TEMA XIV: INTERVALOS DE CONFIANZA

Intervalo de Confianza

A estimación puntual dun parámetro, desenrolada no tema anterior, permíte-nos calcular un valor para ese parámetro, pero non precisa o grau de certeza de dita estimación. Agora imos ver como se fai para calcular un intervalo tal que, escollendo unha probabilidade prefixada, o verdadeiro valor do parámetro se atope nese intervalo.

Intervalo de Confianza para o Parámetro q con Nivel de Confianza 1-a

é o intervalo $[\theta_1,\theta_2]$ tal que os valores θ_1,θ_2 (que dependen da mostra) se calculan de maneira que, se construímos muitos intervalos —cada un deles a partir dunha mostra diferente— o $(1-\alpha)\cdot 100\%$ deles conterá o verdadeiro valor do parámetro θ.

Ex.

$\alpha=0.10 \Rightarrow$ se obtemos 100 intervalos, cada un dunha mostra distinta, 90 deles conterán o valor verdadeiro do parámetro θ.

Fixado α, para achar θ_1, θ_2 necesíta-se un estatístico pivote $\omega=g(\theta,\vec{x})$, variábel aleatória función do parámetro e da mostra cuia distribución é coñecida. O estadístico pivote é v.a. contínua e monótona respeito a θ.

É posíbel achar a, b/$P(a\leq g(\theta,\vec{x})\leq b)=1-\alpha$ por ser contínua e monótona $P(g^{-1}(a,\vec{x})\leq\theta\leq g^{-1}(b,\vec{x}))=1-\alpha$ e chamando $\theta_1=g^{-1}(a,\vec{x})$, $\theta_2=g^{-1}(b,\vec{x})$ obtén-se o intervalo de confianza $[\theta_1,\theta_2]/P(\theta_1\leq\theta\leq\theta_2)=1-\alpha$

Problemas:

1º) Como achar $\omega=g(\theta,\vec{x})$? Para $n\to\infty$: $\boldsymbol{\omega=\dfrac{\theta-\hat{\theta}_{mv}}{\sigma(\hat{\theta}_{mv})}\in N(0,1)}$

2º) Como son a e b? Tales que o intervalo tén lonxitude mínima

sexa simétrica ou asimétrica tomamos intervalo centrado ($\alpha/2$ a cada lado)

3º) Como elexir α? Arbitrariamente, segundo a precisión desexada.

(I) Intervalo de Confianza para a Média μ dunha Povoación Normal*

...coñecendo a varianza povoacional: $x\in N(\mu,\sigma)$, σ coñecido

$\theta=\mu$; $\hat{\theta}=\bar{x}$; $\bar{x}\in N(\mu,\sigma/\sqrt{n})$; $\omega=\bar{x}-\mu/(\sigma/\sqrt{n})\in N(0,1)$

$P(-z_{\alpha/2}\leq\omega\leq z_{\alpha/2})=1-\alpha$ onde $z_{\alpha/2}/P(z>z_{\alpha/2})=\alpha/2$; $z\in N(0,1)$

$\boxed{\bar{x}-\sigma/\sqrt{n}\cdot z_{\alpha/2}\leq\mu\leq\bar{x}+\sigma/\sqrt{n}\cdot z_{\alpha/2}}$

...descoñecendo a varianza povoacional: $x\in N(\mu,\sigma)$, σ descoñecido

$\theta=\mu$; $\hat{\theta}=\bar{x}$; $\bar{x}\in N(\mu,\hat{s}/\sqrt{n})$; $\omega=\bar{x}-\mu/(\hat{s}/\sqrt{n})\in t_{n-1}$

$P(-t_{\alpha/2}\leq\omega\leq t_{\alpha/2})=1-\alpha$ onde $t_{\alpha/2}/P(t>t_{\alpha/2})=\alpha/2$; $t\in t_{n-1}$

$\boxed{\bar{x}-\hat{s}_{n-1}/\sqrt{n}\cdot t_{\alpha/2}^{n-1}\leq\mu\leq\bar{x}+\hat{s}_{n-1}/\sqrt{n}\cdot t_{\alpha/2}^{n-1}}$ [$\hat{s}_{n-1}=\sigma_{n-1}$ na calculadora]

nota: $n\geq 30\Rightarrow\omega\in N(0,1)$ substitúe-se σ por s

*se a povoación non é normal, o I.C. é aproximado e non exacto como até este ponto.

(II) Intervalo de Confianza para a Desviación Típica

(1)...coñecendo a média μ

$$\boxed{\sqrt{\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\chi^2_{n,1-\alpha/2}}}\ ,\ \sqrt{\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\chi^2_{n,\alpha/2}}}}$$

(2)...descoñecendo a média μ

$$\boxed{\sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{\chi^2_{n-1,1-\alpha/2}}}\ ,\ \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{\chi^2_{n-1,\alpha/2}}}}$$

(III) Intervalo de Confianza para a Diferéncia de Médias dunha Povoación Normal

(1)...varianzas iguais

$\vec{x_1}=(x_1^1,x_2^1,\ldots,x_n^1)\in N(\mu_1,\sigma)$
$\vec{x_2}=(x_1^2,x_2^2,\ldots,x_m^2)\in N(\mu_2,\sigma)$

(1A)...e coñecendo σ

$\bar{x}_1\in N(\mu_1,\sigma/\sqrt{n})$
$\bar{x}_2\in N(\mu_2,\sigma/\sqrt{m})$
$\Rightarrow\ \bar{x}_1-\bar{x}_2\in N(\mu_1-\mu_2,\sqrt{(\sigma^2/n+\sigma^2/m)})$

$$\boxed{\omega=\frac{(\mu_1-\mu_2)(\bar{x}_1-\bar{x}_2)}{\sqrt{(\sigma^2/n+\sigma^2/m)}}\in N(0,1)}$$

(1B)...e descoñecendo σ

estíma-se a partir das cuasivarianzas mostrais:

$$*\hat{s}^2_T=\frac{(n-1)\cdot\hat{s}^2_1+(m-1)\cdot\hat{s}^2_2}{n-1+m-1}$$

$$\boxed{\omega=\frac{(\mu_1-\mu_2)(\bar{x}_1-\bar{x}_2)}{\sqrt{(\hat{s}^2_T/n+\hat{s}^2_T/m)}}\in t_{n+m-2}}$$

(continua)

(cont.) (II) Intervalo de Confianza para a Diferéncia de Médias dunha Povoación Normal

(2)...*varianzas distintas*

$\vec{x_1}=(x_1^1,x_2^1,\ldots,x_n^1)\in N(\mu_1,\sigma_1)$
$\vec{x_2}=(x_1^2,x_2^2,\ldots,x_m^2)\in N(\mu_2,\sigma_2)$

(2A)...*e coñecendo s_1, s_2*

$\bar{x}_1\in N(\mu_1,\sigma_1/\sqrt{n})$
$\bar{x}_2\in N(\mu_2,\sigma_2/\sqrt{m})$
$\Rightarrow \bar{x}_1-\bar{x}_2\in N(\mu_1-\mu_2,\sqrt{(\sigma_1^2/n+\sigma_2^2/m)})$

$$\omega=\frac{(\mu_1-\mu_2)-(\bar{x}_1-\bar{x}_2)}{\sqrt{(\sigma_1^2/n+\sigma_2^2/m)}}\in N(0,1)$$

$$\boxed{(\bar{x}-\bar{y})\pm z_{\alpha/2}\sqrt{[\sigma_1^2/n+\sigma_2^2/m]}}$$

(2B)...*e descoñecendo s_1, s_2*

- *supoñendo que son iguais*

$$\boxed{(\bar{x}-\bar{y})\pm t_{\alpha/2}^{n+m-2}\sqrt{[((n-1)\hat{s}_{n-1}^2+(m-1)\hat{s}_{m-1}^2)/(n+m-2)]}\sqrt{[1/n+1/m]}}$$

- *con tamaños mostrais grandes (n, m³30)*

$$\omega=\frac{(\mu_1-\mu_2)-(\bar{x}_1-\bar{x}_2)}{\sqrt{(s_1^2/n+s_2^2/m)}}\in N(0,1)$$

$$\boxed{(\bar{x}-\bar{y})\pm z_{\alpha/2}\sqrt{[\hat{s}_{n-1}^2/n+\hat{s}_{m-1}^2/m]}}$$

- *con tamaños mostrais pequenos (n, m<30)*

$$\omega=\frac{(\mu_1-\mu_2)-(\bar{x}_1-\bar{x}_2)}{\sqrt{(\hat{s}_1^2/n+\hat{s}_2^2/m)}}\in t_{n+m-2-\Delta}$$

*se falla hipótese de normalidade $n\to\infty$ $\omega\approx N(0,1)$ e non exacto como até este ponto

$$\boxed{(\bar{x}-\bar{y})\pm t_{\alpha/2}^{n+m-2-\Delta}\sqrt{[\hat{s}_{n-1}^2/n+\hat{s}_{m-1}^2/m]}}$$

sendo Δ o enteiro mais próximo a:

$$[(m-1)\hat{s}_{n-1}^2/n-(n-1)\hat{s}_{m-1}^2/m]^2\ /\ [(m-1)(\hat{s}_{n-1}^2/n)^2+(n-1)(\hat{s}_{m-1}^2/m)^2]$$

(3)...*con dados apareados*

x_i,y_i independentes entre si (x's con x's; y's con y's)
pero x_k,y_k para un k determinado están relacionados
(medicións distintas do mesmo individuo no que cambiou unha condición)

calcúla-se $D_i=x_i-y_i\ \forall i$, e fai-se a média $D_i=\bar{x}_D$

e despois, como antes: $(\mu_D-\bar{x}_D)/(\sigma/\sqrt{n})$ $\begin{cases}\in t_{n-1} & \text{se hai que estimar } \sigma \text{ como } \hat{s} \text{(cuasivarianza)}\\ \in N(0,1) & \text{se nos dan } \sigma \text{ (da povoación)}\end{cases}$

(IV) Intervalo de Confianza para a Varianza en Povoacións Normais

$$\boxed{\omega=\frac{ns^2}{\sigma^2}=\frac{(n-1)\hat{s}^2}{\sigma^2}\in X_{n-1}^2}$$

(V) Intervalo de Confianza para a Razón de Varianzas

$$\boxed{\omega=\frac{\hat{s}_1^2}{\hat{s}_2^2}\in F_{n_1-1,n_2-1}}$$

ollo! para $\theta=\sigma_2^2/\sigma_1^2$

como $\dfrac{(n_1-1)\hat{s}_1^2}{\sigma_1^2}\in X^2_{n_1-1}$, $\dfrac{(n_2-1)\hat{s}_2^2}{\sigma_2^2}\in X^2_{n_2-1}$

entón:

$$\frac{\hat{s}_1^2/\sigma_1^2}{\hat{s}_2^2/\sigma_2^2}=\omega\in F_{n_1-1,n_2-1}$$

$F_{n,m}(x)=1/[F_{m,n}(1-x)]$

(VI) Intervalo de Confianza para unha Proporción

$$\boxed{(\hat{p}\pm Z_{\alpha/2}\sqrt{[\hat{p}(1-\hat{p})/n]})}$$

$$\omega=\frac{f-P}{\sqrt{[f(1-f)/n]}}\approx N(0,1)$$

$f=1/n\Sigma_{i=1}^{n}x_i$ proporción mostral $\hat{p}$ de individuos con unha característica nunha povoación dividida entre os que teñen a característica e os que non.

$P=\theta$ proporción povoacional

$\Sigma_{i=1}^{n}x_i=S_n\in B(n,p)\approx$(cando $n\to\infty$) $N(np,\sqrt{(npq)})\Rightarrow f=S_n/n\approx N(np/n,\sqrt{(npq)}/n)=N(p,\sqrt{(pq/n)})$

e normalizando $[f-p]/[\sqrt{(pq/n)}]\approx N(0,1)$ e estimando $[f-p]/[\sqrt{f(1-f)/n}]\approx$(por TCL) $N(0,1)$

(VII) Intervalo de Confianza para a Diferéncia de Proporcións

$$\boxed{\omega=\frac{(f_A-f_B)-(P_A-P_B)}{\sqrt{[f_A(1-f_A)/n_A+f_B(1-f_B)/n_B]}}\in N(0,1)}$$

$X_1^A, X_2^A, ..., X_{nA}^A \Rightarrow f_A=1/n_A\Sigma x_i^A$
$X_1^B, X_2^B, ..., X_{nB}^B \Rightarrow f_B=1/n_B\Sigma x_i^B$
aproximando de binomial a normal
$f_A\approx N(p_A,\sqrt{(p_Aq_A/n_A)})$
$f_B\approx N(p_B,\sqrt{(p_Bq_B/n_B)})$
$\Rightarrow f_A-f_B\approx(p_A-p_B,\sqrt{[p_Aq_A/n_A+p_Bq_B/n_B]})$
normalizando e estimando $\Rightarrow\omega$

(VIII) Intervalos Asintóticos

Sexa θ parámetro dunha distribución dunha povoación e θ_{mv} o seu estimador máximo verosímil.
Asintoticamente ($n\to\infty$): $E(\theta_{mv})\to\theta$; $Var(\theta_{mv})\to -1/[d^2L(\theta_{mv})/d\theta^2]$.
Entón un intervalo asintótico póde-se obter a partir de:

$$\boxed{\omega = \frac{\theta-\theta_{mv}}{\sigma(\theta_{mv})} \approx N(0,1)}$$

sexa cal sexa a distribución da povoación

▪ Exercicio [XIV.1] *Quére-se estimar a média m dunha variábel aleatória normal con varianza 2.5. Tóma-se unha mostra de 100 valores independentes cuia média é 4.3.*

a) Píde-se o intervalo de confianza ao 95%.

b) Cal deberia ser o tamaño da mostra para que a lonxitude do intervalo disminua en 0.08 unidades?

a)

$\sigma^2=2.5 \Rightarrow \sigma=\sqrt{2.5}$

$n=100$

$\bar{x}=4.3$

$1-\alpha=0.95 \Rightarrow \alpha=0.05 \Rightarrow \alpha=0.025 \Rightarrow$

$\Rightarrow z_{0.025}$ é tal que $P(z>z_{0.025})=0.025$

$\Leftrightarrow z_{0.025}$ é tal que $P(z<z_{0.025})=0.975$

$\Rightarrow z_{0.025}=1.96$ [táboas distribución normal estándar, p.336]

$\bar{x}-\sigma/\sqrt{n}\cdot z_{\alpha/2}\le\mu\le \bar{x}+\sigma/\sqrt{n}\cdot z_{\alpha/2}$

$\Rightarrow 4.3-1.58/10\cdot 1.96\le\mu\le 4.3+1.58/10\cdot 1.96$

$\Rightarrow 3.99\le\mu\le 4.61$

$\Rightarrow \mu \in [3.99,4.61]$

b)

$\bar{x}-\sigma/\sqrt{n}\cdot z_{\alpha/2}\le\mu\le \bar{x}+\sigma/\sqrt{n}\cdot z_{\alpha/2}$

lonxitude do intervalo$=\sigma/\sqrt{n}\cdot z_{\alpha/2}=\sqrt{[2.5/100]}\cdot 1.96=0.3099032$

$\bar{x}-0.3099032\le\mu\le \bar{x}+0.3099032$

$\Rightarrow \bar{x}-0.3099032+0.04\le\mu\le \bar{x}+0.3099032-0.04$

$\Rightarrow \bar{x}-(0.3099032-0.04)\le\mu\le \bar{x}+(0.3099032-0.04)$

$\Rightarrow 0.3099032-0.04=\sigma/\sqrt{n}\cdot z_{\alpha/2}=\sqrt{2.5}/\sqrt{n}\cdot 1.96$

$\Rightarrow n=131.83663$

$\Rightarrow n\approx 132$

▪ Exercicio [XIV.2] *Nun sondeo sobre as preferéncias políticas dun colectivo atopou-se que a terceira parte das 60 persoas escollidas apoian ao partido A. Determinar un intervalo de confianza para a proporción p de persoas que apoian ao partido A a un nivel de confianza do 99%.*

x=“nº persoas que apoian ao partido A” $\in B(60,1/3)$

$$\omega=\frac{f-p}{\sqrt{[f(1-f)/n]}}\approx N(0,1)$$

$1-\alpha=0.99 \Leftrightarrow \alpha=0.01 \Leftrightarrow \alpha/2=0.005 \Leftrightarrow P(z>z_{\alpha/2})=P(z>z_{0.005})=P(z<z_{0.995})$
$\Rightarrow z_{\alpha/2}=2.58$

$(\hat{p}\pm Z_{\alpha/2}\sqrt{[\hat{p}(1-\hat{p})/n]})=(1/3\pm 2.58\cdot\sqrt{[1/3\cdot 2/3\cdot 1/60]})=[0.1763195,0.4903471]$

▪ Exercicio [XIV.3] *Extraiu-se unha mostra de 60 observacións dunha distribución de Poisson, obténdo-se unha média de 2.1. Construir o correspondente intervalo de confianza ao nivel do 10% para a média povoacional.*

$n=60$
$\bar{x}=2.1=\lambda_{poisson}=Var(x) \Rightarrow \bar{x}=\sqrt{2.1}$
$\alpha=0.1 \Rightarrow \alpha/2=0.05 \Rightarrow P(z>z_{0.05})=P(z<z_{0.95}) \Rightarrow z_{0.95}=1.65$

$\bar{x}-\sigma/\sqrt{n}\cdot z_{\alpha/2}\leq\mu\leq \bar{x}+\sigma/\sqrt{n}\cdot z_{\alpha/2}$
$2.1-\sqrt{2.1}/\sqrt{60}\cdot 1.65\leq\mu\leq 2.1+\sqrt{2.1}/\sqrt{60}\cdot 1.65$
$\mu \in [1.7913133,2.4086867]$

▪ Exercicio [XIV.4] *O peso de 7 paquetes de arroz son: 9.8, 10.2, 10.4, 9.8, 10.0, 10.2, 9.6. Atopar un intervalo de confianza ao 95% para a média de todos os paquetes.*

$n=7$
$1-\alpha=0.95 \Rightarrow \alpha=0.05 \Rightarrow \alpha/2=0.025 \Rightarrow t_{0.025}^{6}=2.447$
↑ [táboa p.621, t de Student]
$\bar{x}=10$
non temos a desviación típica povoacional σ
$\bar{x}-\hat{s}_{n-1}/\sqrt{n}\cdot t_{\alpha/2}^{n-1}\leq\mu\leq \bar{x}+\hat{s}_{n-1}/\sqrt{n}\cdot t_{\alpha/2}^{n-1}$ [$\hat{s}_{n-1}=\sigma_{n-1}$ na calculadora]
$10-0.2828427/\sqrt{7}\cdot 2.447\leq\mu\leq 10+0.2828427/\sqrt{7}\cdot 2.447$
$\mu \in [9.7384047,10.261595]$

Se non queremos cometer un erro superior a 10 gr., o número de paquetes de arroz necesário será aquel que implique unha reducción da amplitude do intervalo en 10 gr.

Cando n é mui elevado $t_{\alpha/2}^{n-1}$ aproxíma-se pola Normal ($z_{\alpha/2}$) supoñendo ademais que a desviación vai ser similar á cuasivarianza.

$0.2828/\sqrt{n}\cdot z_{0.025}=0.01$

▪ Exercicio [XIV.5] *Aplicou-se un test a 81 estudantes e obtívo-se unha desviación típica nas puntuacións de 12. Achar un intervalo de confianza do 95% para s.*

$n=81$
$s=12$
$1-\alpha=0.95 \Rightarrow \alpha=0.05 \Rightarrow \alpha/2=0.025 \Rightarrow X^2_{80,0.975}=106.6$
$X^2_{80,0..025}=57.2$

$\sqrt{\Sigma_{i=1}^{n}(x_i-\bar{x})^2}$ $\sqrt{\Sigma_{i=1}^{n}(x_i-\bar{x})^2}$

|————— , |—————

$\sqrt{\ } \chi^2_{n-1,1-\alpha/2}$ $\sqrt{\ } \chi^2_{n-1,\alpha/2}$

$1/n\cdot\Sigma_{i=1}^{n}(x_i-\bar{x})^2=s^2 \Rightarrow \Sigma_{i=1}^{n}(x_i-\bar{x})^2=n\cdot s^2=81\cdot 12^2=11{,}664$

$(\sqrt{[11{,}664/106.6]},\sqrt{[11{,}664/57.2]})=(10.460324,14.279919)$

▪ Exercicio [XIV.6] *Coa fin de comparar o rendemento de duas máquinas A e B tóman-se duas mostras, obténdo-se os seguintes resultados:*

máquina A (unid./dia): 20, 24, 23, 22, 22, 20, 23.

máquina B (unid./dia): 19, 22, 20, 18, 20, 22, 20, 20.

Supoñendo que o rendemento de calquera das duas máquinas segue unha distribución normal, atopar un intervalo de confianza ao 99% para a diferéncia de médias.

$1-\alpha=0.99 \Rightarrow \alpha=0.01 \Rightarrow \alpha/2=0.005$

$\bar{x}=22,\ n=7,\ \hat{s}_{n-1}^2=2.3333333$

$\bar{y}=20.125,\ m=8,\ \hat{s}_{m-1}^2=1.8392857$

$(\bar{x}-\bar{y}) \pm t_{\alpha/2}^{n+m-2-\Delta} \sqrt{[\hat{s}_{n-1}^2/n+\hat{s}_{m-1}^2/m]}$

sendo Δ o enteiro mais próximo a: $[(m-1)\hat{s}_{n-1}^2/n-(n-1)\hat{s}_{m-1}^2/m]^2 / [(m-1)(\hat{s}_{n-1}^2/n)^2+(n-1)(\hat{s}_{m-1}^2/m)^2]=$

$=[(8-1)\cdot 2.3333333/7-(7-1)\cdot 1.8392857/8]^2$

$/ [(8-1)(1.8392857/7)^2+(7-1)(1.8392857/8)^2]$

$=[2.3333333-6\cdot 1.8392857/8]^2 / [7\cdot(1.8392857/7)^2+6\cdot(1.8392857/8)^2]=$

$=0.9098661/0.8004353=1.1367141$

$\Rightarrow \Delta=1$

$n+m-2-\Delta=7+8-2-1=12$

$t_{\alpha/2}^{n+m-2-\Delta}=t_{0.005}^{12}=3.055$

$(\bar{x}-\bar{y}) \pm t_{\alpha/2}^{n+m-2-\Delta} \sqrt{[\hat{s}_{n-1}^2/n+\hat{s}_{m-1}^2/m]}=$

$=(22-20.125)\pm 3.055\cdot\sqrt{[2.3333333/7+1.8392857/8]}=[-0.4177648,4.1677649]$

▪ Exercício [XII.7] *Nun experimento binomial obsérvan-se x éxitos en n ensaios independentes. Propóñen-se as duas seguintes funcións como estimadores do parámetro p: $T_1=x/n$; $T_2=(x+1)/(n+2)$.*

a) Obter e comparar os erros meios cuadráticos para T_1 e T_2.

b) Demostrar que T_1 é un estimador consistente de p.

a)

$x\in B(n,p)\Rightarrow E(x)=np;\ Var(x)=npq$

$ECM(T)=Sesgo(T)^2+Var(T)=[E(T)-p]^2+Var(T)$

$ECM(T_1)=[E(T_1)-p]^2+Var(T_1)=$
$=[E(x/n)-p]^2+Var(x/n)=$
$=[1/n\cdot E(x)-p]^2+Var(x/n)=$
$=[1/n\cdot np-p]^2+1/n^2\cdot Var(x)=$
$=1/n^2\cdot npq=$
$=p(1-p)/n$

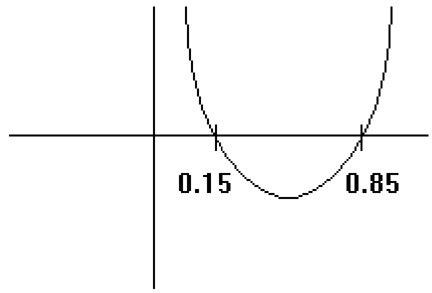
$ECM(T_2)=[E(T_2)-p]^2+Var(T_2)=$
$=[E((x+1)/(n+2))-p]^2+Var(x/(n+2)+1/(n+2))=$
$=[1/(n+2)\cdot E(x)+1/(n+2)-p]^2+1/(n+2)^2\cdot Var(x)=$
$=[(np+1)/(n+2)-p]^2+np(1-p)/(n+2)^2=$
$=[(np+1-pn-2p)/(n+2)]^2+np(1-p)/(n+2)^2=$
$=(1-2p)^2/(n+2)^2+np(1-p)/(n+2)^2=$
$=[(1-2p)^2+np(1-p)]/(n+2)^2$

$$¿\ \frac{p(1-p)}{n} < \frac{[(1-2p)^2+np(1-p)]}{(n+2)^2}\ ?$$

$\Leftrightarrow n(0.25+2p^2-2p) < p(1-p)$

Analisamos as posíbeis solucións:

- caso trivial: $0.25+2p^2-2p=0 \Leftrightarrow p=0.85$ ou 0.15
  no intervalo (0.15,0.85) T_2 é mellor que T_1



- caso contrário: depende de n: se $n\to\infty$, T_1 é mellor que T_2

b)

$$\lim_{n\to\infty} ECM(T_1)= \lim_{n\to\infty} p(1-p)/n = 0 \Rightarrow T_1$$ é un estimador consistente de p

---

- <u>Exercício</u> [XII.8] *Sexa x_1, x_2, x_3, x_4 unha mostra aleatória de tamaño 4 dunha povoación de distribución expoñencial con parámetro q descoñecido.*

  a) Das seguintes funcións, cales son estimadores insesgados de q?

  $T_1=1/6(x_1+x_2)+1/3(x_3+x_4)$

  $T_2=(x_1+2x_2+3x_3+4x_4)/5$

  $T_3=(x_1+x_2+x_3+x_4)/4$

  b) De entre os estimadores insesgados de q, determinar cal é o que ten a varianza mais pequena.

  c) Cales son as eficácias relativas dos demais estimadores insesgados con respeito ao que ten a varianza mais pequena?

a)

Distribución expoñencial:

$$P(x)=\begin{cases} e^{-\lambda x} & \text{se } x\geq 0 \\ 0 & \text{se } x<0 \end{cases} \underset{\lambda=1/\theta}{\overset{\uparrow}{=}} \begin{cases} e^{-1/\theta\cdot x} & \text{se } x\geq 0 \\ 0 & \text{se } x<0 \end{cases}$$

$$f(x)=\begin{cases} 1/\theta\cdot e^{-1/\theta\cdot x} & \text{se } x\geq 0 \\ 0 & \text{se } x<0 \end{cases}$$

$E(x)=1/\lambda=\theta$

$Var(x)=1/\lambda^2=\theta^2$

$E(T_1)=E(1/6(x_1+x_2)+1/3(x_3+x_4))=1/6E(x_1)+1/6E(x_2)+1/3E(x_2)+1/3E(x_4)=$
$=(2/6+1/3)\cdot\theta=\theta \Rightarrow T_1$ é un estimador insesgado.

$E(T_2)=E[(x_1+2x_2+3x_3+4x_4)/5]=(1/5+2/5+3/5+4/5)\cdot\theta=2\theta$
$\Rightarrow T_2$ é un estimador sesgado.

$E(T_3)=E((x_1+x_2+x_3+x_4)/4)=\theta \Rightarrow T_3$ é un estimador insesgado b)

$Var(T_1)=Var(1/6(x_1+x_2)+1/3(x_3+x_4))=1/36x_1+1/36x_2+1/9x_3+1/9x_4=15/8\cdot\theta^2$

↑

$1/6x_1$... son independentes porque o son os x_1... por separado

$Var(T_3)=Var((x_1+x_2+x_3+x_4)/4)=\theta^2/4$

$Var(T_1)>Var(T_3) \Rightarrow$ a varianza de T_3 é menor

c)

A eficiéncia (eficácia) é a inversa do ECM. Como ECM=sesgo^2+varianza e T_1, T_3 son insesgados (sesgo=0) daquela pergúntan-nos o cociente da eficácia de T_1 entre a de T_3 (menor varianza, mais eficaz):

$$\frac{Efic(T_1)}{Efic(T_3)} = \frac{1/Var(T_1)}{1/Var(T_3)} = Var(T_3)/Var(T_1) = [\theta^2/4]/[5/18\cdot\theta^2] = 18/20 = 0.9$$

$\Rightarrow$ T_1 ten o 90% de efiéncia que ten T_3