Inferencia Estadística

Manuel Gijón Agudo

Octubre 2018 -

Índice

1. Tema 0: Introducción a la inferencia estadística

2. Tema 1: Muestreo

2.1. Definiciones

Definiciones:

- Denominamos **población** al conjunto que presenta la característica que estamos interesados en estudiar.
- **Muestra** es un subconjunto de la población. La intención al tomarlo es que sea **representativo**, esto es que cada individuo sea elegido de manera aleatorio, todo stienen las mismas probabilidades de serlo. Cada subconjunto de k individuos debe tener las mismas probabilidades de ser elegido que cualquier otro conjunto de k individuos. A esta técnica y proceso se le denomina **Muestreo aleatorio**.
- **Muestreo**: proceso por el que tomamos una muestra.

Algunas razones para realizar muestreo podían ser las siguientes:

- Económicas.
- Temporales.
- De destrucción de la muestra tras su análisis.

Entre los métodos de muestre se encuentran los siguientes:

- **Muestreo aleatorio simple**:
- **Muestreo sistemático**:
- **Muestreo estratificado**:
- **Muestreo de clústering**:
- **'Quata sampling'**
- **'Panel sampling'**

Definición: Decimos que una muestra es **aleatoria simple** cuando cumple lo siguiente:

- Cada elemento de la población y todos los posibles subconjuntos de la población tienen la misma probabilidad de ser elegidos. Esto nos asegura la **representatividad**.
- Seleccionar un elemento no condiciona el seleccionar otro. En esto consiste la **independencia**.

Una muestra aleatoria simple $X_1, X_2, ..., X_n$ es una colección de n variables aleatorias tales que:

- Son independientes.
- Siguen la misma distribución de probabilidad.

Obs: las variables aleatorias que conforman una muestra aleatoria simple son idénticas e igualmente distribuidas (iid).

Definición: el conjunto de n observaciones $(x_1, ..., x_n)$ provenientes de $(X_1, ..., X_n)$ se denomina **realización muestral**.

Definición: la **distribución conjunta** de una muestra aleatoria viene dada por la siguiente función de densidad:

$$l(\theta; x_1, ..., x_n) = \prod_{i=1}^{n} f(x_i|\theta)$$

que se denomina **función de densidad conjunta (likelihood function)**. Usaremos este término tanto en el caso discreto como en el continuo.

Ejemplo: $X \sim N(\mu, \sigma)$

$$f_X(x|\mu, \sigma) = \frac{1}{\sqrt[2]{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

Conocemos la distribución de X pero no los parámetros $\theta = (\mu, \sigma)$. La función de distribución para n variables aleatorias iid será:

$$\begin{aligned} l(\mu, \sigma; x_1, ..., x_n) &= \prod_{i=1}^{n} f(x_i|\mu, \sigma) \\ &= \frac{1}{(2\pi\sigma^2)(n/2)} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2} \end{aligned}$$

Definición: denominamos a una función T que solamente depende de los valores de una muestra aleatoria $X_1, ..., X_n$ un **estadístico**. Destacar que solamente depende de los valores observados pero no de los parámetros que determinan la variable aleatoria que los ha generado. Por supuesto un estadístico es también una variable aleatoria.

Definición: Los estadísticos que utilizamos para estimar el valor de la variable θ son denominados **estimadores**.

Definición: La distribución que sigue $Y = T(X_1, ..., T_n)$ es denominada **distribución muestral**.

2.2. Métodos de muestreo

We usually work with samples due mainly to causes:

- Economics
- Time. If we work with a large population we need to much time to analyze and maybe the characteristic to understand may change in this period
- Destruction. When we measure the characteristic we use a destructive trial.

Algunos métodos de muestreo

- **Simple random sample**: all such subsets of the frame are given an equal probability. Furthermore, any given pair of elements has the same chance of selection as any other such pair (and similarly for triples, and so on). This minimises bias and simplifies analysis of results. It will be the equivalent to the extraction of n balls with replacement from an urn. If we consider a large population, replacement "do not matter". If the population is finite and $n/N > 0{,}1$ and we don't do replacement we have to apply some corrections in the inference.
- **Systematic sampling**: Systematic sampling relies on arranging the study population according to some ordering scheme and then selecting elements at regular intervals through that ordered list. Systematic sampling involves a random start and then proceeds with the selection of every kth element from then onwards.
- **Stratified sampling**: Where the population embraces a number of distinct categories (stratas), the frame can be organized by these categories into separate "strata."Each stratum is then sampled as an independent sub-population, out of which individual elements can be randomly selected. There are several potential benefits to stratified sampling. The sample size of each strata can be assigned by some methods: proportional, optimum, ...
- **Cluster sampling**: Sometimes it is more cost-effective to select individuals in groups or clusters. Sampling is often clustered by geography, or by time periods. For instance, if surveying households within a city, we might choose to select 100 neighborhoods and then interview every household within the selected blocks.
- **Quota sampling**
- **Panel sampling**

2.2.1. Muestreo aleatorio simple

sample is a simple random sample when:

- Each population item, and all the possible subsets of population items, has the same probability to be selected. That assures REPRESENTATIVITY.
- The selection of an individual should not influence the selection of another. This assures the INDEPENDENCE.

A simple random sample of size n, $X_1, X_2, ..., X_n$, is a collection of n random variables

- **Independientes**
- **Con la misma distribución de probabilidad**

The random variables from a S.R.S, $(X_1, X_2, ..., X_n)$, are independent and identically distributed (i.i.d).

The set $(x_1, x_2...., x_n)$ of specific observations of $(X_1, X_2, ..., X_n)$ are known as realisation of the sample (data).

The joint distribution of a simple random sample of X is given by the density function.

$$l(\theta; x_1, \cdots, x_n) = f_{\overline{X}}(\overline{x}; \theta) := \prod_{i=1}^{n} f(x_i|\theta)$$

This function is called likelihood function of the sample. We will use this term both in continuous situation and in discrete situation.

Ejemplo: $X \sim N(\mu, \sigma^2)$

$$f_X(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

La distribución de X es conocida, no así los parámetros μ y σ, por tanto tendremos $\theta = (\mu, \sigma)$. La función de densidad conjunta de n variables i.i.d. es:

$$\begin{aligned} l(\mu,\sigma;x_1,x_2,\cdots,x_n) &= f_{\overline{X}}(\overline{x};\mu,\sigma) \\ &= \prod_{i=1}^{n} f(x_i|\mu,\sigma) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2} \end{aligned}$$

2.3. Distribuciones de muestreo

The functions, T, that only depends on the values of a simple random sample $X_1, X_2, ..., X_n$ are known as STATISTICS. It depends on the observed values but not on the unknown parameters that determine the distribution of X_i . A statistic is also a random variable.

Estimador: When a statistic is used in order to estimate a parameter θ we name it estimator of θ.

Ejemplos:

- Estimador de la media muestral:
$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$
- Estimador de la varianza poblacional:
$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$$

In Statistics Inference we will work with estimators. We want these estimators sufficients. A Statistic is sufficient if all the information contained in the sample is catched in that. In chapter 3 we will see other interesting properties of estimators.

$(X_1, X_2, ..., X_n)$ is a random variable, then $Y = T(X1, X2, ..., Xn)$ will be also a random variable. The distribution of Y is called Sampling distribution.

To derive the characteristics of the different estimators we need to derive the distribution of the different statistics. The sampling distribution depends on the statistic $Y = T(X_1, X_2, ..., X_n)$ and also on the statistical model of the study variable X.

In some cases we can deduce the exact distribution. In some others, we can find a distribution when $n \longrightarrow \infty$.

Ejemplos: Lanzamos una moneda equilibrada $P(\text{Cara}) = P(\text{Cruz}) = 0{,}5$, tenemos la muestra $(x_1, ..., x_n)$ y sabemos que el lanzamiento sigue la siguiente distribución $X \sim B(p = 0{,}5)$.

Trabajaremos cone el siguiente estadístico para contestar la siguiente pregunta: ¿Cuál es la probabilidad de obtener un número de caras entre 490 y 510?

$$Y = T(X_1, ..., X_n) = \sum_{i=1}^{n} x_i$$

- cuya distribución exacta conocemos:

$$Y \sim Bin(n = 1000, p = 0 - 5)$$

  entonces: $P(490 < Y < 510) = 1 - P(Y > 510) = 0{,}4729$

- como el tamaño muestral es grande, también podemos utilizar una distribución asintótica:

$$Y \sim N(\mu = np = 500, \sigma^2 = npq = 250)$$

  entonces: $P(490 < Y < 510) = 1 - P(Y > 510) = 0{,}4729$

2.3.1. Distribución del estimador media muestral

Sea $X_1, ..., X_n$ una serie de variables aleatorias idénticamente distribuídas que comparten $E(X) = \mu$ y $0 < Var(X) = \sigma^2 < \infty$. La media muestral se define como:

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

- $E(\overline{X}) = E\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \frac{\sum_{i=1}^{n} E(X_i)}{n} = \frac{n\mu}{n} = \mu$
- $Var(\overline{X}) = Var\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \frac{Var(\sum_{i=1}^{n} X_i)}{n^2} = \frac{\sum_{i=1}^{n} Var(X_i)}{n^2} = \frac{\sigma^2}{n}$
- La distribución de $\overline{X}$ es más simétrica y tiene menos dispersión que las distribuciones de X_i.
- Cuando n es grande, $n > 30$, la distribución de $\overline{X}$ se aproxima a la de una distribución normal con media μ y varianza $\frac{\sigma^2}{n}$ (**Teorema central del límite**).
- Si X sigue una distribución normal $\Rightarrow$ la distribución exacta de $\overline{X}$ es una normal (la combinación lineal de distribuciones normales es una normal).
- Cuando n es pequeño la distribución exacta de $\overline{X}$ depende de las distribuciones de los X_i.

2.3.2. Distribución del estimador varianza muestral

Sea $X_1, ..., X_n$ una serie de variables aleatorias idénticamente distribuídas que comparten $E(X) = \mu$ y $0 < Var(X) = \sigma^2 < \infty$. La varianza muestral se define como:

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

$$\begin{aligned}
(n-1)S^2 &= \sum_{i=1}^{n}(X_i - \overline{X})^2 = \sum_{i=1}^{n}(X_i - \mu + \mu - \overline{X})^2 \\
&= \sum_{i=1}^{n}(X_i - \mu)^2 + n(\mu - \overline{X})^2 + 2(X_i - \mu)n(\mu - \overline{X}) \\
&= \sum_{i=1}^{n}(X_i - \mu)^2 - n(\mu - \overline{X})^2
\end{aligned}$$

Luego la variabilidad en los datos se divide entre la variabilidad debida a la media muestral y la debida a la variabilidad entre la media muestral y la media de la población.

$$\begin{aligned}
E(S^2) &= E\left(\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}\right) \\
&= \frac{1}{n-1}E\left(\sum_{i=1}^{n}(X_i - \mu)^2 - n(\mu - \overline{X})^2\right) \\
&= \frac{1}{n-1}\left(\sum_{i=1}^{n}\sigma^2 - nE\Big(\big(\overline{X} - \mu\big)^2\Big)\right) \\
&= \frac{1}{n-1}\Big(n\sigma^2 - \frac{n\sigma^2}{n}\Big) = \sigma^2
\end{aligned}$$

$$\begin{aligned}
Var(S^2) &= HACERMUHAHAHAHAHAHAHAHA \\
&= \frac{2\sigma^4}{n-1}
\end{aligned}$$

Observación: La distribución de S^2 no es simétrica y su forma depende tanto de n como de las distribuciones de los X_i. No puede ser aproximada por una distribución normal.

2.3.3. Muestreo de distribuciones normales: χ^2

Sean $X_1, X_2, ..., X_n$ variables aleatorias independientes tales que $X_i \sim N(0,1)$.

La distribución **Chi-cuadrado con ν grados de libertad** se define como:

$$\boxed{\sum_{i=1}^{\nu} X_i = Y \sim \chi^2_\nu}$$

2.3.4. Muestreo de distribuciones normales: T-Student

Sean Z e Y variables aleatorias independientes tales que $Z \sim N(0,1)$ e $Y \sim \chi^2_\nu$.

La distribución **t-Student con ν grados de libertad** se define como:

$$\boxed{t = \frac{Z}{\sqrt{Y/\nu}}}$$

2.3.5. Muestreo de distribuciones normales: F de Fisher

Sean U y V dos variables aleatorias independientes tales que $U \sim \chi^2_m$ y $V \sim \chi^2_n$.

La distribución **F-Fisher con m y n grados de libertad** se define como:

$$\boxed{F = \frac{U/m}{V/n}}$$

2.3.6. Teorema de Fisher y sus consecuencias

Tenemos $X_1, X_2, ..., X_n$ variables aleatorias independientes que siguen una distribución $X_i \sim N(\mu, \sigma^2)$. Sabemos lo siguiente:

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

Por el Teorema de Fisher:

- $\overline{X}$ y S^2 son **independientes**, así como $\overline{X}$ y $\widehat{S}^2$.
- Si σ^2 **es conocido**, entonces el siguiente estadístico cumple:
$$\frac{(n-1)S^2}{\sigma^2} = \frac{n\widehat{S}^2}{\sigma^2} \sim \chi^2_{n-1}$$
- Si μ **es conocido**, entonces el siguiente estadístico cumple:
$$\sqrt{n-1}\frac{\overline{X}-\mu}{S} = \sqrt{n}\frac{\overline{X}-\mu}{\widehat{S}} \sim t_{n-1}$$
- Dadas dos muestras aleatorias simples, $X_1, ..., X_{n_1} \underbrace{\sim}_{iid} N(\mu_1, \sigma_1)$ y $Y_1, ..., Y_{n_2} \underbrace{\sim}_{iid} N(\mu_2, \sigma_2)$ de las que conocemos σ_1 y σ_2 y son independientes, el estadístico siguiente:
$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{n_1-1, n_2-1}$$
En particular si $\sigma_1^2 = \sigma_2^2 \Rightarrow \frac{S_1^2}{S_2^2} \sim F_{n_1-1,n_2-1}$

- Dadas dos muestras aleatorias simples, $X_1, ..., X_{n_1} \underbrace{\sim}_{iid} N(\mu_1, \sigma_1)$ y $Y_1, ..., Y_{n_2} \underbrace{\sim}_{iid} N(\mu_2, \sigma_2)$ de las que conocemos μ_1, μ_2, σ_1 y σ_2 y son independientes, el estadístico siguiente:

$$\frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\frac{(n_1-1)S_1^2}{\sigma_1^2} + \frac{(n_2-1)S_2^2}{\sigma_2^2}} \sqrt{\frac{n_1 + n_2 - 2}{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \sim t_{n_1+n_2-2}$$

3. Tema 2: Estimación de parámetros

3.1. Definiciones y propiedades de los estimadores

Definiciones:

- Sean $X_1, ..., X_n$ una secuencia de variables aleatorias independientes idénticamente distribuidas tales que $X \sim f(x; \theta)$ $\theta \in \Theta$.
- Definimos la **estimación puntual** el parámetro θ como el proceso de seleccionar un estadístico[1] T que mejor estima el valor del parámetro para esa población.
- Llamaremos a este estadístico $T = T(X_1, ..., X_n)$ que utilizamos para estimar θ un **estimador**.

Observaciones:

- Los estimadores son variables aleatorias.
- Usaremos sus propiedades estadísticas para estudiar su calidad y comparar entre ellos varios estimadores.
- Siempre tendremos un error en la estimación, nuestro objetivo será minimizarlo.

Definición: Decimos que un estimador $T_n = T(X_1, ..., X_n)$ para el parámetro θ es **consistente** cuando $\forall \epsilon > 0$:

$$\lim_{n\to\infty} P(|T_n - \theta| \geq \epsilon) = 0$$

Ejemplo: Consistencia de la media aritmética $\overline{X}_n$ como estimador

- Directamente:
  Recordemos primero $P_X(x) = P(X \leq x)$ y observemos lo siguiente:

$$X \sim N(\mu, \sigma) \Rightarrow \overline{X}_n \sim N\Big(\mu, \frac{\sigma}{\sqrt{n}}\Big)$$

  Si normalizamos tenemos $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$

$$\Rightarrow F_X(a) = P(X \leq a) = P\Big(Z \leq \frac{a-\mu}{\sigma}\Big) = P_Z\Big(\frac{a-\mu}{\sigma}\Big)$$

$$\begin{aligned} P(|\overline{X}_n - \mu| \geq \epsilon) &= P(\overline{X}_n \geq \epsilon + \mu) + P(\overline{X}_n \leq \mu - \epsilon) \\ &= 1 - F_{\overline{X}_n}(\mu + \epsilon) + F_{\overline{X}_n}(\mu - \epsilon) \\ &= 1 - F_Z\left(\frac{\mu + \epsilon - \mu}{\sigma/\sqrt{n}}\right) + F_Z\left(\frac{\mu - \epsilon - \mu}{\sigma/\sqrt{n}}\right) \\ &= 1 - F_Z\Big(\frac{\epsilon\sqrt{n}}{\sigma}\Big) + F_Z\Big(\frac{-\epsilon\sqrt{n}}{\sigma}\Big) \end{aligned}$$

  Sabemos dos cosas que utilizaremos ahora:

[1]**Estadístico:** es una función medible que tiene como espacio de salida $(X_1, ..., X_n)$ una muestra estadística de valores.

- $\lim_{x\to-\infty} F_X(x) = 0$
- $\lim_{x\to+\infty} F_X(x) = 1$

Así concluímos:

$$\lim_{x\to+\infty} 1 - F_Z\Big(\frac{\epsilon\sqrt{n}}{\sigma}\Big) + F_Z\Big(\frac{-\epsilon\sqrt{n}}{\sigma}\Big) = 0$$

- Utilizando Chebychev's:

  Primero enunciamos el resultado:

$$\boxed{P\Big(|X - E(X)| \geq \epsilon\Big) \leq \frac{\sigma^2}{\epsilon^2}}$$

Sabemos que:

- $E(\overline{X}_n) = \mu$
- $Var(\overline{X}_n) = \frac{\sigma^2}{n}$

Ahora aplicamos directamente el resultado:

$$\lim_{n\to\infty} P\Big(|\overline{X}_n - \mu| \geq \epsilon\Big) \leq \frac{\sigma^2}{n\epsilon^2} \longrightarrow 0$$

Observación: Dependiendo del libro o el material de consulta que estemos utlizando, puede darse confisuón, como norma general en este texto utilizaremos:

- $$S^2 = \frac{\sum(x_i - \overline{X}_n)^2}{n}$$
- $$\widehat{S}^2 = \frac{\sum(x_i - \overline{X}_n)^2}{n-1}$$

También podríamos encontrarnos lo siguiente:

- Cuasivarianza
$$S^* = \frac{\sum(x_i - \overline{X}_n)^2}{n}$$
- $$S^2 = \frac{\sum(x_i - \overline{X}_n)^2}{n-1}$$

Ejemplo: Tengamos $X \sim N(\mu, \sigma)$, calculemos is el estimador S^2 es insesgado o no. Hagamos tres observaciones primero, en la primera la normalidad y la definición teórica de la varianza serán las claves:

- $$\begin{aligned} Var(X) &= E(X^2) - \Big(E(X)\Big)^2 \\ \Rightarrow E(X^2) &= Var(X) + \Big(E(X)\Big)^2 \\ &= \mu^2 + \sigma^2 \end{aligned}$$

- $2\overline{X}_n \sum x_i = 2\overline{X}_n \cdot n\overline{X}_n$
- $E(\overline{X}_n^2) = \frac{\sigma^2}{n} + \mu^2$

$$\begin{aligned} E(S^2) &= E\left(\frac{\sum x_i - 2x_i\overline{X}_n + \overline{X}_n}{n}\right) \\ &= \frac{nE(x^2) - 2nE(\overline{X}_n^2) + nE(\overline{X}_n^2)}{n} \\ &= \frac{n(\sigma^2 + \mu^2) - n\left(\frac{\sigma^2}{n} + \mu^2\right)}{n} \\ &= \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n}\sigma^2 \end{aligned}$$

Luego es un estimador sesgado, sin embargo el siguiente no lo será:

$$\frac{n}{n-1}S^2$$

$$E\left(\frac{n}{n-1}S^2\right) = \sigma^2 = \widehat{S}^2$$

Teorema: si T_n es una secuencia de estimadores tales que $E(T_n) \longleftrightarrow \theta$ y $V(T_n) \longleftrightarrow 0$ cuando $n \to \infty$ entonces T_n es consistente para el parámetro θ.

Definiciones:

- Definimos la **desviación** de un estimador T como:

  $$bias(T) = E(T) - \theta$$

- Sea T un estimador para θ. Decimos que el estimador es **no desviado** si $\forall\theta \in \Theta$:

  $$E(T) = \theta$$

  En caso contrario decimos que es **desviado**. Es obvio que en este caso $bias(T) \neq 0$.a

Para introducir el siguiente concepto usaremos un ejemplo concreto. Sean $X_1, ..., X_n$ una muestra aleatoria de una variable tal que $E(X) = \mu$ y $V(X) = \sigma^2$. Probar que:

$$E(\overline{X_n}) = \mu$$

$$E(S^2) = \frac{n-1}{n}\sigma^2$$

ENCONTRAR ESTA MIERDA Y CONTINUAR A PARTIR DE AQUÍ, MUHAHHHAHHHAHHHAHHHA

Corrección de la desviación:

$$\widehat{S^2} = \frac{n}{n-1}S^2 \Rightarrow E(\widehat{S^2}) = \sigma^2$$

Definición: Sea T_n un estimador, decimos que es un estimador de θ asintóticamente no desviado si, para $n \to \infty$:

$$E(T_n) \to \theta$$

Sea $X \sim Unif(0,\theta)$ y sea el estimador para θ $T = \max X_1, ..., X_n = X_{(n)}$. Verificar que es no desviado,¿ es consistente?

Es fácil comprobar los siguiente:

$$\begin{aligned} F_{X_{(n)}} &= P(X_{(n)} \leq x) = P(X_1 \leq x, ..., X_n \leq x) \\ &= \prod_{i=1}^{n} P(X_i \leq x) = \frac{x^n}{\theta^n} \end{aligned}$$

Para $0 < x < \theta$:

$$f_{X_{(n)}} = n\frac{x^{n-1}}{\theta^n}$$

$$E(X_{(n)}) = \int_0^\theta xn\frac{x^{n-1}}{\theta^n}dx = \theta\frac{n}{n+1} < \theta$$

$$\begin{aligned} E(X_{(n)}^2) = \int_0^\theta x^2 n\frac{x^{n-1}}{\theta^n}dx = \theta^2\frac{n}{n+2} \\ \Rightarrow Var(X_{(n)}) = \frac{\theta^2 n}{(n+1)^2(n+2)} \end{aligned}$$

DESDE AQUÍ CONTINUAMOS EXPLICÁNDO EL RESULTADO DETALLADAMENTE

Observaciones:

- En general, si el momento poblacional k-ésimo m_k existe, entonces el momento muestral k-ésimo es no desviado para m_k.
- Si T es no desviado para θ, $g(T)$ no lo es, en general, el estimador $g(\theta)$ REPASARLO POR QUE NO ENTIENDO QUE DICE
- Los estimadores no desviados no siempre existen.
- En ocasiones el uso de estimadores no desviados puede ser absurdo.

Definición: Decimos que el estimador T_1 es más **eficiente** que el estimador T_2 (ambos no desviados) si:

$$Var(T_1) < Var(T_2)$$

Definición: Definimos la **eficiencia** del estimador T_1 relativa al edstimador T_2 (ambos no desviados) como:

$$\text{eff}(T_1|T_2) = \frac{Var(T_1)}{Var(T_2)}$$

Observemos que T_1 es más eficiente que T_2 si $eff(T_1|T_2) < 1$.

Definición: Decimos que T es el estimador de mínima varianza no desviado para θ si $E(T) = \theta$ y para cualquier otro estimador T' tal que $E(T') = \theta$ ocurre:

$$Var(T) \leq Var(T')$$

Podemos encontrar una cota inferiro para la varianza de un estimador no desviado:

Teorema, Cota de Cramer-Rao (CRB): bajo ciertas condiciones de regularidad y siendo $X_1, ..., X_n$ variables aleatorias idénticamente distribuidas que siguen la función de densidad $f(x;\theta)$, si T_n es un estimador no desviado para θ, entonces:

$$Var(T_n) \geq \frac{1}{nE\Big(\Big(\frac{d}{d\theta}\ln f(x;\theta)\Big)^2\Big)}$$

Obs: Podemos devinir la eddiciencia absoluta de un estimador no desviado T_n como:

$$eff(T_n) = \frac{CRB}{Var(T_n)}$$

Definición: Denominamos a la siguiente cantidad **información de Fisher**, $\mathcal{I}_X(\theta)$:

$$E\Big(\Big(\frac{d}{d\theta}\ln f(x;\theta)\Big)^2\Big) = -E\Big(\frac{d^2}{d\theta^2}\ln f(x;\theta)\Big)$$

Esta es una medida de la cantidad de información que la variable X contiene del parámetro θ. Cuanto mayor sea la cantidad de Fisher menor será la varianza y en consecuencia, la estimación será más precisa.

Si $\mathcal{X} = (X_1, ..., X_n)$ es una muestra aleatoria entonces $\mathcal{I}_{\mathcal{X}(\theta} = n\mathcal{I}_X(\theta)$ es la información de Fisher que la muestra aporta sobre el parámetro.

Una forma más general de límite puede ser obtenida considerando un estimador no desviado $T(\mathcal{X})$ de una función $\psi(\theta)$ del parámetro θ.

$$Var(T_n) \geq \frac{(\psi'(\theta))^2}{n\mathcal{I}_{\mathcal{X}}(\theta)}$$

3.2. Métodos para la obtención de estimadores

3.2.1. Método de los momentos

3.2.2. Método de la máxima verosimilitud

3.3. Métodos de remuestreo

4. Tema 3: Intervalos de confianza

4.1. Definiciones

4.2. Construcción de intervalos

4.2.1. Cantidades pivotales

Ejemplo: Tenemos una muestra de valores independientes $x_1, ..., x_n$ de una variable que sigue una distribución exponencial $X \sim Ex(\lambda)$.

- ¿$Q = \lambda \sum_{i=1}^{n} x_i$ es una cantidad pivotal?

  Para saberlo deberemos hallar su distribución y ver de qué parámetros depende esta.

  Recordemos primero una serie de cosas sobre la distribución exponencial:

$$X \sim Ex(\lambda) \Rightarrow \begin{cases} E(X) = \frac{1}{\lambda} \\ V(X) = \frac{1}{\lambda^2} \end{cases}$$

  A partir de las propiedades de la exponencial y de la suma de variables aleatorias independientes llegamos a la conclusión de que:

$$\lambda x \sim Ex(1)$$
$$\sum_{i=1}^{n} x_i \sim Gamma(n, \lambda)$$
$$\Rightarrow Q \sim Gamma(n, 1) \Rightarrow \quad \text{Q es una cantidad pivotal}$$

- ¿Cómo construyo la cantidad pivotal?

$$1 - \alpha = P\Big(q_1 \leq Q \leq q_2\Big)$$
$$1 - \alpha = P\left(q_1 \leq \lambda \sum_{i=1}^{n} x_i \leq q_2\right)$$
$$1 - \alpha = P\left(\frac{q_1}{\sum_{i=1}^{n} x_i} \leq \lambda \leq \frac{q_2}{\sum_{i=1}^{n} x_i}\right)$$

  INSERTAR IMÄGEN AL FINAL

4.2.2. Intevalos usuales

4.2.3. Intervalos de confianza asintónticos

4.3. Evaluación de intervalos de confianza

4.4. Determinación del tamaño muestral

5. Tema 4: Contraste de Hipótesis

5.1. Nociones fundamentales sobre el contraste de hipótesis

5.2. Test de máxima verosimilitud

$X \sim f(x, \theta) \quad \theta \in \Theta$

$$H_0 : \theta \in \Theta_0 \quad (\theta = \Theta_0)$$
$$H_1 : \theta \in \Theta_0^c \quad (\theta \neq \Theta_0)$$

Denominaremos a la hipótesis representada por H_0 **Hipótesis nula** (*null composite*) y a su complementaria representada H_1 **Hipótesis alternativa** (*alternative composite*).

Ejemplos:

- $X \sim N(\mu, \sigma)$

$$H_0 : \mu = 0 \quad (X \sim N(0, \sigma)\ \sigma > 0)$$
$$H_1 : \mu \neq 0 \quad (X \sim N(\mu, \sigma) \quad \mu \neq 0, \sigma > 0)$$

- $X \sim B(p_1),\ Y \sim B(p_2)$

$$H_0 : p_1 = p_2$$
$$H_1 : p_1 > p_2$$

Sean $X_1, ..., X_n$ variable aleatorias idénticamente distribuídas $X \sim f(x; \theta)$, la **función de verosimilitud** se define como:

$$l(\theta; X) = l(x_1, x_2, ..., x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

Sean:

- El **estimador de máxima verosimilitud para** $\theta \in \Theta$:

$$l(\widehat{\Theta}) = \sup_{\theta \in \Theta} l(\theta; X)$$

- El **estimador de máxima verosimilitud para** $\theta \in \Theta_0$:

$$l(\widehat{\Theta_0}) = \sup_{\theta \in \Theta_0} l(\theta; X)$$

Llamamos **ratio de verosimilitud** *Likelihood ratio* a la función definida de la siguiente manera:

$$\Lambda = \frac{l(\Theta_0)}{l(\Theta)}$$

Observemos que el valor de la función se encuentra entre 0 y 1:

$$0 \leq \Lambda \leq 1$$

- $\lambda \approx 1$ apoya H_0.
- $\lambda \approx 0$ apoya H_1.

Una vez hemos seleccionado el nivel de significación α, definimos la **región crítica** como:

$$W := \{(x_1, ..., x_n) | \Lambda \leq k_\alpha\}$$

donde k_α es una constante tal que:

$$P(\Lambda \leq k_0 | H_0) = \alpha \quad 0 < k_\alpha < 1$$

Importante: Para encontrar el valor de k_α tenemos que conocer la distribución de Λ o de una transformación monótona de la misma.

Ejemplo: Test para la media μ de una población normal con varianza σ conocida:

$X \sim N(\mu, \sigma_0)$

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu \neq \mu_0$$

$$l(\widehat{\Theta}) = \left(\frac{1}{\sigma_0\sqrt{2\pi}}\right)^n e^{-\frac{\sum(x_i - \overline{X})^2}{2\sigma_0^2}}$$
$$l(\widehat{\Theta_0}) = \left(\frac{1}{\sigma_0\sqrt{2\pi}}\right)^n e^{-\frac{\sum(x_i - \mu_0)^2}{2\sigma_0^2}}$$

$$\Lambda = e^{-\frac{n}{2}\frac{(\overline{X}-\mu_0)^2}{\sigma_0^2}}$$

Nuestro estadístico en este caso será:

$$Z_{exp} = \sqrt{-2\ln(\Lambda)} = \sqrt{n}\frac{(\overline{X} - \mu_0)}{\sigma_0} \sim N(0, 1)$$

Y nuestra región crítica:

$$W = \{(x_1, ..., x_n) || Z_{exp}| \geq z_{\alpha/2}\}$$

donde $\frac{\alpha}{2} = P(Z > z_{\alpha/2})$ y $Z \sim N(0,1)$.

Teorema de Wilks: Sea k el número de parámetros independientes de Θ y r el número de parámetros independientes de Θ_0. Bajo ciertas condiciones de regularidad y para muestras grandes[2], si H_0 es cierto entonces:

$$U = -2\ln(\Lambda) \longrightarrow \chi^2_{k-r}$$

Criterios de decisión:

- Si $U = -2\ln(\Lambda) < \chi^2_\alpha$ aceptamos H_0.
- Si $-2\ln(\Lambda) \geq \chi^2_\alpha$ rechazamos H_0 y aceptamos H_1.

donde $\frac{\alpha}{2} = P(\chi > \chi^2_\alpha)$, siendo $\chi \sim \chi^2_{k-r}$.

Ejemplo:

$X \sim Poiss(\lambda)$

$$H_0 : \lambda = \lambda_0$$
$$H_1 : \lambda \neq \lambda_0$$

- En Θ_0: $\lambda = \lambda_0$
- En Θ: $\widehat{\lambda} = \overline{X}$

$$\Lambda = e^{-n\lambda_0 + n\overline{X}} \left(\frac{\lambda_0}{\overline{X}}\right)^{n\overline{X}}$$

$$\Rightarrow U = -2\ln(\lambda) = -2n(\overline{X} - \lambda_0) - 2n\overline{X}\ln\left(\frac{\lambda_0}{\overline{X}}\right) \longrightarrow \chi^2_1$$

5.2.1. Estadístico de Pearson y el estadístico de máxima verosimilitud

$$P = \sum_{i=1}^{k} \frac{(n_i - np_{i_0})^2}{np_{i_0}}$$

Consideramos ahora la expansión de Taylor de la función $x\ln\left(\frac{x}{a}\right)$ alrededor del punto $x = a$:

[2] Mirar Anexo

$$x \ln\left(\frac{x}{a}\right) = (x-a) + \frac{(x-a)^2}{2a} + \frac{(x-a)^3}{6a^2} + \cdots$$

Ahora tomamos $x = n_i$ y $a = np_{i_0}$ quedándonos entonces:

$$n_i \ln\left(\frac{n_i}{np_{i_0}}\right) = (n_i - np_{i_0}) + \frac{(n_i - np_{i_0})^2}{2(np_{i_0})} + \frac{(n_i - np_{i_0})^3}{6(np_{i_0})^2} + \cdots$$

Resultando:

$$\sum_{i=1}^{k} n_i \ln\left(\frac{n_i}{np_{i_0}}\right) = \frac{1}{2}\sum_{i=1}^{n}\frac{(n_i - np_{i_0})^2}{np_{i_0}} + \cdots \backsimeq \frac{1}{2}P$$

También:

$$\Lambda \backsimeq P$$

5.3. Test no paramétricos robustos

Los test clásicos no paramétricos producen buenos resultados solamente si se cumplen las hi'pótesis que los sustentas, cosa que raramente ocurre en la vida real. Los **test paramétricos robustos** minimizan los problemas que ocurren al violar las hipótesis.

Importante:

- No usar con varianzas muy distintas.
- No usar para varianzas discretas.

- ***Rank-based non-parametric tests***: existen muchos, por el momento solo nombraremos dos:
  - El test de **Mann-Whitney-Wilcoxon** para la independencia entre muestras.
  - ***Wilcoxon signed-rank test*** para muestras pareadas.

  En ambos casos tendremos:
  - H_0: Las dos muestras vienen de poblaciones con idénticas distribuciones.
  - H_1: evidentemente, el suceso contrario.

  Observemos que el test de Mann-Whitney-Wilcoxon es más eficiente que el test T para distribuciones no normales y casi tan eficiente como este mismo para distribuciones normales. Sin embargo, estos test no son apropiados si no estamos analizando datos simples o *one-way layouts*, en ningún caso para diseños factoriales.

- ***Rank transform***: propuesto por Conover e Iman (1981), este proceso consta de dos pasos.
  - Convertimos los datos en rangos, esto es ordenarlos de menor a mayor y quedarnos con el orden de cada elemento.
  - Ahora ejecutamos un test paramétrico estándard en los datos ordenados en lugar de en los originales.

La hipóteis de igualdad de varinzas se aplica igualemnte. En muchas circunstancias es menos potente que los test clásicos o los test no paramétricos, por tanto la norma general es **no utilizarlos**.

- ***Randomization (permutation) tests***
    - Se aplican a situaciones en las que queremos comparar grupos.
    - La hipótesis nula es que no existen diferencias entre los grupos.
      TRADUCIR
    - A permutation test gives a simple way to compute the sampling distribution for any test statistic, under the null hypothesis.
    - Permutation tests exist for any test statistic, regardless of whether or not its distribution is known. Thus one is always free to choose the statistic which best discriminates.
    - If the null hypothesis is true, then any random arrangement of observations to groups is equally possible
    - The ranking of the real test statistic among the shuffled test statistics gives a p-value

Steps to perform a permutation test for the means: Let's assume that we want to compare two populations and we have a sample of size n_1 from population 1 and a sample of size n_2 from population 2.

- Calculate the difference between the mean of the two groups (D_0).
- Randomly reassign the $n_1 + n_2$ observations so that n_1 are in the first group and n_2 are in the second group and calculate the difference between the means of the two groups (D_1).
- Repeat this step a large number of times, each time calculating the (D_i) (With 1000 permutations the smallest possible p-value is 0.001)
- Calculate the proportion of all the (D_i)'s that in absolute value are greater than or equal to (D_0). This is the "P value".

Observaciones:

- Permutation test is not limited to using the difference in means as a test statistic.
- In the example above we have also included the Student's T statistic.
- We could use the difference in medians or the sum of the observations or the maximum of the sample as different test statistics.
- When the alternative hypothesis is one sided, the way to compute the p-value changes accordingly.

5.4. Test múltiple

5.5. Test clásicos no paramétricos

5.6. Test de Fisher

5.7. Determinación del tamaño muestral

6. Tema 5: Modelos de Regresión

6.1. Regresión lineal simple

El objetivo es modelar la relación entre una variable respuesta Y y una variable aleatoria explicativa X_1. Más tarde lo generalizaremos a conjuntos de variables explicativas $X_1, ..., X_{p-1}$.

En la práctica, además de la variable explicativa X_1 tendremos también otras variables explicativas $Z_1, ..., Z_r$ que nos serán desconocidas. El hecho de no tener en cuenta esas variables tendrá repercusiones sobre la bondad del modelo.

Nuestros modelos podrán ser tanto deterministas como aleatorios.

- En un **modelo determinista** la variable respuesta Y se relaciona con las variables explicativas mediante una función matemática que involucra constantes β, $f(X_1, X_2, ..., X_{p-1}|\beta)$.

  De acuerdo con estos modelos, fijados los valores de las variables explicativas podemos obtener el valor de la variable respuesta exactamente.

  - Movimiento parabólico

  $$f(t|v_0, G, \alpha) = v_0 + \operatorname{sen} t(\alpha) - \frac{1}{2} \cdot G \cdot t^2$$

  - Ley de Ohm

  $$f(I|R) = I \cdot R = Y$$

- En los **modelos estadísticos** el valor de la variable respuesta es una combinación de los valores obtenidos mediante las variables explicativas más otro término de **ruido**.

$$\boxed{Y = \text{Señal} + \text{ruido}}$$

Observación: a partir de este punto asumiremos durante todo el capítulo que solo contamos con una única variable explicativa $X_1 = X$ y que nuestro modelo es el siguiente:

$$Y_i = \beta_0 + \beta_1 \cdot X_i + \epsilon$$

donde $\beta_0, \beta_1 \in \mathbb{R}$ son desconocidos, son los parámetros a ajustar y los subíndices i hacen referencia a la observación i-ésima. Al tener una única variable el modelo se corresponde con una recta, al tener más lo hará con planos de diferentes dimensiones.

Suposiciones sobre los ϵ:

- $H_1 : E(\epsilon_i) = 0, \quad i = 1, 2, ..., n$
- $H_2 : V(\epsilon_i) = \sigma^2, \quad i = 1, 2, ..., n$
- $H_3 : \epsilon_i \sim N(0, \sigma)$
- $H_4 : \epsilon_i$ es independiente de $\epsilon_j \quad \forall i \neq j \Rightarrow cov(\epsilon_i, \epsilon_j) = corr(\epsilon_i, \epsilon_j) = 0$

Notar que σ^2 es un tercer parámetro desconocido del modelo.

Estas suposiciones son equivalentes a las siguientes:

- $H_1 : E(y_i|x_i) = \beta_0 + \beta_1 x_1$ (**Linealidad**)
- $H_2 : V(y_i|x_i) = \sigma^2$ (**Varianza constante**)
- $H_3 : y_i|x_i \approx$ Normal (**Normalidad**)
- $H_4 : y_i|x_i$ independiente de $y_j|x_j$ (**Independencia**)

- $E(\epsilon_i) = 0 \Rightarrow E(y_i|x_i) = \beta_0 + \beta_1 x_1$ (**Linealidad**) [3]

$$\begin{aligned} E(Y_i) &= E(\beta_0 + \beta_1 x_i + \epsilon_i) \\ &= E(\beta_0 + \beta_1 x_i) + E(\epsilon_i) \\ &= \beta_0 + \beta_1 x_i + 0 = \beta_0 + \beta_1 x_i \end{aligned}$$

- $V(\epsilon_i) = \sigma^2 \Rightarrow V(y_i|x_i) = \sigma^2$ (**Varianza constante**) [4]

$$\begin{aligned} V(Y_i) &= V(\beta_0 + \beta_1 x_i + \epsilon_i) \\ &= V(\beta_0 + \beta_1 x_i) + V(\epsilon_i) \\ &= 0 + V(\epsilon_i) = 0 + 0 = 0 \end{aligned}$$

- $\epsilon_i \sim N(0, \sigma) \Rightarrow y_i|x_i \approx$ Normal (**Normalidad**)
  DEMOSTREAR ESTA MERDA
- ϵ_i es independiente de $\epsilon_j \quad \forall i \neq j \Rightarrow y_i|x_i$ independiente de $y_j|x_j$ (**Independencia**)
  DEMOSTRAR ESTA MIERDA

Como veremos más adelante las dos primeras (linealidad y varianza constante) son la suposiciones con más importancia.

De nuevo otra equivalencia, esto es equivalente:

$$\boxed{Y_i|X_i \sim N(\beta_0 + \beta_1 X_1, \sigma) \ \text{ independientes}}$$

No es necesaria una normalidad global, pero sí una normalidad en los valores de y_i que comparten el mismo x_i. La consecuencia es que Y debe ser continuo (o casi continuo).

[3] Propiedades de la **esperanza**:

- $E(X + c) = E(X) + c$
- $E(X + Y) = E(X) + E(Y)$
- $E(aX) = aE(X)$
- $E(X \cdot Y) = E(X) \cdot E(Y)$ si son independientes.

[4] Propiedades de la **varianza**:

- $V(X) \geq 0$
- $V(aX + b) = a^2 V(X)$
- $V(X + Y) = V(X) + V(Y) + 2 \cdot Cov(X, Y)$
- $V(X - Y) = V(X) + V(Y) - 2 \cdot Cov(X, Y)$
- $V(Y) = E(V(Y|X)) + V(E(Y|X))$ (Varianza por Pitágoras).

6.1.1. Ajuste del modelo

Una vez tenemos las parejas de datos (x, y) queremos ajustar el modelo:

$$Y_i = b_0 + b_1 \cdot X_1$$

Observación: β_0 y β_1 son valores desconocidos pero fijos.

Resumiendo, tendremos:

- Nuestro modelo teórico:
$$Y_i = \beta_0 + \beta_1 \cdot X_1 + \epsilon_i$$
- El modelo ajustado con el que trabajaremos:
$$Y_i = b_0 + b_1 \cdot X_1 + e_i$$
- A la diferencia la llamaremos **residuo**:
$$e_i = Y_i - (b_0 + b_1 \cdot X_1)$$

Criterio de los mínimos cuadrados: el criterio que utilizaremos para hallas los valores. Nuestro siguiente problema consistirá en minimizar las siguientes expresiones:

- Minimizar: $\sum_{i=1}^{n} |e_i|$
- Minimizar: $Q(b_0, b_1) = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n}(Y_i - b_0 - b_1 \cdot X_1)^2$

Haciendo los cálculos llegamos a la conclusión de que los estimadores de mínimos cuadrados para b_0 y b_1 son:

$$\frac{\partial Q}{\partial b_0} = -2 \cdot \sum_{i=1}^{n}(Y_i - b_0 - b_1 \cdot X_1) = 0$$
$$\frac{\partial Q}{\partial b_1} = -2 \cdot \sum_{i=1}^{n}(Y_i - b_0 - b_1 \cdot X_1) \cdot X_1 = 0$$

A este conjunto de ecuaciones $Q(b_0, b_1) = SQ_R(b_0, b_1)$ se les denomina **ecuaciones normales** (*normal equations*).

Podemos reescribir ambas ecuaciones respectivamente como:

(PROBARLO QUEDA PENDIENTE)

$$\boxed{\begin{aligned} \frac{\partial Q}{\partial b_0} &= \sum_{i=1}^{n} e_i = 0 \\ \frac{\partial Q}{\partial b_1} &= \sum_{i=1}^{n} e_i \cdot X_1 = 0 \end{aligned}}$$

Las soluciones de las ecuaciones normales son las siguientes:

(PROBARLO QUEDA PENDIENTE)

$$b_0 = \frac{\sum_{i=1}^{n}(X_i - \overline{X}) \cdot (Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2} = \frac{\sum_{i=1}^{n}(X_i - \overline{X}) \cdot Y_i}{\sum_{i=1}^{n}(X_i - \overline{X})^2} = \sum_{i=1}^{n} C_i \cdot Y_i$$

$$b_1 = \overline{Y} - b_1 \cdot \overline{X} = \frac{\sum_{i=1}^{n} Y_i - b_1 \cdot \sum_{i=1}^{n} X_i}{n} = \sum_{i=1}^{n} C_i' \cdot Y_i$$

Usando este criterio siempre tendremos (en otros casos no es seguro):

- b_0 y b_1 son combinaciones lineales de y_i.
- La media muestral de los residuos es 0.
- La recta ajustada siempre pasa por el punto $(\overline{X}, \overline{Y})$ y es la siguiente:

$$\widehat{Y}_i = \overline{Y} + b_1 \cdot (X_i - \overline{X})$$

6.1.2. Tabla ANOVA

Es claro que $Y_i - \overline{Y} = Y_i - \overline{Y} + \widehat{Y}_i - \widehat{Y}_i$, observemos que $Y_i - \widehat{Y}_i$ son los residuos.

$$\sum_{i=1}^{n}(Y_i - \overline{Y}) = \sum_{i=1}^{n}(Y_i - \widehat{Y}_i) + \sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})$$

También es posible demostrar lo siguiente:

HACER LOS CÁLCULOS DETALLADAMENTE...

$$\sum_{i=1}^{n}(Y_i - \overline{Y})^2 = \sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2 + \sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})^2$$

O lo que es lo mismo:

Variabilidad total del proceso = variabilidad residual (no explicada por el modelo) + variabilidad explicada por el modelo

AQUÑI IRÁ UNA TABLA MUHAHAHAHAHA

La **Suma total de cuadrados** SQ_T solamente depende de y y en absoluto de x. Notemos también que: $SQ_T = (n-1)s_y^2$

Sin embargo, tanto SQ_E y SQ_R dependen de x_i y por tanto del modelo.

Sin no realizamos ninguna transformación en Y, cuanto mayor sea SQ_E y menor SQ_R **mejor será el modelo**.

6.1.3. Bondad del ajuste

Definimos el **coeficiente de determinación** R como:

$$R = \frac{\text{Suma de cuadrados explicados por la regresión}}{\text{Total de cuadrados}}$$

$$R = \frac{SQ_E}{SQ_T} = \frac{SQ_T - SQ_R}{SQ_T} = 1 - \frac{SQ_R}{SQ_T} = 1 - \frac{S_R^2 \cdot (n-2)}{S_Y^2 \cdot (n-1)}$$

$$0 \leq R^2 \leq 1$$

Normalmente el coeficiente de determinación se expresa en forma de porcentaje.

R^2 es el porcentaje de variabilidad de la respuesta explicado por el modelo.

Observación: Solamente en el caso del modelo lineal simple el valor de R^2 se corresponde con el cuadrado de la correlación muestral entre X e Y.[5]

La covarianza y el coeficiente de correlación muestral miden el grado de **dependencia lineal** entre dos variables. Definimos la **covarianza** como:

$$C_{xy} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{n-1}$$

El **coeficiente de correlación muestral**, que es adimensional, se expresa de la siguiente manera:

$$r_{xy} = \frac{C_{xy}}{S_x \cdot S_y} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2 \cdot \sum_{i=1}^{n}(Y_i - \overline{Y})^2}} = \frac{\sum_{i=1}^{n}(X_i - \overline{X}) \cdot Y_i}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2 \cdot \sum_{i=1}^{n}(Y_i - \overline{Y})^2}}$$

Propiedad: $-1 \leq r_{xy} \leq 1$

Observación: En el caso de la regresión lineal múltiple R^2 es el cuadrado del coeficiente de correlación entre Y y la mejor combinación lineal de todas las variables explicativas que se corresponde con el valor predicho.

Es también simple comprobar:

$$b_1 = \frac{c_{xy}}{S_x^2} = r_{xy} \cdot \left(\frac{S_y}{S_x}\right)$$

HACER, EJERCICIO

6.1.4. Distribución de los coeficientes

Hasta ahora no hemos utilizando las hipótesis del modelo, varianza constante, normalidad e independencia. Estas nos serán necesarias a la hora de encontrar las distribuciones de los valores ajustados b_0 y b_1.

[5] $R^2 = (r_{xy})^2$

Necesitaremos conocer estas distribuciones para poder hacer inferencia sobre β_0 y β_1. Construir intervalos de confianza y test de significación para β_0 y β_1 y también hacer predicciones sobre los futuros valores de y.

Para un conjunto de X_i, hemos simulado gran cantidad de muestras (X_i, Y_i) basándonos en el modelo teórico $Y_i = \beta_0 + \beta_1 \cdot X_i + \epsilon_i$ donde los parámetros β_0, β_1 y σ^2 son conocidos para el simulador.

Para cada conjunto, calculamos utilizando mínimos cuadrados los valores b_0, b_1 y S_R^2.

Para encontrar las distribuciones teóricas de b_0, b_1 y S_R^2 necesitaremos las cuatro hipótesis en las que basamosl el modelo teórico:

$$Y_i|X_i \sim N(\beta_0 + \beta_1 \cdot X_i, \sigma) \text{ independientes}$$

b_0 y b_1 son combinaciones lineales de las y_i's y si estas siguen una distribución normal entonces si el modelo es correcto las b_0 y b_1 también seguirán una distribución normal.

Si el modelo es correcto entonces:

- $E(b_0) = \beta_0$
- $V(b_0) = \frac{\sigma^2}{n} \cdot \frac{\sum_{i=1}^n X_i^2}{\sum_{i=1}^n (X_i - \overline{X})^2}$

y también:

- $E(b_1) = \beta_1$
- $V(b_1) = \frac{\sigma^2}{\sum_{i=1}^n (X_i - \overline{X})^2}$

Siendo la covarianza entre ambos:

$$cov(b_0, b_1) = \frac{-\overline{X} \cdot \sigma}{\sum_{i=1}^n (X_i - \overline{X})^2}$$

Entonces nos queda:

$$b_0 \sim N\left(\beta_0; \sqrt{\frac{\sigma^2}{n} \cdot \frac{\sum_{i=1}^n X_i^2}{\sum_{i=1}^n (X_i - \overline{X})^2}}\right)$$

$$b_1 \sim N\left(\beta_1; \sqrt{\frac{\sigma^2}{\sum_{i=1}^n (X_i - \overline{X})^2}}\right)$$

Cuanto menor sea $V(b_1)$ más cercano estará b_1 al verdadero valor de β_0.

¿Cómo minimizaríamos $V(b_1)$? RESPONDER ¿por qué la covarianza entre b_0y b_1 no es cero? RESPONDER

$V(b_0)$ y $V(b_1)$ dependen de σ^2, desconocido. Para estimar estas cantidades remplazaremso σ^2 por su estimador S_R^2.

$$S_{b_0}^2 = \frac{S_R^2}{n} \cdot \frac{\sum_{i=1}^n X_i^2}{\sum_{i=1}^n (X_i - \overline{X})^2}$$
$$S_{b_1}^2 = \frac{S_R^2}{\sum_{i=1}^n (X_i - \overline{X})^2}$$

En la tabla ANOVA podemos encontrar la varianza residual:

$$S_R^2 = \frac{\sum_{i=1}^n e_i^2}{n-2}$$

El estimador está dividido por $n-2$ en lugar de por $n-1$ o n para que sea **insesgado**:

$$E(S_R^2) = \sigma^2$$

PROBAR ESTA COSA AÑADIENDO AQUÏ LOS CÄLCULOS OPORTUNOS

6.1.5. Intervalos de confianza para los coeficientes

Usando las propiedades de la distribución normal podemos crear intervalos de confianza para β:

$$[b_i - t_{\frac{\alpha}{2}, n-2} \cdot S_{b_i}; b_i + t_{\frac{\alpha}{2}, n-2} \cdot S_{b_i}]$$

con un nivel de confianza $(1-\alpha)$. [6]

¿Podría ser $\beta_1 = 1$ en el modelo teórico?

RESOLVER

¿Por qué $cov(b_0, b_1) = \neq 0$?

RESOLVER

6.1.6. Tests de significación para los coeficientes

$$H_0 : \beta_1 = 0$$
$$H_1 : \beta_1 \neq 0$$

(de igual manera para β_0).

Si H_0 es cierta, entonces $t_1 = \frac{b_1}{S_{b_1}} \sim$ t-Student, con $\nu = n-2$.

INSERTAR IMÁGEN DE T CON DOS COLAS!!

- Si $|t_1| > 2$ y el modelo es correcto, entonces rechazamos H_0 y **no** podemos **eliminar** la variable x_1 del modelo.

[6] El intervalo de confianza contendrá el verdadero valor de β_i un $(1-\alpha)$ por ciento de las veces.

- Si $|t_1| < 2$ y el modelo es correcto, entonces no podemos rechazar H_0 y podemos **eliminar** la variable x_1 del modelo.

Con lo anterior podemos discriminar si las variables nos sirven o no para predecir el valor de Y. Usaremos este procedimiento para simplificar el modelo.

Para averiguar si β_1 pudiera ser igual a 0 o no, necesitamos saber si $|t_1| = |\frac{b_1}{S_{b_1}}|$ es grande o pequeño.

$$H_0 : \beta_1 = a$$
$$H_1 : \beta_1 \neq a$$

Si H_0 es cierta, entonces $t = \frac{b_i - a}{S_{b_i}} \sim$ t-Student, con $\gamma = n - 2$.

INSERTAR IMÁGEN DE T CON DOS COLAS!!

- Si $|t| > 2$ (el **p-valor** es más pequeño que 0,05) y el modelo es correcto entonces **rechazamos** H_0.
- Si $|t| < 2$ (el **p-valor** es mayor que 0,05) y el modelo es correcto entonces **no podemos rechazar** H_0.

6.1.7. Predicción

El **intervalo de confianza** al $(1-\alpha)$ para predecir el valor de Y en X_0, $E(Y|X_0)$, es:

$$\left[(b_0 + b_1 \cdot X_0) - t_{\frac{\alpha}{2},n-2} \cdot \sqrt{\frac{1}{n} + \frac{(X_0 - \overline{X})^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2}}; (b_0 + b_1 \cdot X_0) + t_{\frac{\alpha}{2},n-2} \cdot \sqrt{\frac{1}{n} + \frac{(X_0 - \overline{X})^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2}}\right]$$

EXPLICAR EL POR QUÉ ES ASÍ CON DETALLE

Si el valor de X_0 se encuentra dentro del rango de valores de X que hemos usado para ajustar el modelo, nos encontramos frente a una **interpolación**.

De otra manera, la predicción es una **extrapolación** y puede ser justificada **si y solo sí** asumimos que el modelo teórico es válido no solo en el rango de las variables con las que lo hemos ajustado si no más allá. **Debemos tener cuidado al asumir esto**.

6.1.8. Correlación, causalidad e interpretación

Analysing a bivariate diagram or adjusting a linear model between two variables means that they are related but doesn't mean that X is cause of Y .

You can have another predictor hidden which is the one who does changes in both X and Y .

¿Cómo podríamos saber si la relación entre dos variables es causal o no?

RESPONDER

7. Tema 6: ANOVA

7.1. Único factor

7.1.1. Introducción

Poniendo las cosas en contexto:

- **Modelo estadísticos**

$$\text{variable respuesta} = \text{modelo} + \text{error}$$

- **Modelos lineales**

$$y_i = \underbrace{\underbrace{\beta_0}_{\text{intercepto}} + \underbrace{\beta_1}_{\text{pendiente}} \cdot \underbrace{x_1}_{\text{variable predictora}} + \beta_2 \cdot x_2 + \cdots}_{\text{modelo}} + \epsilon_i$$

- **Modelos para la variables categóricas predictivas: modelo para los efectos** (*effects model*)

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$
$$y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk}$$

  donde k representa el conjunto de réplicas.

7.1.2. Estimación de los parámetros en los modelos lineales

- Los parámetros pueden ser estimados utilizando cualquiera de los métodos de estimación, entre los más usuales se encuentras **Mínimos cuadrados** (*Ordinary lest squares, OLS*) y **Máxima verosimilitud** (*maximun likelihod, ML* o también *REML).*
- Los modelos que asumen una distribución normal para los errores, y que frecuentemente utilizan OLS, son denominados de forma genérica **modelos lineales generalizados** (*'general' linear models*). Nos encontramos con varios tipos:
  - **Modelos de Regresión** (*Regression models*): variables predictivas continuas.
  - **ANOVA** (*Analysis of variance*) variables predictivas categóricas.
  - **ANCOVA** (*Analysis of covariance*): incorporan tanto variables predictivas categóricas como continuas.
- El estimador por máxima verosimilitud se utiliza en ocasiones también para los modelos lineales generalizados, luego no se restringe a modelos donde tanto los residuos como la respuesta siguen una distribución normal.

7.1.3. Factores fijos vs aleatorios

Factores fijos (*Fixed factors*): son factores cuyos niveles representan la población de interés completa.

Factores aleatorios (*Random factors*): son factores cuyos niveles han sido aleatoriamente elegidos de entre todos los posibles niveles de interés, representan muestras aleatorias de los mismos.

7.1.4. *One way ANOVA fixed factor*

Notación:

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij} \quad i = 1, ..., a \quad j = 1, ..., n_i$$

- μ: *overall mean.*
- α_i: efecto del grupo i, factor fijo.
- ϵ_{ij}: error aleatorio.

Terminología:

- **Replicación** (*Replicate*): observaciones hechas bajo las mismas condiciones experimentales.
- **Diseño balanceado** (*Balance design*): diseño experiental donde todas las células tienen el mismo número de réplicas.
- **Factor**: una fuente de variación controlada.
- **Nivel** (*Level*): cada uno de los diferentes valores de un factor.

El modelo **ANOVA de un factor fijo** nos queda:

$$\boxed{y_{ij} = \mu + \alpha_i + \epsilon_{ij} \quad i = 1, ..., a \quad j = 1, ..., n_i}$$

Donde debemos asumir:

$$\boxed{\epsilon_{ij} \ \text{ iid } \ \epsilon_{ij} \sim N(0, \sigma^2)}$$

Los **residuos** son:

- Independientes los unos de los otros.
- Normalmente distribuídos.
- Tienen la propiedad de la **homeostacididad** (tienen la misma varianza).

Además el modelo cumple:

$$E(\overline{Y}) = \frac{N\mu + \sum_{i=0}^{a} n_i \alpha_j}{N} = \mu \Rightarrow \sum_{i=0}^{a} n_i \alpha_j = 0$$

donde $N = \sum_{i=0}^{a} n_i$.

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij} = \mu + \underbrace{(\mu_i - \mu)}_{\alpha_i} + \underbrace{(y_{ij} - \mu_i)}_{\epsilon_{ij}} \Rightarrow$$

$$\underbrace{y_{ij} - \mu}_{\text{Total}} = \underbrace{(\mu_i - \mu)}_{\text{Efecto de los factores}} + \underbrace{(y_{ij} - \mu_i)}_{\text{Interacciones}}$$

Estimación OLS:

$$\epsilon_{ij} = y_{ij} - \mu - \alpha_i$$

$$||\epsilon_{ij}||^2 = ||y_{ij} - \mu - \alpha_i||^2 = \textstyle\sum_{i=1}^{a} \sum_{j=1}^{n_i} (y_{ij} - \mu - \alpha_i)^2$$

$$\frac{\partial ||\epsilon_{ij}||^2}{\partial \mu} = -2N(\overline{Y} - \mu) = 0 \Rightarrow \widehat{\mu} = \overline{Y}$$
$$\frac{\partial ||\epsilon_{ij}||^2}{\partial \alpha_i} = -2n_i(\overline{Y_{i\cdot}} - \overline{Y} - \alpha_i) = 0 \Rightarrow \widehat{\alpha}_i = \overline{Y_{i\cdot}} - \overline{Y} \Rightarrow \widehat{\mu}_i = \overline{Y_{i\cdot}}$$

7.1.5. Partición de la varianza

$$y_{ij} - \mu = (\mu_i - \mu) + (y_{ij} - \mu_i)$$

$$y_{ij} - \overline{Y} = (\overline{Y_{i\cdot}} - \overline{Y}) + (y_{ij} - \overline{Y_{i\cdot}})$$

$$\begin{aligned} \overbrace{\sum_i \sum_j (y_{ij} - \overline{Y})^2}^{\text{SST}} &= \sum_i \sum_j (\overline{Y_{i\cdot}} - \overline{Y} + y_{ij} - \overline{Y_{i\cdot}})^2 \\ &= \underbrace{\sum_i n_i (\overline{Y_{i\cdot}} - \overline{Y})^2}_{\text{SSB}} + \underbrace{\sum_i \sum_j (y_{ij} - \overline{Y_{i\cdot}})^2}_{\text{SSR}} \end{aligned}$$

7.1.6. Hipótesis nula

En un test ANOVA de un solo factor nuestra hipótesis nula es la siguiente:

$$H_0 : \alpha_1 = \alpha_2 = ... = \alpha_a = 0 \ \ \text{El efecto de cada grupo es equivalente a 0}$$

Equivalentemente:

$$H_0 : \mu_1 = \mu_2 = ... = \mu_a = \mu$$

En este punto es evidente, pero aclaramos:

$$H_1 : \exists i : \alpha_i \neq 0$$

7.1.7. Esperanza de la suma de cuadrados: SSR

$$SSR = \sum_i \sum_j (y_{ij} - \overline{Y_{i\cdot}})^2 = \sum_i (n_i - 1)\widehat{S_i}^2$$
$$E(SSR) = \sum_i (n_i - 1)\sigma^2 = (N - a)\sigma^2$$

Utilizando el hecho de que $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_a^2$

$$E\Big(\frac{SSR}{N-a}\Big) = \sigma^2$$

es siempre un **estimador insesgado de** σ^2.

Asumiendo además la **independencia de los residuos** tenemos que:

$$\frac{SSR}{\sigma^2} \sim \chi^2_{N-a}$$

7.1.8. Esperanza de la suma de cuadrados: SSB

$$SSB = \sum_i n_i(\overline{Y_{i\cdot}} - \overline{Y})^2$$

Utilizando álgebra es posible demostrar los siguiente:

$$E\Big(\frac{SSB}{a-1}\Big) = \frac{\sum_i n_i \alpha_i^2}{a-1} + \sigma^2$$

DEMOSTRARLO

Bajo H_0, $\frac{SSB}{a-1}$ es un **estimador insesgado** de σ^2 y además:

$$\frac{SSB}{\sigma^2} \sim \chi^2_{a-1}$$

y es independiente de $\frac{SSR}{\sigma^2}$.[7]

7.1.9. Test para la hipótesis nula

Recordermos que $H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_a = 0$. Si esto es cierto tendremos:

$$\frac{\frac{\frac{SSB}{\sigma^2}}{a-1}}{\frac{\frac{SSR}{\sigma^2}}{N-a}} = \frac{\frac{SSB}{a-1}}{\frac{SSR}{N-a}} \sim F_{a-1,N-a}$$

[7] Teorema de Cochran. bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla

INCLUIR QQUÍ LA TABLA ANOVA MUAHAHAHHAHHAHAH

Rechazamos la hipótesis nula sí $F \geq F_{a-1,N-a,\alpha}$.

7.2. Suposiciones sobre el modelo

Asumimos lo siguiente para los residuos:

- Normalmente distribuídos.
- Homogeneidad de varianzas.
- Independientes uno a uno.

Sobre el hecho de que estén normalmente distribuídos:

- Asumimos que los términos de error dentro de cada grupo provienen de poblaciones normalmente distribuídas ($\epsilon_{ij} \sim N(0, \sigma_\epsilon)$).
- Si los tamaños muestrales y las varianzas son similares el test ANOVA es muy robusto respecto a esta suposición.
- Debemos comprobar la presencia de outliders, simetrías y bimolaridad.
- Podemos comprobar la normalidad de varias formas:
  - *Boxplots* de las observaciones de los residuos.
  - *Probability plots* de los residuos.
  - Test de normalidad.
    - Test de Wilks Shapiro.
    - Test de bondad de ajuste.

Consideraciones sobre la homeostacidad:

- Una hipótesis muy importante es asumir que las varianzas de los términos de error (y las observaciones en las poblaciones consideradas) son aproximadamente iguales en cada grupo.
- La anterior es más importante que la normalidad: si las varianzas no son similares el resultado del test F ANOVA se verá gravemente afectado.
- Un diseño balanceado ayuda a mitigar los efectos de la heteroestacidad.
- Si contamos con tamaños muestrales similares y el ratio entre la varinaza mayor y la menor no excede $3:1$, ANOVA es razonablemente robusto en este sentido.
- Comprobando la homeodasticidad:
  - *Boxplots* de los residuos deberían tener una dispersión similar.
  - Existen tests para comprobar, como H_0, que las varianzas entre las poblaciones son las mismas entre los grupos.
    - Test de Bartlett.
    - Test de Hartley.
    - Test de Cochran.

   - Test de Levene.

Consideraciones sobre la independencia:

- Los términos de error y las observaciones deben de ser independientes.
  - Correlación positiva entre las réplicas dentro de los grupos deviene en una subestimación de la varianza real e incrementa el ratio de Errores de Tipo I.
  - Correlación negativa entre las ré‘plicas dentro de los grupos implica una sobreestimación de la varianza real e incremente el ratio de Errores de Tipo II.
- Esta hipótesis debe ser alcanzada durante las fases de diseño del experimento y recolección de datos (gran importancia de las técnicas de muestro empleadas) y no puede ser enmendada después.

7.2.1. *Multiple testing*

Rechazar la hipótesis nula indica que al menos una de las medias poblacionales difiere de las otros, lo **no nos indica** cuál es la diferente.

- *Post-hoc unplanned pairwise comparisons*: se trata de comparar sistemáticamente todas las parejas entre sí.
- *Planned comparisons*: planeamos, normalmente durante el diseño del experimento, que poblaciones comparar entre sí.

unplanned pairwise comparisons:

- Existen una gran variedad de procedimientos para controlar el ratio de Errores del Tipo I, de este modo minimizamos estos errores.
- Estos procedimientos reducen la importancia de las comparaciones por parejas (incrementando los Errores de Tipo II), lo que está directamente relacionado con el número de grupos a comparar.

  página 33**CONTINUAR DESDE AQUÍ!!!!**

7.2.2. ANOVA de un factor aleatorio

$$\mu + \alpha_i = \mu_i$$

Reagrupando parámetros muhahahahhhahhha

Notas sobre la probabilidad de equivocarse o acertar enc omparaciones 2 a 2

1 - 0.95*0.95 más o menos 0.1 (probabilidad de equivocarse en al menos 1 de ellos)

Si F ¿1 podemos rechazar la hipótesis (MIRAR TODA ESTA MIERDA BIEN FUERTE)

alphas = media de cada tratamiento - media global

E(SSR /N-a) = sigma?2 siempre se cunmple muhehehehehhe (REPSAR EST AMIERDA GORDA)

BLA BLA BLA BLA BLA BLA BLA BLA BLA BLA BLA BLA BLA BLA BLA BLA

La normalidad la comprobamos con los residuos, si unsamos la variable original (la de densidad conjunta) podemos detectar no normalidad cuando la hay por que las mus sean distintas.

SI P ¡ALPHA RECHAZAMOS LA HIPÓTESIS NULA

Multiple testing = comparacinoes múltiples muhahahahahahahahahahahahahahahaha

CASO RANDOM Observemos que lo que comparamos es si los hospitales añaden variabilidad o nop

La table y lo demás no varían

Luego deberemso **estimar las componentes de la varianza**

intraclass correlation = correlación intraclásica

7.3. Multifactor

7.3.1. Interpretando interacciones

Si se cruzan la slíneas la combinación es signigicativa pero n hay interacción INsertar muchos gráficos cuquies aquí, muhahahahahahah

8. EJERCICIOS

8.1. Tema 1

8.1.1. Métodos de muestreo

8.1.2. Distribuciones de muestreo

8.2. Tema 2

8.2.1. Definiciones

8.2.2. Propiedades de los estimadores

Demostrar que la media aritmética es un estimador consistente para el parámetro μ en una distribución $N(\mu, \sigma)$. Nota: usar la desigualdad de Chebychev: $P(|X - E(X)| >\geq \epsilon) \leq \frac{\sigma^2 x}{\epsilon^2}$

La igualdad de Chebychev nos dice lo siguiente, bajo la condición de que la varianza de una cierta variable aleatoria X sea finita:

$$\boxed{P\Big(|X - E[X]| \geq \epsilon\Big) \leq \frac{\sigma^2}{\epsilon^2}}$$

En el caso que nos ocupa, $X \sim N(\mu, \sigma)$, la variable aleatoria $\overline{X}_n \sim N\Big(\mu, \frac{\sigma}{\sqrt{n}}\Big)$, luego tendremos:

$$E[\overline{X}_n] = \mu$$
$$Var(\overline{X}_n) = \frac{\sigma^2}{n}$$

Resta sustituir en la expresión:

$$P\Big(|\overline{X}_n - \mu| \geq \epsilon\Big) \leq \frac{\sigma^2}{n\epsilon^2} \longrightarrow 0$$

8.2.3. Métodos para la obtención de estimadores

Mostrar que el estadístico $T = \sum_{i=0}^{n} x_i$ es suficiente para el parámetro p perteneciente a una muestra $x_1, ..., x_n$ que sigue una distribución $B(1, p)$.

AQUÍ IRÁ SU SOLUCIÓN MUHHAHHHHAHHA

Estimador por el método de máxima verosimilitud para una distribución Gamma de parámetros k y θ.

Su función de densidad es la siguiente:

$$f(x; k, \theta) := \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-\frac{x}{\theta}}$$

donde $\Gamma(k)$ es la función Gamma: $\Gamma(z) = \int_0^{+\infty} t^{z-1} e^{-t} dt$

$$f(x_1, ..., x_n; k, \theta) = \prod_{i=1}^{n} f(x_i; k, \theta) = \prod_{i=1}^{n} \frac{1}{\Gamma(k)\theta^k} x_i^{k-1} e^{-\frac{x_i}{\theta}}$$
$$= . - . - ....calcularestamierda$$

COMPROBAR TODOS LOS RESULTADOS:::: PUESTOS EN DUDA HASTA QUE LOS COMPRUEBE

$$L(k,\ theta; \overrightarrow{x}) = \sum_{i=1}^{n} \log\left(f(x_i|\theta)\right)$$

$$= -nklog(\theta) - nlog(\Gamma(k)) + (k-1)\sum_{i=1}^{n} ln(x_i) = \frac{\sum x_i}{\theta}$$

ESTO NO ESTÁ ACABADO NI A TIROS, ES SOLO UN ESBOZO

Estimador por máxima verosimilitud para el parámetro λ de una Poisson:

ESTO ES UN DESASTRE QUE ME SUENA HABER HECHO YA, COMPROBARLO Y SI NO ESTÁ ACABADO

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

$$L(\lambda | x_1, ..., x_n) = \prod_{i=1}^{n} f(x_i|\lambda) = e^{-n\lambda} \frac{\lambda^{\sum x_i}}{\prod x_i!}$$

Dada una muestra aleatoria simple de una distribución $N(\mu, 2)$, verificar que la media muestral es un estimador suficiente para el parámetro μ.

AQUÍ IRÁ SU SOLUCIÓN MUHHAHHHHAHHA

Encontrar un estadístico suficiente por el método de la máxima verosimilitud para θ para la distribución con la siguiente función de densidad bajo las condiciones $\theta > 0$ y $0 < x < 1$:

$$f_\theta(x) = \theta x^{\theta - 1}$$

Primero calculamos la función de densidad conjunta (**función de verosimilitud**) que, asumiendo independencia, es el producto de las funciones de densidad para cada x_i.

$$\begin{aligned} l(\theta; x_1, ..., x_n) &= \prod_{i=1}^{n} f_\theta(x_i) \\ &= \prod_{i=1}^{n} \theta x_i^{\theta-1} \\ &= \theta^n \prod_{i=1}^{n} x_i^{\theta-1} \qquad (1) \end{aligned}$$

Calculamos la el logaritmo de la función de verosimilitud para hacernos más sencillo calcular el **estimador de máxima verosimilitud (MLE)** $\widehat{\theta}$:

$$\begin{aligned} L(\theta; x_1, ..., x_n) &= \ln(l)(\theta; x_1, ..., x_n) \\ &= \ln\left(\theta^n \prod_{i=1}^{n} x_i^{\theta-1}\right) \\ &= n\ln(\theta) + (\theta-1)\sum_{i=1}^{n} \ln(x_i) \qquad (2) \end{aligned}$$

Hallamos el mínimo de la función para encontrar el MLE y comprobamos que es mínimo (pasos obviados aquí).

$$L_\theta(\theta; x_1, ..., x_n) = \frac{d}{d\theta} L(\theta; x_1, ..., x_n) = \frac{n}{\theta} + \sum_{i=1}^{n} \ln x_i$$

$$\begin{aligned} L_\theta(\theta; x_1, ..., x_n) &= \frac{d}{d\theta} L(\theta; x_1, ..., x_n) = 0 \\ \Rightarrow \widehat{\theta} &= \frac{-n}{\sum_{i=1}^{n} \ln x_i} \qquad (3) \end{aligned}$$

Por el Teorema de Fisher–Neyman sabemos que nuestro estimador será suficiente sí y solo sí:

$$f_\theta(x) = h(x) g_\theta(T(x)) = h(x_1, ..., x_n) g(T, \theta) = f(x_1, ..., x_n; \theta)$$

Donde T es el estimador. Observemos que aquí tenemos como estimador (3) y todo se reduce por el teorema a conseguir escribir la función dfe densidad como una combinación de otras dos, una que solo dependa de la muestra y otra que dependa de la muestra y del parámetro θ.

Sea $x_1, ..., x_n$ una muestra de una distribución normal $N(\mu, \sigma)$. Encontrar el método de los momentos para μ y para σ.

Recordemos primero algunas de las propiedades de la distribución normal:

- Función de densidad:

$$f(\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \sigma > 0, x \in \mathbb{R}, \mu \in \mathbb{R}$$

- Media: $E[X] = \mu$
- Varianza: $Var(X) = \sigma^2$

Ahora igualaremos los momentos muestrales y teóricos de orden 1 y 2:

- $$\frac{\sum_{i=1}^{n} x_i}{n} = E[X] = \mu \Rightarrow \widehat{\mu} = \overline{X}_n$$
- $$\frac{\sum_{i=1}^{n} x_i^2}{n} = E[X^2] = \mu^2 + \sigma^2 \Rightarrow \widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{X}_n = S^2$$

Encontrar el estimador por momentos del parámetro b de una distribución uniforme $U(0, b)$.

Recordemos primero algunas de las propiedades de la distribución uniforme continua $X \sim U(a, b)$:

- Función de densidad:
$$f(x) = \frac{1}{b-a} \quad a \leq x \leq b$$
- Función de distribución:
$$F(x) = P(X \leq x) = \frac{x-a}{b-a} \quad a \leq x \leq b$$
- Media: $E[X] = \frac{a+b}{2}$
- Varianza: $Var(X) = \frac{(b-a)^2}{12}$
Ahora realizamos el procedimiento:
$$\frac{1}{n}\sum_{i=1}^{n} x_i = E[X] = \frac{b}{2} \Rightarrow \widehat{b} = 2\overline{X}_n$$

Encontrar el estimador por el método de los momentos en los siguientes casos:

- Parámetro p en una distribución de Bernulli $B(1, p)$.
- Parámetro α en una distribución Exponencial $Exp(\alpha)$.
- Parámetros μ y σ en una Distribución Normal $N(\mu, \sigma)$.
- Parámetro b en una Distribución Uniforme $U(0, b)$.

- Parámetro p en una distribución de Bernulli $B(1, p)$.
Primero recordaremos algunos datos sobre la distribución de Bernulli:
  - Recalcar que es una distribución **discreta**.

- Función de probabilidad:

$$f(x) = p^x(1-p)^{1-x} = p^x q^{1-x} \quad x \in \{0,1\} \quad p \in [0,1]$$

- Media: $E[X] = p$
- Varianza: $Var(X) = pq$

$$\frac{1}{n}\sum_{i=1}^{n} x_i = E[X] = p \Rightarrow \widehat{p} = \overline{X}_n$$

- Parámetro α en una distribución Exponencial $Exp(\alpha)$.

  Primero recordaremos algunos datos sobre la distribución Exponencial ($\lambda > 0$):

  - Su función de probabilidad:

  $$f(x) = P(x) = \begin{cases} \lambda e^{-\lambda x} & si \quad x \geq 0 \\ 0 & si \quad x < 0 \end{cases}$$

  - Su función de distribución:

  $$F(x == P(X \leq x) = \begin{cases} 1 - e^{-\lambda x} & si \quad x \geq 0 \\ 0 & si \quad x < 0 \end{cases}$$

  - Su media: $E[X] = \frac{1}{\lambda}$
  - Su varianza: $Var(X) = \frac{1}{\lambda^2}$

  $$\frac{1}{n}\sum_{i=1}^{n} x_i = E[X] = \frac{1}{\lambda} \Rightarrow \widehat{\lambda} = \frac{1}{\overline{X}_n}$$

- Parámetros μ y σ en una Distribución Normal $N(\mu, \sigma)$.

  Sabemos que $\mu = E(X)$ y que $Var(X) = \sigma^2$ por la definición de distribución normal. Pero por definición de varianza también sabemos que $Var(X) = E(X^2) - E(X) = m_2 - m_1$.

  $$\Rightarrow \widehat{\mu} = \overline{X}_n$$

  $$\Rightarrow \widehat{\sigma}^2 = \frac{\sum_{i=1}^{n} X_i^2}{n} - (\overline{X}_n)^2 = S^2$$

- Parámetro b en una Distribución Uniforme $U(0, b)$.

  Sabemos por la definición de la distribución normal que $E(X) = \frac{b}{2} \to b = 2E(X) = 2m_1$, entonces:

  $$\Rightarrow \widehat{b} = 2\overline{X}_n$$

Encontrar el estimador por el método de máxima verosimilitud en los siguientes casos:

- Parámetro p en una distribución de Bernulli $B(1, p)$.
- Parámetro α en una distribución Exponencial $Exp(\alpha)$.
- Parámetros μ y σ en una Distribución Normal $N(\mu, \sigma)$.

- Parámetro b en una Distribución Uniforme $U(0,b)$.

- Parámetro p en una distribución de Bernulli $B(1,p)$.

  MLE para le parámetro p en una distribución de Bernulli $X \sim B(1,p)$.

  Sabemos que su función de densidad, para $x \in \{0,1\}$ y $0 \leq p \leq 1$ es:

$$f(x) = p^x(1-p)^{1-x}$$

  Tenemos una muestra de tamaño n: $x_1, ..., x_n$.

  Calculamos la función de verosimilitud:

$$l(p; x_1, ..., x_n) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}$$

  Tomamos logaritmos:

  Ahora buscaremos el máximo de la función L de la manera habitual, derivamos para hallar un candidato a máximo, que será nuestro estimador $\widehat{p}$

$$\frac{dL}{dp} = \frac{\sum x_i}{p} - \frac{n - \sum x_i}{1-p} = 0$$

$$\rightarrow \sum x_i - p\sum x_i - np + p\sum x_i \Rightarrow p^* = \frac{\sum x_i}{n}$$

  Comprobamos ahora que se trata de un mínimo:

$$\frac{d^2L}{dp^2} = \frac{-\sum x_i}{p^2} - \frac{(n - \sum x_i)}{(1-p)^2} < 0$$

  Podemos concluir pues que $\widehat{p} = p^*$

- Parámetro α en una distribución Exponencial $Exp(\alpha)$.

  Sabemos que su función de densidad es $f(x) = \frac{1}{\alpha}e^{-\frac{x}{\alpha}}$

$$l(\alpha; x_1, ..., x_n) = \frac{1}{\alpha^n}e^{\frac{-\sum_{i=1}^{n} x_i}{n}}$$

$$L(\alpha; x_1, ..., x_n) = \ln(l)(\alpha; x_1, ..., x_n) = -n\ln(\alpha) - \frac{\sum_{i=1}^{n} x_i}{\alpha}$$

$$\frac{\partial L}{\partial \alpha} = \frac{-n}{\alpha} + \frac{\sum_{i=1}^{n} x_i}{\alpha^2} = 0$$

$$-n\alpha + \sum x_i = 0 \Rightarrow \widehat{\alpha} = \frac{\sum x_i}{n} = \overline{x_n}$$

$$\frac{\partial^2 L}{\partial \alpha^2} = \frac{n}{\alpha^2} - \frac{2\alpha\sum x}{\alpha^4} = \frac{n}{\alpha^2} - \frac{2\sum x}{\alpha^3} \overbrace{<}^{?} 0$$

  Teniendo en cuenta el hecho de que $x > 0$ nos fijamos en el numerador de la expresión anterior para poder llegar al objetivo.

$$n\alpha - 2\sum x \Rightarrow n\overline{x_n} - 2nx = \frac{-n\overline{x_n}}{\overline{x_n}^3} < 0$$

- Parámetros μ y σ en una Distribución Normal $N(\mu, \sigma)$.
  Sabemos que si $X \sim N(\mu, \sigma)$ entonces:

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

$$l(\mu, \sigma; x_1, \cdots, x_n) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{\frac{-\sum(x_i-\mu)^2}{2\sigma^2}}$$

$$L(\mu, \sigma; x_1, \cdots, x_n) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \sum \frac{(x_i-\mu)^2}{2\sigma^2}$$

$$\begin{cases} \frac{\partial L}{\partial \mu} & = -\frac{2\sum(x_i-\mu)}{2\sigma^2} = -\sum \frac{x_i-\mu}{\sigma^2} = 0 \\ \frac{\partial L}{\partial \sigma} & = -\frac{n}{2\sigma^2} + \sum \frac{(x_i-\mu)^2}{2(\sigma^2)^2} = 0 \end{cases}$$

  - De la primera de las ecuaciones obtenemos:

$$\sum(x_i - \mu) = 0 \Rightarrow \widehat{\mu} = \overline{x_n}$$

  - De la segunda (quedando cálculos pendientes):

$$-n\sigma^2 + \sum(x_i - \mu)^2 = 0$$

$$-n\sigma^2 + \sum(x_i - \overline{x_n})^2 = 0$$

$$\Rightarrow \widehat{\sigma}^2 = \sum \frac{(x_i - \overline{x_n})^2}{n} = S^2$$

- Parámetro b en una Distribución Uniforme $U(0, b)$.
  Como sabemos la función de densidad de una variable aleatoria $X \sim U(0, b)$ es:

$$f(x) = \frac{1}{b} \quad 0 < x < b$$

$$x_1, x_2, ..., x_n \longrightarrow l(x_1, ..., x_n) = \prod_{i=1}^{n} f(x_i) = \frac{1}{b^n}$$

$$\ln(l(x_1, ..., x_n)) = -n\ln(b)$$

$$\frac{\partial L}{\partial b} = \frac{-n}{b} \neq 0 \text{ para cualquier valor de b}$$

$\Rightarrow$ La ecuación de verosimilitud NO TIENE SOLUCIÓN!!

En este caso nuestro estimador será el siguiente (aunque no llegaremos a él por este método):

$$\widehat{b} = \max\{x_1, x_2, \cdots, x_n\}$$

Consideremos el siguiente modelo de regresión, siendo $e \sim N(0, \sigma)$:

$$y_i = \alpha + \beta x_i + e$$

Probar que los estimadores por máxima verosimilitud y por el método de los momentos para los parámetros α y β resultan el mismo.

solución

Sea X una variable aleatoria que sigue una distribución exponencial de parámetro α.

$$f(x) = \frac{1}{\alpha} e^{\frac{-x}{\alpha}}$$

Para $x \geq 0$.

- Hallar el estimador máximo verosímil de α para una muestra aleatoria de tamaño n.
- Referido al estimador anterior:
  - ¿Es un estimador insesgado?
  - Hallar el error cuadrático medio del estimador.
  - ¿Cuál es la eficiencia absoluta del estimador?
- Hay tres tipos de babosas: verdes, púrpuras y rayadas. El tiempo de vida de las babosas verdes sigue una distribución exponencial de parámetro α. El tiempo de vida de las babosas púrpura sigue una distribución exponencial de parámetro 4α. El tiempo de vida de las babosas rayadas sigue una distribución exponencial de parámetro 16α. Hemos observado la siguiente muestra:
  - 1 babosa verde con un tiempo de vida de 39.
  - 2 babosas púrpura con tiempos de vida de 45 y 165.
  - 1 babosa rayada con un tiempo de vida de 900.

  Utilizar el método de la máxima verosimilitud para hallar α.

- solución
- solución
  - solución
  - solución
  - solución
-
  - solución
  - solución
  - solución

Sean $X_1, ..., X_n$ variables aleatorias independientes e indénticamente distribuídas con la siguiente función de densidad para cada X_i (con $x > 0$):

$$f(x;\theta) = \frac{x^2}{2\theta^3} e^{\frac{-x}{\theta}}$$

Como dato sabemos que $E(X) = 3\theta$ y que $Var(X) = 3\theta^2$.

- Hallar el estimador por el método de los momentos de θ.
- Calcular la función de verosimilitud para una muestra aleatoria de tamaño n.
- Hallar el estimador máximo verosímil de θ.
- Demostrar que ambos estimadores son insesgados.
- Demostrar que $\sum_{i=1}^{n} X_i$ es un estimador suficiente para el parámetro θ.
- Halla la cota de Cramer-Rao para la varianza de un estimador insesgado de θ.
- ¿Cuál es el valor de la información de Fisher en una observación individual de esta densidad?
- ¿Es el estimador de máxima verosimilitud el estimador insesgado de la varianza mínima de θ?
- Hallar el estadístico del test de la razón de verosimilitud para contrastar la hipótesis $H_0 : \theta = \theta_0$ contra la alternativa $H_1 : \theta \neq \theta_0$.
- Supongamos que tenemos una muestra aleatoria de tamaño $n = 27$ produce $\sum_{i=1}^{27} x_i = 108$. Utilizando el contraste obtenido antes, ¿podemos rechazar la hipótesis $H_0 : \theta = 1$ frente a la alternativa $H_1 : \theta \neq 1$ con un nivel de significación de $\alpha = 0{,}1$?
- BONUS: comprobar los valores de la varianza y de la esperanza.

- Se trata de igualar el momento muestral $(\sum_{i=1}^{n} \frac{x_i}{n}, \sum_{i=1}^{n} \frac{x_i^2}{n}, \cdots, \sum_{i=1}^{n} \frac{x_i^k}{n})$ con el teórico $(E[X], E[X^2], \cdots, E[X^k])$. En nuestro caso como solo tenemos un parámetro serán los momentos de primer orden.

$$\overline{X}_n = \sum_{i=1}^{n} \frac{x_i}{n} = E[X] = 3\theta$$

$$\Rightarrow \widehat{\theta}_{MOM} = \frac{\overline{X}_n}{3}$$

-

$$l(\theta) = f(x_1, ..., x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

$$= \frac{\prod_{i=1}^{n} x_i^2}{2^n \theta^{3n}} e^{\frac{-\sum_{i=1}^{n} x_i}{\theta}}$$

- Partimos del apartado anterior:

$$L = \ln\Big(l(\theta)\Big) = \ln\Big(f(x_1, ..., x_n; \theta)\Big)$$

$$= 2\sum_{i=1}^{n} \ln(x_i) - n\ln(2) - 3n\ln(\theta) - \frac{\sum_{i=1}^{n} x_i}{\theta}$$

$$\frac{\partial L}{\partial \theta} = \frac{-3n}{\theta} + \frac{\sum_{i=1}^{n} x_i}{\theta^2} = 0$$

$$\theta\left(\frac{-1}{3n} + \frac{1}{\sum x_i}\theta\right) = 0$$

Tenemos dos soluciones, la primera es $\theta = 0$ y la descartamos por motivos obvios, luego nos queda:

$\Rightarrow \widehat{\theta}_{MLE} = \frac{\overline{X}_n}{3}$

- Como ambos estimadores tienen la misma expresión, lo calcularemos en el caso de uno de ellos:

$$E(\overline{X}_n/3) = \frac{3\theta}{3} = \theta$$

- Para demostrar que $T = \sum_{i=1}^{n} x_i$ es un estimador suficiente aplicamos el **Criterio de factorización de Neyman-Fisher**:

  Esto quiere decir que si T es un estimador insesgado y es suficiente debemos se capaces de encontrar dos funciones no negativas tales que:

$$f(x_1, ..., x_n; \theta) = g(T; \theta) \cdot h(x_1, ..., x_n)$$

  Lo haremos escribiendo la función de densidad conjunta y sustituyendo la expresión de T:

$$\begin{aligned} f(x_1, ..., x_n; \theta) &= \frac{\prod_{i=1}^{n} x_i^2}{2^n \theta^{3n}} e^{\frac{-\sum x_i}{\theta}} \\ &= \frac{e^{\frac{-T}{\theta}}}{2^n \theta^{3n}} \cdot \prod_{i=1}^{n} x_{i=1}^{n} \end{aligned}$$

$$\begin{aligned} &\Rightarrow g(T; \theta) = \frac{e^{\frac{-T}{\theta}}}{2^n \theta^{3n}} \\ &\Rightarrow h(x_1, ..., x_n) = \prod_{i=1}^{n} x_{i=1}^{n} \end{aligned}$$

- $$\ln\left(f(x; \theta)\right) = 2\ln(x) - \ln(2) - 3\ln(\theta) + \frac{-x}{\theta}$$

$$\begin{aligned} \frac{\partial f}{\partial \theta} &= -\frac{3}{\theta} + \frac{x}{\theta^2} \\ \frac{\partial^2 f}{\partial \theta^2} &= \frac{3}{\theta^2} - \frac{2x}{\theta^3} \end{aligned}$$

$$E\left(\frac{3}{\theta^2} - \frac{2x}{\theta^3}\right) = -\frac{3}{\theta^2}$$

$$\Rightarrow CRB(\theta) = \frac{1}{\frac{3n}{\theta^2}} = \frac{\theta^2}{3n}$$

- DETALLAR MÁS ESTE APARTADO

$$\begin{aligned} \mathcal{I}_X(\theta) &= E\left(\left(\frac{\partial}{\partial \theta}\ln(f)\right)^2\right) \\ &= \frac{3}{\theta^2} \end{aligned}$$

- Calculamos su varianza y la comparamos con la cota de Cramer-Rao, si tenemos el caso de igualdad es el mejor estimador con el que podíamos contar:

$$Var\left(\frac{\overline{X}_n}{3}\right) = \frac{3\theta^2}{9n} = \frac{\theta^2}{3n} = CRB(\theta)$$

- Recordemos la hipótesis que estamos considerando:

$$H_0 : \theta = \theta_0$$
$$H_1 : \theta \neq \theta_0$$

Recordemos también algunas de las definiciones sobre las que se basarán los cálculos:

$$l(\widehat{\Theta}) = \sup_{\theta\in\Theta} l(\theta; X)$$
$$l(\Theta) = \sup_{\theta\in\Theta_0} l(\theta; X)$$

$$\Lambda := \frac{l(\Theta_0)}{l(\Theta)}$$

Antes de realizar los cálculos, fijémonos primero en los espacios que nos ocupan:

$$\Theta_0 = \{\theta : \theta = \theta_0 \wedge \theta \in \mathbb{R}\}$$
$$\Theta = \{\theta : \theta \neq \theta_0 \wedge \theta \in \mathbb{R}\}$$

$$l(\Theta_0) =$$

$$l(\Theta) =$$

$$\Lambda = \frac{l(\Theta_0)}{l(\Theta)} = \frac{\frac{e^{-\sum_{i=1}^n x_i/\theta_0}}{\theta_0^{3n}}}{\frac{e^{-3n}}{\widehat{\theta}^{3n}}} = \left(\frac{\widehat{\theta}}{\theta_0}\right)^{3n} e^{3n-\frac{\sum x_i}{\theta_0}}$$

- DETALLAR MÁS ESTE APARTADO
  $\widehat{\theta} = \frac{108}{81}$

$$\Lambda = \left(\frac{108}{81}\right)^{81} e^{81-108} = 0{,}024779$$
$$U = -2\ln(\Lambda) = 7{,}3955$$
$$\chi_{1,0,1} = 2{,}70554 \Rightarrow \text{Rechazamos } H_0$$

- QUEDA PENDIENTE DE HACER

Sea $X_1, ..., X_n$ una muestra aleatoria de una distribución $X \sim N(\mu, \sigma)$. Dados los siguientes estimadores ?¿cuál es preferible desde el punto de vista de MLE?.

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

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

solución

Tenemos una muestra aleatoria de la variable definida por la siguiente función de densidad (para $0 < x < \beta$):

$$f(x, \beta) = \frac{2}{\beta^2}(\beta - x)$$

Encontrar los estimadores por el método de máxima verosimilitud y por momentos para β. Encontrar la desviación de cada uno de los estimadores y también la eficiencia relativa del procedente de la máxima verosimilitud respecto al hallado por momentos.

solución

8.2.4. Métodos de remuestreo

8.3. Tema 3

8.3.1. Cantidades pivotales

8.3.2. Intevalos usuales

8.3.3. Intervalos de confianza asintónticos

8.3.4. Evaluación de intervalos de confianza

8.3.5. Determinación del tamaño muestral

8.4. Tema 4

8.4.1. Test de máxima verosimilitud

Consideremos la muestra $x_1, ..., x_n$ de una variable aleatoria $X \sim N(\mu, \sigma_0)$ con σ_0 conocido y calculemos el test de razón de verosimulitud para el siguiente test:

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu \neq \mu_0$$

Notemos primero como es la estructura de los espacios de parámetros en los que trabajaremos, esta estará condicionada por el hecho de que la varianza es un parámetro fijo σ_0:

- $\Theta = \{(\mu, \sigma_0) : \mu \in \mathbb{R}\}$.
- $\Theta_0 = \{(\mu_0, \sigma_0)\}$.

$$l(\theta; x_1, ..., x_n) = \left(\frac{1}{\sigma_0\sqrt{2\pi}}\right)^n e^{-\frac{\sum(x_i-\mu)^2}{2\sigma_0}}$$

$$l(\Theta_0) = \left(\frac{1}{\sigma_0\sqrt{2\pi}}\right)^n e^{-\frac{\sum(x_i-\mu_0)^2}{2\sigma_0}}$$

Ahora buscamos el máximo de la función de verosimilitud sin ninguna restricción.

$$L = -n\ln(\sigma_0\sqrt{2\pi}) - \frac{\sum(x_i-\mu)^2}{2\sigma_0^2}$$
$$\frac{\partial L}{\partial\mu} = \frac{2\sum(x_i-\mu)}{2\sigma_0^2} = 0$$
$$\Rightarrow \widehat{\mu} = \frac{\sum x_i}{n} = \overline{x}$$

$$\Rightarrow l(\Theta_0) = \left(\frac{1}{\sigma_0\sqrt{2\pi}}\right)^n e^{-\frac{\sum(x_i-\overline{x})^2}{2\sigma_0}}$$

Para los cálculos posteriores, utilizaremos que, bajo H_0:

$$\frac{\overline{x}-\mu_0}{\sigma_0/\sqrt{n}} \sim N(0,1)$$

por que sabemos que: $\overline{x} \sim N(\mu, \frac{\sigma_0}{\sqrt{n}})$

Llamaremos $Z = .\frac{\overline{x}-\mu_0}{\sigma_0/\sqrt{n}}$.

Así nos queda:

$$\Lambda = \frac{l(\widehat{\Theta_0})}{l(\widehat{\Theta})} = e^{-\frac{n}{2}\frac{(\overline{x}-\mu_0)^2}{\sigma_0^2}}$$
$$= e^{-\frac{Z^2}{2}}$$

INSERTAR IMÁGEN AQUI DE QUÉ ES LO QUE BUSCAMOS, EN EJE Y $\Lambda(Z)$

Ahora se tratará de calcular el valor de $k_\alpha = \Lambda(-z) = \Lambda(+z)$ tal que $P(\Lambda \leq K_\alpha | H_0) = \alpha$. Usaremos el conocimiento de la distribución de Z para ahorrarnos el cálculo de k_α.

Sea $x_1, \cdots, x_n$ una muestra aleatoria de valores independientes de la siguiente variable: $X \sim N(\mu, \sigma)$.

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu \neq \mu_0$$

Observemos que en nuestros espacios serán $\Theta := \{(\mu, \sigma) : \mu \neq \mu_0\}$ y $\Theta_0 := \{(\mu_0, \sigma)\}$.

$$l(\mu, \sigma; x_1, \cdots, x_n) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{\sum(x_i-\mu)^2}{2\sigma^2}}$$

- En Θ:

$$\widehat{\mu} = \overline{x}$$
$$\widehat{\sigma^2} = S^2 = \frac{\sum(x_i - \overline{x})^2}{n}$$

- En Θ_0:

$$l = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{\sum(x_i-\mu_0)^2}{2\sigma^2}}$$
$$L = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{\sum(x_i - \mu_0)^2}{2\sigma^2}$$
$$L_\theta = -\frac{1}{2\sigma^2} + \frac{\sum(x_i - \mu_0)^2}{2(\sigma^2)^2} = 0$$
$$\Rightarrow -n\sigma^2 + \sum_{i=1}(x_i - \mu_0)^2 = 0$$
$$\Rightarrow \widehat{\sigma^2} = \frac{\sum(x_i - \mu_0)^2}{n}$$

Sabemos por la teoría que:

$$\Lambda = \frac{l(\widehat{\Theta_0})}{l(\widehat{\Theta})}$$

$$l(\widehat{\Theta_0}) = \left(\frac{1}{\sqrt{2\pi}\sqrt{\frac{\sum(x_i-\mu_0)^2}{n}}}\right)^n e^{\frac{-\sum(x_i-\mu_0)^2}{2\frac{\sum(x_i-\mu_0)^2}{n}}}$$

El denominador es similar pero utilizando $\overline{x}$.

Llegado cierto punto, utilizaremos también el hecho de que $\frac{\overline{x}-\mu_0}{S/n-1} \sim T_{n-1}$.

$$\begin{aligned}
\Lambda = \frac{l(\widehat{\Theta_0})}{l(\widehat{\Theta})} = \left(\frac{\sum(x_i-\overline{x})^2}{\sum(x_i-\mu_0)^2}\right)^{n/2} &= \left(\frac{nS^2}{\sum(x_i^2 - 2\mu_0 \underbrace{\sum x_i}_{n\overline{x}} + n\mu_0^2)}\right)^{n/2} \\
&= \left(\frac{nS^2}{\sum x_i^2 - 2n\overline{x}\mu_0 + n\mu_0^2 + n\overline{x}^2 - n\overline{x}^2}\right)^{n/2} \\
&= \left(\frac{nS^2}{nS^2 + \sum n\mu_0^2 - 2\mu_0 n\overline{X}}\right)^{n/2} = \left(\frac{nS^2}{nS^2 + n(\overline{x}-\mu_0)^2}\right)^{n/2} \\
&= \left(\frac{1}{1+\left(\frac{\overline{x}-\mu_0}{S}\right)^2}\right)^{n/2} = \left(\frac{1}{1+\left(\frac{T}{\sqrt{n-1}}\right)^2}\right)^{n/2} \\
&= \Lambda(T)
\end{aligned}$$

Donde tenemos que:

$$T \sim T_{n-1}$$

INSERTAR AQUÍ IMÁGEN DE LA T DE STUDENT CON 2 COLAS, DONDE EL EJE Y ES $\Lambda(T)$ y el X es T.

$\boxed{\text{Cálculo de } k_\alpha \text{ mediante la relación funcional}}$

Basándonos en parte en el ejemplo anterior (muestra aleatoria independiente $x_1, ..., x_n$ de una variable $X \sim N(\mu, \sigma)$), utilizaremos la relación funcional para calcular k_α. Supongamos ahora que nuestro test es el siguiente:

$$H_0 : \mu \leq \mu_0$$
$$H_1 : \mu > \mu_0$$

Definamos primero los espacios en los que trabajaremos:

- $\Theta = \{(\mu, \sigma) : \mu \in \mathbb{R} \wedge \sigma \in \mathbb{R}^+\}$
- $\Theta_0 = \{(\mu, \sigma) : \mu \in \mathbb{R} \wedge \mu \leq \mu_0 \wedge \sigma \in \mathbb{R}^+\}$
- En Θ:

$$\widehat{\mu} = \overline{x}$$
$$\widehat{\sigma^2} = S^2$$

- En Θ_0:

$$\widehat{\mu} = \begin{cases} \overline{x} & si \quad \mu_0 > \overline{x} \\ \mu_0 & si \quad \mu_0 < \overline{x} \end{cases}$$

$$\widehat{\sigma^2} = \frac{\sum (x_i - \widehat{\mu})^2}{n}$$

Ahora ya podemos calcular la función que nos interesa, basándonos en el caso anterior:

$$\Lambda = \begin{cases} 1 & si \quad \mu_0 > \overline{x} \\ \Lambda(T) & si \quad \mu_0 < \overline{x} \text{ Nos encontramos en el caso anterior} \end{cases}$$

INCLUIR AQUÍ EL GRÁFICO APROPIADO

Sean $x_1, ..., x_{n_1}$ e $y_1, ..., y_{n_2}$ dos muestras independientes de, respectivamente, las variables aleatorias:

$$X \sim N(\mu_1, \sigma_0) \qquad Y \sim N(\nu_2, \sigma_0)$$

donde σ_0 es conocido y el test es el siguiente:

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

Observemos por la descripción del test que los espacios son los siguiente:

$$\Theta = \{(\mu_i, \mu_j) : \ \mu_i, \mu_j \in \mathbb{R}\}$$
$$\Theta_0 = \{(\mu_i, \mu_i) : \ \mu_i \in \mathbb{R}\}$$

$$l = \left(\frac{1}{\sigma_0\sqrt{2\pi}}\right)^{n_1} e^{-\frac{\sum(x_i-\mu_1)^2}{2\sigma_0}} \cdot \left(\frac{1}{\sigma_0\sqrt{2\pi}}\right)^{n_2} e^{-\frac{\sum(x_i-\mu_2)^2}{2\sigma_0}}$$

- En Θ:

$$\widehat{\mu_1} = \overline{x}$$
$$\widehat{\mu_2} = \overline{y}$$

- En Θ_0:

$$L = K_1 - \frac{\sum^{n_1}(x_i - \mu)^2}{2\sigma_0^2} + K_2 - \frac{\sum^{n_2}(y_i - \mu)^2}{2\sigma_0^2}$$

$$L_\mu = \frac{\sum^{n_1}(x_i - \mu)^2}{\sigma_0^2} + \frac{\sum^{n_2}(y_i - \mu)^2}{\sigma_0^2} = 0$$

$$\Rightarrow \sum x_i - n_1\mu + \sum y_i - n_2\mu = 0$$

$$\Rightarrow \widehat{\mu} = \frac{n_1\overline{x} + n_2\overline{y}}{n_1 + n_2}$$

Antes de proseguir los cálculos, observemos que, bajo H_0:

$$\overline{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n_1}}\right)$$

$$\overline{Y} \sim N\left(\mu, \frac{\sigma}{\sqrt{n_2}}\right)$$

$$\overline{X} - \overline{Y} \sim N\left(\mu, \sigma \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)$$

$$\Rightarrow \frac{\overline{X} - \overline{Y}}{\sigma \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = Z \sim N(0,1)$$

Ahora podemos proseguir con los cálculos:

$$\Lambda = \frac{l(\widehat{\Theta_0})}{l(\widehat{\Theta})} = \frac{e^{\frac{-\sum_{i=1}^{n_1}(x_i-\widehat{\mu})^2 - \sum_{j=1}^{n_2}(y_j-\widehat{\mu})^2}{2\sigma_0^2}}}{e^{\frac{-\sum_{i=1}^{n_1}(x_i-\overline{x})^2 - \sum_{j=1}^{n_2}(y_j-\overline{y})^2}{2\sigma_0^2}}}$$

$$= e^{\frac{1}{2\sigma_0}\left(\frac{n_1 \cdot n_2}{n_1+n_2}(\overline{X}-\overline{Y})^2\right)} = e^{-Z^2/2}$$

$$= \Lambda(Z)$$

INSERTAR AQUÍ LA IMÁGEN CORRESPONDIENTE MUHAHAHAHAHAH

Test de máxima verosimilitud para la varianza de distribuciones normales ($X \sim M(\mu_i, \sigma_i)$ $i = 1, 2, ..., k$) con el siguiente test:

$$H_0 : \sigma_1 = \sigma_2 = \cdots = \sigma_k = \sigma$$
$$H_1 : \exists i, j \ : \sigma_i \neq \sigma_j$$

$$l = \prod_{i=1}^{k} \left(\frac{1}{\sigma_i \sqrt{2\pi}}\right)^{n_i} e^{\frac{-\sum_{j=1}^{n_i}(x_{ij}-\mu_i)^2}{2\sigma_i^2}}$$

Ahora tomaremos logaritmos y derivaremos:

- En Θ:

$$\widehat{\mu_i} = \overline{x_i}$$
$$\widehat{\sigma_i} = S_i^2 = \frac{\sum_{j=1}^{n_i}(x_{ij} - \overline{x_i})^2}{n_i}$$

- En Θ_0:
  $n_1 + n_2 + \cdots + n_k = N$

$$l = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^N e^{\frac{-\sum_{i=1}^{k}\sum_{j=1}^{n_j}(x_{ij}-\mu_i)^2}{2\sigma^2}}$$

$$\widehat{\mu_i} = \overline{x_i}$$

$$L = -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma^2) - \sum_{i=1}^{k}\sum_{j=1}^{n_i}\frac{(x_{ij}-\overline{x_i})^2}{2\sigma^2}$$

$$\frac{\partial L}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{\sum_i\sum_j(x_{ij}-\overline{x_i})^2}{2(\sigma^2)^2} = 0$$

$$\Rightarrow -N\sigma^2 + \sum_i\sum_j(x_{ij}-\overline{x_i})^2 = 0$$

$$\Rightarrow \widehat{\sigma}^2 = \frac{\sum_i\sum_j(x_{ij}-\overline{x_i})^2}{N}$$

$$\widehat{\sigma}^2 = \frac{\sum_i n_i S_i^2}{N}$$

$$\Lambda = \frac{l\left(\widehat{\Theta_0}\right)}{l\left(\widehat{\Theta}\right)} = \cdots$$

$$= \frac{\prod_{i=1}^{k}(S_i^2)^{\frac{n_i}{2}}}{(S^2)^{\frac{N}{2}}}$$

Ahora utilizaremos el Teorema de Wills:

$U = -2\ln\Lambda = -2\left(\sum_{i=1}^{k}\frac{n_i}{2}\ln(S_i^2) - \frac{N}{2}\ln(S^2)\right)$

Bajo H_0: $U \sim \chi^2_{2k-(k+1)} = \chi^2_{k-1}$

(Tenemos $2k$ parámetros en Θ (su dimensión) y $k+1$ en Θ_0)

$\boxed{\text{En el siguiente ejemplo justificaremos el TEST ANOVA a partir de la Razón de Verosimilitud}}$

Test de máxima verosimilitud para la media de distribuciones normales ($X \sim M(\mu_i, \sigma_i)$ $i = 1, 2, ..., k$) con el siguiente test:

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k = \mu$$
$$H_1 : \exists i, j \ : \mu_i \neq \mu_j$$

$$l = \left(\frac{1}{\sigma\sqrt{2\pi}}\right) e^{\frac{-\sum_i\sum_j x_{ij}(x_{ij}-\mu_i)^2}{2\sigma^2}}$$

- En Θ:

$$\widehat{\mu_i} = \overline{x_i}$$

Ahora utilizaremos la terminología ANOVA (que veremos en próximos temas).

$$\widehat{\sigma}^2 = \frac{\sum_i \sum_j (x_{ij} - \overline{x_i})^2}{N} = \frac{SSW}{N}$$

- En Θ_0:

$$l = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^N e^{\frac{-\sum_i \sum_j (x_{ij}-\mu)^2}{2\sigma^2}}$$

Utilizaremos la siguiente igualdad aunque no la demostremos aquí:

$$SST = \underbrace{\sum_i n_i(\overline{x_i} - \overline{x})^2}_{SSB} + \underbrace{\sum_i \sum_j (x_{ij} - \overline{x_i})^2}_{SSW}$$

$$\begin{aligned} \widehat{\mu} &= \frac{\sum_i \sum_j x_{ij}}{N} = \overline{x} \\ \widehat{\sigma}^2 &= \frac{\sum_i \sum_j (x_{ij} - \overline{x})^2}{N} = \frac{SST}{N} \\ &= \frac{SSB + SSW}{N} \end{aligned}$$

$$\Rightarrow \Lambda = \frac{\left(\frac{SSW}{N}\right)^{N/2}}{\left(\frac{SSB+SSW}{N}\right)^{N/2}} = \left(\frac{SSW}{SSB + SSW}\right)^{\frac{N}{2}}$$

Rechazaremos la hipótesis si $\Lambda < k_\alpha$:

$$\begin{aligned} &\Rightarrow \frac{SSB + SSW}{SSW} > c' \\ &\Rightarrow \frac{SSB}{SSW} + 1 > c' \Longleftrightarrow \frac{SSB}{SSW} > c'' \end{aligned}$$

Ya sabíamos lo que esto implica, $\frac{nS^2}{\sigma^2} \sim \chi^2_{n-1}$

$SSW = \sum n_i S_i$

Las dos siguientes variables aleatorias serán independientes:

- $\frac{SSW}{\sigma^2} = \frac{\sum_i n_i S_i^2}{\sigma^2} \sim \chi^2_{N-k}$
- $\frac{n_i S_i^2}{\sigma^2} \sim \chi^2_{n_i - 1}$

Análogamente es demostrable lo siguiente:

$$\frac{SSB}{\sigma^2} \sim \chi^2_{k-1}$$

$$\Rightarrow \frac{\left(\frac{SSB}{\sigma^2}\right)/k-1}{\left(\frac{SSW}{\sigma^2}\right)/N-k} = \frac{SSB/k-1}{SSW/N-k} \sim \mathfrak{F}_{k-1,N-k}$$

$X \sim f(x;\theta)$

$$H_0 : \theta = \theta_0$$
$$H_1 : \theta = \theta_1$$

Observemos que el test nos indica lo siguiente:

$$\Theta = \{\theta_0, \theta_1\}$$

Así tenemos:

$$\Lambda = \frac{l(\widehat{\Theta_0})}{l(\widehat{\Theta})} = \frac{l(\theta_0)}{\text{máx}\,\{l(\theta_0), l(\theta_1)\}} = \begin{cases} 1 & si \quad l(\theta_0) > l(\theta_1) \\ \frac{l(\theta_0)}{l(\theta_1)} & si \quad l(\theta_0) < l(\theta_1) \end{cases}$$

$$W = \{1 \leq k_\alpha\}$$

Sea $x_1, ..., x_n$ un conjunto de muestras independientes de una variable $X \sim Poiss(\lambda)$ y el siguiente test:

$$H_0 : \lambda = \lambda_0$$
$$H_1 : \lambda = \lambda_1$$

Asumimos $\lambda_1 > \lambda_0$.

- Encontrar el test de máxima verosimilitud.
- Aplicar el anterior apartado al caso en el que el proceso observado durante 60 minutos tiene un total de 327 casos de éxito y contamos con el siguiente test:

$$H_0 : \lambda = \lambda_0 = 5$$
$$H_1 : \lambda = \lambda_1 = 6$$

- $$l(\lambda; x_1, ..., x_n) = e^{-n\lambda} \frac{\lambda^{\sum x_i}}{\prod x_i!}$$

  $$\frac{l(\lambda_0; x_1, ..., x_n)}{l(\lambda_1; x_1, ..., x_n)} = \frac{e^{-n\lambda_0} \frac{\lambda_0^{\sum x_i}}{\prod x_i!}}{e^{-n\lambda_1} \frac{\lambda_1^{\sum x_i}}{\prod x_i!}} = e^{-n(\lambda_1 - \lambda_0)} \left(\frac{\lambda_0}{\lambda_1}\right)^{\sum x_i}$$

  $$\Lambda \leq k_\alpha \Longleftrightarrow \sum_{i=1}^{n} x_i > C$$

  $W = \{\Lambda \leq k_\alpha\}$

  $P(W|H_0) = \alpha$

  Bajo H_0 sabemos que, en este caso:

  $$\sum_{i=1}^{n} x_i \sim Poiss(n\lambda_0)$$

  y usaremos esto para hallar el valor de p.

- En nuestro caso tenemos $n = 60$ y $\sum x_i = 3$.

  Si H_0 cierto: $\sum^{60} x_i \sim Poiss(300)$.

  Utilizamos el software R para calcular el resultado: *qpois(*$0,95,300$*)*$= 329 = p_{0,05}$

8.4.2. Test no paramétricos robustos

8.4.3. Test múltiple

8.4.4. Test clásicos no paramétricos

8.4.5. Test de Fisher

8.4.6. Determinación del tamaño muestral

8.5. Tema 5

8.5.1. Ajuste del modelo

8.5.2. Table ANOVA

8.5.3. Bondad del ajuste

8.5.4. Distribución de los coeficientes

8.5.5. Intervalos de confianza para los coeficientes

8.5.6. Predicción

8.5.7. Correlación, causalidad e interpretación

8.6. Tema 6

8.6.1. Estimación de parámetros (un factor)

8.6.2. Partición de la varianza (un factor)

8.6.3. Hipótesis nula y tésts para la hipótesis nula

9. TEOREMAS Y DEMOSTRACIONES