Universidad Rey Juan Carlos

Tema 1. Estadísticos y sus distribuciones

Estadística II. Grado de Economía y ADE

Dr. Francisco Rabadán Pérez

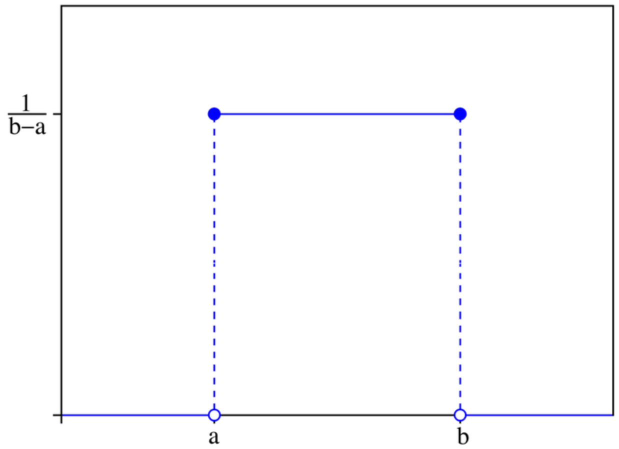
Universidad Rey Juan Carlos

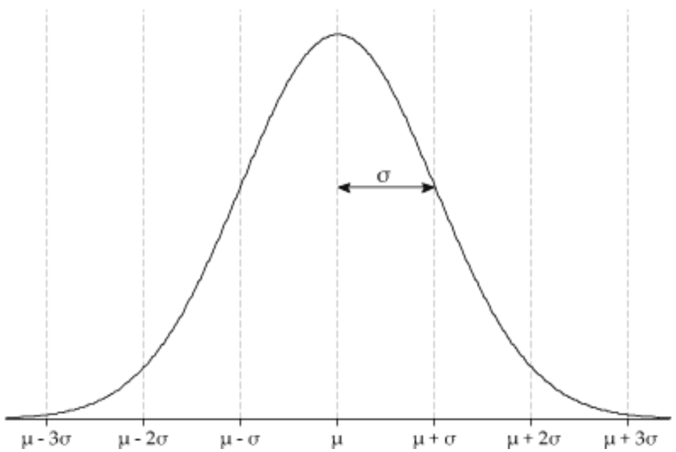
Esquema

1. Parámetros poblacionales
2. Distribución conjunta de la muestra
3. Estadístico: concepto
4. Distribución de estadísticos en el muestreo
   1. $N(\mu, \sigma)$ siendo σ conocida
      1. Distribución de $\bar{x}$
      2. Distribución de S^2 y S_1^2
      3. Distribución de $\bar{x}$ - $\bar{y}$
   2. $N(\mu, \sigma)$ siendo σ desconocida
      1. Distribución de $\bar{x}$
      2. Distribución de $\bar{x}$ - $\bar{y}$
   3. $B(1, \pi)$ siendo σ desconocida
      1. Distribución de p
      2. Distribución de p_1 -p_2

1. Parámetros poblacionales

- Los parámetros poblacionales "θ" son las características numéricas de una población.
  - Definen la **forma** de la distribución, el valor de la **esperanza**, **varianza**, etc.
  - Describen de forma parcial o total la **función de probabilidad** de la variable poblacional que estamos estudiando.
- Ya hemos visto distribuciones de probabilidad paramétricas, ahora definen poblaciones:
  - $\mathrm{N}(\mu, \sigma)$,
  - $Poisson(\lambda)$,
  - $U[a, b]$, …

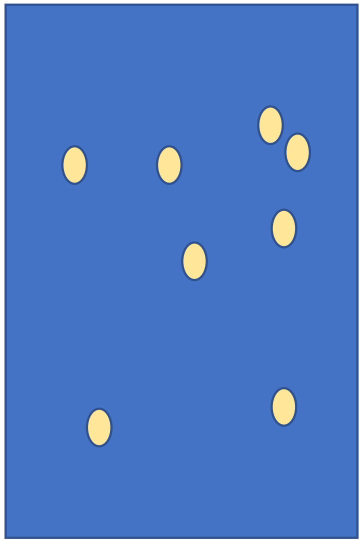




2. Distribución conjunta de la muestra

- **El problema**: conocemos la forma de la distribución de probabilidad de la población, pero ésta depende de un(os) parámetro(s) desconocido(s): θ

$$\xi, población(\theta) \xrightarrow{m.a.s.(n)} muestra\ X = \{x_1, x_2, \ldots, x_n\}$$



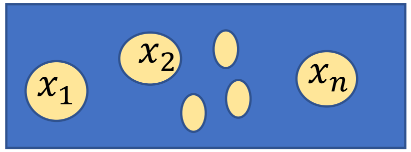
MUESTREO ALEATORIO SIMPLE de tamaño n. Garantiza: x_i son V.A.I.I.D

1. $\boldsymbol{x_i}$ **se distribuye como** $\boldsymbol{\xi}$: misma función de probabilidad.

2. x_i **independiente** de $x_j \quad \forall i \neq j$

3. $\mathbf{E}(\boldsymbol{x_i}) = \mathrm{E}(\xi) = \boldsymbol{\mu}$

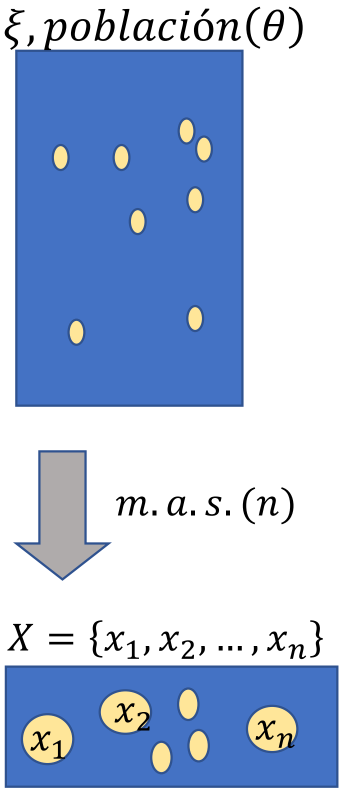
4. $\mathbf{V}(\boldsymbol{x_i}) = \mathrm{V}(\xi) = \boldsymbol{\sigma}^2$



$\boldsymbol{x_i}$ **es v.a.** $\forall \boldsymbol{i} = \mathbf{1, 2, \ldots,} \boldsymbol{n}$

Universidad Rey Juan Carlos

2. Distribución conjunta de la muestra



A partir de ahora designamos las funciones muestrales conjuntas como

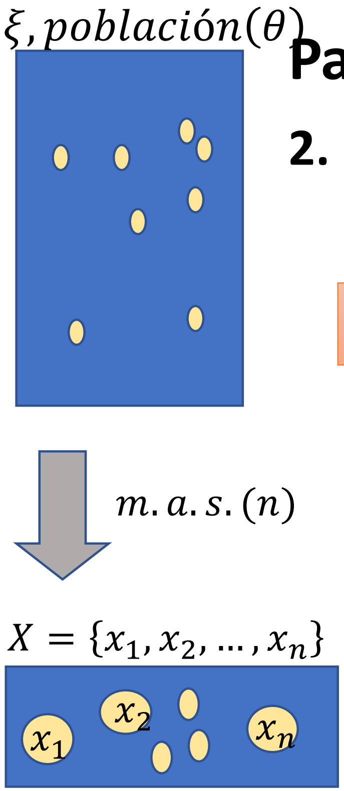
$$\boldsymbol{L(X,\theta) = L(x_1, x_2, \ldots, x_n; \theta)}$$

a) **Si $\boldsymbol{\theta}$ es fijo y $\boldsymbol{X}$ depende** de la extracción aleatoria m.a.s(n), entonces $L(X,\theta)$ es **función de probabilidad conjunta**,

b) **Si $\boldsymbol{X}$es fijo** (tenemos una muestra concreta de tamaño n) **y $\boldsymbol{\theta}$ es desconocido**, entonces, $L(X,\theta)$ es **función de verosimilitud conjunta**.

[no podemos hablar de función de probabilidad de θ porque θ no es V.A.]

2. Distribución conjunta de la muestra



Para $X = (x_1, x_2, \dots, x_n)$ obtenida por m.a.s. (n):

2. **Si ξ es V.A. CONTINUA, su función de probabilidad es una <u>función de densidad</u> $f(x)$. Y la probabilidad de ocurrencia de la muestra es**

$$\mathrm{P(X)} = \mathrm{f}(x_1, x_2, \dots, x_n)dx_1 dx_2 \dots dx_n = \mathrm{f}(x_1)dx_1 * \mathrm{f}(x_2)dx_2 * \cdots * \mathrm{f}(x_n)dx_n$$

$x_i\ indep$

es función de densidad conjunta de la muestra

$$\mathrm{f}(x_1, x_2, \dots, x_n)dx_1 dx_2 \dots dx_n$$

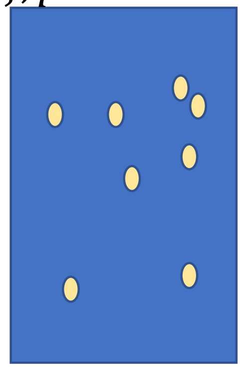
producto de las funciones de densidad marginales

$$\mathrm{f}(x_1)dx_1 * \mathrm{f}(x_2)dx_2 * \cdots * \mathrm{f}(x_n)dx_n$$

¿ Qué pasaría en un muestreo aleatorio sin reemplazamiento?

2. Distribución conjunta de la muestra

$\xi, población(\theta)$

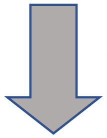


Para $\boldsymbol{X} = (\boldsymbol{x_1}, \boldsymbol{x_2}, \dots, \boldsymbol{x_n})$ obtenida por m.a.s. (n):

1. **Si ξ es V.A. DISCRETA**, su función de probabilidad es una **<u>función de cuantía</u>** $\mathrm{P}(\xi = X)$. Y la probabilidad de ocurrencia de la muestra es

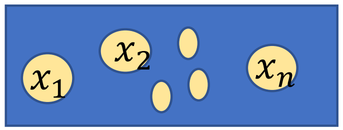
$$\mathrm{P(X)} = \mathrm{P}(x_1 \cap x_2 \cap \dots \cap x_n) = \mathrm{P}(\xi = x_1) * \mathrm{P}(\xi = x_2) * \dots * \mathrm{P}(\xi = x_n)$$

$x_i \; indep$



$m.a.s.(n)$

$X = \{x_1, x_2, \dots, x_n\}$



es función de cuantía conjunta de la muestra

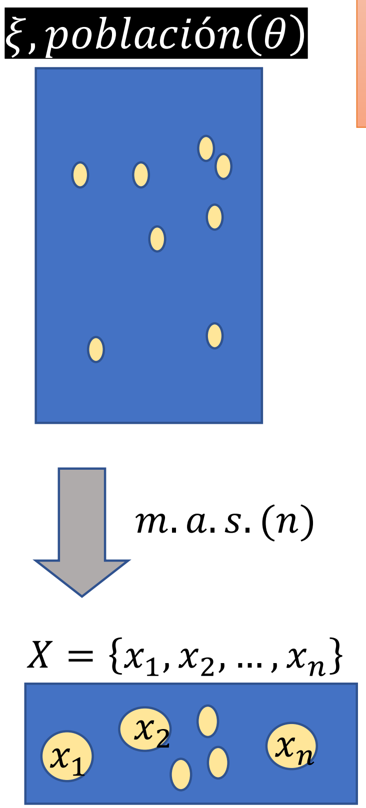
$$\mathrm{P(X)} = \mathrm{P}(x_1 \cap x_2 \cap \dots \cap x_n)$$

Es producto de las funciones de cuantía marginales

$$\mathrm{P}(\xi = x_1) * \mathrm{P}(\xi = x_2) * \dots * \mathrm{P}(\xi = x_n)$$

Universidad Rey Juan Carlos

3. ESTADÍSTICO $T(X) = \mathrm{T}(x_1, x_2, \dots, x_n)$



T(x): es cualquier función real de los elementos muestrales
(*no puede contener parámetros desconocidos*).

	Parámetro poblacional	Estadístico muestral
Suma	Masa de ξ (población)	T(x)= $x_1 + x_2 + \cdots + x_n = \sum_{i=1}^{n} x_i = n * a_1$
Suma de cuadrados	Masa cuadrática de ξ	T(x)= $x_1^2 + x_2^2 + \cdots + x_n^2 = \sum_{i=1}^{n} x_i^2 = n * a_2$
Media	$E(\xi) = \mu = \alpha_1$	T(X)=$a_1 = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$
Varianza	$V(\xi) = \sigma^2 = \alpha_2 - \alpha_1^2$	T(X)=$s_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \left(\frac{1}{n}\sum_{i=1}^{n}(x_i)^2\right) - \bar{x}^2$ $= a_2 - a_1^2$
Cuasivarianza	$V_1(\xi) = \frac{n}{n-1}\sigma^2$	T(X)=$s_{1x}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{n}{n-1} s_x^2$
Proporción	P o ∏ : proporción de éxitos poblacional	P o π: proporción de éxitos en un m.a.s.(n)

Universidad Rey Juan Carlos

3. ESTADÍSTICO $T(X) = \mathrm{T}(x_1, x_2, \dots, x_n)$

La distribución en probabilidad de un estadístico T(x) recibe el nombre de **distribución de probabilidad en el muestreo**

Para cualquier distribución: sea ξ población con $E(\xi) = \mu$ y $V(\xi) = \sigma^2$

- $\boldsymbol{E(\overline{x}) = \mu}$ porque

$$E(\bar{x}) = E\left(\frac{x_1 + x_2 + \cdots + x_n}{n}\right) = \frac{1}{n} n\boldsymbol{E}(\boldsymbol{x_i}) = \frac{1}{n} n\boldsymbol{\mu} = \mu$$

- $\boldsymbol{V(\overline{x}) = \frac{\sigma^2}{n}}$ porque

$$V(\bar{x}) = V\left(\frac{x_1 + x_2 + \cdots + x_n}{n}\right) = \frac{1}{n^2} n\boldsymbol{V}(\boldsymbol{x_i}) = \frac{1}{n^2} n\boldsymbol{\sigma}^2 = \frac{\sigma^2}{n}$$

Universidad Rey Juan Carlos

3. ESTADÍSTICO $T(X) = \mathrm{T}(x_1, x_2, \ldots, x_n)$

La distribución en probabilidad de un estadístico T(x) recibe el nombre de **distribución de probabilidad en el muestreo**

Para cualquier distribución: sea ξ población con $E(\xi) = \mu$ y $V(\xi) = \sigma^2$

$$\boldsymbol{s^2 = \frac{\sum_i (x_i - \overline{x})^2}{n}}$$

- $\boldsymbol{E(s^2) = \frac{n-1}{n}\sigma^2}$ porque

$$E(\boldsymbol{s^2}) = E\left[\frac{\sum(x_i - \mu + \mu - \overline{\boldsymbol{x}})^2}{n}\right] = \frac{1}{n}E\left[\sum(x_i - \mu)^2 - \sum(\bar{x} - \mu)^2\right] =$$

$$= \frac{1}{n}\left[\sum E[(x_i - \mu)^2] - \sum E[(\bar{x} - \mu)^2]\right] = \frac{1}{n}\left[n\sigma^2 - \mathrm{n}\frac{\sigma^2}{\mathrm{n}}\right] = \sigma^2 - \frac{\sigma^2}{\mathrm{n}} = \boldsymbol{\frac{n-1}{n}\sigma^2}$$

Universidad Rey Juan Carlos

3. ESTADÍSTICO $T(X) = \mathrm{T}(x_1, x_2, \dots, x_n)$

La distribución en probabilidad de un estadístico T(x) recibe el nombre de **distribución de probabilidad en el muestreo**

Para cualquier distribución: sea ξ población con $E(\xi) = \mu$ y $V(\xi) = \sigma^2$

$$\boldsymbol{s_1^2} = \frac{\boldsymbol{\sum_i (x_i - \overline{x})^2}}{\boldsymbol{n-1}} = \frac{\mathrm{n}}{\mathrm{n}-1} \frac{\sum_\mathrm{i}(\mathrm{x_i} - \overline{\mathrm{x}})^2}{\mathrm{n}} = \frac{\boldsymbol{n}}{\boldsymbol{n-1}} \boldsymbol{s^2}$$

- $\boldsymbol{E\left(s_1^2\right) = \sigma^2}$ porque

$$\boldsymbol{E\left(s_1^2\right)} = \frac{\mathrm{n}}{\mathrm{n}-1} \mathrm{E}(\mathrm{s}^2) = \frac{\mathrm{n}}{\mathrm{n}-1} \frac{\mathrm{n}-1}{\mathrm{n}} \sigma^2 = \boldsymbol{\sigma^2}$$

Universidad Rey Juan Carlos

3. Distribución de los estadísticos muestrales

(Ejercicio)

En una empresa dedicada al transporte de mercancías se estudia el absentismo laboral, es decir, el número de días que ha faltado al trabajo un empleado elegido aleatoriamente de la plantilla total. Se sabe que durante el último año, han causado baja un día el 50% de los trabajadores, dos días el 40% y tres días el resto. Si se toma una m.a.s. de tamaño dos (X1 X2), se pide:

1. Distribución de probabilidad de la variable aleatoria "número de días que ha faltado al trabajo un empleado", así como su media y varianza.
2. Distribución de probabilidad del estadístico media muestra!, así como su esperanza y varianza.
3. Distribución de probabilidad del estadístico varianza muestral, y su esperanza.
4. Calcular la probabilidad de que el estadístico media muestra! sea menor que 2.

Universidad Rey Juan Carlos

2. Distribución de los estadísticos muestrales

(Ejercicio)

POBLACIÓN

ξ: nº dias falta al trabajo	x=1	x=2	x=3
$P(\xi=xi)$	0,50	0,40	0,10

PARÁMETROS POBLACIONALES:

$$E(\xi) = \mu = \sum_{\forall xi} x_i P(\xi = x_i) = 1*0{,}50 + 2*0{,}40 + 3*0{,}10 = 1{,}6$$

$$V(\xi) = \sigma^2 = \sum_{\forall xi} x_i^{\,2} P(\xi = x_i) - \mu^2 = (1^2*0{,}50 + 2^2*0{,}40 + 3^2*0{,}10) - (1{,}6)^2 = 0{,}44$$

$$V_1(\xi) = (n/n\text{-}1)*\sigma^2=(2/1)*0{,}44 = 0{,}88$$

2. Distribución de los estadísticos muestrales

(Ejercicio)

M.A.S. n=2

Muestras (x_1 x_2)	Media muestral $\overline{x_i} = \frac{(x_1 + x_2)}{2}$	Varianza muestral $S^2_{xi} = \sum_{i=1}^{2} \frac{(x_i - \overline{x})^2}{2}$	Probabilidades P(X=x_1; X=x_2)
(1,1)	$\overline{x_i} = \frac{(1+1)}{2} = 1$	$S^2_{xi} = \frac{(1-1)^2 + (1-1)^2}{2} = 0$	$P(x_1 = 1; x_2 = 1) = 0,5 * 0,5 = 0,25$
(1,2)	$\overline{x_i} = \frac{(1+2)}{2} = 1,5$	$S^2_{xi} = \frac{(1-1,5)^2 + (2-1,5)^2}{2} = 0,25$	$P(x_1 = 1; x_2 = 2) = 0,5 * 0,4 = 0,20$
(1,3)	2	1	0,05
(2,1)	1,5	0,25	0,20
(2,2)	2	0	0,16
(2,3)	2,5	0,25	0,04
(3,1)	2	1	0,05
(3,2)	2,5	0,25	0,04
(3,3)	3	0	0,01

Universidad Rey Juan Carlos

2. Distribución de los estadísticos muestrales

(Ejercicio)

Según los anteriores resultados, la distribución de probabilidad del estadístico media muestral $\bar{x}$ sería:

$\bar{x}i$	1	1,5	2	2,5	3
P_i	0,25	0,40	0,26	0,08	0,01

Esperanza y varianza de la media muestral:

$$E(\bar{x}) = \sum_{\forall xi} \bar{x}_i P(\xi = \bar{x}_i) = 1 * 0{,}25 + 1{,}5 * 0{,}40 + 2 * 0{,}26 + 2{,}5 * 0{,}08 + 3 * 0{,}01 = 1{,}6$$

$$V(\bar{x}) = \left[\sum_{\forall xi} \bar{x}_i^{\,2} P(\xi = \bar{x}_i) \right] - (E(\bar{x}))^2 = (1^2 * 0{,}25 + 1{,}5^2 * 0{,}40 + 2^2 * 0{,}26 + 2{,}5^2 * 0{,}08 + 3^2 * 0{,}01) - (1{,}6)^2 = 0{,}22$$

RELACIONES ENTRE LOS PARÁMETROS POBLACIONALES Y EL ESTADÍSTICO MEDIA MUESTRAL:

Se cumple que: $E(\bar{x}) = \mu = E(\xi)$ $\qquad V(\bar{x}) = \dfrac{\sigma^2}{n}$

La distribución de probabilidad del estadístico varianza muestral S^2_x sería:

S^2_{xi}	0	0,25	1
P_i	0,42	0,48	0,10

Esperanza de la varianza muestral: $E(S_x^{\,2}) = (0*0{,}42)+(0.25*0{,}48)+(1*0{,}10) = 0{,}22$

Universidad Rey Juan Carlos

2. Distribución de los estadísticos muestrales

(Ejercicio)

RELACIONES ENTRE LOS PARÁMETROS POBLACIONALES Y EL ESTADÍSTICO VARIANZA MUESTRAL:

Se cumple que: $E(S_x^2) = \frac{n-1}{n}\sigma^2$

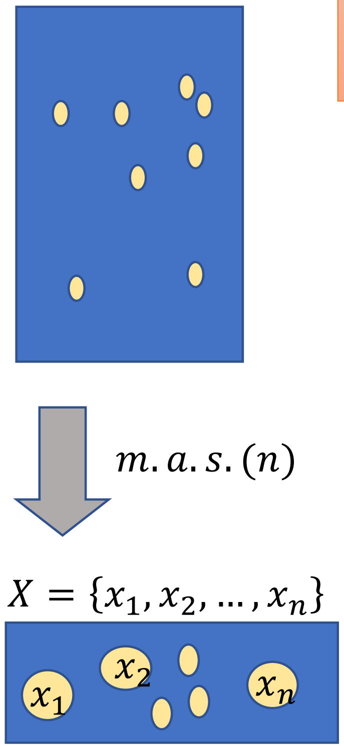
La varianza de la Varianza Muestral se obtiene como:

$V(S_x^2)=(0^2*0,42)+(0.25^2*0,48)+(1^2*0,10)-(0.22^2) = 0,0816$

Universidad Rey Juan Carlos

4.1. Muestreo en poblaciones normales (σ conocida)

$\xi, : N(\mu, \sigma)$



4.1.1. Media muestral $\overline{x}$ es Normal porque

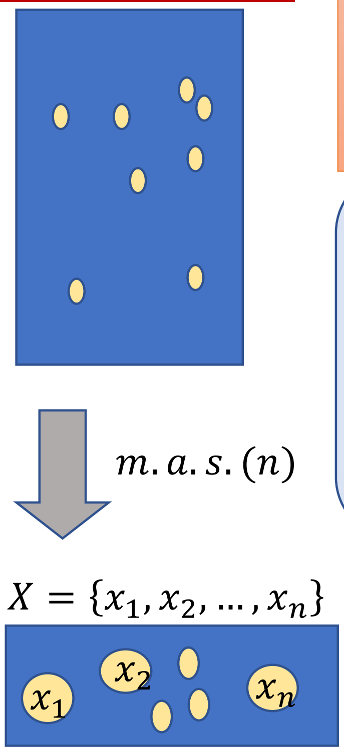
es suma de normales independientes multiplicadas por una constante $\left(\frac{1}{n}\right)$

$$\left.\begin{array}{l} E(\overline{x}) = \mu \\ V(\overline{x}) = \dfrac{\sigma^2}{n} \end{array}\right\} \rightarrow \boldsymbol{\overline{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)}$$

Universidad Rey Juan Carlos

4.1. Muestreo en poblaciones normales (σ conocida)

$\xi, : N(\mu, \sigma)$



4.1.2. Varianza muestral $s^2 = \frac{\sum_i (x_i - \bar{x})^2}{n}$

El lema de Fisher-Cochran garantiza

1. $\frac{ns^2}{\sigma^2} = \chi^2_{n-1}$

pq la suma de cuadrados de N(0,1)=z independientes es una χ^2_n y la última z no es independiente.

2. s^2 y $\bar{x}$ son linealmente independientes.

Si $\frac{ns^2}{\sigma^2} = \chi^2_{n-1}$ entonces como $s_1^2 = \frac{n}{n-1} s^2 \rightarrow \frac{n \frac{n-1}{n} s_1^2}{\sigma^2} = \chi^2_{n-1} \rightarrow \frac{(n-1)s_1^2}{\sigma^2} = \chi^2_{n-1}$

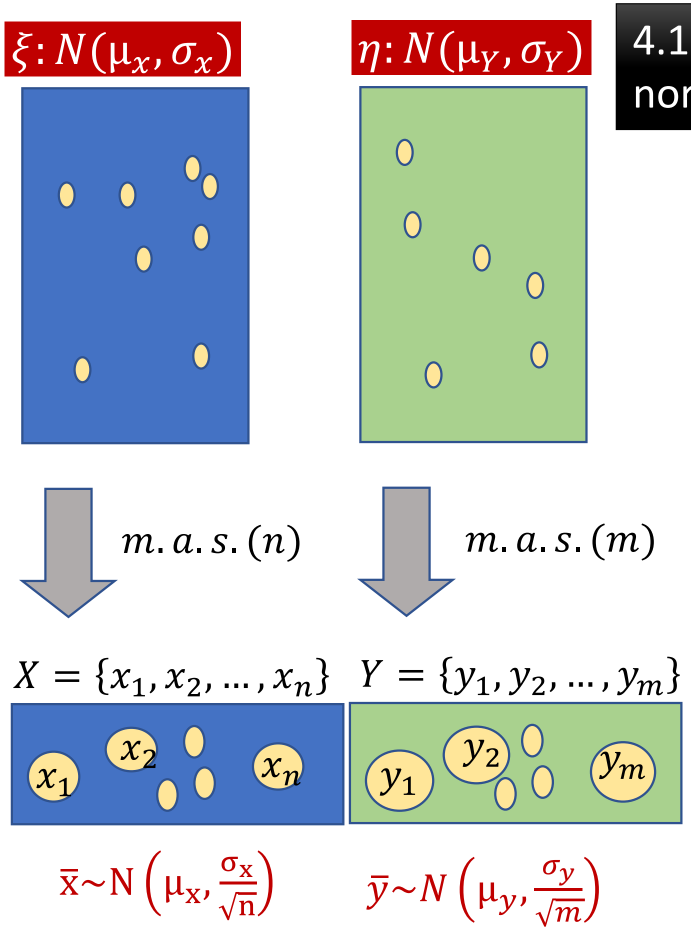
Distribución de la **varianza** muestral

$$\frac{ns^2}{\sigma^2} = \chi^2_{n-1}$$

$$\frac{ns_1^2}{\sigma^2} = \chi^2_n$$

Distribución de la **cuasivarianza** muestral

4.1. Muestreo en poblaciones normales



4.1.3. Distribución de $\bar{x}$ - $\bar{y}$ con parametros conocidos en poblaciones normales

$$\left.\begin{array}{l} \bar{x} \sim N\left(\mu_x, \frac{\sigma_x}{\sqrt{n}}\right) \\ \bar{y} \sim N\left(\mu_y, \frac{\sigma_y}{\sqrt{m}}\right) \end{array}\right\} \bar{x} - \bar{y} \sim N\left(\mu_x - \mu_y, \sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}\right)$$

porque la diferencia de normales es una normal de...

- $E(\bar{x} - \bar{y}) = E(\bar{x})\text{-}\, E(\bar{y}) = \mu_x - \mu_y$
- $V(\bar{x} - \bar{y}) = E(\bar{x}) + E(\bar{y}) = \frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}$

$$Z = \frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}}$$

y por tanto..

$$\boxed{(\bar{x} - \bar{y}) = (\mu_x - \mu_y) + z\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}}$$

Universidad Rey Juan Carlos

4.1. Muestreo en poblaciones normales

4.1.3. Distribución de $\bar{x}$ - $\bar{y}$ con parametros conocidos en poblaciones normales (Ejemplo)

Elemplo 1. De una población $\xi: N(5, \sqrt{4})$ se extrae un m.a.s. (16). Se pide:

1. Distribución de la media muestral

$$\bar{x}: N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) = \mathrm{N}\left(5, \frac{2}{\sqrt{16}}\right) = \mathrm{N}\left(5, \frac{1}{2}\right)$$

2. Distribución de la varianza muestral

$$\frac{ns^2}{\sigma^2} = \chi^2_{n-1} \leftrightarrow s^2 = \frac{\sigma^2 \chi^2_{n-1}}{n} \rightarrow s^2: \frac{4\chi^2_{16-1}}{16} = \frac{1}{4}\chi^2_{15}$$

3. Tamaño muestral necesario para que la media muestral se aleje menos de 0,5 unidades de la esperanza

$$n?\ si\ P(|\bar{x} - \mu| \leq 0'5) \geq 0{,}95$$

$$P(|\bar{x} - \mu| \leq 0'5) = P(-0'5 \leq \bar{x} - \mu \leq 0'5) = P\left(\frac{-0'5}{1/_{2\sqrt{n}}} \leq \frac{\bar{x} - \mu}{\sigma/_{\sqrt{n}}} \leq 0' \frac{0'5}{1/_{2\sqrt{n}}}\right)$$

$$= P(-0'5 * 2\sqrt{n} \leq z \leq 0'5 * 2\sqrt{n}) = P(-\sqrt{n} \leq z \leq \sqrt{n});$$

$$P(-z_0 \leq z \leq z_0) = 0{,}95 \rightarrow P(z_0 \leq z) = 0{,}975 \rightarrow z_0 = 1{,}96 = \sqrt{n}$$

$$\boldsymbol{n = 1{,}96^2 = 3{,}84 \cong 4}$$

En EXCEL =INV.NORM.ESTAND(0,975)

Universidad Rey Juan Carlos

4.1. Muestreo en poblaciones normales

4.1.3. Distribución de $\bar{x}$ - $\bar{y}$ con parametros conocidos en poblaciones normales (Ejemplo 2)

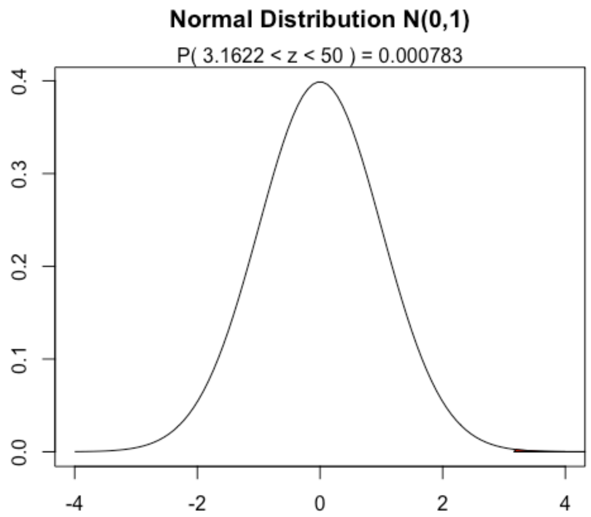
Elemplo 2. En una empresa, el salario de los directivos se distribuye como una N(10,1), sobre estos se realiza un m.a.s(10). El salario del resto de empleados se distribuye como N(4,2) y sobre estos realizamos un m.a.s. (20). Se pide.

a) Probabilidad de que el salario de los directivos supere las 11 u.m. $\boldsymbol{P(\bar{x} > 11)}$

$$\bar{x}: \mathrm{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right) = \mathrm{N}\left(10, \frac{1}{\sqrt{10}}\right); P(\bar{x} > 11) = P\left(\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} > \frac{11-\mu}{\frac{\sigma}{\sqrt{n}}}\right) = P\left(z > \frac{11-10}{0'3162}\right) = P(z > 3{,}1622) = 0'0027$$

Plotting in R

En EXCEL =DISTR.NORM.ESTAND.N(3,1622;FALSO)



R Code

```
mean=0; sd=1
lb=3.1622; ub=50

x <- seq(-4,4,length=100)*sd + mean
hx <- dnorm(x,mean,sd)

plot(x, hx, type="n", xlab="z values", ylab="",
  main="Normal Distribution N(0,1)", axes=TRUE)

i <- x >= lb & x <= ub
lines(x, hx)
polygon(c(lb,x[i],ub), c(0,hx[i],0), col="red")

area <- pnorm(ub, mean, sd) - pnorm(lb, mean, sd)
result <- paste("P(",lb,"< z <",ub,") =",
  signif(area, digits=3))
mtext(result,3)
axis(1, at=seq(40, 160, 20), pos=0)
```

Universidad Rey Juan Carlos

4.1. Muestreo en poblaciones normales

4.1.3. Distribución de $\bar{x}$ - $\bar{y}$ con parametros conocidos en poblaciones normales (Ejemplo 2)

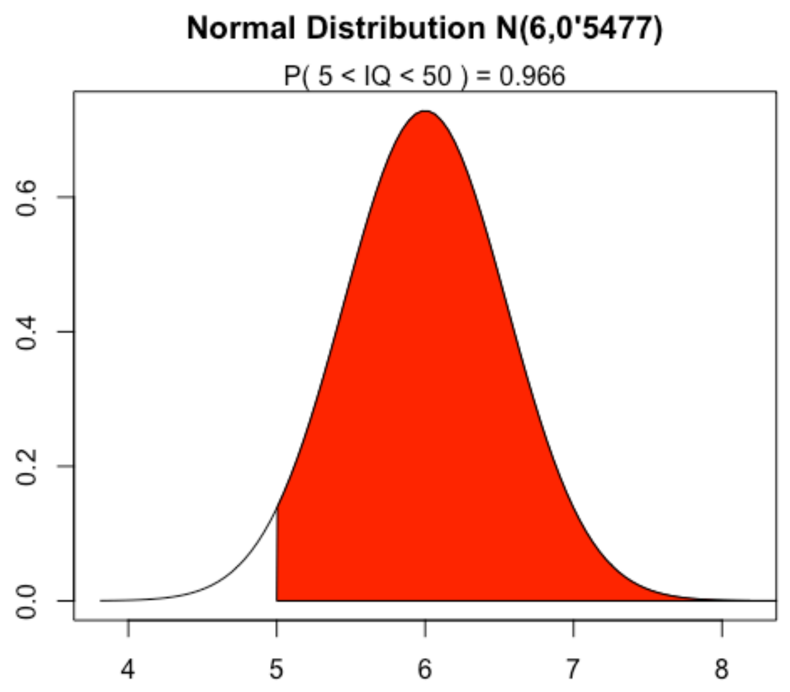
Elemplo 2. En una empresa, el salario de los directivos se distribuye como una N(10,1), sobre estos se realiza un m.a.s(10). El salario del resto de empleados se distribuye como N(4,2) y sobre estos realizamos un m.a.s. (20). Se pide.

B) Probabilidad de que el salario medio de los directivos supere en 5 u.m. al salario medio del resto. $\boldsymbol{P(\bar{x} > \bar{y} + 5)}$

$$\left.\begin{array}{l}\bar{x}: \mathrm{N}\left(10, \frac{1}{\sqrt{10}}\right) \\ \bar{y}: \mathrm{N}\left(4, \frac{2}{\sqrt{20}}\right)\end{array}\right\} \bar{x} - \bar{y}: \mathrm{N}\left(10-4, \sqrt{\frac{1}{10} + \frac{4}{20}}\right) = N(6{,}0'5477)$$

$$\boldsymbol{P(\bar{x} > \bar{y} + 5)} = P(\bar{x} - \bar{y} > 5) = P\left(\frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}} > \frac{5-6}{0.5477}\right) =$$

$$= P(z > -1{,}82) = 0{,}9656$$

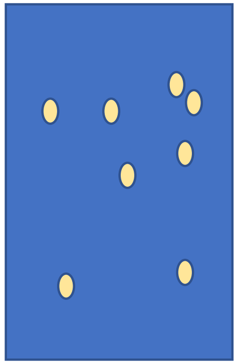


En EXCEL =DISTR.NORM.ESTAND.N(1,82;VERDADERO) POR SIMETRÍA

Universidad Rey Juan Carlos

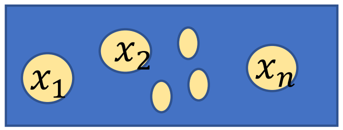
4.2. Muestreo en poblaciones normales (σ =?)

$\xi, : N(\mu, \sigma)$



$m.a.s.(n)$

$X = \{x_1, x_2, \dots, x_n\}$



$\bar{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$

4.2.1. Distribución de $\bar{x}$ con σ^2 en poblaciones normales

$\bar{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ pero σ^2 es desconocido. Hay que recurrir a distribuciones independientes de la varianza: la t de Student.

$$\left.\begin{aligned} z &= \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \\ \chi^2_{n-1} &= \frac{ns^2}{\sigma^2} \end{aligned}\right\} t_{n-1} = \frac{z}{\sqrt{\frac{\chi^2_{n-1}}{n-1}}} = \frac{\frac{\bar{x} - \mu}{\sigma / \sqrt{n}}}{\sqrt{\frac{\frac{ns^2}{\sigma^2}}{n-1}}} = \frac{(\bar{x} - \mu)\sqrt{n-1}}{\frac{\sigma}{n}\sqrt{\frac{ns^2}{\sigma^2}}} = \frac{(\bar{x} - \mu)\sqrt{n-1}}{s}$$

y por tanto..

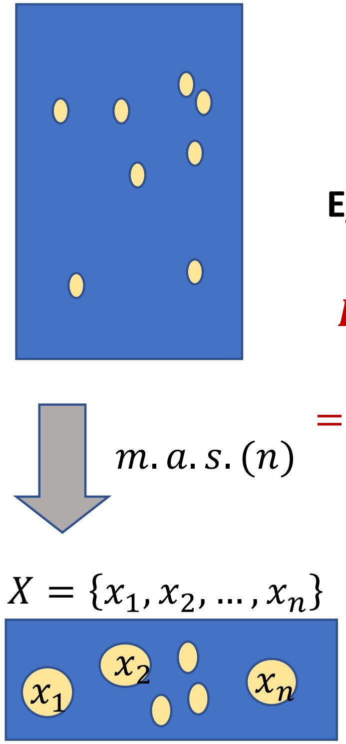
$$\bar{x} = \mu + \frac{s}{\sqrt{n-1}} t_{n-1}$$

Universidad Rey Juan Carlos

4.2. Muestreo en poblaciones normales (σ =?)

$\xi, : N(\mu, \sigma)$

4.2.1. Distribución de $\bar{x}$ con σ^2 en poblaciones normales



$\bar{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$

$$\bar{x} = \mu + \frac{s}{\sqrt{n-1}} t_{n-1}$$

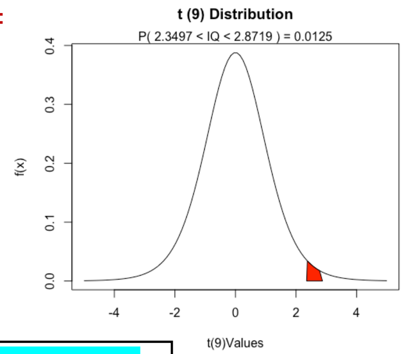
Ejemplo: $\xi: N(0, \sigma) \xrightarrow[mas(10)]{} X = \{x_1, x_2, \dots, x_n\}$; $\bar{x} = ?$, $s = 2{,}298$. **calcular:** $P(1'8 \leq \bar{x} \leq 2'2)$

$$P(1'8 \leq \bar{x} \leq 2'2) = P\left(1'8 \leq \frac{2{,}298}{\sqrt{10-1}} t_{10-1} \leq 2'2\right) = P\left(\frac{\sqrt{10-1}}{2{,}298} 1'8 \leq t_9 \leq \frac{\sqrt{10-1}}{2{,}298} 2'2\right) =$$

$$= P\left(\frac{3}{2{,}298} 1'8 \leq t_9 \leq \frac{3}{2{,}298} 2'2\right) \cong P(1{,}3054 * 1'8 \leq t_9 \leq 1{,}3054 * 2'2) =$$

$$= P(1{,}3054 * 1'8 \leq t_9 \leq 1{,}3054 * 2'2) = P(2{,}3497 \leq t_9 \leq 2{,}8719) =$$

$= P(t_9 \leq 2{,}8719)$- $P(t_9 \leq 2{,}3497)$= 0’9908-0’9783=0,0125

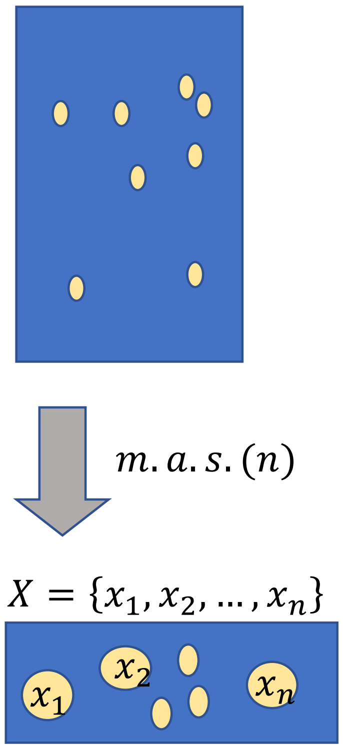


En EXCEL =DISTR.T.N(2,8719;9;VERDADERO)- DISTR.T.N(2,3497;9;VERDADERO)

Universidad Rey Juan Carlos

4.2. Muestreo en poblaciones normales (σ =?)

$\xi, : N(\mu, \sigma)$

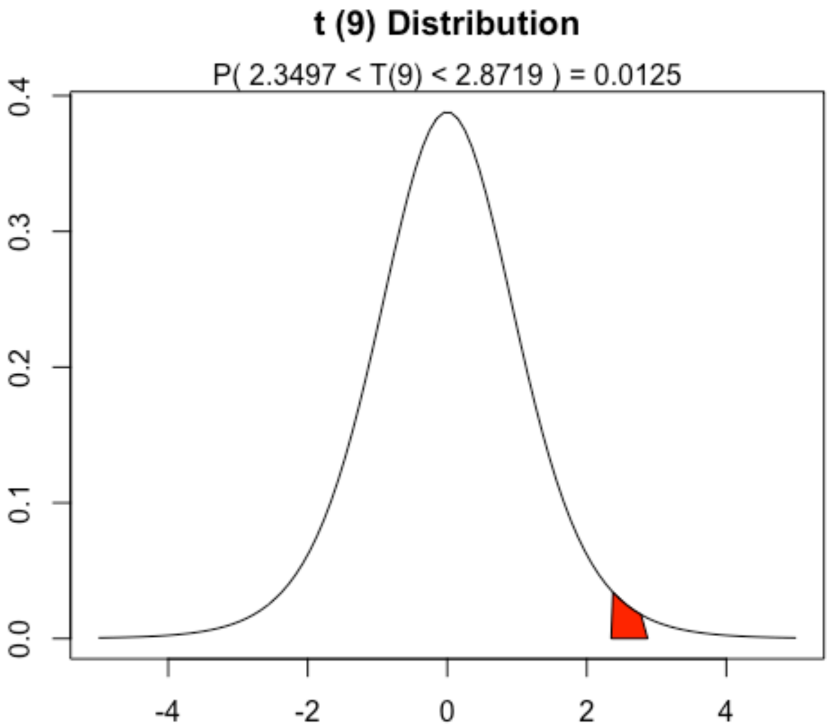


$\bar{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$

4.2.1. Distribución de $\bar{x}$ con σ^2 en poblaciones normales

$\boldsymbol{P(1'8 \leq \bar{x} \leq 2'2)} = \boldsymbol{P}(2{,}3497 \leq t_9 \leq 2{,}8719)$0,0125

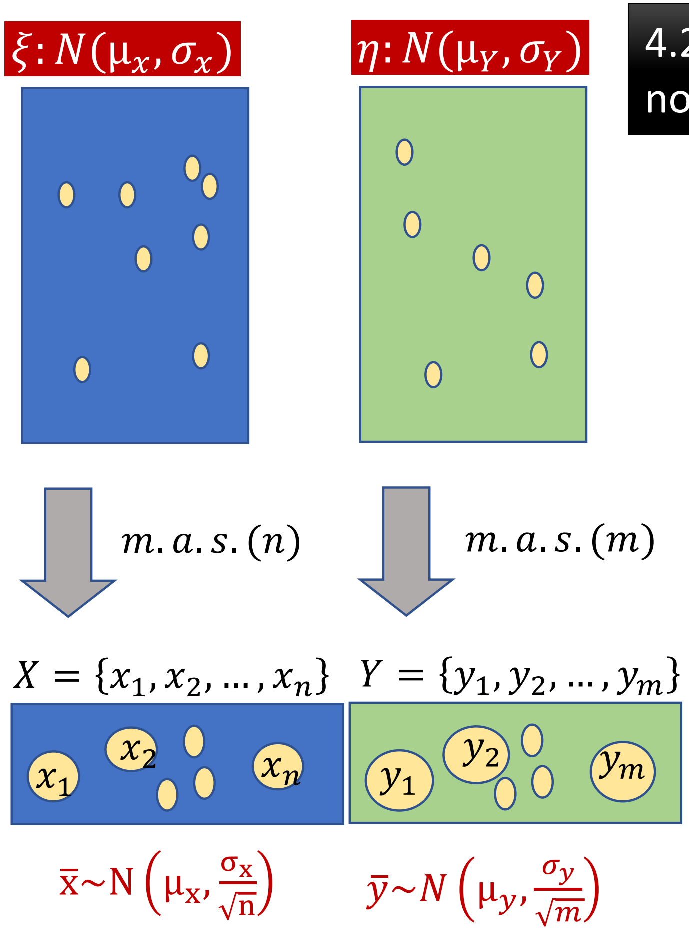
Plotting in R



R Code

```
lb=2.3497; ub=2.8719

x <- seq(-5,5,length=100)
hx <- dt(x,9)

plot(x, hx, type="n", xlab="t(9)Values", ylab="",
  main="t (9) Distribution", axes=TRUE)

i <- x >= lb & x <= ub
lines(x, hx)
polygon(c(lb,x[i],ub), c(0,hx[i],0), col="red")

area <- pt(ub,9) - pt(lb,9)
result <- paste("P(",lb,"< T(9) <",ub,") =",
   signif(area, digits=3))
mtext(result,3)
```

4.2. Muestreo en poblaciones normales (σ =?)



4.2.2. Distribución de $\bar{x}$ - $\bar{y}$ con σ_x^2 y σ_y^2 desconocidas en poblaciones normales

$$\left.\begin{array}{l} \bar{x} \sim N\left(\mu_x, \frac{\sigma_x}{\sqrt{n}}\right) \\ \bar{y} \sim N\left(\mu_y, \frac{\sigma_y}{\sqrt{m}}\right) \end{array}\right\} \bar{x} - \bar{y} \sim N\left(\mu_x - \mu_y, \sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}\right)$$

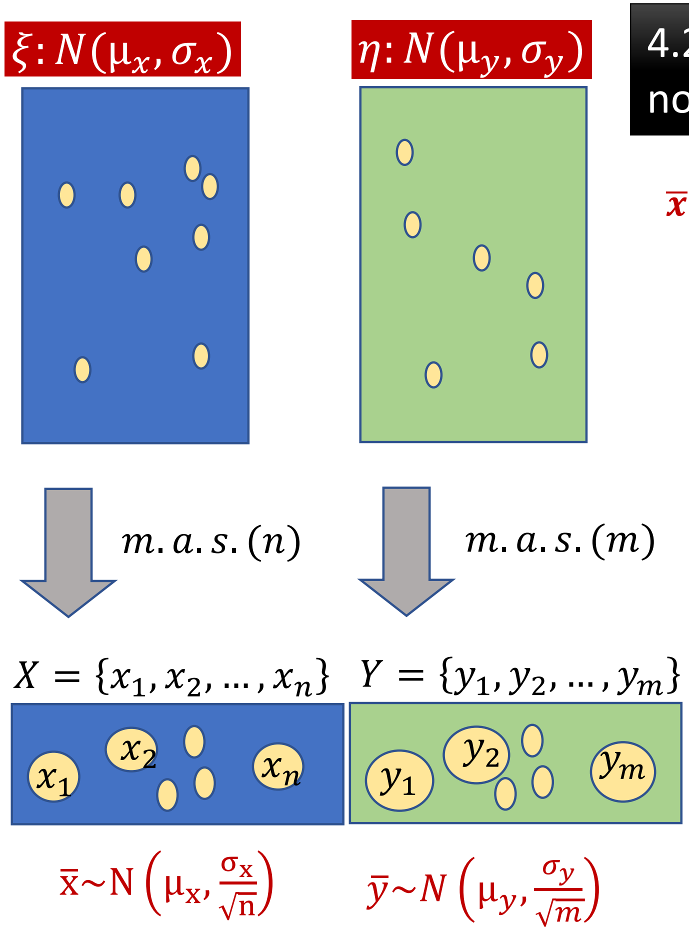
porque la diferencia de normales es una normal de…

- $E(\bar{x} - \bar{y}) = E(\bar{x})$- $E(\bar{y}) = \mu_x - \mu_y$
- $V(\bar{x} - \bar{y}) = V(\bar{x})$+ $V(\bar{y}) = \frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}$

Hay que construir una t …

Universidad Rey Juan Carlos

4.2. Muestreo en poblaciones normales (σ =?)



4.2.2. Distribución de $\bar{x}$ - $\bar{y}$ con σ_x^2 y σ_y^2 desconocidas en poblaciones normales

$$\bar{x} - \bar{y} \sim N\left(\mu_x - \mu_y, \sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}\right)$$

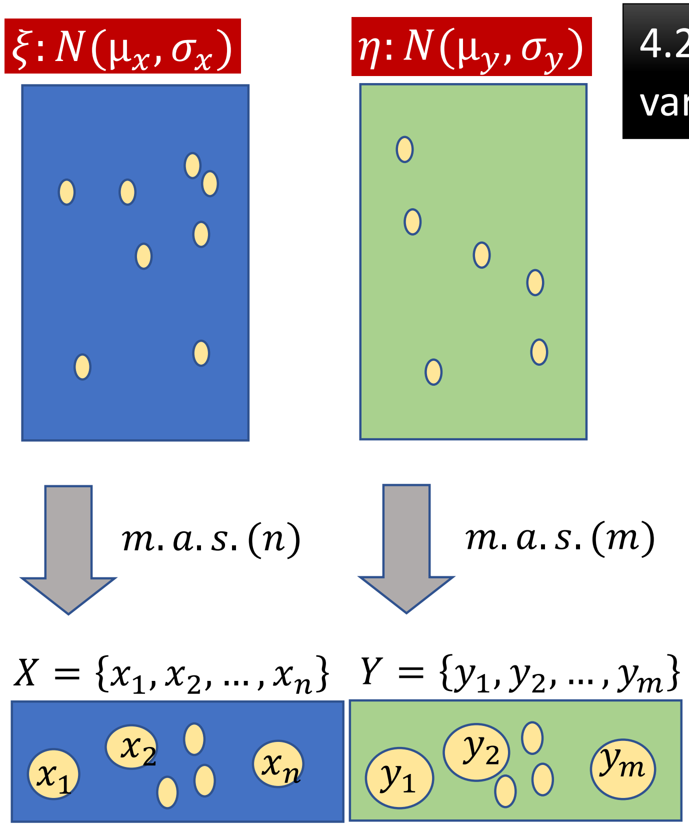
$$\left.\begin{array}{l} z = \dfrac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{\sqrt{\dfrac{\sigma_x^2}{n} + \dfrac{\sigma_y^2}{m}}} \\ \left.\begin{array}{l} \dfrac{ns_x^2}{\sigma_x^2} = \chi_{n-1}^2 \\ \dfrac{ms_y^2}{\sigma_y^2} = \chi_{m-1}^2 \end{array}\right\} \chi_{m+n-2}^2 = \dfrac{ns_x^2}{\sigma_x^2} + \dfrac{ms_y^2}{\sigma_y^2} \end{array}\right\} t_{m+n-2} = \frac{\dfrac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{\sqrt{\dfrac{\sigma_x^2}{n} + \dfrac{\sigma_y^2}{m}}}}{\sqrt{\dfrac{\dfrac{ns_x^2}{\sigma_x^2} + \dfrac{ms_y^2}{\sigma_y^2}}{m + n - 2}}}$$

y por tanto, **TENEMOS UN PROBLEMA**.. Porque no desaparecen las varianzas poblacionales… **a no ser que ocurra que $\sigma_x^2 = \sigma_y^2 = \sigma^2$** es decir, que las varianzas de ambas poblaciones sean iguales.

Universidad Rey Juan Carlos

4.2. Muestreo en poblaciones normales (σ =?)

4.2.2. Distribución de $\bar{x}$ - $\bar{y}$ con $\boldsymbol{\sigma_x^2 = \sigma_y^2 = \sigma^2}$; varianzas poblacionales desconocidas pero iguales



$$\bar{x} \sim N\left(\mu_x, \frac{\sigma}{\sqrt{n}}\right) \qquad \bar{y} \sim N\left(\mu_y, \frac{\sigma}{\sqrt{m}}\right)$$

$$\bar{x} - \bar{y} \sim N\left(\mu_x - \mu_y, \sqrt{\frac{\sigma^2}{n} + \frac{\sigma^2}{m}}\right)$$

$$z = \frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma^2}{n} + \frac{\sigma^2}{m}}}$$

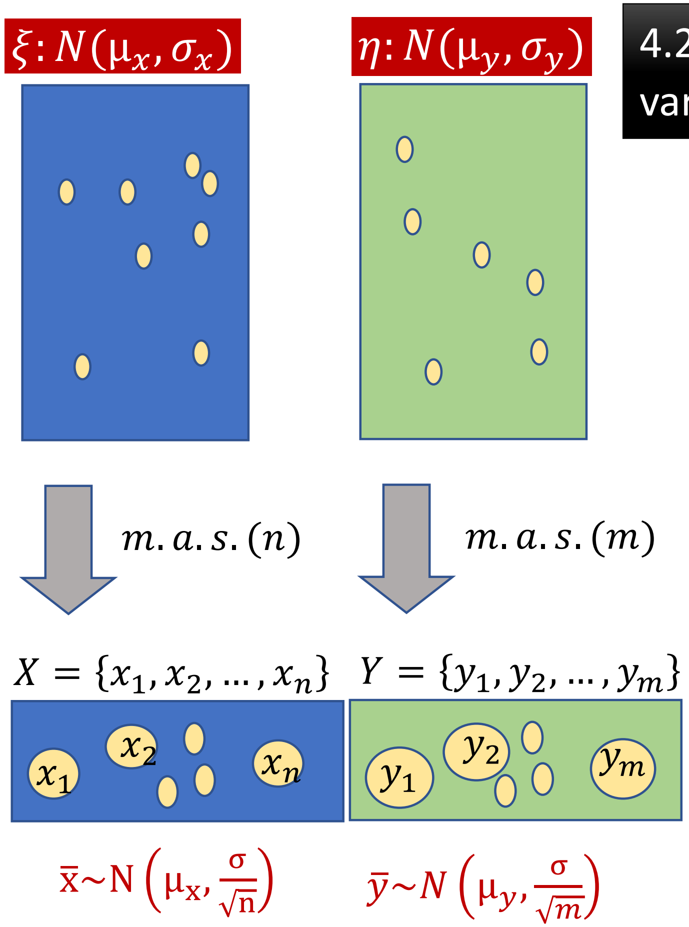
$$\left.\begin{array}{l} \frac{ns_x^2}{\sigma^2} = \chi^2_{n-1} \\ \frac{ms_y^2}{\sigma^2} = \chi^2_{m-1} \end{array}\right\} \chi^2_{m+n-2} = \frac{ns_x^2}{\sigma^2} + \frac{ms_y^2}{\sigma^2}$$

$$t_{m+n-2} = \frac{\frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma^2}{n} + \frac{\sigma^2}{m}}}}{\sqrt{\frac{\frac{ns_x^2}{\sigma^2} + \frac{ms_y^2}{\sigma^2}}{m+n-2}}}$$

$$\boxed{\begin{aligned} \boldsymbol{t_{m+n-2}} &= \frac{[(\bar{x} - \bar{y}) - (\mu_x - \mu_y)]\sqrt{m+n-2}}{\sqrt{\frac{\sigma^2}{n} + \frac{\sigma^2}{m}}\sqrt{\frac{ns_x^2}{\sigma^2} + \frac{ms_y^2}{\sigma^2}}} = \\ &= \boldsymbol{\frac{[(\bar{x} - \bar{y}) - (\mu_x - \mu_y)]\sqrt{m+n-2}}{\sqrt{\frac{1}{n} + \frac{1}{m}}\sqrt{ns_x^2 + ms_y^2}}} \end{aligned}}$$

Universidad Rey Juan Carlos

4.2. Muestreo en poblaciones normales (σ =?)



4.2.2. Distribución de $\bar{x}$ - $\bar{y}$ con $\sigma_x^2 = \sigma_y^2 = \sigma^2$; varianzas poblacionales desconocidas pero iguales

$$\bar{x} - \bar{y} \sim N\left(\mu_x - \mu_y, \sqrt{\frac{\sigma^2}{n} + \frac{\sigma^2}{m}}\right)$$

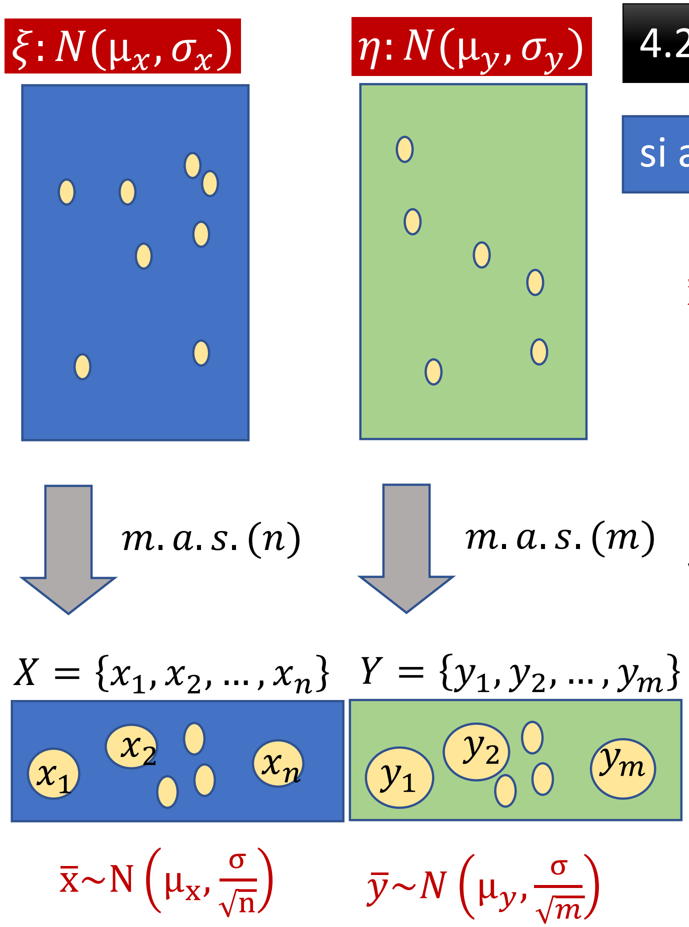
$$t_{m+n-2} = \frac{\left[(\bar{x} - \bar{y}) - (\mu_x - \mu_y)\right]\sqrt{m+n-2}}{\sqrt{\frac{1}{n} + \frac{1}{m}}\sqrt{n s_x^2 + m s_y^2}}$$

Ya podríamos calcular probabilidades despejando $(\bar{x} - \bar{y})$

$$\bar{x} - \bar{y} = (\mu_x - \mu_y) + \frac{t_{m+n-2}\sqrt{\frac{1}{n} + \frac{1}{m}}\sqrt{n s_x^2 + m s_y^2}}{\sqrt{m+n-2}}$$

Universidad Rey Juan Carlos

4.2. Muestreo en poblaciones normales (σ =?)



4.2.2. Distribución de $\bar{x}$ - $\bar{y}$ con $\boldsymbol{\sigma_x^2} = \boldsymbol{\sigma_y^2} = \boldsymbol{\sigma^2}$ desconocidas

si además $\mu_x = \mu_y = \mu$ son iguales aunque desconocidas

$$\bar{x} - \bar{y} \sim N\left(0, \sqrt{\frac{\sigma^2}{n} + \frac{\sigma^2}{m}}\right)$$

porque

$$\mu_x - \mu_y = \mu - \mu = 0$$

y entonces

$$t_{m+n-2} = \frac{(\bar{x} - \bar{y})\sqrt{m+n-2}}{\sqrt{\frac{1}{n} + \frac{1}{m}}\sqrt{ns_x^2 + ms_y^2}}$$

$$(\bar{x} - \bar{y}) = \frac{t_{m+n-2}\sqrt{\frac{1}{n} + \frac{1}{m}}\sqrt{ns_x^2 + ms_y^2}}{\sqrt{m+n-2}}$$

4.2. Muestreo en poblaciones normales (σ =?)

Universidad Rey Juan Carlos

Ejemplo: La empresa A fabrica un producto cuya demanda sigue una $N(200, \sigma)$ mientras que la empresa B fabrica el mismo producto con demanda $N(100, \sigma)$. Se efectúa un m.a.s. (125) para cada empresa y se obtiene $S_a = 300$ y $S_b = 250$. Calcular la probabilidad de que la media de la empresa A supere en 50 unidades a la media de la empresa b.

$$P(\bar{x} - \bar{y} < 50) = ? \qquad \bar{x} - \bar{y}: N\left(100, \sigma\sqrt{\frac{2}{125}}\right)$$

$$\left.\begin{array}{l} z = \dfrac{(\bar{x} - \bar{y}) - 100}{\sigma\sqrt{\dfrac{2}{125}}} \\ \left.\begin{array}{l} \dfrac{125 s_a^2}{\sigma^2} = \chi^2_{124} \\ \dfrac{125 s_b^2}{\sigma^2} = \chi^2_{124} \end{array}\right\} \dfrac{125}{\sigma^2}(s_a^2 + s_b^2) = \chi^2_{248} \end{array}\right\} t_{248} = \frac{\dfrac{(\bar{x} - \bar{y}) - 100}{\sigma\sqrt{\dfrac{2}{125}}}}{\sqrt{\dfrac{\dfrac{125}{\sigma^2}(s_a^2 + s_b^2)}{248}}} = \frac{\dfrac{(\bar{x} - \bar{y}) - 100}{0'13\sigma}}{\sqrt{\dfrac{\dfrac{125}{\sigma^2}(300^2 + 250^2)}{248}}} = \frac{(\bar{x} - \bar{y}) - 100}{36{,}04}$$

$$\frac{(\bar{x} - \bar{y}) - 100}{36{,}04} = t_{248} \rightarrow (\bar{x} - \bar{y}) = 36{,}04 t_{248} + 100$$

Recordemos $t_n \xrightarrow[TCL\ n\geq 30]{} N\left(0, \sqrt{\frac{n}{n-2}}\right)$ en nuestro caso $t_{248} \xrightarrow[TCL\ n\geq 30]{} N\left(0, \sqrt{\frac{248}{248-2}}\right) = \mathrm{N}\left(0, \sqrt{1'0081}\right)$

4.2. Muestreo en poblaciones normales (σ =?)

Ejemplo: La empresa A fabrica un producto cuya demanda sigue una $N(200, \sigma)$ mientras que la empresa B fabrica el mismo producto con demanda $N(200, \sigma)$. Se efectúa un m.a.s. (125) para cada empresa y se obtiene $S_a = 300$ y $S_b = 250$. Calcular la probabilidad de que la media de la empresa A supere en 50 unidades a la media de la empresa b.

$$P(\bar{x} - \bar{y} < 50) = ? \qquad \bar{x} - \bar{y} : N\left(100, \sigma\sqrt{\frac{2}{125}}\right)$$

$$\frac{(\bar{x} - \bar{y}) - 100}{36{,}04} = t_{248} \rightarrow (\bar{x} - \bar{y}) = 36{,}04 t_{248} + 100$$

$$P(\bar{x} - \bar{y} < 50) = \mathrm{P}(36{,}04 t_{248} + 100 < 50) = \mathrm{P}\left(t_{248} < \frac{50-100}{36{,}04}\right) = \mathrm{P}(t_{248} < -1'3872) = 0{,}0833$$

En EXCEL =DISTR.T.N(-1,3872;248;VERDADERO)

Recordemos $t_n \xrightarrow[TCL\ n\geq 30]{} N\left(0, \sqrt{\frac{n}{n-2}}\right)$ en nuestro caso $t_{248} \xrightarrow[TCL\ n\geq 30]{} N\left(0, \sqrt{\frac{248}{248-2}}\right) = \mathrm{N}\left(0, \sqrt{1'0081}\right)$

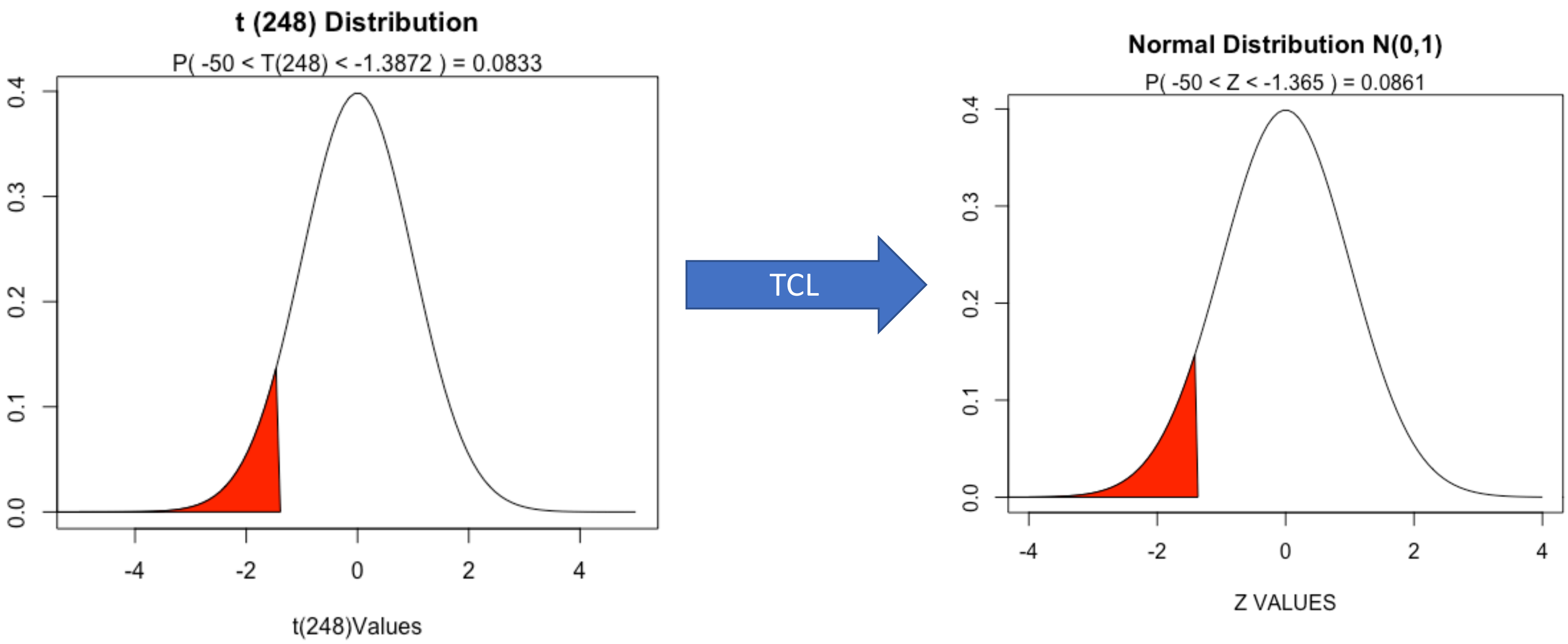
$$P(\bar{x} - \bar{y} < 50) = \mathrm{P}\left(t_{248} < \frac{50-100}{36{,}04}\right) \cong \mathrm{P}\left(z < \frac{-1'3872}{1'0162}\right) = \mathrm{P}(z < -1{,}365) = 0{,}0861$$

En EXCEL =DISTR.NORM.ESTAND.N(-1,365;VERDADERO)

4.2. Muestreo en poblaciones normales (σ =?)

Universidad Rey Juan Carlos

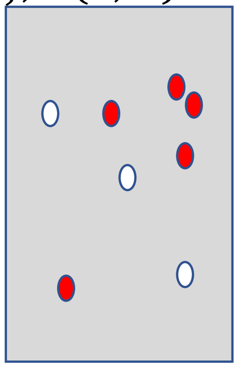
Recordemos $t_n \xrightarrow[TCL\ n\geq 30]{} N\left(0, \sqrt{\frac{n}{n-2}}\right)$ en nuestro caso $t_{248} \xrightarrow[TCL\ n\geq 30]{} N\left(0, \sqrt{\frac{248}{248-2}}\right) = \mathrm{N}(0, \sqrt{1'0081})$



Universidad Rey Juan Carlos

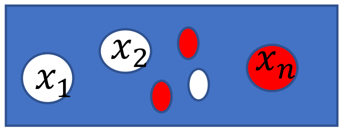
4.3.1. ξ: $B(1,\pi)$, distribución de la proporción muestral p

$\xi, B(1,\Pi)$



$m.a.s.(n)$

$X=\{0,\ldots,0,1,\ldots 1\}$



- ξ es la población que se distribuye como $B(1,\pi)$: formada por infinitas realizaciones de un experimento aleatorio de unos y ceros. Luego **no es necesario conocer el tamaño población**.
- π refleja la proporción de éxitos poblacional: proporción de individuos que cumplen una característica.

$$\pi=\frac{n^{o}\ de\ elementos\ de\ la\ población\ que\ cumplen\ la\ característica}{n^{o}\ total\ de\ la\ población}$$

fenómeno	x_i	P_i
éxito	1	π
fracaso	0	1-π

$$B(1,\pi) \xrightarrow{m.a.s.(n)} X=\left\{\begin{matrix}0\ \ldots\ 1\ \ldots\ 0\ \ldots\ 0\ \ldots\ 1\ \ldots\ 1\\ n\ extracciones\end{matrix}\right\}$$

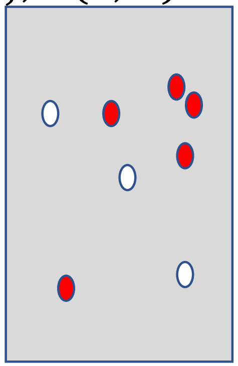
$$p=\frac{n^{o}\ de\ elementos\ de\ la\ muestra\ que\ cumplen\ la\ característica}{tamaño\ de\ la\ muestra\ (n)}=\frac{\sum_{i=1}^{n}\{0;1\}}{n}=\frac{\sum_{i=1}^{n}x_i}{n}$$

Da lo mismo sumar los elementos muestrales que no poseen la característica porque toman el valor 0.

Universidad Rey Juan Carlos

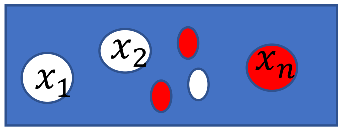
4.3.1. $\xi: B(1,\pi)$, distribución de la proporción muestral p

$\xi, B(1,\Pi)$



$m.a.s.(n)$

$X = \{0, \dots, 0, 1, \dots 1\}$



Al realizar un m.a.s (n), p es variable aleatoria:

1. P no es binomial porque la media de binomiales no es binomial.
2. $E(p) = E\left[\sum \frac{x_i}{n}\right] = \frac{1}{n} E\left[\sum x_i\right] \boxed{=} \frac{1}{n} \sum E(x_i) = \frac{1}{n} \sum E(\xi) = \frac{1}{n} n\pi = \pi$
3. $V(p) = V\left[\sum \frac{x_i}{n}\right] = \frac{1}{n^2} V\left[\sum x_i\right] \boxed{=} \frac{1}{n^2} \sum V(x_i) = \frac{1}{n^2} \sum V(\xi) = \frac{n\pi\,(1-\pi)}{n^2} = \frac{\pi\,(1-\pi)}{n}$

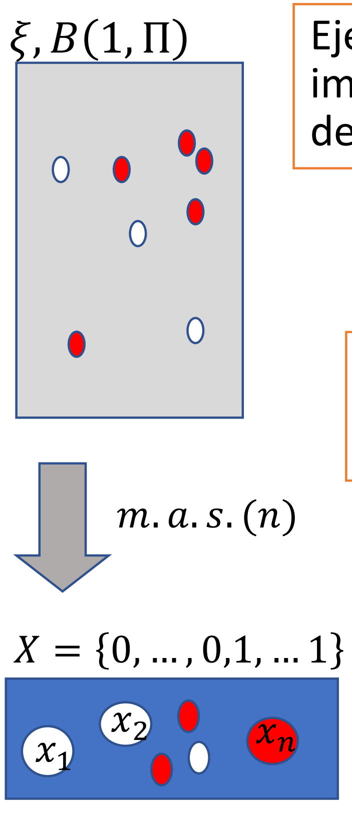
Podemos asegurar el = por propiedades del **m.a.s.** -> VAIID

Aunque p no es binomial, si lo es n*p=número de éxitos muestral

$$(n \le 30)\quad n * p = \sum x_i = \frac{x_1 + x_2 + \dots + x_n}{x_i: B(1,\pi)} = \mathrm{B}(n,\pi)$$

$$(n > 30)\quad p \xrightarrow[TCL\ n \ge 30]{} N\left(\pi, \sqrt{\frac{\pi(1-\pi)}{n}}\right)$$

4.3.1 ξ: $B(1,\pi)$, distribución de la proporción muestral p

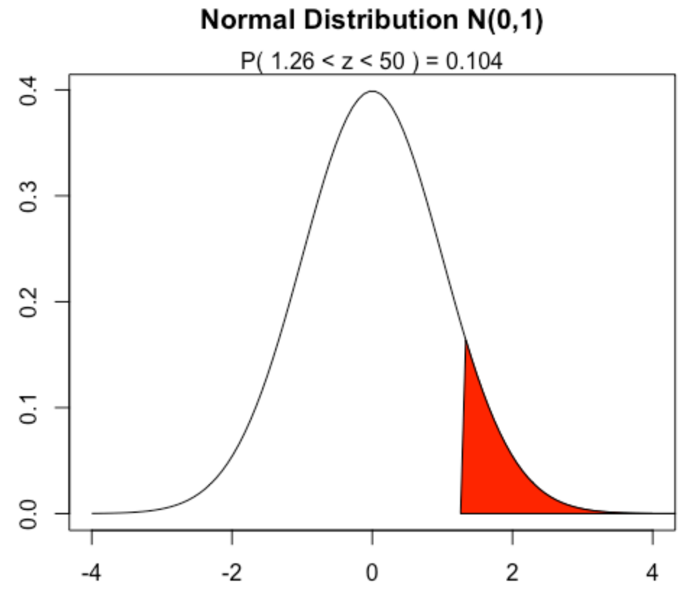


Ejemplo: Se ha estimado que el 43% de los graduados en Economía consideran importante un curso de Ética profesional. Calcule la probabilidad de que más de la mitad de 80 licenciados obtenidos por m.a.s. opinen de este modo.

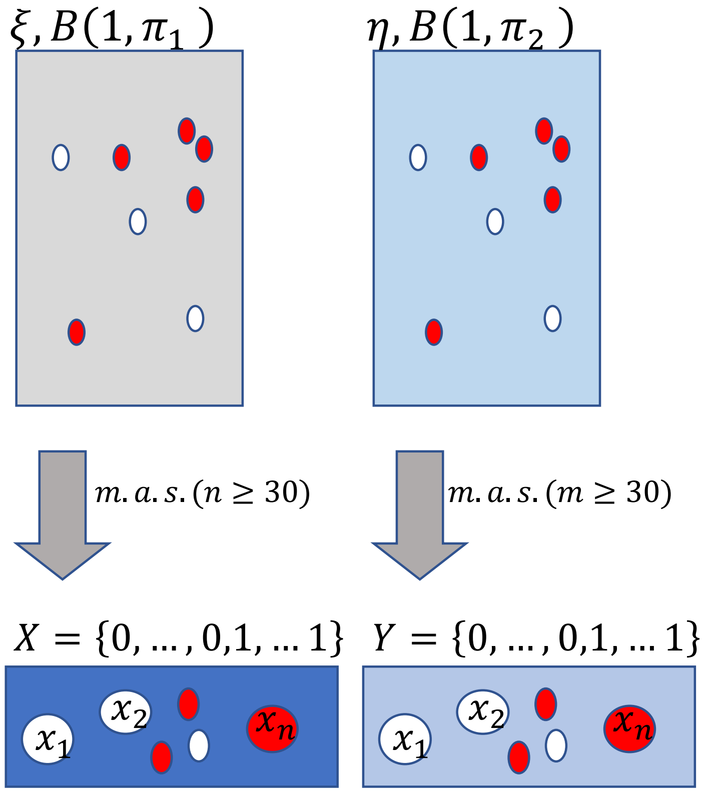
$$Sea\ \xi = \begin{cases} 1\ Sí\ Ética \\ 0\ No\ Ética \end{cases} \qquad \xi: B(1,\pi) = B(1{,}0'43)$$

$$\xi \xrightarrow[m.a.s.(80)]{} X = \{x_1, x_2, \dots, x_n\} \qquad p \xrightarrow[TCL.n\geq 30]{} N(0'43{,}0'055)$$

$$P(p > 0{,}5) \underset{n>30}{=} P\left(\frac{p-\pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} > \frac{0'5 - 0{,}43}{\sqrt{\frac{0{,}43(1-0{,}43)}{80}}}\right) = \\ = P(z > 1{,}26) = 0{,}104$$



4.3.2 Distribución de la diferencia de proporciones muestrales



$$\left.\begin{array}{l} p_1 \xrightarrow[TCL]{} N\left(\pi_1, \sqrt{\dfrac{\pi_1(1-\pi_1)}{n}}\right) \\ p_2 \xrightarrow[TCL]{} N\left(\pi_2, \sqrt{\dfrac{\pi_2(1-\pi_2)}{m}}\right) \end{array}\right\}$$

$$\boxed{\boldsymbol{p_1 - p_2 : N\left[(\pi_1 - \pi_2), \sqrt{\frac{\pi_1(1-\pi_1)}{n} + \frac{\pi_2(1-\pi_2)}{m}}\right]}}$$

Universidad Rey Juan Carlos

4.3.2 Distribución de la diferencia de proporciones muestrales

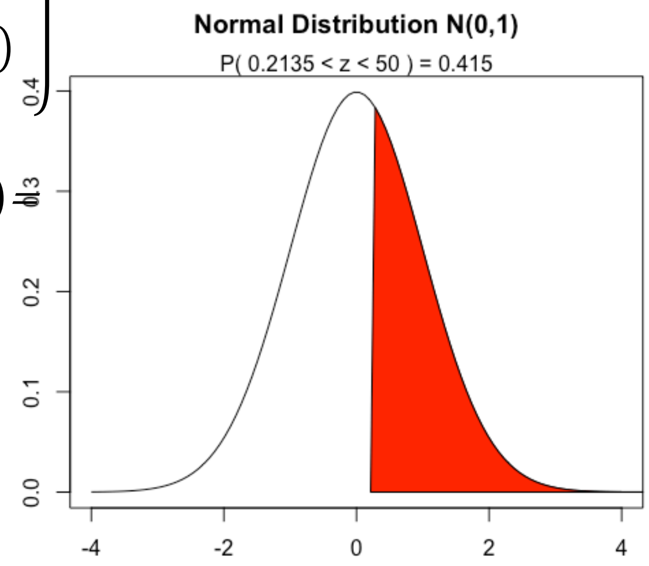
- Ejemplo: El porcentaje de votantes con preferencia a un determinado partido es del 5% en A y 150 de B. Consultados 100 electores de A y 150 de B, determine la probabilidad de que el % de electores consultados favorables al partido en A supere en 0’02 al % de electores favorables en B.

$$A \to \xi_1: B(1,0'05); mas(100);\ p_1 \xrightarrow[TCL]{} N\left(0'05, \sqrt{\frac{0'05 * 0'95}{100}}\right) = N\left(0'05, \sqrt{0'000475}\right)$$

$$B \to \xi_2: B(1,0'10); mas(150);\ p_2 \xrightarrow[TCL]{} N\left(0'10, \sqrt{\frac{0'10 * 0'90}{150}}\right) = N\left(0'10, \sqrt{0'0006}\right)$$

$$p_1 - p_2: N\left[(\pi_1 - \pi_2), \sqrt{\frac{\pi_1(1-\pi_1)}{n} + \frac{\pi_2(1-\pi_2)}{m}}\right] = N\left(0'05 - 0'10, \sqrt{0'000475 + 0'0006}\right) =$$
$$= N(-0'05, 0'03279)$$

$$P(p_1 - p_2 > 0{,}02) = P\left(\frac{(p_1-p_2)-(-0'05)}{0'3279} > \frac{0{,}02-(-0'05)}{0'3279}\right) = P(z > 0{,}2135) = 0{,}415$$



Universidad Rey Juan Carlos

Referencias

- Martín Pliego, Fundamentos de Inferencia Estadística, Editorial AC, 3ªEd, 2005

- Gracias a mis antiguas compañeras, Pilar Ordás del Amo y Mercedes Casas.

- Otros recursos en www.pacorabadan.com y aula virtual URJC
- R from CRAN
- Rstudio + ggplot
- https://www.statmethods.net/advgraphs/probability.html