Distribuciones en el muestreo.

Mª Eugenia Cruces, Salvador J. Molina y Mª Dolores Sarrión

UNIVERSIDAD DE MÁLAGA
Departamento de Estadística y Econometría
Parcialmente financiado a través del PIE13-024 (UMA).

4 de julio de 2014

DISTRIBUCIONES EN EL MUESTREO

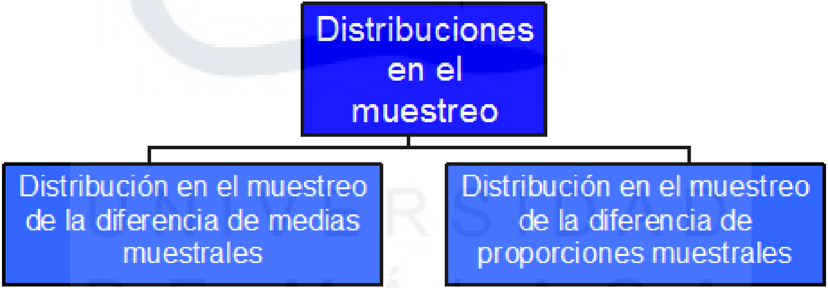
1. Distribuciones en el muestreo.
   - Introducción.
   - Distribución de la media muestral.
   - Distribución de la proporción muestral.
   - Distribución de la varianza muestral.
   - Distribución de la diferencia de medias.
   - Distribución de la diferencia de proporciones.

Lo que vamos a estudiar:

Para una muestra:

Lo que vamos a estudiar:

Para dos muestras:



DISTRIBUCIONES EN EL MUESTREO

Distribución de la media muestral

Estudiamos dos casos:

Distribución de la media muestral

Estudiamos dos casos:

- Población normal
- Población no normal

Distribución de la media muestral

Estudiamos dos casos:

- Población normal
  - Distribución exacta.
- Población no normal

Distribución de la media muestral

Estudiamos dos casos:

- Población normal
    - Distribución exacta.
- Población no normal
    - Principales características
    - Distribución asintótica

Distribución de la media muestral: Población normal

Si la población es normal ($X \sim N(\mu, \sigma)$):

Distribución de la media muestral: Población normal

Si la población es normal ($X \sim N(\mu, \sigma)$):

- $X_1, X_2, \cdots, X_n$ m.a.s. de $X \sim N(\mu, \sigma)$

Distribución de la media muestral: Población normal

Si la población es normal ($X \sim N(\mu, \sigma)$):

- $X_1, X_2, \cdots, X_n$ m.a.s. de $X \sim N(\mu, \sigma)$
- $\overline{X}$ la media muestral de esa muestra,

Distribución de la media muestral: Población normal

Si la población es normal ($X \sim N(\mu, \sigma)$):

- $X_1, X_2, \cdots, X_n$ m.a.s. de $X \sim N(\mu, \sigma)$
- $\overline{X}$ la media muestral de esa muestra, $\overline{X} = \dfrac{\sum_{i=1}^{n} X_i}{n}$

Distribución de la media muestral: Población normal

Si la población es normal ($X \sim N(\mu, \sigma)$):

- $X_1, X_2, \cdots, X_n$ m.a.s. de $X \sim N(\mu, \sigma)$
- $\overline{X}$ la media muestral de esa muestra, $\overline{X} = \dfrac{\sum_{i=1}^{n} X_i}{n}$

Distribución de la media muestral: Población normal

Si la población es normal ($X \sim N(\mu, \sigma)$):

- $X_1, X_2, \cdots, X_n$ m.a.s. de $X \sim N(\mu, \sigma)$
- $\overline{X}$ la media muestral de esa muestra, $\overline{X} = \dfrac{\sum_{i=1}^{n} X_i}{n}$

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

Distribución de la media muestral: Población normal

Si la población es normal ($X \sim N(\mu, \sigma)$):

- $X_1, X_2, \cdots, X_n$ m.a.s. de $X \sim N(\mu, \sigma)$
- $\overline{X}$ la media muestral de esa muestra, $\overline{X} = \dfrac{\sum_{i=1}^{n} X_i}{n}$

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

Se utiliza para estimar por intervalos la media poblacional o verificar hipótesis relativas a la media poblacional (μ) cuando la población es normal y la varianza poblacional (σ^2) es conocida.

Distribución de la media muestral: Población normal

Es decir,

- Si la población es normal con media μ y desviación típica σ, la media muestral para una m.a.s. de tamaño n es también normal con

$$\boxed{\mu_{\overline{X}} = \mu} \quad \text{y} \quad \boxed{\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}}$$

Distribución de la media muestral: Población normal

Es decir,

- Si la población es normal con media μ y desviación típica σ, la media muestral para una m.a.s. de tamaño n es también normal con

$$\boxed{\mu_{\overline{X}} = \mu} \quad \text{y} \quad \boxed{\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}}$$

Error estándar de $\overline{X}$

Distribución de la media muestral: Población normal

Es decir,

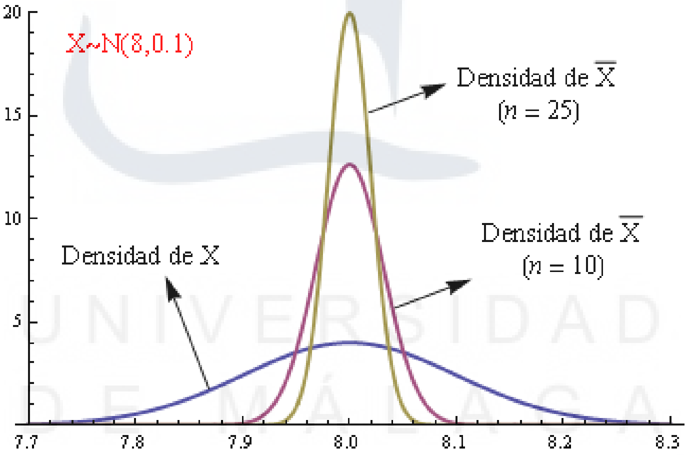
- Si la población es normal con media μ y desviación típica σ, la media muestral para una m.a.s. de tamaño n es también normal con

$$\boxed{\mu_{\overline{X}} = \mu} \quad \text{y} \quad \boxed{\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}}$$

Error estándar de $\overline{X}$

- El error estandar disminuye al aumentar el tamaño de muestra (n)

Distribución de la media muestral: Población normal



Distribución de la media muestral: Principales características

- Incluso cuando la población no es normal:

$$\boxed{\mu_{\overline{X}} = \mu} \quad \text{y} \quad \boxed{\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}}$$

Distribución de la media muestral: Principales características

- Incluso cuando la población no es normal:

$$\mu_{\bar{X}} = \mu \quad \text{y} \quad \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

$\bar{X}$ es insesgada para μ

Teorema Central del Límite: Distribución asintótica de la media muestral

- $\{X_n\}_{n\in\mathbb{N}}$ sucesión de v.a. iid con media μ y varianza σ^2.
- $\overline{X}_n$ la media de las n primeras

Teorema Central del Límite: Distribución asintótica de la media muestral

- $\{X_n\}_{n\in\mathbb{N}}$ sucesión de v.a. iid con media μ y varianza σ^2.
- $\overline{X}_n$ la media de las n primeras

Es decir, definida por

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad n \in \mathbb{N}.$$

Teorema Central del Límite: Distribución asintótica de la media muestral

- $\{X_n\}_{n\in\mathbb{N}}$ sucesión de v.a. iid con media μ y varianza σ^2.
- $\overline{X}_n$ la media de las n primeras

Entonces,

$$\left\{ \frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \right\}_{n\in\mathbb{N}}$$

converge en distribución a una variable aleatoria $N(0,1)$.

Teorema Central del Límite: Distribución asintótica de la media muestral

- $\{X_n\}_{n\in\mathbb{N}}$ sucesión de v.a. iid con media μ y varianza σ^2.
- $\overline{X}_n$ la media de las n primeras

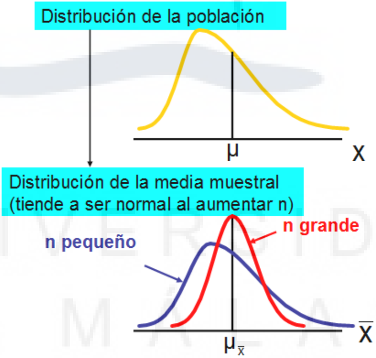
Entonces,

$$\left\{\frac{\overline{X}_n-\mu}{\frac{\sigma}{\sqrt{n}}}\right\}_{n\in\mathbb{N}}$$

converge en distribución a una variable aleatoria $N(0,1)$. Para n grande la media muestral es aproximadamente normal.

Distribución asintótica de la media muestral

- Aunque la población no sea normal



Distribución asintótica de la media muestral

¿Cómo de grande tiene que ser n?

- Para la mayoría de las distribuciones es suficiente $n \geq 25$

DISTRIBUCIONES EN EL MUESTREO

1. Distribuciones en el muestreo.
   - Introducción.
   - Distribución de la media muestral.
   - Distribución de la proporción muestral.
   - Distribución de la varianza muestral.
   - Distribución de la diferencia de medias.
   - Distribución de la diferencia de proporciones.

Distribución de la proporción muestral

La propoción muestral ($\widehat{P}_X$) la hemos definido como

$$\widehat{P}_X = \frac{X_1 + X_2 + \ldots + X_n}{n},$$

Distribución de la proporción muestral

La propoción muestral ($\widehat{P}_X$) la hemos definido como

$$\widehat{P}_X = \frac{X_1 + X_2 + ... + X_n}{n},$$

siendo $X_1, X_2, \cdots, X_n$ una m.a.s. de $X \sim B(1, p)$

Distribución de la proporción muestral

La propoción muestral ($\widehat{P}_X$) la hemos definido como

$$\widehat{P}_X = \frac{X_1 + X_2 + ... + X_n}{n},$$

siendo $X_1, X_2, \cdots, X_n$ una m.a.s. de $X \sim B(1, p)$

$\widehat{P}_X$ es, por lo tanto, una media muestral.

Distribución de la proporción muestral

Aplicando lo estudiado para la distribución de la media:

Distribución de la proporción muestral

Aplicando lo estudiado para la distribución de la media:

$$E\left[\widehat{P}_X\right] = p$$

Distribución de la proporción muestral

Aplicando lo estudiado para la distribución de la media:

$$E\left[\widehat{P}_X\right] = p$$

$$Var\left(\widehat{P}_X\right) = \frac{p(1-p)}{n}$$

Distribución de la proporción muestral

Aplicando lo estudiado para la distribución de la media:

$$E\left[\widehat{P}_X\right] = p$$

$$Var\left(\widehat{P}_X\right) = \frac{p(1-p)}{n}$$

La desviación típica de $\widehat{P}_X$ (error típico) es $\sigma_{\widehat{P}_X} = \sqrt{\frac{p(1-p)}{n}}$

Distribución de la proporción muestral

Aplicando lo estudiado para la distribución de la media:

$$E\left[\widehat{P}_X\right] = p$$

La desviación típica de $\widehat{P}_X$ (error típico) es $\sigma_{\widehat{P}_X} = \sqrt{\dfrac{p(1-p)}{n}}$ y, por el Teorema Central del Límite, para n grande,

Distribución de la proporción muestral

Aplicando lo estudiado para la distribución de la media:

$$E\left[\widehat{P}_X\right] = p$$

La desviación típica de $\widehat{P}_X$ (error típico) es $\sigma_{\widehat{P}_X} = \sqrt{\dfrac{p(1-p)}{n}}$ y, por el Teorema Central del Límite, para n grande,

$$Z = \frac{\widehat{P}_X - p}{\sigma_{\widehat{P}_X}} = \frac{\widehat{P}_X - p}{\sqrt{\dfrac{p(1-p)}{n}}} \approx N(0,1)$$

Distribución de la proporción muestral

En la estimación de la proporción poblacional (p), el error típico

Distribución de la proporción muestral

En la estimación de la proporción poblacional (p), el error típico

$$\sigma_{\widehat{P}_x} = \sqrt{\frac{p(1-p)}{n}}$$

Distribución de la proporción muestral

En la estimación de la proporción poblacional (p), el error típico

$$\sigma_{\widehat{P}_X} = \sqrt{\frac{p(1-p)}{n}}$$

se desconoce.

Distribución de la proporción muestral

En la estimación de la proporción poblacional (p), el error típico se desconoce. Un estimador de $\sigma_{\widehat{P}_X}$ es

Distribución de la proporción muestral

En la estimación de la proporción poblacional (p), el error típico se desconoce. Un estimador de $\sigma_{\widehat{P}_X}$ es

$$\widehat{\sigma}_{\widehat{P}_X} = \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

Distribución de la proporción muestral

En la estimación de la proporción poblacional (p), el error típico se desconoce. Un estimador de $\sigma_{\widehat{P}_X}$ es

$$\widehat{\sigma}_{\widehat{P}_X} = \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

Para n grande,

Distribución de la proporción muestral

En la estimación de la proporción poblacional (p), el error típico se desconoce. Un estimador de $\sigma_{\widehat{P}_X}$ es

$$\widehat{\sigma}_{\widehat{P}_X} = \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

Para n grande,

$$Z = \frac{\widehat{P}_X - p}{\widehat{\sigma}_{\widehat{P}_X}} = \frac{\widehat{P}_X - p}{\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}} \approx N(0,1)$$

DISTRIBUCIONES EN EL MUESTREO

1. Distribuciones en el muestreo.
   - Introducción.
   - Distribución de la media muestral.
   - Distribución de la proporción muestral.
   - Distribución de la varianza muestral.
   - Distribución de la diferencia de medias.
   - Distribución de la diferencia de proporciones.

Distribución de la varianza muestral

$X_1, X_2, \cdots, X_n$ m.a.s. de X con $E[X] = \mu$ y $Var(X) = \sigma^2$.

Distribución de la varianza muestral

$X_1, X_2, \cdots, X_n$ m.a.s. de X con $E[X] = \mu$ y $Var(X) = \sigma^2$.

La varianza muestral (S_X^2) la hemos definido como

Distribución de la varianza muestral

$X_1, X_2, \cdots, X_n$ m.a.s. de X con $E[X] = \mu$ y $Var(X) = \sigma^2$.

La varianza muestral (S_X^2) la hemos definido como

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

Distribución de la varianza muestral

$X_1, X_2, \cdots, X_n$ m.a.s. de X con $E[X] = \mu$ y $Var(X) = \sigma^2$.

La varianza muestral (S_X^2) la hemos definido como

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

La media de la distribución en el muestreo de S_X^2 es

Distribución de la varianza muestral

$X_1, X_2, \cdots, X_n$ m.a.s. de X con $E[X] = \mu$ y $Var(X) = \sigma^2$.

La varianza muestral (S_X^2) la hemos definido como

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

La media de la distribución en el muestreo de S_X^2 es

$$E\left[S_X^2\right] = \frac{n-1}{n}\sigma^2.$$

Distribución de la varianza muestral

$X_1, X_2, \cdots, X_n$ m.a.s. de X con $E[X] = \mu$ y $Var(X) = \sigma^2$.

La varianza muestral (S_X^2) la hemos definido como

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

$$E\left[S_X^2\right] = \frac{n-1}{n}\sigma^2.$$

La varianza muestral es un estimador sesgado de σ^2.

Distribución de la varianza muestral

Las inferencias relativas a la varianza poblacional las basamos en la cuasivarianza muestral ($\widehat{S}_X^2$).

Distribución de la varianza muestral

Las inferencias relativas a la varianza poblacional las basamos en la cuasivarianza muestral ($\widehat{S}_X^2$).

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

Distribución de la varianza muestral

Las inferencias relativas a la varianza poblacional las basamos en la cuasivarianza muestral ($\widehat{S}_X^2$).

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

Distribución de la varianza muestral

Las inferencias relativas a la varianza poblacional las basamos en la cuasivarianza muestral ($\widehat{S}_X^2$).

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

La cuasivarianza muestral es un estimador insesgado de σ^2

Distribución de la varianza muestral

Las inferencias relativas a la varianza poblacional las basamos en la cuasivarianza muestral ($\widehat{S}_X^2$).

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

La cuasivarianza muestral es un estimador insesgado de σ^2

$$E\left[\widehat{S}_X^2\right] = \sigma^2$$

Distribución de la varianza muestral (Población normal)

Si la población es normal, es decir,

Distribución de la varianza muestral (Población normal)

Si $X_1, X_2, \cdots, X_n$ es una m.a.s. de $X \sim N(\mu, \sigma)$,

Distribución de la varianza muestral (Población normal)

Si $X_1, X_2, \cdots, X_n$ es una m.a.s. de $X \sim N(\mu, \sigma)$,

- La varianza y la cuasivarianza muestral están relacionadas con σ^2 a través de la distribución χ^2:

Distribución de la varianza muestral (Población normal)

Si $X_1, X_2, \cdots, X_n$ es una m.a.s. de $X \sim N(\mu, \sigma)$,

- La varianza y la cuasivarianza muestral están relacionadas con σ^2 a través de la distribución χ^2:

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

sigue una distribución χ^2 con n-1 grados de libertad.

Distribución Chi-cuadrado o Ji-dos

Recordemos que:

Distribución Chi-cuadrado o Ji-dos

Recordemos que:

- $Z \sim N(0, 1) \quad \Longrightarrow \quad Z^2 \sim \chi_1^2$

Distribución Chi-cuadrado o Ji-dos

Recordemos que:

- $Z \sim N(0,1) \quad \Longrightarrow \quad Z^2 \sim \chi^2_1$
- Z_1, Z_2, ...,Z_n variables iid $N(0,1)$ y X está definida por

$$X = Z_1^2 + Z_2^2 + \cdots Z_n^2.$$

Distribución Chi-cuadrado o Ji-dos

Recordemos que:

- $Z \sim N(0,1) \quad \Longrightarrow \quad Z^2 \sim \chi^2_1$
- Z_1, Z_2, ...,Z_n variables iid $N(0,1)$ y X está definida por

$$X = Z_1^2 + Z_2^2 + \cdots Z_n^2.$$

Entonces,

$$X \sim \chi^2_n$$

Distribución de S_μ^2

$X_1, X_2, \cdots, X_n$ m.a.s. de $X \sim N(\mu, \sigma)$.

Distribución de S_μ^2

$X_1, X_2, \cdots, X_n$ m.a.s. de $X \sim N(\mu, \sigma)$.

$$S_\mu^2 = \frac{\sum_{i=1}^{n} (X_i - \mu)^2}{n}.$$

Distribución de S_μ^2

$X_1, X_2, \cdots, X_n$ m.a.s. de $X \sim N(\mu, \sigma)$.

$$S_\mu^2 = \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{n}.$$

Entonces

$$\frac{nS_\mu^2}{\sigma^2} \sim \chi_n^2.$$

Distribución de S_μ^2

$X_1, X_2, \cdots, X_n$ m.a.s. de $X \sim N(\mu, \sigma)$.

$$S_\mu^2 = \frac{\sum_{i=1}^{n} (X_i - \mu)^2}{n}.$$

Entonces

$$\frac{nS_\mu^2}{\sigma^2} \sim \chi_n^2.$$

Para hacer inferencias sobre σ^2 cuando μ es conocida

La relación entre las distribuciones de $\overline{X}$ y S_X^2 a través de la distribución t_{n-1}

$X_1, X_2, \cdots, X_n$ m.a.s. de $X \sim N(\mu, \sigma)$ y $\overline{X}$ y S_X^2 la media y la varianza muestral.

La relación entre las distribuciones de $\overline{X}$ y S_X^2 a través de la distribución t_{n-1}

$X_1, X_2, \cdots, X_n$ m.a.s. de $X \sim N(\mu, \sigma)$ y $\overline{X}$ y S_X^2 la media y la varianza muestral. Entonces

$$\frac{\overline{X} - \mu}{\frac{S_X}{\sqrt{n-1}}} \sim t_{n-1}.$$

La relación entre las distribuciones de $\overline{X}$ y S_X^2 a través de la distribución t_{n-1}

$X_1, X_2, \cdots, X_n$ m.a.s. de $X \sim N(\mu, \sigma)$ y $\overline{X}$ y S_X^2 la media y la varianza muestral. Entonces

$$\frac{\overline{X} - \mu}{\frac{S_X}{\sqrt{n-1}}} \sim t_{n-1}.$$

Se utiliza para hacer inferencias sobre μ cuando σ^2 es desconocida

DISTRIBUCIONES EN EL MUESTREO

1. Distribuciones en el muestreo.
   - Introducción.
   - Distribución de la media muestral.
   - Distribución de la proporción muestral.
   - Distribución de la varianza muestral.
   - Distribución de la diferencia de medias.
   - Distribución de la diferencia de proporciones.

Distribución de la diferencia de medias muestrales

$X_1, X_2, \cdots, X_n$ e $Y_1, Y_2, \cdots, Y_m$ muestras aleatorias simples y tomadas con independencia de $X \sim N(\mu_X, \sigma_X)$ e $Y \sim N(\mu_Y, \sigma_Y)$, respectivamente.

Distribución de la diferencia de medias muestrales

$X_1, X_2, \cdots, X_n$ e $Y_1, Y_2, \cdots, Y_m$ muestras aleatorias simples y tomadas con independencia de $X \sim N(\mu_X, \sigma_X)$ e $Y \sim N(\mu_Y, \sigma_Y)$, respectivamente. Entonces

$$\overline{X} - \overline{Y} \sim N\left(\mu_X - \mu_Y, \sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}\right).$$

Distribución de la diferencia de medias muestrales

$X_1, X_2, \cdots, X_n$ e $Y_1, Y_2, \cdots, Y_m$ muestras aleatorias simples y tomadas con independencia de $X \sim N(\mu_X, \sigma_X)$ e $Y \sim N(\mu_Y, \sigma_Y)$, respectivamente. Entonces

$$\overline{X} - \overline{Y} \sim N\left(\mu_X - \mu_Y, \sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}\right).$$

Se utiliza para hacer inferencias sobre la diferencia de medias poblacionales $\mu_X - \mu_Y$ cuando las varianzas poblacionales son conocidas

Consecuencia

$X_1, X_2, \cdots, X_n$ e $Y_1, Y_2, \cdots, Y_m$ muestras aleatorias simples y tomadas con independencia de $X \sim N(\mu_X, \sigma)$ e $Y \sim N(\mu_Y, \sigma)$, respectivamente.

Consecuencia

$X_1, X_2, \cdots, X_n$ e $Y_1, Y_2, \cdots, Y_m$ muestras aleatorias simples y tomadas con independencia de $X \sim N(\mu_X, \sigma)$ e $Y \sim N(\mu_Y, \sigma)$, respectivamente. Entonces

$$\frac{\overline{X} - \overline{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{nS_X^2 + mS_Y^2}{n+m-2}}\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n+m-2}.$$

Consecuencia

$X_1, X_2, \cdots, X_n$ e $Y_1, Y_2, \cdots, Y_m$ muestras aleatorias simples y tomadas con independencia de $X \sim N(\mu_X, \sigma)$ e $Y \sim N(\mu_Y, \sigma)$, respectivamente. Entonces

$$\frac{\overline{X} - \overline{Y} - (\mu_X - \mu_Y)}{\sqrt{\dfrac{nS_X^2 + mS_Y^2}{n+m-2}}\sqrt{\dfrac{1}{n} + \dfrac{1}{m}}} \sim t_{n+m-2}.$$

Se utiliza para hacer inferencias sobre la diferencia de medias poblacionales $\mu_X - \mu_Y$ cuando las varianzas poblacionales son desconocidas, pero iguales

DISTRIBUCIONES EN EL MUESTREO

1. Distribuciones en el muestreo.
   - Introducción.
   - Distribución de la media muestral.
   - Distribución de la proporción muestral.
   - Distribución de la varianza muestral.
   - Distribución de la diferencia de medias.
   - Distribución de la diferencia de proporciones.

Distribución asintótica de la diferencia de proporciones

$X_1, X_2, \cdots, X_n$ e $Y_1, Y_2, \cdots, Y_m$ muestras aleatorias simples y tomadas con independencia de $X \sim B(1, p_X)$ e $Y \sim B(1, p_Y)$, respectivamente.

Distribución asintótica de la diferencia de proporciones

$X_1, X_2, \cdots, X_n$ e $Y_1, Y_2, \cdots, Y_m$ muestras aleatorias simples y tomadas con independencia de $X \sim B(1, p_X)$ e $Y \sim B(1, p_Y)$, respectivamente. $\widehat{P}_X$ y $\widehat{P}_Y$ las proporciones muestrales correspondientes a dichas muestras.

Distribución asintótica de la diferencia de proporciones

$X_1, X_2, \cdots, X_n$ e $Y_1, Y_2, \cdots, Y_m$ muestras aleatorias simples y tomadas con independencia de $X \sim B(1, p_X)$ e $Y \sim B(1, p_Y)$, respectivamente. $\widehat{P}_X$ y $\widehat{P}_Y$ las proporciones muestrales correspondientes a dichas muestras. Entonces

$$\frac{\widehat{P}_X - \widehat{P}_Y - (p_X - p_Y)}{\sqrt{\frac{p_X(1-p_X)}{n} + \frac{p_Y(1-p_Y)}{m}}} \approx_{n,m\to\infty} N(0,1)$$

Consecuencia

$X_1, X_2, \cdots, X_n$ e $Y_1, Y_2, \cdots, Y_m$ muestras aleatorias simples y tomadas con independencia de $X \sim B(1, p_X)$ e $Y \sim B(1, p_Y)$, respectivamente.

Consecuencia

$X_1, X_2, \cdots, X_n$ e $Y_1, Y_2, \cdots, Y_m$ muestras aleatorias simples y tomadas con independencia de $X \sim B(1, p_X)$ e $Y \sim B(1, p_Y)$, respectivamente. $\widehat{P}_X$ y $\widehat{P}_Y$ las proporciones muestrales correspondientes a dichas muestras.

Consecuencia

$X_1, X_2, \cdots, X_n$ e $Y_1, Y_2, \cdots, Y_m$ muestras aleatorias simples y tomadas con independencia de $X \sim B(1, p_X)$ e $Y \sim B(1, p_Y)$, respectivamente. $\widehat{P}_X$ y $\widehat{P}_Y$ las proporciones muestrales correspondientes a dichas muestras. Entonces

$$\frac{\widehat{P}_X - \widehat{P}_Y - (p_X - p_Y)}{\sqrt{\dfrac{\widehat{P}_X(1-\widehat{P}_X)}{n} + \dfrac{\widehat{P}_Y(1-\widehat{P}_Y)}{m}}} \approx_{n,m\to\infty} N(0,1)$$

Consecuencia

$X_1, X_2, \cdots, X_n$ e $Y_1, Y_2, \cdots, Y_m$ muestras aleatorias simples y tomadas con independencia de $X \sim B(1, p_X)$ e $Y \sim B(1, p_Y)$, respectivamente. $\widehat{P}_X$ y $\widehat{P}_Y$ las proporciones muestrales correspondientes a dichas muestras. Entonces

$$\frac{\widehat{P}_X - \widehat{P}_Y - (p_X - p_Y)}{\sqrt{\dfrac{\widehat{P}_X(1-\widehat{P}_X)}{n} + \dfrac{\widehat{P}_Y(1-\widehat{P}_Y)}{m}}} \approx_{n,m\to\infty} N(0,1)$$

Para estimar por intervalos la diferencia de proporciones $p_X - p_Y$ a partir de muestras independientes de tamaños grandes