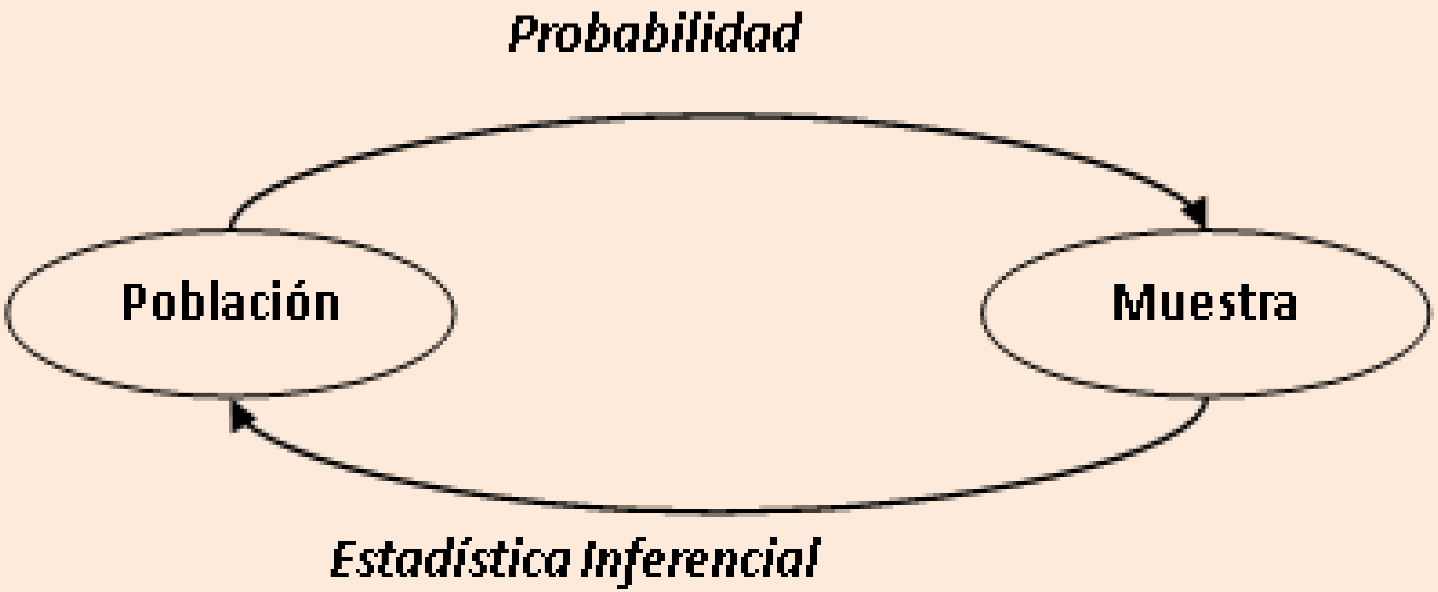
Muestras aleatorias
Distribuciones de muestreo

Ingeniería

UM

Revisión



Probabilidad
Población
Muestra
Estadística Inferencial

Población

- Se llama **población** al conjunto total de elementos en discusión y sobre los cuáles se quiere tener alguna información.
- Esta información esta representada por una (o varias) variable aleatoria
- EJEMPLO 1.
- En un estudio sobre la variación del precio de el

  Cemento en el mes de marzo en Mendoza.

  La población es el conjunto de negocios donde se vende cemento.
- ***La variable aleatoria*** que representa a esta población en este estudio es ***¨El precio del cemento¨.***

EJEMPLO 2

- En un estudio sobre los salarios docentes
- ***La población*** es el conjunto de todos los docentes a los que va dirigido el estudio, los primarios, por ejemplo. ***La variable de interés*** es el salario de un docente.

Muestra

- Se tiene una población de la cual se quiere tener alguna información. Como se dijo antes, a veces es imposible o poco práctico, observar toda la población, entonces se toma parte de ella (**muestra**) y después de analizar esta parte se infieren los resultados a la población total.

- Como la inferencia estadística se formula con base en una muestra de objetos de la población de interés, el proceso por medio del cual se obtiene será aquél que asegure la selección de una “buena” muestra.

- Una manera de obtenerla es cuando el proceso de muestreo proporciona, a cada objeto en la población, una oportunidad igual e independiente de ser incluido en la muestra. Este concepto conduce a lo que se conoce como **muestra aleatoria**.

- Imagínese probando una población de 1. 000.000 de circuitos hasta que fallen antes de comercializarlos! Mejor es tomar algunos de ellos, observar la proporción que falla y luego inferir este resultado al total circuitos. Evidentemente este resultado no será nunca “exacto” pero puede resultar interesante si se lo relaciona con el concepto de probabilidad. Es decir, si se puede establecer una cierta confianza en nuestra inferencia.

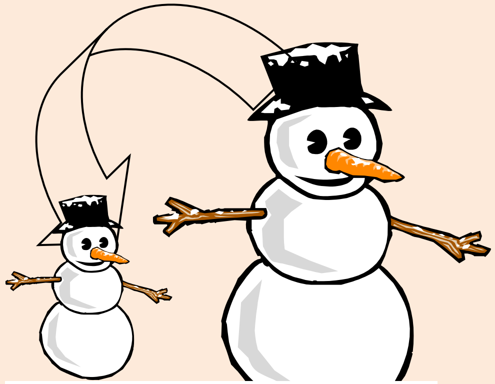
Muestra

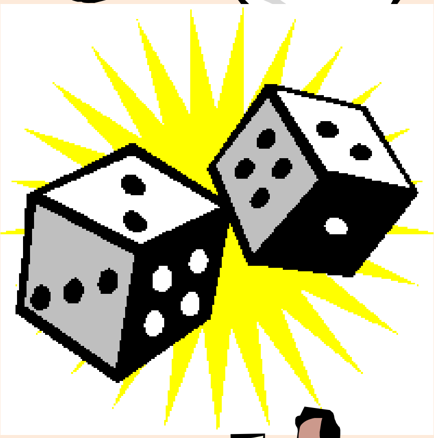
Propiedades que debe tener una buena muestra:

repesentativa

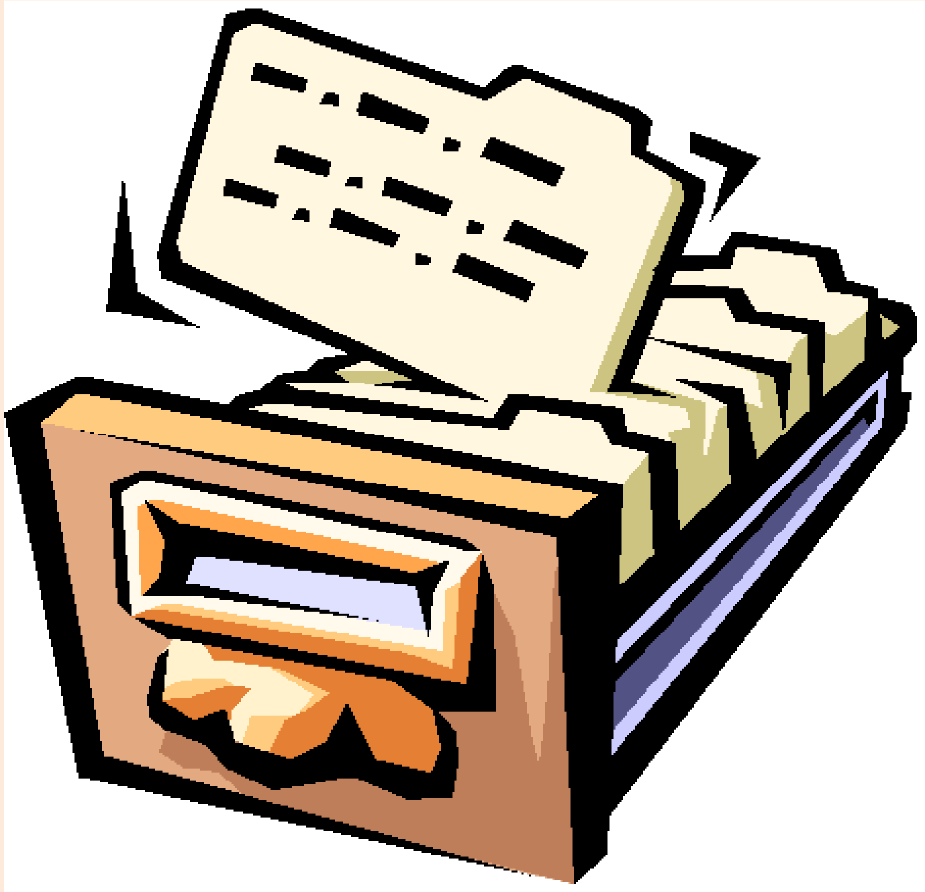
aleatoria

independencia



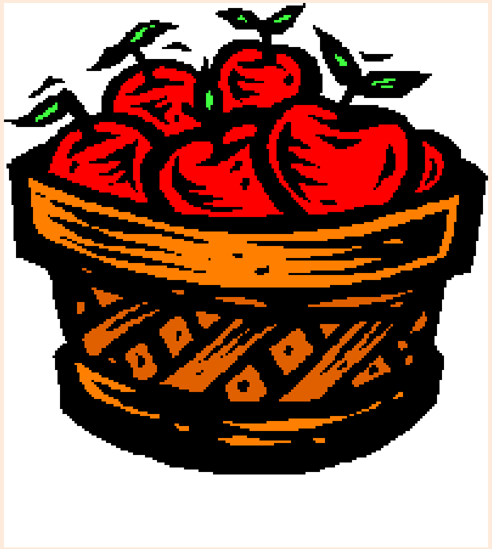


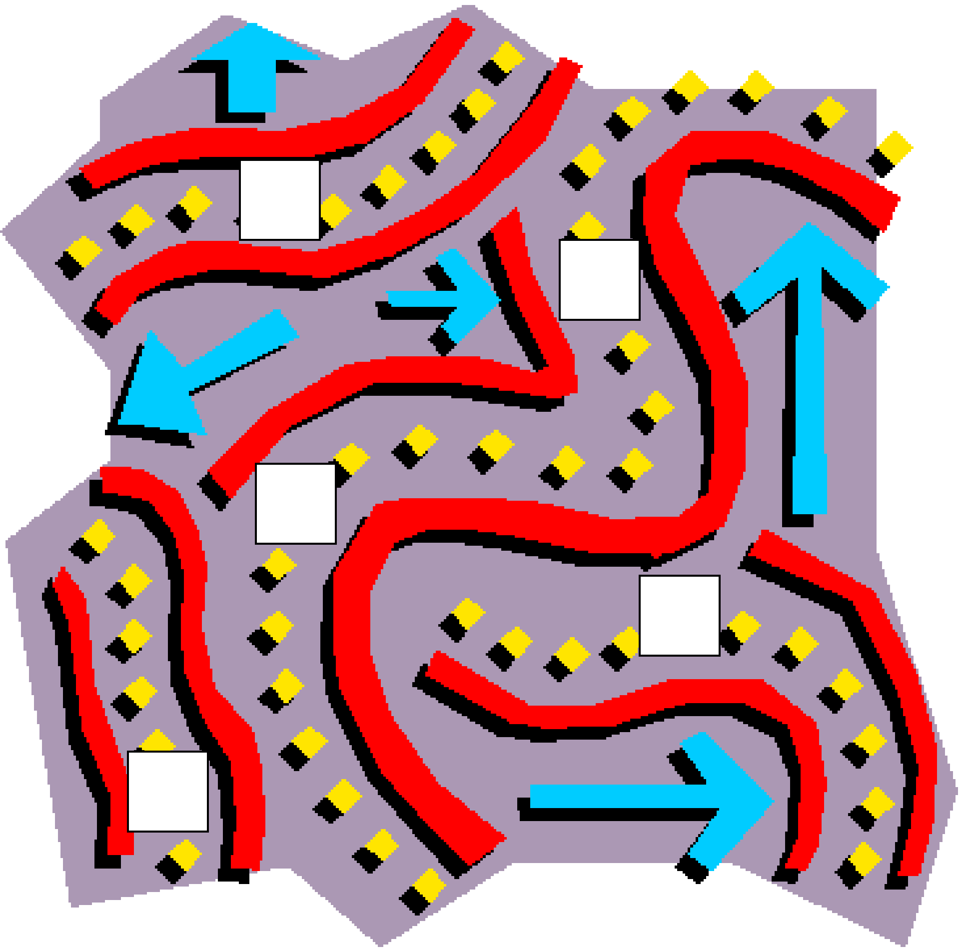
MUESTREO ALEATORIO SIMPLE



MUESTREO SISTEMÁTICO

MUESTREO ESTRATIFICADO





MUESTREO DE CONGLOMERADOS

Concepto de muestra

- Una **muestra aleatoria** de tamaño “n” de una población representada por la variable X, con función de densidad *f* es un conjunto de “n” variables aleatorias independientes y cada una con idéntica distribución de la población. Simbólicamente
- $$X_1, X_2, \ldots X_n \overset{i.id.}{\sim} f_X(x, \theta)$$
- donde el símbolo *i.id.* hace referencia a la independencia y a la idéntica distribución de las variables . θ Representa el o los parámetros poblacionales y $\underline{x} = x_1, x_2, \ldots, x_n$

FUNCION DENSIDAD DE LA MUESTRA

- Sea $(X_1, X_2, .., X_n)$ una muestra aleatoria de una población
- representada por la variable aleatoria X con función densidad
- fx (x; θ).
- La función densidad conjunta de la muestra

$$f_{X_1,X_2,..,X_n}(x_1, x_2, \ldots, x_n; \theta) \underline{\underline{ind.}} f_{X_1}(x_1, \theta) \cdot f_{X_2}\left(x_2, \theta\right) \cdot \ldots \cdot f_{X_n}\left(x_n, \theta\right)$$

Anotaremos resumidamente

$$f_{\underline{\mathbf{X}}}(\underline{x}; \theta) = f_{X_1}(x_1, \theta) \cdot f_{X_2}\left(x_2, \theta\right) \cdot \ldots \cdot f_{X_n}\left(x_n, \theta\right)$$

Ejemplo

- $(X_1, X_2, .., X_n)$ muestra de tamaño n
- $x_1, x_2, \ldots, x_n$ un valor observado de la muestra
- :
- :
- $x'_1, x'_2, \ldots, x'_n$ otro valor observado de la muestra

Estadísticos

- Podemos decir que uno de los problemas más frecuentes en Estadística es estudiar una población *f*, donde *f* depende de un parámetro θ desconocido.

- Es decir se conoce la forma de la distribución pero no exactamente cuál es la densidad *f* por no conocerse el valor del parámetro θ.

- Por ejemplo se sabe que la distribución es exponencial pero no se conoce el valor del parámetro .

- El problema a resolver es estimar de la mejor manera posible el valor de este parámetro.

Estadístico

- Para abordar este problema tomamos una muestra aleatoria $X_1, X_2, \ldots, X_n$ de esta población con densidad *f* y obtenemos los valores $x_1, x_2, \ldots, x_n$.

- Luego se construye alguna función $\ell(X_1, X_2, \ldots X_n)$ de manera que sirva para estimar al parámetro desconocido.

- Esta función recibe el nombre de **estadístico** (o estadística).

Estadístico definición

- Si $X_1, X_2, \ldots, X_n$ es una muestra de tamaño n de una población representada por la variable X, se llama estadístico a cualquier función $T = \hat{\Theta} = \ell(X_1, X_2, \ldots X_n)$, que no dependa de parámetros desconocidos.

- Un estadístico es una función de variables aleatorias observables y en consecuencia el mismo es una variable aleatoria.

Ejemplos de estadísticos

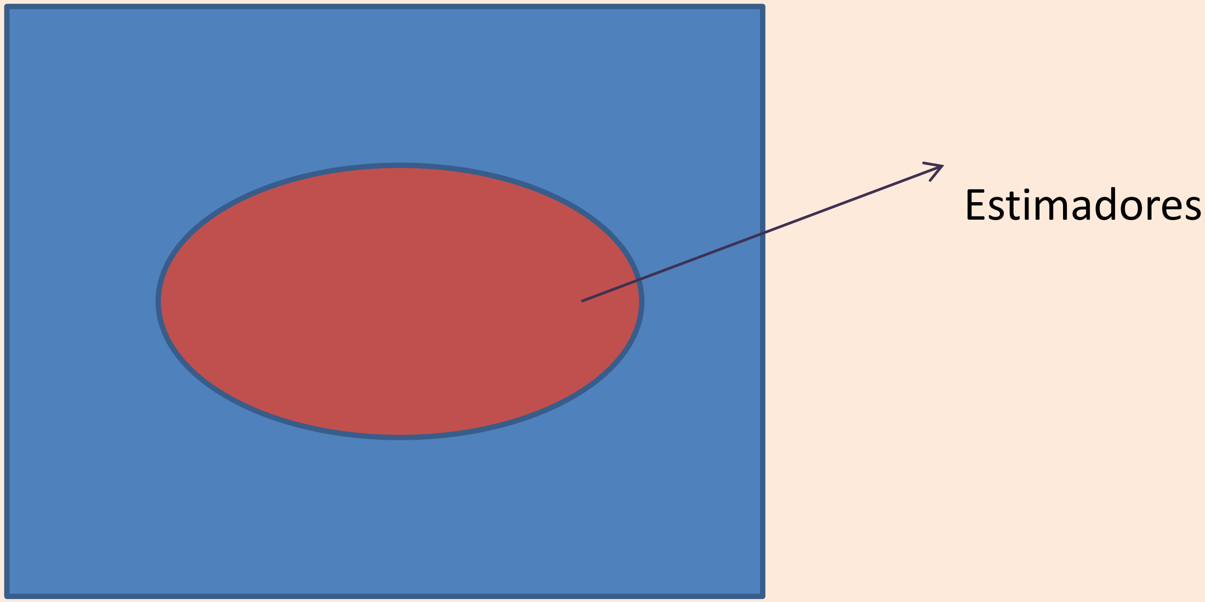
- Sea $X_1, X_2,\ldots,.X_n$ una muestra aleatoria de una población con densidad $f_X(x,\theta)$, donde θ es un parámetro desconocido.
- $T_1 = \hat{\Theta}_1 = X_2 \cdot X_5 = l(X_1, X_2,\ldots,.X_n)$ es un estadístico
- $t_1 = \hat{\theta}_1 = x_1 x_2 = l(x_1, x_2,\ldots, x_n)$ es un valor observado del estadístico o un valor puntual del estadístico
- $\hat{\Theta}_2 = X_1 + X_2 + n$ es un estadístico
- $\hat{\Theta}_3 = X_1{}^2 + lnX_2$ es un estadístico

Estadísticos y no estadísticos

- $X-\theta$ no es un estadístico
- $\dfrac{X}{\theta}-3X_1+X_2$ no es un estadístico
- $\hat{\Theta}_3=\dfrac{1}{n}\sum_{i=1}^{n}X_i=\overline{X}$ El promedio de las variables de la
- muestra es un estadístico
- $\hat{\Theta}_4=\dfrac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2=S^2$ es un estadístico

Estimadores

- Estadísticos



- Aquellos estadísticos que sirven para estimar un parámetro desconocido, reciben el nombre de estimadores

EJEMPLO DE ESTIMADOR

- Sea $X_1, X_2, \ldots, X_n \sim N(\mu, \sigma^2)$ con μ y σ^2 desconocidos.
- Ejemplo, $T_1 = \ell_1(X_1, X_2, \ldots, X_n) = \overline{X}$
- donde $\ell_1 = \ell_1(x_1, x_2, \ldots x_n) = \overline{x}$

luego $\hat{\mu} = \overline{x}$

- Un estimador siempre es una estadístico

ESTIMACIÓN DE LOS PARÁMETROS DE LA POBLACIÓN

PARÁMETRO	ESTADÍSTICA	ESTIMACION PUNTUAL
θ	$\hat{\Theta}$	$\hat{\theta}$
μ	$\overline{X}$	$\overline{x}$
σ^2	S^2	s^2
p	$\hat{P}$	$\hat{p}$
	ESTIMADOR	ESTIMACIÓN

Revisando

Parámetro: es una constante

Un estadístico es una variable aleatoria y como v.a. tiene una distribución de probabilidades

- A la distribución de probabilidades de un estadístico se le llama **distribución muestral del estadístico**.

Estadístico Media Muestral- Esperanza y Varianza de la Media Muestral

- Sea $X_1, X_2, \ldots, X_n$ una m.a. de una población representada por X con función densidad fx y con esperanza μ <∞ y varianza σ^2 <∞
- Estadístico Media muestral

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

- La distribución de la media muestral $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

depende de la distribución de la variable X en la población.

- Sin embargo su $E\left(\overline{X}\right)$ y $\operatorname{var}(\overline{X})$ siempre tienen las mismas relaciones con la esperanza y varianza de la población

Esperanza y varianza de la media muestral

- Esperanza

$$E\left(\overline{X}\right) = \mu$$

- Varianza

$$\operatorname{var}(\overline{X}) = \frac{\sigma_X^2}{n}$$

Varianza muestral

- Es un estadístico pero como sirve para estimar el parámetro poblacional varianza lo denominaremos estimador

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

- $E(S^2) = \sigma^2$ si la varianza poblacional es finita

- Otro estimador que nos interesa es la desviación estándar muestral

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

Distribución muestral de $\overline{X}$

- Determinaremos la distribución de $\overline{X}$ a partir de un ejemplo de muestras de tamaño 2 con reemplazo de una variable aleatoria discreta (población) con distribución uniforme.
- Probaremos que se cumplen las relaciones
- $$\boxed{E\left(\overline{X}\right) = \mu}$$ $$\boxed{\operatorname{var}(\overline{X}) = \frac{\sigma_X^2}{n}}$$

Distribución muestral de el estadístico $\overline{X}$

- Primeramente determinaremos la esperanza y varianza de la población.
- Consideremos como ejemplo una población representada por una variable aleatoria discreta con distribución uniforme

$$f_X(x) = \begin{cases} \dfrac{1}{5} & sí \quad x = 2,3,4,5,6 \\ 0 & otro \quad caso \end{cases}$$

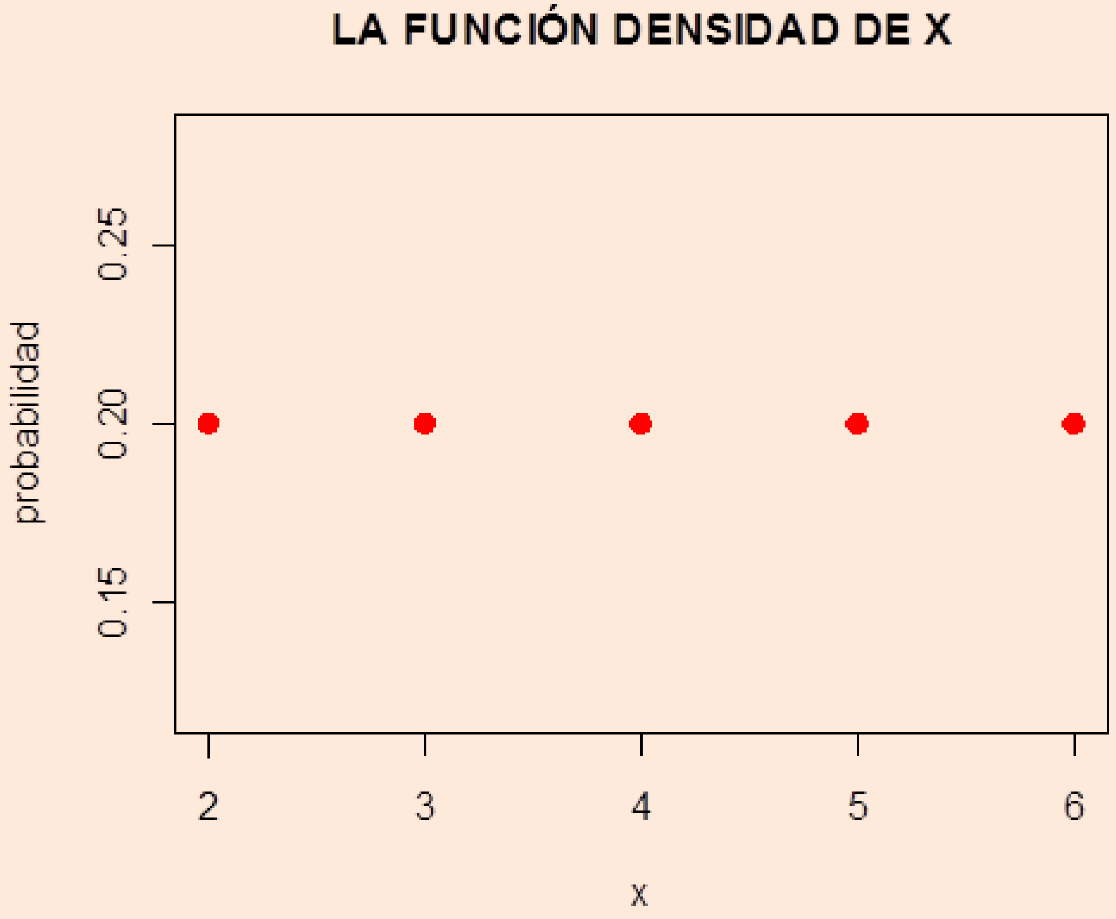
- Esperanza poblacional

$$\mu = E(X) = \sum_{x_i=2}^{6} x_i f_X(x_i) = \frac{1}{5}(2+3+4+5+6) = \frac{20}{5} = 4$$

- Varianza poblacional

$$\sigma^2 = E(X-\mu)^2 = \sum_{x_i=2}^{6} (x_i-\mu)^2 f_X(x_i) = \frac{1}{5}\left(4+1+0+1+4\right) = \frac{10}{5} = 2$$

Función densidad de la población



Distribución muestral del estadístico $\overline{X}$

- Para determinar $f_{\overline{X}}(\overline{x})$ tenemos en cuenta las muestras obtenidas y en cada una de ellas determinamos el valor observado de la media muestral.

$$f_X(x) = \begin{cases} \frac{1}{5} & sí \quad x = 2,3,4,5,6 \\ 0 & otro \quad caso \end{cases}$$

Muestras	2,2	2,3	2,4	2,5	2,6	3,2	3,3	3,4	3,5	3,6	4,2	4,3	4,4	4,5	4,6
$\overline{x}$	2	2.5	3	3.5	4	2.5	3	3.5	4	4.5	3	3.5	4	4.5	5

Muestras	5,2	5,3	5,4	5,5	5,6	6,2	6,3	6,4	6,5	6,6
$\overline{x}$	3.5	4	4.5	5	5.5	4	4.5	5	5.5	6

Distribución muestral del Estadístico $\overline{X}$

Ya que todas estas muestras tienen la misma probabilidad de ser obtenida y son 25 muestras, la probabilidad de cada una es 1/25. Luego, la distribución del estadístico es:

$\overline{x}$	2	2.5	3	3.5	4	4.5	5	5.5	6
$f_{\overline{X}}\left(\overline{x}\right)$	$\frac{1}{25}$	$\frac{2}{25}$	$\frac{3}{25}$	$\frac{4}{25}$	$\frac{5}{25}$	$\frac{4}{25}$	$\frac{3}{25}$	$\frac{2}{25}$	$\frac{1}{25}$

$$f_{\overline{X}}(x) = \begin{cases} \dfrac{1}{25} \; si \;\; x = 2; 6 \\ \dfrac{2}{25} \; si \;\; x = 2.5; 5.5 \\ \dfrac{3}{25} \; si \;\; x = 3; 5 \\ \dfrac{4}{25} \; si \;\; x = 3.5; 4.5 \\ \dfrac{5}{25} \; si \;\; x = 4 \end{cases}$$

Esperanza y Varianza de $\overline{X}$

- Esperanza de $\overline{X}$

$$E(\overline{X}) = \sum_{x_i=2}^{6} \overline{x}_i f_{\overline{X}}(x_i) = (2+6)\frac{1}{25} + (2.5+5.5)\frac{2}{25} + (3+5)\frac{3}{25} + (3.5+4.5)\frac{4}{25} + 4\frac{5}{25} =$$

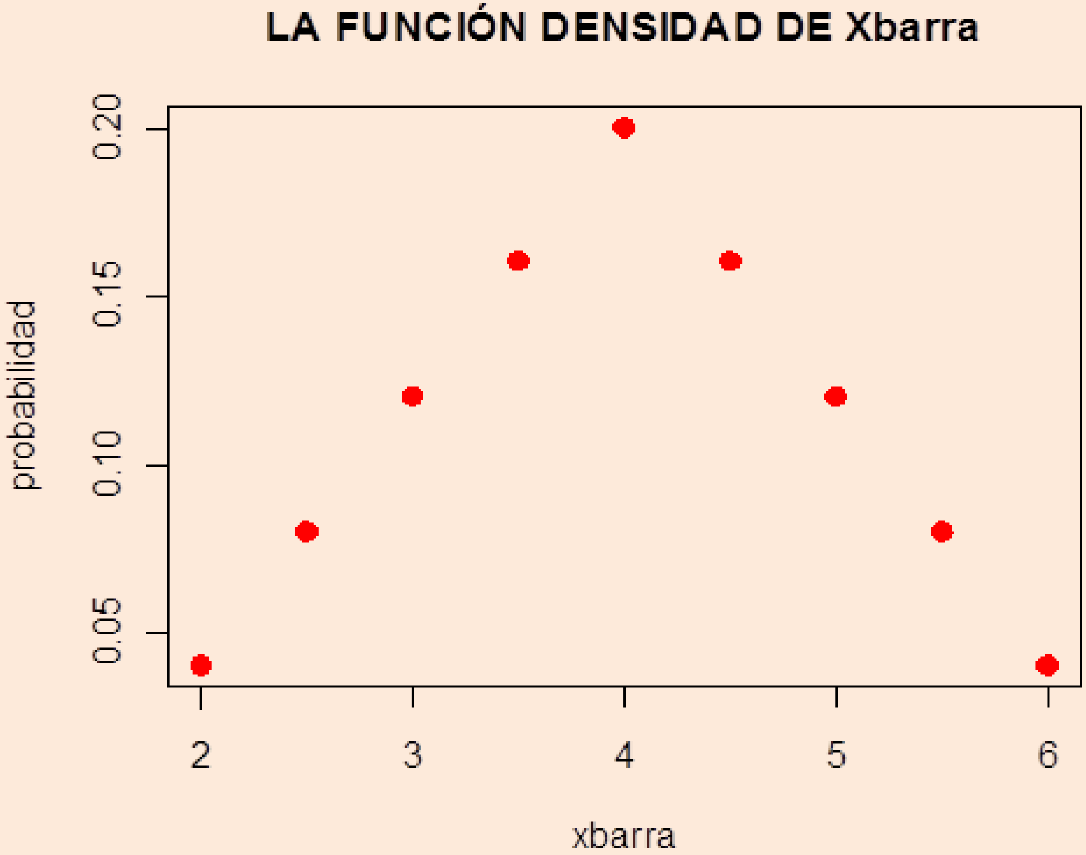
$$= \frac{8}{25} + \frac{16}{25} + \frac{24}{25} + \frac{32}{25} + \frac{20}{25} = 4 = \mu = E(X)$$

- Varianza de $\overline{X}$

$$\sigma^2_{\overline{X}} = \operatorname{var}(\overline{X}) = E(\overline{X} - \mu)^2 = \sum_{x_i=2}^{6} (\overline{x}_i - \mu)^2 f_{\overline{X}}(\overline{x}_i) =$$

$$= 2\frac{1}{25}4 + 2\frac{2}{25}2.25 + 2\frac{3}{25} + 2\frac{4}{25}0.25 = 1 = \frac{\sigma^2}{2}$$

Función de probabilidad de $\overline{X}$



Vemos que los valores de se distribuyen alrededor de en forma simétrica y con menor varianza que la variable original de la población.

Distribución de la media muestral

- El problema es que se observa una gran variabilidad, Esto hace que los valores observados en muestras de tamaño 2 no presenten un buen comportamiento para darnos información respecto del parámetro desconocido de la población bajo estudio μ.

- Si tomáramos muestras más grandes, la distribución tendría mejores características, esto puede verse en el hecho de que la varianza y, por lo tanto, su raíz cuadrada, llamada error estándar, disminuyen a medida que aumenta el tamaño de muestra.

Distribución muestral de $\overline{X}$
Ejemplo 2

- Se ha medido las alturas de cuatro personas, en centímetros, que serán nuestra "población". Siendo esta población de tamaño 4, podemos seleccionar 16 muestras aleatorias de tamaño 2

1	*2*	*3*	*4*	μ	σ
183	185	188	190	**186,5**	**2,6926**

- Partimos de una distribución uniforme para X

$$f_X(x) = \begin{cases} \dfrac{1}{4} & sí \quad x = 183, 185, 188, 190 \\ 0 & otro \quad caso \end{cases}$$

Distribución de la media muestral con reemplazo

- Muestras de tamaño 2

Muestra	Observación 1	Observación 2	$\overline{x}_i$
n_1	183	183	183,0
n_2	183	185	184,0
n_3	183	188	185,5
n_4	183	190	186,5
n_5	185	183	184,0
n_6	185	185	185,0
n_7	185	188	186,5
n_8	185	190	187,5
n_9	188	183	185,5
n_{10}	188	185	186,5
n_{11}	188	188	188,0
n_{12}	188	190	189,0
n_{13}	190	183	186,5
n_{14}	190	185	187,5
n_{15}	190	188	189,0
n_{16}	190	190	190,0

Distribución de la media muestral con reemplazo

- Hemos partido de una distribución uniforme para X con f(x) = 1/4

$\overline{x}_i$	$f(\overline{x}_i)$
183,0	0,063
184,0	0,125
185,0	0,063
185,5	0,125
186,5	0,250
187,5	0,125
188,0	0,063
189,0	0,125
190,0	0,063

$$\mu_{\overline{X}} = 186{,}5$$

$$\sigma_{\overline{X}} = 1{,}9039$$

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{2{,}6926}{\sqrt{2}} = 1{,}9039$$

Obtenemos para $\overline{X}$ una distribución simétrica con media

$$\mu_{\overline{X}} = \mu$$

$$\operatorname{var}(\overline{X}) = \frac{\sigma_X^2}{n}$$

Distribuciones de la media muestral en poblaciones finitas o sin reemplazo

- En el ejemplo, bajo un muestreo sin reemplazo, el número de muestras es 6:

Muestra	Observación 1	Observación 2	$\overline{x}_i$
n_1	183	185	184,0
n_2	183	188	185,5
n_3	183	190	186,5
n_4	185	188	186,5
n_5	185	190	187,5
n_6	188	190	189,0

$$\mu_{\overline{X}} = E(\overline{X}) = 186{,}5$$

$$\sigma_{\overline{X}} = 1{,}5546$$

- El valor de la desviación estándar no es 1,9039 (valor que debería dar)
- En este caso, la varianza de la media muestral no es igual a la varianza poblacional dividido el tamaño de la muestra. Existe una relación entre éstas y está dada por el factor de corrección por finitud:

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}} \qquad \sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}} = \frac{2{,}6926}{\sqrt{2}} \cdot \sqrt{\frac{4-2}{4-1}} = 1{,}5546$$

TEOREMA DEL LIMITE CENTRAL

- Sea *X* una variable aleatoria con función densidad de media μ y varianza σ^2 finitas, si se toma una muestra aleatoria de tamaño *n* y se obtiene $\bar{X}_n$, luego la distribución de $\bar{X}_n$ tiende a una distribución normal cuando $n \to \infty$.

$$\boxed{\bar{X}_n \xrightarrow[n\to\infty]{} N\left(\mu, \frac{\sigma^2}{n}\right)}$$

- Si estandarizamos $\bar{X}_n$, obtenemos una variable Z: $Z = \dfrac{\bar{X} - \mu}{\dfrac{\sigma}{\sqrt{n}}}$
-

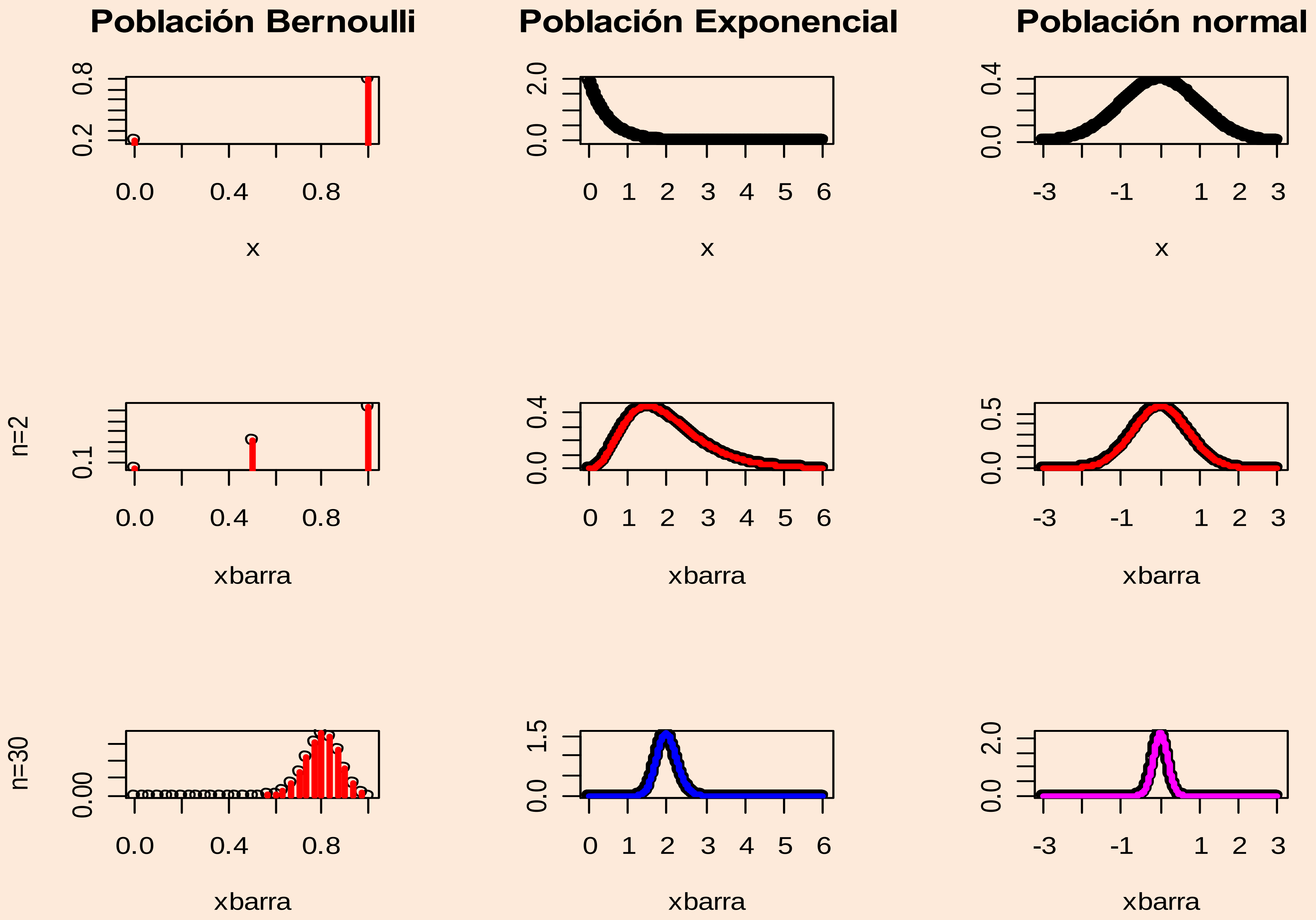
$$\boxed{\bar{Z} \to N(0,1)}$$

$$n \to \infty$$

Teorema del límite central

- El teorema del límite central se puede aplicar para una muestra aleatoria de cualquier distribución siempre que μ y σ^2 sean finitos y el tamaño de la muestra sea grande.

- En general, la aproximación será buena si $n \geq 30$
- Si n<30 , la distribución muestral de $\bar{X}_n$ será normal **sólo si** la distribución de *X* es normal.



Población Bernoulli
Población Exponencial
Población normal
0.2
0.8
0.0
0.4
0.8
x
0.0
2.0
0
1
2
3
4
5
6
x
0.0
0.4
-3
-1
1
2
3
x
n=2
0.1
0.0
0.4
0.8
xbarra
0.0
0.4
0
1
2
3
4
5
6
xbarra
0.0
0.5
-3
-1
1
2
3
xbarra
n=30
0.00
0.0
0.4
0.8
xbarra
0.0
1.5
0
1
2
3
4
5
6
xbarra
0.0
2.0
-3
-1
1
2
3
xbarra

Teorema del Límite central

- En la primera fila se representan las distribuciones de tres poblaciones diferentes. En las filas siguientes se representan las distribuciones muestrales de cuando el tamaño de las muestras consideradas (*n*) es igual a 2 y 30 respectivamente, tomadas estas de las poblaciones (infinitas) indicadas en la primera fila.
- Observemos que:
- 1.- Cuando la población es normal la distribución de $\overline{X}_n$ es normal, cualquiera sea *n*.
- 2.- Las distribuciones que están en la misma columna tienen todas la misma media μ, pero la varianza decrece a medida que crece n.
- 3.- Cuando la distribución de la población es simétrica más rápidamente tiene una distribución aproximadamente normal.
- 4.- Cuando *n*= 30 la distribución de es aproximadamente normal, cualquiera sea la población de la cual provienen las muestras.
- Se considera entonces que para tamaños muestrales n≥30, la aproximación a la distribución normal será bastante buena.

Ejemplo de aplicación del TLC

- De acuerdo con la información que suministra la compañía telefónica, el pago mensual promedio de todos los abonados de la Ciudad de Mendoza es de $153 con una desviación estándar de $41, Se toma una muestra de tamaño 36 de esa población ¿cuál es la probabilidad de que el pago promedio sea inferior a $140?
- Como el tamaño de muestra es mayor que 30 se puede considerar que la distribución de $\bar{X}$ es aproximadamente normal, entonces
- X: Pago de los abonados de una compañía de Teléfono de la Ciudad Mendoza
- $\bar{X}_n$: Pago promedio de los abonados de una compañía de Teléfono de Mza

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{41}{\sqrt{36}} = 6{,}83$$

$$P(\bar{X} < 140) = pnorm(140{,}153{,}6{,}83) \cong 0{,}0287$$

Corolario

- Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria, la distribución de la variable aleatoria $\sum_{i=1}^{n} X_i$ es asintóticamente normal con media $n \cdot \mu$ y varianza $n \cdot \sigma^2$
- En símbolos:
- cuando

$$\boxed{\sum_{i=1}^{n} X_i \underset{n \to \infty}{\to} N(n\,\mu, \quad n\,\sigma^2)}$$

Muestreo de una población normal
DISTRIBUCIÓN DE LA MEDIA MUESTRAL

- Sea $X_1, X_2, ..., X_n$ una muestra aleatoria de una población $N(\mu, \sigma^2)$ entonces

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

Muestreo de una población normal

- Propiedad

$$X_1, X_2, \ldots, X_n \overset{i.id.}{\sim} N(\mu, \sigma^2) \Rightarrow \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

Muestreo de una población normal Varianza muestral

Si S^2 es la varianza muestral en una muestra aleatoria de tamaño n tomada de una población normal con varianza σ^2, entonces la relación

$$U = \frac{S^2(n-1)}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{X_i - \overline{X}}{\sigma}\right)^2 \sim \chi^2_{n-1}$$

Tiene distribución ji-cuadrado con ν=n-1 grados de libertad

Muestreo de una población normal Propiedades

- Si $Z \sim N(0,1)$ $U \sim \chi_n^2$ *Z* y *U* son independientes
- Entonces $$T = \frac{Z}{\sqrt{\frac{U}{n}}} \sim t_n$$
- Una consecuencia $$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1}$$
- demostración
- $Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ $U = \frac{(n-1)S^2}{\sigma^2}$ Z y U son variables a.i.

Muestreo de una población normal Propiedades

Continuación demostración

$$Z \sim N(0,1) \qquad U \sim \chi^2_{n-1}$$

- Por propiedad anterior $$T = \frac{Z}{\sqrt{\frac{U}{n}}} \sim t_n$$

$$T = \frac{Z}{\sqrt{\frac{U}{n-1}}} = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \cdot \frac{\sqrt{n-1}}{\frac{\sqrt{(n-1)S^2}}{\sigma}} = \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1}$$

$$\boxed{T = \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1}}$$

Muestreo de dos poblaciones Normales
Propiedad para obtener una variable F que tiene distribución f Fischer-Snedecor

- Si $\left.\begin{array}{l} U \sim \chi^2_m \\ V \sim \chi^2_n \\ \text{U y V v.a.i.} \end{array}\right\} \Rightarrow F = \dfrac{U/m}{V/n} \sim \mathrm{f}_{\mathrm{m,n}}$

Muestreo de dos poblaciones Normales
Corolario

- Sea $X_1, X_2, \ldots, X_m$ una muestra de una población normal $N(\mu_X, \sigma_X{}^2)$

  $Y_1, Y_2, \ldots, Y_n$ una muestra (independiente de la anterior) de una población normal $N(\mu_Y, \sigma_Y{}^2)$

$$\left.\begin{array}{l} U = \dfrac{(m-1)S_X^2}{\sigma_X^2} \sim \chi^2_{m-1} \\ V = \dfrac{(n-1)S_Y^2}{\sigma_Y^2} \sim \chi^2_{n-1} \\ \text{U y V v.a.i} \end{array}\right\} \Longrightarrow F = \frac{U/(m-1)}{V/(n-1)} = \frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} \sim f(m-1, n-1)$$

Muestreo de una población Normal
Distribución F-FISHER

TEOREMA

Si F tiene una distribución f con ν_1 y ν_2 grados de libertad, entonces F´=1/F tiene una distribuciónn f con ν_2 y ν_1

$$f_{(1-\alpha);(\nu_1,\nu_2)} = \frac{1}{f_{\alpha;(\nu_2,\nu_1)}}$$

Muestreo de dos poblaciones Normales Distribución de la Diferencia de Medias

1 Caso: Sea Población normal y varianzas conocidas

$$\left.\begin{array}{l} X_1 \sim N(\mu_1, \sigma_1^2)\ se\ toma\ una\ muestra\ de\ tamaño\ n_1 \\ \\ X_2 \sim N(\mu_2, \sigma_2^2) se\ toma\ una\ muestra\ de\ tamaño\ n_2 \end{array}\right\} \rightarrow \begin{array}{l} \bar{X}_1 \sim N\left(\mu_1, \dfrac{\sigma_1^2}{n_1}\right) \\ \bar{X}_2 \sim N(\mu_2, \dfrac{\sigma_2^2}{n_2}) \end{array}$$

$$\left.\begin{array}{l} \bar{X}_1 \sim N\left(\mu_1, \dfrac{\sigma_1^2}{n_1}\right) \\ \bar{X}_2 \sim N(\mu_2, \dfrac{\sigma_2^2}{n_2}) \end{array}\right\} \rightarrow \bar{X}_1 - \bar{X}_2 \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$$

$$Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \sim Z(0,1)$$

Muestreo de dos poblaciones Normales

Distribución de la Diferencia de Medias

- 2 Caso: Población normal y varianzas desconocidas iguales
- 3 Caso: Población normal y varianzas desconocidas y distintas
- Ver como se distribuye la diferencia de medias por el libro subido a la página de la cátedra

Estadístico para Diferencia de Medias

4to Caso: Población con distribución desconocida

$$\left.\begin{array}{l} X_1 \sim f(\mu_{1,}\ \sigma^2{}_1) \\ X_2 \sim f(\mu_{2,}\ \sigma^2{}_2) \end{array}\right\}$$

Se toman nuestras a.i. de tamaño n_1 y n_2 y se obtienen $\overline{X}_1$ y $\overline{X}_2$

Si consideramos la v.a. $\overline{X}_1 - \overline{X}_2$ se puede definir Z:

.

$$Z = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \rightarrow Z(0,1)$$

La distribución Z tiende a una distribución normal estándar cuando n$\rightarrow\infty$ .Es decir $\overline{X}_1 - \overline{X}_2$ es asintóticamente normal con media μ1 y μ2 y error estándar $\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$

Distribución F-FISHER

TEOREMA

Sea S^2_1 y S^2_2 son las varianzas de muestras aleatorias independientes de tamaño n_1 y n_2 tomadas de poblaciones normales con varianzas σ^2_1 y σ^2_2 entonces

$$F = \frac{U/\nu}{V/\nu} = \frac{\dfrac{S_1^2(n-1)}{\sigma_1^2(n-1)}}{\dfrac{S_2^2(m-1)}{\sigma_2^2(m-1)}} = \frac{\dfrac{S_1^2}{\sigma_1^2}}{\dfrac{S_2^2}{\sigma_2^2}} = \frac{S_1^2}{\sigma_1^2}\cdot\frac{\sigma_2^2}{S_2^2} = \frac{\sigma_2^2}{\sigma_1^2}\cdot\frac{S_1^2}{S_2^2} \sim f_{n_1-1,n_2-1}$$

Tiene una distribución F con $\nu_1=n_1-1$ y $\nu_2=n_2-1$ (gl)