

Especialización en Estadística Aplicada

Curso Propedéutico: Estadística

Javier Santibáñez

IIMAS, UNAM

`jsantibanez@sigma.iimas.unam.mx`

Junio 2017

En general, se puede representar a los fenómenos que ocurren en la naturaleza como sigue:



La representación anterior es conocida como *modelo*.

Un modelo es una abstracción que simplifica la realidad y permite estudiarla con mayor detalle.

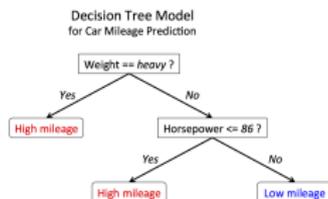
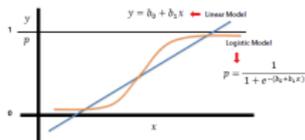
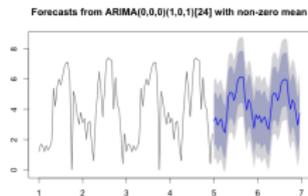
Existen dos corrientes científicas para describir cómo ocurren los fenómenos:

- **Determinismo:** las condiciones en las que se da el fenómeno determinan completamente su resultado.
- **Indeterminismo:** se admite la existencia de fenómenos en los que no es posible saber de antemano cuál será su resultado, aún controlando las condiciones en las que se desarrolla.

P. S. Laplace

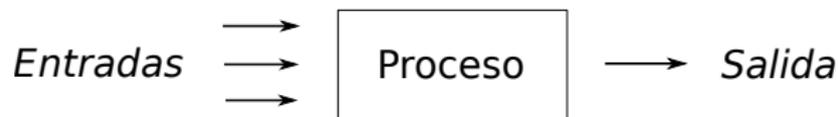
Si fuera posible conocer todas las leyes que rigen todos los procesos que ocurren en el universo y el estado actual de todos los objetos que lo componen además, si fuera posible analizar tales datos, se podría abarcar en una sola fórmula los movimientos de todos los objetos en el universo; nada resultaría incierto y tanto el futuro como el pasado estarían presentes ante nuestros ojos.

Desarrollo de expresiones matemáticas para describir, en algún sentido, la relación que existe entre un conjunto de variables que describen las condiciones en que se desarrolla un fenómeno y una variable asociada su resultado.



Modelos deterministas

En el caso determinista, las entradas determinan completamente la salida:



Si x_1, \dots, x_p son las entradas y y la salida o respuesta, podemos modelar matemáticamente el proceso como:

$$y = f(x_1, \dots, x_p)$$

El objetivo es determinar una buena elección para f , se pueden utilizar distintos criterios para elegir a un candidato.

Ejemplo. Caída libre

Cuenta la historia que Galileo Galilei descubrió experimentalmente que la distancia que recorre un cuerpo en caída libre, despreciando la resistencia del aire, está dada por:

$$d(t) = \frac{1}{2}gt^2$$

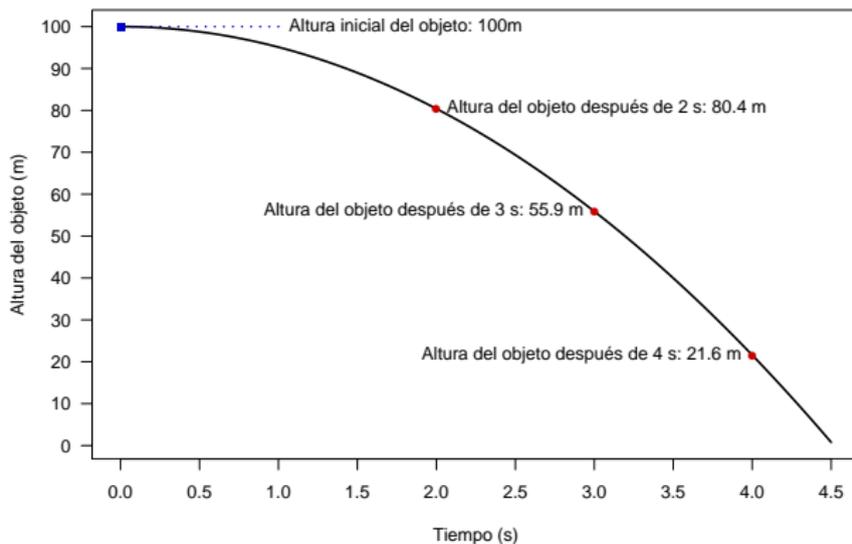
donde:

- t tiempo transcurrido,
- $g = 9.8ms^{-2}$ (aceleración de la gravedad),
- $d(t)$ distancia recorrida al tiempo t .

Caída libre (cont.)

Ejemplo

Se deja caer un objeto desde 100 m altura. ¿Cuál será la altura del objeto después de $t = 2, 3, 4$ s?



Crecimiento poblacional

- El crecimiento exponencial se puede utilizar para modelar el crecimiento o decrecimiento de una población.
- Si el crecimiento/decrecimiento se da a una tasa proporcional al tamaño de la población, entonces se puede modelar el tamaño de la población al tiempo t como

$$N(t) = N_0 \exp^{-\lambda t}$$

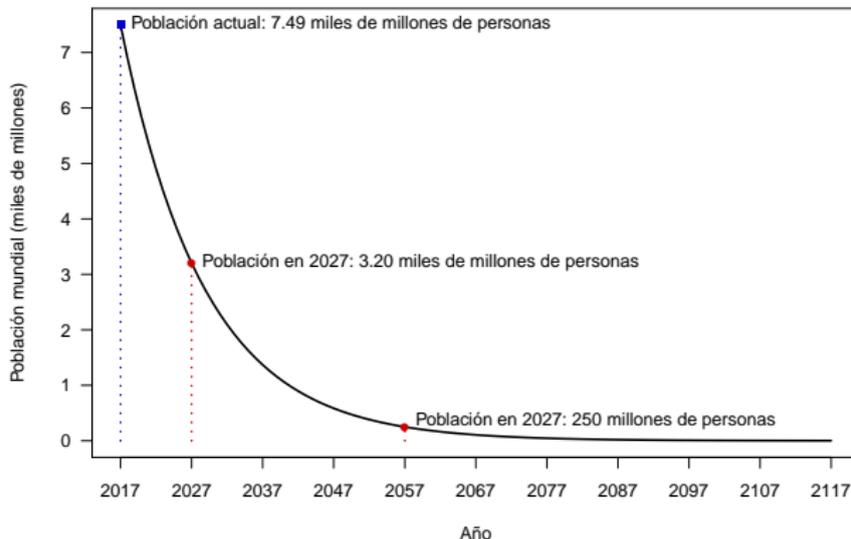
donde:

- N_0 es la población inicial (tiempo $t = 0$),
- t es el tiempo transcurrido,
- λ es la tasa de crecimiento/decrecimiento,
- $N(t)$ es la población al tiempo t .

Crecimiento poblacional (cont.)

Ejemplo

La población de mundial el 02/02/2017 es aproximadamente 7,494 millones de personas. Si en la población se detuvieran los nacimientos y la tasa de mortalidad se mantuviera constante en 8.5% anual:



En el caso estocástico, el objetivo es modelar matemáticamente la relación entre las condiciones en que se desarrolla un fenómeno y la incertidumbre de sus posibles respuestas.



Si x_1, \dots, x_p son las entradas y y la salida o respuesta, podemos modelar matemáticamente el proceso como:

$$y \sim F(x_1, \dots, x_p)$$

El objetivo es encontrar un buen modelo probabilista F , estamos ante un problema típico de inferencia estadística.

Ejemplo. El estado del tiempo

Es posible modelar el estado del tiempo como una *cadena de Markov*. Consideremos tres estados posibles $S = \{1, 2, 3\}$ donde: 1 es soleado, 2 es nublado, 3 es lluvioso. Suponemos que la matriz de transición P está dada por:

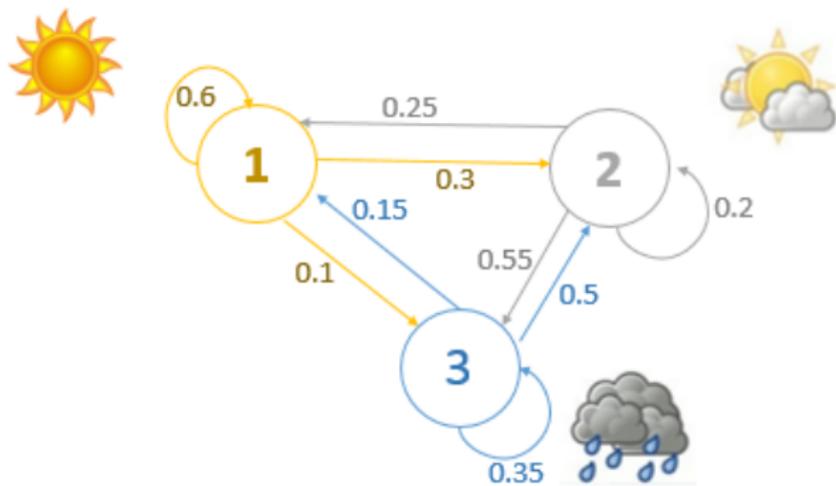
$$P = \begin{bmatrix} 0.60 & 0.30 & 0.10 \\ 0.25 & 0.20 & 0.55 \\ 0.15 & 0.50 & 0.35 \end{bmatrix}$$

La i, j -ésima entrada de P representa la probabilidad de que el tiempo de mañana sea j dado que hoy es i , $i, j = 1, 2, 3$. Por ejemplo:

$$P(\text{soleado}|\text{soleado}) = 0.60$$

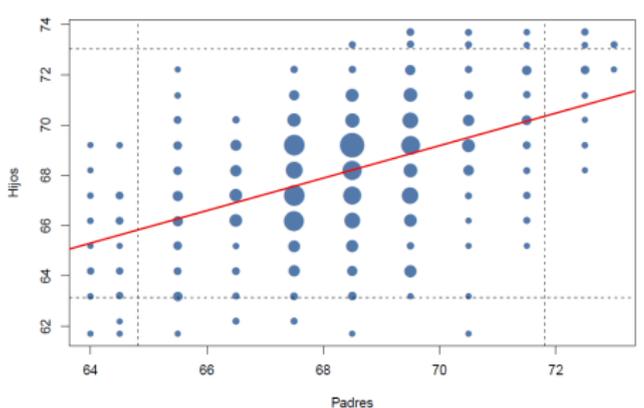
$$P(\text{nublado}|\text{lluvioso}) = 0.50$$

$$P(\text{lluvioso}|\text{soleado}) = 0.10$$



Ejemplo. Regresión de Galton

Francis Galton en el siglo XIX observó la relación que había entre la altura de padres e hijos, notó que aunque existe una tendencia de que un padre alto tenga un hijo alto y padre bajo tenga un hijo bajo, la distribución de las estaturas, no cambia drásticamente de una generación a otra (a este fenómeno se le conoce como regresión a la media).



Los modelos matemáticos se usan para describir la relación funcional entre las variables de entrada (x_1, x_2, \dots, x_p) y la variable respuesta y .

$$Y = f(x_1, x_2, \dots, x_p) \quad (\text{Determinista})$$

$$Y \sim F(x_1, x_2, \dots, x_p) \quad (\text{Probabilístico})$$

- ¿Qué función f (o F) tomar?
- ¿El modelo es adecuado? ¿Existe el mejor modelo?
- ¿Qué supuestos debemos tomar en cuenta?
- ¿Las variables x_i explican a y ?

¿Qué es la estadística?

Estadística es un conjunto de técnicas para la colección, manejo, descripción y análisis de información de manera que las conclusiones obtenidas de ella tengan un grado de confiabilidad especificado.

Supongamos que se nos proponen los siguientes *juegos*:

- 1 Lanzar dos dados y si la suma de sus resultados es menor a 5 o mayor a 9 ganamos 1,000 MXN, de otro modo perdemos 1,000 MXN.
- 2 Lanzar 10 volados y si caen entre 4 y 6 águilas, ganamos 1,000 MXN, de otro modo perdemos 1,000 MXN.

Si asumimos que los dados y la moneda no están cargados ¿en qué juego nos conviene participar?

Hay dos tipos de variables: categóricas y numéricas.

- 1 **Categóricas:** corresponden a mediciones no cuantificables (color de ojos, estado civil, nacionalidad, nivel educativo, entre otras). Se clasifican en dos tipos:
 - **Ordinales:** cuando el registro de la medición se expresa en grados de intensidad que tienen un orden, pero no se puede determinar el incremento entre los grados (nivel educativo, grado de satisfacción bueno-regular-malo, calificaciones MB-B-S-NA).
 - **Nominales:** cuando las categorías sólo se les da un nombre pero no tienen un orden entre ellas. Además las categorías de una variable nominal deben ser mutuamente excluyentes exhaustivas (color de ojos, estado civil, nacionalidad, código postal).

- 1 Numéricas: corresponden a mediciones cuantificables (edad, estatura, peso, años de escolaridad, entre otras).
 - Discretas: tienen rango de valores *a lo más numerable* (puede ser infinito, pero se puede contar). Estas variables corresponden generalmente a conteos (el número de estrellas en cierta región del cielo, el número de cabellos en la cabeza de una persona, el número de células en un determinado organismo).
 - Continuas: tienen rango de valores no numerable (no es posible enumerar sus elementos). Virtualmente toda medición de una magnitud física es continua sin embargo, ésta se vuelve discreta debido a la precisión de los aparatos de medición (masa, peso, longitud, tiempo, velocidad).

Las escalas de medición están determinadas por el tipo de variable sus características. Hay cuatro escalas de medición

- Nominal: incluye a las variables nominales. La única relación permitida es la igualdad. Sólo es posible hacer conteos de frecuencias.
- Ordinal: incluye a las variables ordinales. Se pueden hacer comparaciones de igualdad y de orden. Es posible hacer conteos de frecuencias y calcular estadísticos de orden (mediana y percentiles).
- De intervalos: incluye a las variables numéricas. Existe la noción de distancia entre sus distintos valores. Es posible hacer cualquier tipo de operación aritmética. Si incluye un cero éste es arbitrario.
- De razón: incluye las variables numéricas que tiene un cero absoluto, que indica la ausencia de la característica.

En general, en estadística se estudian variables en escala de intervalo y de razón, aunque hay modelos y métodos para estudiar observaciones en las otras escalas.

Considerar la descripción de las bases de datos del Módulo de Condiciones Socioeconómicas de la ENIGH 2015 (disponible [aquí](#)). Identificar los tipos y escalas de medición de las variables incluidas en la tabla de viviendas (p. 27).

- En estadística el término *población* se utiliza para representar al conjunto formado por los elementos que se tiene interés en estudiar. Aunque es un concepto fundamental, no siempre se especifica adecuadamente y se asume que está bien definido.
- Generalmente el interés se centra en estudiar una o varias características de los elementos de la población. A las observaciones o mediciones correspondientes a tales características se les conoce como *variables*.
- Las poblaciones de interés son *grandes*, incluso pueden ser infinitas. La estadística se encarga de hacer inferencias sobre una o varias variables en la población entera a partir de observar éstas variables en un subconjunto de la población al que se llama *muestra*.
- La forma de estudiar las poblaciones depende del número de elementos que contienen. Si la población tiene un número finito de elementos entonces se conoce como *población finita*. Hacer inferencias en poblaciones finitas es el objeto de estudio del *muestreo*.

- Se utiliza el término población infinita para representar al modelo de población que potencialmente pudiera tener infinitos elementos.
- Ejemplo, cuando se habla de la probabilidad de que una moneda caiga en águila en un volado, nos referimos a una característica de una potencial sucesión de infinitos volados.
- Para estudiar una característica de una población infinita se utilizan los modelos de probabilidad. De manera que se asume que la variable correspondiente sigue tal o cual distribución.
- De acuerdo con la interpretación frecuentista de la probabilidad, si se observa una muestra lo suficientemente grande de elementos de la población, las frecuencias en la que aparecen los valores de la variables de interés corresponden a las probabilidades dadas por la distribución.
- La forma de estudiar las poblaciones infinitas depende del modelo que se utilice y del tipo de observaciones que se hagan.

- Nuestro interés será, en la mayoría de los casos, estudiar una variable X de una población (infinita) de la cual se asume que la frecuencia de valores de X puede modelarse con una distribución de probabilidades $F(\cdot | \theta)$, la cuál no está completamente especificada y se deben hacer inferencias sobre la parte desconocida a partir de observar X para algunos elementos de la población.
- En el planteamiento anterior se asume que hay conjunto de elementos la población que serán observados o medidos. Las mediciones aún no realizadas se denotan por X_1, X_2, X_3, \dots . Este conjunto se conoce como muestra aleatoria y se denota por $S = \{X_1, X_2, \dots, X_n\}$, asumiendo que hay n de tales elementos.
- Formalmente, una *muestra aleatoria* de una población infinita es un conjunto de variables *independientes e idénticamente distribuidas* según el modelo de probabilidad asumido para la población.
- Se denota por $s = \{x_1, x_2, \dots, x_n\}$ para representar a la muestra *observada*, en la que x_i corresponde a la medición realizada al i -ésimo elemento en la muestra, es decir, x_i es la realización de X_i , para $i = 1, \dots, n$.

- La *estadística descriptiva* es el conjunto de técnicas analíticas y gráficas que se utilizan para describir un determinado conjunto de datos, posiblemente una muestra de una población finita o infinita. Generalmente el interés se centra en estudiar el comportamiento de una cierta variable en el conjunto de datos o estudiar posibles asociaciones entre pares de variables.
- Hay dos tipos de características que son importantes para describir la *distribución* o comportamiento de una variable: la tendencia central y la dispersión. La idea de tendencia central es: el valor de la variable al que *mejor* describe todo el conjunto de valores. La idea de la dispersión es: qué tanto difieren las observaciones con respecto a su valor *central*.
- Las medidas de asociación buscan estudiar cómo se relacionan la ocurrencia de ciertos valores entre un par de variables. por ejemplo, si valores altos en la variable X están asociados con valores altos (o bajos) de la variable Y .
- El tipo de estadísticas descriptivas a utilizar depende de la escala de medición de las variables.

- Los estadísticos de orden de una variable numérica sirven para identificar la posición de una medición con respecto al orden de las mediciones. Suponer que tenemos una muestra aleatoria $S = \{X_1, X_2, \dots, X_n\}$, de alguna distribución.
- El primer estadístico de orden se denota por $X_{(1)}$, corresponde a la menor observación de toda la muestra y se le llama *mínimo*.
- El último (n -ésimo) estadístico de orden se denota por $X_{(n)}$, corresponde a la mayor observación de la muestra y se le llama *máximo*.
- El i -ésimo estadístico de orden se denota por $X_{(i)}$ y corresponde a la observación de la muestra que ocupa de i -ésima posición cuando ésta se ordena de menor a mayor.
- La distribución de los estadísticos de orden $X_{(i)}$ no necesariamente corresponde a la distribución de las variables originales X_i .
- Una vez que se ha observado la muestra los estadísticos de orden se pueden *calcular* y se denotan por $x_{(1)}, \dots, x_{(n)}$.

Medidas de tendencia central

Suponer que se tienen las siguientes observaciones de una variable: x_1, x_2, \dots, x_n . Se definen las siguientes medidas de tendencia central.

- **Media:** sólo aplica para variables numéricas, se calcula como promedio aritmético de las observaciones y se denota por \bar{x} .

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i.$$

- **Mediana:** aplica para variables ordinales y numéricas. Cuando la variable es ordinal, se define como la categoría tal que el 50% o más de las observaciones tiene una categoría menor y el 50% o más tiene una categoría mayor. Cuando las variables son numéricas, se define la mediana como sigue:

$$\text{Mediana} = \begin{cases} x_{(\frac{n+1}{2})} & \text{si } n \text{ es impar} \\ \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2} & \text{si } n \text{ es par} \end{cases}$$

- **Moda:** aplica para variables de cualquier escala y se define como el valor más frecuente, por lo que puede no existir o bien tomar más de un valor. Cuando hay un empate de k valores o categorías más frecuentes, se dice entonces que hay k modas, excepto cuando $k = n$, todos los valores o categorías tienen la misma frecuencia, entonces se dice que no hay moda.
- Es común que en las variables categóricas haya más de una moda y que en las variables numéricas no exista.

Medidas de dispersión

- **Varianza:** aplica sólo para variables numéricas, se denota por s^2 y se calcula como

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2.$$

- **Desviación estándar:** es la raíz cuadrada de la varianza, tiene la ventaja que está expresada en las mismas unidades que las mediciones originales.
- **Desviación absoluta:** aplica sólo para variables numéricas. Se calcula como la suma de las desviaciones absolutas de las observaciones con respecto a su media.

$$\text{Desv. Abs.} = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|.$$

Tiene la ventaja que está expresada en las mismas unidades que las variables originales.

- **Rango:** aplica sólo para variables numéricas, es la longitud del menor intervalo que contiene a todas las observaciones y se calcula como

$$\text{Rango} = x_{(n)} - x_{(1)}.$$

Ejemplo.

- Identificar el tipo y la escala de medición de las variables del siguiente conjunto de datos.
- Calcular los estadísticos descriptivos que correspondan.

Nombre	Sexo	Edad	Estado civil	Educación	Ingreso (MXN)
Alicia	F	30	Soltero	Licenciatura	18,000.00
Beto	M	45	Viudo	Secundaria	14,000.00
Carlos	M	28	Soltero	Preparatoria	11,000.00
Diana	F	29	Soltero	Licenciatura	12,000.00
Elena	F	40	Casado	Secundaria	15,500.00
Felipe	M	42	Casado	Preparatoria	16,000.00

- **Cuartiles:** son una medida de tendencia central, aplican sólo para variables numéricas. Son tres y se denotan por Q_1 , Q_2 y Q_3 .
 - ★ el 25 % de las observaciones es menor o igual a Q_1 .
 - ★ Q_2 es la mediana, es decir, 50 % de las observaciones es menor o igual a Q_2 .
 - ★ el 75 % de las observaciones es menor o igual a Q_r .
- **Rango intercuartil:** es una medida de dispersión, aplica sólo para variables numéricas. Se calcula como

$$\text{Rango intercuartil} = Q_3 - Q_1.$$

- **Coefficiente de variación:** es una medida de dispersión, aplica sólo para variables numéricas y se calcula como

$$CV = \frac{s}{\bar{x}}, \quad \text{si } \bar{x} \neq 0$$

donde s es la desviación estándar. Generalmente se expresa como porcentaje.

Si $n \geq 4$:

- Para calcular Q_1 primero se obtiene el cociente $n/4$. Si la parte entera es e y la parte decimal es d , entonces:

$$Q_1 = x_{(e)} + d \times (x_{(e+1)} - x_{(e)}).$$

- Para calcular Q_3 primero se obtiene el cociente $3n/4$. Si la parte entera es e y la parte decimal es d , entonces:

$$Q_3 = x_{(e)} + d \times (x_{(e+1)} - x_{(e)}).$$

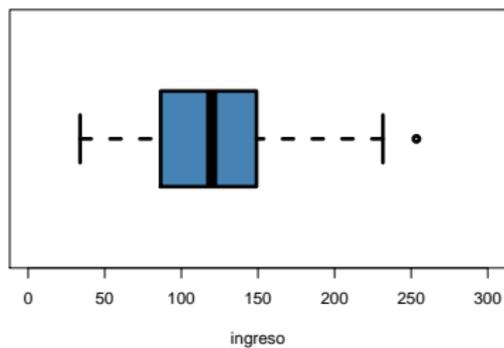
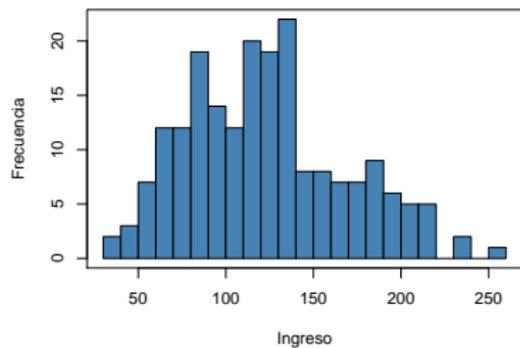
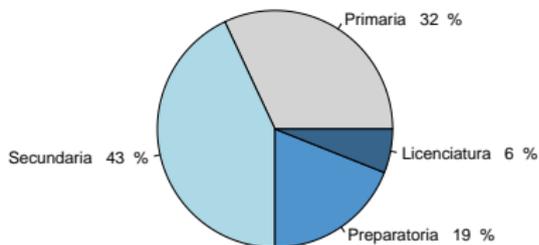
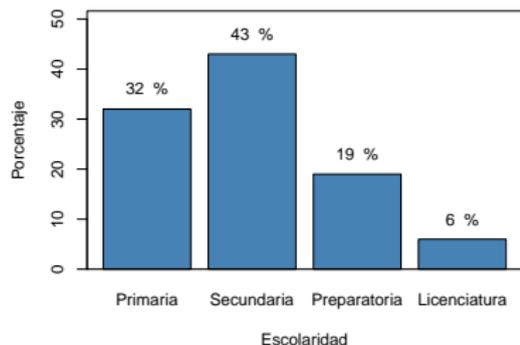
Considerar el siguiente conjunto de observaciones:

$$s = \{10, 4, 3, 7, 9, 3, 6, 5, 6, 19, 15, 3, 14, 25\}$$

Calcular: media, mediana, moda, rango, varianza, cuartiles, rango intercuartil y coeficiente de variación.

- **Barras:** se utiliza para representar frecuencias (absolutas o relativas) de mediciones de una variable binaria. En el eje X se ubican las posibles categorías y se utilizan barras verticales cuya altura representa la frecuencia.
- **Histograma:** se utiliza para estudiar la dispersión de una variable numérica. Es un gráfico de barras para una variable numérica que se vuelve categórica, se debe definir el número de categorías y estas se forman igualmente espaciadas en todo el rango de la variable.
- **Caja:** se utiliza para estudiar la distribución de una variable numérica. También conocido como gráfico de caja y bigotes. La caja está delimitada por Q_1 y Q_3 , al centro de agrega Q_2 y los bigotes se calculan como $Q_1 - 1.5RI$ y $Q_3 + 1.5IR$. Se agregan puntos más allá de los bigotes si hay observaciones con valores más extremos.
- **Dispersión:** se utiliza para estudiar la asociación entre un par de variables. Es un gráfico con dos ejes, en el que se representa con puntos a los pares de observaciones (x_1, y_2) . Se utiliza para estudiar asociaciones entre pares de variables.

Ejemplos



- El objetivo de la Inferencia Estadística es hacer inferencias sobre una característica de una población, modelada con una función de distribución $F(\cdot | \theta)$ donde θ es un parámetro desconocido.
- El conjunto de valores posibles de θ se conoce como espacio parametral y se denota por Θ .
- Las inferencias sobre θ se basan en una muestra aleatoria $S = \{X_1, X_2, \dots, X_n\}$, que es un conjunto de variables aleatorias *independientes e idénticamente distribuidas* como $F(\cdot | \theta)$.

Definición (estadístico)

Un *estadístico* es cualquier función de una muestra aleatoria S , que no depende de valores desconocidos.

Como los estadísticos son funciones de la muestra aleatoria, también son variables aleatorias.

Definición (estimador)

Sea una población modelada por $F(\cdot | \theta)$ con $\theta \in \Theta$ desconocido y una muestra aleatoria $S = \{X_1, X_2, \dots, X_n\}$. Un *estimador* de θ es un estadístico con el que se pretende *aproximar* el verdadero valor de θ . Lo usual es denotar por $\hat{\theta}$ a un estimador para θ .

En el curso se *Inferencia Estadística* se ven distintas técnicas de estimación, es decir, distintos criterios para determinar *buenos* estimadores, de momento nos limitaremos a estudiar algunas de sus propiedades más generales.

Definición (Error Cuadrático Medio)

Sea una población modelada por $F(\cdot | \theta)$ con $\theta \in \Theta$ desconocido y una muestra aleatoria $S = \{X_1, X_2, \dots, X_n\}$ y $\hat{\theta}$ un estimador de θ . Se define el *Error Cuadrático Medio* de $\hat{\theta}$ como sigue

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

- El ECM de un estimador es una medida de dispersión con respecto al verdadero valor del parámetro.
- Se utiliza el ECM para comparar dos estimadores de un mismo parámetro, de manera que se prefieren los estimadores con menor ECM.
- Generalmente el ECM es función del o los parámetros desconocidos, por lo que dados dos estimadores $\hat{\theta}_1$ y $\hat{\theta}_2$ de θ , $\hat{\theta}_1$ puede tener menor ECM que $\hat{\theta}_2$ para ciertos valores de θ , pero para otros valores de θ , $\hat{\theta}_2$ puede tener menor ECM que $\hat{\theta}_1$.

Definición (Sesgo)

Sea una población modelada por $F(\cdot | \theta)$ con $\theta \in \Theta$ desconocido y una muestra aleatoria $S = \{X_1, X_2, \dots, X_n\}$ y $\hat{\theta}$ un estimador de θ . Se define el *sesgo* de $\hat{\theta}$ como sigue

$$\text{Sesgo}(\hat{\theta}) = \theta - E(\hat{\theta}).$$

- El sesgo de un estimador es una medida de la distancia entre su valor esperado y el parámetro desconocido, de manera que lo ideal es tener sesgos pequeños en valor absoluto.
- Si el sesgo de un estimador es cero, significa que el valor esperado de $\hat{\theta}$ coincide con el verdadero parámetro que está estimando.

Definición (estimador insesgado)

Sea una población modelada por $F(\cdot | \theta)$ con $\theta \in \Theta$ desconocido y una muestra aleatoria $S = \{X_1, X_2, \dots, X_n\}$. Se dice que un estimador $\hat{\theta}$ es insesgado para θ si $\text{Sesgo}(\hat{\theta}) = 0$ o de manera equivalente $E(\hat{\theta}) = \theta$.

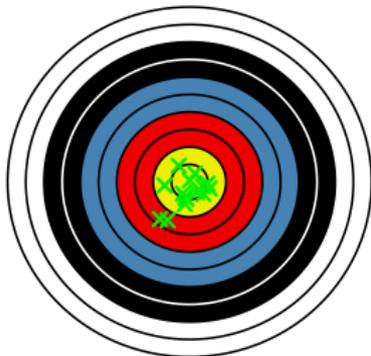
El ECM de un estimador se puede escribir de la siguiente forma

$$ECM(\hat{\theta}) = Var(\hat{\theta}) + Sesgo^2(\theta).$$

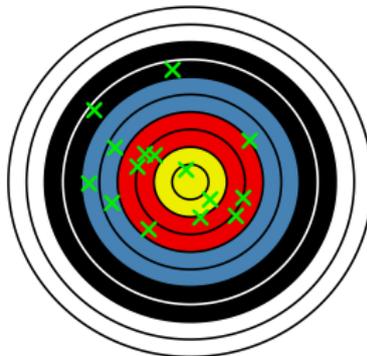
La ecuación anterior nos dice que el ECM de un estimador depende de su propia variabilidad, medida $Var(\hat{\theta})$, y de su posición con respecto al verdadero parámetro, medida por $Sesgo^2(\hat{\theta})$. De manera que los mejores estimadores serán aquellos que tengan poca varianza y poco sesgo.

Ejemplo.

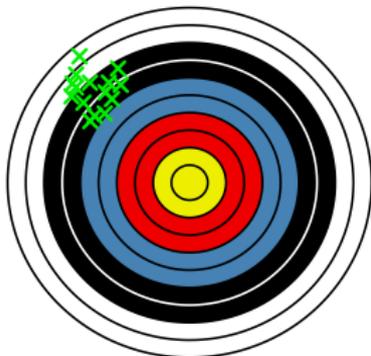
ECM pequeño



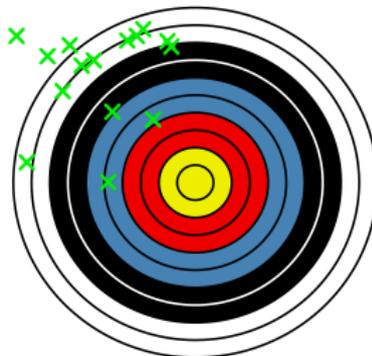
ECM moderado



ECM moderado



ECM grande



- Dada una población $F(\cdot | \theta)$, no existe un estimador $\hat{\theta}$ que tenga menor ECM que cualquier otro estimador de θ .
- Uno de los enfoques clásicos en inferencia estadística es restringir la búsqueda de estimadores únicamente a aquellos que son insesgados.
- Si $\hat{\theta}$ es un estimador insesgado para θ , entonces $ECM(\hat{\theta}) = Var(\hat{\theta})$. Por lo que el mejor estimador insesgado es aquel que tiene la varianza mínima.

Resultado

Si X y Y variables aleatorias *independientes* y a, b, c constantes conocidas, entonces se cumple

- 1 $E(a + bX + cY) = a + bE(X) + cE(Y)$.
- 2 $Var(a + bX + cY) = b^2 Var(X) + c^2 Var(Y)$.

Del resultado anterior se siguen las igualdades:

- $E(a) = a$.
- $E(bX) = bE(X)$.
- $E(X + Y) = E(X) + E(Y)$.
- $Var(a) = 0$.
- $Var(bX) = b^2 Var(X)$.
- $Var(X + Y) = Var(X) + Var(Y)$, sólo vale si X y Y son independientes.

- ① Sea $S = \{X_1, X_2, \dots, X_{10}\}$ una muestra aleatoria de una distribución con esperanza μ y varianza σ^2 , con μ y σ^2 los parámetros desconocidos. Considerar los siguientes estimadores de μ :

$$\hat{\mu}_1 = X_1, \quad \hat{\mu}_2 = \frac{X_1 + X_2}{2} \quad \text{y} \quad \hat{\mu}_3 = \frac{X_1 + X_2 + \dots + X_{10}}{10}.$$

Calcular el sesgo y el ECM de los tres estimadores ¿cuál es el mejor estimador para μ ?

- ② Si $S = \{X_1, X_2, \dots, X_5\}$ es una muestra aleatoria de una población $Poisson(\lambda)$, donde λ es el parámetro desconocido. Considerar las siguientes funciones de S :

$$\begin{aligned} \hat{\lambda}_1 &= \frac{X_1 + X_2}{2}, & \hat{\lambda}_2 &= \frac{X_1 + 3X_2 + X_3}{5} \\ \hat{\lambda}_3 &= \lambda \frac{X_1 + X_3 + X_5}{3}, & \hat{\lambda}_4 &= \frac{X_1 + X_2 + \dots + X_5}{5}. \end{aligned}$$

¿Cuáles son estimadores de λ ? Calcular el sesgo y el ECM de los estimadores ¿cuál es el mejor estimador para μ ?

Distribución normal

Se dice que una variable aleatoria X con soporte en $(-\infty, \infty)$, tiene o sigue una distribución normal con parámetros μ y σ^2 si su función de densidad está dada por

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{1}{2\sigma^2}(x - \mu)^2 \right\}.$$

Lo anterior se denota como $X \sim N(\mu, \sigma^2)$. Cuando $\mu = 0$ y $\sigma^2 = 1$ se dice que la distribución es normal estándar y en ese caso

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2}x^2 \right\}.$$

La distribución $N(\mu, \sigma^2)$ tiene las siguientes características:

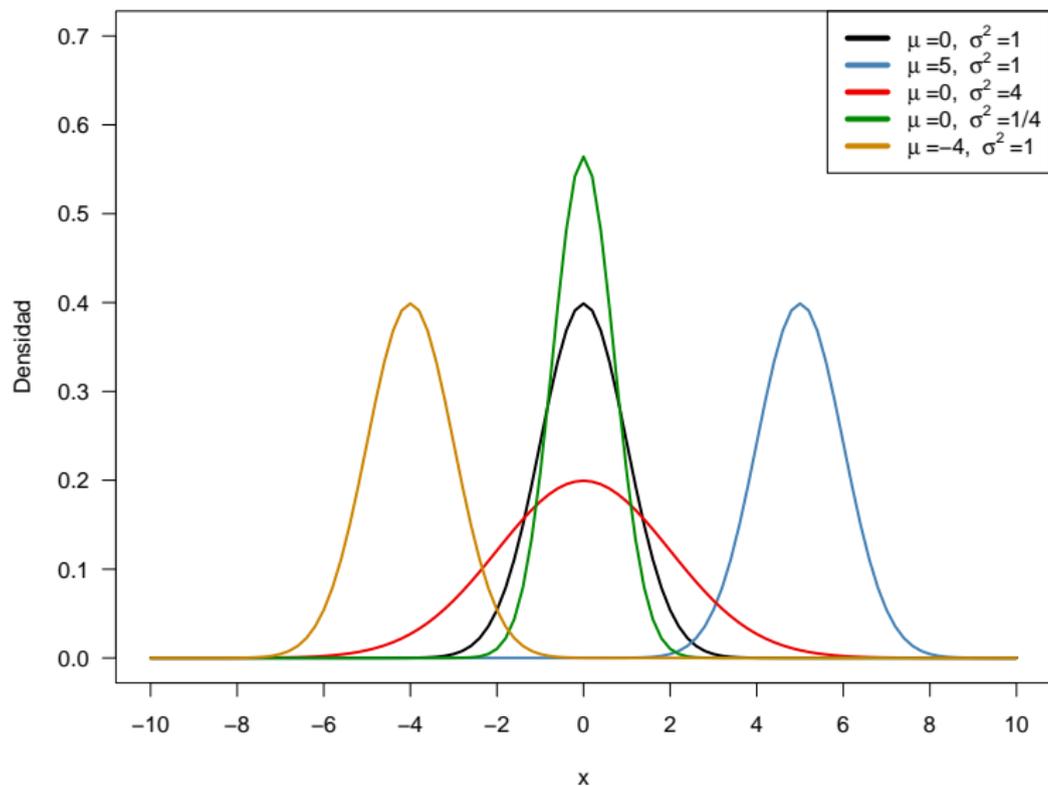
- Tiene esperanza μ y varianza σ^2 .
- Es simétrica alrededor de su esperanza.
- Si $X \sim N(\mu, \sigma^2)$, entonces

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

A esta operación se le conoce como estandarización.

- Si $X \sim N(0, 1)$, entonces $Y = \mu + \sigma X \sim N(\mu, \sigma^2)$.

Distribución normal

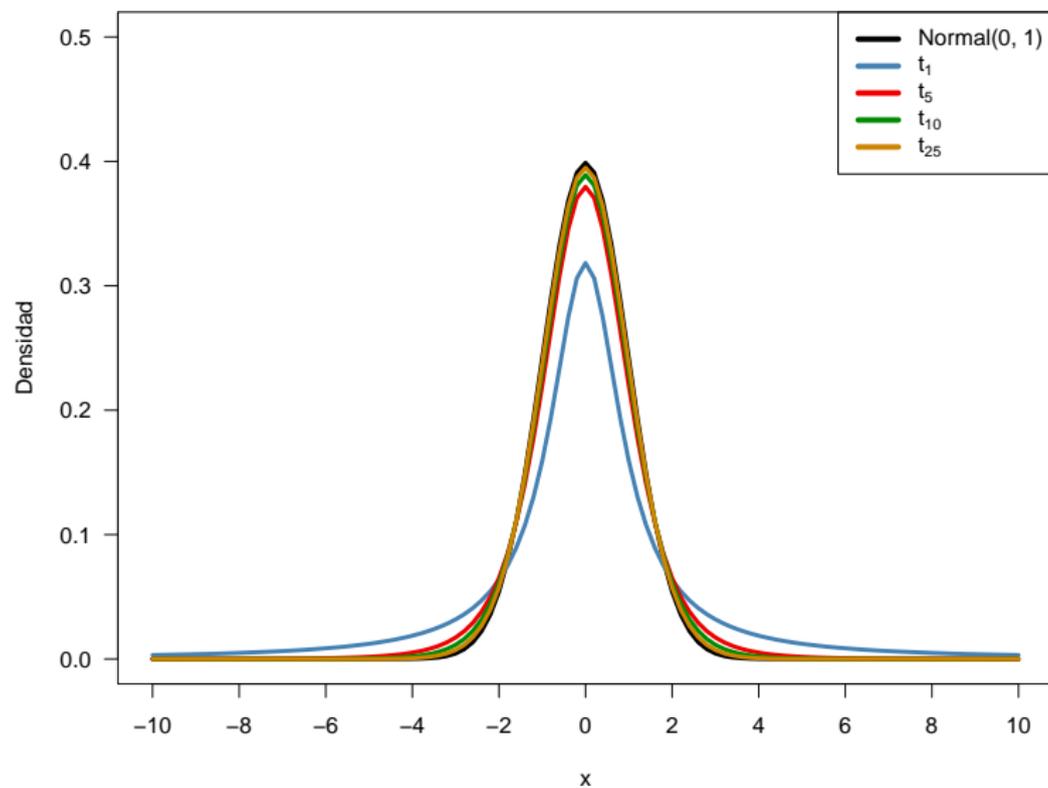


La distribución t tiene una función de distribución bastante complicada por lo que no vale la pena mostrarla en este momento.

- Está relacionada con la distribución de ciertos estadísticos de muestras aleatorias normales.
- El parámetro que caracteriza esta distribución recibe el nombre de *grados de libertad*.
- La distribución t estándar está centrada en cero y su forma es similar a la distribución normal estándar.
- A medida que los grados de libertad crecen, la distribución t se parece más a la normal estándar.
- Si $X \sim t_n$, entonces

$$E(X) = 0 \quad \text{y} \quad \text{Var}(X) = 2n.$$

Distribución normal



Distribución de la media muestral

Sea $S = \{X_1, X_2, \dots, X_n\}$ una muestra aleatoria de una población $F(\cdot)$, con esperanza μ y varianza σ^2 . La media muestral se define como

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}.$$

Se cumple lo siguiente

$$E(\bar{X}) = \mu \quad \text{y} \quad \text{Var}(\bar{X}) = \frac{\sigma^2}{n}.$$

Si se asume que la población es normal, entonces

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

De igual manera, si la población es normal y se desconoce σ^2 ,

$$\frac{\sqrt{n}(\bar{X} - \mu)}{S} \sim t_{n-1}$$

donde t_{n-1} denota la distribución t de Student con $n-1$ grados de libertad.

Resultado

Si $S = \{X_1, X_2, \dots, X_n\}$ una muestra aleatoria de una población $F(\cdot)$, con esperanza μ y varianza σ^2 , entonces cuando n es grande

$$\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \sim N(0, 1).$$

donde \sim representa que la distribución es aproximada.

- El TLC permite aproximar la distribución de la media (aritmética) de una muestra aleatoria a través de la distribución normal, sin importar cuál sea la distribución de las observaciones originales, sean continuas o discretas.
- Que tan grande debe ser n depende de la forma de la distribución, generalmente $n \geq 30$ proporciona buenos resultados.

- ① Sea $S = \{X_1, X_2, \dots, X_n\}$ una muestra aleatoria de una población $Poisson(\lambda)$, entonces

$$E(X_i) = \lambda \quad \text{y} \quad \text{Var}(X_i) = \lambda.$$

Entonces, para n grande

$$\frac{\sqrt{n}(\bar{X} - \lambda)}{\sqrt{\lambda}} \sim N(0, 1) \quad \text{o bien} \quad \bar{X} \sim N\left(\lambda, \frac{\lambda}{n}\right)$$

- ② Sea $S = \{X_1, X_2, \dots, X_n\}$ una muestra aleatoria de una población $Bin(m, p)$ entonces

$$E(X_i) = mp \quad \text{y} \quad \text{Var}(X_i) = mp(1 - p).$$

Entonces, para n grande

$$\frac{\sqrt{n}(\bar{X} - mp)}{\sqrt{mp(1 - p)}} \sim N(0, 1) \quad \text{o bien} \quad \bar{X} \sim N\left(mp, \frac{mp(1 - p)}{n}\right).$$

Cuantiles distribucionales

La función de densidad de una variable aleatoria continua nos permite calcular probabilidades relacionadas con la variable aleatoria. Por ejemplo, si X es una v.a. continua con función de densidad $f(\cdot)$, entonces

$$Pr(a \leq X \leq b) = \int_a^b f(x) dx.$$

Dada una constante $\alpha \in (0, 1)$, se define el *cuantil inferior* α como el número real q_α tal que

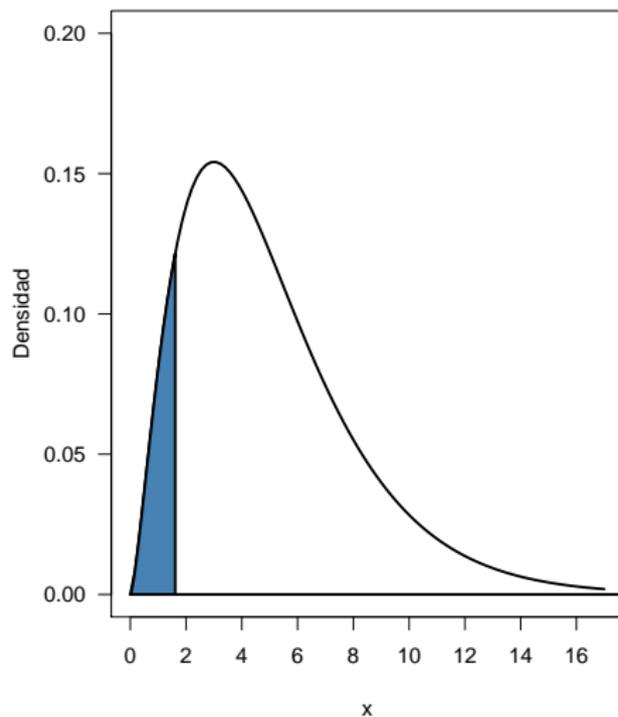
$$Pr(X \leq q_\alpha) = \alpha \quad \Leftrightarrow \quad \int_{-\infty}^{q_\alpha} f(x) dx = \alpha.$$

Se define el *cuantil superior* α como el número real $q^{(\alpha)}$ tal que

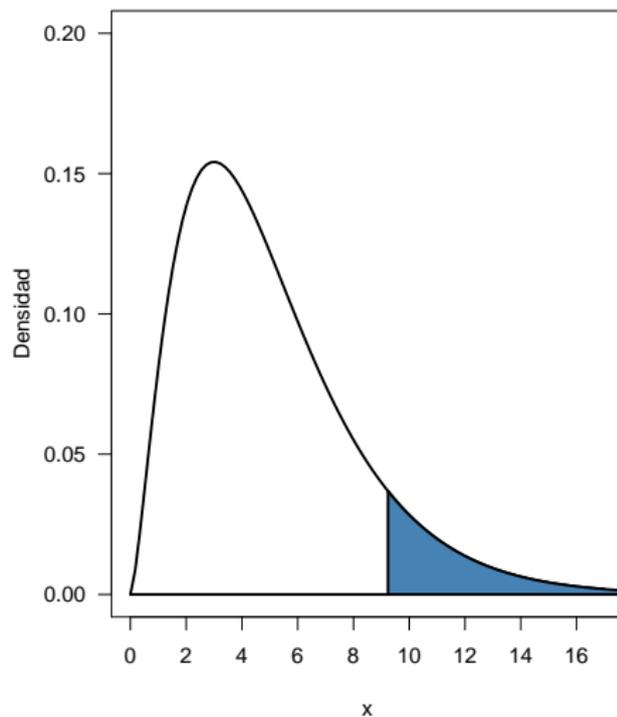
$$Pr(X > q^{(\alpha)}) = \alpha \quad \Leftrightarrow \quad \int_{q^{(\alpha)}}^{\infty} f(x) dx = \alpha.$$

Cuantiles distribucionales

Cuantil inferior 0.1



Cuantil superior 0.1



Los cuantiles distribucionales cumplen las siguientes propiedades:

- $q_\alpha = q^{(1-\alpha)}$, es decir, el cuantil inferior α es igual al cuantil superior $1 - \alpha$.
- $q^{(\alpha)} = q_{1-\alpha}$, es decir, el cuantil superior α es igual al cuantil inferior $1 - \alpha$.
- Si la distribución es simétrica alrededor de 0, entonces $q^{(\alpha)} = -q_\alpha$.

Intervalos de confianza

- La estimación por intervalos consiste en proponer un conjunto de valores posibles, contenidos dentro del espacio parametral, para un parámetro desconocido.
- La medida de error en este caso se llama *confianza* y se expresa como porcentaje. Una confianza cercana a 100 % implica menor frecuencia de cometer errores, que el parámetro no esté contenido en el conjunto de valores propuesto.
- En el caso de tener un sólo parámetro desconocido, los conjuntos toman la forma de intervalos, por lo que quedan determinados por sus extremos inferior y superior.

Definición (intervalos de confianza)

Sea muestra aleatoria $S = \{X_1, X_2, \dots, X_n\}$ de una población $F(\cdot | \theta)$ con θ desconocido. Un intervalo de confianza $100(1 - \alpha)\%$ para θ está formado por dos estadísticos $L(S)$ y $U(S)$ tales que

$$Pr \{L(S) \leq \theta \leq U(S)\} \geq 1 - \alpha.$$

$L(S)$ se conoce como límite inferior y $U(S)$ como límite superior.

Intervalo de confianza para la media de una población normal, σ^2 conocida

Sea $S = \{X_1, X_2, \dots, X_n\}$ una muestra aleatoria de una población $N(\mu, \sigma^2)$ con μ desconocida y σ^2 conocida. Se sabe que

$$\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \sim N(0, 1).$$

De lo anterior se sigue que

$$Pr \left\{ -z^{(\alpha/2)} \leq \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \leq z^{(\alpha/2)} \right\} = 1 - \alpha.$$

donde $z^{(\alpha/2)}$ es el cuantil superior $\alpha/2$ de una distribución $N(0, 1)$, $\alpha \in (0, 1)$. De lo anterior se puede despejar

$$Pr \left\{ \bar{X} - z^{(\alpha/2)}\sigma/\sqrt{n} \leq \mu \leq \bar{X} + z^{(\alpha/2)}\sigma/\sqrt{n} \right\} = 1 - \alpha.$$

Intervalo de confianza para la media de una población normal, σ^2 conocida

Entonces, un intervalo de confianza $100(1 - \alpha)\%$ para μ tiene límites

$$L(S) = \bar{X} - z^{(\alpha/2)}\sigma/\sqrt{n}$$

$$U(S) = \bar{X} + z^{(\alpha/2)}\sigma/\sqrt{n}.$$

Intervalo de confianza para la media de una población normal, σ^2 desconocida

Sea $S = \{X_1, X_2, \dots, X_n\}$ una muestra aleatoria de una población $N(\mu, \sigma^2)$ con μ y σ^2 desconocidas. El siguiente resultado es similar al anterior excepto que se reemplaza σ^2 por S^2 y la distribución normal por la distribución t . Se sabe que

$$\frac{\sqrt{n}(\bar{X} - \mu)}{S} \sim t_{n-1}.$$

De lo anterior se sigue que

$$Pr \left\{ -t_{n-1}^{(\alpha/2)} \leq \frac{\sqrt{n}(\bar{X} - \mu)}{S} \leq t_{n-1}^{(\alpha/2)} \right\} = 1 - \alpha.$$

donde $t_{n-1}^{(\alpha/2)}$ es el cuantil superior $\alpha/2$ de una distribución t_{n-1} , $\alpha \in (0, 1)$. De lo anterior se puede despejar

$$Pr \left\{ \bar{X} - t_{n-1}^{(\alpha/2)} S / \sqrt{n} \leq \mu \leq \bar{X} + t_{n-1}^{(\alpha/2)} S / \sqrt{n} \right\} = 1 - \alpha.$$

Intervalo de confianza para la media de una población normal, σ^2 desconocida

Entonces, un intervalo de confianza $100(1 - \alpha)\%$ para μ tiene límites

$$L(S) = \bar{X} - t_{n-1}^{(\alpha/2)} S / \sqrt{n}$$

$$U(S) = \bar{X} + t_{n-1}^{(\alpha/2)} S / \sqrt{n}.$$

- 1 Una muestra de 100 empleados se seleccionó, se registraron sus salarios y se encontró que $\bar{Y} = 7750$ y se sabe de estudios similares que $\sigma = 900$. Encontrar un intervalo de confianza 95 % para el salario promedio μ .
- 2 Los rendimientos para 9 botes de pintura (la misma presentación) en m^2 son: 5.52, 3.75, 4.31, 3.27, 5.99, 4.76, 3.87, 4.45, 4.70. Se requiere construir un intervalo de confianza del 90% el rendimiento promedio de la pintura. De la muestra se obtiene que $\bar{X} = 4.51$ y $S^2 = 0.73$.

Tamaño de muestra para estimar la media de una población normal

- La longitud de los intervalos de confianza es importante para hacer comparaciones y una forma de controlarla es a través del tamaño de muestra.
- El intervalo de confianza para la media de una población $N(\mu, \sigma^2)$, con σ^2 conocida, está dado por

$$L(S) = \bar{X} - z^{(\alpha/2)}\sigma/\sqrt{n}$$
$$U(S) = \bar{X} + z^{(\alpha/2)}\sigma/\sqrt{n}.$$

y su longitud es

$$U(S) - L(S) = \bar{X} + z^{(\alpha/2)}\sigma/\sqrt{n} - \left(\bar{X} - z^{(\alpha/2)}\sigma/\sqrt{n} \right)$$
$$= 2z^{(\alpha/2)}\sigma/\sqrt{n}.$$

- El requisito usual es pedir que el error de estimación de μ sea menor a cierta cantidad $d > 0$. Esto es equivalente a pedir un intervalo de confianza con una longitud menor a $2d$.

Tamaño de muestra para estimar la media de una población normal

- Se puede despejar el tamaño de muestra n necesario para cubrir con tal requisito.

$$2d \geq \frac{2z^{(\alpha/2)}\sigma}{\sqrt{n}}$$
$$\sqrt{n} \geq \frac{z^{(\alpha/2)}\sigma}{d}$$
$$n \geq \left(\frac{z^{(\alpha/2)}\sigma}{d} \right)^2$$

- Por lo tanto, una vez que se fija el nivel de confianza $100(1 - \alpha)\%$, el error de estimación d y se conoce la varianza de la población σ^2 (o se tienen una buena estimación), se puede calcular el tamaño de muestra requerido n .

Ejemplos. Tamaño de muestra

- 1 Se tiene interés en realizar un muestreo para estimar el peso promedio (en kg) de los hombres mayores de 18 años en México. Si se asume que el peso de esta población se puede modelar con una distribución $N(\mu, \sigma^2)$ con σ^2 conocida igual a 36. Calcular el tamaño de muestra necesario para estimar la media μ con un error $d = 0.5$ kg y confianza 95 %.
- 2 Un psicólogo desea estimar el tiempo medio de reacción a un determinado estímulo, con un error $d = 0.01$ y confianza 95 %. Él supone que la desviación estándar de las mediciones es 0.05 segundos. Determinar el tamaño de muestra necesario para cubrir los requisitos de estimación.

Intervalo de confianza para la diferencia de medias, varianzas conocidas

- En algunas ocasiones el interés es identificar diferencias en la media de dos poblaciones. Por ejemplo: comparar dos esquemas de producción a través de la vida útil de un lote de productos; comparar dos métodos de enseñanza a través del desempeño de grupos de estudiantes, entre otros casos.
- En estos casos se asume que hay dos poblaciones $N(\mu_1, \sigma_1^2)$ y $N(\mu_2, \sigma_2^2)$, con μ_1 y μ_2 desconocidas y σ_1^2 y σ_2^2 conocidas. El objetivo es hacer inferencias sobre $\mu_1 - \mu_2$ a partir de una muestra aleatoria $\{X_1, \dots, X_n\}$ de la primera población y una muestra aleatoria $\{Y_1, \dots, Y_m\}$ de la segunda población; se asume además que las muestras son independientes.
- Por propiedades de la distribución normal

$$\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}\right).$$

Intervalo de confianza para la diferencia de medias, varianzas conocidas

- De lo anterior se sigue que

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}} \sim N(0, 1).$$

- Luego, para $\alpha \in (0, 1)$ se cumple que

$$Pr \left\{ -z^{(\alpha/2)} \leq \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}} \leq z^{(\alpha/2)} \right\} = 1 - \alpha.$$

- De lo anterior se puede despejar un intervalo de confianza $100(1 - \alpha)\%$ para $\mu_1 - \mu_2$:

$$L(S) = \bar{X} - \bar{Y} - z^{(\alpha/2)} \left(\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m} \right) \quad y \quad U(S) = \bar{X} - \bar{Y} + z^{(\alpha/2)} \left(\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m} \right).$$

Intervalo de confianza para la diferencia de medias, varianza común desconocida

- En la mayoría de las aplicaciones se desconoce la varianza de las poblaciones, por lo que debe ser estimada. En el caso más sencillo se asume que ambas poblaciones tienen la misma varianza desconocida σ^2 .
- Bajo este supuesto, el estimador de σ^2 es la varianza combinada

$$S_p^2 = \frac{(n-1)S_x^2 + (m-1)S_y^2}{n+m-2}$$

donde $S_x^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ y $S_y^2 = \frac{1}{m-1} \sum_{i=1}^m (Y_i - \bar{Y})^2$.

Resultado

Si $\{X_1, \dots, X_n\} \sim N(\mu_1, \sigma^2)$, $\{Y_1, \dots, Y_m\} \sim N(\mu_2, \sigma^2)$ y las muestras son independientes, entonces

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{S_p \sqrt{\left(\frac{1}{n} + \frac{1}{m}\right)}} \sim t_{n+m-2}.$$

Intervalo de confianza para la diferencia de medias, varianza común desconocida

A partir del resultado anterior y siguiendo un procedimiento análogo a los anteriores se obtiene un intervalo de confianza $100(1 - \alpha) \%$ para $\mu_1 - \mu_2$:

$$L(S) = \bar{X} - \bar{Y} - t_{n+m-2}^{(\alpha/2)} S_p \sqrt{\left(\frac{1}{n} + \frac{1}{m}\right)}$$

$$U(S) = \bar{X} - \bar{Y} + t_{n+m-2}^{(\alpha/2)} S_p \sqrt{\left(\frac{1}{n} + \frac{1}{m}\right)}$$

- ① Se hizo una comparación sobre las cualidades de duración de dos tipos de llantas, se rodaron $n_1 = n_2 = 100$ llantas de cada tipo. Se registró el número de kilómetros hasta que se notara un desgaste predeterminado. Los resultados fueron los siguientes:

$$\bar{X} = 26400, \quad \bar{Y} = 25100, \quad S_x^2 = 1440000, \quad S_y^2 = 1960000.$$

Calcular un intervalo confianza 90% para la diferencia de medias:

- ① asumiendo a S_x^2 y S_y^2 como las verdaderas varianzas poblacionales.
 - ② asumiendo que la varianza de las poblaciones es común pero desconocida.
- ② Se entrenaron dos grupos de nueve empleados nuevos, uno usando el método tradicional y el otro con un método nuevo. Se midió en minutos el tiempo que tardó cada empleado en montar el dispositivo de interés. Los resultados fueron:

$$\bar{X} = 35.22, \quad \bar{Y} = 31.56, \quad S_x^2 = 24.44, \quad S_y^2 = 20.03.$$

Calcular un intervalo de confianza 95% para la diferencia de medias:

- ① asumiendo a $\sigma_1^2 = \sigma_2^2 = 23$.
- ② asumiendo que la varianza de las poblaciones es común pero desconocida.

Interpretación de los intervalos de confianza

- Supongamos que $(2.5, 7.3)$ es un intervalo de confianza 95 % para la media de una población $N(\mu, 1)$, ¿qué significa realmente este resultado? ¿Qué es la confianza?
- De la definición de intervalo de confianza y por la interpretación frecuentista de la probabilidad sabemos lo siguiente: si pudieramos observar un número grande de muestras, bajo las mismas condiciones, y para cada realización calculamos el intervalo de confianza 95 %, entonces esperamos que 95 % de los intervalos calculados contengan al verdadero μ .
- En nuestro ejemplo, $(2.5, 7.3)$ es sólo una realización, por lo que contiene o no a μ , no lo sabemos porque μ es desconocido, sin embargo *confiamos* en que sea uno de los que sí lo contiene.

Intervalos de confianza y pruebas de hipótesis

- Los intervalos de confianza son conjuntos de valores posibles, contruidos con información de una muestra aleatoria.
- Cualquier valor contenido en un intervalo de confianza es admisible para el parámetro desconocido.
- Este hecho se puede utilizar para *contrastar hipótesis* sobre el verdadero valor de un parámetro desconocido.
 - En el caso de un intervalo de confianza $100(1 - \alpha)\%$ para la media μ de una población normal, es de interés contrastar la hipótesis $H_0 : \mu = \mu_0$, donde μ_0 puede ser cualquier constante conocida. El procedimiento consiste en rechazar H_0 si el intervalo no contiene a μ_0 , de lo contrario la hipótesis no se rechaza.
 - Para comparación de medias la hipótesis más común es $H_0 : \mu_1 - \mu_2 = 0$, que es equivalente a $H_0 : \mu_1 = \mu_2$. Si el intervalo de confianza para $\mu_1 - \mu_2$ no contiene al 0, entonces se rechaza la hipótesis de que las medidas de las poblaciones sean iguales. Si el intervalo contiene sólo valores positivos, se puede concluir que *estadísticamente* $\mu_1 > \mu_2$; la conclusión es análoga si sólo contiene valores negativos.
- En el caso de contrastes de hipótesis se habla de *significancia* y es complementaria a la confianza. La significancia es la frecuencia con la que se rechaza una hipótesis nula verdadera.

Consideremos que se puede realizar uno de los siguientes experimentos:

- 1 Se lanzan cinco monedas y se registra el número de águilas que ocurren.
- 2 Se lanza un dado y se registra el resultado.

Sólo se nos informará del resultado del experimento X pero no cuál se realizó y se nos pide diseñar un conjunto de reglas de decisión para inferir qué experimento se realizó. Por ejemplo, una de tales reglas puede ser: *si $X \leq 3$ entonces concluimos que se lanzó la moneda, en caso contrario concluimos que se lanzó el dado*. Notemos que los posibles valores de X son $0, 1, 2, \dots, 6$.

Podemos plantear este problema como uno de contraste de hipótesis estadísticas. Consideremos, por ejemplo que nuestro interés es contrastar si se lanzó el dado y lo establecemos como hipótesis nula. En la hipótesis alternativa ponemos al lanzamiento de la moneda.

$$H_0 : \text{Se lanzo el dado.} \quad \text{vs.} \quad H_1 : \text{Se lanzó la moneda}$$

Una prueba de hipótesis consiste en una regla decisión que establece en qué casos rechazamos H_0 en favor de H_1 , o de otro modo no rechazamos H_0 . Por ejemplo, la regla anterior establece rechazar H_0 si $X \leq 3$, de otro modo no se rechaza H_0 .

Tomemos como ejemplo la regla de decisión anterior. Existe la posibilidad de haber lanzado el dado y obtener un resultado menor o igual a 3, en cuyo caso estaríamos rechazando la hipótesis nula siendo verdadera. Otra posibilidad es haber lanzado la moneda y obtener un resultado mayor a 3, en tal caso estaríamos *aceptando* la hipótesis nula siendo falsa. En estadística estos errores se conocen como *Error tipo I* y *Error tipo II*.

A la probabilidad de cometer *Error tipo I* se le denota como α y se le llama *tamaño de la prueba*. A la probabilidad de cometer *Error tipo II* denota como β y no recibe un nombre, sin embargo, su complemento $1 - \beta$ que representa la probabilidad de rechazar H_0 cuando es falsa, se le conoce como *potencia de la prueba*. Lo ideal es utilizar pruebas o reglas de decisión con el menor tamaño y la mayor potencia posibles.

En nuestro ejemplo, con la regla rechazar H_0 si $X \leq 3$ tenemos lo siguiente:

- La probabilidad de cometer *Error tipo I* es la probabilidad de obtener 1, 2 o 3 en el lanzamiento de un dado. Entonces el tamaño de la prueba es $\alpha = 0.5$.
- La probabilidad de cometer *Error tipo II* es la probabilidad de obtener 4 o 5 águilas en el lanzamiento de 5 monedas. Entonces $\beta = 0.19$. Luego la potencia de la prueba es $1 - 0.19 = 0.81$.

Estos resultados nos indican que: si realmente se lanzó el dado (H_0 verdadera), la probabilidad que tenemos de concluir mal es $\alpha = 0.5$ (tamaño de la prueba) y si realmente se lanzó la moneda (H_0 falsa), la probabilidad que tenemos de concluir bien es $1 - \beta = 0.81$ (potencia de la prueba).

Pruebas de hipótesis

En la siguiente tabla se muestran los valores de α y β para reglas de la forma rechazar H_0 si $X \leq c$ con $c = 0, 1, 2, \dots, 6$.

c	0	1	2	3	4	5	6
α	0.00	0.17	0.33	0.50	0.67	0.83	1.00
β	0.97	0.85	0.50	0.19	0.03	0.00	0.00

Podemos observar como al disminuir la probabilidad del *error tipo I* se incrementa la probabilidad de *error tipo II* y viceversa. Visto de otra forma, al disminuir el tamaño de la prueba disminuye también la potencia.