

Estadística 2

Andrés Martínez

Departamento de Matemáticas
Universidad Externado

4 de abril de 2021

Introducción

Introducción

Variable Aleatoria y Distribuciones

Distribución normal estándar

Muestreo

Distribuciones Muestrales

Estimación

Métodos de Estimación

Estimación por Intervalos de Confianza

Pruebas de Hipótesis

Análisis de Datos Categóricos

ANOVA

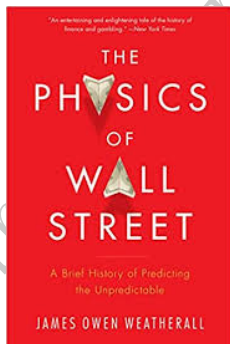
Regresión Lineal

Referencias

- ▶ Contenido
- ▶ Evaluación
- ▶ Horario de Atención
- ▶ Proyecto Final

Andrés Martínez

■ Puedo predecir con exactitud los movimientos planetarios, pero no puedo predecir la locura de las masas ■ Isaac Newton



Definición: Siendo Ω un conjunto no vacío con una colección de subconjuntos \mathcal{F} . Se dice que \mathcal{F} es una σ álgebra con las siguientes propiedades[Llinás Solano, 2015]:

- ▶ El conjunto vacío \emptyset pertenece a \mathcal{F}
- ▶ Como sea que un conjunto \mathcal{A} pertenece a \mathcal{F} , \mathcal{A}^c también pertenece a \mathcal{F} .
- ▶ Como sea que unos conjuntos $\mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_n$ pertenecen a \mathcal{F} , la unión \cup de $\mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_n$ también pertenece a \mathcal{F} .

Propiedades: Siendo Ω un conjunto no vacío con una colección de subconjuntos \mathcal{F} . Una medida de probabilidad \mathbb{P} es una función que para cada conjunto de $\mathcal{A} \in \mathcal{F}$ asigna un número entre $[0, 1]$ llamado probabilidad de \mathcal{A} y escrito como $\mathbb{P}(\mathcal{A})$ [Llinás Solano, 2015]:

- ▶ $\mathbb{P}(\Omega) = 1$
- ▶ Como sea que $\mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_n$ es una secuencia de conjuntos disyuntos en \mathcal{F}



$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} \mathcal{A}_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(\mathcal{A}_n)$$

Por lo tanto, cualquier variable aleatoria dentro de un proyecto financiero se mide a partir del espacio de probabilidad triple $(\Omega, \mathcal{F}, \mathbb{P})$. [Llinás Solano, 2015]

Teorema: Sea $(\Omega, \mathcal{F}, \mathbb{P})$ un espacio de probabilidad. Para A, B $(A_n)_{n \in \mathbb{N}}$ y $(B_n)_{n \in \mathbb{N}} \in \mathcal{F}$

- ▶ $\mathbb{P}(\emptyset) = 0$
- ▶ $P(A^c) = 1 - P(A)$
- ▶ Si $A \subseteq B$ entonces $P(A) \leq P(B)$
- ▶ $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

σ algebra: La dupla (Ω, \mathcal{F}) se llama espacio medible si $\Omega \neq \emptyset$ y \mathcal{F} una σ álgebra, es decir, un sistema de subconjuntos de un conjunto que posee las siguientes propiedades:[Llinás Solano, 2015]

- a $\Omega \in \mathcal{F}$
- b Si $A \in \mathcal{F}$, entonces $\bar{A} := \Omega \setminus A \in \mathcal{F}$.
- c Si $A_1, A_2, \dots \in \mathcal{F}$ entonces $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$

Espacio Borel: Siendo X una variable aleatoria en un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$. La medida de distribución de X es la medida de probabilidad μ_X que asigna a cada subconjunto Borel \mathcal{B} de \mathbb{R} la medida $\mu_X(B) = \mathbb{P}\{X \in B\}$ [Llinás Solano, 2015].

Definición: La menor σ álgebra sobre \mathbb{R} que contiene todos los intervalos de la forma $(-\infty, a]$ con $a \in \mathbb{R}$ se llama σ álgebra de Borel y se denota por \mathcal{B} . Los elementos de \mathcal{B} se llaman conjuntos de Borel. [Llinás Solano, 2015]

Experimento de Laplace

El experimento de laplace está compuesto por la probabilidad del conjunto A que es igual a los elementos de A dividido en los elementos de Ω . [Llinás Solano, 2015]

$$P(A) = \frac{|A|}{|\Omega|}$$

Probabilidad condicionada [Llinás Solano, 2015] Para $P(A) \neq 0$ existe una probabilidad condicionada de dos subconjuntos en Ω por lo tanto la probabilidad condicionada de B con respecto a A es igual:

$$P_A(B) \equiv P(B | A) = \frac{P(B \cap A)}{P(A)} \text{ para } P(A) > 0$$

Independencia de Resultados

Dos conjuntos A y B son independientes siempre que:

$$P(A \cap B) = P(A)P(B)$$

Probabilidad Total

[Llinás Solano, 2015] Si no hay independencia en los resultados se dice que el evento la probabilidad del evento B $P(B)$ es la suma de todos los eventos que suceden en A_1, \dots, A_n de esta forma:

$$P(B) = \sum_i^n P(A_i)P(B | A_i)$$

Teorema de Bayes

Los métodos bayesianos son herramientas de análisis de datos que se derivan de los principios de inferencia bayesiana [Llinás Solano, 2015].

- ▶ Estimaciones de parámetros con buenas propiedades estadísticas
- ▶ Descripciones parsimoniosas de los datos observados
- ▶ Predicciones para datos faltantes y pronósticos de datos futuros
- ▶ Un marco computacional para la estimación, selección y validación de modelos

$$P(A_i | B) = \frac{P(A_i)P(B | A_i)}{\sum_i^n P(A_i)P(B | A_i)}$$

Definición Sean (Ω, P) un espacio de probabilidad finito. Una variable es una función real definida en Ω y Ω dos espacios medibles [Llinás Solano, 2015].

- a Una función $f : \Omega \rightarrow \Omega'$ se llama \mathcal{F} medible.
- b Si, además, está definida una probabilidad P sobre (Ω, \mathcal{F}) entonces X se mide en un espacio de probabilidad (Ω, \mathcal{F}, P) .

Una función de densidad $f(x)$ de una variable aleatoria X tiene las siguientes propiedades:

a $f(x) \geq 0$

b $\int_{-\infty}^{\infty} f(x) dx = 1$

c $P(a \leq X \leq b) = \int_a^b f(x) dx$

[Llinás Solano, 2015] Si se quiere obtener la distribución acumulada de una variable aleatoria X discreta o continua, se puede usar la función de distribución acumulada que tiene las siguientes propiedades:

- a $0 \leq F(x) \leq 1$
- b $F(x_1) \leq F(x_2)$
- c $\lim_{t \rightarrow -\infty} F(x) = 0$
- d $\lim_{t \rightarrow \infty} F(x) = 1$
- e $F(x)$ es continua por la derecha

Definición Sean (Ω, \mathcal{F}, P) un espacio de probabilidad y X una variable aleatoria numérica [Llinás Solano, 2015].

- a Si X es una variable aleatoria discreta con valores x_1, x_2, \dots se dice que la esperanza de X existe si
- $$\sum_{k=1}^{\infty} |x_k| = P(\{X = x_k\}) < \infty.$$

El valor esperado de una variable aleatoria se obtiene sumando todos sus resultados y multiplicándolo por la probabilidad

$$\mathbb{E}(X) = \sum_x x \mathbb{P}(x) \quad (1)$$

- b Si X es una variable aleatoria continua con una función de densidad $f(x)$, se dice que la esperanza de X existe si $\int_{-\infty}^{\infty} |x|f(x)dx < \infty$

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} xf(x)dx \quad (2)$$

- c Si $E[X] < \infty$, entonces $Var[X] := E([X - E(X)]^2)$ y $\sigma = \sqrt{Var(X)}$

Propiedades del Valor Esperado: Sean (Ω, \mathcal{F}, P) un espacio de probabilidad y X una variable aleatoria real [Llinás Solano, 2015].

- ▶ $E(a) = a$ para todo $a \in \mathbb{R}$.
- ▶ si $X_1 \leq X_2$, entonces $E(X_1) \leq E(X_2)$
- ▶ $E(aX_1 + bX_2) = aE(X_1) + bE(X_2)$ para todo $a, b \in \mathbb{R}$
- ▶ $|E(X)| \leq E(|X|)$

Propiedades de la Varianza: Sean (Ω, \mathcal{F}, P) un espacio de probabilidad y X una variable aleatoria real.

- ▶ $Var(X) = E(X^2) - (E(X))^2$
- ▶ $Var(aX + b) = a^2 Var(X)$ para $a, b \in \mathbb{R}$

$$STD[X] = \sqrt{V[X]} \equiv \sigma_X$$

Distribución Binomial Es una distribución para una variable aleatoria discreta, la probabilidad se obtiene con

$$P(X = x) = \binom{n}{x} p^x q^{n-x} \quad (3)$$

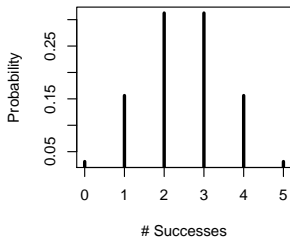
Donde p y q son probabilidades n el número de escenarios y x el valor que se desea obtener.

El valor esperado y la varianza son:

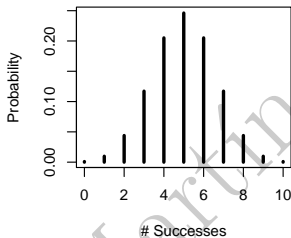
$$\mathbf{E}(X) = np$$

$$\mathbf{V}(X) = np(1 - p)$$

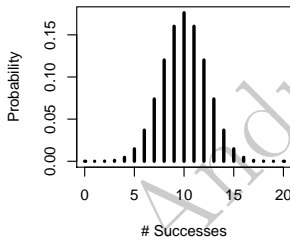
Binomial Distribution ($n=5$, $p=0.5$)



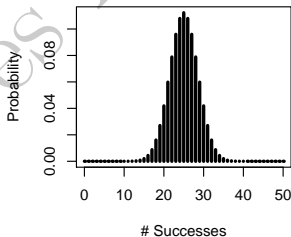
Binomial Distribution ($n=10$, $p=0.5$)



Binomial Distribution ($n=20$, $p=0.5$)



Binomial Distribution ($n=50$, $p=0.5$)



Su función de probabilidad es:

$$P(x) = \frac{\lambda^x}{x!} e^{-\lambda}$$

Mientras que la función de densidad se define como:

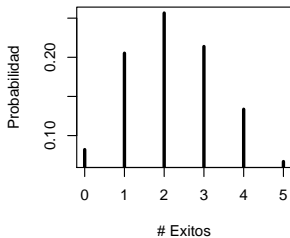
$$F(x) = \sum_{x=0}^n P(x) = e^{-\lambda} \sum_{x=0}^n \frac{\lambda^x}{x!}$$

Con un valor esperado y una varianza de:

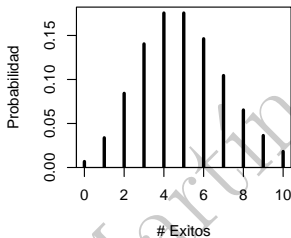
$$\mathbf{E}(X) = \lambda$$

$$\mathbf{V}(X) = \lambda$$

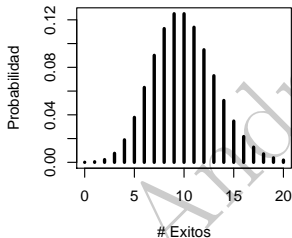
Distribución Poisson lambda = 2.5



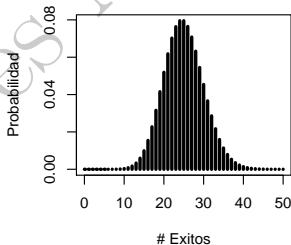
Distribución Poisson lambda = 5



Distribución Poisson lambda = 10



Distribución Poisson lambda = 25



La distribución mas sencilla de la teoría de probabilidad, se encuentra entre dos intervalos $[a, b]$ y su distribución igual se denota de la forma $X \sim \mathbb{U}(a, b)$ lo que quiere decir que esta función se distribuye de forma uniforme con parámetros a y b que son constantes en el espacio cartesiano.

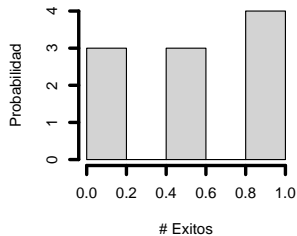
$$f(x) = \begin{cases} \frac{1}{b-a} & \text{para } a \leq x \leq b \\ 0 & \text{de otra forma} \end{cases}$$

El valor esperado y la varianza son:

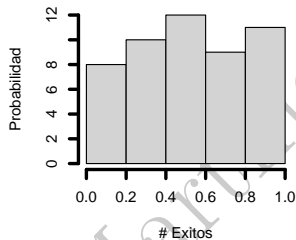
$$\mathbf{E}(X) = \frac{b + a}{2}$$

$$\mathbf{V}(X) = \frac{(b - a)^2}{12}$$

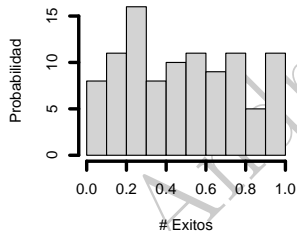
Distribución Uniforme $n=10$



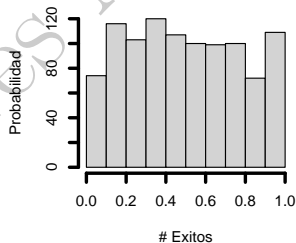
Distribución Uniforme $n=50$



Distribución Uniforme $n=100$



Distribución Uniforme $n=1000$



Es la más importante de las distribuciones continuas. Su característica principal es la simetría, ya que la media, la mediana y la moda tienen el mismo valor; su sesgo es cero y su curtosis es 3, lo que significa que es poco probable que presente escenarios extremos.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad (\mu \in (-\infty, \infty), \sigma > 0)$$

$$X \sim N(\mu; \sigma^2)$$

Función de distribución acumulada

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) du$$

$$\mu = 0, \sigma = 1: X \sim N(0; 1).$$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x)^2}{2}\right) (\mu \in (-\infty, \infty), \sigma > 0)$$

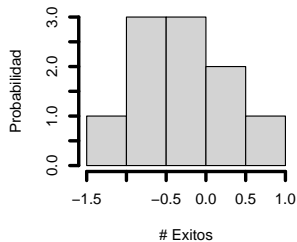
$$X \sim Z(0; 1)$$

Función de distribución acumulada

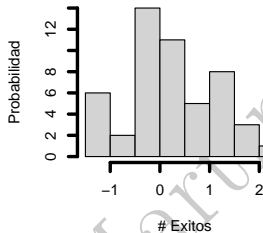
$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{(x)^2}{2}\right) du$$

$$Z = \frac{x - \mu}{\sigma}$$

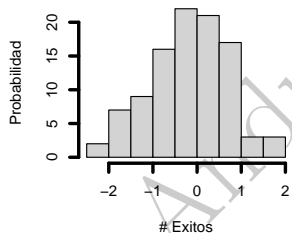
Distribución Normal $n=10$



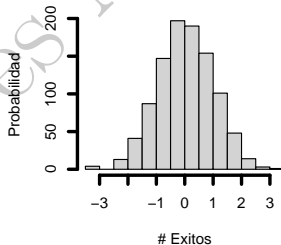
Distribución Normal $n=50$



Distribución Normal $n=100$



Distribución Normal $n=1000$



Los precios de los instrumentos financieros tendrán este tipo de distribución, y esta es una de las razones por las cuales se trabaja mejor con rendimientos dado que este tipo de medición es más complejo. El resultado de la variable siempre es positivo como en los precios de los activos $x > 0$ y la variable aleatoria se distribuye $X \sim LN(0; 1)$.

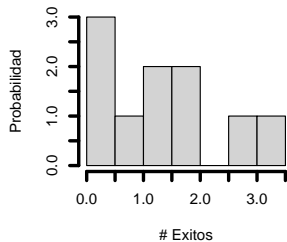
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \frac{1}{x} \exp\left(-\frac{(\ln(x) - \mu)^2}{2\sigma^2}\right)$$

Para toda variable aleatoria que se distribuye de forma normal, $X \sim LN(\mu; \sigma^2) \leftrightarrow \exp(Y) = X \sim LN(\mu; \sigma^2)$.

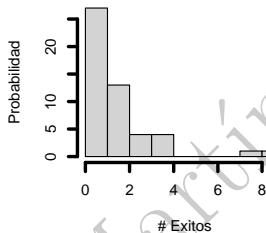
$$\mathbf{E}(X) = e^{\mu + \frac{\sigma^2}{2}},$$

$$\mathbf{V}(X) = e^{2\mu + \sigma^2} (e^{\sigma^2} - 1),$$

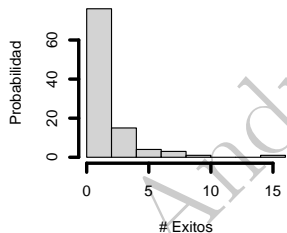
Distribución Log Normal n=10



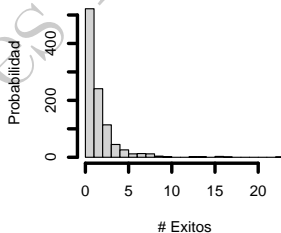
Distribución Log Normal n=50



Distribución Log Normal n=100



Distribución Log Normal n=1000



Este curso parte del supuesto de normalidad para medir y analizar el riesgo financiero, desde la estadística descriptiva se puede evaluar de la siguiente forma

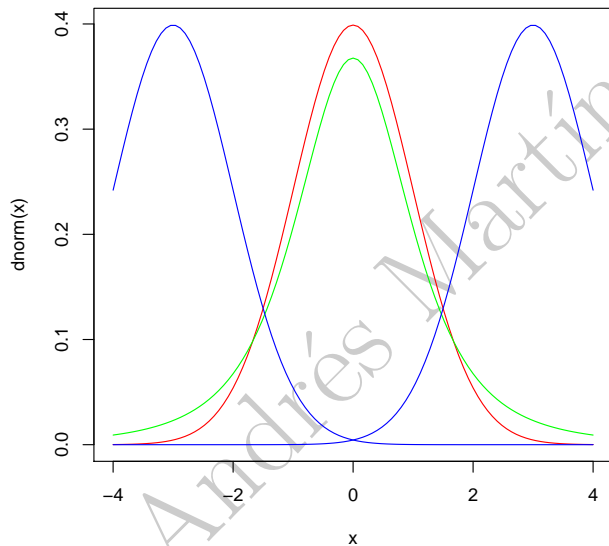
$$Sesgo = \frac{\sum_{i=1}^n (x_i - \mu)^3}{(n-1)\sigma^3} \quad (4)$$

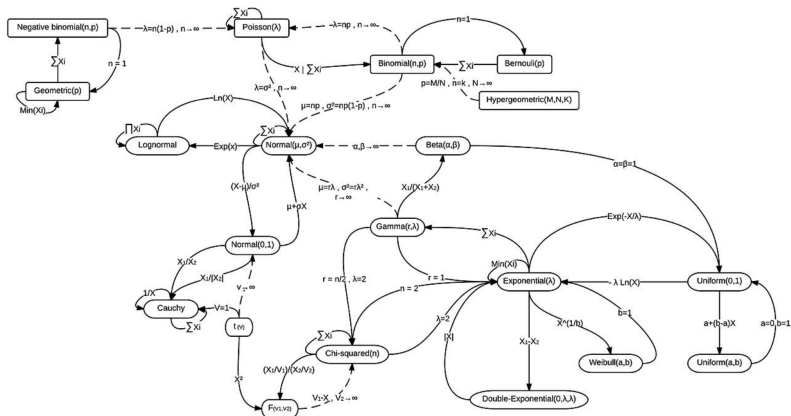
$$Kurtosis = \frac{\sum_{i=1}^n (x_i - \mu)^4}{(n-1)\sigma^4} \quad (5)$$

El estadístico se obtiene con ayuda de la ecuación 4 y 5

$$LM = N \left[\frac{Sesgo^2}{6} + \frac{(Kurtosis - 3)^2}{24} \right] \quad (6)$$

Forma de la distribución





Métodos de muestreo

La distribución de probabilidad asociada con la población (de la cual muestra) se llama distribución de población y distribución de probabilidad asociado con nuestra estadística se llama distribución muestral; claramente, los dos están interrelacionados. Para conocer la distribución de la población, es imperativo saber todo lo que podamos sobre la distribución muestral. Por lo tanto, nos enfocamos en lo que sucede en el caso especial del muestreo de la distribución normal, y en particular, cumplimos con la distribución muestral de \bar{x} y s^2 . [Wackerly et al., 2010]

Usualmente cuando se lleva a cabo una investigación de forma cuantitativa, es necesario contar con variables que permitan generar las pruebas de contraste de hipótesis. Estas variables provienen de observaciones que se obtienen de una muestra dado que es muy difícil trabajar con una población. [Peña, 2014]

- ▶ Población - Parámetros
- ▶ Muestra - Estadístico
- ▶ error de muestra: Diferencia entre el estadístico y el parámetro

Muestreo de acuerdo a:

- ▶ Objeto de estudio
- ▶ Método utilizado (Paramétricos - No paramétricos)
- ▶ Información considerada (Clásico - Bayesiano)

Procedimientos de Inferencia [Peña, 2014]

Objetivo	¿Cómo se distribuye y ?	¿Influye x en y ?
Datos	Muestreo (y_1, \dots, y_n)	Diseño $x_1 : (y_{1,1}, \dots, y_{1,n})$
Método	Paramétrico	No paramétrico
Información	Clásico - Bayesiano	
Resultados	$\hat{\theta} \in (a, b) - \theta \mid y$	Distribución empírica

- ▶ Muestreo aleatorio simple:
 - ▶ Cada elemento de la población tiene la misma probabilidad de ser elegido.
 - ▶ Las observaciones se realizan con reemplazamiento, de esta manera la población es idéntica en todas las observaciones.
- ▶ Muestreo estratificado: Los elementos de la población se dividen en clases o estratos, pero de forma interna la información es homogénea.
- ▶ Muestreo por Conglomerados: En los conglomerados no hay diferencias entre los grupos, pero la información interna de cada grupo es heterogénea.
- ▶ Muestreo sistemático: Variación del aleatorio simple. N/n con k como su entero más próximo.

Teorema 1 Para cualquier población, el valor promedio de todas las posibles medias muestrales calculado a partir de todas las posibles muestras aleatorias de un tamaño dado de la población será igual a la media de la población, $\mu_{\bar{x}} = \mu$

Estimador insesgado: característica de ciertas estadísticas en las que promedio de todos los valores posibles de la estadística de muestra es igual al parámetro

Teorema 2 Para cualquier población, la desviación estándar de las posibles medias muestrales calculado a partir de todas las muestras aleatorias posibles de tamaño n es igual al desviación estándar de la población dividida por la raíz cuadrada del tamaño de la muestra. También llamado error estándar,

$$\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}}$$

Teorema 3 Si una población tiene una distribución normal, con media μ y una desviación estándar σ , la distribución muestral de la media muestral \bar{x} también es normalmente distribuida con una media igual a la media de la población ($\mu_{\bar{x}} = \mu$) y una desviación estándar muestral igual a la desviación estándar de la población dividida por la raíz cuadrada del tamaño de la muestra $\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}}$

Ley de los grandes números

La ley de los grandes números es una propiedad que explica como el incremento del número de experimentos permite que el valor esperado de las simulaciones, se acerque al valor esperado real. [Peña, 2014]

Definición: Sea X_n para $n \in N$ una sucesión de variables aleatorias reales e integrables sobre un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$. Se dice que esta sucesión satisface la ley débil de los grandes números si

$$\frac{1}{n} \sum_{j=1}^n [X_j - E(X_j)] \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0 \quad (7)$$

Teorema del Límite Central: Si X_1, X_2, \dots , son variables aleatorias reales, independientes e idénticamente distribuidas iid con $E(X_n^2) < \infty$. Sean $\mu := E(X_n)$ y $\sigma^2 := \text{Var}(X_n)$. Si $\sigma > 0$ entonces [Llinás Solano, 2016]

$$Z_n := \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow[n \rightarrow \infty]{d} N(0, 1) \quad (8)$$

Sin embargo, hay muchas poblaciones que no son normales. Estadístico a menudo se encuentran tomando muestras de esas poblaciones. ¿Qué se puede decir en este caso? La sorprendente respuesta está contenida en el Teorema del límite central

- ▶ El tamaño de la muestra debe ser "suficientemente grande"
- ▶ Si la población es simétrica, entonces tamaños de muestra tan pequeños como 2 o 3 puede proporcionar una distribución de muestreo normalmente distribuida
- ▶ Si la población está muy sesgada o tiene una forma irregular, el tamaño de muestra requerido será mayor
- ▶ Una definición conservadora de un tamaño de muestra suficientemente grande es $n \geq 30$

La distribución muestral de la media se puede obtener de dos formas. La primera cuando conocemos la varianza σ^2 en donde podemos decir que:

$$\bar{X} \xrightarrow{d} N\left(\mu, \frac{\sigma^2}{n}\right) = \frac{\bar{X}_n - \mu}{\sigma_n / \sqrt{n}} \quad (9)$$

Observe que la varianza depende del tamaño de la muestra n . En caso de que se desconosca la varianza σ^2 , la distribución t de student ofrece una solución para medir la probabilidad de la media muestral \bar{X}

$$\frac{\bar{X}_n - \mu}{S_n / \sqrt{n}} \xrightarrow{d} t(n - 1) \quad (10)$$

La distribución muestral de la media se puede obtener de dos formas. La primera cuando conocemos la varianza σ^2 en donde podemos decir que:

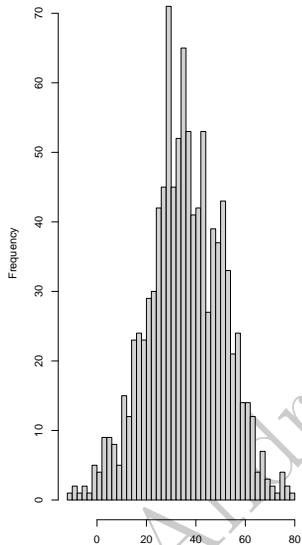
$$z = \frac{\bar{X}_n - \mu}{\sigma / \sqrt{n}} \quad (11)$$

En caso de que se desconosca la varianza σ^2 , la distribución t de student ofrece una solución para medir la probabilidad de la media muestral \bar{X}

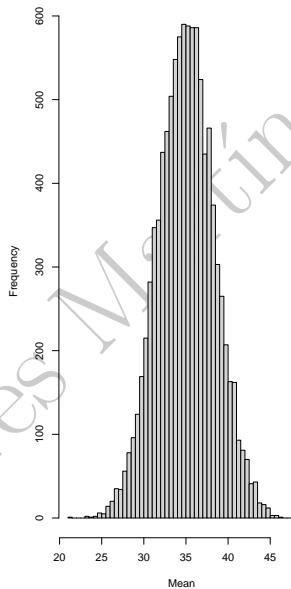
$$z = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{n-1}}} \quad (12)$$

Donde $\frac{N-n}{n-1}$ es un factor de corrección cuando la muestra es mayor al 5 % del tamaño de la población.

Distribución Poblacional



Distribución media muestral



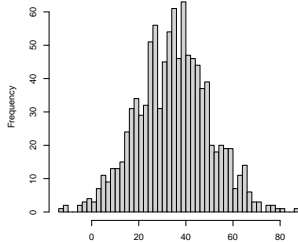
La distribución muestral de la varianza inicia con siguiente teorema.

- ▶ X_1, X_2, \dots, X_n son observaciones de una muestra aleatoria n de una distribución $N(\mu, \sigma^2)$
- ▶ $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ es la muestra de la media de n observaciones.
- ▶ $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ es la varianza de la muestra de n observaciones.

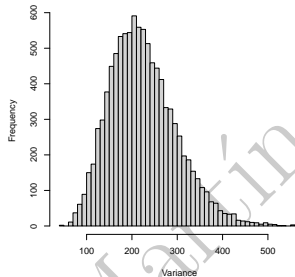
Por lo tanto:

- ▶ \bar{X} y S^2 son independientes.
- ▶ $\frac{n-1S^2}{\sigma^2} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} \sim \chi^2(n-1)$

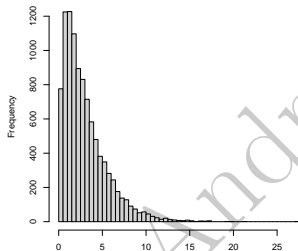
Distribución normal Población



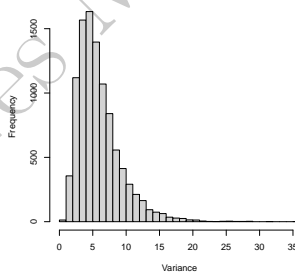
Distribución de la varianza



Distribución Chi Cuadrado



Distribución de la muestra



La población de una proporción es

$$p = \frac{X}{N} \quad (13)$$

Donde

- ▶ p La proporción de la población.
- ▶ X Número de observaciones con los atributos de interés que se distribuyen de forma binomial $B \sim (p, N)$
- ▶ N Población.

La muestra de una proporción es

$$\bar{p} = \frac{x}{n} \quad (14)$$

- ▶ \bar{p} La proporción de la muestra
- ▶ x Número de observaciones con los atributos de se distribuyen de forma binomial $B \sim (p, n)$
- ▶ n Muestra.

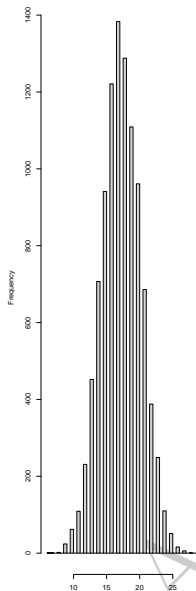
La población de una proporción es

$$p = \frac{X}{N} \quad (15)$$

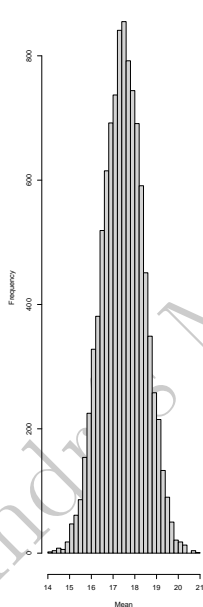
Donde

- ▶ La mejor estimación de la proporción poblacional será \bar{p} , la proporción muestral
- ▶ Cualquier inferencia sobre qué tan cerca está su estimación del valor real de la población se basará en la distribución de esta proporción muestral, \bar{p} .
- ▶ Si el tamaño de la muestra es lo suficientemente grande como para que $np \geq 5$ y $n(1 - p) \geq 5$, entonces la distribución normal se puede utilizar como una aproximación razonable a la distribución binomi

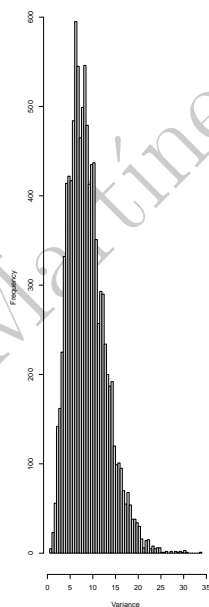
Distribución de una Población



Distribución de la media



Distribución de la Varianza



Media de la muestra de una proporción

$$\mu_{\bar{p}} = p \quad (16)$$

Varianza

$$\sigma_{\frac{\bar{p}}{p}}^2 = \frac{\sigma^2}{n} = \frac{p(1-p)}{n} \quad (17)$$

Media de la muestra de una proporción

Independientemente del valor de la proporción de la población, p , (con las obvias excepciones de $p = 0$ y $p = 1$), la distribución muestral para la proporción muestral, \bar{p} , se distribuirá aproximadamente normalmente con $\mu_{\bar{p}} = p$ y $\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}}$ con $np \geq 5$ y $n(1 - p) \geq 5$. La aproximación a la la distribución normal mejora a medida que aumenta el tamaño de la muestra y p se acerca a 0.50

Aproximación binomial

$$z = \frac{\bar{p} - p}{\sigma} \quad (18)$$

Desviación estandar

$$\sigma = \sqrt{np(1-p)} \quad (19)$$

Corrección de población

$$\sigma = \sqrt{np(1-p)} \sqrt{\frac{N-n}{n-1}} \quad (20)$$

Estimación La estimación estadística es el proceso mediante el cual intentamos determinar el valor de un parámetro de la población sin hacer un censo y a partir de la información de una muestra. [Llinás Solano, 2016]

- ▶ El estimador es el estadístico de la muestra utilizado para hacer una estimación estadística.
- ▶ Una estimación es un posible valor del estimador.

Insesgo Se dice que un estimador $\hat{\theta}$ es insesgado si el valor esperado del estimador es igual al parámetro θ de la población que está estimando es $E(\hat{\theta}) = \theta$. Evidentemente si $E(\hat{\theta}) \neq \theta$ se dice que el estimador es sesgado [Llinás Solano, 2016]

$$B(\hat{\theta}) = E(\hat{\theta}) - \theta \quad (21)$$

Eficiencia Sean $\hat{\theta}_1$ y $\hat{\theta}_2$ dos estimadores insesgados de θ , obtenidos en muestras del mismo tamaño. Entonces [Llinás Solano, 2016]

- ▶ Se dice que $\hat{\theta}_1$ es más eficiente que $\hat{\theta}_2$ si la varianza de la distribución muestral de $\hat{\theta}_1$ es menor que la varianya muestral de $\hat{\theta}_2$. Es decir que $V(\hat{\theta}_1) < V(\hat{\theta}_2)$.
- ▶ La eficiencia relativa de $\hat{\theta}_2$, con respecto a $\hat{\theta}_1$ es el cociente de $\frac{V(\hat{\theta}_2)}{V(\hat{\theta}_1)}$ de sus varianzas.

Consistencia Un estimador puntual $\hat{\theta}$ de θ es consistente para θ si sus valores tienden a acercarse al parámetro poblacional θ conforme se incrementa el tamaño de la muestra.

[Llinás Solano, 2016]

Suficiencia Un estadístico $S(x)$ es suficiente para θ si y solo si para todo valor s de $S(X)$ la distribución condicional de X dado $S(X) = s$ no depende de θ . [Llinás Solano, 2016]

Un estimador es una regla, a menudo expresada como una fórmula, que indica cómo calcular el valor de una estimación con base en las mediciones contenidas en una muestra.[Wackerly et al., 2010]

- ▶ Un estadístico único, determinado a partir de una muestra que se utiliza para estimar el parámetro de población correspondiente
- ▶ El parámetro de población se puede estimar con estadística de muestra.

Error cuadrático de un estimador puntual $\hat{\theta}$ es

$$MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] \quad (22)$$

$$MSE(\hat{\theta}) = V(\hat{\theta}) + [B(\hat{\theta})]^2$$

Cuadro: Estimadores puntuales comunes

Parámetro objetivo θ	Tamaño de la Muestra	Estimador puntual $\hat{\theta}$	$E[\hat{\theta}]$	Error estándar $\sigma_{\hat{\theta}}$
μ	n	\bar{X}	μ	$\frac{\sigma}{\sqrt{n}}$
p	n	$\hat{p} = \frac{Y}{n}$	p	$\sqrt{\frac{pq}{n}}$
$\mu_1 - \mu_2$	n_1 y n_2	$\bar{X}_1 - \bar{X}_2$	$\mu_1 - \mu_2$	$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
$p_1 - p_2$	n_1 y n_2	$\hat{p}_1 - \hat{p}_2$	$p_1 - p_2$	$\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$

El k momento teórico de una variable aleatoria es

$$\mu_k = E(X^k) = \int x^k f(x | \theta) dx \quad (24)$$

$$\mu_k = E(X^k) = \sum x^k f(x | \theta) \quad (25)$$

Si X_1, \dots, X_n es una variable aleatoria iid el momento k es

$$m_k = E(X^k) = \frac{1}{n} \sum x^k \quad (26)$$

Entonces m_k se puede ver como un estimador $m_k \rightarrow \mu_k$ en la medida en que $n \rightarrow \infty$

Suponga que X es una variable discreta con la siguiente función de probabilidad donde θ es un parametro. Con la siguiente muestra obtenga un estimador $\hat{\theta}$ para θ .

X	0	1	2	3
$P(X)$	$2\theta/3$	$\theta/3$	$2(1 - \theta)/3$	$(1 - \theta)/3$

La muestra es (3, 1, 0, 2, 3, 1, 3, 1, 0, 2) Encuentre θ

Use el método de momentos para estimar los parámetros μ y σ^2 para una densidad normal.

$$f(x | \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right) \quad (27)$$

La primera y segunda solución de una distribución teórico

$$\mu_1 = E(X) = \mu \quad y \quad \mu_2 = E(X^2) = \mu^2 + \sigma^2 \quad (28)$$

El primer y segundo momento es:

$$m_1 = \overline{X} \quad y \quad m_2 = \frac{1}{n} \sum_{i=1}^n X_i^2 \quad (29)$$

Resolviendo

$$\mu = \overline{X} \quad y \quad \mu^2 + \sigma^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 \quad (30)$$

Por lo tanto

$$\hat{\mu} = \overline{X} \quad y \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \overline{X}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \overline{X})^2 \quad (31)$$

Obtenga el primer y segundo método de momentos para de una variable aleatoria que se distribuye de forma uniforme

$$X \sim U(a, b)$$

$$f(x) = \begin{cases} 1/(b-a) & \text{si } x \in [a, b] \\ 0 & \text{if } x \notin [a, b] \end{cases}$$

El primer momento es:

$$\int_a^b xf(x)dx = \int_a^b \frac{x}{b-a} dx = \frac{1}{2} \frac{b^2 - a^2}{b-a} = \frac{b+a}{2} \quad (32)$$

El segundo momento es

$$\int_a^b x^2 f(x) dx = \int_a^b \frac{x^2}{b-a} dx = \frac{1}{3} \frac{b^3 - a^3}{b-a} = \frac{b^2 + ba + a^2}{3} \quad (33)$$

Propiedades de los estimadores de momentos

Los estimadores obtenidos por el método de momentos son consistentes, pero no son en general, ni centrados, ni con varianza mínima, ni robustos. La ventaja de estos estimadores es su simplicidad; su inconveniente es que al no tener en cuenta la distribución de la población que genera los datos no utiliza toda la información de la muestra. [Peña, 2014]

Suponga que tiene las siguientes variables aleatorias Y_1, \dots, Y_n de una muestra aleatoria con la siguiente distribución $f(y | \theta)$; Si Y es una variable aleatoria continua, $f(y | \theta)$ es pdf, si Y es una variable aleatoria discreta, $f(y | \theta)$ es una función de masa de probabilidad pmf. La distribución depende de un parámetro θ , donde θ es un valor real de un parámetro desconocido. Para cada muestra aleatoria y_1, \dots, y_n

$$f(y_1, \dots, y_n | \theta) = f(y_1 | \theta) \dots f(y_n | \theta) \quad (34)$$

Si $f(y | \theta)$ es pdf, $f(y_1, \dots, y_n | \theta)$ es la función de densidad conjunta; Si $f(y | \theta)$ es pmf, $f(y_1, \dots, y_n | \theta)$ es la probabilidad conjunta. Donde $f(y_1, \dots, y_n | \theta)$ es la función de probabilidad que depende de un parámetro θ desconocido y siempre se denota como $L(\theta)$.

MLE requiere obtener la mayor probabilidad de la función $L(\theta)$ con respecto al parámetro desconocido θ . $L(\theta)$ es definido como el producto de n términos, que debe ser maximizado.

Maximizando $L(\theta)$ es equivalente a maximizar $\log L(\theta)$ porque \log es una función monótona creciente. Definimos $L(\theta)$ como la función de verosimilitud donde $l(\theta)$ es la función de verosimilitud logarítmica.

$$l(\theta) = \log L(\theta) = \log \prod_{i=1}^n f(Y_i | \theta) = \sum_{i=1}^n \log(f(Y_i | \theta)) \quad (35)$$

Maximizando $l(\theta)$ con respecto a θ se obtiene la estimación MLE para el parámetro θ .

Suponga que la función de verosimilitud depende de k parámetros $\theta_1, \theta_2, \dots, \theta_k$. Escoja como estimaciones los valores de los parámetros que maximicen la verosimilitud $L(y_1, y_2, \dots, y_n \mid \theta_1, \theta_2, \dots, \theta_k)$ [Wackerly et al., 2010].

Un experimento binomial consistente en n ensayos resultó en las observaciones y_1, y_2, \dots, y_n , donde $y_i = 1$ si el i -ésimo intento fue un éxito y $y_i = 0$ en cualquier otro punto. Encuentre el MLE de p .

$$f(y) = \left(\frac{n!}{y_i!(n - y_i)!} \right) p^{y_i} (1 - p)^{n - y_i} \quad (36)$$

$$L(p) = \prod_{i=1}^n f(y_i) = \prod_{i=1}^n \left(\frac{n!}{y_i!(n - y_i)!} \right) p^{y_i} (1 - p)^{n - y_i} \quad (37)$$

$$L(p) = \left(\prod_{i=1}^n \left(\frac{n!}{y_i!(n-y_i)!} \right) \right) p^{\sum_{i=1}^n y_i} (1-p)^{n-\sum_{i=1}^n y_i} \quad (38)$$

$$\ln L(p) = \sum_{i=1}^n y_i \ln(p) + \left(n - \sum_{i=1}^n y_i \right) \ln(1-p) \quad (39)$$

$$\frac{d \ln L(p)}{dp} = \frac{1}{p} \sum_{i=1}^n y_i + \frac{1}{1-p} \left(n - \sum_{i=1}^n y_i \right) = 0 \quad (40)$$

$$(1-\hat{p}) \sum_{i=1}^n y_i + p \left(n - \sum_{i=1}^n y_i \right) = 0 \quad (41)$$

$$\hat{p} = \frac{\sum_{i=1}^n y_i}{n} = \frac{k}{n} \quad (42)$$

Un experimento binomial consistente en n ensayos resultó en las observaciones y_1, y_2, \dots, y_n , donde $y_i = 1$ si el i -ésimo intento fue un éxito y $y_i = 0$ en cualquier otro punto. Encuentre el MLE de p .

con $\lambda = np$

$$f(y) = \frac{\lambda^y e^{-\lambda}}{y!} \quad y = 0, 1, 2, \dots \quad (43)$$

$$L(\lambda) = \prod_{i=1}^n \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} = e^{-n\lambda} \frac{\lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} \quad (44)$$

$$\ln L(\lambda) = -n\lambda + \sum_{i=1}^n x_i \ln(\lambda) - \ln\left(\prod_{i=1}^n x_i!\right) \quad (45)$$

$$\frac{d \ln L(\lambda)}{d\lambda} = -n + \sum_{i=1}^n x_i \frac{1}{\lambda} \quad (46)$$

$$\hat{\lambda} = \frac{\sum_{i=1}^n x_i}{n}$$

Un experimento que se realiza n veces continuo, resultó en las observaciones y_1, y_2, \dots, y_n , donde $y_i = 1$ si el i -ésimo intento fue un éxito y $y_i = 0$ en cualquier otro punto. Encuentre el MLE de μ y σ .

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} \quad (48)$$

$$f(y_1, \dots, y_n \mid \sigma, \mu) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i - \mu)^2}{2\sigma^2}} \quad (49)$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2} \quad (50)$$

$$\log(f(y_1, \dots, y_n \mid \sigma, \mu)) = \log \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2} \quad (51)$$

$$= n \log \frac{1}{\sigma\sqrt{2\pi}} - \frac{1}{2\sigma^2} \sum (y_i - \mu)^2 \quad (52)$$

$$= -\frac{n}{2} \log(2\pi) - n \log \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2$$

Derivamos $f(y_1, \dots, y_n \mid \sigma, \mu)$

$$\frac{\partial \mathcal{L}}{\partial \mu} = -\frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \mu) = \frac{n\bar{y} - n\mu}{\sigma^2} = 0 \quad (54)$$

$$\hat{\mu} = \frac{\sum_{i=1}^n y_i}{n} \quad (55)$$

$$\frac{\partial \mathcal{L}}{\partial \sigma} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2 = 0 \quad (56)$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (y_i - \mu)^2}{n} \quad (57)$$

Matriz Hessiana y la matriz de información de fisher

$$\begin{bmatrix} \frac{\partial^2 l}{\partial \mu^2} & \frac{\partial l}{\partial \mu \partial \sigma} \\ \frac{\partial^2 l}{\partial \mu \partial \sigma} & \frac{\partial^2 l}{\partial \sigma^2} \end{bmatrix} \quad (58)$$

$$- \begin{bmatrix} -\frac{n}{\sigma^2} & -\frac{\sum(y_i - \mu)}{\sigma^4} \\ -\frac{\sum(y_i - \mu)}{\sigma^4} & -\frac{\sum(y_i - \mu)}{\sigma^6} + \frac{n}{2\sigma^4} \end{bmatrix} \quad (59)$$

$$\begin{bmatrix} \frac{n}{\sigma^2} & \frac{\sum(y_i - \mu)}{\sigma^4} \\ \frac{\sum(y_i - \mu)}{\sigma^4} & \frac{\sum(y_i - \mu)}{\sigma^6} - \frac{n}{2\sigma^4} \end{bmatrix} \quad (60)$$

Matriz Hessiana y la matriz de información de fisher

$$\mathcal{I}(\theta) = \begin{bmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{bmatrix} \quad (61)$$

$$\mathcal{I}(\theta)^{-1} = \begin{bmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{2\sigma^4}{n} \end{bmatrix} \quad (62)$$

Para $\sigma^2 = \text{Var}(Y) = E(Y^2) - (E(Y))^2$

$$E(\hat{\sigma}^2) = \frac{1}{n} E\left(\sum_{i=1}^n (y_i - \hat{\mu})^2\right) \quad (63)$$

$$= \frac{1}{n} E\left(\sum_{i=1}^n ((y_i^2 - n\hat{\mu}^2))\right) \quad (64)$$

$$= \frac{1}{n} E\left(\sum_{i=1}^n \left((y_i^2 - \frac{(\sum y_i)^2}{n})\right)\right) \quad (65)$$

$$= \frac{1}{n} \left(\sum_{i=1}^n E((y_i^2)) - \frac{1}{n} E(\sum y_i)^2\right) \quad (66)$$

Para $\sigma^2 = \text{Var}(Y) = E(Y^2) - (E(Y))^2$

$$E(\hat{\sigma}^2) = \frac{1}{n} \left(\sum_{i=1}^n (\sigma^2 + \mu^2) - \frac{1}{n} [n\sigma^2 n\mu^2] \right) \quad (68)$$

$$= \frac{1}{n} (n\sigma^2 + n\mu^2 - \sigma^2 - \mu^2) \quad (69)$$

$$= \frac{n-1}{n} \sigma^2 \quad (70)$$

$$\neq \sigma^2 \quad (71)$$

Propiedades [Peña, 2014]

- ▶ Asintoticamente cerrados
- ▶ Con distribución asintoticamente normal
- ▶ Si existe un estadístico suficiente para el parámetro, el estimador es máximo verosimil
- ▶ Invariantes

"Un intervalo de confianza puede interpretarse como una estimación por un intervalo alrededor de la estimación puntual y dará una mejor idea sobre la precisión de la estimación"[Llinás Solano, 2016, p. 110].

Para construir un intervalo de confianza se debe seguir el siguiente procedimiento:

- ▶ Según el problema se tiene un parámetro $\theta \in \Theta \subseteq \mathbb{R}^k$, .estimar.el valor $\eta = q(\theta)$, siendo $q : \Theta \rightarrow \mathbb{R}$ una dimensión unidimensional.
- ▶ Se toma una muestra $X = (X_1, \dots, X_n)^t$ de tamaño n , cuya distribución f_θ debe ser conocida para cada θ , observando un dato concreto $x = (x_1, \dots, x_n)^t$.
- ▶ Se buscan dos estadísticos unidimensionales, $U(X)$ y $W(X)$, con $U(X) \leq W(X)$, para cada posible observación x de X tal que:

$$P(U(X) < \eta < W(X)) \geq 1 - \alpha \quad (79)$$

Para cierto α fijo y lo más pequeño posible

Definición Considere la metodología explicada anteriormente para construir un intervalo en la ecuación 72. Al intervalo aleatorio $I(X) = (U(X), W(X))$ se le llama intervalo de confianza para η con un grado de confianza de $(1 - \alpha)100\%$. También a su valor se le llama intervalo de confianza.[Llinás Solano, 2016]

Para estimar el intervalo entonces se toma el estimador puntal $\hat{\eta} = q(\hat{\theta})$

$$U = \eta - D$$

$W = \eta + D$ Con un estadístico $D = D(X)$ que cumple con la ecuación 72.

De esta forma un intervalo se define como $I = (\hat{\eta} - l$

En un escenario en donde no conocemos la media, pero además sabemos que podemos obtener diferentes valores de acuerdo al tipo de muestra que estamos hallando, entonces establecemos un intervalo de acuerdo a las siguientes condiciones:

- ▶ La población es normal con varianza conocida
- ▶ La población es normal con varianza desconocida

De esta forma el intervalo se define como $I(\bar{X}_n - D, \bar{X}_n + D)$, donde $\hat{\mu} = \hat{\mu}(X) = \bar{X}$

$$P(\mu \in I) = P(\bar{X}_n - D < \mu < \bar{X}_n + D) = 1 - \alpha \quad (73)$$

$$-\frac{D}{\sigma/\sqrt{n}} < \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} < \frac{D}{\sigma/\sqrt{n}} \quad (74)$$

$$1 - \alpha = P(\bar{X}_n - D < \mu < \bar{X}_n + D) = \Phi(z) - \Phi(-z) = 2\Phi(z) - 1 \quad (75)$$

donde $\Phi(z) = 1 - \frac{\alpha}{2}$ y $D = Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$

$$\bar{X}_n - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X}_n + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \quad (76)$$

Caso de las muestras pequeñas. Cuando σ^2 no se conoce, podemos usar el ejemplo que aparece en [Llinás Solano, 2016, p. 113] para encontrar el intervalo bajo las siguientes condiciones:

$$t = \frac{\bar{X}_n - \mu}{S_n / \sqrt{n}} \rightarrow t(n-1) \quad (77)$$

$$S_n^2 = \frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{n-1} \quad (78)$$

$$D = t_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

$$\bar{X} - t_{\frac{\alpha}{2}} \frac{S_n}{\sqrt{n}} < \mu < \bar{X} + t_{\frac{\alpha}{2}} \frac{S_n}{\sqrt{n}} \quad (79)$$

De igual forma podemos establecer un intervalo de confianza acorde a la estimación puntual para la proporción, donde $\hat{p} \sim N(p, \frac{p(1-p)}{n})$ y con $\hat{p}(1 - \hat{p}) \rightarrow p(1 - p)$

$$\hat{p} - Z_{\frac{\alpha}{2}} \frac{\sqrt{\hat{p}(1 - \hat{p})}}{\sqrt{n}} < p < \hat{p} + Z_{\frac{\alpha}{2}} \frac{\sqrt{\hat{p}(1 - \hat{p})}}{\sqrt{n}} \quad (80)$$

En donde el tamaño de la muestra debe ser $n \geq 30$ o que $n\hat{p} > 5$ y $n(1 - \hat{p}) > 5$. En este caso $Z_{\frac{\alpha}{2}} = \frac{\bar{X} - p}{\sqrt{p(1-p)/n}}$

En el procedimiento de muestreo, o diseño experimental, la pregunta que surge es ¿Cuanta información necesitamos para el proyecto o experimento?. En este caso usamos el error de estimación α que nos puede ayudar a determinar el tamaño de una muestra. Usando la regla empírica podemos saber una valor aproximado si las observaciones se distribuyen de forma normal.

$$\frac{2\sigma}{\sqrt{n}} = B \quad \text{o} \quad n = \frac{4\sigma^2}{B^2} \quad (81)$$

Donde B es el valor límite que esperamos de la distribución de la media muestral.

¿Es común el uso del parámetro poblacional σ^2 para obtener la muestra?

Sean las X_1, \dots, X_n variables aleatorias independientes, con una muestra de tamaño n que se distribuyen de forma normal $N \sim (\mu_x, \sigma_x^2)$, sea Y_1, \dots, Y_m otro conjunto de variables aleatorias independientes, con una muestra de tamaño m que se distribuyen de forma normal $N \sim (\mu_y, \sigma_y^2)$.
en[Llinás Solano, 2016]

Dos estimadores para diferencia de medias se puede construir bajo las siguientes condiciones:

- ▶ Varianzas poblacionales conocidas o desconocidas y muestras grandes.
- ▶ Varianzas poblacionales iguales, desconocidas y muestras pequeñas.
- ▶ Varianzas poblacionales diferentes, desconocidas pequeñas.

$$\bar{X} \sim N\left(\mu, \frac{\sigma_x^2}{n}\right) \quad y \quad \bar{Y} \sim N\left(\mu, \frac{\sigma_y^2}{m}\right) \quad (82)$$

$$\bar{X}_n - \bar{Y}_m \sim N\left(\mu_x - \mu_y, \frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}\right) \quad (83)$$

$$Z = \frac{(\bar{X}_n - \bar{Y}_m) - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}} \quad (84)$$

$$\bar{X}_n - \bar{Y}_m - Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}} < \mu_x - \mu_y < \bar{X}_n - \bar{Y}_m + Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}$$

$$\frac{(n-1)S_n^2}{\sigma^2} \sim \chi_{(n-1)}^2 \quad y \quad \frac{(m-1)S_m^2}{\sigma^2} \sim \chi_{(m-1)}^2 \quad (86)$$

$$\frac{(n-1)S_n^2 + (m-1)S_m^2}{\sigma^2} \sim \chi_{(n+m-2)}^2 \quad (87)$$

$$t_{(n+m-2)} = \frac{(\bar{X}_n - \bar{Y}_m) - (\mu_x - \mu_y)}{\sqrt{\frac{S_{n,m}^2}{n} + \frac{S_{n,m}^2}{m}}} \quad (88)$$

$$S_{(n,m)}^2 = \frac{(n-1)S_n^2 + (m-1)S_m^2}{(n+m-2)} \quad (89)$$

$$\bar{X}_n - \bar{Y}_m - t_{\frac{\alpha}{2}}(\nu)c < \mu_x - \mu_y < \bar{X}_n - \bar{Y}_m + t_{\frac{\alpha}{2}}(\nu)c \quad (90)$$

$$\nu = (n + m - 2) \text{ y } c = \sqrt{\frac{S_{n,m}^2}{n} + \frac{S_{n,m}^2}{m}}$$

Para el tercer caso la única variación es:

$$t_{(\nu)} = \frac{(\bar{X}_n - \bar{Y}_m) - (\mu_x - \mu_y)}{\sqrt{\frac{S_n^2}{n} + \frac{S_m^2}{m}}} \quad (91)$$

$$\nu = \frac{\left(\frac{S_n^2}{n} + \frac{S_m^2}{m}\right)^2}{\frac{(\frac{S_n^2}{n})^2}{n-1} + \frac{(\frac{S_m^2}{m})^2}{m-1}} \quad (92)$$

Siguiendo el mismo proceso para las proporciones para estimación puntual e intervalo para proporciones. Para $\bar{p} = \bar{p}_1 - \bar{p}_2$

$$b = Z_{\frac{\alpha}{2}} \sqrt{\frac{\bar{p}_1(1 - \bar{p}_1)}{n} + \frac{\bar{p}_2(1 - \bar{p}_2)}{m}} \quad (93)$$

Karl Popper " El hombre de ciencia, ya sea teórico o experimental, propone enunciados o sistemas de enunciados y los contrasta paso a paso. En particular, en el campo de las ciencias empíricas construye hipótesis o sistemas de teorías y las contrasta con la experiencia por medio de observaciones y experimentos"[Mardones, 2007, p. 86]

- ▶ Hipótesis nula
- ▶ Hipótesis Alternativa
- ▶ Error tipo I y Error tipo II
- ▶ P value
- ▶ Región de rechazo

La hipótesis nula, que se simboliza con H_0 , es la hipótesis que se debe comprobar. La hipótesis alternativa, simbolizada con H_1 , es la hipótesis elegida como contraste a H_0 .

[Llinás Solano, 2016, p. 136]

En general, si θ es un parámetro poblacional y k es cualquier número real, entonces la hipótesis alternativa es $H_1 : \theta \neq k$, se llama alertnativa bilateral y cada una de las hipótesis alternativas $H_1 : \theta < k$ y $H_1 : \theta > k$ es alternativa unilateral.[Llinás Solano, 2016, p. 136]

Un prueba de hipótesis tiene los siguientes pasos
[Llinás Solano, 2016]:

- ▶ La hipótesis se plantea de acuerdo al problema y la función de distribución y se contrasta el parámetro de interés θ .
- ▶ El modelo estadístico se formula con una muestra aleatoria de tamaño n con distribución conocida.
- ▶ Se escoge un estadístico de prueba $T(X)$ unidimensional, de tal manera que se pueda rechazar H_0 con base en x si y solo si $T(X) \geq c$, donde c es el valor crítico del test de hipótesis.

Error tipo I: Para un c se escoge un $\alpha \in (0, 1)$ se cumple que:

$$P(T(X) \geq c/H_0) = \sup_{\theta \in \Theta} P(T(X) \geq c/\theta) \leq \alpha \quad (94)$$

" $T(X) \geq c/H_0$ "significa " rechazar H_0 a pesar de que H_0 fuese correcta.

Error tipo II: Para un c se escoge un $\beta \in (0, 1)$ se cumple que:

$$P(T(X) \geq c/H_1) = \sup_{\theta \in \Theta} P(T(X) \geq c/\theta) \leq \beta \quad (95)$$

" $T(X) \geq c/H_1$ "significa " no rechazar H_0 a pesar de que H_0 no fuese correcta.

El p valor es el mínimo nivel de significancia en el cual la hipótesis nula H_0 sería rechazada cuando se utiliza un procedimiento de prueba especificado con un conjunto dado de información.

- ▶ Si el $P - \text{valor} \geq \alpha$, entonces se rechaza H_0 al nivel α .
- ▶ Si el $P - \text{valor} < \alpha$, entonces no se rechaza H_0 al nivel α .

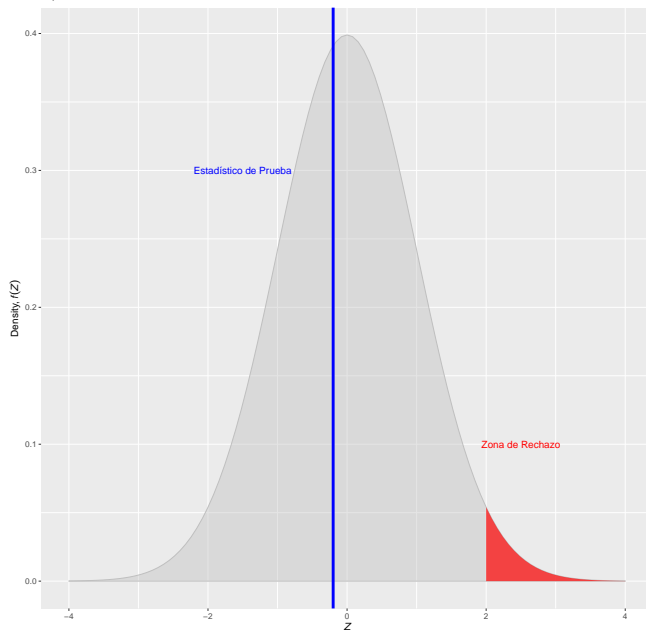
$H_0 : \theta = \theta_0$ [Wackerly et al., 2010, p. 500]

$$H_a = \begin{cases} \theta > \theta_0 & \text{alternativa de cola superior} \\ \theta < \theta_0 & \text{alternativa de cola inferior} \\ \theta \neq \theta_0 & \text{alternativa de dos colas} \end{cases}$$

Estadístico de prueba: $Z = \frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}}$

$$\text{Región de rechazo} = \begin{cases} \{z > z_{\alpha}\} & \text{Cola superior} \\ \{z < -z_{\alpha}\} & \text{Cola inferior} \\ \{|z| > z_{\alpha/2}\} & \text{Dos colas} \end{cases}$$

Hipótesis a una cola



$$H_0 : \mu = \mu_0 [\text{Wackerly et al., 2010, p. 521}]$$

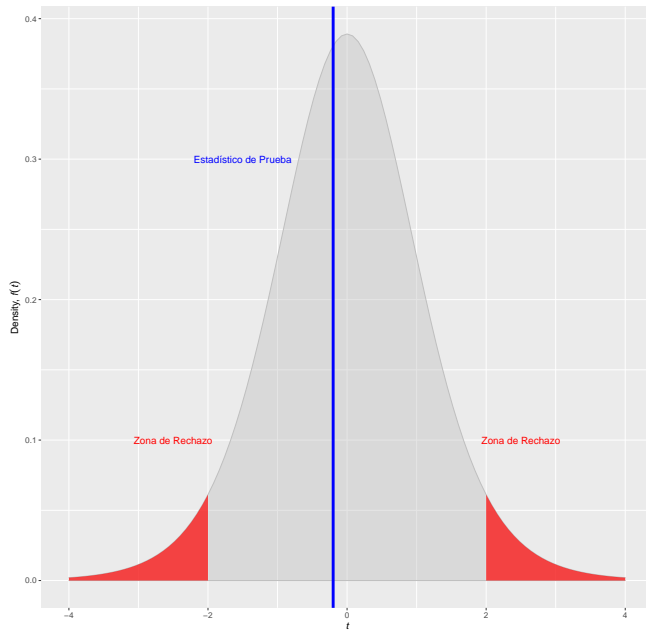
$$H_a = \begin{cases} \theta > \theta_0 & \text{alternativa de cola superior} \\ \theta < \theta_0 & \text{alternativa de cola inferior} \\ \theta \neq \theta_0 & \text{alternativa de dos colas} \end{cases}$$

$$\text{Estadístico de prueba: } T = \frac{\bar{Y} - \mu_0}{S/\sqrt{n}}$$

$$\text{Región de rechazo} = \begin{cases} \{t > t_\alpha\} & \text{Cola superior} \\ \{t < -t_\alpha\} & \text{Cola inferior} \\ \{|t| > t_{\alpha/2}\} & \text{Dos colas} \end{cases}$$

$$\nu = n - 1 \text{ grados de libertad.}$$

Hipótesis a dos colas



$$H_0 : \mu_1 - \mu_2 = D_0 [\text{Wackerly et al., 2010, p. 522}]$$

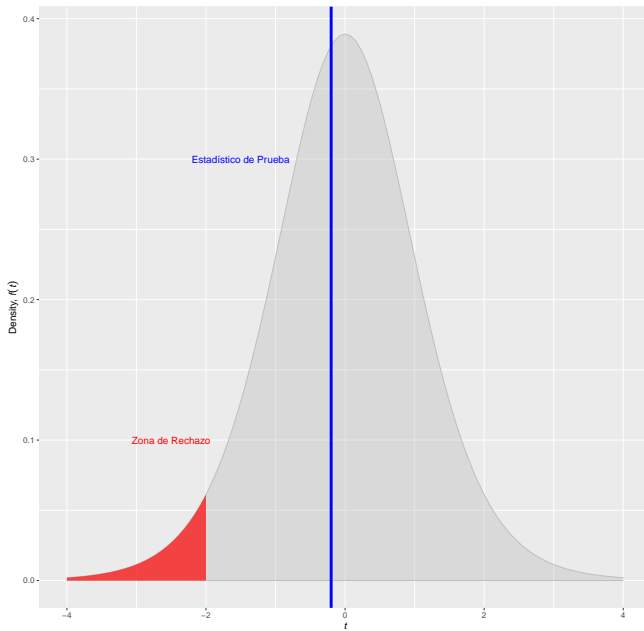
$$H_a = \begin{cases} \mu_1 - \mu_2 > D_0 & \text{alternativa de cola superior} \\ \mu_1 - \mu_2 < D_0 & \text{alternativa de cola inferior} \\ \mu_1 - \mu_2 \neq D_0 & \text{alternativa de dos colas} \end{cases}$$

$$\text{Estadístico de prueba: } T = \frac{(\bar{Y}_1 - \bar{Y}_2) - D_0}{S_{n,m} / \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

$$\text{Región de rechazo} = \begin{cases} \{t > t_\alpha\} & \text{Cola superior} \\ \{t < -t_\alpha\} & \text{Cola inferior} \\ \{|t| > t_{\alpha/2}\} & \text{Dos colas} \end{cases}$$

$$\nu = n + m - 2 \text{ grados de libertad.}$$

Hipótesis a una cola



Siendo $X \sim N(\mu, \sigma^2)$.

$H_0 : \sigma^2 = \sigma_0^2$ [Wackerly et al., 2010, p. 532]

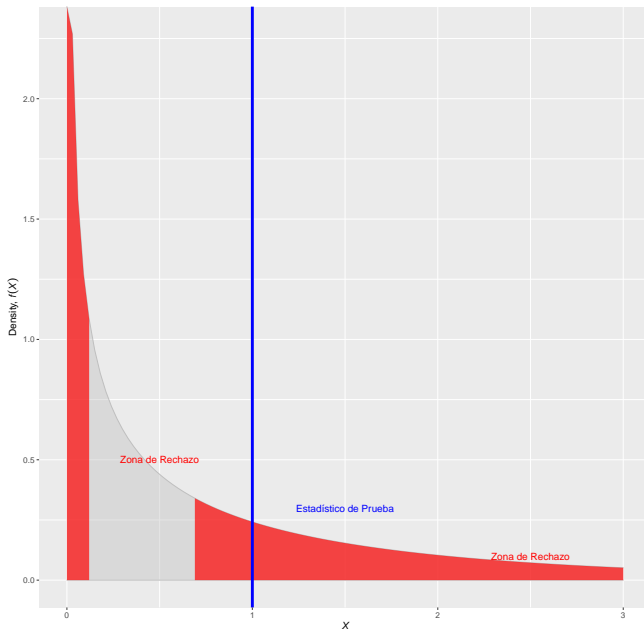
$$H_a = \begin{cases} \sigma^2 > \sigma_0^2 & \text{alternativa de cola superior} \\ \sigma^2 < \sigma_0^2 & \text{alternativa de cola inferior} \\ \sigma^2 \neq \sigma_0^2 & \text{alternativa de dos colas} \end{cases}$$

Estadístico de prueba: $\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$

$$\text{Región de rechazo} = \begin{cases} \{\chi^2 > \chi_{\alpha}^2\} & \text{Cola superior} \\ \{\chi^2 < \chi_{1-\alpha}^2\} & \text{Cola inferior} \\ \{\chi^2 > \chi_{\alpha/2}^2\} \text{ o } \{\chi^2 < \chi_{1-\alpha/2}^2\} & \text{Dos colas} \end{cases}$$

$\nu = n - 1$ grados de libertad.

Hipótesis a dos colas



Para dos muestras aleatorias X y Y independientes con distribuciones normales y medias desconocidas donde $V(X) = \sigma_x^2$ y $V(Y) = \sigma_y^2$ donde σ_x^2 y σ_y^2 se desconocen.

$H_0 : \sigma_x^2 = \sigma_y^2$ y la alterna $H_1 : \sigma_x^2 > \sigma_y^2$. Usando las varianzas muestrales S_x^2 y S_y^2 .

$$\text{Zona de Rechazo} = \frac{S_x^2}{S_y^2} > k \quad (96)$$

Usamos la distribución F con parámetros $\nu_1 = n_1 - 1$ y $\nu_2 = n_2 - 1$ para calcular el valor crítico y el estadístico de prueba se obtiene con:

$$F = \frac{(n_1 - 1)S_x^2}{\sigma_x^2(n_1 - 1)} / \frac{(n_2 - 1)S_y^2}{\sigma_y^2(n_2 - 1)} = \frac{S_x^2}{S_y^2}$$

Están relacionados con variables cualitativas que se contabilizan de forma ordinal o nominal. Los experimentos se diseñan buscan promedios o diferencias entre cada una de las categorías y esto se puede hacer con ayuda de procesos binomiales o multinomiales[Wackerly et al., 2010].

- ▶ El experimento consiste en n ensayos
- ▶ El resultado de cada ensayo cae exactamente en k categorías o celdas distintas.
- ▶ La probabilidad de caer en una celda i es p_i , donde $i = 1, 2, \dots, k$
$$p_1 + p_2 + p_3 + \dots + p_k = 1$$
- ▶ Los ensayos son independientes.
- ▶ Los ensayos indican el número de veces que el resultado cayó en cada celda $n_1, n_2, \dots, n_k = n$

El objetivo principal es comprobar si la hipótesis nula se ajusta a lo que presentan los datos y a esto se le conoce como prueba de bondad de ajuste.[Johnson and Bhattacharyya, 2019]

$$H_0 = p_1 = p_{10}, \dots, p_k = p_{k0} \quad (98)$$

Celdas	1	2	k	Total
O	n_1	n_2	...	n_k	n
Probabilidad H_0	p_{10}	p_{20}	...	p_{k0}	1
E	np_{10}	np_{20}	...	np_{k0}	n

O = Frecuencia Observada, Probabilidad bajo H_0 y E = Valor esperado de la Frecuencia bajo H_0 .

Prueba χ^2 Basada en una muestra grande n

Hipótesis Nula

$$H_0 = p_1 = p_{10}, \dots, p_k = p_{k0} \quad (99)$$

Estadístico de Prueba

$$\chi^2 = \sum_{i=1}^k \frac{(n_i - np_{i0})^2}{np_{i0}} = \sum_{cells} \frac{(O - E)^2}{E} \quad (100)$$

Zona de Rechazo

$$\chi^2 \geq \chi_{\alpha}^2 \quad (101)$$

Grados de libertad $df = k - 1$

Muestras independientes clasificadas en varias categorías

Para comparar la efectividad de dos dietas A y B, se incluyeron 150 bebés en un estudio. La dieta A se administró a 80 lactantes seleccionados al azar y la dieta B a los otros 70 lactantes. Posteriormente, se observó la salud de cada bebé y se clasificó en una de las tres categorías "excelente", "promedio" o "deficiente". A partir de los recuentos de frecuencia registrados en la Tabla, deseamos probar la hipótesis nula de que no hay diferencia entre la calidad de las dos dietas. [Johnson and Bhattacharyya, 2019, p. 516]

	Excelente	Promedio	Pobre	Muestra
Dieta A	37	24	19	80
Dieta B	17	33	20	70
Total	54	57	39	150

	Excelente	Promedio	Pobre	Total
Dieta A	0.46	0.30	0.24	1
Dieta B	0.24	0.47	0.29	1

Las proporciones o probabilidades (desconocidas) de la población se ingresan en la Tabla. Nos permiten describir la hipótesis nula con mayor claridad. La hipótesis nula de “no diferencia” es equivalente a la afirmación de que, para cada categoría de respuesta, la probabilidad es el mismo para la dieta A y la dieta B. En consecuencia, formulamos

$$H_0 = p_{A1} = p_{B1} \quad p_{A2} = p_{B2} \quad p_{A3} = p_{B3} \quad (102)$$

	Excelente	Promedio	Pobre	Total
Dieta A	p_{A1}	p_{A2}	p_{A3}	1
Dieta B	p_{B1}	p_{B2}	p_{B3}	1

1. Calculamos las probabilidades por columna

$$\hat{p}_1 = \frac{54}{150} \quad \hat{p}_2 = \frac{57}{150} \quad \hat{p}_3 = \frac{39}{150} \quad (103)$$

2. Se obtiene el valor esperado de la frecuencia de la dieta A.

$$80 \frac{54}{150} = 28,8 \quad 80 \frac{57}{150} = 30,4 \quad 80 \frac{39}{150} = 20,8 \quad (104)$$

3. Se obtiene el valor esperado de la frecuencia de la dieta B.

$$70 \frac{54}{150} = 25,2 \quad 70 \frac{57}{150} = 26,6 \quad 70 \frac{20}{150} = 18,2 \quad (105)$$

Cuadro: Valores de $(O - E)^2/E$

	Excelente	Promedio	Pobre	
Dieta A	2.335	1.347	0.156	
Dieta B	2.668	1.540	0.178	
				$8.224 = \chi^2$

$$\chi^2 = 2,335 + 1,347 + 0,156 + 2,668 + 1,540 + 0,178 = 8,224 \quad (106)$$

Valor crítico de $\chi^2 = 3 - 1$ con $\alpha = 0,05$ es 5.991, por lo tanto se rechaza la hipótesis nula y existen diferencias entre categorías.

Para $\hat{p}_1 = \frac{X}{n_1}$ y $\hat{p}_2 = \frac{Y}{n_2}$ La hipótesis nula es $H_0 : p_1 = p_2$

	Sucesos	Fallas	Muestra
Población 1	X	$n_1 - X$	n_1
Población 2	Y	$n_2 - X$	n_2

Donde

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad \hat{p} = \frac{X + Y}{n_1 + n_2} \quad (107)$$

H_0 : Cada celda de probabilidad es el producto de sus correspondientes pares de probabilidad marginal

	Excelente	Promedio	Pobre	Probabilidad marginal
Dieta A	p_{A1}	p_{A2}	p_{A3}	P_A
Dieta B	p_{B1}	p_{B2}	p_{B3}	P_B
Probabilidad marginal	p_1	p_2	p_3	1

Usando el mismo proceso para determinar la diferencia de medias, podemos rechazar la H_0 y por lo tanto las probabilidades no son independientes.

Análisis de Varianza es un modelo que permite percibir diferencias sistemáticas de las medias k de una o varias poblaciones dependiendo del tratamiento que tienen sus observaciones. Para comparar estas medias existen dos recursos

- ▶ Diferencia entre medias o tratamientos
- ▶ Diferencia dentro de la variación de la población (error)

Para esto el experimento debe ser aleatoria y los resultados deben ser independientes.

Para un número de tratamientos k se obtiene una media \bar{y} y una media por cada tratamiento, así como la suma de cuadrados que es la suma cuadrada de las diferencias de las observaciones con respecto a la media de cada tratamiento. [Johnson and Bhattacharyya, 2019]

	Observaciones	Media	Suma de Cuadrados
Tratamiento 1	$y_{11}, y_{12}, \dots, y_{1n_1}$	\bar{y}_1	$\sum_{j=1}^{n_1} (y_{1j} - \bar{y}_1)^2$
Tratamiento 2	$y_{21}, y_{22}, \dots, y_{2n_2}$	\bar{y}_2	$\sum_{j=1}^{n_2} (y_{2j} - \bar{y}_2)^2$
	\vdots	\vdots	\vdots
Tratamiento k	$y_{k1}, y_{k2}, \dots, y_{kn_k}$	\bar{y}_k	$\sum_{j=1}^{n_k} (y_{kj} - \bar{y}_k)^2$
Gran Media \bar{y}	$\frac{\sum y_i}{n_1 + n_2 + \dots + n_k}$	$\frac{n_1 \bar{y}_1 + \dots + n_k \bar{y}_k}{n_1 + n_2 + \dots + n_k}$	

Descomposición de las observaciones es igual a la gran media más las desviaciones con respecto a los tratamientos más el residual.

$$y_{ij} = \bar{y} + (\bar{y}_i - \bar{y}) + (y_{ij} - \bar{y}_i) \quad (108)$$

- Suma de los cuadrados del tratamiento

$$SSt = \sum_i^k n_i (\bar{y}_i - \bar{y})^2$$

- Suma de los cuadrados del error $SSE = \sum_i^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$

- Suma total de los cuadrados $SST = \sum_i^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2$

Descomposición de las observaciones es igual a la gran media más las desviaciones con respecto a los tratamientos más el residual.

$$\sum_i^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2 = \sum_i^k n_i (\bar{y}_i - \bar{y})^2 + \sum_i^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 \quad (109)$$

donde los grados de libertad para SST: $df = \sum_{i=1}^k n_i - 1$, para SSt: $df = k - 1$ y para SSE: $df = \sum_{i=1}^k n_i - k$.

$$\text{Media de Cuadrados} = \frac{\text{Suma de Cuadrados}}{df} \quad (110)$$

La hipótesis nula es; $H_0 : \mu_1 = \mu_2 = \dots = \mu_k$

	Suma de Cuadrados	df	Media de Cuadrados
Tratamiento	SSt	$k - 1$	MSt
Error	SSE	$\sum_{i=1}^k n_i - k$	MSE
Total	SST	$\sum_{i=1}^k n_i - 1$	

El estadístico de prueba es:

$$F = \frac{\text{Between treatments}}{\text{within population}} = \frac{MSt}{MSE} \quad (111)$$

El valor crítico c se obtiene de la distribución F con $df = (k - 1, n - k)$. Si El estadístico de prueba es mayor a c entonces se rechaza H_0 .

Se compararán cuatro modelos de auriculares de gama alta para la reproducción de sonido. Hay 5 juegos de audífonos A, 4 de audífonos B, 7 de audífonos C y 6 de audífonos D disponibles para probar.[Johnson and Bhattacharyya, 2019]

	Observaciones	Media	Suma de Cuadrados
A	10,15,8,12,15	12	38
B	14,18,21,15	17	30
C	17,16,14,15,17,15,18	16	12
D	12,15,17,15,16,15	15	14
Gran Media \bar{y}		15	

$$y_{ij} = \bar{y} + (\bar{y}_i - \bar{y}) + (y_{ij} - \bar{y}_i)$$

$$\begin{bmatrix} 10 & 15 & 8 & 12 & 15 \\ 14 & 18 & 21 & 15 \\ 17 & 16 & 14 & 15 & 17 & 15 & 18 \\ 12 & 15 & 17 & 15 & 16 & 15 \\ 15 & 15 & 15 & 15 & 15 \\ 15 & 15 & 15 & 15 \\ 15 & 15 & 15 & 15 & 15 & 15 & 15 \\ 15 & 15 & 15 & 15 & 15 & 15 \end{bmatrix} = \\
 + \\
 \begin{bmatrix} -3 & -3 & -3 & -3 & -3 \\ 2 & 2 & 2 & 2 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ -2 & 3 & -4 & 0 & -3 \\ -3 & 1 & 4 & -2 \\ 1 & 0 & -2 & -1 & 1 & -1 & 2 \\ -3 & 0 & 2 & 0 & 1 & 0 \end{bmatrix} +$$

$$SSt = \sum_i^k n_i (\bar{y}_i - \bar{y})^2 = 5(-3)^2 + 4(2)^2 + 7(1)^2 + 6(0)^2 = 68$$

$$SSE = \sum_i^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 = (-2)^2 + 3^2 + (-4)^2 + \dots + 1^2 + 0^2 = 94$$

Usando la matriz de observaciones menos la gran media al cuadrado y sumando se obtiene

$$SST = \sum_i^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2 = (-5)^2 + 0^2 + (-7)^2 + \dots + 0^2 = 162$$

Para $H_0 : \mu_A = \mu_B = \mu_C = \mu_D$ entonces

	Suma de Cuadrados	df	Media de Cuadrados
Tratamiento	68	3	22.666
Error	94	18	5.222
Total	162	21	

El estadístico de prueba es $F = \frac{22,66}{5,22} = 4,34$ mientras que con un nivel de significancia $\alpha = 0,05$ el valor crítico es $F(4 - 1, 22 - 4) = 3,16$ lo que significa que se rechaza la hipótesis nula con un p value de 0.018 y se tiene por lo tanto una diferencia sistemática en el promedio de la calidad del sonido.

Un modelo de regresión lineal estudia la relación entre dos o más variables. Aunque tiene algunas limitaciones la simplicidad de su implementación lo convierte en uno de los más utilizados para cualquier tipo de investigación

- ▶ Parámetros lineales.
- ▶ Muestra Aleatoria
- ▶ No hay perfecta colineldidad
- ▶ El valor esperado de los errores es cero y es indepnediente
 $E[\epsilon] = E[\epsilon | X] = E[x\epsilon] = Cov(x, \epsilon) = 0$
- ▶ El error es homocedástico. σ^2

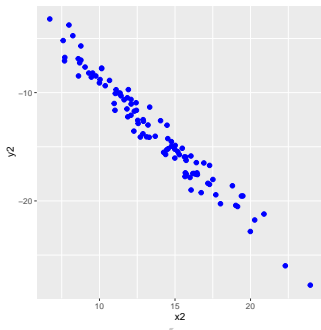
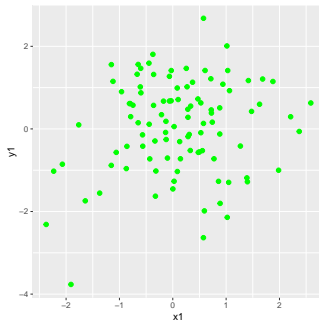
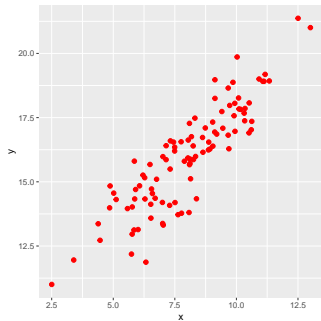
Partiendo del siguiente modelo poblacional:

$$Y = \beta_0 + \beta_1 X + \epsilon \quad (113)$$

Se desea saber la magnitud en la que X la variable independiente afecta en promedio a Y la variable dependiente. En donde ϵ representa lo que X no puede explicar y que también afecta a Y . [Wooldridge, 2016]

El coeficiente β_0 mide en promedio cuanto se puede mover Y cuando X o ϵ no tienen ningún cambio.

El coeficiente β_1 mide en promedio cuanto se puede mover Y cuando existe una variación de X .



El experimento se realiza con una muestra y por lo tanto el objetivo es representar el modelo poblacional de la ecuación 113, demostrando también que hay una relación entre X y Y .

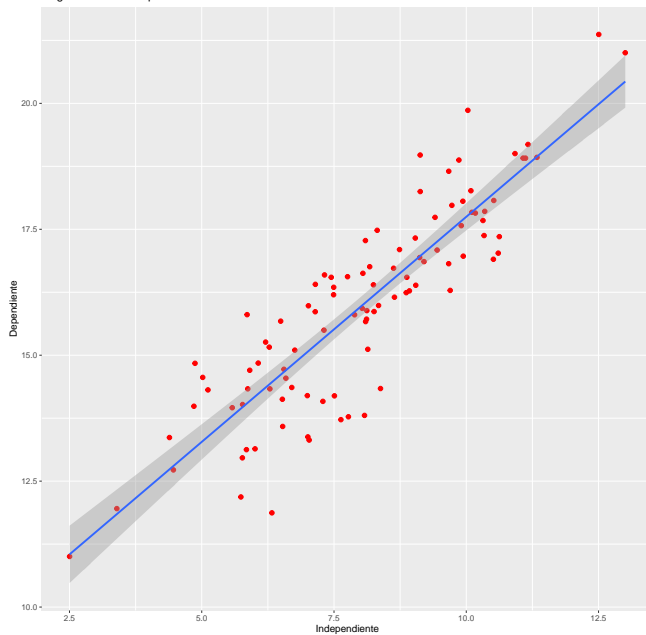
$$y = \hat{\beta}_0 + \hat{\beta}_1 x + \epsilon \quad (114)$$

El proposito es encontrar en promedio cuanto x puede afectar a y através de los estimadores $\hat{\beta}_0$ y $\hat{\beta}_1$.

$$E[Y | X] = \beta_0 + \beta_1 X \quad (115)$$

Para logralo se busca disminuir el error ϵ .

Regresión lineal simple



Usando el supuesto de $E[\epsilon] = 0$ y despejando la ecuación 113 se puede estimar β_0 .

$$E[y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i] = 0 \quad (116)$$

$$n^{-1} \sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \quad (117)$$

$$\beta_0 = \bar{Y} - \beta_1 \bar{X} \quad (118)$$

Usando el supuesto de $E[\epsilon x_i] = 0$ y despejando la ecuación 113 se puede estimar β_1 .

$$E[x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)] = 0 \quad (119)$$

$$n^{-1} \sum x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \quad (120)$$

$$\sum x_i(y_i - \bar{y}) = \hat{\beta}_1 \sum x_i(x_i - \bar{x}) \quad (121)$$

$$\hat{\beta}_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \quad (122)$$

$$f(y_i, x_i, \sigma, \beta_0, \beta_1) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{(y_i - \beta_0 - \hat{\beta}_1 x_i)^2}{2\sigma^2}} \quad (123)$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n e^{\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \beta_0 - \hat{\beta}_1 x_i)^2} \quad (124)$$

$$\log L(y_i, x_i, \sigma, \beta_0, \beta_1) = \log \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n e^{\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \beta_0 - \hat{\beta}_1 x_i)^2} \quad (125)$$

$$= n \log \frac{1}{\sigma\sqrt{2\pi}} - \frac{1}{2\sigma^2} \sum (y_i - \beta_0 - \hat{\beta}_1 x_i)^2 \quad (126)$$

$$= -\frac{n}{2} \log(2\pi) - n \log \sigma \quad (127)$$

$$- \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \beta_0 - \hat{\beta}_1 x_i)^2$$

Derivamos $f(y_i, x_i, \sigma, \beta_0, \beta_1)$

$$\frac{\partial \mathcal{L}}{\partial \beta_0} = -\frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \beta_0 - \hat{\beta}_1 x_i) = 0 \quad (129)$$

$$\frac{\partial \mathcal{L}}{\partial \beta_1} = -\frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \beta_0 - \hat{\beta}_1 x_i) x_i = 0 \quad (130)$$

$$\frac{\partial \mathcal{L}}{\partial \sigma^2} = -\frac{n}{\sigma^2} + \frac{1}{2\sigma^2} \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \beta_0 - \hat{\beta}_1 x_i)^2 = 0 \quad (131)$$

$$(132)$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \quad \hat{\beta}_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \quad (133)$$

$$S^2 = \frac{n-2}{n} S^2$$

-  Johnson, R. A. and Bhattacharyya, G. K. (2019).
Statistics: principles and methods.
John Wiley & Sons.
-  Llinás Solano, H. (2015).
Introducción a la teoría de la probabilidad.
Universidad del Norte.
-  Llinás Solano, H. (2016).
Introducción a la estadística matemática.
Universidad del Norte.
-  Mardones, J. M. (2007).
Filosofía de las ciencias humanas y sociales: materiales para
una fundamentación científica.
Filosofía de las ciencias humanas y sociales, pages 1–416.
-  Peña, D. (2014).
Fundamentos de estadística.
Alianza editorial.
-  Wackerly, D. D., Muñoz, R., Humbertotr, J., et al.



Wooldridge, J. M. (2016).

Introductory econometrics: A modern approach.

Nelson Education.

Andrés Martínez