

Inferencia Estadística

Manuel Gijón Agudo

Octubre 2018 -

Índice

1. Tema 0: Introducción a la inferencia estadística	3
2. Tema 1: Muestreo	4
2.1. Definiciones	4
2.2. Métodos de muestreo	5
2.2.1. Muestreo aleatorio simple	5
2.3. Distribuciones de muestreo	5
3. Tema 2: Estimación de parámetros	6
3.1. Definiciones y propiedades de los estimadores	6
3.2. Métodos para la obtención de estimadores	9
3.3. Métodos de remuestreo	9
4. Tema 3: Intervalos de confianza	10
5. Tema 4: Contraste de Hipótesis	11
6. Tema 5: Modelos de Regresión	12
6.1. Regresión lineal simple	12
6.1.1. Ajuste del modelo	14
6.1.2. Tabla ANOVA	15
6.1.3. Bondad del ajuste	15
6.1.4. Distribución de los coeficientes	16
6.1.5. Intervalos de confianza para los coeficientes	18
6.1.6. Tests de significación para los coeficientes	18
6.1.7. Predicción	19
7. Tema 6: ANOVA	20
7.1. Tema 1	20
7.1.1. Métodos de muestreo	20
7.1.2. Distribuciones de muestreo	20
7.2. Tema 2	20

<i>Inferencia Estadística</i>	2
7.2.1. Definiciones	20
7.2.2. Propiedades de los estimadores	20
7.2.3. Métodos para la obtención de estimadores	21
7.2.4. Métodos de remuestreo	27
7.3. Tema 3	27
7.4. Tema 4	27
7.5. Tema 5	27
7.6. Tema 6	27

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 tienen 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 muestreo se encuentran los siguientes:

- **Muestreo aleatorio simple:**
- **Muestreo sistemático:**
- **Muestreo estratificado:**
- **Muestreo de clústering:**
- 'Quota 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, \dots, 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, \dots, x_n) provenientes de (X_1, \dots, X_n) se denomina **realización muestral**.

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

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

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

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

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

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

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

Definición: denominamos a una función T que solamente depende de los valores de una muestra aleatoria X_1, \dots, 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 $Y = T(X_1, \dots, T_n)$ es denominada **distribución muestral**.

2.2. Métodos de muestreo

2.2.1. Muestreo aleatorio simple

2.3. Distribuciones de muestreo

3. Tema 2: Estimación de parámetros

3.1. Definiciones y propiedades de los estimadores

Definiciones:

- Sean X_1, \dots, 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 estadístico¹ T que mejor estima el valor del parámetro para esa población.
- Llamaremos a este estadístico $T = T(X_1, \dots, X_n)$ 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 $T_n = T(X_1, \dots, X_n)$ para el parámetro θ es **consistente** cuando $\forall \epsilon > 0$:

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

Ejemplo de la media aritmética como estimador, usando chebychev's

Teorema: si T_n es una secuencia de estimadores tales que $E(T_n) \longleftrightarrow \theta$ y $V(T_n) \longleftrightarrow 0$ cuando $n \rightarrow \infty$ 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 $bias(T) \neq 0$.

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

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

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

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

ENCONTRAR ESTA MIERDA Y CONTINUAR A PARTIR DE AQUÍ, MUHAHH-HAHHHAHHHAHHHA

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 $n \rightarrow \infty$:

$$E(T_n) \rightarrow \theta$$

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

Es fácil comprobar los siguiente:

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

Para $0 < x < \theta$:

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

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

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

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 estimador T_2 (ambos no desviados) como:

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

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 inferior para la varianza de un estimador no desviado:

Teorema, Cota de Cramer-Rao (CRB): bajo ciertas condiciones de regularidad y siendo X_1, \dots, 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:

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

Obs: Podemos definir la eficiencia 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**, $\mathcal{I}_X(\theta)$:

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

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 $\mathcal{X} = (X_1, \dots, X_n)$ 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 .

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

3.2. Métodos para la obtención de estimadores

Método de los momentos

Método de la máxima verosimilitud

3.3. Métodos de remuestreo

4. Tema 3: Intervalos de confianza

5. Tema 4: Contraste de Hipótesis

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, \dots, X_{p-1} .

En la práctica, además de la variable explicativa X_1 tendremos también otras variables explicativas Z_1, \dots, 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 , $f(X_1, X_2, \dots, X_{p-1}|\beta)$.

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

- Movimiento parabólico

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

- 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**.

$$Y = \text{Señal} + \text{ruido}$$

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 :

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

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**)
- $H_3 : y_i|x_i \approx \text{Normal}$ (**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**) ²

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

■ $V(\epsilon_i) = \sigma^2 \Rightarrow V(y_i|x_i) = \sigma^2$ (**Varianza constante**) ³

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

■ $\epsilon_i \sim N(0, \sigma) \Rightarrow y_i|x_i \approx \text{Normal}$ (**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 las suposiciones con más importancia.

De nuevo otra equivalencia, esto es equivalente:

$Y_i X_i \sim N(\beta_0 + \beta_1 X_1, \sigma)$ 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).

²Propiedades 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.

³Propiedades de la **varianza**:

- $V(X) \geq 0$
- $V(aX + b) = a^2 V(X)$
- $V(X + Y) = V(X) + V(Y) + 2 \cdot \text{Cov}(X, Y)$
- $V(X - Y) = V(X) + V(Y) - 2 \cdot \text{Cov}(X, Y)$
- $V(Y) = E(V(Y|X)) + V(E(Y|X))$ (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 hallar 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:

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

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)

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

Las soluciones de las ecuaciones normales son las siguientes:

(PROBARLO QUEDA PENDIENTE)

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

$$b_1 = \bar{Y} - b_0 \cdot \bar{X} = \frac{\sum_{i=1}^n Y_i - b_0 \cdot \sum_{i=1}^n X_i}{n} = \sum_{i=1}^n C'_i \cdot Y_i$$

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 (\bar{X}, \bar{Y}) y es la siguiente:

$$\hat{Y}_i = \bar{Y} + b_1 \cdot (X_i - \bar{X})$$

6.1.2. Tabla ANOVA

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

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

También es posible demostrar lo siguiente:

HACER LOS CÁLCULOS DETALLADAMENTE...

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

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 .⁴

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

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

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

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

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 .

⁴ $R^2 = (r_{xy})^2$

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 basamos 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:

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

y también:

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

Siendo la covarianza entre ambos:

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

Entonces nos queda:

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

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

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_0 y b_1 no es cero? RESPONDER

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

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

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

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 AQUÍ LOS CÁLCULOS 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 $(1 - \alpha)$.⁵

¿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 $t_1 = \frac{b_1}{S_{b_1}} \sim t\text{-Student}$, 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.

⁵El 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 $|t_1| = \left| \frac{b_1}{S_{b_1}} \right|$ es grande o pequeño.

$$H_0 : \beta_1 = a$$

$$H_1 : \beta_1 \neq a$$

Si H_0 es cierta, entonces $t = \frac{b_1 - a}{S_{b_1}} \sim t\text{-Student}$, 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:

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

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 si** 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

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

Reagrupando parámetros muhahahhhahhha

Notas sobre la probabilidad de equivocarse o acertar en comparaciones 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) = sigma² siempre se cumple muhehehehehe (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 usamos 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 = comparaciones múltiples muhahahahahahahahahahahahahahahahahahah

CASO RANDOM Observemos que lo que comparamos es si los hospitales añaden variabilidad o no

La table y lo demás no varían

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

intraclass correlation = correlación intraclásica

EJERCICIOS

7.1. Tema 1

7.1.1. Métodos de muestreo

7.1.2. Distribuciones de muestreo

7.2. Tema 2

7.2.1. Definiciones

7.2.2. 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 Chevychev: $P(|X - E(X)| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}$

7.2.3. 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, \dots, x_n que sigue una distribución $B(1, p)$.

AQUÍ IRÁ SU SOLUCIÓN MUHHAHHHHHAHHA

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 MUHHAHHHHHAHHA

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 .

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

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

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

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

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

$$L_{\theta}(\theta; x_1, \dots, x_n) = \frac{d}{d\theta} L(\theta; x_1, \dots, x_n) = 0$$

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

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

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

Donde T es el estimador. Observemos que aquí tenemos como estimador (3) y todo se reduce por el teorema a conseguir escribir la función de 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, \dots, x_n una muestra de una distribución normal $N(\mu, \sigma)$. Encontrar el método de los momentos para μ y para σ .

solución

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

solución

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 $B(1, p)$.
- solución

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

- 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 $Var(X) = E(X^2) - E(X)^2 = m_2 - m_1^2$.

$$\Rightarrow \hat{\mu} = \bar{X}_n$$

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

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

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

$$\Rightarrow \hat{b} = 2\bar{X}_n$$

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 $B(1, p)$.

MLE para el 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, \dots, x_n .

Calculamos la función de verosimilitud:

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

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 \hat{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 $\hat{p} = p^*$

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

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

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

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

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

CONTINUAR

- Parámetros μ y σ en una Distribución Normal $N(\mu, \sigma)$.
- Parámetro b en una Distribución Uniforme $U(0, b)$.

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, \dots, X_n variables aleatorias independientes e idénticamente distribuidas 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.

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

Sea X_1, \dots, 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.

7.2.4. Métodos de remuestreo

7.3. Tema 3

7.4. Tema 4

7.5. Tema 5

7.6. Tema 6