INTRODUCCIÓN A LA INFERENCIA ESTADISTICA

Daniel E. González
Pontificia Universidad
Javeriana Cali

¿QUE ES....

CUANDO HACER...

COMO LO HACEMOS...

INTRODUCCIÓN A LA INFERENCIA ESTADISTICA

CONCEPTOS BÁSICOS

- POBLACIÓN
- CENSO
- PARÁMETRO (θ)

MUESTRA

MUESTREO

ESTIMADOR ($\hat{\theta}$)

TAMAÑO DE MUESTRA

TIPOS DE MUESTREO

DISTRIBUCIONES MUESTRALES

- N
- t-Student
- χ^2
- F

MODELOS DE PROBABILIDAD

$f(x)$ $F(x)$ $E(X)$ $V(X)$

$E(XY)$ $COV(XY)$ ρ_{xy}

- NORMAL (μ, σ^2)
- UNIFORME (a, b)
- EXPONENCIAL (λ)
- WEIBULL (α, β)

- BINOMIAL (n, p)
- POISSON (λ)
- GEOMETRICA (p)

- **PUNTUAL**

- **INTERVALOS DE CONFIANZA**

- MÉTODOS DE ESTIMACIÓN
  - M. DE MOMENTOS
  - M. DE MAX. VEROSIMILITUD
- PROPIEDAD DE LOS ESTIMADORES
  - INSESGADEZ
  - EFICIENCIA
  - CONSISTENCIA

- $\bar{X}$ $E(\bar{X})$ $V(\bar{X})$
- $\hat{p}$ $E(\hat{p})$ $V(\hat{p})$

- TEOREMA DEL LIMITE CENTRAL

- **PRUEBAS DE HIPÓTESIS**

INTRODUCCIÓN A LA INFERENCIA ESTADÍSTICA

CONCEPTOS BÁSICOS

- POBLACIÓN
- CENSO
- PARÁMETRO

MUESTRA

MUESTREO

ESTIMADOR

- PUNTUAL
- POR INTERVALOS DE CONFIANZA

- TAMAÑO DE MUESTRA (n)
- TIPOS DE MUESTREO
- DISTRIBUCIONES MUESTRALES
  - N
  - χ^2
  - t-student
  - F

PROPIEDADES DE LOS ESTIMADORES

- INSESGADEZ

EFICIENCIA

CONSISTENCIA

- MODELOS DE PROBABILIDAD

$f(x)$ $F(x)$ $E(x)$ $V(x)$

$E(XY)$ $Cov(XY)$ ρ_{XY}

BINOMIAL(n, p)
POISSON(λ)
GEOMETRICA(p)

NORMAL(μ, σ^2)
UNIFORME(a, b)
EXPONENCIAL(λ)
WEIBULL(α, β)

$\bar{X}$ $E(\bar{X})$ $V(\bar{X})$

$\hat{P}$ $E(\hat{p})$ $V(\hat{p})$

METODOS DE ESTIMACIÓN

M. DE MOMENTOS

M. DE MAX. VEROSIMILITUD

TEOREMA CENTRAL DEL LIMITE

- PRUEBAS DE HIPÓTESIS

DISTRIBUCIONES MUESTRALES

NORMAL $X \sim N(\mu, \sigma^2)$

NORMAL ESTANDAR

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

t-Student

$$T = \frac{Z}{\sqrt{\chi^2/v}} \sim t_{v=n-1}$$

$$\chi^2 = \sum_{i=1}^{n} Z_i^2 \sim \chi^2_{v=n-1}$$

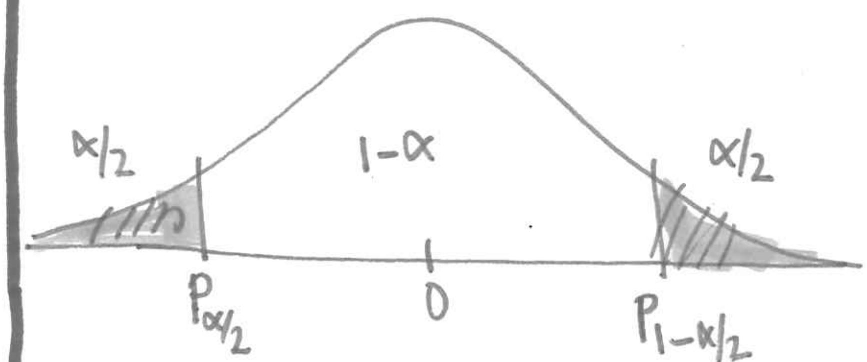
F

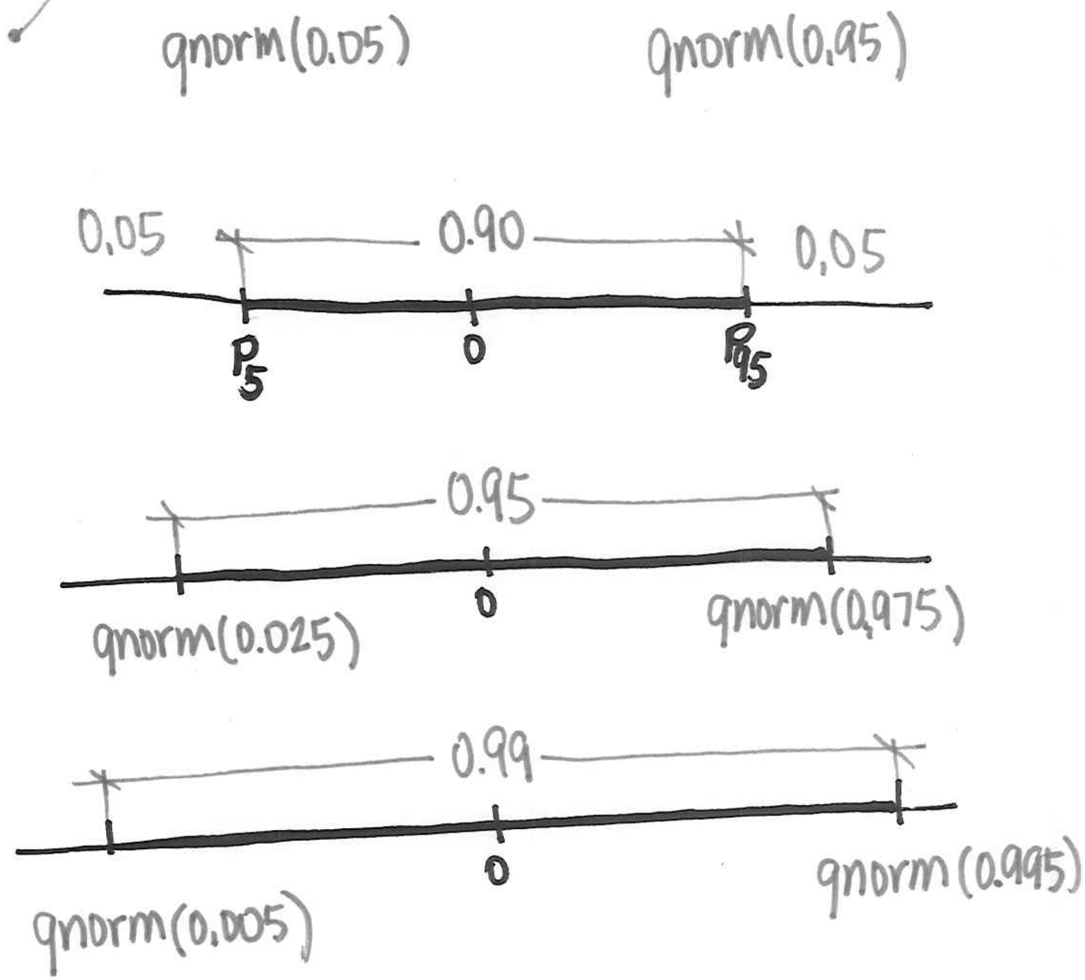
$$\frac{\chi_1^2/v_1}{\chi_2^2/v_2} \sim f_{\substack{v_1=n_1-1 \\ v_2=n_2-1}}$$

NORMAL $X \sim N(\mu=100, \sigma^2=625)$

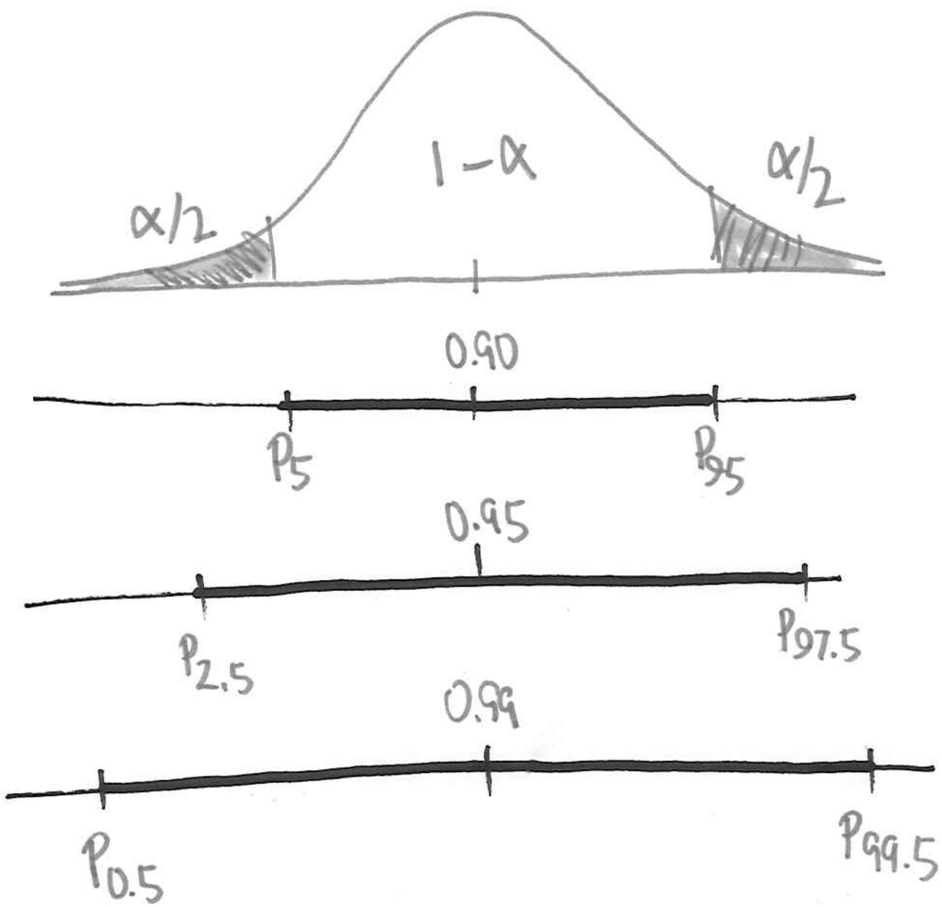
NORMAL ESTANDAR

$$Z = \frac{X-100}{25} \sim N(0,1)$$





t-Student $T_{v:n-1}$

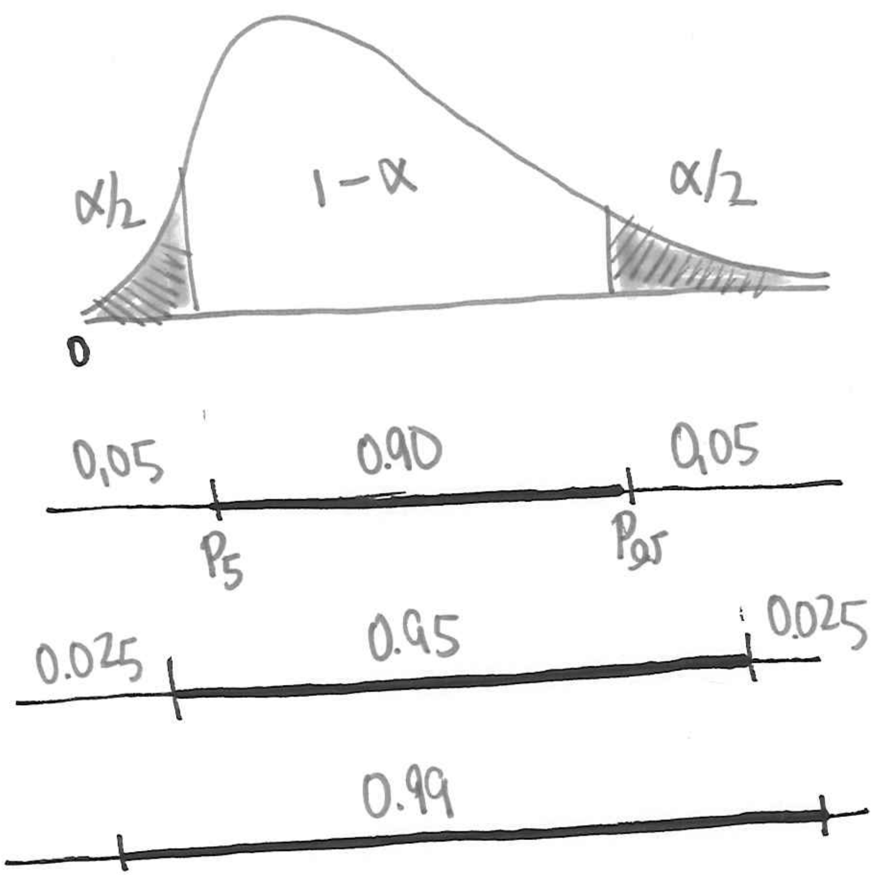


$n=16$

qt(0.005,15)

qt(0.995,15)

CHI-CUADRADO

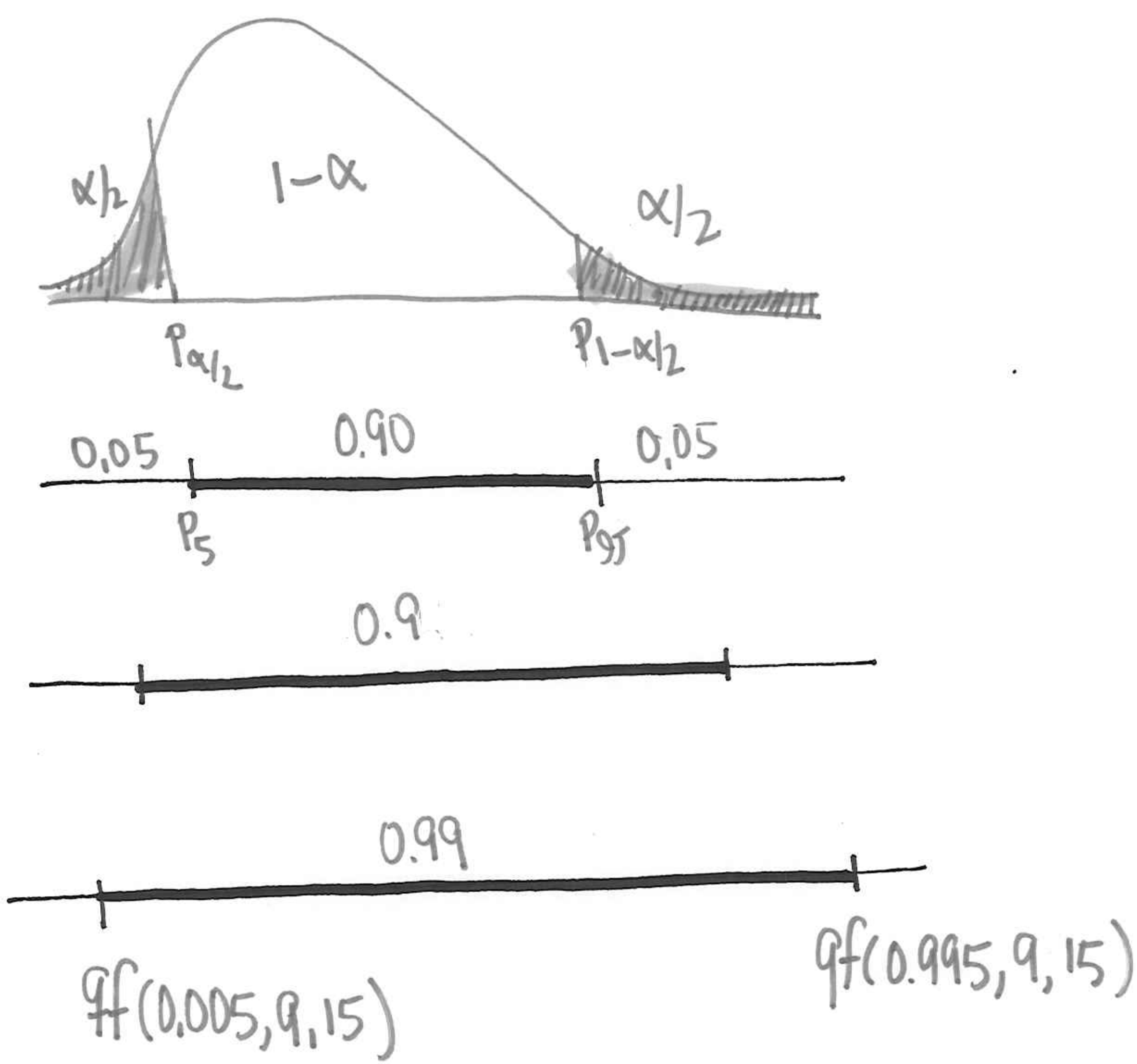


qchisq(0.995, 15)

qchisq(0.005, 15)

n=16

F de Fisher



$n_1=10$

$n_2=16$

MEDIA MUESTRAL $\bar{X}$

DISTRIBUCIÓN DE LA MEDIA MUESTRAL ($\bar{X}$)

SUPUESTO: $E(X) = \mu$

$V(X) = \sigma^2$

$E(\bar{X})$

$V(\bar{X})$

$$E(\bar{X}) = \frac{1}{n}\left(E(X_1 + X_2 + \cdots X_n)\right)$$

$$= \frac{1}{n}\left(E(X_1) + E(X_2) + \cdots + E(X_n)\right)$$

$$= \frac{1}{n}\left(\mu + \mu + \mu + \cdots \mu\right) = \frac{1}{n} n\mu = \mu$$

$$V(\bar{X}) = V\left(\frac{1}{n}(X_1 + X_2 + \cdots X_n\right) = \frac{1}{n^2}\left(V(X_1) + V(X_2) + \cdots + V(X_n)\right)$$

$$= \frac{1}{n^2}\left(\sigma^2 + \sigma^2 + \cdots \sigma^2\right) = \frac{1}{n^2} n\sigma^2$$

$$= \frac{\sigma^2}{n}$$

PROPORCIÓN MUESTRAL $\hat{p}$

DISTRIBUCIÓN DE LA PROPORCIÓN MUESTRAL $\hat{p}$

$$\hat{p} : \frac{\text{\# DE EVENTOS QUE CUMPLEN UNA CARACTERISTICA}}{n} = \frac{\text{SUMA DE VARIABLES BERNOULLI}}{n} = \frac{\sum X_i}{n}$$

$X \sim$ BERNOULLI

$E(X) = p$

$V(X) = pq$

$$E(\hat{p}) = \frac{1}{n} E(\sum X_i) = \frac{1}{n}\left(E(X_1) + E(X_2) + \cdots + E(X_n)\right)$$

$$\frac{1}{n}\left(p + p + \cdots p\right) = \frac{1}{n} np = p$$

$$V(\hat{p}) = \frac{1}{n^2} V\left(\sum X_i\right) = \frac{1}{n^2}\left(V(X_1) + V(X_2) + \cdots V(X_n)\right)$$

$$\frac{1}{n^2}\left(pq + pq + \cdots pq\right) = \frac{1}{n^2} npq = \frac{pq}{n}$$

MÉTODOS DE ESTIMACIÓN

MÉTODO DE MOMENTOS

MÉTODO:

MOMENTOS POBLACIONALES	MOMENTOS MUESTRALES
$\mu' = E(X) = \mu$	$m' = \frac{\sum X_i}{n} = \bar{X}$
$\mu'' = E(X^2)$	$m'' = \frac{\sum X_i^2}{n}$
$\vdots$	$\vdots$
$\mu^k = E(X^k)$	$m^k = \frac{\sum X_i^k}{n}$

NOTA

$$E(X^k) = \sum_{R_X} x^k f(x)$$

VA DISCRETAS

$$E(X) = \int_{-\infty}^{\infty} x^k f(x)\,dx$$

VA CONTINUAS

MÉTODO DE MÁXIMA VEROSIMILITUD

FUNCIÓN DE VEROSIMILITUD

$$L(\theta, X_1, X_2, \dots X_n) = f(x_1)\, f(x_2) \dots f(x_n)$$

MÉTODO: MÁXIMIZAR L

- ENCONTRAR LOS VALORES DE θ QUE MAXIMICEN

  $$L = P(X_1 = x_1,\ X_2 = x_2, \dots X_n = x_n)$$

- EQUIVALENTE A REALIZAR EL PROCESO CON $\ln L$

ENCONTRAR EL ESTIMADOR DE MÁX VEROSIMILITUD PARA μ EN UNA DISTRIBUCIÓN NORMAL

$$f(x) = 1/\sqrt{2\pi\sigma^2}\, e^{-1/2\sigma^2 (x-\mu)^2}$$

$$L = \prod_{i=1}^{n} f(x_i) = (2\pi\sigma^2)^{-n/2}\, e^{-1/2\sigma^2 \sum (x_i-\mu)^2}$$

$$\ln L = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum (x_i-\mu)^2$$

PROPIEDADES DE LOS ESTIMADORES

PROPIEDADES DE LOS ESTIMADORES

- **INSESGADEZ** UN ESTIMADOR $\hat{\theta}$ ES INSESGADO SI $E(\hat{\theta}) = \theta$

- **EFICIENCIA** UN ESTIMADOR $\hat{\theta}_1$ ES MÁS EFICIENTE QUE OTRO ESTIMADOR $\hat{\theta}_2$ CUANDO

$$V(\hat{\theta}_1) < V(\hat{\theta}_2)$$

- **CONSISTENCIA** CUANDO UN ESTIMADOR SIENDO SESGADO SE CONVIERTE EN INSESGADO CUANDO AUMENTA EL TAMAÑO DE LA MUESTRA SE DICE QUE ESTE ESTIMADOR ES CONSISTENTE

Ej. PARA UNA POBLACIÓN CON $E(X) = \mu$

$$V(X) = \sigma^2$$

VERIFICAR SI $\bar{X}$ ES UN ESTIMADOR INSESGADO

Ej. PARA UNA MUESTRA OBTENIDA DE UNA POBLACIÓN EXPONENCIAL CON PARAMETRO β $(E(X) = \beta)$

$$V(X) = \beta^2$$

EXAMINAR LOS SIGUIENTES ESTIMADORES PARA UNA MUESTRA (X_1, X_2, X_3, X_4)

$$T_1 = \frac{1}{6}(X_1 + X_2) + \frac{1}{3}(X_4 + X_3)$$

$$T_2 = \frac{1}{10}(X_1 + 2X_2 + 3X_3 + 4X_4)$$

$$T_3 = \frac{1}{4}(X_1 + X_2 + X_3 + X_4)$$

$$E(T_1) = E(\tfrac{1}{6}(X_1+X_2)) + \tfrac{1}{3}(X_3+X_4))$$

$$\tfrac{1}{6}(E(X_1)+E(X_2)) + \tfrac{1}{3}(E(X_3)+E(X_4))$$

$$\tfrac{1}{6}(\beta+\beta) + \tfrac{1}{3}(\beta+\beta) = \tfrac{2}{6}\beta + \tfrac{2}{3}\beta = \frac{2+4}{6}\beta = \beta$$

$$E(T_2) = \tfrac{1}{10}(E(X_1+2X_2+3X_3+4X_4))$$

$$\tfrac{1}{10}(E(X_1)+2E(X_2)+3E(X_3)+4E(X_4))$$

$$\tfrac{1}{10}(\beta+2\beta+3\beta+4\beta) = \tfrac{1}{10}\,10\beta = \beta$$

$$E(T_3) = \tfrac{1}{4}(E(X_1+X_2+X_3+X_4)) = \tfrac{1}{4}(E(X_1)+E(X_2)+E(X_3)+E(X_4))$$

$$= \tfrac{1}{4}(\beta+\beta+\beta+\beta) = \tfrac{1}{4}\,4\beta = \beta.$$

$$V(T_1) = V\left(\frac{1}{6}(X_1 + X_2) + \frac{1}{3}(X_3 + X_4)\right) = \frac{1}{36}\left(V(X_1) + V(X_2)\right) + \frac{1}{9}\left(V(X_3) + V(X_4)\right)$$

$$= \frac{1}{36}(\beta^2 + \beta^2) + \frac{1}{9}(\beta^2 + \beta^2) = \frac{1}{36}(2\beta^2) + \frac{1}{9}(2\beta^2) = \frac{5}{18}\beta^2$$

$$V(T_2) = V\left(\frac{1}{10}(X_1 + 2X_2 + 3X_3 + 4X_4\right) = \frac{1}{100}(\beta^2 + 4\beta^2 + 9\beta^2 + 16\beta^2)$$

$$= \frac{1}{100}(30\beta^2) = \frac{3}{10}\beta^2$$

$$V(T_3) = V\left(\frac{1}{4}(X_1 + X_2 + X_3 + X_4)\right) = \frac{1}{16}(\beta^2 + \beta^2 + \beta^2 + \beta^2) = \frac{1}{4}\beta^2$$

DISTRIBUCIONES MUESTRALES Y SU RELACIÓN CON LOS PRINCIPALES ESTIMADORES MUESTRALES

- $\bar{X}$ CUANDO $X \sim N(\mu, \sigma^2)$ — $\bar{X} \sim N(\mu, \sigma^2/n)$ SUPUESTO σ^2 CONOCIDA

  └ $\bar{X} \sim t_{v=n-1}$ σ^2 DESCONOCIDA

  CUANDO $X \sim ?$ $\bar{X} \sim N(\mu, \sigma^2/n)$ $\frac{\bar{X}-\mu}{s/\sqrt{n}}$

  $n >>$

- $\hat{p}$ $p \sim N(\hat{p}, \hat{p}\hat{q}/n)$ $n >>$

- s^2 $\frac{(n-1)s^2}{\sigma_0^2} \sim \chi^2_{v=n-1}$

  $X \sim N(\mu, \sigma^2)$

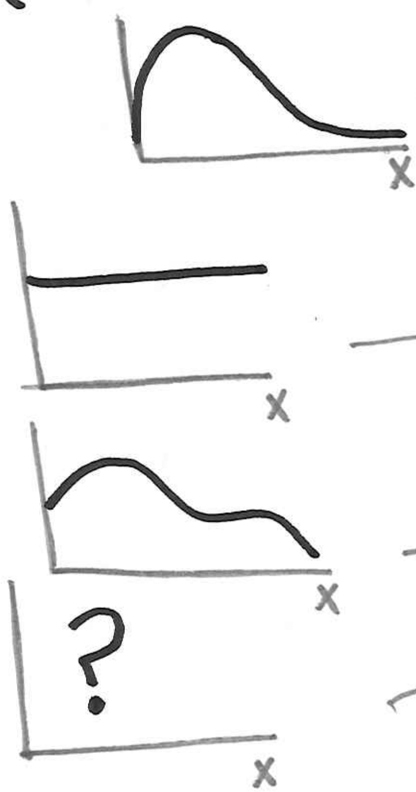
TEOREMA CENTRAL DEL LÍMITE

TEOREMA CENTRAL DEL LIMITE

$X \sim N(\mu, \sigma^2)$

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

$X \sim$



$n >>$

TCL en R

Daniel Enrique González Gómez
Dep. Ciencias Naturales y Matemáticas
Facultad de Ingeniería y Ciencias
Pontificia Universidad Javeriana
Cali