

1 - Conceptos máis importantes

1.1.- Poboación e mostra

A inferencia estatística trata de coñecer **características dunha poboación** (variable aleatoria) a través do estudo dunha mostra. **Poboación** é o **conxunto de individuos** obxecto do estudo. **Mostra** é un **subconxunto** extraído da poboación.

As características descoñecidas da poboación adoitan ser os parámetros que definen a distribución da variable na poboación, como a **media** dun dato ou a **proporción** coa que se presenta unha certa característica.

É necesario recorrer a mostras para inferir datos dunha poboación porque a poboación é moi grande, difícil de controlar, inaccesible ou se tardaría demasiado en consultar a todos os individuos. A mostra debe ser representativa dos datos.

O obxectivo da estatística inferencial é deducir o comportamento dunha poboación a partir do comportamento dunha mostra de tamaño máis manexable. Nesta unidade centrarémonos en estimar o valor da media e proporción dunha poboación a partir dunha mostra usando un estatístico.

1.2.- Estimadores ou estatísticos.

Sexa θ un parámetro da poboación (por exemplo, a media). Dicimos que $\hat{\theta}$ é un estimador seu se é un parámetro obtido a partir dunha mostra e cuxo fin é asignar un valor aproximado a θ . Hai dous tipos de estimadores:

Estimación puntual: é o resultado da estimación con un **valor numérico** (con un certo grado de confianza).

Estimación por intervalos de confianza: é o resultado da estimación con un **conxunto de valores** (con un certo grado de confianza). Estudámola na seguinte unidade.

2 - Mostra dunha poboación. Estimación de parámetros estatísticos

Sexa X a **variable aleatoria** que imos estudar **dunha poboación** a través dunha mostra (peso, ingresos, fuma ou non fuma, reacciona ao tratamento ou non, taxa de desintegracións dun átomo, etc). Ao traballar con mostras, hai que diferenciar os parámetros observados na mostra (estatísticos) dos parámetros reais correspondentes á poboación (parámetros). Os parámetros estatísticos que usaremos son os seguintes:

Parámetros	Poboación (Parámetro poboacional)	Mostra (Estatístico)	Significado	Cálculo na mostra
Media de X ou Esperanza de X	$\mu = E(X)$	$\bar{X}$	Media de todos os datos	$\bar{X} = \sum x_i \cdot p_i$
Proporción ($\approx$ porcentaxe)	p	$\hat{p}$	Razón entre individuos con certa característica e o nº total de datos	$\hat{p} = \frac{\sum x_i}{n} \quad (*)$
Desviación típica	σ	S_n ou S_{n-1}	Só traballamos con σ , que será un dato ou unha incógnita	$s_n = \sqrt{\frac{\sum x_i^2}{n} - \bar{X}^2}$ (non imos usala)

(*) onde $\sum x_i$ é o nº de veces que se observa a característica na mostra

2.1.- Estimación puntual da media

Dada unha **mostra aleatoria simple** de tamaño n procedente dunha variable aleatoria X con media poboacional μ , pódese utilizar o valor da media da mostra $\bar{X}$ para estimar a media da poboación μ . Esta estimación chámase **estimación puntual da media**, que debe acompañarse da probabilidade de que isto se cumpra.

2.2.- Estimación puntual da proporción

Sexa unha mostra aleatoria simple de tamaño n procedente dunha poboación na que a proporción de certa característica é p , pódese utilizar o valor da proporción da mostra $\hat{p}$ para estimar a proporción da poboación. Esta estimación chámase **estimación puntual da proporción**, que debe acompañarse da probabilidade de que isto se cumpra.

2.3.- Representatividade da mostra

A fiabilidade das conclusións acerca da poboación, obtidas a partir dunha mostra, dependerá da representatividade da mostra elixida. Á forma de elixir unha mostra chámase mostraxe. Unha condición necesaria para que unha mostra sexa representativa é que os seus elementos se elixan aleatoriamente. Chámase **mostraxe aleatoria**.

3 - Distribución da media na mostraxe

Teorema Central do Límite

Sexa $\bar{X}$ unha variable aleatoria medida nunha poboación (non necesariamente normal nin continua).

Sexa μ a media de X e σ a desviación típica.

Entón a **distribución da media na mostraxe**, usando mostras de tamaño n :

- Ten a mesma media que a poboación: $\mu_{\bar{X}} = E(\bar{X}) = \mu$
- A súa desviación típica é $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$ e, polo tanto, diminúe ó aumentar n .
- Cando $n > 30$, a distribución de $\bar{X}$ é practicamente normal: $\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$
- Se a variable X na poboación de partida é normal, entón a distribución da media na mostraxe tamén será normal, aínda que tomemos mostras pequenas: $X \sim N(\mu, \sigma) \Rightarrow \bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ para calquera tamaño mostral n

NOTA: sumando VA normais, se suman medias e varianzas: $X \sim N(\mu, \sigma) \Rightarrow \sum X \sim N(n \cdot \mu, \sigma \cdot \sqrt{n})$

Exemplo 1.- A duración en minutos das chamadas recibidas nun despacho profesional ten unha duración media $\mu = 3'5$ minutos e unha desviación típica $\sigma = 1'4$ minutos. Un día determinado elíxese unha mostra de 49 chamadas, cal é a probabilidade de que a duración media oscile entre 3 e 4 minutos?

Sexa $X =$ "duración dunha chamada". A distribución de X non é normal, pero o tamaño da mostra, $n = 49$, é suficientemente grande e, polo teorema central do límite, a duración media segue unha distribución normal. Logo:

$$\text{Media}(X) = 3,5; \text{Desv}(X) = 1,4 \Rightarrow \bar{X} \sim N\left(3,5; \frac{1,4}{\sqrt{49}}\right) = N(3,5; 0,2)$$

Como se pide a probabilidade de que a duración media oscile entre 3 e 4 minutos, calcularemos:

$$P(3 < \bar{X} < 4) = P\left(\frac{3-3,5}{0,2} < \frac{\bar{X}-3,5}{0,2} < \frac{4-3,5}{0,2}\right) = P(-2,5 < Z < 2,5) = 0,9938 - 0,0062 = 0,9876$$

Resposta: a probabilidade de que a duración media das chamadas oscile entre 3 e 4 minutos é de 0'9876.

1 Un banco estima que un determinado produto financeiro lle xera unhas ganancias que se distribúen segundo unha lei normal de media 20 millóns de euros e unha desviación típica de 2 millóns de euros. Selecciona unha mostra de 9 sucursais, e desexa saber a probabilidade de que a ganancia media obtida por ese produto financeiro se manteña ente 18 e 23 millóns de euros.

2 O contido en nicotina dos cigarros da marca Matados segue unha distribución de media 0,6 mg cunha desviación típica de 0,2mg. Coa fin de verificar este dato elíxese ao chou unha mostra de 50 cigarros desa marca e analízase o contido en nicotina. Cal é a probabilidade de que o contido medio en nicotina da mostra supere os 0,65mg?

3 A distribución da media na mostraxe para mostras de tamaño 100 dunha certa variable X ten unha desviación típica igual a 2. Cal é a desviación típica da variable de partida?

Hai exercicios da PAU de aplicación directa do Teorema Central do Límite marcados como [TCL] (só ata o 2019).

4 - Intervalo de confianza para a media μ dunha poboación con desviación típica σ coñecida

Un intervalo de confianza é un rango de valores entre os cales se estima que estará o valor verdadeiro dun parámetro cun determinado nivel de confianza.

O noso obxectivo será atopar un intervalo I , obtido a partir dunha mostra, tal que conteña o parámetro μ .

◦ En primeiro lugar hai que fixar a probabilidade $1 - \alpha$ coa que desexamos que o intervalo I buscado conteña a μ .

$\alpha =$ nivel de significación

$1 - \alpha =$ nivel de confianza

◦ Se esiximos moita seguridade de que a media estea no intervalo ($1 - \alpha$ GRANDE) obteremos un intervalo maior.

◦ Se esiximos menos seguridade de que a media estea no intervalo ($1 - \alpha$ pequeno) obteremos un intervalo menor

◦ Para estimar μ usaremos $\bar{X}$, sabendo que se X é $N(\mu, \sigma)$ ou $n > 30$ entón $\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$.

Posto que $\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ se tipificamos a variable $\bar{X}$, obteremos unha variable $Z \sim N(0,1)$.

4.1.- Intervalos característicos e valores críticos

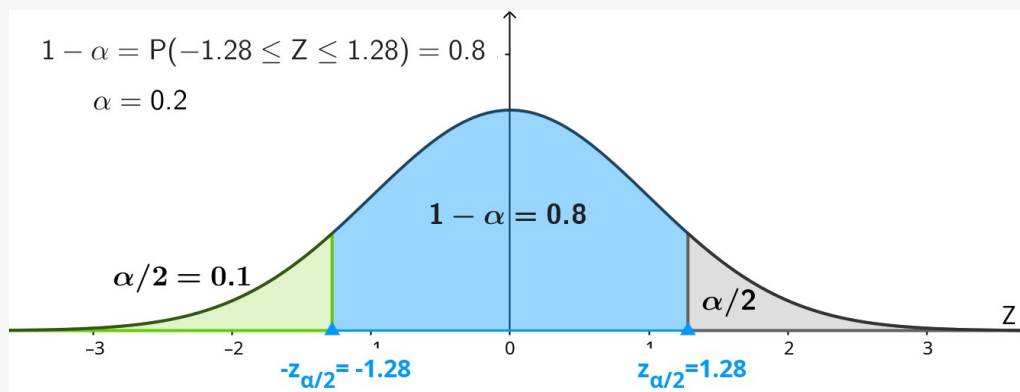
Usando a táboa da $N(0,1)$ podemos calcular o intervalo $[-z_{\alpha/2}, z_{\alpha/2}]$ tal que $P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1 - \alpha$.

Estes intervalos, que son sempre o mesmo para cada valor de $1 - \alpha$, chámanse **intervalos característicos**, podemos denotalos por I_α .

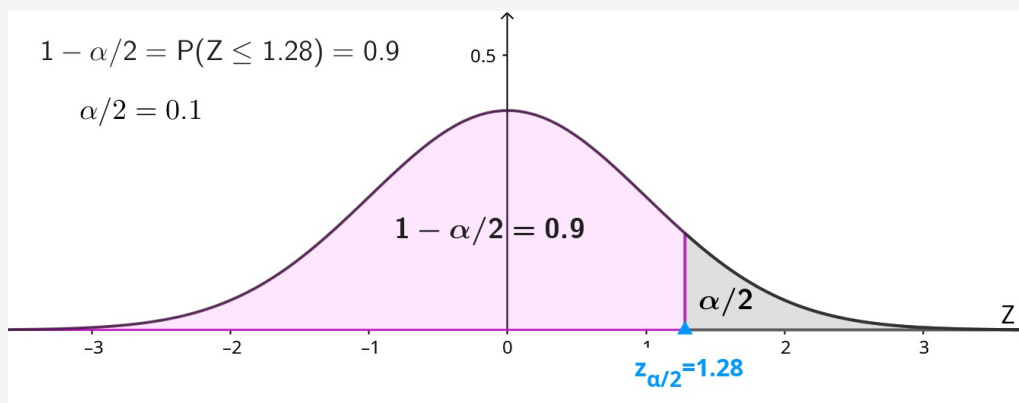
O valor $z_{\alpha/2}$, que determina os extremos do intervalo, chámase **valor crítico** do nivel de confianza $1 - \alpha$.

Exemplo 2: Busquemos o valor crítico e o intervalo característico para un nivel de confianza de $1 - \alpha = 80\%$.

Graficamente temos que localizar dous puntos, simétricos entre si que determinan unha probabilidade de 0,8. Unha vez feito o cálculo veremos que estes puntos son respectivamente $z_{\alpha/2} = 1,28$ e o seu oposto, $-z_{\alpha/2} = -1,28$



Para localizar o noso número $z_{\alpha/2} = 1,28$ temos que buscar o valor de z que deixa á esquerda unha área total igual á suma das áreas azuis e verde (neste caso 0,90) ou o que é o mesmo, a área complementaria a $\alpha/2$ (neste caso $1 - 0,1$)



Para buscar na táboa, debemos localizar o valor z_0 que se corresponde con unha probabilidade de 0,9

Neste caso non hai ningún valor que resulte na probabilidade exacta de 0,9, pois temos:

$P(Z < 1,28) = 0,8997$ que está a 0,3 milésimas

$P(Z < 1,29) = 0,9015$ que está a 1,5 milésimas

Como $z = 1,28$ é un valor moito máis cercano ao necesario, é o que tomamos como $z_{\alpha/2}$

Valor crítico: $z_{\alpha/2}$ é o valor da variable $N(0,1)$ que ten a probabilidade acumulada de $1 - \alpha/2$: $P(Z \leq z_{\alpha/2}) = 1 - \frac{\alpha}{2}$

4.2.- Niveis de confianza máis usados

Os niveis de confianza máis empregados son 90%, 95% e 99%. Completa a táboa cos valores críticos correspondentes:

Nivel de confianza $1 - \alpha$	Nivel de significación α	Probabilidade da cola $\alpha/2$	Valor crítico $z_{\alpha/2}$
$1 - \alpha = 0,90$	$\alpha = 0,10$	$\frac{\alpha}{2} = 0,05$	$z_{\alpha/2} = 1,645$
$1 - \alpha = 0,95$	$\alpha = 0,05$	$\frac{\alpha}{2} = 0,025$	$z_{\alpha/2} = 1,96$
$1 - \alpha = 0,99$	$\alpha = 0,01$	$\frac{\alpha}{2} = 0,005$	$z_{\alpha/2} = 2,575$

4.3.- Intervalo de confianza para a media

Unha vez que coñecemos o valor $z_{\alpha/2}$ xa podemos obter o intervalo que andabamos buscando.

Para estimar a media μ dunha poboación con σ coñecida recórrese a unha mostra de tamaño n da cal se obtén unha media $\bar{X}$ da mostra.

Se a poboación de partida é normal ou se $n > 30$, entón o intervalo de confianza de μ cun nivel de confianza de $1 - \alpha$ é:

$$I = (\bar{X} - E, \bar{X} + E) = \left(\bar{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \right). \text{ Resumidamente, os extremos son } \bar{X} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = \bar{X} \pm E$$

O valor $E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$ chámase erro máximo admisible, erro da estimación ou simplemente erro

Exemplo 3: Unha máquina encárgase de encher caixas de cereais. O peso destas caixas segue unha distribución normal con desviación típica de 4 g. Nunha mostra aleatoria simple de 50 caixas obtívose un peso medio de 498 g. Achar un intervalo de confianza do 95% para o peso medio de todas as caixas enchidas por esa máquina.

Solución:

- A poboación segue unha distribución $N(\mu, 4)$, e a mostra de tamaño $n=50$ ten unha media mostral $\bar{X} = 498$ g.
- Se o nivel de confianza é 0,95, entón $1 - \alpha = 0,95 \rightarrow \alpha = 0,05 \rightarrow \alpha/2 = 0,025$

○ *Xa que $P(Z < z_{\alpha/2}) = 1 - \frac{\alpha}{2} = 0,975$, consultando a táboa $N(0, 1)$ obtense $z_{\alpha/2} = 1,96$*

O intervalo para a media poboacional μ cun nivel de confianza do 95% é o seguinte:

$$I = \left(\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \right) = \left(498 - 1,96 \cdot \frac{4}{\sqrt{50}} ; 498 + 1,96 \cdot \frac{4}{\sqrt{50}} \right) = (496,89 ; 499,11)$$

4.4.- Exercicios

4 O salario, en miles de euros, dos empregados dunha multinacional, é unha variable normal de media μ e desviación típica 0,3. Tómake unha mostra aleatoria simple de 36 empregados para os que se obtén un soldo medio de 2230 €. Pídese:

a) Determina un intervalo de confianza para μ a un nivel de confianza de 0,90.

b) Acha a amplitude do intervalo de confianza se o nivel de confianza é 0,99.

5 O aforro anual (débeda en caso de valores negativos) das familias de certa localidade caracterizouse por unha distribución normal con 2000€ de desviación típica. Elíxese unha mostra aleatoria de 25 familias, sendo o aforro medio observado de 500€. Determina o intervalo de confianza ao 95% para o aforro medio anual familiar da localidade.

6 As cualificacións de certa materia seguen unha distribución normal de media descoñecida e varianza 4. Calcula, cun nivel de confianza do 98%, un intervalo para a puntuación media sabendo que nunha mostra de 64 estudantes se obtivo unha puntuación media de 5.

5 - Distribución da proporción na mostraxe

A estimación de proporcións é un caso particular do caso anterior con datos non normais. O estimador neste caso será a proporción mostral.

Por exemplo, quérese estimar a proporción de daltónicos en España. A poboación é o conxunto de habitantes de España e a característica estudada é $C = \text{"ser daltónico"}$.

5.1.- Variación da proporción ao cambiar a mostra

Para facer estimacións da proporción poboacional p usaremos o estatístico “proporción mostral” $\hat{p}$.

Supoñamos que na poboación hai unha proporción p de individuos que teñen a característica C que nos interesa. ¿Que sucederá cando seleccionemos mostras de tamaño n desta poboación?

Para cada mostra seleccionada $M_1, M_2, \dots$, podemos calcular que proporción de individuos coa característica C nesa mostra $\hat{p}_1, \hat{p}_2, \dots$. Con todos estes valores definimos unha nova variable aleatoria $\hat{p}$, que a cada mostra seleccionada da poboación asigne a proporción de individuos coa característica C : $M_i \xrightarrow{\hat{p}} \hat{p}_i$

Exemplo: tomamos 4 mostras de tamaño 1000

Mostras de tamaño 1000	X="Nº de individuos daltónicos"	Proporción de daltónicos en cada mostra
M_1	30	$\hat{p}_1 = 30/1000 = 0,030 = 3,0\%$
M_2	37	$\hat{p}_2 = 37/1000 = 0,037 = 3,7\%$
M_3	35	$\hat{p}_3 = 35/1000 = 0,035 = 3,5\%$
M_4	27	$\hat{p}_4 = 27/1000 = 0,027 = 2,7\%$

A variable aleatoria $\hat{p}$ denomínase **proporción dunha mostra** e a distribución que segue esta variable aleatoria chámase **distribución da proporción na mostraxe**.

Se sacamos unha nova mostra de 1000 individuos M_5 .

- ¿É pouco probable ou moi probable que o nº de individuos con daltonismo estea entre 0 e 10?
- ¿É pouco probable ou moi probable que o nº de individuos con daltonismo estea entre 25 e 35?
- ¿É pouco probable ou moi probable que o nº de individuos con daltonismo estea entre 20 e 40?

É dicir, estémonos preguntando polo valor de $P(0 \leq \hat{p} \leq 0,1)$, $P(2,5 \leq \hat{p} \leq 3,5)$ e $P(2 \leq \hat{p} \leq 4)$, respectivamente.

A fórmula que determina a probabilidade de cada valor posible de $\hat{p}$ chámase **distribución da proporción na mostraxe**.

5.2.- Distribución da proporción no mostreo

Partimos dunha variable X ="número de individuos coa característica C que hai nunha mostra de tamaño n".

Entón $X \sim B(n, p) \approx N(n \cdot p, \sqrt{n \cdot p \cdot q})$ entón $\hat{p} = \frac{X}{n} \sim N\left(\frac{np}{n}, \sqrt{\frac{npq}{n}}\right)$; é dicir; $\hat{p} \sim N\left(p, \sqrt{\frac{pq}{n}}\right)$.

Dada unha poboación na que p é a proporción de individuos coa característica C;

- Tomamos unha mostra de tamaño n desa poboación
- Contamos cantos individuos desa mostra teñen a característica C e calculamos $\hat{p}$, a proporción de C na mostra
- Entón $\hat{p} \sim N\left(p, \sqrt{\frac{pq}{n}}\right)$

7 Unha máquina produce parafusos. Sábese que o 5% deles son defectuosos. Empaquéitanse en caixas de 400 parafusos. Calcula:

- Cantos parafusos defectuosos cabe esperar que haxa nunha caixa.
- Que porcentaxe de caixas conterán máis de 20 parafusos defectuosos.
- Se nun palé hai 200 caixas de parafusos, cantas cabe esperar que teñan máis de 20 parafusos defectuosos?
- Que distribución segue a proporción de parafusos defectuosos nunha caixa de 400 elixida ao chou?

5.3.- Estimación por intervalos

A estimación por intervalo de confianza funciona de forma análoga a como funcionaba coa media. O nivel de confianza α ten o mesmo significado e os valores críticos $z_{\alpha/2}$ calcúlanse da mesma forma.

O intervalo de confianza $1-\alpha$ para a proporción de individuos da poboación que teñen unha certa característica C constrúese a partir da proporción dunha mostra de tamaño n e é:

$$I = (\hat{p} - E, \hat{p} + E) = \left(\hat{p} - z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}}, \hat{p} + z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}} \right)$$

8 [Setembro 2011] Unha compañía telefónica A afirma que a proporción de fogares que contratan o seu servizo de ADSL é, polo menos, do 26%. Sen embargo, outra compañía da competencia B sostén que actualmente a proporción de usuarios da compañía A é menor do 26%. Para comprobalo fai unha enquisa a 400 clientes que teñen nos seus fogares o servizo ADSL e deles 85 manifestan que teñen contratado dito servizo á compañía A.

- Utilizando a información obtida na enquisa, calcula un intervalo de confianza do 95% para a proporción de fogares que contratan actualmente o seu servizo de ADSL coa compañía telefónica A.

9 [Xuño 2013] Un estudo sobre o hábito de fumar entre os habitantes adultos dunha cidade informa que o intervalo da proporción de fumadores se estima entre un 30% e un 40%.

a) Determina a proporción mostral de fumadores observada, segundo o devandito estudo.

b) O estudo engade que os datos obtéñenos dunha enquisa aleatoria realizada a 364 habitantes adultos da cidade. Cal é entón o nivel de confianza do devandito intervalo de estimación da proporción de fumadores?

10 [Setembro 2013] Nun estudo sobre hixiene dental, a porcentaxe de nenos que presentaron indicios de carie utilizando un dentífrico tradicional foi de, polo menos, o 10%. Nun grupo de 500 nenos elixidos aleatoriamente que utilizaron un novo dentífrico, presentaron indicios de carie 35 deles.

a) Calcula un intervalo do 95% de confianza para a proporción de nenos con indicios de carie utilizando o novo dentífrico.

5.4.- Nivel de confianza, erro admisible e tamaño da mostra

O valor $E = z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}}$ é o **erro máximo admisible** para a estimación da proporción p .

O erro máximo admisible para a estimación da proporción ten o grave inconveniente de que depende de $\hat{p}$, valor obtido nesa mostra concreta.

Cando o que pretendemos é calcular o tamaño de mostra n necesario para manter o erro dentro de certo rango aínda non se extraeu a mostra, así que non temos un valor da proporción mostral $\hat{p}$ para poder usar no cálculo do erro.

O que se fai na práctica nestas circunstancias é poñernos no peor dos casos posible, que non é un caso tan malo: podemos ver o erro E como unha función que depende de p (e tamén do nivel de confianza e do tamaño da mostra) desta

forma: $E_{\alpha,n}(p) = z_{\alpha/2} \cdot \sqrt{\frac{pq}{n}} \approx z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$. Esta función faise máxima cando $p = q = \frac{1}{2}$

O que se fai na práctica cando non coñecemos p nin tampouco temos un valor de $\hat{p}$ porque aínda non sacamos a mostra, é poñerse no caso máis desfavorable; é dicir, tomar $p = q = \frac{1}{2}$.

O tamaño de mostra que nos asegure que o erro se vai manter menor có noso límite cando $p = 1/2$ tamén nos manterá o erro por debaixo do noso límite para calquera outro valor de p .

11 [Setembro 2014] A proporción de mulleres dunha poboación portadoras de hemofilia é descoñecida. Para estimala elíxese unha mostra aleatoria de 500 mulleres entre as que se encontran 80 portadoras da enfermidade.

a) Calcula un intervalo do 95% de confianza para a proporción de mulleres portadoras de hemofilia desa poboación

b) Supoñendo que aínda non se tomou a mostra e queremos facer a estimación cometendo un erro non superior ao 2%, cun 95% de confianza, de que tamaño debería ser a devandita mostra?

Estimación de parámetros estadísticos		
Media A media da variable X na poboación é μ A media da variable X na mostra é $\bar{X}$ $\bar{X} = \sum x_i \cdot p_i$	Proporción A proporción da caract. C na poboación é p A proporción da caract. C na mostra é $\hat{p}$ $\hat{p} = 1/n \sum x_i$	Desviación típica A DT de X na poboación é σ A DT de X na mostra é S_{n-1} NON SE ESTUDA EN BACH
Teorema Central do Límite		
Distribución da media na mostraxe		
X é VA de calquera tipo medida nunha poboación. O tamaño da mostra é n. Entón: $\mu_{\bar{X}} = E(\bar{X}) = \mu$; $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$		
Se $n > 30$, $\bar{X}$ é aproximadamente unha VA normal: $n > 30 \Rightarrow \bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$	Se X é normal, entón $\bar{X}$ tamén é VA normal $X \sim N(\mu, \sigma) \Rightarrow \bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$	Se $X = \text{"Éxitos nunha mostra n"}$ entón $\hat{p} = \frac{X}{n}$ é VA normal Se $X \sim B(n, p)$ entón $\hat{p} \sim N\left(p, \sqrt{\frac{pq}{n}}\right)$
Intervalos de confianza		
$1 - \alpha$ é o nivel de confianza α é o nivel de significación $\alpha/2$ é a probabilidade da cola	Valor crítico $z_{\alpha/2}$ é o valor $N(0,1)$ que ten unha probabilidade acumulada de $1 - \alpha/2$: $P(z \leq z_{\alpha/2}) = 1 - \alpha/2$	Intervalos característicos $I_{\alpha} = [-z_{\alpha/2}, z_{\alpha/2}]$ $P(-z_{\alpha/2} \leq z \leq z_{\alpha/2}) = 1 - \alpha = P(z \in I_{\alpha})$
Intervalo de confianza para a media $I = (\bar{X} - E, \bar{X} + E) =$ $= \left(\bar{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \right)$ sendo $E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$ o erro característico	Intervalo de confianza para a proporción $I = (\hat{p} - E, \hat{p} + E) = \left(\hat{p} - z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}}, \hat{p} + z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}} \right)$ sendo $E = z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}}$ o erro característico e sendo $E = z_{\alpha/2} \cdot \sqrt{\frac{0,5 \cdot (1 - 0,5)}{n}}$ o erro se $\hat{p}$ é descoñecida	

Problemas de ABAU de Galicia [2024-2017]

12 [xuño 2025 P4.2] — Unha enquisa realizada a 100 individuos dunha poboación revela que 80 deles están satisfeitos co servizo da súa compañía eléctrica.

a) 4.2.1. Calcule un intervalo cun 95% de confianza para a proporción de individuos satisfeitos co servizo da súa compañía eléctrica.

b) [TCL] 4.2.2. Se se sabe que 8 de cada 10 individuos están satisfeitos co servizo da súa compañía eléctrica e se toma unha mostra de 100 individuos, cal é a probabilidade de que a proporción de individuos satisfeitos co servizo da súa compañía eléctrica sexa superior ao 87%?

a) $I_{95\%} = (0,72; 0,88)_{95\%}$

b) A probabilidade de que a proporción de individuos satisfeitos sexa superior ao 87% é 0,0401

13 [xuño 2025 P1] — CONTEXTO Unha das principais novidades das probas PAU 2025 é que o exame de cada materia debe incluír un exercicio obrigatorio e de carácter “máis competencial”. Aínda que as notas se publican a semana seguinte de realizarse o exame, os membros do grupo de traballo da materia Matemáticas Aplicadas ás Ciencias Sociais II están interesados en determinar canto antes se se produciron cambios relevantes na nota media da materia que coordinan con relación ás notas de cursos pasados.

Con este obxectivo contactaron previamente cun grupo de correctores, dos que cada un deles se comprometeu a corrixir un máximo de 25 exames o primeiro día. Polos datos doutros cursos, as notas desta materia pode supoñerse que seguen unha distribución normal con desviación típica igual a 1,5.

Responda estes tres apartados: 1.1., 1.2. e 1.3.

a) 1.1. Se se quere estimar esta nota media cun erro máximo de 0,25, empregando un nivel de confianza do 95%, cal é o número mínimo de correctores que se necesitan?

b) 1.2. Unha vez corrixidos os 100 primeiros exames, a nota media resultou ser igual a 7,2. A partir desta mostra, calcule un intervalo de confianza con nivel de confianza do 95% da nota media.

1.3. Unha vez corrixidos todos os exames, elíxense 25 ao azar, cal é a probabilidade de que a nota media destes 25 exames sexa superior a 7 se sabemos que $\mu=7,3$?

14 [xuño 2024 P6] — Pode supoñerse que o tempo de formación, en horas, que precisa un empregado dunha empresa para poder traballar nunha nova planta segue unha distribución normal con desviación típica igual a 15.

a) Se nunha mostra de 25 empregados, o tempo medio precisado foi de 97 horas, calcule un intervalo de confianza cun 95% de confianza para a media do tempo de formación precisado.

b) [TCL] Se a media do tempo de formación precisado é $\mu=97$ horas, cal é a probabilidade de que o tempo medio precisado de mostras de 36 traballadores se atope entre 90 e 104 horas?

$$a) I_{95\%} = (91,12 ; 102,88)_{95\%} \quad b) P(90 < \bar{X} < 104) = 0,9948$$

15 [xullo 2024 P6] — A lonxitude (en centímetros) dos listóns de madeira que se producen nunha industria distribúese normalmente cunha desviación típica de $\sigma = 6$ centímetros.

a) Calcule un intervalo do 98% de confianza para a lonxitude media dos listóns tendo en conta que nun lote de 9 listóns se observou unha lonxitude media de 244 centímetros.

b) [TCL] Se a lonxitude media dos listóns producidos é de $\mu = 244$ centímetros, cal é a probabilidade de que a lonxitude media dos listóns dun lote de $n = 16$ listóns sexa inferior a 242 centímetros?

$$a) I_{98\%} = (239,34 ; 248,66)_{98\%} \quad b) P(\bar{X} < 242) = 0,0918$$

16 [xuño 2023 P6] — Nunha mostra aleatoria de 120 empresas inspeccionadas, de entre as visitadas un ano polos inspectores de traballo dunha provincia, sancionouse a 30 delas.

a) Calcule, con un nivel de confianza do 90%, un intervalo de confianza para a **proporción** de empresas sancionadas pola Inspección de Traballo.

b) [TCL] Se ignoramos os datos iniciais e cun nivel de confianza do 95%, cal é o tamaño mínimo da mostra necesaria para estimar a proporción de empresas sancionadas cun erro máximo do 2%?

$$a) I = (0,25 - 0,065 ; 0,25 + 0,065) = (0,185 ; 0,315) \quad b) \text{O tamaño mínimo da mostra para estimar dita proporción cun erro máximo do 2\% e un nivel de confianza do 95\% debe ser de 2401 empresas.}$$

17 [xullo 2023 P6] — O salario (en €) dos traballadores dunha empresa distribúese normalmente con desviación típica $\sigma=300$ €. Preguntouse a 36 traballadores elixidos ó chou, e estableceuse que o salario medio dos traballadores da empresa oscila entre 1552 € e 1748 €.

a) Cal foi o salario medio dos traballadores da mostra? Con que nivel de confianza se estableceu o intervalo anterior?

b) [TCL] Se o salario medio dos traballadores da empresa é $\mu=1650$ €, cal é a **probabilidade de que o salario medio** de mostras de 36 traballadores sexa superior a 1590 €?

- a) O salario medio dos traballadores foi de 1650 €. O nivel de confianza co que se estableceu o intervalo e do 95%
 b) A probabilidade de que o salario medio de mostras de 36 traballadores sexa superior a 1590€ é 0,8849

18 [xuño 2022 P6] — Tomamos unha mostra aleatoria de 36 facturas de consumo mensual de luz (en euros) e o intervalo de confianza obtido ao 95% para o consumo mensual medio é [60.1, 69.9]. Segundo esta información:

a) Cal foi o consumo medio mostral de luz? b) Cal e o erro máximo cometido?

b) Determine un intervalo de confianza ao 90% para o consumo medio de luz.

- a) 65 € consumo medio mostral de luz. b) 4, 9 € erro máximo cometido. c) Intervalo de confianza $(65 \pm 1,645 \cdot 15/\sqrt{36})$. Ao 90% de confianza, o consumo medio de luz estará comprendido entre 60,8875 e 69,4125 €

19 [xullo 2022 P6] — Sábese que a idade dos traballadores nas fábricas dunha zona segue unha distribución normal de desviación típica 10 anos. Cunha mostra de traballadores da zona o intervalo de confianza ao 90% para a media de idade obtido é (39.25, 44.75)

a) Cal foi o tamaño da mostra utilizada?

b) Canto vale a media da mostra?

c) Cal sería o erro cometido a un nivel de confianza do 95%?

- a) O tamaño de mostra utilizada foi de 36 traballadores. b) A idade media da mostra e de 42 anos. c) O erro cometido sería de 3,26 anos

20 xuño 2021 P6 — Unha compañía de seguros quere determinar que proporción dos seus clientes estaría disposta a aceptar unha subida de tarifas a cambio dun incremento nas súas prestacións. Unha enquisa previa indica que esta proporción está en torno ao 15%.

a) De que tamaño mínimo debería ser a mostra se se quere estimar dita proporción cun erro inferior a 0,08 e un nivel de confianza do 95%?

Finalmente, realízase o estudo cunha mostra de 196 clientes, dos que 37 manifestaron a súa conformidade coa proposta.

b) Calcule un intervalo de confianza, ao 92%, para a proporción de clientes da compañía que aceptaría dita proposta. Cal e o erro máximo cometido?

- a) O tamaño mínimo da mostra para estimar dita proporción cun erro inferior a 0,08 e un nivel de confianza do 95% debe ser de 77 clientes.
 b) o I.C ao 92%, para a proporción de clientes que aceptaría a proposta (0,141;0,239). O erro máximo cometido $e=0,049 \cong 5\%$

21 xullo 2021 P6 — O peso das laranxas para zume colleitadas por un produtor é unha variable aleatoria que se distribúe normalmente cunha media de $\mu = 200$ gramos e unha desviación típica $\sigma = 50$

a) [TCL] Se tomamos unha mostra aleatoria unha mostra aleatoria de $n= 25$ laranxas, ¿cal é a probabilidade de que o seu peso medio estea comprendido entre 175 e 215 gramos?

b) [TCL] De que tamaño se tomou outra mostra aleatoria se a probabilidade de que o peso medio sexa inferior a 210 gramos é do 97.72%?

- a) O peso medio da mostra de 25 laranxa estará entre 175g e 215g con unha probabilidade 0,9270. b) A mostra tomada e de $n=100$ laranxas.

22 xuño 2020 P6 — A produción diaria de leite, medida en litros, dunha granxa pódese aproximar por unha variable normal de media μ descoñecida e desviación típica $\sigma=50$ litros.

a) Determine o tamaño mínimo de mostra para que o correspondente intervalo de confianza para μ ao 95% teña unha amplitude como máximo de 8 litros.

b) [TCL] Tómanse os datos de produción de 25 días, calcule a probabilidade de que a media das producións obtidas sexa menor ou igual a 930 litros se sabemos que $\mu=950$ litros.

a) *Tamaño mínimo da mostra 601 días.* b) *A Probabilidade pedida, que a media das producións obtidas sexa menor ou igual a 930 litros, e 0,0228*

23 xuño 2020 P6 — Unha editorial desexa coñecer o impacto que terá a publicación dunha nova obra dun recoñecido novelista. Tras entrevistar a 100 persoas afeccionadas á lectura, 80 delas recoñecen que adquirirán esa nova obra.

a) ¿Con que nivel de confianza se pode afirmar que a proporción de afeccionados á lectura que adquirirán a obra está entre o 69,7% e o 90,3%?

b) [TCL] Se se sabe que 8 de cada 10 persoas afeccionadas á lectura adquirirán a obra e eliximos unha mostra de $n = 144$ desas persoas, calcule a probabilidade de que a proporción de afeccionados á lectura que adquirirán a obra sexa superior ó 75%.

a) *Pódese afirmar cun nivel de confianza do 99%* b) *A probabilidade de que a proporción de afeccionados a lectura que adquirirán a obra sexa superior o 75%, para mostras de $n=144$ persoas, e 0,9525*

24 xuño 2020 Modelo A4 – Nun grupo de 400 estudantes de bacharelato escollidos ao azar, 320 deles manifestan acudir a clases particulares.

a) Calcula un **intervalo do 95% de confianza para a proporción** de estudantes de bacharelato que acoden a clases particulares. Interpreta o resultado.

b) Calcula o nivel de confianza co que se tería que construír o intervalo para que o erro cometido sea inferior ao 4,34%.

25 xuño 2020 Modelo B4 – O tempo de ensamblaxe dun xoguete segue unha **distribución normal** con unha desviación típica de $\sigma = 15$ minutos.

a) Se se afirma que o **tempo medio** de ensamblaxe deses xoguetes **está entre 69,06 min e 74,94 min a partir dunha mostra dada de 100** xoguetes, ¿cal é o valor da media mostral? ¿cal é o nivel de confianza co que se fai esa afirmación?

b) [TCL] Se se sabe que o tempo medio de ensamblaxe é de $\mu = 70$ minutos, ¿cal é a probabilidade de que o tempo medio de ensamblaxe sexa inferior a 73 minutos, en mostras de 64 xoguetes?

26 xuño 2019 A4 - Un estudo electoral cunha mostra de 400 electores obtén un **intervalo para a proporción** de votantes dun partido de $[0.23, 0.31]$.

a) Canto vale a proporción mostral?

b) Cal é o nivel de confianza co que se estableceu o intervalo?

c) Cal é o erro máximo cometido co intervalo dado?

a) $\hat{p}=0,27$ b) *Nivel de confianza $1-\alpha=0,9282 \rightarrow 92,82\%$* c) $e=0,04 \rightarrow e=4\%$

27 xuño 2019 B4 – Logo de anos de utilizalo sábese que a puntuación dun test de uso habitual en certa rama industrial segue unha **distribución normal de media 74 e desviación típica 16**. Nunha empresa decídese realizalo a 100 dos seus empregados.

a) [TCL] Cal é a probabilidade de que se obteña unha media mostral superior a 78 puntos, de seguirse a pauta xeral?

b) [TCL] E a probabilidade de que a media mostral sexa inferior a 74 puntos?

a) $P(\bar{X}>78) = 1 - 0,9938 = 0,0062$ b) $P(\bar{X}<74) = 0,5$

28 xullo 2019 A4 – Tomouse unha mostra aleatoria de 100 mozos e medíuselles o nivel de glicosa en sangue obténdose unha **media mostral** de 105 mg/cm³. Sábese que a desviación típica na poboación é de 15mg/cm³.

a) Obtén un intervalo de confianza, ao 95%, para o nivel medio da glicosa en sangue na poboación.

b) Canto vale o erro máximo no intervalo anterior?

c) Que ocorre co a amplitude do intervalo se o nivel de confianza é do 99%?

a) *Intervalo de confianza para o nivel medio de glicosa (102,06 ; 107,94) a un n.c do 95%* b) $e = 107,94 - 105 = 2,94$ mg/cm³ c) *O intervalo (101,14; 108,86) a un n.c do 99% é mais amplo que con un n.c do 95%*

29 xullo 2019 B4

a) Nunha mostra aleatoria de estudantes $n = 25$ de bacharelato, o 75% afirman querer realizar estudos universitarios. Calcula un **intervalo de confianza para a proporción** de estudantes de bacharelato que queren realizar estudos universitarios cun nivel de confianza do 90%.

b) [TCL] Se se sabe que 8 de cada 10 estudantes de bacharelato afirman querer realizar estudos universitarios e tomamos unha mostra aleatoria de $n = 100$ estudantes, cal é a probabilidade de que a proporción de estudantes da mostra que queren realizar estudos universitarios sexa superior ao 65%?

a) *A proporción de estudantes que manifestan querer realizar estudos universitarios estará entre (0,6075 , 0,8925) a un nivel de confianza do 90%* b) $P(\hat{p} > 0,65) = 0,9999$

30 xuño 2018 A3 - O peso (en gramos) das empanadas que saen dun forno segue unha distribución normal cunha desviación típica de 120 gramos. Se se estableceu o intervalo (1499,9; 1539,1) como **intervalo de confianza para a media** a partir dunha mostra de 144 empanadas

a) cal é o valor da media mostral?, con que nivel de confianza se construíu o intervalo?

b) Cantas empanadas, como mínimo, deberíamos pesar para que o nivel de confianza do intervalo anterior sexa do 99%?

a) $\bar{X} = 1519,5$ g media mostral; $\alpha = 0,05 \rightarrow$ n.c 95% b) *Deberíamos pesar ao menos 249 empanadas*

31 xuño 2018 B4 - Un consumidor cre que o peso medio dun produto é distinto do que indica o envase. Para estudar este feito, o consumidor toma unha mostra aleatoria simple de 100 produtos nos que se observou un peso medio de 245 g. Suponse ademais que o peso do produto por envase segue unha distribución normal con desviación típica 9 g.

a) Constrúe un intervalo de confianza para o peso medio dese produto ao 95 % de confianza.

b) Cal sería o tamaño mostral mínimo necesario para estimar o verdadeiro peso medio a partir da media mostral cun erro de estimación máximo de 2 g e un nivel de confianza do 90 %?

a) *O intervalo pedido é (243,236; 246,764)_{95%}* b) *Necesitaríase un tamaño de mostra de ao menos 55 produtos.*

32 setembro 2018 A4 - Nun estanque deséxase estimar a porcentaxe de peixes dourados. Para iso, tómase unha mostra aleatoria de 700 peixes e atópase que exactamente 70 deles son dourados.

a) Acha, cun nivel de confianza do 99 %, un **intervalo para estimar a proporción** de peixes dourados no estanque

b) No intervalo anterior, canto vale o erro de estimación?

c) Considerando dita mostra, que lle ocorrería ao erro de estimación se aumentase o nivel de confianza? Xustifica a resposta.

a) *A un nivel de confianza do 99% a proporción de peixes dourados estará entre 7,07% e 12,92%*
b) $e = 0,0292 \rightarrow e = 2,92\%$ c) *Calculando de novo o valor do erro, e aumenta.*

33 setembro 2018 B4 - Nunha empresa quérese racionalizar o gasto en teléfono móbil dos seus axentes comerciais. Para iso faise un estudo sobre unha mostra dos devanditos axentes e obtense: "cunha confianza do 95%, a **media** do

gasto mensual en teléfono móvil **está entre 199,71 e 220,29 euros**". Supoñendo que o gasto en teléfono móvil é unha variable normal

a) Calcula o gasto medio mostral e o erro cometido na estimación.

b) Se a desviación típica é de 42 euros, que tamaño ten a mostra?

a) $\hat{\mu}=210\text{ €}$ gasto medio; $e= 10,29\text{ €}$ b) $n = \text{tamaño mostra} = 64$

34 xuño 2017, opción A - Unha empresa informática lanzou ao mercado un produto do que sabe que a súa vida útil, en anos, segue unha distribución normal de media μ e desviación típica $\sigma = 1,6$ anos.

a) Para unha mostra aleatoria de 100 produtos, a vida media útil foi de 4,6 anos. Calcula un intervalo do 95% de confianza para estimar a vida media útil do produto. Interpreta o intervalo obtido.

b) Supoñamos que a vida útil do produto segue unha distribución $N(4,6 ; 1,6)$ e tómasse unha mostra aleatoria de 64 produtos. Calcula a probabilidade de que a vida media útil da mostra estea entre 4,25 e 4,95 anos.

a) Estímase a vida media útil do produto informático entre, aproximadamente, 4'29 anos e 4'91 anos, cun 95% de confianza. b) $P(4,25 < \bar{X} < 4,95) = 0'9198 \approx 0'92$

35 xuño 2017, opción B - Como resultado dunha enquisa na que se utilizou o suposto de máxima indeterminación ($p = 1 - p = 1/2$) afirmase que, cun 97,56% de confianza, a **porcentaxe de individuos dunha poboación** que considera o alcol e/ou as drogas como causa principal dos accidentes de tráfico, está entre o 57,5% e o 62,5%.

a) Calcula o número de individuos desa poboación aos que se lles realizou a enquisa.

b) Dos que se lles realizou a enquisa, ¿cantos contestaron que a causa principal dos accidentes é o alcol e/ou as drogas?

a) A enquisa realizóuselles a 2025 individuos desa poboación. b) Contestaron que o alcol e/ou as drogas son a causa principal dos accidentes de tráfico 1215 individuos, da mostra de 2025 enquisados

36 setembro 2017, opción A -

a) Nunha mostra aleatoria de 200 clientes dun centro comercial, 150 efectúan as súas compras utilizando a tarxeta propia do centro. Calcula un **intervalo do 95% de confianza para a proporción** de clientes que efectúan as compras utilizando a tarxeta propia do centro. Interpreta o intervalo obtido.

b) Se se sabe que 8 de cada 10 clientes do centro comercial utilizan para as súas compras a tarxeta propia do centro e tomamos unha mostra aleatoria de 100 clientes, ¿cal é a probabilidade de que a proporción de clientes da mostra que utilizan a tarxeta propia do centro sexa superior a 0,75?

a) $I=(0,69;0,81)$. Cun 95% de confianza, nese centro comercial entre o 69% e o 81% (aproximadamente) dos seus clientes utilizan para as súas compras a tarxeta propia do centro. b) $P(\hat{P} > 0'75) = 0'8944$

37 setembro 2017, opción B - O tempo de formación, en horas, que necesita un empregado dunha empresa para poder traballar nunha nova planta segue unha distribución $N(\mu, \sigma= 15)$.

a) Elixida unha mostra de 36 empregados da empresa, **obtense o intervalo de confianza (321,1 ; 330,9) para a media μ** . Calcula o tempo medio de formación dos empregados da mostra e o nivel de confianza co que se construíu o intervalo.

b) Supoñamos que o tempo de formación, en horas, que necesita un empregado desa empresa para poder traballar nunha nova planta segue unha distribución $N(\mu=326, \sigma= 15)$. Calcula a probabilidade de que o tempo medio de formación non supere as 330 horas, en mostras de 36 empregados.

a) $\bar{X} = 326$; $1 - \alpha = 0'95$, sendo o intervalo dun 95% de confianza b) $P(\bar{X} \leq 330) = 0'9452$