Estadística I (notas breves)

<http://www.est.uc3m.es/svegas>

1 Distribuciones en el muestreo

1.1 Muestras aleatorias

- Es necesario seleccionar un subconjunto representativo de la población, al que se denominará muestra.
- Si la muestra está bien escogida es posible inferir característica de la población a partir de los datos.
- Los estimadores son valores que se obtienen a partir de los datos y aproximan el verdadero valor del parámetro que nos interesa estudiar, siempre y cuando la muestra escogida sea representativa de la población.
- Existen diversos procedimientos de muestreo que nos ayudan a conseguir que la muestra sea representativa:
 - muestreo aleatorio simple: una muestra se denomina muestra aleatoria simple si cada elemento de la población tiene la misma probabilidad de ser elegido y los elementos se seleccionan de forma independiente.
 - muestreo estratificado: una muestra se denomina muestra estratificada si los elementos de la población se dividen en clases o estratos y cada uno de éstos es homogéneo (p.ej. edades).
 - muestra por conglomerados: una muestra se denomina muestra por conglomerados si los elementos de la población se dividen en clases, siendo cada clases tan heterogéneas como la población y las clases homogéneas entre sí (p.ej. provincias).
- Nos centraremos en las muestras aleatorias simples o m.a.s. Así, una muestra aleatoria simple es una colección de n variables aleatorias v.a. $(X_1, X_2, \dots, X_n)$ tal que $(X_1, X_2, \dots, X_n)$ son independientes y cada X_i tiene una función de distribución $F_\theta(x)$, siendo θ la colección de parámetros de la distribución que nos interesa conocer.
- Suponemos que la forma de la distribución $F_\theta(x)$ es conocida salvo por θ . Por ejemplo, si X sigue una distribución Normal, $X \sim \mathcal{N}(\mu, \sigma^2)$, los parámetros de interés son $\theta = (\mu, \sigma^2)$. Así, para estimar el parámetro μ escogeremos una m.a.s y calcularemos su media $\bar{x}$.
- Al número n le denominaremos tamaño muestral.
- La función de distribución conjunta de la m.a.s $(X_1, X_2, \dots, X_n)$ es

$$F_\theta(x_1, x_2, \dots, x_n) = F_\theta(x_1) \cdot F_\theta(x_1) \cdots F_\theta(x_n)$$

- La función de densidad conjunta de la m.a.s $(X_1, X_2, \dots, X_n)$ es

$$f_\theta(x_1, x_2, \dots, x_n) = f_\theta(x_1) \cdot f_\theta(x_1) \cdots f_\theta(x_n)$$

- A una realización concreta de una m.a.s $(X_1, X_2, \dots, X_n)$ la denominaremos $(x_1, x_2, \dots, x_n)$

1.2 Estadísticos

- Un estadístico es cualquier función de la muestra, luego también es una variable aleatoria: su valor va cambiando a medida que se obtienen diferentes muestras.
- Denotaremos a los estadísticos por $\hat{\theta} = g(X_1, X_2, \dots, X_n)$

- Estadísticos habituales:

· media muestral: $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ (su correspondiente característica poblacional es la media poblacional o esperanza: $\mu = E(X) = \int x f(x) dx$)

· varianza muestral: $\hat{S}_X^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ (su correspondiente característica poblacional es la varianza poblacional $\sigma^2 = Var(X) = \int (x - \mu)^2 f(x) dx$)

· cuasi-varianza muestral: $S_X^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$

· mínimo: $\hat{\theta} = \min\{X_1, X_2, \dots, X_n\}$

· máximo: $\hat{\theta} = \max\{X_1, X_2, \dots, X_n\}$

· logaritmo: $\hat{\theta} = \frac{1}{n} \sum_{i=1}^n \log X_i$

1.3 Distribución muestral

- La distribución de probabilidad correspondiente a un estadístico se denomina distribución muestral.

- La esperanza, la varianza y la desviación típica de la v.a. media muestral $\bar{X}$ de una m.a.s con $E(X_i) = \mu$ y $Var(X_i) = \sigma^2$ toman la forma siguiente,

$$\begin{aligned} E(\bar{X}) &= \frac{1}{n} \sum_{i=1}^n E(X_i) = \mu \\ Var(\bar{X}) &= \frac{1}{n^2} Var(\sum_{i=1}^n X_i) = \frac{\sigma^2}{n} \\ sd(\bar{X}) &= \frac{\sigma}{\sqrt{n}} \end{aligned}$$

- La esperanza de la v.a. varianza muestral $\hat{S}_X^2$ de una m.a.s con $E(X_i) = \mu$ y $Var(X_i) = \sigma^2$ es

$$E(\hat{S}_X^2) = E\left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2\right) = \frac{n-1}{n} \sigma^2$$

- A medida que aumenta el tamaño de la muestra el error estándar (o desviación típica) disminuye, por lo que cuanto mayor sea el tamaño muestral mejor será la inferencia sobre la media poblacional.

- La distribución Chi-cuadrado con n grados de libertad, $\mathcal{X}_n^2$, se define como la suma de n normales independientes $\mathcal{N}(0, 1)$ al cuadrado,

$$\begin{aligned} \mathcal{X}_n^2 &= \mathcal{N}(0, 1)^2 + \dots + \mathcal{N}(0, 1)^2 \\ E(\mathcal{X}_n^2) &= n \\ Var(\mathcal{X}_n^2) &= 2n \end{aligned}$$

- La distribución t de Student con n grados de libertad, t_n , se define como el cociente de dos distribuciones independientes, una de ellas $\mathcal{N}(0, 1)$ y la otra $\mathcal{X}_n^2$,

$$\begin{aligned} t_n &= \frac{\mathcal{N}(0,1)}{\sqrt{\frac{1}{n} \mathcal{X}_n^2}} \\ E(t_n) &= 0 \\ Var(t_n) &= \frac{n}{n-2} \\ t_n &\xrightarrow{n \rightarrow \infty} \mathcal{N}(0, 1) \end{aligned}$$

- **Teorema Central del Límite (TCL):** Sean $(X_1, X_2, \dots, X_n)$ v.a.i. con $E(X_i) = \mu_i$ y $Var(X_i) = \sigma_i^2$, entonces,

$$i) \frac{\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i}{\sqrt{\sum_{i=1}^n \sigma_i^2}} \xrightarrow{n \rightarrow \infty} \mathcal{N}(0, 1)$$

- ii) si las variables $(X_1, X_2, \dots, X_n)$ son además idénticamente distribuidas (v.a.i.i.d.) con $\mu_i = \mu$ y $\sigma_i^2 = \sigma^2$, entonces $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \xrightarrow{n \rightarrow \infty} \mathcal{N}(0, 1)$

- **Lema de Fisher:** Sean $(X_1, X_2, \dots, X_n)$ v.a.i.i.d $\mathcal{N}(\mu, \sigma^2)$, entonces

$$i) \bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right) \text{ o equivalentemente } \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$$

$$ii) \frac{(n-1)}{\sigma^2} S_X^2 \sim \chi_{n-1}^2$$

$$iii) \bar{X} \text{ e } S_X^2 \text{ son independientes}$$

- **Lema de Student:** Sean $(X_1, X_2, \dots, X_n)$ v.a.i.i.d $\mathcal{N}(\mu, \sigma^2)$, entonces

$$\frac{\bar{X} - \mu}{S_X/\sqrt{n-1}} = \frac{\bar{X} - \mu}{S_X/\sqrt{n}} \sim t_{n-1}$$

Observaciones generales:

- Se suele considerar que n es suficientemente grande si $n > 30$
- La esperanza y la varianza de la v.a. cuasi-varianza muestral S_X^2 de una m.a.s con $E(X_i) = \mu$ y $Var(X_i) = \sigma^2$ toman la forma siguiente

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

$$Var(S_X^2) = \frac{\sigma^4}{(n-1)^2} Var(\chi_{n-1}^2) = \frac{2\sigma^4}{n-1}$$

- Sea α cualquier número entre 0 y 1, definimos

- z_α como el valor de la normal estándar $Z \sim \mathcal{N}(0, 1)$ tal que $P(Z > z_\alpha) = 1 - \Phi(z_\alpha) = \alpha$, siendo Φ la función de distribución de la normal estándar.

- $t_{n,\alpha}$ como el valor de la t de Student t_n tal que $P(t_n > t_{n,\alpha}) = \alpha$

- $\chi_{n,\alpha}^2$ como el valor de la χ_n^2 tal que $P(\chi_n^2 > \chi_{n,\alpha}^2) = \alpha$

2 Estimación puntual

- Inferencias sobre la población basándose en la información que proporciona la muestra. Para ello extraemos una muestra de la población y realizamos una estimación del parámetro de interés.

- Repaso:

- Supondremos $(X_1, \dots, X_n)$ una m.a.s. de una distribución conocida $F_\theta(x)$, salvo por sus parámetros.

- Los estadísticos son funciones de la muestra $\hat{\theta} = g(X_1, \dots, X_n)$ y se utilizan para estimar el verdadero valor del parámetro θ

- Los estadísticos son v.a. y tienen una función de distribución generalmente difícil de obtener.

- Veremos criterios de comparación de estimadores y las propiedades que necesita un estadístico para proporcionar buenas estimaciones.

2.1 Estimadores insesgados

- Un estimador $\hat{\theta}$ de θ se dice que es *insesgado o centrado* si $E(\hat{\theta}) = \theta$
- Un estimador se dice que es *sesgado* cuando no es insesgado, y se define el *sesgo* del estimador como $sesgo(\hat{\theta}) = E(\hat{\theta}) - \theta$
- Nos interesan los estimadores centrados con poca varianza y los estimadores con varianza baja y poco sesgo.

2.2 Error cuadrático medio

- Combina las propiedades del sesgo y la varianza.
- El *error cuadrático medio* ECM de un estimador se define como $ECM(\hat{\theta}) = E((\hat{\theta} - \theta)^2) = Var(\hat{\theta}) + sesgo(\hat{\theta})^2$
- Criterio: “nos quedaremos con el estimador de θ con el menor error cuadrático medio”
- El problema con este criterio es que muchos estimadores no se pueden comparar entre sí, de manera que no podemos seleccionar uno frente al resto.

2.3 Estimadores insesgados de mínima varianza

- Dentro de los estimadores insesgados buscaremos los de mínima varianza (los de mínimo ECM)
- Se dice que el estimador $\hat{\theta}_1$ es *más eficiente* que $\hat{\theta}_2$ si $Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$
- Se define la *eficiencia* de un estimador $\hat{\theta}$ como $efic(\hat{\theta}) = \frac{1}{Var(\hat{\theta})}$
- Se define la *eficiencia relativa* de $\hat{\theta}_2$ con respecto a $\hat{\theta}_1$ como $ER(\hat{\theta}_2|\hat{\theta}_1) = \frac{efic(\hat{\theta}_2)}{efic(\hat{\theta}_1)} = \frac{Var(\hat{\theta}_1)}{Var(\hat{\theta}_2)}$

2.4 Consistencia y robustez

- La consistencia de un estimador nos habla de su comportamiento cuando la muestra es grande.
- Un estimador $\hat{\theta}$ es *consistente* si

$$\begin{aligned} E(\hat{\theta}) &\longrightarrow_{n \rightarrow \infty} \theta \\ Var(\hat{\theta}) &\longrightarrow_{n \rightarrow \infty} 0 \end{aligned}$$

- Así, desde el punto de vista del error cuadrático medio y para muestras de tamaño grande, es deseable emplear estimadores consistentes.
- Un estimador es *robusto* si, para pequeñas modificaciones en el modelo, el estimador continua siendo razonablemente bueno.

3 Estimadores de Máxima Verosimilitud

3.1 Caso discreto

- Supongamos que tenemos una m.a.s. $(X_1, \dots, X_n)$ de una v.a. discreta con distribución $P(x|\theta)$ conocida salvo por los parámetros θ
- Una vez realizada la muestra tenemos $(x_1, \dots, x_n)$ y estamos interesados en encontrar un estimador $\hat{\theta}$ del parámetro θ que maximice la probabilidad de la muestra.
- Para una muestra dada, la función de probabilidad de la muestra es

$$P(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^n P(X_i = x_i) = \prod_{i=1}^n P(x_i|\theta) = l(\theta|(x_1, \dots, x_n)) \equiv l(\theta)$$

y la denominaremos *función de verosimilitud*.

3.2 Caso continuo

- Supongamos ahora que tenemos una m.a.s $(X_1, \dots, X_n)$ de una v.a. *continua* con función de densidad $f(x|\theta)$ conocida salvo por los parámetros θ
- Una vez realizada la muestra tenemos $(x_1, \dots, x_n)$ y estamos interesados en encontrar un estimador $\hat{\theta}$ del parámetro θ que maximice la función de densidad conjunta de la muestra.
- Para una muestra dada, la función de densidad conjunta de la muestra es

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

y la denominaremos *función de verosimilitud*.

3.3 Propiedades

- Así, el método de máxima verosimilitud consistirá en encontrar el valor $\hat{\theta}$ que maximiza la función de verosimilitud $l(\theta)$, que coincide con el valor $\hat{\theta}$ que maximiza $\log l(\theta)$
- El *estimador de máxima verosimilitud* $\hat{\theta}_{MV}$ se calcula igualando la primera derivada de $\log l(\theta)$ a 0 y comprobando que la segunda derivada en sus soluciones es negativa.
- El estimador de máxima verosimilitud $\hat{\theta}_{mv}$ tiene, bajo ciertas condiciones generales, las siguientes propiedades:
 - es *asintóticamente centrado*: a medida que crece el tamaño muestral el sesgo tiende a cero.
 - sigue *asintóticamente* una distribución *normal* con media θ y varianza $\frac{-1}{(\log l(\theta))''}$.
 - es un estimador *asintóticamente eficiente*: el de todos los estimadores asintóticamente centrados, el de máxima verosimilitud tiene menor varianza
 - es *invariante*: si $\hat{\theta}_{mv}$ es el estimador de máxima verosimilitud de θ , entonces $g(\hat{\theta}_{mv})$ será el estimador de máxima verosimilitud de $h(\theta)$, para cualquier función h continua y biyectiva.

3.4 Distribuciones conocidas

- $X \sim \mathcal{B}(p)$ Bernoulli

$$l(p) = \prod_{i=1}^n P(x_i|p) = \prod_{i=1}^n (p^{x_i}(1-p)^{1-x_i}) = p^{\sum_i x_i} (1-p)^{n-\sum_i x_i}$$

$$\log l(p) = \log p^{\sum_i x_i} (1-p)^{n-\sum_i x_i} = (\sum_i x_i) \log p + (n - \sum_i x_i) \log(1-p)$$

$$\hat{p} = \bar{x}$$

- $X \sim \mathcal{B}(k, p)$ Binomial, con k conocido

$$l(p) = \prod_{i=1}^n P(x_i|p) = \prod_{i=1}^n \binom{k}{x_i} p^{x_i} (1-p)^{k-x_i} = p^{\sum_i x_i} (1-p)^{nk-\sum_i x_i} \cdot \left(\prod_{i=1}^n \binom{k}{x_i} \right)$$

$$\log l(p) = p^{\sum_i x_i} (1-p)^{nk-\sum_i x_i} \cdot \log \left(\prod_{i=1}^n \binom{k}{x_i} \right) = cte + (\sum_i x_i) \log p + (nk - \sum_i x_i) \log(1-p)$$

$$\hat{p} = \frac{1}{nk} \sum_i x_i$$

- $X \sim \mathcal{N}(\mu, \sigma^2)$ Normal

$$l(\mu, \sigma^2) = \prod_{i=1}^n f(x_i | \mu, \sigma^2) = \prod_{i=1}^n \left(\frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2\sigma^2}(x_i - \mu)^2} \right) = \frac{1}{(\sqrt{2\pi\sigma})^n} e^{-\frac{1}{2\sigma^2} \sum_i (x_i - \mu)^2}$$

$$\log l(\mu, \sigma^2) = \log \frac{1}{(\sqrt{2\pi\sigma})^n} e^{-\frac{1}{2\sigma^2} \sum_i (x_i - \mu)^2} = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_i (x_i - \mu)^2$$

$$\hat{\mu} = \bar{x}$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum (x_i - \bar{x})^2$$

- $X \sim \exp(\lambda)$ Exponencial

$$l(\lambda) = \prod_{i=1}^n f(x_i | \lambda) = \prod_{i=1}^n (\lambda e^{-\lambda x_i}) = \lambda^n e^{-\lambda \sum_i x_i}$$

$$\log l(\lambda) = \log \lambda^n e^{-\lambda \sum_i x_i} = n \log(\lambda) - \lambda \sum_i x_i$$

$$\hat{\lambda} = \frac{1}{\bar{x}}$$

4 Intervalos de confianza

- Nos interesa dar una medida de la precisión de la estimación que hemos hecho del parámetro. Daremos un rango de valores entre los que debería encontrarse el verdadero valor del parámetro.
- *Intervalo de confianza*: rango de valores entre los que posiblemente se encuentre el verdadero valor del parámetro.
- Un intervalo de confianza tiene una *probabilidad* de 1 (si contiene al parámetro) ó 0 (si no lo contiene), y una *confianza* del 95% de contener al parámetro.

4.1 Intervalos de confianza para una población normal

- Intervalo de confianza para la media de una población normal con varianza conocida: Sea $X \sim \mathcal{N}(\mu, \sigma^2)$ con varianza σ^2 conocida. Entonces $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$, con lo que $P(-z_{\alpha/2} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < z_{\alpha/2}) = 1 - \alpha$. Si despejamos μ obtenemos lo siguiente (siendo $\bar{x}$ una realización particular de $\bar{X}$):

$$IC(\mu)_{(1-\alpha)100\%} = [\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}]$$

- Intervalo de confianza para la media de una población normal con varianza desconocida: Sea $X \sim \mathcal{N}(\mu, \sigma^2)$ con varianza σ^2 desconocida. Entonces $\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$, luego $P(-t_{n-1, \alpha/2} < \frac{\bar{X} - \mu}{S/\sqrt{n}} < t_{n-1, \alpha/2}) = 1 - \alpha$. Si $\bar{x}$ es una realización particular de $\bar{X}$ y s una realización particular de S , obtenemos:

$$IC(\mu)_{(1-\alpha)100\%} = [\bar{x} - t_{n-1, \alpha/2} \frac{s}{\sqrt{n}}, \bar{x} + t_{n-1, \alpha/2} \frac{s}{\sqrt{n}}]$$

- Intervalo de confianza para la varianza de una población normal: Sea $X \sim \mathcal{N}(\mu, \sigma^2)$. Como sabemos que $\frac{n-1}{\sigma^2} S^2 \sim \chi_{n-1}^2$, tenemos que $P(-\chi_{n-1, 1-\alpha/2}^2 < \frac{n-1}{\sigma^2} S^2 < \chi_{n-1, \alpha/2}^2) = 1 - \alpha$, luego, si s es una realización particular de S

$$IC(\sigma^2)_{(1-\alpha)100\%} = \left[\frac{n-1}{\chi_{n-1, \alpha/2}^2} s^2, \frac{n-1}{\chi_{n-1, 1-\alpha/2}^2} s^2 \right]$$

4.2 Intervalos de confianza para dos poblaciones normales

- Extracción de muestras: *datos pareados* (las muestras se eligen por pares y cada par de variables corresponde al mismo elemento de la muestra) o *muestras independientes* (las muestras de cada población son independientes entre sí)
- Intervalo de confianza para la diferencia de medias de dos poblaciones normales basado en datos pareados: Sean $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ e $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$. Entonces la variable $D = X - Y$ es $\mathcal{N}(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2 - 2 \cdot \text{cov}(X, Y))$, con lo que $\frac{\bar{D} - (\mu_1 - \mu_2)}{S_d / \sqrt{n}} \sim t_{n-1}$ y tenemos que $\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{S / \sqrt{n}} \sim t_{n-1}$, luego $P(-t_{n-1, \alpha/2} < \frac{\bar{D} - (\mu_1 - \mu_2)}{S_d / \sqrt{n}} < t_{n-1, \alpha/2}) = 1 - \alpha$. Si $\bar{d}$ es una realización particular de $\bar{D}$ y s_d una realización particular de S_d tenemos:

$$IC(\mu_1 - \mu_2)_{(1-\alpha)100\%} = [\bar{d} - t_{n-1, \alpha/2} \frac{s_d}{\sqrt{n}}, \bar{d} + t_{n-1, \alpha/2} \frac{s_d}{\sqrt{n}}]$$

- Intervalo de confianza para la diferencia de medias de dos poblaciones normales con varianzas conocidas basado en muestras independientes: Sean $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ e $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$, con varianzas σ_1^2 y σ_2^2 conocidas. Sabemos que $\bar{X} - \bar{Y} \sim \mathcal{N}(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$, donde n_i es el tamaño de la muestra i , con lo que $P(-z_{\alpha/2} < \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} < z_{\alpha/2}) = 1 - \alpha$. De esta manera:

$$IC(\mu_1 - \mu_2)_{(1-\alpha)100\%} = [\bar{x} - \bar{y} - z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, \bar{x} - \bar{y} + z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}]$$

- Intervalo de confianza para la diferencia de medias de dos poblaciones normales con varianzas desconocidas pero iguales basado en muestras independientes: Sean $X \sim \mathcal{N}(\mu_1, \sigma^2)$ e $Y \sim \mathcal{N}(\mu_2, \sigma^2)$, con varianzas desconocidas e iguales a σ^2 . En este caso $\bar{X} - \bar{Y} \sim \mathcal{N}(\mu_1 - \mu_2, \sigma^2 \frac{n_1 + n_2}{n_1 n_2})$, con lo que $\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\sigma^2 \frac{n_1 + n_2}{n_1 n_2}}} \sim \mathcal{N}(0, 1)$.

Para crear el intervalo de confianza sustituiremos el valor de σ^2 por una ponderación de las dos cuasi-varianzas muestrales $S^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$, ya que de esta manera obtenemos que $\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{S^2 \frac{n_1 + n_2}{n_1 n_2}}} \sim t_{n-2}$.

Por tanto, $P(-t_{n-2, \alpha/2} < \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{S^2 \frac{n_1 + n_2}{n_1 n_2}}} < t_{n-2, \alpha/2}) = 1 - \alpha$, y así

$$IC(\mu_1 - \mu_2)_{(1-\alpha)100\%} = [\bar{x} - \bar{y} - t_{n-2, \alpha/2} \sqrt{S^2 \frac{n_1 + n_2}{n_1 n_2}}, \bar{x} - \bar{y} + t_{n-2, \alpha/2} \sqrt{S^2 \frac{n_1 + n_2}{n_1 n_2}}]$$

4.3 Intervalos de confianza para muestras grandes

- Supondremos que el tamaño de la muestra es suficientemente grande y así poder utilizar el Teorema Central del Límite.

- Intervalo de confianza para una proporción: como ya hemos visto, la proporción muestral $\hat{p}$ coincide con la media muestral $\bar{x}$, luego por el TCL tenemos que $\frac{\hat{p} - p}{\sqrt{\hat{p}(1-\hat{p})/n}} \sim \mathcal{N}(0, 1)$, con lo que

$$IC(p)_{(1-\alpha)100\%} = [\hat{p} - z_{\alpha/2} \sqrt{\hat{p}(1-\hat{p})/n}, \hat{p} + z_{\alpha/2} \sqrt{\hat{p}(1-\hat{p})/n}]$$

- Intervalo de confianza para la diferencia de dos proporciones: por el TCL tenemos que $\frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\hat{p}_1(1-\hat{p}_1)/n_1 + \hat{p}_2(1-\hat{p}_2)/n_2}} \sim \mathcal{N}(0, 1)$, luego

$$IC(p_1 - p_2)_{(1-\alpha)100\%} = [\hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2} \sqrt{\hat{p}_1(1-\hat{p}_1)/n_1 + \hat{p}_2(1-\hat{p}_2)/n_2}]$$

- Intervalo de confianza para la media de cualquier distribución: por el TCL tenemos que

$$IC(\mu)_{(1-\alpha)100\%} = [\bar{x} - z_{\alpha/2} \frac{s}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \frac{s}{\sqrt{n}}]$$

- Intervalo de confianza para un parámetro cualquiera de cualquier distribución: Sea $\hat{\theta}_{MV}$ el estimador de máxima verosimilitud de θ , entonces, independientemente de la distribución y siendo $L(\hat{\theta})$ la función soporte, tenemos que

$$IC(\theta)_{(1-\alpha)100\%} = [\hat{\theta} - z_{\alpha/2} \sqrt{\frac{-1}{L''(\hat{\theta})}}, \hat{\theta} + z_{\alpha/2} \sqrt{\frac{-1}{L''(\hat{\theta})}}]$$

4.4 Determinación del tamaño muestral

- Hemos visto que el intervalo de confianza para la media μ de una población normal con varianza conocida es $IC(\mu)_{(1-\alpha)100\%} = [\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}]$.
- Así, el tamaño muestral necesario para garantizar una longitud determinada para el intervalo de confianza es $n = (\frac{z_{\alpha/2} \sigma}{A})^2$, siendo A la amplitud del intervalo (def. la mitad de la longitud del intervalo)

5 Contraste de hipótesis

- Nos plantearemos hipótesis y tendremos que rechazarlas o no a partir de la información que proporciona la muestra.
- La hipótesis a validar se denomina *hipótesis nula* H_0 y la enfrentamos a la *hipótesis alternativa* H_1 , dando lugar a un contraste *unilateral* o *bilateral*.
- Si los datos no muestran suficiente evidencia para rechazar la hipótesis nula esta se mantiene como cierta (diremos que no hay suficiente evidencia para rechazarla evitando decir que la aceptamos)
- Por ser las características de la población desconocidas, podemos cometer dos tipos de errores: *error de tipo I* (si decidimos rechazar H_0 cuando es cierta) y *error de tipo II* (si decidimos aceptar H_0 cuando es falsa)
- A la probabilidad de cometer un error de tipo I se le denomina *nivel de significación del contraste* y se denota por $\alpha = P(\text{error tipo I}) = P(\text{rechazar } H_0 | H_0 \text{ cierta})$. Así, la probabilidad de aceptar H_0 cuando es cierta vale $P(\text{aceptar } H_0 | H_0 \text{ cierta}) = 1 - \alpha$
- La probabilidad de cometer un error de tipo II se denota por $\beta = P(\text{error tipo II}) = P(\text{aceptar } H_0 | H_0 \text{ falsa})$. Así, a la probabilidad de rechazar H_0 cuando es falsa se le denomina *potencia del contraste* y vale $P(\text{rechazar } H_0 | H_0 \text{ falsa}) = 1 - \beta$
- El nivel a partir del cual se pasa de aceptar la hipótesis nula H_0 a rechazarla se denomina *nivel crítico del contraste* o *p-valor del contraste*. Así, para cualquier contraste, la regla de decisión es *Rechazar H_0 si p-valor $< \alpha$*
- En la práctica fijaremos la probabilidad de cometer un error de tipo I (nivel de significación) $\alpha \in \{0.1, 0.05, 0.01\}$

5.1 Contrastes para una población normal

- Contrastes para la media de una población normal con varianza conocida: Sea $X_1, \dots, X_n$ una m.a.s de una población normal $\mathcal{N}(\mu, \sigma^2)$ con varianza conocida.

$$\text{Unilateral} \quad \left| \begin{array}{l} H_0 : \mu = \mu_0 \\ H_1 : \mu > \mu_0 \end{array} \right. \implies (*) \quad \text{Rechazar } H_0 \text{ si } \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} > z_\alpha$$

$$\text{Bilateral} \quad \left| \begin{array}{l} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{array} \right. \implies (*) \quad \text{Rechazar } H_0 \text{ si } \frac{|\bar{x} - \mu_0|}{\sigma/\sqrt{n}} > z_{\alpha/2}$$

(*) Sabemos que bajo H_0 (es decir, si H_0 es cierta) $\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$. Así, basta fijar el nivel de significación α para obtener la regla del contraste.

- Contrastes para la media de una población normal con varianza desconocida: Sea $X_1, \dots, X_n$ una m.a.s de una población normal $\mathcal{N}(\mu, \sigma^2)$ con varianza desconocida.

$$\text{Unilateral} \quad \left| \begin{array}{l} H_0 : \mu = \mu_0 \\ H_1 : \mu > \mu_0 \end{array} \right. \implies (*) \quad \text{Rechazar } H_0 \text{ si } \frac{\bar{x} - \mu_0}{s/\sqrt{n}} > t_{n-1, \alpha}$$

$$\text{Bilateral} \quad \left| \begin{array}{l} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{array} \right. \implies (*) \quad \text{Rechazar } H_0 \text{ si } \frac{|\bar{x} - \mu_0|}{s/\sqrt{n}} > t_{n-1, \alpha/2}$$

(*) Bajo H_0 se cumple $\frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1}$

- Contrastes para la varianza de una población normal: Sea $X_1, \dots, X_n$ una m.a.s de una $\mathcal{N}(\mu, \sigma^2)$.

$$\text{Unilateral} \quad \left| \begin{array}{l} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 < \sigma_0^2 \end{array} \right. \implies (*) \quad \text{Rechazar } H_0 \text{ si } \frac{n-1}{\sigma_0^2} s^2 < \chi_{n-1, 1-\alpha}$$

$$\text{Bilateral} \quad \left| \begin{array}{l} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 \neq \sigma_0^2 \end{array} \right. \implies (*) \quad \text{Rechazar } H_0 \text{ si } \left\{ \frac{n-1}{\sigma_0^2} s^2 > \chi_{n-1, \alpha/2} \right. \\ \left. \text{ó } \frac{n-1}{\sigma_0^2} s^2 < \chi_{n-1, 1-\alpha/2} \right\}$$

(*) Bajo H_0 se cumple $\frac{n-1}{\sigma_0^2} S^2 \sim \chi_{n-1}^2$

5.2 Contrastes para dos poblaciones normales

- Contrastes para la diferencia de medias de dos poblaciones normales basado en datos pareados.

$$\text{Unilateral} \quad \left| \begin{array}{l} H_0 : \mu_1 - \mu_2 = d_0 \\ H_1 : \mu_1 - \mu_2 > d_0 \end{array} \right. \implies (*) \quad \text{Rechazar } H_0 \text{ si } \frac{\bar{d} - d_0}{s_D/\sqrt{n}} > t_{n-1, \alpha}$$

$$\text{Bilateral} \quad \left| \begin{array}{l} H_0 : \mu_1 - \mu_2 = d_0 \\ H_1 : \mu_1 - \mu_2 \neq d_0 \end{array} \right. \implies (*) \quad \text{Rechazar } H_0 \text{ si } \frac{|\bar{d} - d_0|}{s_D/\sqrt{n}} > t_{n-1, \alpha/2}$$

(*) Bajo H_0 se cumple que, si definimos $D = X - Y$, tenemos $\frac{\bar{D} - (\mu_1 - \mu_2)}{S_D/\sqrt{n}} \sim t_{n-1}$

- Contrastes para la diferencia de medias de dos poblaciones normales con varianzas conocidas basado en muestras independientes.

$$\text{Unilateral} \quad \left| \begin{array}{l} H_0 : \mu_1 - \mu_2 = d_0 \\ H_1 : \mu_1 - \mu_2 > d_0 \end{array} \right. \implies (*) \quad \text{Rechazar } H_0 \text{ si } \frac{\bar{x} - \bar{y} - d_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} > z_\alpha$$

$$\text{Bilateral} \quad \left| \begin{array}{l} H_0 : \mu_1 - \mu_2 = d_0 \\ H_1 : \mu_1 - \mu_2 \neq d_0 \end{array} \right. \implies (*) \quad \text{Rechazar } H_0 \text{ si } \frac{|\bar{x} - \bar{y} - d_0|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} > z_{\alpha/2}$$

(*) Bajo H_0 se cumple que $\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim \mathcal{N}(0, 1)$

- Contrastes para la diferencia de medias de dos poblaciones normales con varianzas desconocidas pero iguales basado en muestras independientes.

$$\text{Unilateral} \quad \left\{ \begin{array}{l} H_0 : \mu_1 - \mu_2 = d_0 \\ H_1 : \mu_1 - \mu_2 > d_0 \end{array} \right. \implies (*) \quad \text{Rechazar } H_0 \text{ si } \frac{\bar{x} - \bar{y} - d_0}{\sqrt{s_{12}^2 \frac{n_1 + n_2}{n_1 n_2}}} > t_{n_1 + n_2 - 2, \alpha}$$

$$\text{Bilateral} \quad \left\{ \begin{array}{l} H_0 : \mu_1 - \mu_2 = d_0 \\ H_1 : \mu_1 - \mu_2 \neq d_0 \end{array} \right. \implies (*) \quad \text{Rechazar } H_0 \text{ si } \frac{|\bar{x} - \bar{y} - d_0|}{\sqrt{s_{12}^2 \frac{n_1 + n_2}{n_1 n_2}}} > t_{n_1 + n_2 - 2, \alpha/2}$$

(*) Bajo H_0 se cumple que $\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{s_{12}^2 \frac{n_1 + n_2}{n_1 n_2}}} \sim t_{n_1 + n_2 - 2}$ siendo $S_{12}^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$

- Contrastes para la diferencia de varianzas de dos poblaciones normales basado en muestras independientes.

$$\text{Unilateral} \quad \left\{ \begin{array}{l} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 < \sigma_2^2 \end{array} \right. \implies (*) (**) \quad \text{Rechazar } H_0 \text{ si } \frac{s_1^2}{s_2^2} < F_{n_1 - 1, n_2 - 1, 1 - \alpha}$$

$$\text{Bilateral} \quad \left\{ \begin{array}{l} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 \neq \sigma_2^2 \end{array} \right. \implies (*) (**) \quad \text{Rechazar } H_0 \text{ si } \left\{ \frac{s_1^2}{s_2^2} < F_{n_1 - 1, n_2 - 1, 1 - \alpha/2} \right. \\ \left. \text{ó } \frac{s_1^2}{s_2^2} > F_{n_1 - 1, n_2 - 1, \alpha/2} \right\}$$

(*) Bajo H_0 se cumple que $\left(\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}\right) = \frac{S_1^2}{S_2^2} \sim F_{n_1 - 1, n_2 - 1}$

(**) La distribución $F_{n,m}$ se define como $F_{n,m} = \frac{\chi_n^2}{\frac{\chi_m^2}{m}}$ donde χ_n^2 y χ_m^2 son distribuciones chi-cuadrado independientes.

5.3 Contrastes para muestras grandes

- Observación: Especificaremos en estas notas breves sólo los contrastes bilaterales.
- Contraste bilateral para una proporción.

$$\text{Bilateral} \quad \left\{ \begin{array}{l} H_0 : p = p_0 \\ H_1 : p \neq p_0 \end{array} \right. \implies (*) \quad \text{Rechazar } H_0 \text{ si } \frac{|\hat{p} - p_0|}{\sqrt{p_0(1-p_0)/n}} > z_{\alpha/2}$$

(*) Bajo H_0 se cumple que $\frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} \sim \mathcal{N}(0, 1)$

- Contraste bilateral para la diferencia de dos proporciones.

$$\text{Bilateral} \quad \left\{ \begin{array}{l} H_0 : p_1 - p_2 = 0 \\ H_1 : p_1 - p_2 \neq 0 \end{array} \right. \implies (*) \quad \text{Rechazar } H_0 \text{ si } \frac{|\hat{p}_1 - \hat{p}_2|}{\sqrt{q(1-q)/n_1 + q(1-q)/n_2}} > z_{\alpha/2}$$

(*) Bajo H_0 se cumple que $\frac{\hat{p}_1 - \hat{p}_2}{\sqrt{q(1-q)/n_1 + q(1-q)/n_2}} \sim \mathcal{N}(0, 1)$, siendo $q = \frac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n_1 + n_2}$

- Contraste bilateral para la media de cualquier distribución.

$$\text{Bilateral} \quad \left\{ \begin{array}{l} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{array} \right. \implies (*) \quad \text{Rechazar } H_0 \text{ si } \frac{|\bar{x} - \mu_0|}{s/\sqrt{n}} > z_{\alpha/2}$$

(*) Bajo H_0 se cumple que $\frac{\bar{x} - \mu_0}{s/\sqrt{n}} \xrightarrow{n \rightarrow \infty} \mathcal{N}(0, 1)$

- Contraste bilateral para un parámetro θ de cualquier distribución.

$$\text{Bilateral} \quad \left| \begin{array}{l} H_0 : \theta = \theta_0 \\ H_1 : \theta \neq \theta_0 \end{array} \right. \implies (*) \quad \text{Rechazar } H_0 \text{ si } \frac{|\hat{\theta}_{MV} - \theta_0|}{\sqrt{\frac{-1}{L''(\hat{\theta}_{MV})}}} > z_{\alpha/2}$$

(*) Bajo H_0 se cumple que $\hat{\theta}_{MV} \xrightarrow{n \rightarrow \infty} \mathcal{N}(\theta_0, \frac{-1}{L''(\hat{\theta}_{MV})})$

6 Diagnosís del modelo

- Estudiaremos, mediante contrastes de bondad de ajuste, si la distribución que suponemos es adecuada para los datos que tenemos.

6.1 Contraste de la chi-cuadrado χ^2

- Comparamos el histograma muestral con el teórico. Si existe mucha discrepancia entre los valores muestrales y los valores esperados, es razonable pensar que el modelo utilizado no es el correcto.
- En la práctica dividimos la muestra en k clases:

clase	1	...	k	total
frecuencia observada	O_1	...	O_k	n
probabilidad supuesta	p_1	...	p_k	1
frecuencia esperada	$E_1 = np_1$	...	$E_k = np_k$	n

- Así, la regla del contraste es

$$\left| \begin{array}{l} H_0 : X \sim \mathcal{F} \\ H_1 : X \not\sim \mathcal{F} \end{array} \right. \implies (*) \quad \text{Rechazar } H_0 \text{ si } \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \geq \chi_{k-r-1, \alpha}^2$$

(*) Si la muestra es grande o cada $E_i \geq 5$, tenemos que $\sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \sim \chi_{k-r-1}^2$ siendo r el número de parámetros no conocidos (y estimados) del modelo.

6.2 Contraste de Kolmogorov-Smirnov

- Comparamos la distribución muestral con la distribución teórica. Así, si existe mucha discrepancia entre la distribución muestral y la teórica, es razonable pensar que el modelo utilizado no es el correcto.
- Calcularemos el estadístico $\{D_n = \max |F_n(x) - F(x)|\}$ siendo la distribución empírica de la muestra

$$F_n(x) = \begin{cases} 0 & \text{si } x < x_{(1)} \\ \frac{i}{n} & \text{si } x_{(i)} \leq x_{(i+1)} \\ 1 & \text{si } x \geq x_{(n)} \end{cases}$$

donde $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ es la muestra inicial ordenada de menor a mayor.

- Para el contraste

$$\left| \begin{array}{l} H_0 : X \sim \mathcal{F} \\ H_1 : X \not\sim \mathcal{F} \end{array} \right. \implies (*) \quad \text{Rechazar } H_0 \text{ si } D_n \geq D_{n, \alpha}$$

(*) Bajo H_0 se cumple que D_n sigue una distribución perfectamente tabulada

- En la práctica calculamos D_n mediante la siguiente tabla, $D_n = \max\{|D_i^-|, |D_i^+|\}$

...	...	...	...	...
$x_{(i)}$	$F_n(x_{(i)}) = \frac{i}{n}$	$F(x_{(i)})$	$D_i^- = \frac{i}{n} - F(x_{(i)})$	$D_i^+ = F(x_{(i)}) - \frac{i-1}{n}$
...	...	...	...	...

6.3 Contraste de Kolmogorov-Smirnov-Lilliefors

- Caso particular del contraste de Kolmogorov-Smirnov para contrastar normalidad.

$$\left| \begin{array}{l} H_0 : X \sim \mathcal{N}(\mu, \sigma^2) \\ H_1 : X \not\sim \mathcal{N}(\mu, \sigma^2) \end{array} \right. \implies \text{Rechazar } H_0 \text{ si } D_n \geq D_{n,\alpha}^{(*)}$$

(*) siendo $D_{n,\alpha}$ el valor que encontramos en las tablas Kolmogorov-Smirnov-Lilliefors

6.4 Análisis de homogeneidad de la muestra. Tablas de contingencia

- Utilizaremos las tablas de contingencia para ver la posible homogeneidad de la muestra en variables cualitativas. Diremos que una muestra es *heterogénea* si existen al menos dos elementos que no proceden de la misma población.
- Contrastaremos

$$\left| \begin{array}{l} H_0 : \text{no existe asociación entre los atributos} \\ H_1 : \text{existe asociación entre los atributos} \end{array} \right.$$

mediante la siguiente tabla de contingencia $I \times J$

$X Y$	1	...	J	total
1	O_{11}	...	O_{1J}	n_1
...	...	...	...	...
I	O_{I1}	...	O_{IJ}	n_I
total	n_1	...	n_J	n

calculamos el valor esperado en cada celda como $\hat{E}_{ij} = \frac{n_i n_j}{n}$ y si existe mucha discrepancia entre los valores observados y los esperados inferiremos que existe relación entre los atributos.

- Así, la regla del contraste es

$$\left| \begin{array}{l} H_0 : \nexists \text{ asociación} \\ H_1 : \exists \text{ asociación} \end{array} \right. \implies^{(*)} \text{Rechazar } H_0 \text{ si } \sum_{i=1}^I \sum_{j=1}^J \frac{(O_{ij} - \hat{E}_{ij})^2}{\hat{E}_{ij}} \geq \chi_{(I-1)(J-1), \alpha}^2$$

(*) Si la muestra es grande o cada $\hat{E}_{ij} \geq 5$, tenemos que $\sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \sim \chi_{(I-1)(J-1)}^2$