] = L$$

$$\text{rango } E[\mathbf{Z}'\mathbf{X}] = k$$

y la condición de orden

$$L > k$$

Mínimos cuadrados en dos etapas: 2SLS

Tomando una muestra aleatoria de tamaño N de la población, $\{y_i, \mathbf{x}_i=(1, x_{i1}, x_{i2}, \dots, x_{ik}), \mathbf{z}_i=(1, z_{i1}, z_{i2}, \dots, z_{iL})\}_{i=1, 2, \dots, N}$, el estimador 2SLS para los parámetros se obtiene en función de los momentos muestrales:

$$\hat{\boldsymbol{\beta}} = \left[\left(\sum_{i=1}^N \mathbf{x}_i' \mathbf{z}_i \right) \left(\sum_{i=1}^N \mathbf{z}_i' \mathbf{z}_i \right) \left(\sum_{i=1}^N \mathbf{z}_i' \mathbf{x}_i \right) \right]^{-1} \left(\sum_{i=1}^N \mathbf{x}_i' \mathbf{z}_i \right) \left(\sum_{i=1}^N \mathbf{z}_i' \mathbf{z}_i \right) \left(\sum_{i=1}^N \mathbf{z}_i' y_i \right) = [(\mathbf{x}' \mathbf{z})(\mathbf{z}' \mathbf{z})(\mathbf{z}' \mathbf{x})]^{-1} (\mathbf{x}' \mathbf{z})(\mathbf{z}' \mathbf{z})(\mathbf{z}' \mathbf{y})$$

donde:

$$\mathbf{x} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1j} & \cdots & x_{1k} \\ \vdots & \vdots & & \vdots & & \vdots \\ 1 & x_{i1} & \cdots & x_{ij} & \cdots & x_{ik} \\ \vdots & \vdots & & \vdots & & \vdots \\ 1 & x_{N1} & \cdots & x_{Nj} & \cdots & x_{Nk} \end{pmatrix}, \mathbf{z} = \begin{pmatrix} 1 & z_{11} & \cdots & z_{1h} & \cdots & z_{1L} \\ \vdots & \vdots & & \vdots & & \vdots \\ 1 & z_{i1} & \cdots & z_{ih} & \cdots & z_{iL} \\ \vdots & \vdots & & \vdots & & \vdots \\ 1 & z_{N1} & \cdots & z_{Nh} & \cdots & z_{NL} \end{pmatrix} \text{ e } \mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_i \\ \vdots \\ y_N \end{pmatrix}$$

Mínimos cuadrados en dos etapas: 2SLS

Si además se cumple el supuesto débil de homocedasticidad

$$E[u^2 \mathbf{Z}' \mathbf{Z}] = \sigma^2 E[\mathbf{Z}' \mathbf{Z}]$$

que implica que el error cuadrático está incorrelado con cada Z_i , Z_i^2 o cualquier producto cruzado de la forma $Z_i Z_j$, $\forall 1 \leq i < j \leq L$, podemos asegurar que, asintóticamente:

$$\sqrt{N}(\hat{\beta} - \beta) \xrightarrow{\ell, N \rightarrow \infty} N(0, \sigma^2 A^{-1})$$

donde

$$A = E[\mathbf{X}' \mathbf{Z}] E[\mathbf{Z}' \mathbf{Z}]^{-1} E[\mathbf{Z}' \mathbf{X}].$$

Ejercicio: problema de endogeneidad

Considere su diseño del programa PIPE. ¿Cree que puede haber algún problema de endogeneidad? En caso afirmativo:

- ▶ ¿Qué variable o variables pueden presentar dicho problema?
- ▶ ¿Qué instrumento o instrumentos podría utilizar para subsanar el problema de endogeneidad?
- ▶ Tiempo: 15 minutos
- ▶ Ejercicio en grupo

Bibliografía

Introductory Econometrics: A Modern Approach, fifth international edition

Autor: Jeffrey M. Wooldridge

South-Western *CENGAGE* Learning, 2013

Econometric Analysis of Cross Section and Panel Data, second edition

Autor: Jeffrey M. Wooldridge

MIT Press, 2010

Econometric Analysis (7th Edition)

Autor: William H. Greene

Prentice Hall, 2012

¡Gracias por su atención!

Joaquín María Núñez Varo
Analista del Departamento de Evaluación
ICEX España Exportación e Inversiones
joaquin.m.nunez@icex.es