10 CÁLCULO DE PROBABILIDADES

O inicio do cálculo de probabilidades é atribuído a **Pascal** e **Fermat** (século XVII), que ó enfrontarse con certos problemas propostos por un xogador profesional, sentaron as primeiras bases desta teoría.

Laplace, a comezos do século XIX, realizou importantes investigacións que recompilou no seu libro *Teoría analítica das probabilidades.* Entre outras moitas cousas estudiou o teorema "sobre a probabilidade das causas", enunciado por **Bayes** no século XVIII.

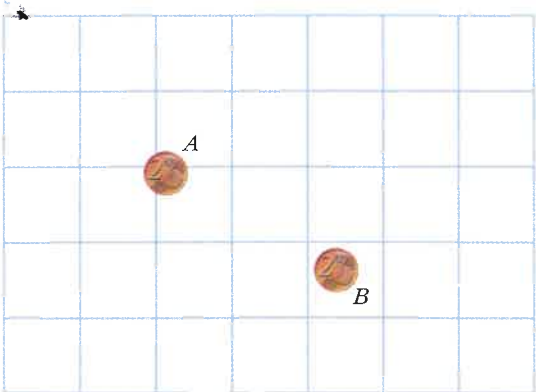
No século XX, coa intervención da teoría de conxuntos, o cálculo de probabilidades axiomatízase, co que gaña en rigor e xeneralidade.

PARA EMPEZAR, REFLEXIONA E RESOLVE

Cálculo de probabilidades

Cuadriculamos un folio con cadrados de 3 cm × 3 cm.

Se deixamos caer sobre el unha moeda de 2 céntimos de euro pode caer "tocando raia", como A, ou "sen tocar raia", como B.



Imos calcular a probabilidade do suceso S:

S = "A MOEDA CAE SEN TOCAR RAIA"

1º MÉTODO. OBTENCIÓN DA PROBABILIDADE MEDIANTE EXPERIMENTACIÓN

Realízase a experiencia moitas veces cun certo tipo de moeda e calcúlase a frecuencia relativa do suceso S.

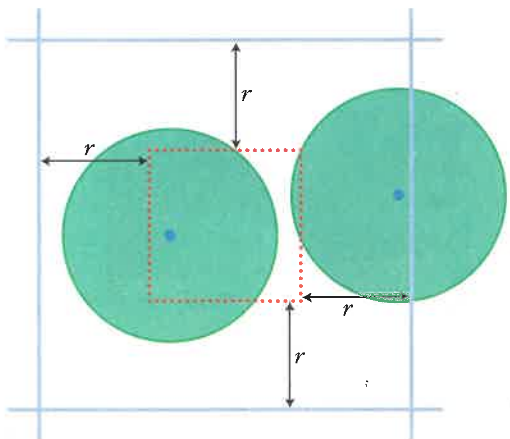
Supoñamos que 30 persoas (os alumnos dunha clase) realízano 100 veces cada unha (un total de 3000 experiencias), e que se contabiliza que o suceso S aconteceu 511 veces. A frecuencia relativa será:

$$fr\,(S) = \frac{511}{3\,000} = 0{,}17$$

Inducimos así que $P[S] \approx 0{,}17$. Ou o que é igual, que o suceso S acontecerá, por termo medio, unha de cada seis veces (1 : 6 ≈ 0,17).

2º MÉTODO. CÁLCULO MATEMÁTICO DA PROBABILIDADE

A posición da moeda queda determinada polo seu centro.



¿En que puntos da cuadrícula debe quedar o centro da moeda para que esta non toque raia? É claro que debe estar no interior do cadrado pequeno, é dicir, a unha distancia da raia que sexa superior ó raio da moeda.

Supoñamos que o diámetro da moeda é de 18 mm.

Área do cadrado grande: $3^2 = 9\ \text{cm}^2$

Área do cadrado pequeno: $(3 \angle 1{,}8)^2 = 1{,}44\ \text{cm}^2$

$$P[S] = \frac{1{,}44}{9} = 0{,}16$$

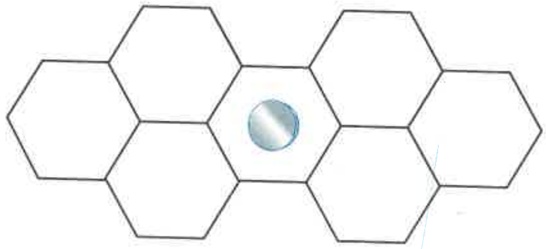
A probabilidade obtida por este método (0,16) foi sensiblemente igual á obtida polo método experimental (0,17). Pode pensarse que era este mesmo tipo de moedas de 18 mm de diámetro co que experimentaban os alumnos.

- Calcula matematicamente cál é a probabilidade de que a moeda "non toque raia" na cuadrícula de 3 cm × 3 cm unha moeda de 1 cm de diámetro.

- ¿De que tamaño debe ser un disco para que a probabilidade de que "non toque raia" nunha cuadrícula de 4 cm × 4 cm sexa de 0,2?

- Nunha cuadrícula de 4 cm × 4 cm deixamos caer 5 000 veces unha moeda e contabilizamos que "non toca raia" en 1 341. Estima cál é o diámetro da moeda.

- Sobre un piso de baldosas hexagonais regulares de 12 cm de lado déixase caer un disco de 10 cm de diámetro.



¿Cal é a probabilidade de que "non toque raia"?

NESTA UNIDADE VERÁS

- Lembraremos en qué consisten as **experiencias aleatorias,** os **sucesos** e as operacións entre estes para, con eles, estudiar as propiedades da **probabilidade.**

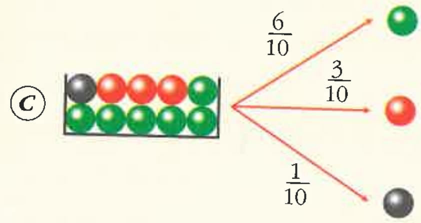
- Estudiaremos a relación entre a *frecuencia* relativa dun suceso e a súa probabilidade (**Lei dos grandes números**).

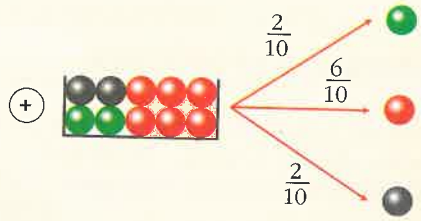
- Practicaremos coa **regra de Laplace** como aplicación para o cálculo de probabilidades de sucesos pertencentes a experiencias realizadas con instrumentos regulares.

- Aprenderemos a identificar e a calcular **probabilidades condicionadas,** tanto en experiencias compostas como en táboas de continxencia.

{1, 2}	12	3	5	20
{3, 4, 5, 6}	12	20	8	40
	24	23	13	60

- Utilizaremos o diagrama en árbore para describir **probabilidades compostas** e sobre el calcularemos **probabilidades totais** e **probabilidades "a posteriori"** (as "probabilidades das causas", que dicía Bayes).





10.1 EXPERIENCIAS ALEATORIAS. SUCESOS

Se lanzamos unha pedra e coñecemos as condicións iniciais de altura, velocidade, etc., sabemos con seguridade ónde caerá, cánto tempo tardará, etc. É unha *experiencia determinista.*

Se botamos un dado sobre unha mesa, ignoramos qué cara quedará arriba. O resultado depende do azar. É unha *experiencia aleatoria.*

SUCESOS ALEATORIOS

O número de coches que pasarán por un tramo de estrada certo día, ¿é un suceso aleatorio? Poderiamos preguntarlle, previamente, a cada un dos posibles viaxeiros e realizar con certa exactitude a previsión. ¿Imaxinas a inmensa dificultade da consulta? Sen embargo, os expertos, tratando como aleatorio o suceso e recorrendo a experiencias anteriores, prevén con notable aproximación o que vai acontecer.

Experiencia aleatoria é aquela na que o resultado depende do azar.

Suceso aleatorio é un acontecemento que ocorrerá ou non dependendo do azar.

A vida cotiá está chea de sucesos aleatorios. Moitos deles, de tipo sociolóxico (viaxes, accidentes, número de persoas que acudirán a un grande almacén...), aínda que son suma de moitas decisións individuais, poden ser estudiados, moi vantaxosamente, como aleatorios.

Para comezar, fixámonos en experiencias aleatorias sinxelas como lanzar dados, extraer cartas dunha baralla, sacar bólas de urnas...

Espacio mostral

Chámaselle **espacio mostral** ó conxunto de todos os posibles resultados dunha experiencia aleatoria. En adiante designarémolo por E.

Por exemplo: Nun dado, $E = \{1, 2, 3, 4, 5, 6\}$

Nunha moeda, $E = \{C, +\}$

Sucesos

Chámaselle **suceso** a calquera subconxunto de E.

Os elementos de E chámanse **sucesos individuais** ou **sucesos elementais.** Tamén se chaman **casos**.

Tamén son sucesos o suceso baleiro ou **suceso imposible**, $\varnothing$, e o propio E, **suceso seguro**.

Ó conxunto de todos os sucesos dunha experiencia aleatoria chamarémoslle $\mathcal{S}$.

Se E ten un número finito, n, de elementos, o número de sucesos de E é 2^n.

Por exemplo: Nun dado, $\{1, 2\}, \{2, 4, 6\}, \{3, 5\}$ son sucesos.

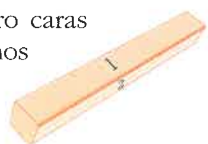
$\{1\}, \{2\}, \{3\}, ...$ son sucesos individuais.

Hai $2^6 = 64$ sucesos.

Nunha moeda hai $2^2 = 4$ sucesos: $\mathcal{S} = \{\varnothing, \{C\}, \{+\}, \{C, +\}\}$

EXERCICIOS PROPOSTOS

1. Numeramos con 1, 2, 3 e 4 as catro caras alongadas dunha regreta. Deixamos caer a regreta e anotamos o número da cara superior.



a) ¿Cal é o espacio mostral?

b) Escribe un suceso elemental e tres non elementais.

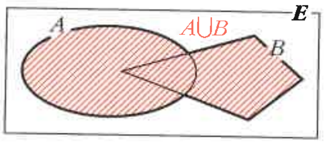
c) ¿Cantos sucesos ten esta experiencia?

Dise que un suceso *se verifica*, ou que *ocorre*, cando ó realizar a experiencia aleatoria correspondente, o resultado é un dos elementos (casos) dese suceso.

Por exemplo, se ó tirar un dado sae un 2, verificáronse entre outros, os sucesos $\{2\}$, $\{2, 4, 6\}$, $\{2, 5\}$, E.

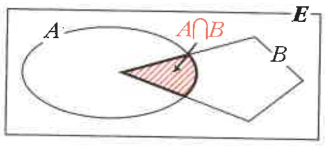
Operacións con sucesos

Dados dous sucesos, A e B, chámase:



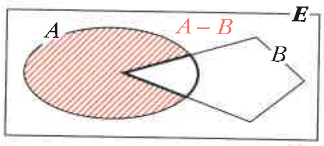
UNIÓN. $A \cup B$ (lese *A unión B*) é o suceso formado por todos os elementos de A e de B.

O suceso $A \cup B$ verifícase cando ocorre un dos dous, A ou B, ou ambos.



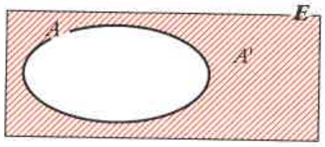
INTERSECCIÓN. $A \cap B$ (lese *A intersección B*) é o suceso formado por todos os elementos que son, á vez, de A e de B.

O suceso $A \cap B$ verifícase cando ocorren simultaneamente A e B.



DIFERENCIA. $A - B$ (lese *A menos B*) é o suceso formado por todos os elementos de A que non son de B.

$A - B$ verifícase cando o fai A e non B.



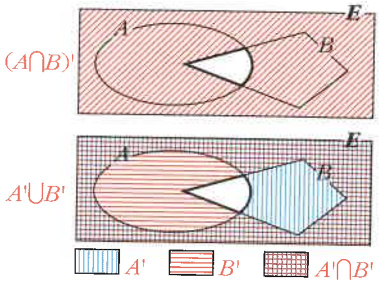
COMPLEMENTARIO. O suceso $A' = E - A$ chámase **suceso contrario** ou **complementario de** A.

O suceso A' verifícase sempre e cando non se verifique A.

SUCESOS INCOMPATIBLES. Dous sucesos A e B, chámanse **incompatibles** cando non teñen ningún elemento común. É dicir, cando $A \cap B = \varnothing$.

Os sucesos incompatibles non se poden verificar simultaneamente.

$(A \cap B)' = A' \cup B'$. XUSTIFICACIÓN



Propiedades das operacións con sucesos

As operacións anteriores cumpren, entre outras, as seguintes propiedades:

Distributivas

$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

De simplificación

$A \cup (B \cap A) = A$

$A \cap (B \cup A) = A$

Co complementario

$(A')' = A$ $\quad$ $A - B = A \cap B'$

$(A \cup B)' = A' \cap B'$ $\quad$ $(A \cap B)' = A' \cup B'$

EXERCICIOS PROPOSTOS

2. Xustifica graficamente a seguinte igualdade:

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

3. Xustifica graficamente a seguinte igualdade:

$$A - B = A \cap B'$$

10.2 FRECUENCIA E PROBABILIDADE

Frecuencia absoluta e frecuencia relativa dun suceso

Realizamos N veces unha experiencia aleatoria. Recordemos as seguintes definicións.

Chámaselle **frecuencia absoluta** dun suceso S ou, simplemente, frecuencia de S, ó *número de veces que ocorre* S. Desígnase por $f(S)$. Chámaselle **frecuencia relativa** de S á *proporción de veces que ocorre* S.

$$fr(S) = \frac{f(S)}{N}$$

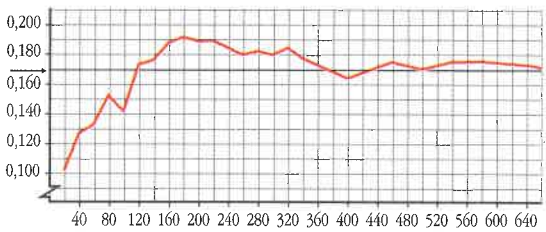
Lei dos grandes números

Ó realizar reiteradamente unha experiencia aleatoria, a frecuencia relativa dun certo suceso, $fr(S)$, vai tomando distintos valores. Estes valores ó principio sofren grandes oscilacións pero, pouco a pouco, vanse estabilizando (oscilan cada vez menos). Cando N crece moito, aproxímanse a un certo valor que é a probabilidade de S, $P[S]$.

$$\lim_{N \to \infty} fr(S) = P[S] \qquad \text{LEI DOS GRANDES NÚMEROS}$$

Por exemplo, ó lanzar un dado obtéñense os valores de $fr(3)$ para N = 20, 40, 60, 80, 100, 120, ...

Os valores de $fr(3)$ vanse estabilizando ó redor de $1/6 = 0{,}166$.



¿Por que a frecuencia relativa tende a estabilizarse?

Ó lanzar reiteradamente un dado, ¿é posible obter o suceso "3" cinco veces consecutivas? Por suposto que si; é difícil, pero posible. ¿Como lle afecta esta sucesión de tiradas á $fr(3)$? Pois depende de cántas experiencias se leven. Por exemplo, supoñamos que antes desa serie consecutiva era $fr(3) = 0{,}16$:

Para	$N = 100$	$f(3) = 16$	$fr(3) = 0{,}16$	Coa serie de 5 veces "3", $fr(3)$ aumenta 4 centésimas
	$N = 105$	$f(3) = 21$	$fr(3) = 0{,}20$	
Para	$N = 1\,000$	$f(3) = 160$	$fr(3) = 0{,}16$	$fr(3)$ aumenta 4 milésimas
	$N = 1\,005$	$f(3) = 165$	$fr(3) = 0{,}164$	
Para	$N = 10\,000$	$f(3) = 1\,600$	$fr(3) = 0{,}16$	$fr(3)$ aumenta 4 dezmilésimas
	$N = 10\,005$	$f(3) = 1\,605$	$fr(3) = 0{,}1\,604$	

É dicir, ***fr*(3) estabilízase porque, ó ser o denominador cada vez máis grande, ó cociente aféctanlle cada vez menos as fluctuacións do numerador.**

Propiedades das probabilidades

As probabilidades dos distintos sucesos dunha mesma experiencia aleatoria deben cumprir as propiedades que damos a continuación. As tres primeiras impóñense (é dicir, son **axiomas**), inspirándonos nas propiedades das frecuencias relativas. As seguintes dedúcense das primeiras (é dicir, son **teoremas**).

RAZOA

É claro que as frecuencias relativas cumpren as seguintes propiedades:

- $fr(S) \geq 0$, calquera que sexa S.
- Se $A \cap B = \varnothing$,
  entón $f(A \cup B) = f(A) + f(B)$
  e, por tanto,
  $fr(A \cup B) = fr(A) + fr(B)$
- $fr(E) = 1$, pois $f(E) = N$

AXIOMAS

A probabilidade de cada suceso é un número. Hanse de cumprir os seguintes axiomas:

Ax.1 Calquera que sexa o suceso S, $P[S] \geq 0$.

Ax.2 Se dous sucesos son incompatibles, a probabilidade da súa unión é igual á suma das súas probabilidades.

$$A \cap B = \varnothing \Rightarrow P[A \cup B] = P[A] + P[B]$$

Ax.3 A probabilidade total é 1: $P[E] = 1$

En esencia, estas tres propiedades indican que a cantidade total de probabilidades é igual á 1 e repártese aditivamente entre os distintos sucesos.

DEMOSTRACIÓN DOS TEOREMAS

Estes teoremas están encadeados. Cada un deles pode demostrarse a partir dos axiomas e dos teoremas anteriores. Abaixo vén a demostración do **T.6.** As demostracións dos outros son máis sinxelas.

TEOREMAS

T.1 $P[A'] = 1 - P[A]$

T.2 $P[\varnothing] = 0$

T.3 Se $A \subset B$, daquela $P[B] = P[A] + P[B - A]$

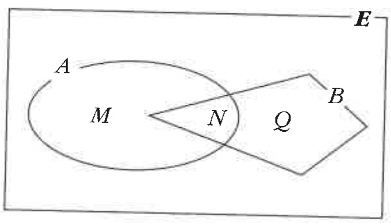
T.4 Se $A \subset B$, daquela $P[A] \leq P[B]$

T.5 Se $A_1, A_2, \ldots, A_k$, son incompatibles dous a dous, daquela:
$P[A_1 \cup A_2 \ldots \cup A_k] = P[A_1] + P[A_2] + \ldots + P[A_k]$

T.6 $P[A \cup B] = P[A] + P[B] - P[A \cap B]$

T.7 Se o espacio mostral E é infinito e un suceso é $S = \{x_1, x_2, \ldots, x_k\}$, daquela:

$$P[S] = P[x_1] + P[x_2] + \ldots + P[x_k]$$



Demostración del T.6

Descompoñemos dos sucesos A e B en sucesos disxuntos (ou incompatibles):

$A \cap B = N$, $A = M \cup N$, $B = N \cup Q$

Como $B = N \cup Q$ con N e Q incompatibles:

$P[B] = P[N] + P[Q] \Rightarrow P[Q] = P[B] - P[N]$

$A \cup B = A \cup Q$, A e Q incompatibles.

Por tanto:

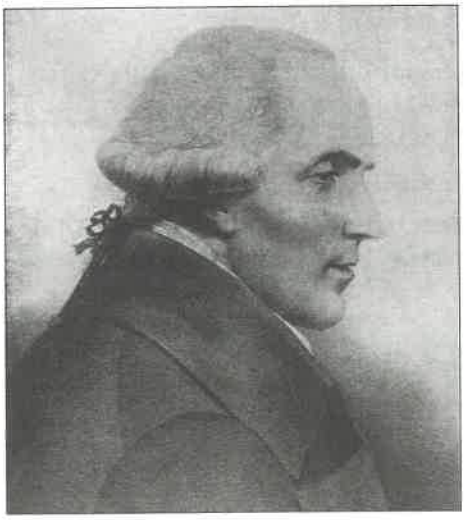
$$P[A \cup B] = P[A] + P[Q] = P[A] + (P[B] - P[N]) =$$
$$= P[A] + P[B] - P[A \cap B]$$

10.3 LEI DE LAPLACE

A propiedade **T.7** da páxina anterior permite calcular a probabilidade dun suceso S coñecendo as probabilidades dos sucesos elementais que o compoñen.

Pero se o espacio mostral consta de n sucesos elementais equiprobables (todos eles coa mesma probabilidade $1/n$), entón a probabilidade de S só depende do número de sucesos elementais que o compoñen:

$$P[S] = \underbrace{\frac{1}{n} + \frac{1}{n} + \ldots + \frac{1}{n}}_{\text{tantos sumandos como elementos ten } S}$$



Pierre Simón Laplace (1749-1827)

Lei de Laplace

Se $E = \{x_1, x_2, \ldots, x_n\}$ y $P[x_1] = P[x_2] = \ldots = P[x_n]$, entón:

$$P[S] = \frac{\text{número de elementos de } S}{n}, \text{ é dicir,}$$

$$P[S] = \frac{\text{número de "casos favorables" a } S}{\text{número de "casos posibles"}}$$

Dise que un instrumento aleatorio é **de Laplace** cando a probabilidade de todos os seus sucesos elementais (casos) é a mesma. Por exemplo, un dado correcto, unha moeda, unha baralla...

EXERCICIOS RESOLTOS

1. *Nunha baralla de 40 cartas, calcula:*

a) $P[\text{AS}]$

b) $P[\text{OUROS}]$

$$P[\text{AS}] = \frac{\text{número de ASES}}{\text{n}^{\text{o}} \text{ total de cartas}} = \frac{4}{40} = 0{,}1$$

$$P[\text{OUROS}] = \frac{\text{número de OUROS}}{\text{n}^{\text{o}} \text{ total de cartas}} = \frac{10}{40} = 0{,}25$$

2. *Nunha baralla suprimimos varias cartas. Entre as que quedan, danse as seguintes probabilidades de ser extraídas:* $P[\text{REI}] = 0{,}15$; $P[\text{BASTOS}] = 0{,}3$; $P[\text{carta que non sexa REI nin BASTOS}] = 0{,}6$.

a) ¿Está entre elas o REI de BASTOS? En caso afirmativo, dá a súa probabilidade.

b) ¿Cantas cartas hai?

a) O suceso "NIN REI NIN BASTOS" é o complementario de "REI OU BASTOS", pois $(R \cup B)' = R' \cap B'$.

$P[\text{NIN REI NIN BASTOS}] = 0{,}6 \Rightarrow P[\text{REI OU BASTOS}] = P[\text{REI} \cup \text{BASTOS}] = 0{,}4$

$P[\text{REI} \cup \text{BASTOS}] = P[\text{REI}] + P[\text{BASTOS}] - P[\text{REI} \cap \text{BASTOS}]$

Substituíndo:

$0{,}4 = 0{,}15 + 0{,}3 - P[\text{REI} \cap \text{BASTOS}] \Rightarrow P[\text{REI} \cap \text{BASTOS}] = 0{,}05$

Por tanto, o REI de BASTOS está e a súa probabilidade é:

$$P[\text{REI de BASTOS}] = 0{,}05 = \frac{1}{20}$$

b) Unha porción de cartas dunha baralla é un instrumento aleatorio "de Laplace", pois a probabilidade de extraer cada unha delas é a mesma. Se neste montón a probabilidade do rei de bastos é 1/20, é porque hai 20 cartas.

Cándo non se pode aplicar a lei de laplace

A lei de Laplace pódese aplicar cando todos os sucesos elementais (elementos do espacio mostral) teñen a mesma probabilidade. Pero hai moitos casos en que isto non ocorre. Vexamos dúas situacións tipo.

INSTRUMENTOS IRREGULARES

Hai instrumentos aleatorios claramente irregulares: un dado mal construído, unha chincheta, unha chuca. ¿Como avaliar as probabilidades dos sucesos nestes casos?

Se o instrumento é irregular, para avaliar a probabilidade dun suceso, S, recorremos á *lei dos grandes números*. Efectuaremos un número, N, grande de experiencias, obteremos a frecuencia, $f(S)$, e así, a frecuencia relativa $fr(S) = f(S)/N$ será unha medida aproximada da probabilidade. Canto maior sexa N, máis confianza teremos na estimación.

Por exemplo, se deixamos caer 100 chinchetas do mesmo tipo 10 veces, obteremos 1 000 resultados.

Supoñamos que $f($$) = 243$, $f($$) = 757$.

As frecuencias relativas sérvennos para estimar as probabilidades:

$P[$$] \approx fr($$) = 0{,}243$ $\quad P[$$] \approx fr($$) = 0{,}757$

Se quixeramos ter máis seguridade nas estimacións, acumulariamos máis experiencias: 2 000, 5 000, 10 000, …

INSTRUMENTOS REGULARES, SUCESOS ELEMENTAIS NON EQUIPROBABLES

Lanzamos dous dados correctos e sumamos os seus resultados. O espacio mostral desta experiencia é {2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12}. Pero a experiencia ou o sentido común fannos ver que non é igualmente probable que a suma sexa 2 ou 7 ou 10. Sen embargo podemos modificar a descrición da experiencia de modo que os sucesos elementais sexan equiprobables, tal como facemos no cadro.

2º DADO \ 1º DADO						
+	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Deste modo, o espacio mostral consta de 36 sucesos elementais (1, 1), (1, 2), …, (3, 5), …, (6, 6), todos eles coa mesma probabilidade: 1/36. Os resultados "SUMA 2", "SUMA 7", "SUMA 10", que antes eran considerados sucesos elementais, na nova descrición son sucesos non elementais. As súas probabilidades obtéñense sinxelamente:

$P[\text{SUMA } 2] = 1/36$ $\quad P[\text{SUMA } 7] = 6/36 = 1/6$ $\quad P[\text{SUMA } 10] = 3/36 = 1/12$

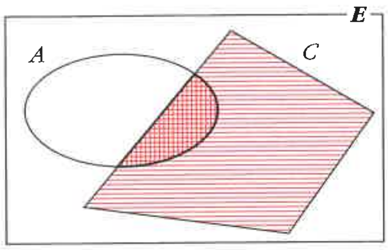
EXERCICIOS PROPOSTOS

1. Lanzamos un dado "chapuceiro" mil veces. Obtermos $f(1) = 117$, $f(2) = 302$, $f(3) = 38$, $f(4) = 234$, $f(5) = 196$, $f(6) = 113$. Estima as probabilidades das distintas caras. ¿Cales son as probabilidades dos sucesos PAR, MENOR CA 6, {1, 2}?

2. ¿Cal é a probabilidade de obter 12 ó multiplicar os resultados de dous dados correctos?

3. ¿Cal é a probabilidade de que ó lanzar dous dados correctos a diferencia das súas puntuacións sexa 3?

10.4 PROBABILIDADE CONDICIONADA. SUCESOS INDEPENDENTES



Dados dous sucesos, A e C, chámaselle **probabilidade de A condicionada a C** e escríbese $P[A/C]$ a:

$$P[A/C] = \frac{P[A \cap C]}{P[C]}$$

Mide a proporción de veces que ocorre A de entre as que ocorre C.

Da expresión anterior dedúcese que:

$$P[A \cap C] = P[C] \cdot P[A/C]$$

SUCESOS INDEPENDENTES

Se A e C son independentes:

$P[A/C] = P[A]$

$P[A \cap C] = P[C] \cdot P[A/C] =$

$= P[C] \cdot P[A]$

Dous sucesos, A e C, dise que son **independentes** cando:

$$P[A/C] = P[A] \quad \text{e} \quad P[C/A] = P[C]$$

Cando dous sucesos son independentes a probabilidade da súa intersección é igual ó producto das súas probabilidades:

$$A \text{ e } C \text{ independentes} \Rightarrow P[A \cap C] = P[A] \cdot P[C]$$

Por exemplo, se nunha bolsa temos as seguintes bólas

téñense as seguintes probabilidades:

$$P[\text{PAR/VERDE}] = \frac{P[2]}{P[\text{VERDE}]} = \frac{1}{3}$$ (Hai unha bóla par entre as tres verdes).

$$P[\text{PAR/VERMELLO}] = \frac{P[4, 6]}{P[\text{VERMELLO}]} = \frac{2}{4} = \frac{1}{2}$$ (Hai dous pares entre as catro vermellas).

$$P[\text{PAR/NEGRO}] = \frac{P[8]}{P[\text{NEGRO}]} = \frac{1}{1} = 1$$ (A única bóla negra é par).

ACLARACIÓN PARA O EXEMPLO

Sabendo que a bóla é verde, a probabilidade de que sexa par é distinta de 1/2. Por iso os sucesos VERDE e PAR non son independentes.

Sabendo que a bóla é roxa, a probabilidade de que sexa par segue sendo 1/2. Por iso os sucesos ROXA e PAR son independentes.

Os sucesos PAR e VERMELLO son independentes, pois:

$$P[\text{PAR/VERMELLO}] = P[\text{PAR}] = \frac{1}{2}$$

É dicir, a proporción de PARES entre as bólas VERMELLAS é igual á proporción de PARES no conxunto total. Por tanto, o dato "a bóla é vermella", non modifica a expectativa de obter "par". Por iso os sucesos son independentes.

Outro exemplo: Nos seguros de coches ás MULLERES cóbranlles menos cós HOMES. É debido a que P[ACCIDENTE/MULLER] < P[ACCIDENTE]. É dicir, os sucesos TER ACCIDENTE e SER MULLER non son independentes. Sen embargo, non lles cobran nin máis nin menos ós que teñan Rh+ ou sexan LOUROS. É lóxico, pois o suceso TER ACCIDENTE é independente tanto do suce-

Estudio de probabilidades en táboas de continxencia

Seguíuselle a pista a 100 000 coches durante un ano, de tres marcas distintas: SETA, VOVO e ADI. Uns tiveron algún accidente serio (AC) e outros non (NON AC).

	SETA	VOVO	ADI
AC	400	200	400
NO AC	49 600	19 800	29 600

Repártense como ves na primeira táboa da esquerda.

Unha táboa deste tipo chámase **táboa de continxencia**. O colectivo total (coches) repártese segundo dúas características (marcas e se tiveron ou non accidente). O estudio dos datos permite pescudar as relacións entre ambas as características.

	SETA	VOVO	ADI	
AC	400	200	400	1 000
NO AC	49 600	19 800	29 600	99 000
	50 000	20 000	30 000	100 000

Para analizar a táboa de continxencia debemos comezar completándoa coas súas parciais e coa suma total (segunda táboa).

As sumas parciais permítennos ver que, por exemplo, hai 20 000 coches da marca VOVO. Ou que hai 1 000 coches que tiveron un accidente.

Se entre os mil coches que tiveron accidente, AC, escollemos un ó chou, é fácil atopar as probabilidades condicionadas.

$P[\text{SETA}/\text{AC}] = \dfrac{400}{1\,000} = 0{,}4$ (Sabendo que o coche tivo accidente, ¿cal é a probabilidade de que sexa SETA?

Análogamente, $P[\text{VOVO}/\text{AC}] = \dfrac{200}{1\,000} = 0{,}2$, $P[\text{ADI}/\text{AC}] = \dfrac{400}{1\,000} = 0{,}4$

Tamén se pode calcular a probabilidade de que un coche dunha certa marca teña accidente:

$P[\text{AC}/\text{SETA}] = \dfrac{400}{50\,000} = 0{,}008$, $P[\text{AC}/\text{VOVO}] = \dfrac{200}{20\,000} = 0{,}01$

$P[\text{AC}/\text{ADI}] = \dfrac{400}{30\,000} = 0{,}0133$

Estes resultados fan ver que a marca máis segura é SETA e a menos, ADI.

Comparando estas últimas probabilidades condicionadas coa probabilidade total, $P[\text{AC}] = \dfrac{1\,000}{100\,000} = 0{,}01$, advertimos que os sucesos VOVO e AC son independentes mentres que SETA e ADI non son independentes con AC.

EXERCICIOS PROPOSTOS

1. Observa as bólas que hai na urna.

a) Completa o cadro de dobre entrada no que se reparten as bólas segundo a cor (V, R, N) e o número (1, 2).

	V	R	N	
1		2		
2		3		
		5		

b) Calcula a probabilidade de RUBIO, NEGRO, VERDE, 1 e 2, sen máis que observar a composición da urna.

c) Comproba que as probabilidades obtidas en b) poden obterse sumando filas ou columnas do cadro formado en a).

d) Calcula as probabilidades condicionadas: $P[1/\text{RUBIO}]$, $P[1/\text{VERDE}]$, $P[1/\text{NEGRO}]$, $P[2/\text{RUBIO}]$, $P[2/\text{VERDE}]$, $P[2/\text{NEGRO}]$

e) Di se algún dos caracteres RUBIO, NEGRO, VERDE é independente de 1 ou de 2.

10.5 PROBAS COMPOSTAS

Hai experiencias nas que facilmente podemos distinguir dúas ou máis etapas. Chámanse **probas compostas**. Nelas o cálculo de probabilidades de sucesos compostos simplifícase moito calculando as probabilidades dos seus compoñentes.

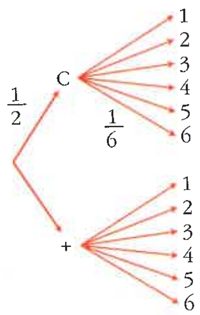
Dúas probas compostas son **independentes** cando o resultado dunha non inflúe na outra. Se non é así, chámanse **independentes**.

- Exemplo 1. Lanzamos unha moeda e un dado. É evidente que o resultado dunha non inflúe na outra. Son independentes.
- Exemplo 2. Extraemos dúas cartas dunha baralla. O resultado da segunda si depende da primeira (por exemplo, se a primeira é AS, é menos probable que a segunda o sexa). Son dependentes.

Experiencias independentes

Analicemos o exemplo 1 de arriba. Neste caso, o cálculo de probabilidades é especialmente sinxelo se supoñemos que a moeda e o dado son de *Laplace*.

	1	2	3	4	5	6	
C	C,1	C,2	C,3	C,4	C,5	C,6	1/2
+	+,1	+,2	+,3	+,4	+,5	+,6	1/2
	1/6	1/6	1/6	1/6	1/6	1/6	



$$P[\text{C y } 3] = \frac{1}{2} \cdot \frac{1}{6} = \frac{1}{12}$$

$$P[(\text{C}, 3)] = P[\text{C}] \cdot P[3] = \frac{1}{2} \cdot \frac{1}{6} = \frac{1}{12}$$

A probabilidade do **suceso composto** é igual ó producto das probabilidades dos sucesos compoñentes. Por exemplo:

$$P[(\text{C e PAR})] = P[\text{C}] \cdot P[\text{PAR}] = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

Dise que **dous ou máis probas son independentes** cando o resultado de cada unha delas non inflúe nas probabilidades dos distintos resultados das outras. Polo tanto, os sucesos correspondentes á primeira son independentes dos sucesos correspondentes á segunda.

Se n probas son independentes e os sucesos $S_1, S_2, \ldots, S_n$ corresponden respectivamente, a cada unha delas correspóndelle que:

$$P[S_1 \text{ na } 1^{\text{a}} \text{ e } S_2 \text{ na } 2^{\text{a}} \text{ e} \ldots S_n \text{ na } n\text{-ésima}] =$$
$$= P[S_1] \cdot P[S_2] \cdot \ldots \cdot P[S_n]$$

TEN EN CONTA

Hai ocasións nas que as probas non son sucesivas senón simultáneas. Pero, se se pode, resulta moi vantaxoso pensar nelas como se se sucedesen no tempo.

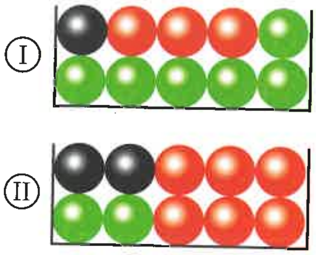
EXERCICIOS PROPOSTOS

1. Calcula a probabilidade de obter tres CATROS ó lanzar tres dados.

2. Calcula a probabilidade de NINGÚN SEIS ó lanzar catro dados (catro veces NON SEIS).

3. Calcula a probabilidade de obter ALGÚN SEIS ó lanzar catro dados. (ALGÚN SEIS é o suceso contrario de NINGÚN SEIS).

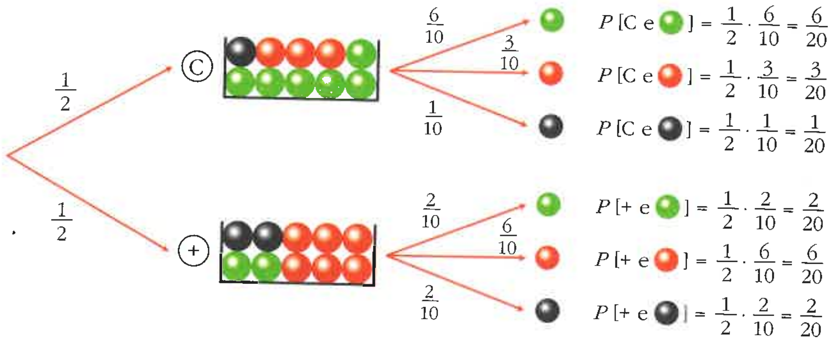
4. Calcula a probabilidade de obter ALGÚN SEIS ó lanzar seis dados.



Experiencias dependentes

Temos unha moeda e dúas urnas con bólas. Lanzamos a moeda e se sae cara, extraemos unha bóla de I e se sae cruz, extraemos unha bóla de II.

Describimos, mediante un diagrama en árbore, a experiencia composta e incluímos as correspondentes probabilidades.



Observemos que as probabilidades das segundas frechas son probabilidades condicionadas.

Por exemplo, $P[\ /C] = \dfrac{6}{10}$

Por tanto: $P[C \text{ e } \ ] = P[C] \cdot P[\ /C] = \dfrac{1}{2} \cdot \dfrac{6}{10} = \dfrac{6}{20}$

Dúas experiencias son **dependentes** cando o resultado da primeira inflúe nas probabilidades dos sucesos da segunda. As probabilidades de sucesos compostos obténense así:

$P[S_1 \text{ na } 1^{a} \text{ e } S_2 \text{ na } 2^{a}] = P[S_1]\, P[S_2/S_1]$. É dicir:

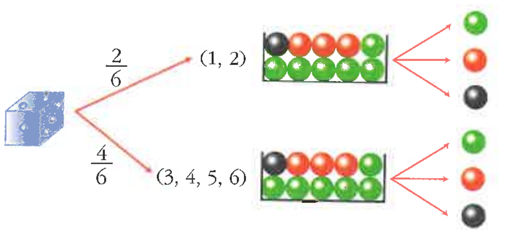
$P[S_1 \text{ na } 1^{a}] \cdot P[S_2 \text{ na } 2^{a} \text{ suposto que ocorreu } S_1 \text{ na } 1^{a}]$

Se se encadean máis de dúas experiencias dependentes, as probabilidades dos sucesos compostos obténense analogamente. Por exemplo, para tres probas:

$$P[S_1 \text{ na } 1^{a} \text{ e } S_2 \text{ na } 2^{a} \text{ e } S_3 \text{ na } 3^{a}] =$$
$$= P[S_1] \cdot P[S_2/S_1] \cdot P[S_3/S_1 \text{ e } S_2]$$

EXERCICIOS PROPOSTOS

5. Temos un dado e as dúas urnas descritas abaixo.

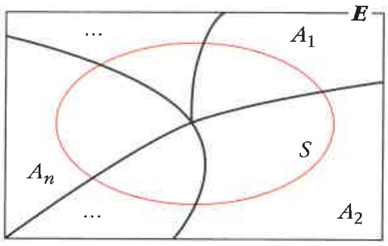


Lanzamos o dado. Se sae 1 ou 2, imos á urna I. Se sae 3, 4, 5 ou 6, acudimos á urna II. Extraemos unha bóla da urna correspondente.

a) Completa as probabilidades no diagrama en árbore.

b) Calcula: $P[\{3, 4, 5, 6\} \text{ e } \ ]$, $P[\ /1]$, $P[\ /5]$ e $P[2 \text{ e } \ ]$.

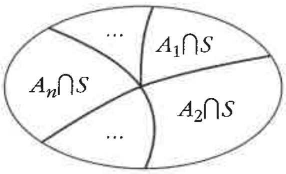
10.6 PROBABILIDADE TOTAL



Temos n sucesos, $A_1, A_2, ..., A_n$, incompatibles dous a dous e tales que $A_1 \cup A_2 \cup ... \cup A_n = E$. Entón, para calquera suceso S cúmprese que:

$$P[S] = P[A_1] \cdot P[S/A_1] + P[A_2] \cdot P[S/A_2] + ... + P[A_n] \cdot P[S/A_n]$$

A probabilidade $P[S]$ descomposta deste modo chámase **probabilidade total**.



Demostración

Descompoñemos S en sucesos incompatibles:

$$S = (A_1 \cap S) \cup (A_2 \cap S) \cup ... \cup (A_n \cap S)$$

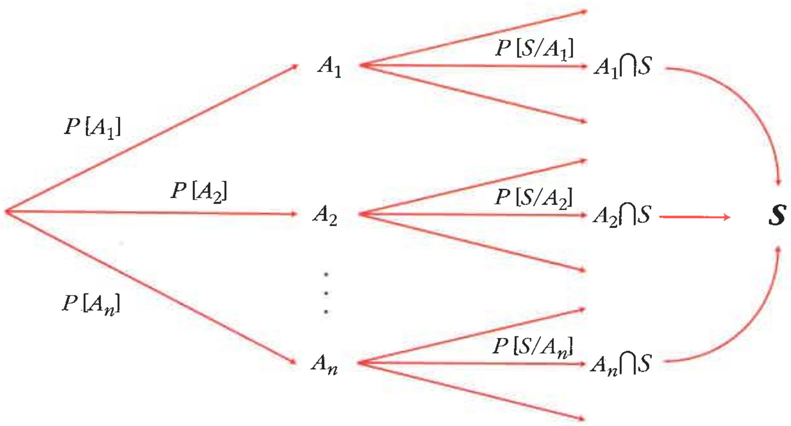
Aplícamos a propiedade **T.5** das probabilidades:

$$P[S] = P[A_1 \cap S] + P[A_2 \cap S] + ... + P[A_n \cap S]$$

Tendo en conta que $P[A_k \cap S] = P[A_k] \cdot P[S/A_k]$, obtense a expresión buscada.

A probabilidade total no caso de probas sucesivas

O resultado anterior é especialmente útil para experiencias compostas que se suceden no tempo. Imos ilustralo no caso de dúas etapas sucesivas:



Os sucesos $A_1, A_2, ..., A_n$ son da primeira etapa e incompatibles dúas a dúas. S é un suceso da segunda etapa. Pode chegarse a S pasando por $A_1, A_2, ...$ ou A_n.

Observando o diagrama en árbore apréciase o significado que teñen, neste caso, as probabilidades seguintes:

$$P[S/A_i],\ P[A_i \cap S] = P[A_i] \cdot P[S/A_i]$$

así como a $P[S]$ como probabilidade total.

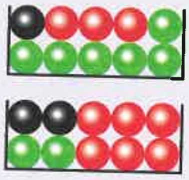
Tamén é claro o interese que ten este proceso para estes casos, pois só se pode chegar ó suceso S pasando por algún dos sucesos $A_1, A_2, ..., A_n$.

EXERCICIOS RESOLTOS

1. *Lanzamos un dado. Se sae 1 ou 2 extraemos unha bóla da urna I, se sae 3, 4, 5 ou 6 extraemos bóla da urna II.*

I

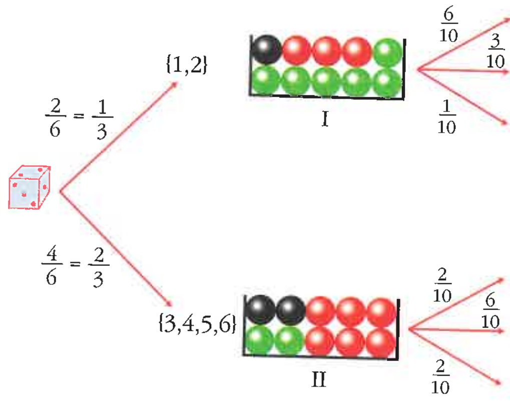
II



a) Calcula P[🔴] (a probabilidade de que a bóla sexa vermella).

b) P[🟢]

c) P[⚫]



🟢 $P[\{1,2\}$ e 🟢$] = \frac{1}{3}\cdot\frac{6}{10}=\frac{6}{30}$

🔴 $P[\{1,2\}$ e 🔴$] = \frac{1}{3}\cdot\frac{3}{10}=\frac{3}{30}$

⚫ $P[\{1,2\}$ e ⚫$] = \frac{1}{3}\cdot\frac{1}{10}=\frac{1}{30}$

🟢 $P[\{3,4,5,6\}$ e 🟢$] = \frac{2}{3}\cdot\frac{2}{10}=\frac{4}{30}$

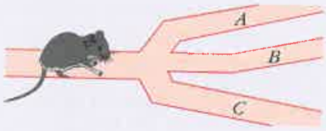
🔴 $P[\{3,4,5,6\}$ e 🔴$] = \frac{2}{3}\cdot\frac{6}{10}=\frac{12}{30}$

⚫ $P[\{3,4,5,6\}$ e ⚫$] = \frac{2}{3}\cdot\frac{2}{10}=\frac{4}{30}$

a) $P[$🔴$] = \frac{3}{30}+\frac{12}{30}=\frac{15}{30}=\frac{1}{2}$ b) $P[$🟢$] = \frac{6}{30}+\frac{4}{30}=\frac{10}{30}=\frac{1}{3}$

c) $P[$⚫$] = \frac{1}{30}+\frac{4}{30}=\frac{5}{30}=\frac{1}{6}$

2. *Un gato persegue a un rato. Este pode entrar nunha das canellas A, B ou C.*



En cada unha delas pode cazalo, +, ou non. Sábese:

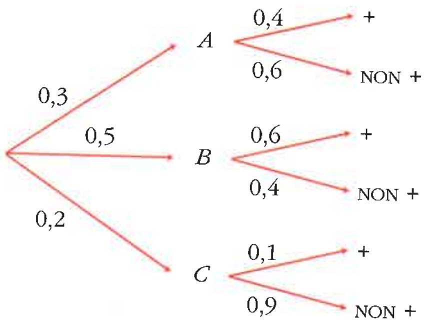
P[entre por A] = P[A] = 0,3

P[B] = 0,5; P[C] = 0,2

P[que o cace entrando en A] = = P[+/A] = 0,4

P[+/B] = 0,6; P[+/C] = 0,1

Calcula a probabilidade de que o gato cace ó rato.



$P[A \cap +] = 0{,}3 \cdot 0{,}4 = 0{,}12$ +

$P[A \cap \text{NON} +] = 0{,}3 \cdot 0{,}6 = 0{,}18$

$P[B \cap +] = 0{,}5 \cdot 0{,}6 = 0{,}30$ +

$P[B \cap \text{NON} +] = 0{,}5 \cdot 0{,}4 = 0{,}20$

$P[C \cap +] = 0{,}2 \cdot 0{,}1 = 0{,}02$ +

$P[C \cap \text{NON} +] = 0{,}2 \cdot 0{,}9 = 0{,}18$

$P[+] = 0{,}12 + 0{,}30 + 0{,}02 = 0{,}44.$

A probabilidade de que o gato acabe cazando ó rato é 0,44.

Observa que razoando sobre o gráfico obtense a fórmula da probabilidade total:

$$P[+] = P[A]\cdot P[+/A] + P[B]\cdot P[+/B] + P[C]\cdot P[+/C] =$$
$$= 0{,}3\cdot 0{,}4 + 0{,}5\cdot 0{,}6 + 0{,}2\cdot 0{,}1 = 0{,}44$$

EXERCICIOS PROPOSTOS

1. Temos dúas urnas:

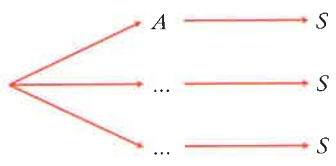
I  II 

A experiencia consiste en extraer unha bóla de I, introducila en II, remover e extraer, finalmente, unha bóla de II. Calcula a probabilidade de que a segunda bóla extraída sexa:

a) vermella b) verde c) negra

10.7 PROBABILIDADES "A POSTERIORI". FÓRMULA DE BAYES

Nunha experiencia composta de dous, se A é un suceso correspondente á primeira e S é un suceso correspondente á segunda, o significado da probabilidade condicionada $P[A/S]$ é interesante.



Pódese chegar ó suceso S pasando por A ou por outros sucesos. Se sabemos que finalmente ocorreu S, ¿cal é a probabilidade de que fora "pasando por A"? $P[A/S]$.

É dicir, das distintas formas en que se pode chegar a S, ¿en que proporción delas se pasa por A?

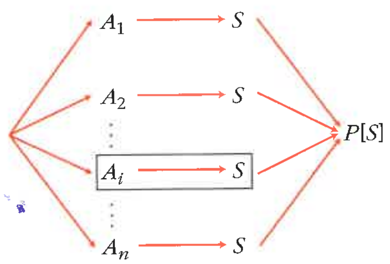
Intuitivamente pode admitirse que dita proporción é:

$$P[A/S] = \frac{P[A \cap S]}{P[S]}$$

A súa demostración non é difícil:

$$\left.\begin{aligned} P[A \cap S] &= P[A] \cdot P[S/A] \\ P[A \cap S] &= P[S] \cdot P[A/S] \end{aligned}\right\} \quad P[A] \cdot P[S/A] = P[S] \cdot P[A/S]$$

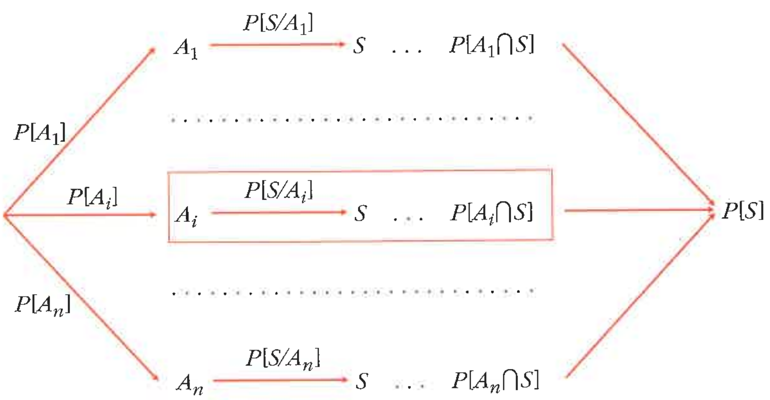
$$\text{Despexando: } P[A/S] = \frac{P[A] \cdot P[S/A]}{P[S]} = \frac{P[A \cap S]}{P[S]}$$



Se expresamos S como probabilidade total, obtense a chamada *fórmula de Bayes*:

$$\boxed{P[A_i/S] = \frac{P[A_i] \cdot P[S/A_i]}{P[A_1] \cdot P[S/A_1] + \ldots + P[A_n] \cdot P[S/A_n]}}$$

Na práctica, máis que a fórmula, resulta moi útil seguir o proceso nun diagrama en árbore:



Se sabemos que ocorreu S, sobre o gráfico obtense que:

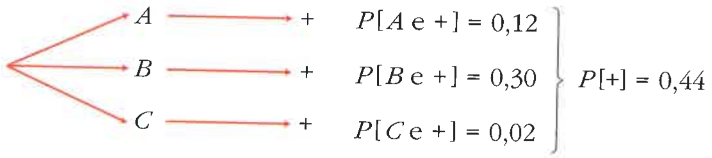
$$\boxed{P[A_i/S] = \frac{P[A_i \cap S]}{P[S]}}$$

EXERCICIOS RESOLTOS

1. *Retomamos o problema do rato e do gato que vimos no exercicio resolto 2 do apartado anterior. supoñamos que vemos a un gato perseguir a un rato. Ó pouco chega con el na boca. ¿En cal dos tres camiños o cazaría?*

Naturalmente, non saberemos ónde o cazou pero si podemos calcular a probabilidade de que o fixera en certa calexa: $P[A/+]$, $P[B/+]$, $P[C/+]$.

Xa coñecemos algúns resultados obtidos no apartado anterior:



O cálculo das probabilidades pedidas é inmediato:

$$P[A/+] = \frac{P[A \text{ e } +]}{P[+]} = \frac{0,12}{0,44} = 0,273$$

$$P[B/+] = \frac{P[B \text{ e } +]}{P[+]} = \frac{0,30}{0,44} = 0,682$$

$$P[C/+] = \frac{P[C \text{ e } +]}{P[+]} = \frac{0,02}{0,44} = 0,045$$

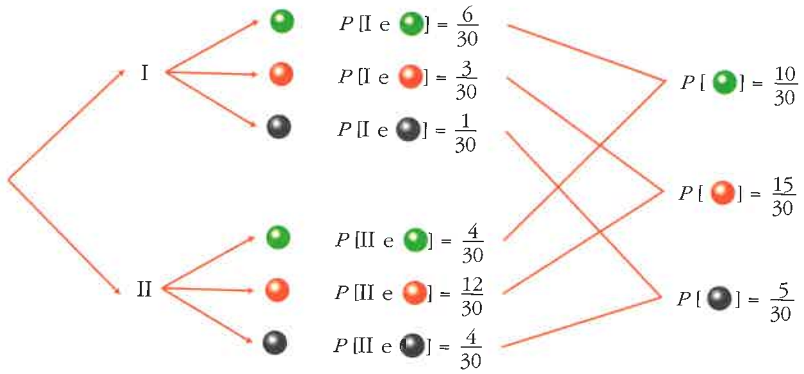
2. *Retomamos o problema do dado e as dúas urnas resolto no apartado anterior.*

a) Sabemos que, finalmente, obtívose unha bóla negra. ¿Cal é a probabilidade de que se extraese da urna II? É dicir P[II/⚫].

b) P[I/🔴]

c) P[I/🟢]

Toma os resultados da resolución do apartado anterior:



a) $P[\text{II}/⚫] = \frac{4/30}{5/30} = \frac{4}{5} = 0,8$

De cada 5 veces obtense bóla negra, en 4 delas provén da urna II.

b) $P[\text{I}/🔴] = \frac{3/30}{15/30} = \frac{3}{15} = 0,2$

c) $P[\text{I}/🟢] = \frac{6/30}{10/30} = \frac{6}{10} = 0,6$

EXERCICIOS PROPOSTOS

1. No exercicio proposto do apartado anterior, calcula:

a) Sabendo que a segunda bóla foi negra, ¿cal é a probabilidade de que a primeira tamén o fora? $P[1^a$⚫$/2^a$⚫$]$

b) Sabendo que a segunda bóla foi vermella, ¿cal é a probabilidade de que primeira fora negra? $P[1^a$⚫$/2^a$🔴$]$

c) ¿Cal é a probabilidade de que primeira fora verde sendo verde a segunda? $P[1^a$🟢$/2^a$🟢$]$

1. Espacio mostral. Sucesos

Nun sorteo de lotería fixámonos na cifra en que termina o "gordo".

a) ¿Cal é o espacio mostral?

b) Describe os sucesos

A = "Menor ca 4"

B = "Par"

C = "Maior ca 5",

escribindo todos os seus elementos.

c) Calcula os sucesos $A \cup B$, $B \cap C$, $A' \cap B'$, $A \cap C$.

d) ¿Cantos sucesos hai?

a) Espacio mostral: $E = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$

b) $A = \{0, 1, 2, 3\}$, $B = \{0, 2, 4, 6, 8\}$, $C = \{6, 7, 8, 9\}$

c) $A \cup B = \{0, 1, 2, 3, 4, 6, 8\}$

$B \cap C = \{6, 8\}$

$$\left.\begin{array}{l} A' = \{4, 5, 6, 7, 8, 9\} \\ B' = \{1, 3, 5, 7, 9\}, \end{array}\right\} A' \cap B' = \{5, 7, 9\}$$

$A \cap C = \varnothing$ (suceso imposible)

d) Hai $2^{10} = 1\,024$ sucesos.

2. Propiedades das probabilidades

Dos sucesos A e B sábese que:

$P[A] = 0{,}4$

$P[B] = 0{,}5$

$P[A' \cap B'] = 0{,}3$

Acha $P[A \cup B]$ e $P[A \cap B]$.

$P[A' \cap B'] = P[(A \cup B)'] = 1 - P[A \cup B]$

Como $P[A' \cap B'] = 0{,}3$, obtemos:

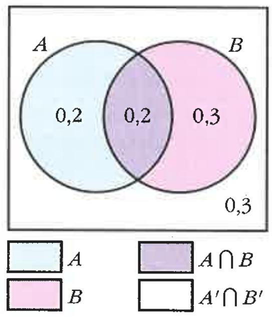
$$\boxed{P[A \cup B] = 1 - 0{,}3 = 0{,}7}$$

Para calcular $P[A \cap B]$ aplicamos a igualdade:

$$P[A \cup B] = P[A] + P[B] - P[A \cap B]$$

$$0{,}7 = 0{,}4 + 0{,}5 - P[A \cap B] \Rightarrow$$

$$\Rightarrow \boxed{P[A \cap B] = 0{,}2}$$



3. Probabilidade condicionada. Sucesos independentes

Considéranse dous sucesos, A e B, asociados a un experimento aleatorio con $P[A] = 0{,}7$, $P[B] = 0{,}6$ e $P[A' \cup B'] = 0{,}58$.

a) ¿Son independentes A e B?

b) Se $M \subset A$, ¿cal é o valor de $P[M'/A']$?

a) Para ver se son independentes, comprobaremos se se cumpre a seguinte igualdade:

$$P[A \cap B] = P[A] \cdot P[B]$$

$$P[A' \cup B'] = P[(A \cap B)'] = 1 - P[A \cap B]$$

Por tanto, $P[A \cap B] = 1 - P[A' \cup B'] = 1 - 0{,}58 = 0{,}42$

$$P[A] \cdot P[B] = 0{,}7 \cdot 0{,}6 = 0{,}42$$

A e B son independentes, pois $P[A \cap B] = P[A] \cdot P[B] = 0{,}42$.

b) $M \subset A \Rightarrow A' \subset M'$. Por tanto:

$$P[M'/A'] = \frac{P[M' \cap A']}{P[A']} = \frac{P[A']}{P[A']} = 1$$

4. Representación gráfica de sucesos

Nunha empresa hai 160 traballadores. Se eliximos un deles ó azar, temos estas probabilidades de que fale os seguintes idiomas:

a) 0,0625, inglés, francés e alemán.

b) 0,175, inglés e francés.

c) 0,15625, inglés e alemán.

d) 0,1375, francés e alemán.

e) 0,375, francés.

f) 0,35625, alemán.

g) 0,425, inglés.

¿Cantos traballadores falan un só dos tres idiomas (inglés, francés ou alemán)? ¿E cantos non falan ningún dos tres idiomas?

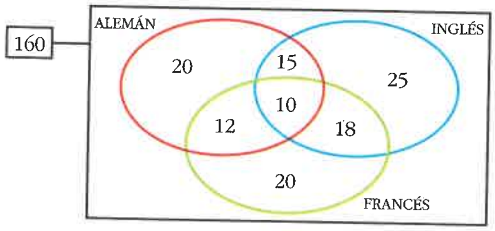
- Comezamos calculando o número de traballadores en cada caso (sabendo que o número total é $n = 160$):

a) $0{,}0625 \cdot 160 = 10$ b) $0{,}175 \cdot 160 = 28$ c) $0{,}15625 \cdot 160 = 25$

d) $0{,}1375 \cdot 160 = 22$ e) $0{,}375 \cdot 160 = 60$ f) $0{,}35625 \cdot 160 = 57$

g) $0{,}425 \cdot 160 = 68$

- Facemos un diagrama cos datos que temos:



- Agora é fácil responder as preguntas:

20 só alemán + 25 só inglés + 20 só francés = 65 falan só un dos tres idiomas.

Por outra parte:

$20 + 15 + 10 + 12 + 18 + 20 + 25 = 120$ falan algún idioma.

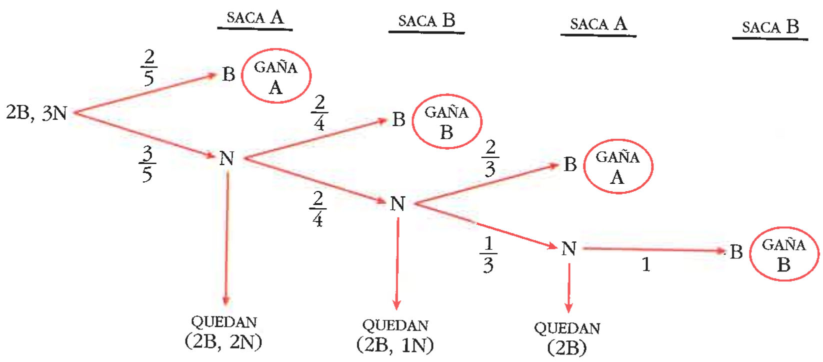
Logo, $160 - 120 = 40$ non falan ningún dos tres idiomas.

5. Experiencias compostas

Nunha urna hai 2 bólas brancas e 3 negras. Dúas persoas sacan, alternativamente, unha bóla de cada unha, sen volta. Gaña a primeira que saque unha bóla branca.

¿Cal é a probabilidade de que gañe a persoa que comeza o xogo?

Vexamos o proceso mediante un diagrama en árbore:



A probabilidade de que gañe a persoa que comeza o xogo é:

$$P[A] = \frac{2}{5} + \frac{3}{5} \cdot \frac{2}{4} \cdot \frac{2}{3} = \frac{3}{5}$$

E a de que gañe a outra persoa: $P[B] = \frac{3}{5} \cdot \frac{2}{4} + \frac{3}{5} \cdot \frac{2}{4} \cdot \frac{1}{3} = \frac{2}{5}$

Polo tanto, é vantaxoso comezar o xogo.

6. Experiencias compostas: Probabilidade total e "a posteriori"

Nunha casa hai tres chaveiros, A, B e C, o primeiro con 5 chaves, o segundo con 7 e o terceiro con 8, das que só unha de cada chaveiro abre a porta da trasteira. Escóllese ó azar un chaveiro e, del, unha chave para intentar abrir a trasteira.

a) ¿Cal será a probabilidade de que se acerte coa chave?

b) ¿Cal será a probabilidade de que o chaveiro escollido sexa o terceiro e a chave non abra?

c) E se a chave escollida é a correcta, ¿cal será a probabilidade de que pertenza ó primeiro chaveiro A?

Describimos o proceso mediante un diagrama en árbore:

A $\left(\frac{1}{3}\right)$:
- $\frac{1}{5}$ SI $\quad P[A \text{ e SI}] = \frac{1}{3}\cdot\frac{1}{5} = \frac{1}{15}$
- $\frac{4}{5}$ NON $\quad P[A \text{ e NON}] = \frac{1}{3}\cdot\frac{4}{5} = \frac{4}{15}$

B $\left(\frac{1}{3}\right)$:
- $\frac{1}{7}$ SI $\quad P[B \text{ e SI}] = \frac{1}{3}\cdot\frac{1}{7} = \frac{1}{21}$
- $\frac{6}{7}$ NON $\quad P[B \text{ e NON}] = \frac{1}{3}\cdot\frac{6}{7} = \frac{6}{21}$

C $\left(\frac{1}{3}\right)$:
- $\frac{1}{8}$ SI $\quad P[C \text{ e SI}] = \frac{1}{3}\cdot\frac{1}{8} = \frac{1}{24}$
- $\frac{7}{8}$ NON $\quad P[C \text{ e NON}] = \frac{1}{3}\cdot\frac{7}{8} = \frac{7}{24}$

a) $P[\text{SI}] = \frac{1}{15} + \frac{1}{21} + \frac{1}{24} = \frac{131}{840}$

b) $P[C \text{ e NON}] = \frac{7}{24}$

c) Resolvemos este apartado por sentido común. De 840 intentos, 131 dan lugar a obter a *chave boa*. E, deles, 56 proveñen do chaveiro *A*, 40 do *B* e 35 do *C*.

Se sabemos que saíu a chave correcta, a probabilidade de que sexa o chaveiro *A* será, pois, $\frac{56}{131}$.

Tamén se pode resolver aplicando a fórmula de Bayes:

$$P[A/\text{SÍ}] = \frac{P[A \cap \text{SÍ}]}{P[\text{SÍ}]} = \frac{1/15}{131/840} = \frac{56}{131}$$

7. Experiencias compostas: Probabilidade total e "a posteriori"

Temos dous dados, un normal e outro trucado. No trucado hai 4 uns e 2 douses. Elíxese un dado ó azar e tírase dúas veces.

a) ¿Cal é a probabilidade de obter un 1 na primeira tirada e un 2 na segunda?

b) Se o resultado da primeira tirada foi 1 e o da segunda 2, calcula a probabilidade de que se escollera o dado trucado.

a)

ELECCIÓN DO DADO:
- $\frac{1}{2}$ DADO CORRECTO (1, 2, 3, 4, 5, 6) $\xrightarrow{\frac{1}{6}}$ $\boxed{1}$ (1ª tirada) $\xrightarrow{\frac{1}{6}}$ $\boxed{2}$ (2ª tirada) $\quad P[\text{CORRECTO e } 1-2] = \frac{1}{2}\cdot\frac{1}{6}\cdot\frac{1}{6}$
- $\frac{1}{2}$ DADO TRUCADO (1, 1, 1, 1, 2, 2) $\xrightarrow{\frac{4}{6}}$ $\boxed{1}$ (1ª tirada) $\xrightarrow{\frac{2}{6}}$ $\boxed{2}$ (2ª tirada) $\quad P[\text{TRUCADO e } 1-2] = \frac{1}{2}\cdot\frac{4}{6}\cdot\frac{2}{6}$

$$P[1-2] = P[\text{CORRECTO e } 1-2] + P[\text{TRUCADO e } 1-2] =$$

$$= \frac{1}{72} + \frac{8}{72} = \frac{9}{72} = \frac{1}{8}$$

b) Aplicamos a fórmula de Bayes:

$$P[\text{TRUCADO}/1-2] = \frac{P[\text{TRUCADO e } 1-2]}{P[1-2]} = \frac{8/72}{(1/72) + (8/72)} = \frac{8}{9}$$

PARA PRACTICAR

1 Lanzamos un dado e unha moeda. Os posibles resultados son (1, C), (1, +), (2, C)…

a) Describe o espacio mostral cos doce elementos dos que consta.

Sexan os sucesos:

A = "Sacar un ou dous no dado"

B = "Sacar + na moeda"

$D = \{(1, C), (2, +), (3, C), (3, +), (6, +)\}$

b) Describe os sucesos A e B mediante todos os elementos.

c) Calcula $A \cup B$, $A \cap B$, $A \cup D'$.

2 Sexa $U = \{a_1, a_2, a_3\}$ o espacio de sucesos elementais dun experimento aleatorio. ¿Cales destas funcións definen unha función de probabilidade? Xustifica a resposta.

a) $P[a_1] = 1/2$, $P[a_2] = 1/3$, $P[a_3] = 1/6$

b) $P[a_1] = 3/4$, $P[a_2] = 1/4$, $P[a_3] = 1/4$

c) $P[a_1] = 1/2$, $P[a_2] = 0$, $P[a_3] = 1/2$

d) $P[a_1] = 2/3$, $P[a_2] = 1/3$, $P[a_3] = 1/3$

3 Determina se son compatibles ou incompatibles os sucesos A e B, sabendo que:

$P[A] = 1/4$, $P[B] = 1/2$, $P[A \cup B] = 2/3$

4 Unha experiencia aleatoria consiste en preguntarlles a tres persoas distintas, elixidas ó azar, se son partidarias ou non de consumir un determinado producto.

a) Escribe o espacio mostral asociado a dito experimento utilizando a letra"s" para as respostas afirmativas e o "n" para as negativas.

b) ¿Que elementos do espacio mostral anterior constitúen o suceso "ó menos dúas das persoas son partidarias de consumir o producto"?

c) Describe o suceso contrario de "máis dunha persoa é partidaria de consumir o producto".

5 En familias de tres fillos, estúdiase a distribución dos seus sexos. Por exemplo (H, M, M) significa que o maior é home e os outros dous mulleres. ¿Cantos elementos ten o espacio mostral E? Describe os seguintes sucesos: A = "A menor é muller", B = "O maior é home". ¿En que consiste $A \cup B$?

6 Lánzanse dous dados. Calcula a probabilidade de que a maior das puntuacións sexa un 1, un 2, un 3, un 4, un 5, un 6.

	1	2	3	4	5	6
1				4		
2	2	2	3	4	5	6
3				4		
4				4		
5				5		
6				6		

Completa esta táboa e razoa sobre ela.

7 Unha clase componse de vinte alumnos e dez alumnas. A metade das alumnas e a metade dos alumnos aproban as matemáticas. Calcula a probabilidade de que, ó elixir unha persoa ó azar, resulte ser:

a) Alumna ou que aproba as matemáticas.

b) Alumno que suspenda as matemáticas.

c) Sabendo que é alumno, ¿cal é a probabilidade de que aprobe as matemáticas?

d) ¿Son independentes os sucesos ALUMNO e APROBA MATEMÁTICAS?

Fai unha táboa de continxencia.

8 Di cál é o espacio mostral correspondente ás seguintes experiencias aleatorias. Se é finito e ten poucos elementos, díos todos, e se ten moitos, descríbeo e di o número total.

a) Extraemos unha carta dunha baralla española e anotamos o número.

b) Extraemos unha carta dunha baralla española e anotamos o pao.

c) Extraemos dúas cartas dunha baralla española e anotamos o pao de cada unha.

d) Lanzamos seis moedas distintas e anotamos o resultado.

e) Lanzamos seis moedas distintas e anotamos o número de caras.

PARA RESOLVER

9 Nunha caixa hai seis bólas numeradas, tres delas con números positivos e as outras tres con números negativos. Extráese unha bóla e despois outra, sen volta.

a) Calcula a probabilidade de que o producto dos números obtidos sexa positivo.

b) Calcula a probabilidade de que o producto dos números obtidos sexa negativo.

10 S Nunha certa cidade, o 40% da poboación ten cabelos castaños, o 25% ten os ollos castaños e o 15% ten cabelos e ollos castaños.

Escóllese unha persoa ó azar:

a) Se ten cabelos castaños, ¿cal é a probabilidade de que tamén teña ollos castaños?

b) Se ten ollos castaños, ¿cal é a probabilidade de que teña cabelos castaños?

c) ¿Cal é a probabilidade de que non teña cabelos nin ollos castaños?

Usa unha táboa como a seguinte:

	OLLOS CAST.	OLLOS NON CAST.	
CAB. CAST.	15		40
CAB. NON CAST.			
	25		100

11 Dúas persoas xogan a obter a puntuación máis alta lanzando os seus dados A e B. O dado A ten catro caras coa puntuación 6 e as outras dúas caras coa puntuación 10. O dado B ten unha cara coa puntuación 3, catro caras con puntuación 6 e a outra con puntuación 12. ¿Que xogador ten máis probabilidade de gañar?

Fai unha táboa na que aparezan as 6 posibilidades do dado A e as do dado B. En cada unha das 36 casas anota quén gaña en cada caso.

12 S Dos sucesos A e B sábese que:

$$P[A] = \frac{2}{5}, \quad P[B] = \frac{1}{3} \quad \text{e} \quad P[A' \cap B'] = \frac{1}{3}$$

Calcula $P[A \cup B]$ e $P[A \cap B]$.

13 S Sexan A e B dous sucesos dun espacio de probabilidade, de maneira que:

$$P[A] = 0{,}4, \quad P[B] = 0{,}3 \quad \text{e} \quad P[A \cap B] = 0{,}1$$

Calcula razoadamente:

a) $P[A \cup B]$ b) $P[A' \cup B']$

c) $P[A/B]$ d) $P[A' \cap B']$

14 A, B, e C son tres sucesos dun mesmo espacio mostral. Expresa en función deles os sucesos:

a) Realízase algún dos tres. b) Non se realiza ningún dos tres. c) Realízanse os tres. d) Realízanse dous dos tres. e) Realízanse, ó menos, dous dos tres.

15 S Un exame consiste en elixir ó azar dous temas de entre os dez do programa e desenvolver un deles.

a) Un alumno sabe 6 temas. ¿Que probabilidade ten de aprobar o exame?

b) ¿Que probabilidade ten o mesmo alumno de saber un dos temas elixidos e o outro non?

16 Lánzase un dado dúas veces. Calcula a probabilidade de que na segunda tirada se obteña un valor maior ca na primeira.

17 S Un estudiante fai dúas probas nun mesmo día. A probabilidade de que pase a primeira proba é 0,6. A probabilidade de que pase a segunda é 0,8 e a de que pase ambas as dúas e 0,5. Pídese:

a) Probabilidade de que pase ó menos unha proba.

b) Probabilidade de que non pase ningunha proba.

c) ¿Son as probas sucesos independentes?

d) Probabilidade de que pase a segunda proba en caso de non superar a primeira.

18 Nunha comarca hai dous periódicos: *O Progresista* e *O Liberal*. Sábese que o 55% das persoas desa comarca le *O Progresista (P)*, o 40% le *O Liberal (L)* e o 25% non le ningún deles.

Expresa en función de P e L estes sucesos:

a) Ler os dous periódicos.

b) Ler só *O Liberal*.

c) Ler só *O Progresista*.

d) Ler algún dos dous periódicos.

e) Non ler ningún dos dous.

f) Ler só un dos dous.

g) Calcula as probabilidades de: P, L, $P \cap L$, $P \cup L$, $P - L$, $L - P$, $(L \cup P)'$, $(L \cap P)'$.

h) Sabemos que unha persoa le *O Progresista*. ¿Que probabilidade hai de que, ademais, lea *O Liberal*? ¿E de que non o lea?

19 Unha urna A ten 3 bólas brancas e 7 negras. Outra urna B ten 9 bólas brancas e 1 negra. Escollemos unha das urnas ó azar e dela extraemos unha bóla. Calcula:

a) $P[\text{BRANCA}/A]$

b) $P[\text{BRANCA}/B]$

c) $P[A \text{ e BRANCA}]$

d) $P[B \text{ e BRANCA}]$

e) $P[\text{BRANCA}]$

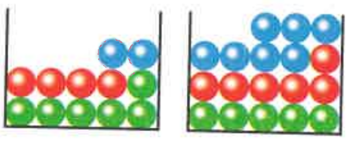
f) Sabendo que a bóla obtida foi branca, ¿cal é a probabilidade de que escolleramos a urna B?

20 Temos as mesmas urnas do exercicio anterior. Sacamos unha bóla de A e botámola en B e, a continuación, sacamos unha bóla de B.

a) ¿Cal é a probabilidade de que a segunda bóla sexa negra?

b) Sabendo que a segunda bóla foi negra, ¿cal é a probabilidade de que tamén a primeira fose negra?

21 Temos dúas urnas con estas composicións:



Extraemos unha bóla de cada urna. ¿Cal é a probabilidade de que sexan da mesma cor? ¿E a probabilidade de que sexan de distinta cor?

22 **S** Un aparato eléctrico está constituído por dous compoñentes A e B. Sabendo que hai unha probabilidade de 0,58 de que non falle ningún dos compoñentes e que no 32% dos casos falla B non fallando A, determina, xustificando a resposta, a probabilidade de que nun de tales aparatos non falle a compoñente A.

23 Dous xogadores botan á vez dúas moedas cada un. ¿Cal é a probabilidade de que ambos obteñan o mesmo número de caras (cero, unha ou dúas)? Razóao.

24 Lánzase un dado repetidas veces e estamos interesados no número de tiradas precisas para obter un 6 por primeira vez.

a) ¿Cal é o espacio mostral?

b) ¿Cal é a probabilidade de que o primeiro 6 se obteña na sétima tirada?

25 Un producto está formado de dúas partes: A e B. O proceso de fabricación é tal, que a probabilidade dun defecto en A é 0,06 e a probabilidade dun defecto en B é 0,07. ¿Cal é a probabilidade de que o producto non sexa defectuoso?

26 **S** Unha urna contén 10 bólas brancas, 6 negras e 4 vermellas. Se se extraen tres bólas, volvéndoas botar logo dentro, ¿cal é a probabilidade de obter 2 brancas e unha vermella?

27 **S** Unha urna A contén 6 bólas brancas e 4 negras. Outra urna B ten 5 brancas e 9 negras. Eliximos una urna ó azar e extraemos dúas bólas, que resultan ser brancas. Determina a probabilidade de que a urna elixida fora a A.

28 **S** Disponse de tres urnas: a A que contén dúas bólas brancas e catro vermellas, a B con tres brancas e tres roxas; e a C cunha branca e cinco vermellas.

a) Elíxese unha urna ó azar e extráese unha bóla dela. ¿Cal é a probabilidade de que esta bóla sexa branca?

b) Se a bóla extraída resulta ser branca, ¿cal é a probabilidade de que proceda da urna B?

29 S Sexan A e B dous sucesos tales que:

$$P[A \cup B] = \frac{3}{4} \qquad P[B'] = \frac{2}{3} \qquad P[A \cap B] = \frac{1}{4}$$

Calcula $P[B]$, $P[A]$, $P[A' \cap B]$.

30 S En certo país onde a enfermidade X é endémica, sábese que un 12% da poboación padece dita enfermidade. Disponse dunha proba para detectar a enfermidade, pero non é totalmente fiable, xa que dá positiva no 90% dos casos de persoas realmente enfermas e tamén dá positiva no 5% de persoas sas. ¿Cal é a probabilidade de que estea sa unha persoa á que a proba lle dera positiva?

31 S En tres máquinas, A, B e C, fabrícanse pezas da mesma natureza. A porcentaxe de pezas que resultan defectuosas en cada máquina é, respectivamente, 1%, 2% e 3%. Mestúranse 300 pezas, 100 de cada máquina, e elíxese unha peza ó azar, que resulta ser defectuosa. ¿Cal é a probabilidade de que fora fabricada na máquina A?

32 S Unha caixa A contén dúas bólas brancas e dúas vermellas, e outra caixa B contén tres brancas e dúas vermellas. Pásase unha bóla de A a B e despois extráese unha bóla de B, que resulta ser branca. Determina a probabilidade de que a bóla trasladada fora branca.

33 S Unha urna A contén 5 bólas brancas e 3 negras. Outra urna B, 6 brancas e 4 negras. Eliximos unha urna ó azar e extraemos dúas bólas, que resultan ser negras. Calcula a probabilidade de que a urna elixida fora a B.

34 S Teño dúas urnas, dúas bólas brancas e dúas bólas negras. Deséxase saber cómo debo distribuír as bólas nas urnas para que, ó elixir unha urna ó azar e extraer dela unha bóla ó azar, sexa máxima a probabilidade de obter bóla branca. A única condición esixida é que cada urna teña ó menos unha bóla.

35 S Sexan A e B dous montóns de cartas. En A hai 8 ouros e 5 espadas e, en B, 4 ouros e 7 espadas. Sacamos dúas cartas do mesmo montón e resulta que ambas as dúas son espadas. Determina a probabilidade de que as sacaramos do montón B.

36 Unha urna contén 25 bólas brancas sen marcar, 75 bólas brancas marcadas, 125 bólas negras sen marcar e 175 bólas negras marcadas.

a) Extráese unha bóla. Calcula a probabilidade de que sexa branca.

b) Extráese unha bóla e está marcada. ¿Cal é a probabilidade de que sexa branca?

c) Extráese unha bóla. ¿Cal é a probabilidade de que sexa negra e estea marcada?

d) ¿Son independentes os sucesos "sacar bóla marcada" e "sacar bóla branca"?

37 Dúas persoas enfróntanse nun xogo no que será vencedor o primeiro que gañe 5 partidas. Pero antes de finalizar o xogo, este interrómpese no momento en que un gañou 4 partidas e outro 3.

¿Como deben repartirse os 4 200 euros que apostaron?

Describe nun diagrama en árbore as posibles continuacións da partida.

38 Nun centro escolar hai tres grupos de Bacharelato. O primeiro está composto por 10 alumnos dos que 7 prefiren a música moderna, 2 prefiren a clásica e 1 que non lle gusta a música. No segundo, composto por 12 alumnos, a distribución de preferencias é 5, 7, 0, respectivamente; e, no terceiro, formado por 14 alumnos, a distribución de preferencias é 6, 6, 2, respectivamente.

Elíxese un grupo ó azar e regálanse 2 entradas para un concerto de música clásica a dous alumnos seleccionados ó azar.

a) Determina a probabilidade de que os dous alumnos elixidos sexan afeccionados á música clásica.

b) Se os dous alumnos agraciados son, efectivamente, afeccionados á música clásica, ¿cal é a probabilidade de que sexan do primeiro grupo?

Organiza os datos nunha táboa.

CUESTIÓNS TEÓRICAS

39 **S** Sexan A e B dous sucesos tales que $P[A] = 0{,}40$; $P[B/A] = 0{,}25$ e $P[B] = b$. Calcula:

a) $P[A \cap B]$ b) $P[A \cup B]$ se $b = 0{,}5$

c) O menor valor posible de b.

d) O maior valor posible de b.

40 Se a probabilidade de que acontezan dous sucesos á vez é p, ¿cal é a probabilidade de que ó menos un dos dous non aconteza? Razóao.

41 Razoa a seguinte afirmación: Se a probabilidade de que ocorran dous sucesos á vez é menor ca 1/2, a suma das probabilidades de ambos (por separado), non pode exceder de 3/2.

42 Sexan A e B dous sucesos dun experimento aleatorio. ¿É posible que p sexa unha probabilidade se: $P[A] = \frac{2}{5}$, $P[B] = \frac{1}{5}$ e $P[A' \cap B'] = \frac{3}{10}$?

43 Sexa A un suceso con $0 < P[A] < 1$.

a) ¿Pode ser A independente do seu contrario A'?

b) Sexa B outro suceso tal que $B \subset A$. ¿Serán A e B independentes?

c) Sexa C un suceso independente de A. ¿Serán A e C' independentes?

Xustifica as respostas.

44 Se A e B son dous sucesos de experimento aleatorio e $P[A] = 0$:

a) ¿Que podemos dicir de $P[A \cap B]$?

b) ¿E de $P[A \cup B]$?

c) Responde as mesmas preguntas se $P[A] = 1$.

45 Ó tirar tres dados, podemos obter unha suma 9 de seis formas distintas:

126, 135, 144, 225, 234, 333

e outras seis de obter suma 10:

136, 145, 226, 235, 244, 334

Sen embargo, a experiencia dinos que é máis fácil obter suma 10 que suma 9. ¿Por que?

PARA PROFUNDAR

46 Un home ten tempo para xogar á ruleta 5 veces, ó sumo. Cada aposta é de 1 euro. O home comeza con 1 euro e deixará de xogar cando perda o euro ou gañe 3 euros.

a) Obtén o espacio mostral dos resultados posibles.

b) Se a probabilidade de gañar ou perder é a mesma en cada aposta, ¿cal é a probabilidade de que gañe 3 euros?

47 **S** Nunha baralla de 40 cartas, tómanse tres cartas distintas. Calcula a probabilidade de que as tres sexan números distintos.

48 **S** Escollidas cinco persoas ó azar, ¿cal é a probabilidade de que ó menos dúas delas naceran no mesmo día da semana (é dicir, en luns, martes, etc.)?

49 **S** Nunha competición de tiro con arco, cada tirador dispón, como máximo, de tres intentos para facer diana. No momento en que o consegue, deixa de tirar e supera a proba e, se non o consegue en ningún dos tres intentos, queda eliminado. Se a probabilidade de facer branco con cada frecha, para un determinado tirador, é 0,8:

a) Calcula a probabilidade de non quedar eliminado.

b) Se sabemos que superou a proba, ¿cal é a probabilidade de que o conseguira no segundo intento?

50 **S** Sexa A o suceso "unha determinada persoa A resolve un determinado problema" e B o suceso "resólveo a persoa B". Sábese que a probabilidade de que o resolvan as dúas persoas é de 1/6; e, a de que non o resolva ningunha das dúas é de 1/3. Sabendo que a probabilidade de que o resolva unha persoa é independente de que o resolva a outra, calcula $P(A)$ e $P(B)$.

51 ¿Que é máis probable, obter algunha vez un 6 lanzando un dado 4 veces ou un dobre 6 lanzado dous dados 24 veces?

11 AS MOSTRAS ESTATÍSTICAS

Hai un século a estatística limitábase ó estudio das grandes masas de datos ("cemiterios de números" chamoulles algún estatístico moderno). Toda a poboación era abarcada en censos minuciosamente elaborados. Ademais de conseguir os datos, o traballo estatístico, consistía en clasificalos, tabulalos, relacionalos. Esta concepción do que é a estatística, que aínda agora segue sendo común entre as persoas correntes (o "home da rúa") cambia drasticamente a partir dos anos trinta do século XX co nacemento da **estatística inductiva.** Con ela búscanse métodos que permitan obter conclusións válidas para toda a poboación a partir do estudio dunha mostra. Para iso, a estatística debe botar man da alta matemática, elaborando procedementos moi específicos. "Antes, a estatística limitábase a unha mera descrición empírica; agora, pode facer unha crítica da situación, porque se desenvolveron as bases matemáticas correspondentes" (Kreiszig, 1970).

A figura máis sobresaliente deste proceso é **Ronald A. Fisher** (1890-1962), pero hai que destacar, tamén, a **Yale, Karl Pearson, Neyman** e **E. Pearson** (fillo de K. Pearson). Estes dous últimos foron os creadores da teoría da mostraxe e dos ensaios de hipóteses.

PARA EMPEZAR, REFLEXIONA E RESOLVE

Unha sondaxe de opinión

Imaxinemos unha cidade na que hai 300 000 votantes. Para ter unha idea das súas opinións políticas facémoslles unha enquisa a 1 200 deles elixidos ó azar.

Posiblemente nos quede a sensación de que as conclusións ás que cheguemos sexan sumamente erróneas, que a imaxe que nos fagamos da realidade sexa deformada, falsa, pois cada individuo da mostra está representando a 250 individuos da poboación (300 000 : 1 200 = 250). ¿Como é posible que co que nos di un individuo pretendamos facernos unha idea do que opinan 250?

Sen embargo, e por chocante que poida parecernos, si que podemos obter unha boa idea da poboación a partir da mostra.

Unhas fotografías

Reflexionemos sobre as fotografías da páxina seguinte.

A da esquerda está formada por 300 000 puntos e dános unha imaxe nidia da realidade. A da dereita é unha mostra extraída da anterior. Partiuse nuns 1 200 cadradiños e cada un deles pintouse da cor dun dos seus puntos elixidos ó azar. Se a miramos de cerca (á distancia á que se mira unha páxina dun libro) o que vemos é deforme. Pero se a observamos a unha certa distancia, a imaxe gaña en nitidez e permítenos ter unha idea bastante clara da realidade.

Fai a proba mirándoas a 4 ou 5 m de distancia e verás cómo ambas se ven aproximadamente igual.

Os dous exemplos descritos na páxina anterior son similares: obter información sobre unha poboación de 300 000 individuos a partir dunha mostra de 1 200 deles. O exemplo das fotografías fainos ver que con tal mostra se consegue unha boa idea de cómo é a poboación. A fotografía que ves sobre estas liñas é unha mostra da da esquerda, se esta se considera como un conxunto de 300 000 puntos de cores. Se miramos esta páxina a unha certa distancia, ambas vense case igual.

NESTA UNIDADE VERÁS

Cos exemplos vistos nesta páxina puidemos apreciar que cunha mostra de tamaño moi inferior ó da poboación conséguese unha imaxe suficientemente boa da mesma. Esta é unha interesante característica das mostras (cun pequeno tamaño é suficiente). Ó longo da unidade prestaremos atención a outros aspectos das mostras e do procedemento para a súa obtención: a mostraxe.

- ¿Por que se recorre ás mostras? Hai multitude de causas que poden xustificar que, para descubrir certas características dunha poboación, se recorra a unha mostra extraída dela.

- Os elementos que compoñen unha mostra deben ser elixidos ó azar, de modo que todos os elementos da poboación teñan a mesma probabilidade (mostraxe aleatoria).

Pero isto pode conseguirse de diversas maneiras:

- Mostraxe aleatoria simple
- Mostraxe aleatoria sistemática
- Mostraxe aleatoria estratificada

A aleatoriedade da mostra, nas distintas versións da mostraxe, pode conseguirse mediante os *números aleatorios* xerados por unha calculadora manual ou por un ordenador.

11.1 O PAPEL DAS MOSTRAS

EXEMPLO DE FICHA TÉCNICA DUNHA SONDAXE

Proxecto e dirección técnica: IMOP

Universo: Poboación española maior de 18 anos.

Mostra: 1278 individuos.

Tipo de mostraxe: aleatoria, mediante entrevistas persoais seguindo un método estratificado por rexións.

Límite máximo de erro: ± 3,1%

Nivel de confianza: 95%

Frecuentemente encontramos nos medios de comunicación referencias a resultados de enquisas de opinión relativas a diversos aspectos da actualidade política ou sociolóxica: *"Valoración de diversos líderes políticos", "Tipo de lectura que se prefire"...*

Estas informacións veñen acompañadas da "ficha técnica" da enquisa correspondente, como a da marxe. Nela menciónanse certos conceptos (universo, mostra, tipo de mostraxe, límite do erro...) que iremos analizando nesta unidade e a seguinte.

Poboación e mostra

Se estamos interesados en coñecer o que opinan os electores sobre algúns líderes políticos, o colectivo que é obxecto do noso interese é o de todos os españois que poden votar: os maiores de 18 anos. É a **poboación** ou **universo**.

Non é posible preguntarlles a todos (sería moi caro e moi lento), polo que recorremos a algúns deles: unha **mostra.** A opinión dunhas persoas sérvenos para facernos unha idea do que opina a totalidade da poboación.

Poboación ou **universo** é o conxunto de todos os individuos obxecto do noso estudio.

Mostra é un subconxunto extraído da poboación. O seu estudio serve para inferir características de toda a poboación.

¿Por que se recorre ás mostras?

Na práctica é moi frecuente ter que recorrer a unha mostra para inferir datos dunha poboación por algún dos seguintes motivos:

- A poboación é excesivamente numerosa.

  Por exemplo, a totalidade dos españois que poden votar.

- A poboación é moi difícil ou imposible de controlar.

  Por exemplo, a totalidade das persoas que entran nuns grandes almacéns ó longo dunha semana.

- O proceso de medición é destructivo.

  Por exemplo, deséxase coñecer a duración media das lámpadas que hai nun almacén. A forma de saber a duración dunha lámpada é deixala acendida ata que se funda e cronometrar o tempo. É claro que só poderemos probar con algunhas delas.

- Deséxase coñecer rapidamente certos datos da poboación e tardaríase demasiado en consultar a todos.

  Por exemplo, as sondaxes electorais ou de opinión.

11.2 ¿COMO DEBEN SER AS MOSTRAS?

Hai dous aspectos das mostras ós que debemos prestarlles moita atención: o seu **tamaño** e **cómo se realiza a selección** dos individuos que a forman.

Respecto ó **tamaño,** é claro que se a mostra é demasiado pequena non poderemos obter dela ningunha conclusión que mereza a pena. Sen embargo, como vimos nas páxinas iniciais desta unidade, conséguense imaxes sorprendentemente boas da realidade con mostras relativamente pequenas. Na próxima unidade aprenderemos a obter con exactitude o tamaño (número de individuos) que debe ter unha mostra para conseguir o que nos propoñamos.

Vexamos, a continuación, **cómo se seleccionan** os elementos da mostra.

Mostraxe

Ó substituír o estudio da poboación polo da mostra, cométense erros. Pero con eles contamos de antemán e poden controlarse.

Sen embargo, se a mostra está mal elixida (non é **representativa**) prodúcense erros adicionais imprevistos e incontrolables (**nesgos**).

A elección da mostra chámase **mostraxe.** Vexamos a continuación cómo debe realizarse a mostraxe para que nos proporcione mostras representativas.

UNHA ANÉCDOTA INTERESANTE

Nas eleccións americanas de 1936, nas que gañou Roosevelt, unha revista fixo unha enquisa de intención de voto a máis de catro millóns dos seus lectores e equivocouse no seu prognóstico.

Outra enquisa realizada só a 4 500 persoas anunciou o éxito de Roosevelt con moita exactitude.

A razón é que no primeiro caso a mostra non era representativa da sociedade americana, pois todos eran lectores dunha mesma revista, mentres que nas 4 500 persoas da segunda estaban ben representados todos os estamentos e ideoloxías de dita sociedade.

Mostraxe aleatoria

Unha condición *case* indispensable para que unha mostra sexa representativa é que os seus elementos se elixiran aleatoriamente, ó azar. Se a elección é subxectiva, os prexuízos de quen fai a elección proxéctanse no resultado da mostra que reflectirá o que esta persoa *cre* que é a realidade.

Por exemplo, imaxina que nun centro escolar se desexa saber o tempo que dedican a estudiar, por temo medio, os 1 300 alumnos e alumnas e para iso extráese unha mostra de 100 deles.

a) Se fora o director ou unha comisión de profesores os que elixiran os alumnos, procurando que houbera alumnos "bos", "medianos", "frouxos"... a mostra sería parcial, pois non reflectiría a realidade senón o que o director ou os profesores creran ver da realidade.

b) Se se elixiran os 100 primeiros alumnos e alumnas que cheguen ó centro un certo día, tamén a mostra estaría *contaminada,* porque é posible que chegar pronto ó centro teña que ver co grao de responsabilidade de ditos alumnos e, por tanto, coa súa dedicación ó estudio.

Dise que unha **mostraxe** é **aleatoria** cando todos os individuos da mostra se elixen ó azar, de modo que todos os individuos da poboación teñen, *a priori,* a mesma probabilidade de ser elixidos.

No apartado seguinte veremos distintos tipos de mostraxes aleatorias.

11.3 TIPOS DE MOSTRAXES ALEATORIAS

Mostraxe aleatoria simple

É o tipo de mostraxe máis sinxela e no que se basean todas as demais. Para obter unha mostra, numéranse os elementos da poboación e selecciónanse ó chou os n elementos que debe conter a mostra. Se os individuos son, por exemplo, parafusos contidos nun caixón, para obter a mostra chega con tomar n deles por simple extracción.

Mostraxe aleatoria sistemática

Móstranse os individuos e, a partir dun deles elixido ó chou, tómanse os seguintes mediante "saltos" numéricos iguais. Por exemplo se o primeiro é o 5°, e o "salto" é de 13, escolleranse 5º, 18º, 31º, 44º...

O "salto" denomínase **coeficiente de elevación**, h, e obtense mediante o cociente enteiro entre o número de individuos da poboación, N, e o número de individuos da mostra, n: $h = N/n$.

O primeiro elemento, chamado **orixe**, escóllese ó chou entre os números 1, 2, 3, ..., h.

Unha vez numerados os N individuos da poboación e sabendo que a mostra ha de ser de tamaño n, o proceso que se segue é:

— Calcúlase o coeficiente de elevación, h, dividindo N entre n.

— Pescúdase o primeiro elemento da mostra, a_1, obténdoo aleatoriamente de entre os h primeiros.

— Obtéñense os restantes elementos da mostra: $a_2 = a_1 + h$, $a_3 = a_2 + h$, $a_4 = a_3 + h$, ...

Esta forma de mostraxe só é válida se o criterio polo que se numeraron os individuos da poboación non ten nada que ver coa característica que se quere estudiar a partir da mostra.

EXERCICIOS RESOLTOS

1. *Nun centro escolar hai 1 300 alumnos. Explica cómo se escolle unha mostra de tamaño 100.*

   a) Mediante mostraxe aleatoria simple.

   b) Mediante mostraxe aleatoria sistemática.

a) Sortéanse 100 números de entre os 1 300. A mostra estará formada polos 100 alumnos ós que lles correspondan estes números.

b) Coeficiente de elevación: $h = \dfrac{1\,300}{100} = 13$

— Sortéase un número do 1 ó 13. Supoñamos que sae o 5.

— Os alumnos seleccionados para a mostra son os que corresponden ós números 5, 18, 31, 44, 57, ..., 1 292.

EXERCICIOS PROPOSTOS

1. Unha gandería ten 3 000 vacas. Quérese extraer unha mostra de 120. Explica cómo se obtén a mostra:

   a) Mediante mostraxe aleatoria simple.

   b) Mediante mostraxe aleatoria sistemática.

Mostraxe aleatoria estratificada

Se a poboación pode dividirse en estratos (por exemplo, por idades: menores de 18 anos; de 18 a 50; máis de 50), ás veces convén elixir a mostra fixando de antemán o número de individuos de cada estrato. Cando estes números son proporcionais ós tamaños dos estratos, dise que a mostraxe é **estratificada con reparto proporcional**.

ESTRATOS	E_1	E_2	E_3	TOTAL
Nº DE INDIV. NA POBOACIÓN	N_1	N_2	N_3	N
Nº DE INDIV. NA MOSTRA	n_1	n_2	n_3	n

$$\frac{n}{N} = \frac{n_1}{N_1} = \frac{n_2}{N_2} = \frac{n_3}{N_3}$$

En cada estrato, os n_i individuos da mostra elíxense aleatoriamente.

Procédese a unha mostraxe aleatoria estratificada cando se supón que a pertenza a un ou outro estrato inflúe na variable que estamos analizando. Por exemplo:

— Pode supoñerse que os alumnos de cursos superiores estudian máis cós demais.

— A idade inflúe nas opinións sobre aspectos sociolóxicos.

— A pertenza a unha ou outra comunidade autónoma pode influír na "renda per cápita", ou na "taxa de paro", ou no prezo da vivenda, ...

EXERCICIOS RESOLTOS

1. *Os 1 300 alumnos dun centro repártense así:*

426 de 1º
359 de 2º
267 de 3º
133 de 4º
115 de 5º

¿Como se elixirá unha mostra de 100 alumnos mediante mostraxe estratificada con reparto proporcional?

Debe cumprirse: $\frac{100}{1\,300} = \frac{n_1}{426} = \frac{n_2}{359} = \frac{n_3}{267} = \frac{n_4}{133} = \frac{n_5}{115}$

Calculamos n_1: $\frac{100}{1\,300} = \frac{n_1}{426} \rightarrow n_1 = \frac{100}{1\,300} \cdot 426 = = 32{,}77$

Analogamente obtéñense os demais:

$$n_2 = 27{,}62, \quad n_3 = 20{,}54, \quad n_4 = 10{,}23, \quad n_5 = 8{,}85$$

A parte enteira destes números suma: $32 + 27 + 20 + 10 + 8 = 97$

Faltan 3 para chegar a 100. Aumentarémoslle unha unidade ós tres cocientes que teñan a parte decimal maior: n_1, n_2 e n_5. Por tanto, os cen individuos da mostra obtéñense elixindo aleatoriamente os seguintes alumnos:

33 de 1º, 28 de 2º, 20 de 3º, 10 de 4º e 9 de 5º

Para que sexa razoable que recorreramos á mostraxe estratificada con reparto proporcional, a característica que se analiza debe depender, nalgunha medida, do curso no que se encontra o alumno. Por exemplo a *estatura,* ou ben o *número de horas semanais de estudio* ou outras.

EXERCICIOS PROPOSTOS

2. Unha gandería ten 2 000 vacas. Son de distintas razas: 853 de A, 512 de B, 321 de C, 204 de D e 110 de E. Queremos extraer unha mostra de 120:

a) ¿Cantas hai que elixir de cada raza para que a mostraxe sexa estratificada con reparto proporcional?

b) ¿Como debe ser a elección dentro de cada estrato?

11.4 TÉCNICAS PARA OBTER UNHA MOSTRA ALEATORIA DUNHA POBOACIÓN FINITA

Xa dixemos nas páxinas anteriores que para obter unha mostra aleatoria "sortéanse" os individuos da poboación para decidir ó azar cáles deles forman parte da mostra. O "sorteo" pode realizarse de diversas formas:

Elección mediante extracción

Nunha caixa introdúcense tantas bólas ou papeletas numeradas como individuos hai na poboación (N). Estes foron previamente numerados (1, 2, 3, ..., N). Escóllense ó azar tantas papeletas como individuos debe ter a mostra (n). Esta operación pode realizarse de dúas formas distintas:

— **Sen volta:** elíxense simultaneamente, ou ben unha a unha, as n papeletas.

— **Con volta:** elíxense unha a unha n papeletas pero, despois de cada extracción, a papeleta elixida (e anotada) devólvese á caixa.

Con ambos os dous métodos conséguese unha mostra aleatoria, pois todos os elementos da poboación teñen, *a priori,* a mesma probabilidade de ser elixidos. Sen embargo, se a elección se realiza con volta, poderiamos obter algún individuo repetido que habería que desbotar e realizar outra extracción. Por iso, *cando se procede por extracción, debe realizarse sen volta,* pois, ademais, o proceso é máis cómodo.

NOMENCLATURA

O sorteo por extracción denomínase, tamén, sorteo **por insaculación**

(Literalmente, insaculación significa *meter nun saco.* Alude ó procedemento de introducir os números ou papeletas nunha bolsa ou caixa).

Obtención de números aleatorios

A calculadora ten unha tecla [RAN#], que se chama **xeradora de números aleatorios,** coa cal se obtén ó azar un número decimal comprendido entre 0,000 e 0,999. Por exemplo:

[RAN#] [0.703]

[RAN#] [0.126]

NOTA

[RAN#] **random** é *aleatorio* en inglés.

Non confundas esta tecla con [RND] (redondeo).

Se multiplicamos un destes números por N (número de elementos da poboación), obtemos un número decimal coa parte enteira comprendida entre 0 e $N - 1$. Por tanto, se tomamos a parte enteira do número obtido mediante a secuencia N [×] [RAN#] [+] 1 [=], obtemos un número elixido ó azar entre 1 e N.

Por exemplo, para $N = 45$:

[RAN#] [0.703] [×] 45 [+] 1 [=] [32.635] → 32

[RAN#] [0.126] [×] 45 [+] 1 [=] [6.67] → 6

Obtivemos, así, dous números (32 e 6) elixidos ó azar entre 1 e 45.

EXERCICIOS PROPOSTOS

1. Obtén aleatoriamente catro números enteiros comprendidos entre 1 e 95.

2. Obtén cinco números enteiros elixidos aleatoriamente entre 1 e 800.

Obtención dunha mostra mediante números aleatorios

Se repetimos n veces o proceso descrito ó final da páxina anterior, obteremos unha mostra de n elementos similar á que se obtería mediante extracción con volta. Teriamos que completar a operación suprimindo os elementos repetidos (que acaso haxa) e obtendo novos elementos que os substitúan.

Para poboacións numerosas este é, evidentemente, o método máis cómodo, pois non hai que andar preparando papeletas en grandes cantidades. Se a poboación tivera máis de 1 000 elementos, habería que obter os números aleatorios con ordenador, pois necesitaremos que teñan máis de tres cifras decimais para poder "separar" todos os elementos da poboación.

Resumo:

— Se se obtén a mostra por **insaculación** (extracción de papeletas) debe realizarse sen volta.

— Se se recorre ós números aleatorios, conséguese unha mostra como se fora con volta. Pode haber elementos repetidos que deben suprimirse e ser substituídos por outros.

EXERCICIOS RESOLTOS

1. *Dunha poboación de 423 individuos, queremos extraer unha mostra de tamaño 5.*

Describe o proceso para obtela mediante números aleatorios.

Para multiplicar por 423 calquera número que apareza en pantalla, procedemos así:

423 ⊗ ⊗ (factor constante)

Agora recorremos ós números aleatorios:

	NÚMERO ALEATORIO				NÚMERO ELIXIDO
RAN#	0.678	=	286.794	→	287
RAN#	0.308	=	130.284	→	131
RAN#	0.292	=	123.516	→	124
RAN#	0.177	=	74.871	→	75
RAN#	0.603	=	255.069	→	256

Os individuos con esta numeración son os que forman a mostra.

EXERCICIOS PROPOSTOS

3. Dunha poboación de $N = 856$ elementos, desexamos extraer unha mostra de tamaño $n = 10$.

Mediante o uso de números aleatorios, designa cáles son os 10 individuos que compoñen a mostra.

4. Dunha poboación de 543 individuos, queremos extraer unha mostra de tamaño 40 mediante números aleatorios.

Obtén os cinco primeiros elementos de dita mostra.

EXERCICIOS E PROBLEMAS RESOLTOS

1. Poboación. Mostra

En cada un dos casos que se mencionan, o colectivo estudiado ¿é poboación ou é mostra?

a) Un fabricante de parafusos, para realizar un control de calidade, recolle un de cada 100 parafusos fabricados nun día e sométeos a diversas probas.

A poboación é o conxunto de todos os parafusos fabricados pola máquina. O fabricante escolle unha mostra.

b) Nuns grandes almacéns, para indagar sobre a eficacia dunha dependenta recentemente contratada, pregúntaselles a todos os clientes atendidos por ela durante o seu primeiro día de traballo.

Posto que se lles pregunta a todos os clientes atendidos por ela, é unha poboación.

c) Noutros grandes almacéns, para indagar sobre a eficacia dos dependentes, pregúntaselles a todos os clientes que saen por unha das portas durante un día.

É unha mostra: trátase dunha parte dos clientes atendidos ese día.

d) Nunhas eleccións locais escrútanse as papeletas.

Nas eleccións, referendos, etc., sempre se recorre á totalidade da poboación (individuos censados e con dereito a voto por ter a idade mínima esixida).

2. Por qué se recorre a unha mostra

Explicar por qué, en cada un dos seguintes casos, é imprescindible —ou case imprescindible— recorrer a unha mostra.

a) Nun almacén hai 4 200 vasos de vidro. Quérese estudiar a súa resistencia á rotura. Para facelo, sométense a presións crecentes ata que parten.

Posto que o proceso de medición é destructivo, é imprescindible recorrer a unha mostra e, ademais, tan pequena como sexa posible (pero procurando que se poidan extraer do seu estudio conclusións fiables).

b) Para estudiar o tempo de reacción de certas substancias, fanse reaccionar en 25 ocasións, tomando medidas en cada unha delas.

Nesta, como noutras experiencias, suponse que se se controlan todas as variables (cantidades das substancias que interveñen na reacción, pureza das mesmas, presión, temperatura...), o resultado sería sempre o mesmo. Sen embargo, o control das variables non é perfecto. Por tanto, cada experimento pode dar lugar a un resultado distinto (supostamente serán moi parecidos a outros). A poboación é infinita, pois componse de todos os experimentos que se poderían realizar. Naturalmente, temos que recorrer a unha mostra.

c) O profesor, para ver se as súas explicacións foron entendidas polos seus alumnos, realiza varias preguntas entre eles.

As preguntas que realiza o profesor na clase son unha mostra que serve para tentear o que saben os alumnos. Mesmo dos exames só se extrae unha mostra dos seus coñecementos, xa que sería imposible preguntárllelo todo.

3. Mostraxe

Dispoñemos do censo electoral dunha poboación. Consta de 27 800 electores. Desexamos extraer unha mostra de 200 individuos.

a) ¿Como se debe realizar mediante mostraxe aleatoria sistemática?

b) ¿Como se debe realizar mediante mostraxe aleatoria simple?

Utiliza a función RAN# da calculadora.

a) **Mostraxe sistemática**

Coeficiente de elevación: h = 27 800/200 = 139

Isto significa que temos que seleccionar un individuo de cada 139. Para saber por cál comezamos, eliximos ó azar un número de 1 a 139. Pode realizarse mediante a función RAN# dunha calculadora:

Obtención do primeiro elemento:

RAN# ⊗ 139 ⊕ 1 ⊜, e quedamos coa parte enteira.

Por exemplo: RAN# 0.534 ⊗ 139 ⊕ 1 ⊜ 75.782

O primeiro elemento será o 75 da lista.

E os seguintes serán: 75 + 139 = 214, 214 + 139 = 353, ...

b) **Mostraxe aleatoria simple**

A secuencia RAN# ⊗ 27 800 ⊕ 1 ⊜ proporciónanos un individuo ó azar do colectivo inicial. Esta secuencia hai que repetila 200 veces para seleccionar os elementos da mostra. Se aparece algún número repetido, suprímese e obtense outro no seu lugar.

Os números aleatorios habería que obtelos con ordenador, para que teñan, ó menos, *cinco* cifras decimais.

4. Mostraxe estratificada

Da poboación anterior sabemos que o 20% teñen entre 18 e 25 anos; o 35% entre 26 e 40 e o 45%, máis de 40.

¿Como se extraería unha mostra de 200 individuos con estratos proporcionais a esas porcentaxes?

$$\left.\begin{array}{l} 20\% \text{ de } 200 = 40 \\ 35\% \text{ de } 200 = 70 \\ 45\% \text{ de } 200 = 90 \end{array}\right\}$$

Elixiranse ó azar 40 individuos de entre os que teñen de 18 a 25 anos; 70 de 26 a 40 anos e 90 de máis de 40 anos.

Para iso, no censo deberán figurar as idades dos individuos.

5. Mostraxe estratificada

Os empregados dunha empresa están clasificados como indica a táboa:

CATEGORÍA	A	B	C	D
Nº EMPREGADOS	400	300	200	100

Para facer unha consulta sobre a modificación do horario laboral, elíxense por sorteo 50 empregados da categoría A, 40 da B, 30 da C e 20 da D.

¿É este un modelo de mostraxe aleatoria estratificada?

¿É proporcional? ¿Por que?

É unha mostraxe aleatoria estratificada.

Non é proporcional, porque

$$\frac{50}{400} \neq \frac{40}{300} \neq \frac{30}{200} \neq \frac{20}{100}$$

Se quixeramos que fora proporcional, cunha mostra de

50 + 40 + 30 + 20 = 140 individuos,

fariamos:

$$\left.\begin{array}{l} 400 + 300 + 200 + 100 = 1\,000; \\ 140 \div 1\,000 = \times \times\ 400 = \ 300 = \ 200 = \ 100 = \\ \qquad\qquad\quad \downarrow \qquad\ \downarrow \qquad\ \downarrow \qquad\ \downarrow \\ \qquad\qquad\quad 56 \qquad 42 \qquad 28 \qquad 14 \end{array}\right\} \begin{array}{l} 56 \text{ de A} \\ 42 \text{ de B} \\ 28 \text{ de C} \\ 14 \text{ de D} \end{array}$$

PARA PRACTICAR

1 En cada un dos casos que se mencionan a continuación, o colectivo ¿é poboación ou é mostra?

Explica por qué.

a) Un campesiño ten 87 galiñas. Para probar a eficacia dun novo tipo de alimentación que acaba de aparecer, pésaas todas antes e despois dos 30 días que dura o tratamento.

b) Un granxeiro proba con 100 das súas galiñas a eficacia dun novo tipo de alimentación que saíu ó mercado.

2 Para preparar un folleto publicitario, un fabricante de elásticos quere estudiar a súa resistencia á rotura. Para iso, estíraos ata que rompen e anota o grao de estiramento que acadan sen romper.

¿Pode realizar dito estiramento sobre a poboación ou é imprescindible realizalo sobre a mostra? ¿Por que?

3 Só un dos seguintes procedementos nos permite obter unha mostra representativa. Di cál é e, nos outros, estudia o sentido do nesgo e a súa importancia:

a) Para estudiar as frecuencias relativas das letras, tómanse ó azar 20 libros da biblioteca dun centro escolar e cóntanse as veces que aparece cada letra na páxina 20 dos libros seleccionados.

b) Para coñecer a opinión dos seus clientes sobre o servicio ofrecido por uns grandes almacéns de certa cidade, selecciónanse ó azar, entre os que posúen tarxeta de compra, 100 persoas entre as que gastaron menos de 1 000 € o último ano, outras 100 entre as que gastaron entre 1 000 € e 5 000 €, e 100 máis entre as que gastaron máis de 5 000 €.

c) Para calcular o número medio de persoas que están adscritas a cada cartilla nun Centro de Saúde da Seguridade Social, os médicos toman nota de todas as cartillas das persoas que acoden ás consultas durante un mes.

4 Dun colectivo de 500 persoas elixe unha mostra de 20 mediante:

a) Unha mostraxe aleatoria sistemática.

b) Unha mostraxe aleatoria simple.

Utiliza a tecla RAN# da calculadora.

5 Nun conxunto de 1 000 conductores hai:

— 50 taxistas.

— 75 camioneiros.

— 25 conductores de autobús.

O resto son conductores de vehículos correntes e repártense así:

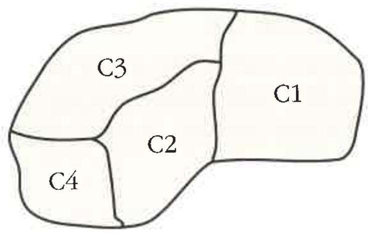
— 250 con máis de 20 anos de experiencia.

— 425 cunha experiencia de entre 5 e 20 anos.

— 175 cunha experiencia de 0 a 5 anos.

Para confeccionar unha mostra de 40 individuos mediante mostraxe aleatoria estratificada proporcional, ¿cantos hai que seleccionar de cada un dos seis estratos?

6 S En determinada provincia hai catro comarcas,



cun total de 1 500 000 persoas censadas. Delas, 300 000 residen en C1, 450 000 en C2 e 550 000 en C3. Quérese realizar un estudio sobre os costumes alimentarios nesa provincia baseado nunha mostra de 3 000 persoas.

a) ¿Que tipo de mostraxe deberiamos realizar se queremos que na mostra resultante haxa representación de todas as comarcas?

b) ¿Que número de persoas habería que seleccionar en cada comarca, atendendo a razóns de proporcionalidade?

c) ¿Como seleccionarías as persoas en cada comarca?

Xustifica as respostas.

7 Nun centro de ensino con 981 alumnos e alumnas, vaise facer unha sondaxe sobre tendencias políticas. Vaise escoller unha mostra de 84 estudiantes. No centro hai 5 cursos (1º, 2º, 3º, 4º e 5º) cun número de alumnos e alumnas en cada un deles de 345, 234, 190, 140 e 72. ¿Cantos alumnos deberemos escoller de cada curso se desexamos que a mostraxe sexa estratificada con reparto proporcional?

8 Queremos seleccionar unha mostra de 50 alumnos de 2º Bacharelato. En cada un dos seguintes casos debes decidir se a mostraxe debe ser aleatoria simple ou estratificada por sexos (rapaces-rapazas) para estudiar as variables indicadas:

a) Estatura.

b) Tempo que empregan os alumnos en ir da súa casa ó instituto.

c) Agudeza visual (porcentaxe de alumnado con lentes).

d) Incidencia de carie dental.

e) Práctica de fútbol.

f) Lectura dalgún xornal.

PARA PROFUNDAR

9 Se contas o número de persoas e o número de cans que viven no teu portal e todos os compañeiros e compañeiras facedes o mesmo, obteredes unha mostra coa que podedes estimar o número de cans que hai na vosa poboación.

a) ¿Como é de fiable esta estimación?

b) ¿É aleatoria a mostra que utilizaches?

c) ¿Ocórreche un procedemento mellor para seleccionar a mostra?

10 Para facer un estudio sobre os hábitos ecolóxicos das familias dunha cidade, seleccionáronse, por sorteo, as direccións, rúa e número que serán visitadas. Se nun portal vive máis dunha familia, sortearase entre elas a que será seleccionada.

¿Obteremos con este procedemento unha mostra aleatoria?

(Pensa se ten a mesma probabilidade de ser incluída na mostra unha familia que vive nunha vivenda unifamiliar, que outra que vive nun bloque de 32 vivendas).

11 A validez da información que nos proporciona unha enquisa depende, en gran medida, da coidadosa elaboración do cuestionario.

Algunhas das características que deben ter as preguntas, son:

— Ser curtas e cunha linguaxe sinxela.

— As súas respostas deben presentar opcións non ambiguas e equilibradas.

— Que non requiran esforzo de memoria.

— Que non levanten prexuízos nos enquisados.

Estudia se as seguintes preguntas son adecuadas para formar parte dunha enquisa e corrixe os erros que observes:

a) ¿Canto tempo adoitas estudiar cada día?

Moito ☐ Pouco ☐ Segundo o día ☐

b) ¿Cantas veces fuches ó cine ó longo do ano pasado?

c) ¿Que opinión tes sobre a xestión do alcalde?

Moi boa ☐ Boa ☐ Indiferente ☐

d) ¿Perden os seus fillos o tempo vendo a televisión?

Si ☐ Non ☐

e) ¿En que grao cre vostede que a instalación da planta de reciclaxe lle afectaría ó emprego e ás condicións de saúde da nosa cidade?

12 Quérense realizar os seguintes estudios:

a) Tipo de transporte que utilizan os veciños dun barrio para acudir ó seu traballo.

b) Estudios que pensan realizar os alumnos e alumnas dun centro escolar ó terminar o Bacharelato.

c) Idade das persoas que viron unha obra de teatro nunha cidade.

d) Número de horas diarias que ven a televisión os nenos e nenas españois con idades comprendidas entre 5 e 10 anos.

Di en cada un destes casos cál é a poboación.

¿En cales deles é necesario recorrer a unha mostra? ¿Por que?

12 INFERENCIA ESTATÍSTICA. ESTIMACIÓN DA MEDIA

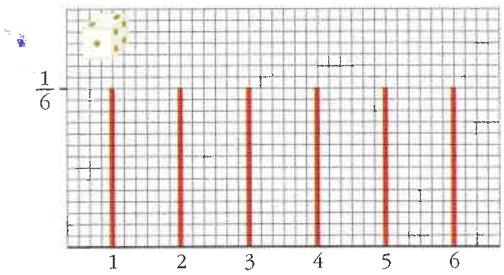
Un dos problemas máis sinxelos da estatística inductiva é o de estimar o valor da media dunha poboación a partir dunha mostra. A estimación realízase de forma aproximada (mediante un intervalo) e cunha certa inseguridade (asignando un "nivel de confianza" ó resultado). O tamaño da mostra inflúe na finura da estimación. Para realizar este proceso bótase man da curva normal.

A distribución normal, como sabes, foi descrita por primeira vez por **De Moivre** (en 1733) e redescuberta por Gauss máis de medio século despois. O propio Gauss xustificou a súa omnipresencia mediante uns razoamentos que son o antecedente do teorema central do límite que estudiaremos nesta unidade.

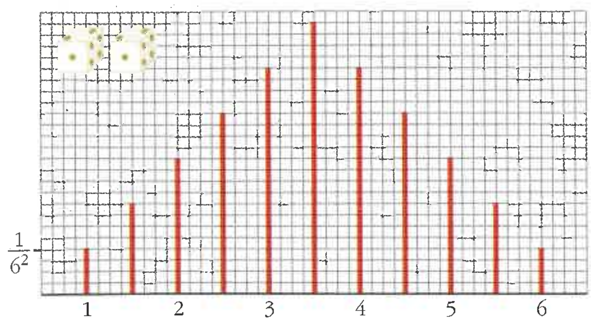
PARA EMPEZAR, REFLEXIONA E RESOLVE

Lanzamento de varios dados

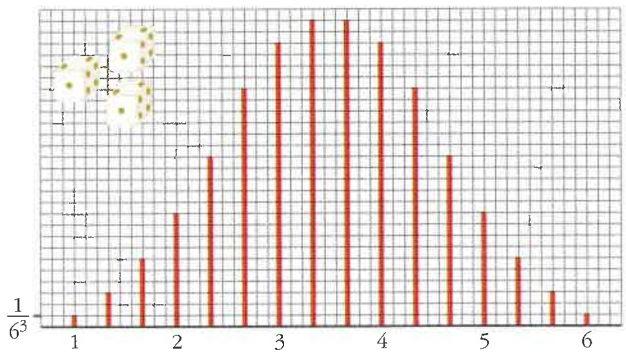
A distribución da probabilidade correspondente ó lanzamento dun dado correcto é:



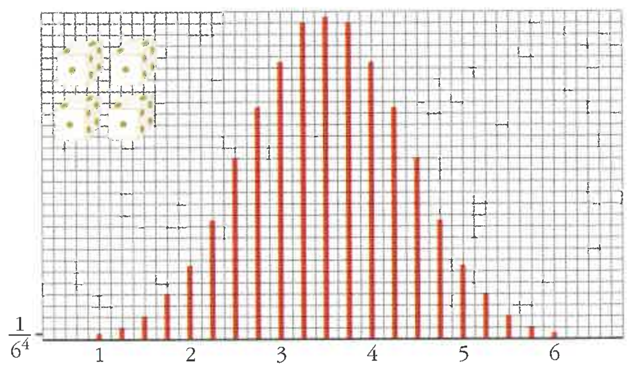
Se lanzamos dous dados e calculamos a media dos seus resultados (por exemplo, un 3 e un 5 fan unha media de 4), a distribución de probabilidades é:



A media dos resultados de tres dados distribúese así:



E a de catro dados:



Imos comparar a media e a desviación típica destas catro distribucións:

	MEDIA	DESVIACIÓN TÍPICA
UN DADO	3,5	1,71
DOUS DADOS (media)	3,5	1,21
TRES DADOS (media)	3,5	0,98
CATRO DADOS (media)	3,5	0,86

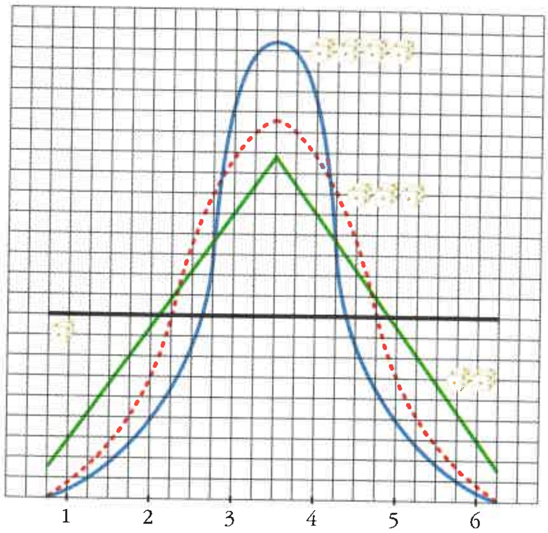
- É evidente, á vista da táboa anterior, que as catro medias son iguais mentres que a desviación típica é tanto menor canto máis dados participan.

Comproba na táboa anterior que a desviación típica para n dados é igual a:

(Desviación típica para un dado) / $\sqrt{n}$

para $n = 2$, $n = 3$ e $n = 4$.

Na gráfica da dereita podes ver cómo comparamos as catro gráficas representándoas xuntas, para un, dous, tres e catro dados.



- Mirando as gráficas, xustifica as seguintes afirmacións:

— Cantos máis dados interveñen, máis se parece a distribución das súas medias á curva normal.

— Todas as distribucións teñen a mesma media.

— Cantos máis dados interveñen, menor desviación típica ten a distribución.

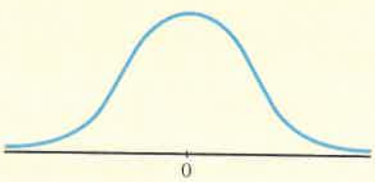
NESTA UNIDADE VERÁS

Os resultados vistos nesta páxina para un, dous, tres ou catro dados describen, de maneira simplificada, o **Teorema Central do Límite,** básico para relacionar a media dunha poboación coa media dunha mostra extraída dela.

- Unha ferramenta básica para a inferencia estatística é a **distribución normal**.

Comezamos a unidade repasando brevemente:

— o cálculo de probabilidades nunha distribución normal $N(0, 1)$,



— o cálculo de probabilidades nunha distribución normal calquera, $N(\mu, \sigma)$,

e sistematizamos a obtención de intervalos centrados na media e que conteñan unha certa proporción de individuos (**intervalos característicos**).

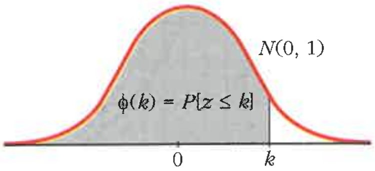
- O **Teorema Central do Límite** permite, coa axuda da distribución normal, describir a **distribución das medias das mostras** dun certo tamaño en función dos parámetros da poboación.

- Os intervalos característicos correspondentes á distribución das medias mostrais permiten, con certa sinxeleza, construír outros intervalos cos que se **estima** o valor da media da poboación a partir da media dunha certa mostra. Son os **intervalos de confianza;** a súa obtención é o obxectivo principal desta unidade.

12.1 DISTRIBUCIÓN NORMAL. REPASO DE TÉCNICAS BÁSICAS

Como vimos nas dúas páxinas coas que se inicia esta unidade e como se ratificará nos próximos apartados, ó tratar coas medias mostrais aparece con moita frecuencia a distribución normal. Imos realizar un breve repaso do cálculo de probabilidades en distribucións normais.

Utilización da táboa da normal $N(0, 1)$



Na distribución $N(0, 1)$, a variable adoita representarse pola letra z. As táboas que aparecen a continuación dannos as probabilidades $P[z \leq k]$ para valores de k de 0 a 4, de centésima en centésima. A estas probabilidades chámaselles $\phi(k)$:

$$\phi(k) = P[z \leq k], \quad z \text{ distribúese } N(0, 1)$$

UNIDADES E DÉCIMAS	CENTÉSIMAS									
k	0	1	2	3	4	5	6	7	8	9
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5754
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6065	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7258	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7518	0,7549
0,7	0,7580	0,7612	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7996	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998
3,5	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998
3,6	0,9998	0,9998	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,7	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,8	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

O valor de k búscase así:

Unidades e **décimas** na columna da esquerda.

Centésimas na fila de arriba.

O número que nos dá a táboa é o valor de:

$$\phi(k) = P[z \leq k]$$

- Exemplos

$P[z \leq 0{,}83] = \phi(0{,}83) = 0{,}7967$

$P[z \leq 2{,}3] = \phi(2{,}30) = 0{,}9893$

$P[z \leq 1] = \phi(1{,}00) = 0{,}8413$

Reciprocamente, se coñecemos o valor da probabilidade $\phi(k)$, pódese saber o valor de k.

- Exemplos

$P[z \leq k] = \phi(k) = 0{,}7190 \rightarrow$

$\rightarrow k = 0{,}58$

$P[z \leq k] = \phi(k) = 0{,}8643 \rightarrow$

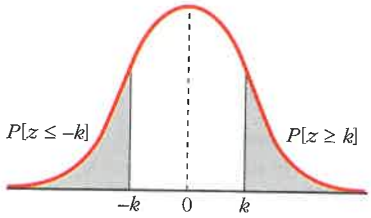
$\rightarrow k = 1{,}1$

$P[z \leq k] = \phi(k) = 0{,}5560 \rightarrow$

$\rightarrow k \approx 0{,}14$

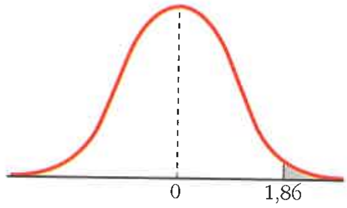
Lembremos que nunha distribución de variable continua as probabilidades puntuais son nulas: $P[x = k] = 0$. Por tanto, $P[x \leq k] = P[x < k]$.

Cálculo de probabilidades nunha distribución $N(0, 1)$

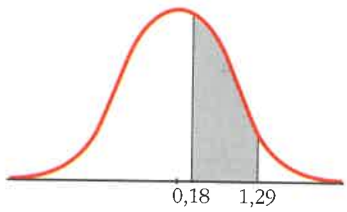


- Se $k \geq 0$, as probabilidades $P[z \leq k] = P[z < k] = \phi(k)$ encóntranse directamente nas táboas.
- $P[z \geq k] = 1 - P[z < k] = 1 - \phi(k)$
- Para abscisas negativas, $P[z \leq -k] = P[z \geq k] = 1 - \phi(k)$.

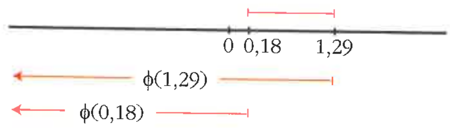
As restantes probabilidades poden obterse a partir destas como se ve nos seguintes exemplos:

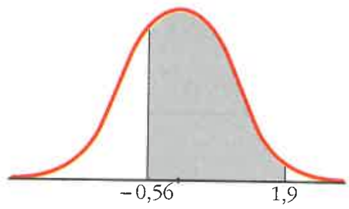


1. $P[z \geq 1,86] = 1 - P[z < 1,86] = 1 - \phi(1,86) = 1 - 0,9686 = 0,0314$



2. $P[0,18 < z \leq 1,29] = \phi(1,29) - \phi(0,18) = 0,9015 - 0,5714 = 0,3301$

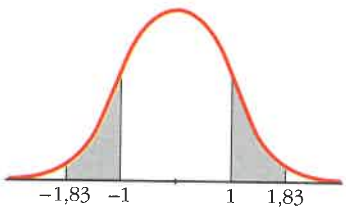




3. $P[-0,56 < z \leq 1,9] = P[z \leq 1,9] - P[z \leq -0,56]$

$P[z \leq -0,56] = P[z \geq 0,56] = 1 - \phi(0,56) = 1 - 0,7123 = 0,2877$

$P[-0,56 < z \leq 1,9] = \phi(1,90) - 0,2877 = 0,9713 - 0,2877 = 0,6836$



4. $P[-1,83 < z < -1] = P[1 < z < 1,83] = \phi(1,83) - \phi(1,00) =$

$= 0,9664 - 0,8413 = 0,1251$

EXERCICIOS PROPOSTOS

1. Calcula as seguintes probabilidades nunha distribución $N(0, 1)$:

a) $P[z > 2,8]$ b) $P[z \leq -1,8]$ c) $P[z > -1,8]$

d) $P[1,62 \leq z < 2,3]$ e) $P[1 \leq z \leq 2]$

f) $P[-0,61 \leq z \leq 1,4]$ g) $P[-1 \leq z \leq 2]$

h) $P[-2,3 < z < -1,7]$ i) $P[-2 \leq z \leq -1]$

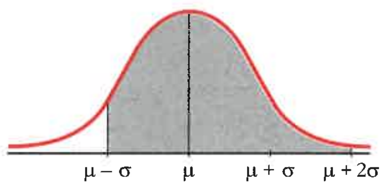
2. Calcula o valor de k (exacta ou aproximadamente) en cada un dos seguintes casos:

a) $P[z \leq k] = 0,5$ b) $P[z \leq k] = 0,8729$

c) $P[z \leq k] = 0,9$ d) $P[z \leq k] = 0,33$

e) $P[z \leq k] = 0,2$ f) $P[z > k] = 0,12$

g) $P[z \geq k] = 0,9971$ h) $P[z \geq k] = 0,6$



Cálculo de probabilidades nunha distribución $N(\mu, \sigma)$

En esencia, todas as distribucións normais son a mesma. Por iso, para calcular a probabilidade de que unha variable x, normal $N(\mu, \sigma)$, estea comprendida entre certos valores só importa saber "cántas desviacións típicas se separan da media" cada un deses valores.

Na gráfica da marxe, o intervalo sinalado caracterízase así: "desde unha desviación típica á esquerda da media ata dúas desviacións típicas á dereita da media".

Isto pode dicirse máis comodamente así: "correspóndelle o intervalo (–1, 2) na distribución $N(0, 1)$".

$$\mu - \sigma < x < \mu + 2\sigma \iff -1 < z < 2$$

Toda distribución $N(\mu, \sigma)$ é unha distribución $N(0, 1)$ se a variable a expresamos en "número de desviacións típicas que se separa da media".

$$\begin{array}{rcl}
x \text{ é } N(\mu, \sigma) & \leftrightarrow & z \text{ es } N(0, 1) \\
a & \leftrightarrow & \dfrac{a-\mu}{\sigma} \\
b & \leftrightarrow & \dfrac{b-\mu}{\sigma} \\
[a, b] & \leftrightarrow & \left[\dfrac{a-\mu}{\sigma}, \dfrac{b-\mu}{\sigma}\right] \\
a < x < b & \leftrightarrow & \dfrac{a-\mu}{\sigma} < z < \dfrac{b-\mu}{\sigma}
\end{array}$$

NOMENCLATURA

A este proceso, pasar dunha variable x $N(\mu, \sigma)$ a unha variable z $N(0, 1)$, chámaselle **tipificación da variable.** As probabilidades da variable tipificada poden calcularse mirando nas táboas.

EXERCICIOS RESOLTOS

1. *Nunha distribución $N(66, 8)$, calcula:*

a) $P[x < 70]$

b) $P[x > 80]$

c) $P[70 < x < 80]$

$$\begin{array}{rcl}
x \text{ é } N(66, 8) & \leftrightarrow & z \text{ é } N(0, 1) \\
70 & \leftrightarrow & \dfrac{70-66}{8} = 0{,}5 \\
80 & \leftrightarrow & \dfrac{80-66}{8} = 1{,}75
\end{array}$$

a) $P[x < 70] \leftrightarrow P[z < 0{,}5] = 0{,}6915$

b) $P[x < 80] \leftrightarrow P[z < 1{,}75] = 0{,}9599$, $P[z > 1{,}75] = 0{,}0401$

c) $P[70 < x < 80] \leftrightarrow P[0{,}5 < z < 1{,}75] = 0{,}9599 - 0{,}6915 = 0{,}2684$

Solucións:

a) $P[x < 70] = 0{,}6915$; b) $P[x < 80] = 0{,}0401$; c) $P[70 < x < 80] = 0{,}2684$

EXERCICIOS PROPOSTOS

3. Nunha distribución $N(18, 4)$, calcula as seguintes probabilidades:

a) $P[x \le 20]$ b) $P[x \ge 16{,}5]$ c) $P[x \le 11]$

d) $P[19 \le x \le 23]$ e) $P[11 \le x < 25]$

4. Nunha distribución $N(6; 0{,}9)$, calcula k para que se dean as seguintes igualdades:

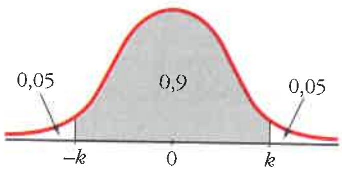
a) $P[x \le k] = 0{,}9772$ b) $P[x \le k] = 0{,}8$

c) $P[x \le k] = 0{,}\dot{3}$ d) $P[x \ge k] = 0{,}6331$

12.2 INTERVALOS CARACTERÍSTICOS

Se a variable x ten unha distribución de media μ, chámase **intervalo característico** correspondente a unha probabilidade p a un intervalo centrado na media, $(\mu - k,\ \mu + k)$, tal que a probabilidade de que x pertenza a dito intervalo é p:

$$P[\mu - k < x < \mu + k] = p$$



Determinemos, por exemplo, o intervalo característico dunha distribución normal $N(0, 1)$ correspondente á probabilidade $p = 0{,}9$.

Se dentro do intervalo hai unha área de 0,9, fóra del haberá 0,1. Posto que o intervalo é simétrico, a área de cada unha das dúas *colas* é de 0,05. Por tanto:

$$P[z > k] = 0{,}05 \ \rightarrow\ P[z \leq k] = 0{,}95$$

Nas táboas encontramos $\phi(1{,}64) = 0{,}9495$, $\phi(1{,}65) = 0{,}9505$.

Por tanto, asignarémoslle a k o punto medio dos valores 1,64 e 1,65. É dicir, $k = 1{,}645$ e, por tanto, $P[-1{,}645 < z < 1{,}645] = 0{,}9$.

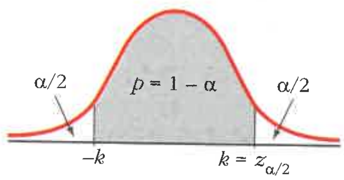
Encontramos un intervalo [–1,645; 1,645], simétrico respecto ó 0, dentro do cal hai unha área do 90% do total.

Intervalos característicos en distribucións $N(0, 1)$

En adiante imos usar con moita frecuencia intervalos característicos de distribucións normais e, en particular, da $N(0, 1)$.

Nunha distribución normal $N(0, 1)$, se $(-k,\ k)$ é o intervalo característico correspondente a unha probabilidade p, é dicir, se

$$P[-k < z < k] = p$$



diremos que k é o **valor crítico** correspondente a p.

Os valores críticos correspondentes ás probabilidades 0,9; 0,95; 0,99 utilízanse moito e son, respectivamente, 1,645; 1,96; 2,575.

Habitualmente desígnase a probabilidade p mediante $1 - \alpha$. O valor crítico correspondente denomínase $z_{\alpha/2}$ e téñense as igualdades:

$$P[z > z_{\alpha/2}] = \alpha/2 \quad ; \quad P[-z_{\alpha/2} < z < z_{\alpha/2}] = 1 - \alpha$$

MOI IMPORTANTE

PRINCIPAIS VALORES CRÍTICOS

$1 - \alpha$	$\alpha/2$	$z_{\alpha/2}$
0,90	0,05	1,645
0,95	0,025	1,96
0,99	0,005	2,575

EXERCICIOS PROPOSTOS

1. Calcula razoadamente os valores críticos correspondentes ás probabilidades 0,95 e 0,99.

2. Calcula os valores críticos correspondentes:

a) $\alpha = 0{,}09$ b) $\alpha = 0{,}21$ c) $\alpha = 0{,}002$

Intervalos característicos en distribucións normais calquera

Sexa x unha variable $N(\mu, \sigma)$. Desexamos encontrar un intervalo centrado na media, $(\mu - k, \mu + k)$ tal que $P[x \in (\mu - k, \mu + k)] = p = 1 - \alpha$. É dicir, un intervalo no que estea o $100 \cdot (1 - \alpha)$% dos individuos da poboación.

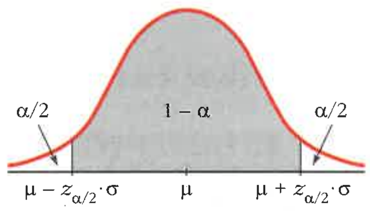
Se x é $N(\mu, \sigma)$, entón $z = \dfrac{x - \mu}{\sigma}$ é $N(0, 1)$.

Tendo en conta os resultados da páxina anterior, o intervalo característico de $z = \dfrac{x - \mu}{\sigma}$ correspondente a $p = 1 - \alpha$ é $(-z_{\alpha/2}, z_{\alpha/2})$.

É dicir, $-z_{\alpha/2} < \dfrac{x - \mu}{\sigma} < z_{\alpha/2}$ cunha probabilidade $1 - \alpha$.

Isto equivale a: $\begin{cases} -z_{\alpha/2} \cdot \sigma < x - \mu < z_{\alpha/2} \cdot \sigma \\ \text{ou ben} \\ \mu - z_{\alpha/2} \cdot \sigma < x < \mu + z_{\alpha/2} \cdot \sigma \end{cases}$

Obtense, así, o intervalo $(\mu - z_{\alpha/2} \cdot \sigma, \ \mu + z_{\alpha/2} \cdot \sigma)$.



Nunha distribución $N(\mu, \sigma)$ o intervalo característico correspondente a unha probabilidade $p = 1 - \alpha$ é:

$$(\mu - z_{\alpha/2} \cdot \sigma, \ \mu + z_{\alpha/2} \cdot \sigma)$$

Por tanto:

Para o 90%, $1 - \alpha = 0{,}9 \rightarrow z_{\alpha/2} = 1{,}645 \rightarrow (\mu - 1{,}645\,\sigma, \ \mu + 1{,}645\,\sigma)$

Para o 95%, $1 - \alpha = 0{,}95 \rightarrow z_{\alpha/2} = 1{,}96 \rightarrow (\mu - 1{,}96\,\sigma, \ \mu + 1{,}96\,\sigma)$

Para 99%, $1 - \alpha = 0{,}99 \rightarrow z_{\alpha/2} = 2{,}575 \rightarrow (\mu - 2{,}575\,\sigma, \ \mu + 2{,}575\,\sigma)$

EXERCICIOS RESOLTOS

1. *Nunha distribución $N(66, 8)$, obtén os intervalos característicos para o 90%, para o 95% e para o 99%.*

90%: $(66 - 1{,}645 \cdot 8; \ 66 + 1{,}645 \cdot 8) = (52{,}84; \ 79{,}16)$

Isto significa que o 90% dos individuos está neste intervalo.

É dicir, que $P[x \in (52{,}84; \ 79{,}16)] = 0{,}9$

95%: $(66 - 1{,}96 \cdot 8, \ 66 + 1{,}96 \cdot 8) = (50{,}32; \ 81{,}68) \rightarrow$

$\rightarrow P[x \in (50{,}32; \ 81{,}68)] = 0{,}95$

99%: $(66 - 2{,}575 \cdot 8, \ 66 + 2{,}575 \cdot 8) = (45{,}4; \ 86{,}6) \rightarrow$

$\rightarrow P[x \in (45{,}4; \ 86{,}6)] = 0{,}99$

EXERCICIOS PROPOSTOS

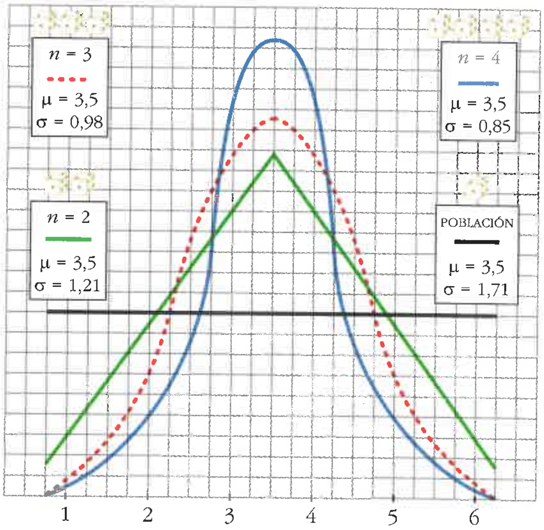
3. Nunha distribución $N(173, 6)$ obtén os intervalos característicos para o 90%, o 95% e o 99%.

4. Nunha distribución $N(18, 4)$ calcula os intervalos característicos para o 95% e o 99,8%.

12.3 DISTRIBUCIÓN DAS MEDIAS MOSTRAIS

Recordemos os resultados que se obtiveron nas dúas primeiras páxinas desta unidade, nas que nos interesabamos polas medias das puntuacións obtidas ó lanzar dous, tres ou catro dados.

O resultado do lanzamento dun dado pode considerarse un individuo dunha poboación infinita: *lanzar un dado indefinidamente.* Lanzar un dado catro veces (ou lanzar catro dados) pode ser considerado como unha mostra de tamaño catro desa poboación.



Segundo ese punto de vista, a experiencia que describimos nas páxinas iniciais da unidade pode resumirse así:

Se da distribución "resultado obtido ó lanzar un dado" extraemos mostras de tamaños $n = 2$, $n = 3$, $n = 4$, ... a distribución das súas correspondentes medias parécese a unha distribución normal tanto máis canto maior sexa n. Ademais, ó aumentar n (o número de dados):

— A media mantense constante.

— A desviación típica vai diminuíndo.

Este resultado relativo ó lanzamento dun dado xeneralízase para calquera distribución segundo o seguinte teorema.

Teorema Central do Límite

Dada unha poboación de media μ e desviación típica σ, non necesariamente normal, **a distribución das medias das mostras** de tamaño n:

- Ten a mesma media, μ, cá poboación.
- A súa desviación típica é $\sigma/\sqrt{n}$ e, por conseguinte, diminúe ó aumentar n.
- Cando $n \geq 30$ é practicamente normal.

NON O ESQUEZAS

$x \to \mu, \sigma$

$\bar{x}$ é $N(\mu, \sigma/\sqrt{n})$ (se $n \geq 30$)

É importante sinalar que este teorema é válido *calquera que sexa a distribución da poboación de partida,* tanto se é discreta como continua.

O grao de aproximación da distribución das medias mostrais á correspondente normal depende do tipo de poboación de partida e do valor de n.

— Se a poboación de partida é normal, tamén o será a distribución das medias mostrais, calquera que sexa o valor de n.

— Aínda que a poboación de partida non sexa normal, a distribución das medias mostrais pode ser moi parecida á normal, mesmo para valores pequenos de n (como vimos nas experiencias anteriores) pero para $n > 30$ é seguro que se consegue unha grande aproximación á normal calquera que sexa a distribución de partida.

MOI IMPORTANTE

Se a poboación de partida é normal, tamén é normal a distribución das medias das mostras de tamaño n, calquera que sexa o valor de n.

Consecuencias do teorema central do límite

Vexamos algunhas das vantaxes que nos achega o teorema central do límite:

I. Control das medias mostrais

Nunha poboación de media μ e desviación típica σ, dispoñémonos a extraer unha mostra de tamaño n. Antes de facelo, sabemos que a distribución das medias, $\bar{x}$, de todas as posibles mostras é $N(\mu, \sigma/\sqrt{n})$ e, por tanto, podemos tratar de coñecer a probabilidade de que a media dunha mostra concreta estea nun certo intervalo.

II. Control da suma de todos os individuos da mostra

TEN EN CONTA

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\Sigma x_i}{n}$$

$$\Downarrow$$

$$\Sigma x_i = n\bar{x}$$

Posto que $\sum_{i=1}^{n} x_i = n\bar{x}$, sabemos que $\sum_{i=1}^{n} x_i$ distribúese normal de media $n\mu$ e desviación típica $n\dfrac{\sigma}{\sqrt{n}} = \sigma\sqrt{n}$:

$$\sum_{i=1}^{n} x_i \text{ é } N(n\mu, \sigma\sqrt{n})$$

Por tanto, podemos calcular cál é a probabilidade de que a suma dos elementos dunha mostra estea, *a priori,* nun certo intervalo.

III. Inferir a media da poboación a partir dunha mostra

Esta é a aplicación máis importante do teorema central do límite. A partir dunha mostra poden extraerse conclusións válidas sobre a media da poboación de partida. Verémolo máis adiante.

EXERCICIOS RESOLTOS

1. *As bolsas de azucre envasadas por unha certa máquina teñen $\mu = 500$ g e $\sigma = 35$ g. As bolsas empaquétanse en caixas de 100 unidades.*
   - *a) Calcula a probabilidade de que a media dos pesos das bolsas dun paquete sexa menor que 495 g.*
   - *b) Obtén o intervalo característico de $\bar{x}$ para unha probabilidade do 95%.*
   - *c) Calcula a probabilidade de que unha caixa de 100 bolsas pese máis de 51 kg.*

A totalidade das bolsas producidas pola máquina é unha poboación de media $\mu = 500$ g e $\sigma = 35$ g.

Cada caixa é unha mostra de 100 individuos.

As medias, $\bar{x}$, dos pesos das bolsas dunha caixa distribúense normal de media $\mu = 500$ g e desviación típica $\dfrac{\sigma}{\sqrt{n}} = \dfrac{35}{\sqrt{100}} = 3{,}5$

$$\bar{x} \text{ é } N(500; 3{,}5)$$

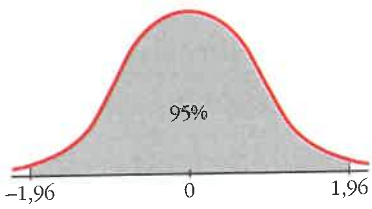
a) $P[\bar{x} < 495] = P\left[z < \dfrac{495 - 500}{3{,}5}\right] = P[z < -1{,}43] = P[z > 1{,}43] =$

$= 1 - \phi(1{,}43) = 1 - 0{,}9236 = 0{,}0764$

b) O valor crítico correspondente a $p = 0{,}95$ é 1,96.

O intervalo característico é:

$(500 \pm 1{,}96 \cdot 3{,}5) = (493{,}1; 506{,}9)$



c) $\sum_{i=1}^{100} x_i$ é $N(50\,000, 350)$. Por tanto:

$$P[\Sigma x_i > 51\,000] = P\left[z > \frac{51\,000 - 50\,000}{350}\right] = P[z > 2{,}86] =$$
$$= 1 - \phi(2{,}86) = 1 - 0{,}9979 = 0{,}0021$$

É dicir, pouco máis de 2 caixas de cada 1 000 pesarán máis de 51 kg.

2. *Os pesos en quilogramos dos soldados dunha quinta seguen unha distribución normal N(69, 8). As gardas nun rexemento están formadas por 12 soldados.*

a) Determina a probabilidade de que a media dos pesos dos soldados dunha garda sexa superior a 71 kg.

b) Obtén o intervalo característico para $\bar{x}$ correspondente a unha probabilidade de 0,9.

c) ¿Cal é a probabilidade de que a suma dos pesos dos soldados dunha garda sexa menor que 800 kg?

d) ¿Cal é a probabilidade de que un membro da garda, elixido ó azar, pese máis de 93 kg?

Supoñemos que as gardas se forman tomando 12 soldados ó azar (ou mediante un procedemento similar en esencia: orde alfabética, por exemplo). En tal caso, a media dos pesos, $\bar{x}$, é normal de media $\mu = 69$ e desviación típica $\sigma = \frac{8}{\sqrt{12}} = 2{,}31$ (aínda $n < 30$, isto é así porque a poboación é normal).

$\bar{x}$ é $N(69; 2{,}31)$

a) $P[\bar{x} > 71] = P\left[z > \frac{71 - 69}{2{,}31}\right] = P[z > 0{,}87] = 1 - \phi(0{,}87) =$

$= 1 - 0{,}8078 = 0{,}1922$

b) O valor crítico correspondente a $p = 0{,}9$ é $1{,}645$. O intervalo é:

$(69 \pm 1{,}645 \cdot 2{,}31) = (65{,}20; 72{,}79)$

Isto significa que o 90% das gardas teñen un peso medio comprendido entre 65,20 kg e 72,79 kg.

c) $12 \times 69 = 828$ $\qquad$ $12 \times 2{,}31 = 27{,}72$

Por tanto, Σx é $N(828; 27{,}2)$

$$P[\Sigma x < 800] = P\left[z < \frac{800 - 828}{27{,}72}\right] = P[z < -1{,}01] =$$
$$= 1 - \phi(1{,}01) = 1 - 0{,}8438 = 0{,}1562$$

d) Un individuo tomado ó azar de entre un grupo que tamén foi elixido ó azar é, sinxelamente, unha extracción ó azar da poboación. É dicir, x é $N(69{,}8)$. Por tanto:

$$P[x > 93] = P\left[z > \frac{93 - 69}{8}\right] = P[z > 3] = 1 - \phi(3) = 0{,}0013$$

EXERCICIOS PROPOSTOS

1. Os parámetros dunha variable son: $\mu = 16{,}4$, $\sigma = 4{,}8$. Dispoñémonos a extraer unha mostra de $n = 400$ individuos:

a) Obtén o intervalo característico para as medias mostrais correspondentes a unha probabilidade $p = 0{,}99$.

b) Calcula $P[16 < \bar{x} < 17]$.

2. Os soldos, en euros, dos empregados dunha fábrica distribúense $N(1\,200, 400)$. Elíxese ó azar unha mostra de 25 deles. ¿Cal é a probabilidade de que a suma dos seus soldos sexa superior a 35 000 €?

Calcula o intervalo característico para as sumas de 25 individuos, correspondentes a unha probabilidade de 0,9.

12.4 EN QUÉ CONSISTE A ESTATÍSTICA INFERENCIAL

Imos analizar tres situacións, parecidas pero moi distintas:

PROBLEMA 1. As estaturas dos soldados dun rexemento teñen unha media $\mu = 175$ e unha desviación típica $\sigma = 5$. ¿Cal é a probabilidade de que a estatura media dos 32 soldados que deben facer garda esta noite estea comprendida entre 174,4 e 175,6?

COÑECEMOS		PREGUNTÁMONOS
A media μ da poboación. No rexemento: $\mu = 175$	➡	A media $\bar{x}$ dunha mostra. Nunha garda: ¿$P[\bar{x} \in (174,4;\ 175,6)]$?

PROBLEMA 2. A estatura media dos 32 soldados que foron seleccionados para facer garda é $\bar{x} = 175$ cm. ¿Cal é a probabilidade de que a media, μ, de todos os soldados do rexemento estea no intervalo (174,4; 175,6)?

COÑECEMOS		PREGUNTÁMONOS
A media $\bar{x}$ dunha mostra. Nunha garda: $\bar{x} = 175$	➡	A media μ da poboación. No rexemento: ¿$P[\mu \in (174,4;\ 175,6)]$?

PROBLEMA 3. Afirman que a media das estaturas dos soldados dun rexemento é $\mu = 175$. Para comprobalo extraemos unha mostra de 32 soldados e calculamos a súa media, $\bar{x} = 175,8$. ¿É razoable admitir a hipótese de que $\mu = 175$?

COÑECEMOS		PREGUNTÁMONOS
A media $\bar{x}$ dunha mostra. Nunha garda: $\bar{x} = 175,8$	➡	¿É admisible a afirmación de que $\mu = 175$?

No **PROBLEMA 1** coñecemos a poboación. A partir de aí pretendemos **deducir o comportamento das mostras.** Isto aprendemos a facelo nos apartados anteriores, baseándonos no Teorema Central do Límite.

No **PROBLEMA 2** coñecemos unha mostra e, a partir dela, pretendemos deducir aspectos da poboación. En concreto, pretendemos **inferir** o valor da **media da poboación** a partir do coñecemento da media dunha mostra.

Este é un típico problema de estatística inferencial: **estimar** o valor dun parámetro da poboación a partir dunha mostra. Dedicarémonos a el no resto desta unidade.

No **PROBLEMA 3** temos unha afirmación, unha **hipótese:** a media da poboación é $\mu = 175$. Pero non temos garantías de que sexa certo. Para contrastalo, extraemos unha mostra e, a partir do seu resultado, $\bar{x} = 175,8$, debemos decidir se a hipótese é ou non admisible.

Este problema é outra situación de estatística inferencial: a **teoría da decisión** (contrastes de hipóteses) que trataremos na unidade 14.

Estimación puntual e estimación por intervalos

Descoñecemos os cocientes intelectuais dos alumnos dunha universidade, pero dispoñemos dos datos dunha mostra de 200 destes individuos.

Calculamos $\bar{x} = 108$ (media do cociente intelectual dos individuos da mostra).

Parece razoable *estimar* que a media μ da poboación será, aproximadamente, igual cá media da mostra, 108. Pero, ¿como de aproximadamente?

Os parámetros da poboación poden estimarse a partir dos da mostra. Así:

A **media mostral**, $\bar{x}$, serve para estimar a media poboacional, μ.

A **desviación típica mostral**, s, é un estimador da desviación típica poboacional, σ.

A **estimación puntual** (o valor de μ é aproximadamente $\bar{x}$), serve de pouco mentres descoñezamos cál é o grao de aproximación de $\bar{x}$ a μ. Por ese motivo procédese á estimación **mediante intervalos.**

OBSERVA

Canto máis pequeno sexa o intervalo de confianza, máis precisos estamos sendo na nosa estimación.

A partir dunha mostra aleatoria de tamaño n podemos **estimar** o valor dun parámetro da poboación do seguinte modo:

- Dando un intervalo dentro do cal *confiamos* que estea o parámetro. Chámase **intervalo de confianza**.
- Calculando a probabilidade de que tal cousa ocorra. A dita probabilidade chámaselle **nivel de confianza**.

Neste curso só faremos estimacións referentes á media (no que queda desta unidade) e á proporción (na unidade 13), de modo que faremos aseveracións do seguinte tipo:

— Estimamos, cun nivel de confianza do 95%, que os ingresos medios mensuais desta poboación están comprendidos entre 956 € e 1 040 €.

— O intervalo de confianza da proporción de daltónicos desta poboación é (0,023; 0,031), e isto sabémolo cun nivel de confianza do 99%.

Canto maior sexa o tamaño da mostra, maior eficacia teremos na nosa estimación. Esta eficacia manifèstase de dúas formas:

— No tamaño do intervalo (canto máis pequeno, máis precisos estamos sendo).

— No nivel de confianza (máis nivel de confianza significa máis seguridade na estimación).

Tamaño da mostra, lonxitude do intervalo e nivel de confianza son tres variables estreitamente relacionadas que manexaremos continuamente ó longo da unidade.

En cada caso deberemos obter unha delas despois de fixar as outras dúas.

12.5 INTERVALO DE CONFIANZA PARA A MEDIA

Deséxase estimar a media, μ, dunha poboación cunha desviación típica, σ, coñecida.

Para isto recórrese a unha mostra de tamaño n na que se obtén unha media mostral, $\bar{x}$.

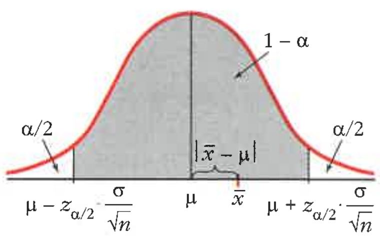
Se a poboación de partida é normal, ou se o tamaño da mostra é $n \geq 30$, entón o **intervalo de confianza** de μ cun nivel de confianza de $(1 - \alpha) \cdot 100\%$ é: $\left(\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}},\ \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right)$

LEMBRA

Teorema Central do Límite

Se $n \geq 30$ ou a poboación é normal, entón

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



Demostración

Temos que demostrar que, baixo as condicións do teorema, cúmprese:

$$(*)\ P\left[\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right] = 1 - \alpha$$

Vexámolo:

Posto que as medias mostrais, $\bar{x}$, se distribúen $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$, obtemos o intervalo característico correspondente a unha probabilidade $1 - \alpha$:

$$P\left[\mu - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \bar{x} < \mu + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right] = 1 - \alpha$$

É dicir, $P\left[|\bar{x} - \mu| < z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right] = 1 - \alpha$. De aquí obtense a relación (*) que é a que se quería demostrar.

OBSERVACIÓNS

- Unha vez extraída a mostra, a súa media $\bar{x}$ estará ou non estará no intervalo $\mu - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$. Pero xa non poderemos falar da probabilidade de que tal cousa aconteza, aínda que por ser μ descoñecida nós ignoremos se acontece ou non. Por iso, en lugar de falar de probabilidade, diremos que temos un **nivel de confianza** $1 - \alpha$ de que μ estea en dito intervalo.

- Se a **desviación típica**, σ, da poboación é **descoñecida,** hai que estimala a partir da mostra.

  A forma máis correcta de facelo é mediante o parámetro mostral $s_{n-1} = \sqrt{\frac{\Sigma (x_i - \bar{x})^2}{n - 1}}$ (non entraremos a explicar por qué).

  Sen embargo, para valores relativamente grandes de n, pode utilizarse a desviación típica da mostra $s_n = s$. O intervalo de confianza para μ quedaría, por tanto, do seguinte modo:

$$\left(\bar{x} - z_{\alpha/2} \cdot \frac{s}{\sqrt{n}},\ \bar{x} + z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}\right)$$

EXERCICIOS RESOLTOS

1. *Desexamos valorar o grao de coñecementos en historia dunha poboación de varios miles de alumnos. Sabemos que $\sigma = 2,3$. Propoñémonos estimar μ pasando unha proba a 100 alumnos.*

a) Calcula o intervalo característico para $\bar{x}$ correspondente a unha probabilidade de 0,95.

Unha vez realizada a proba a 100 alumnos concretos, obtívose unha media $\bar{x} = 6,32$.

b) Obtén o intervalo de confianza de μ cun nivel de confianza do 95%.

As medias, $\bar{x}$, de todas as posibles mostras de tamaño 100 distribúense $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) = N\left(\mu, \frac{2,3}{\sqrt{100}}\right) = N(\mu; 0,23)$.

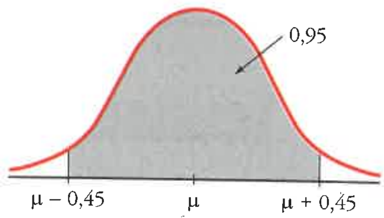
a) $1 - \alpha = 0,95 \rightarrow \alpha = 0,05 \rightarrow z_{\alpha/2} = z_{0,025} = 1,96$

Temos, por tanto:

$P[\bar{x} \in (\mu - 1,96 \cdot 0,23,\ \mu + 1,96 \cdot 0,23)] = 0,95$

É dicir:

$P[\bar{x} \in (\mu - 0,45,\ \mu + 0,45)] = 0,95$



O intervalo característico para $\bar{x}$ correspondente a unha probabilidade de 0,95 é $(\mu - 0,45; \mu + 0,45)$. Isto quere dicir que no 95% das mostras a súa media dista de μ menos de 0,45.

b) O intervalo de confianza para μ (nivel de confianza do 95%) é:

$(\bar{x} - 0,45; \bar{x} + 0,45) \rightarrow (6,32 - 0,45; 6,32 + 0,45) \rightarrow (5,87; 6,77)$

Isto quere dicir que, no 95% das posibles mostras, o intervalo correspondente contén a μ. ¿É este o caso? Non o podemos saber, pero "temos unha confianza do 95% de que é así".

2. *Para estimar a media dos resultados que obterían ó resolver un certo test os alumnos de 4º de ESO de toda unha comunidade autónoma, pásaselles dito test a 400 deles escollidos ó azar.*

Os resultado obtidos veñen dados na táboa da dereita.

A partir deles, estima cun nivel de confianza do 95% o valor da media da poboación.

x_i	f_i
1	24
2	80
3	132
4	101
5	63

Os parámetros estatísticos correspondentes á táboa adxunta son:

$\bar{x} = 3,25; \ s = 1,12$

Posto que desexamos facer a estimación cun nivel de confianza do 95%, temos:

$1 - \alpha = 0,95 \rightarrow \alpha = 0,05 \rightarrow z_{\alpha/2} = 1,96$

O raio do intervalo é:

$$z_{\alpha/2} \cdot \frac{s}{\sqrt{n}} = 1,96 \cdot \frac{1,12}{\sqrt{400}} = 0,10976 \approx 0,11$$

O intervalo buscado é $(3,25 - 0,11; 3,25 + 0,11) = (3,14; 3,36)$.

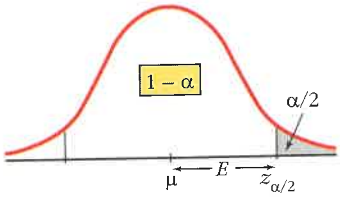
Por tanto, temos unha confianza do 95% de que a nota media da poboación total estea comprendida entre 3,14 e 3,36.

EXERCICIOS PROPOSTOS

1. Dunha variable estatística coñecemos a desviación típica, $\sigma = 8$, pero descoñecemos a media, μ. Para estimala, extraemos unha mostra de tamaño $n = 60$ da que obtemos a media: $\bar{x} = 37$. Estima μ mediante un intervalo de confianza do 99%.

12.6 RELACIÓN ENTRE NIVEL DE CONFIANZA, ERRO ADMISIBLE E TAMAÑO DA MOSTRA

Como acabamos de ver, o $(1 - \alpha) \cdot 100\%$ das mostras cumpren que $|\bar{x} - \mu| < z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$.



$1-\alpha$	α	$z_{\alpha/2}$
0,90	0,10	1,645
0,95	0,05	1,96
0,99	0,01	2,575

O valor $E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$ chámase **erro máximo admisible**.

Depende de α e de n do seguinte modo:

— Canto maior sexa o tamaño da mostra, menor é E (máis estreito é o intervalo, é dicir, máis afinaremos na estimación).

— Canto maior sexa $1 - \alpha$ (é dicir, canto máis seguros queiramos estar da nosa estimación), maior é E.

Na táboa da marxe incluímos os niveis de confianza máis comunmente usados. Neles obsérvase que canto maior é $1 - \alpha$, maior é $z_{\alpha/2}$ e, polo tanto, maior é E.

Calcular o tamaño da mostra dados E e α

Se se nos fixan o erro máximo admisible, E, e o nivel de confianza, $1 - \alpha$, o mínimo tamaño que debe ter unha mostra para conseguir esas condicións obtense despexando n na expresión de E:

$$E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \Rightarrow \sqrt{n} = \frac{z_{\alpha/2} \cdot \sigma}{E} \Rightarrow n = \left(\frac{z_{\alpha/2} \cdot \sigma}{E}\right)^2$$

Observemos que **o tamaño da mostra é**:

— Tanto maior canto maior sexa $z_{\alpha/2}$, ou sexa, canto menor sexa α e maior sexa $1 - \alpha$. É dicir, *para aumentar o nivel de confianza debemos aumentar o tamaño da mostra.*

— Tanto maior canto menor sexa E. É dicir, *para ser máis precisos na estimación temos que aumentar o tamaño da mostra.*

EXERCICIOS RESOLTOS

1. *Da duración dun proceso sabemos que σ = 0,5 s. ¿Cal é o número de medidas que hai que realizar para que, cun 99% de confianza, o erro da estimación non exceda de 0,1 s?*

Ó nivel de confianza do 99% $(\alpha = 0{,}01)$ corresponde: $z_{\alpha/2} = 2{,}575$

Obtención de n:

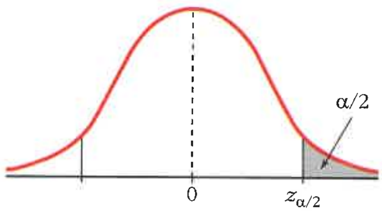
$$0{,}1 = 2{,}575 \cdot \frac{0{,}5}{\sqrt{n}} \rightarrow \sqrt{n} = \frac{2{,}575 \cdot 0{,}5}{0{,}1} = 12{,}875 \rightarrow n = 165{,}76$$

Deben realizarse 166 medidas (o menor enteiro maior que 165,76).

EXERCICIOS PROPOSTOS

1. A desviación típica das estaturas dos soldados é de 5,3 cm.
¿Que tamaño ten que ter a mostra para estimar a estatura media, μ, da poboación cun erro menor de 0,5 cm e cun nivel de confianza do 95%?

Determinar o nivel de confianza coñecendo *E* e *n*



Se se nos fixa o erro máximo admisible, E, e o tamaño da mostra, n, o nivel de confianza co que se realiza a estimación obtense do seguinte modo:

$$E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \Rightarrow z_{\alpha/2} = \frac{E\sqrt{n}}{\sigma}$$

Coñecido $z_{\alpha/2}$, a curva normal daranos o valor de $\frac{\alpha}{2}$.

De aquí obtense o nivel de confianza $1 - \alpha$.

EXERCICIOS RESOLTOS

1. *Ó medir o tempo de reacción, un psicólogo sabe que a desviación típica do mesmo é 0,5 segundos. Desexa estimar o tempo medio de reacción cun erro máximo de 0,1 segundos, para o cal realiza 100 experiencias.*

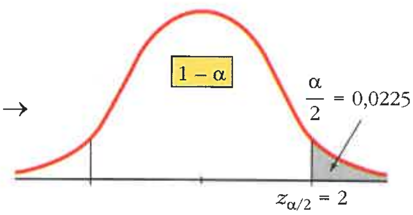
¿Con que nivel de confianza poderá dar o intervalo $(\bar{x} - 0{,}1;\ \bar{x} + 0{,}1)$?

$$0{,}1 = z_{\alpha/2} \cdot \frac{0{,}5}{\sqrt{100}} \Rightarrow z_{\alpha/2} = \frac{0{,}1 \cdot 10}{0{,}5} = 2$$

$P[z < z_{\alpha/2}] = P[z < 2] = 0{,}9772$

$\frac{\alpha}{2} = P[z \geq 2] = 1 - 0{,}9772 = 0{,}0228 \rightarrow$

$\rightarrow \alpha = 0{,}0456 \rightarrow 1 - \alpha = 0{,}9544$



Se $\bar{x}$ é o tempo medio obtido coas 100 experiencias, poderemos asegurar cun nivel de confianza do 95,44% que o tempo de reacción está comprendido entre $\bar{x} - 0{,}1$ e $\bar{x} + 0{,}1$ segundos.

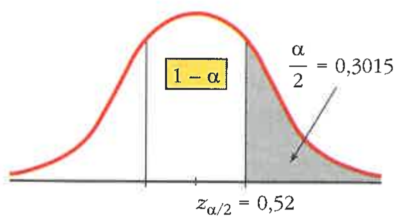
1. *Un coronel desexa estimar a estatura media de todos os soldados do seu rexemento cun erro menor de 0,5 cm utilizando unha mostra de 30 soldados. Sabendo que a desviación típica é $\sigma = 5{,}3$ cm, ¿cal será o nivel de confianza co que se realiza a estimación?*

$$0{,}5 = z_{\alpha/2} \cdot \frac{5{,}3}{\sqrt{30}} \Rightarrow z_{\alpha/2} = \frac{0{,}5 \cdot \sqrt{30}}{5{,}3} = 0{,}52$$

$P[z < z_{\alpha/2}] = P[z < 0{,}52] = 0{,}6985$

$\frac{\alpha}{2} = P[z \geq 0{,}52] = 1 - 0{,}6985 =$

$= 0{,}3015 \rightarrow \alpha = 0{,}6030$



Nivel de confianza: $1 - \alpha = 0{,}3970$

O nivel de confianza sería do 39,7%. ¡Demasiado baixo! Para iso non merece a pena realizar a experiencia.

O baixo nivel de confianza é debido a que se pretendía afinar moito (o erro debía ser menor que medio centímetro) cunha mostra moi pequena (30 individuos).

EXERCICIOS PROPOSTOS

2. Sabemos que a desviación típica dos pesos dos polos adultos é 300 g. Queremos estimar o peso medio dos polos adultos dunha granxa cun erro menor que 100 g e para iso tomamos unha mostra de 50 individuos. ¿Con que nivel de confianza poderemos realizar a estimación?

1. Intervalos característicos

A duración dun tipo de pilas eléctricas ten unha distribución normal con $\mu = 55$ horas e $\sigma = 6$ horas.

Calcula os intervalos característicos correspondentes ás probabilidades $p = 0,75$; $p = 0,90$; $p = 0,95$; $p = 0,99$.

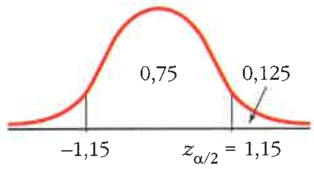
Interpreta o que significan.

- $p = 0,75 \rightarrow p = 1 - \alpha; \quad \alpha = 0,25 \rightarrow \frac{\alpha}{2} = 0,125$

Nunha distribución $N(0, 1)$:

$P[z > z_{\alpha/2}] = 0,125 \rightarrow P[z \leq z_{\alpha/2}] =$

$= 0,875 \rightarrow z_{\alpha/2} = 1,15$



Este último paso realízase buscando nas táboas.

Nunha $N(0, 1)$ o intervalo característico correspondente a $p = 0,75$ é $(-1,15;\ 1,15)$.

Nunha $N(55, 6)$ o intervalo característico correspondente a $p = 0,75$ é:

$(55 - 1,15 \cdot 6;\ 55 + 1,15 \cdot 6) = (55 - 6,9;\ 55 + 6,9) = (48,1;\ 61,9)$

O 75% das pilas teñen unha duración superior a 48,1 horas e inferior a 61,9 horas.

- Para $p = 0,9$; $p = 0,95$; $p = 0,99$, temos en conta que os seus valores críticos, $z_{\alpha/2}$, son coñecidos. Son, respectivamente, 1,645; 1,96; 2,575.

Para a $N(55, 6)$ obtéñense os intervalos característicos seguintes:

PROBABILIDADE	INTERVALO CARACTERÍSTICO
0,90	(45,13; 64,87)
0,95	(43,24; 66,76)
0,99	(39,55; 70,45)

2. Teorema Central do Límite

a) A variable aleatoria x distribúese con media $\mu_x = 17$ e desviación típica $\sigma_x = 5$.

¿Que podemos dicir das medias $\bar{x}$ das mostras de tamaño 100?

b) A variable y distribúese con media $\mu_y = 180$ e desviación típica $\sigma_y = 24$.

¿Que podemos decir das medias, $\bar{y}$, das mostras de tamaño 25?

a) Polo teorema central do límite, a media das mostras de tamaño 100 distribúense normalmente $N\left(\mu_x, \frac{\sigma_x}{\sqrt{n}}\right)$.

É dicir:

$\bar{x}$ é $N(17; 0,5)$

b) Como o tamaño da mostra é menor que 30, só podemos dicir que $\bar{y}$ distribúese con media $\mu_y = 180$ e desviación típica:

$$\frac{\sigma_y}{\sqrt{n}} = \frac{24}{5} = 4,8.$$

Se a poboación de partida, y, fora normal, entón as $\bar{y}$ tamén se distribuirían normal.

É dicir:

Se y é $N(180, 24)$, entón $\bar{y}$ é $N(180; 4,8)$

3. Distribución de medias mostrais

A media de idade das alumnas e os alumnos que se presentan ás probas de acceso á Universidade é de 18,1 anos; e a desviación típica 0,6 anos.

Das alumnas e alumnos anteriores vaise elixir, ó azar, unha mostra de 100, ¿cal é a probabilidade de que a media de idade da mostra estea comprendida entre 17,9 e 18,2 anos?

Como o tamaño da mostra é $n = 100$ ($n \geq 30$), polo teorema central do límite sabemos que as medias mostrais se distribúen segundo unha normal de media $\mu = 18{,}1$ e de desviación típica

$$\frac{\sigma}{\sqrt{n}} = \frac{0{,}6}{\sqrt{100}} = \frac{0{,}6}{10} = 0{,}06. \text{ É dicir: } \bar{x} \sim N(18{,}1;\ 0{,}06)$$

Por tanto, temos que:

$$P[17{,}9 < \bar{x} < 18{,}2] = P\left[\frac{17{,}9 - 18{,}1}{0{,}06} < z < \frac{18{,}2 - 18{,}1}{0{,}06}\right] =$$

$$= P[-3{,}33 < z < 1{,}67] =$$

$$= P[z < 1{,}67] - P[z < -3{,}33] =$$

$$= P[z < 1{,}67] - (1 - P[z < 3{,}33]) =$$

$$= 0{,}9525 - (1 - 0{,}9996) = 0{,}9521$$

4. Distribución de medias mostrais

O cociente intelectual (C.I.) dos alumnos dun centro distribúese N(110, 15). Propoñémonos extraer unha mostra aleatoria de tamaño n = 25.

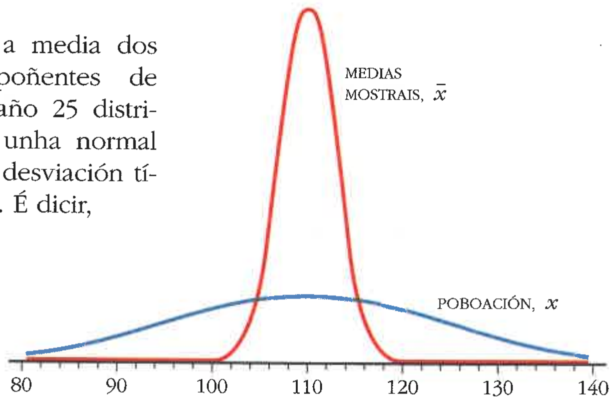
a) ¿Cal é a distribución das medias das mostras que poden extraerse?

b) ¿Cal é a probabilidade de que a media do C.I. dos 25 alumnos dunha mostra sexa superior a 115?

c) Dar o intervalo característico das medias mostrais correspondentes a unha probabilidade 1 – α = 0,95.

a) Posto que a poboación de partida xa é normal, as medias mostrais de tamaño n distribúense $N(\mu, \sigma/\sqrt{n})$, aínda que n sexa menor que 30.

No noso caso, a media dos C.I. dos compoñentes de mostras de tamaño 25 distribúese segundo unha normal de media 110 e desviación típica $15/\sqrt{25} = 3$. É dicir, $\bar{x}$ é $N(110, 3)$.



b) Posto que $\bar{x}$ é $N(110, 3)$

$$P[\bar{x} > 115] = P\left[z > \frac{115 - 110}{3}\right] = P[z > 1{,}67] = 1 - \phi(1{,}67) =$$

$$= 1 - 0{,}9525 = 0{,}0475$$

É dicir, só o 4,75% das mostras de 25 alumnos teñen un C.I. medio superior a 115.

c) Para $1 - \alpha = 0{,}95$ corresponde un valor crítico $z_{\alpha/2} = 1{,}96$. Por tanto, o intervalo característico das medias mostrais correspondente a esta probabilidade é:

$$\left(\mu \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right) = (110 \pm 1{,}96 \cdot 3) = (110 \pm 5{,}88) = (104{,}12;\ 115{,}88)$$

5. Distribución de medias mostrais

Lanzamos 36 dados correctos e calculamos a media dos seus resultados, $\bar{x}$. Se repetiramos esta experiencia de forma reiterada:

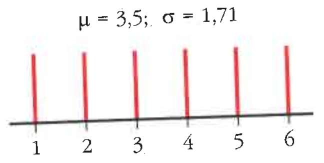
a) ¿Cal sería a distribución das medias, $\bar{x}$?

b) ¿Cal é a probabilidade de que a media dunha das tiradas sexa maior que 4?

c) Calcula un intervalo centrado na media no que se encontre o 99% das medias dos lanzamentos.

a) Ó lanzar un dado correcto, a media das puntuacións é $\mu = 3{,}5$ e a desviación típica, $\sigma = 1{,}71$.

Estamos considerando mostras de tamaño $n = 36$.

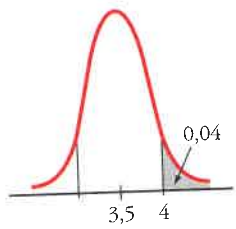


Posto que $n > 30$, as súa medias, $\bar{x}$, distribúense

$N(3{,}5;\ 1{,}71/\sqrt{36}) = N(3{,}5;\ 0{,}285)$

b) $P[\bar{x} > 4] = P\left[z > \dfrac{4 - 3{,}5}{0{,}285}\right] = P[z > 1{,}75] =$

$= 1 - \phi(1{,}75) = 1 - 0{,}9599 = 0{,}0401$



Só no 4% dos casos a media das puntuacións dos 36 dados superará os 4 puntos.

c) Para $1 - \alpha = 0{,}99$ corresponde un valor crítico $z_{\alpha/2} = 2{,}575$. Por tanto, o intervalo característico das medias mostrais correspondente a esta probabilidade é:

$$\left(\mu \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right) = (3{,}5 \pm 2{,}575 \cdot 0{,}285) = (3{,}5 \pm 0{,}73) = (2{,}77;\ 4{,}23)$$

O 99% das medias dos 36 dados estarán entre 2,77 e 4,23.

6. Intervalo de confianza para a media

a) Un gandeiro de reses bravas quere estimar o peso medio dos touros da súa gandería cun nivel de confianza do 95%. Para iso, toma unha mostra de 30 touros e pésaos. Obtén unha media $\bar{x} = 507$ kg e unha desviación típica $s = 32$ kg.

¿Cal é o intervalo de confianza para a media μ da poboación?

b) ¿Cal será o intervalo se queremos que o nivel de confianza sexa do 99%?

a) Posto que descoñecemos a desviación típica, σ, da poboación, estimámola a partir da desviación típica da mostra, $s = 32$.

O raio do intervalo buscado (erro máximo admisible) é:

$$E = z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

Para un nivel de confianza $1 - \alpha = 0{,}95$, o valor crítico é $z_{\alpha/2} = 1{,}96$. Por tanto:

$$E = 1{,}96 \cdot \frac{32}{\sqrt{30}} = 11{,}45$$

O intervalo buscado é: $507 \pm 11{,}45 = (495{,}55;\ 518{,}45)$

Conclusión: O gandeiro pode ter un nivel de confianza do 95% de que o peso medio dos seus touros será superior a 495,55 kg e inferior a 518,45 kg.

b) Para $1 - \alpha = 0{,}99$, o valor crítico é $z_{\alpha/2} = 2{,}575$. Por tanto, o intervalo de confianza será:

$$507 \pm 2{,}575 \cdot 5{,}84 = 507 \pm 15 = (492,\ 522)$$

7. Estimación da media. Tamaño da mostra

O cociente intelectual dun certo colectivo ten unha media μ descoñecida e unha desviación típica $\sigma = 8$.

¿De que tamaño debe ser a mostra coa que se estime a media cun nivel de confianza do 99% e un erro admisible $E = 3$?

O erro máximo admisible é:

$$E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}.$$

Desta expresión coñecemos:

$$E = 3$$

$$1 - \alpha = 0{,}99 \rightarrow \alpha = 0{,}01 \rightarrow z_{\alpha/2} = 2{,}575$$

$$\sigma = 8$$

Por tanto:

$$3 = 2{,}575 \cdot \frac{8}{\sqrt{n}} \rightarrow \sqrt{n} = \frac{2{,}575 \cdot 8}{100} = 6{,}87 \rightarrow n = 47{,}15$$

Haberá que tomar unha mostra de 48 individuos.

8. Estimación da media. Nivel de confianza

Para estimar o peso medio das rapazas de 16 anos dunha cidade, tómase unha mostra aleatoria de 100 delas. Obtéñense os seguintes parámetros:

$\bar{x} = 52{,}5$ kg, $s = 5{,}3$ kg.

Realízase a afirmación seguinte: o peso medio das rapazas de 16 anos desta cidade está entre 51 e 54 kg. ¿Con que nivel de confianza se fai a afirmación?

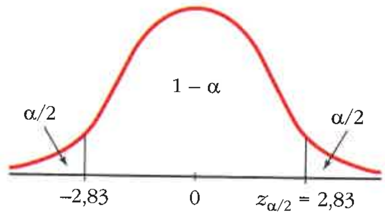
Na fórmula de erro máximo admisible, $E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$, coñecemos:

— E é a metade do intervalo. Por tanto, $E = \frac{54 - 51}{2} = 1{,}5$.

— O valor de σ estímase a partir de $s = 5{,}3$.

— $n = 100$

Obtemos, pois, $z_{\alpha/2}$: $z_{\alpha/2} = \frac{E\sqrt{n}}{\sigma} = \frac{1{,}5\sqrt{100}}{5{,}3} = 2{,}83$



Calculemos $1 - \alpha$ a partir de $z_{\alpha/2} = 2{,}83$:

$$\frac{\alpha}{2} = P[z > z_{\alpha/2}] = P[z > 2{,}83] =$$

$$= 1 - P[z \leq 2{,}83] = 1 - 0{,}9977 = 0{,}0023$$

$$\alpha = 2 \cdot 0{,}0023 = 0{,}0046$$

$$1 - \alpha = 1 - 0{,}0046 = 0{,}9954$$

Tense un nivel de confianza do 99,54% de que o peso medio das rapazas de 16 anos desta cidade está comprendido entre 51 e 54 kg.

EXERCICIOS E PROBLEMAS PROPOSTOS

PARA PRACTICAR

Intervalos característicos

1 Nas distribucións normais das que se dan os parámetros, calcula o intervalo característico que en cada caso se indica:

	a)	b)	c)	d)	e)
MEDIA, μ	0	0	0	0	112
DESV. TÍPICA, σ	1	1	1	1	15
PROBAB. $1-\alpha$	0,95	0,99	0,90	0,80	0,95

	f)	g)	h)	i)
MEDIA, μ	3 512	3 512	3 512	3 512
DESV. TÍPICA, σ	550	550	550	550
PROBAB. $1-\alpha$	0,99	0,95	0,90	0,80

2 Nunha distribución normal con media $\mu = 25$ e desviación típica $\sigma = 5{,}3$; obtén un intervalo centrado na media, $(\mu - k, \mu + k)$, de forma que o 95% dos individuos estean nese intervalo.

3 Nunha distribución $N(10, 4)$, obtén un intervalo centrado na media $(\mu - k, \mu + k)$, tal que:

$$P[\mu - k < x < \mu + k] = 0{,}90$$

4 Nunha distribución normal de media $\mu = 9{,}5$ e varianza $\sigma^2 = 1{,}44$, determina o intervalo característico para o 99%.

Teorema Central do Límite

5 Dunha variable aleatoria x de distribución descoñecida, media $\mu = 23$ e desviación típica $\sigma = 3{,}5$ extráense mostras de tamaño n. ¿Que se pode dicir da distribución das medias mostrais, $\bar{x}$:

a) no caso de que $n = 49$?

b) no caso de que $n = 25$?

6 Unha variable aleatoria x distribúese normal $N(120, 30)$. ¿Que se pode afirmar da distribución das medias $\bar{x}$ das mostras de tamaño n:

a) se $n = 36$? b) se $n = 16$?

7 Di cómo se distribúen as medias mostrais en cada un dos seguintes casos:

POBOACIÓN		a)	b)	c)
POBOACIÓN	DISTRIBUCIÓN	Normal	Desc.	Normal
POBOACIÓN	MEDIA, μ	20	20	3,75
POBOACIÓN	DESV. TÍPICA, σ	4	4	1,2
TAM. MOSTRA, n		16	100	4

POBOACIÓN		d)	e)	f)	g)
POBOACIÓN	DISTRIBUCIÓN	Desc.	Norm.	Desc.	Desc.
POBOACIÓN	MEDIA, μ	3,75	112	112	3 512
POBOACIÓN	DESV. TÍPICA, σ	1,2	15	15	550
TAM. MOSTRA, n		50	100	100	40

Distribución das medias mostrais

8 Unha variable aleatoria distribúese $N(\mu, \sigma)$. Se se extraen mostras de tamaño n:

a) ¿Que distribución ten a variable aleatoria media mostral, $\bar{x}$?

b) Se se toman mostras de tamaño $n = 4$ dunha variable aleatoria x con distribución $N(165, 12)$, calcula $P[\bar{x} > 173{,}7]$.

9 S Nunha distribución $N(20, 6)$, tomamos mostras de tamaño 64.

a) ¿Cal é a distribución das medias das mostras?

b) ¿Cal é a probabilidade de extraer unha mostra cunha media comprendida entre 19 e 21?

10 S Sábese que o cociente intelectual dos alumnos dunha universidade distribúese segundo unha lei normal de media 100 e varianza 729.

a) Calcula a probabilidade de que unha mostra de 81 alumnos teña un cociente intelectual medio inferior a 109.

b) Determina a probabilidade de que unha mostra de 36 alumnos teña un cociente intelectual medio superior a 109.

11 S O tempo de espera, en minutos, dos pacientes nun servicio de urxencias, é $N(14, 4)$.

a) ¿Como se distribúe o tempo medio de espera de 16 pacientes?

b) Nunha media xornada atendéronse 16 pacientes. ¿Cal é a probabilidade de que o tempo medio da súa espera estea comprendido entre 10 e 15 minutos?

12 S Sábese que o peso en quilogramos dos alumnos de Bacharelato de Madrid é unha variable aleatoria, x; que segue unha distribución normal de desviación típica igual a 5 kg. No caso de considerar mostras de 25 alumnos, ¿que distribución ten a variable aleatoria media mostral, $\bar{x}$?

13 S Nunha cidade, a altura media dos seus habitantes ten unha desviación típica de 8 cm. Se a altura media de ditos habitantes fora de 175 cm, ¿cal sería a probabilidade de que a altura media dunha mostra de 100 individuos tomada ó azar fora superior a 176 cm?

Intervalo de confianza para a media

14 A desviación típica dunha variable estatística é $\sigma = 5$. Para estimar a media de dita variable, extraemos unha mostra aleatoria de tamaño $n = 100$ e obtemos $\bar{x} = 2{,}8$. Obtén un intervalo de confianza do 95% para estimar a media da poboación, μ.

15 Os estudiantes de Bacharelato dunha certa comunidade autónoma dormen un número de horas diarias que se distribúe segundo unha lei normal de media μ descoñecida e de desviación típica 3. A partir dunha mostra de tamaño 30 obtívose unha media mostral igual a 7 horas.

Obtén un intervalo de confianza ó 90% para a media de horas de sono, μ.

16 Nunha mostra de 50 mozos encontramos que a dedicación media diaria de ocio é de 400 minutos e a súa desviación típica de 63 minutos. Calcula o intervalo de confianza da media da poboación ó 95% de nivel de confianza.

PARA RESOLVER

17 As notas nun certo exame distribúense normal con media $\mu = 5{,}3$ e desviación típica $\sigma = 2{,}4$.

Calcula a probabilidade de que un estudiante tomado ó azar teña unha nota:

a) Superior a 7.

b) Inferior a 5.

c) Comprendida entre 5 e 7.

Tomamos ó azar 16 estudiantes.

Determina a probabilidade de que a media das notas destes 16 estudiantes:

d) Sexa superior a 7.

e) Sexa inferior a 5.

f) Estea comprendida entre 5 e 7.

g) Calcula k para que o intervalo

$(5{,}3 - k;\ 5{,}3 + k)$

conteña o 95% das notas.

h) Calcula b para que o intervalo

$(5{,}3 - b;\ 5{,}3 + b)$

conteña o 95% das notas medias das mostras de 16 individuos.

18 S A estatura dos mozos dunha cidade segue unha distribución $N(\mu, \sigma)$. Se o 90% das medias das mostras de 81 mozos están en (173,4; 175,8), calcula μ e σ.

ten que ser centrado

19 S Se a distribución da media das alturas en mostras de tamaño 49 dos nenos de 10 anos ten como media 135 cm e como desviación típica 1,2 cm, ¿canto valen a media e a varianza da altura dos nenos desa cidade?

20 S O peso dos paquetes recibidos nun almacén distribúese normal con media 300 kg e desviación típica 50 kg. ¿Cal é a probabilidade de que 25 dos paquetes, elixidos ó azar, excedan o límite de carga do montacargas onde se van meter, que é de 8 200 kg?

21 O peso dos cans adultos dunha certa raza é unha variable aleatoria que se distribúe normalmente cunha media de 7,4 kg e unha desviación típica de 0,6 kg.

Se consideramos mostras de 30 destes animais:

a) ¿Cal é a distribución da media mostral, $\bar{x}$?

b) Calcula $P[6{,}5 < \bar{x} < 7{,}5]$.

c) ¿Cal é a distribución da suma dos pesos dos 30 animais das mostras?

d) Calcula $P\left[\sum_{i=1}^{30} x_i > 225\right]$.

22 Suponse que o peso medio das sandías de certa variedade segue unha distribución normal con $\mu = 6$ kg e $\sigma = 1$ kg. Se empaquetamos as sandías en caixas de 8 unidades:

a) Determina a probabilidade de que a media dos pesos das sandías dunha caixa sexa menor que 5,5 kg.

b) Calcula a probabilidade de que entre as 8 sandías dunha das caixas pesen máis de 50 kg.

23 S Para estimar a estatura media dos mozos entre 15 e 25 anos dunha localidade, medíronse 40 destes mozos, obténdose os seguintes resultados:

ESTATURA (cm)	[148, 153)	[153, 158)	[158, 163)
Nº DE MOZOS	2	4	11

ESTATURA (cm)	[163, 168)	[168, 173)	[173, 178)
Nº DE MOZOS	14	5	4

Estima, cun nivel de confianza do 99%, o valor da estatura media dos mozos entre 15 e 25 anos de dita localidade.

24 S Deséxase estudiar o gasto semanal de fotocopias, en céntimos de euro, dos estudiantes de Bacharelato de certa comunidade. Para iso, elixiuse unha mostra aleatoria de 9 deses estudiantes, resultado os valores seguintes para estes gastos:

100; 150; 90; 70; 75; 105; 200; 120; 80

Suponse que a variable obxecto do estudio segue unha distribución normal de media descoñecida e desviación típica igual a 12. Determina un intervalo de confianza ó 95% para a media do gasto semanal en fotocopias por estudiante.

25 S Tomouse unha mostra aleatoria de 100 individuos ós que se lles mediu o nivel de glicosa en sangue, obténdose unha media mostral de 110 mg/cm^3. Sábese que a desviación típica da poboación é de 20 mg/cm^3.

a) Obtén un intervalo de confianza, ó 90%, para o nivel de glicosa en sangue na poboación.

b) ¿Que erro máximo se comete en a)?

26 A duración das lámpadas fabricadas por unha empresa segue unha distribución normal de media descoñecida e desviación típica 50 horas. Para estimar a duración experiméntase cunha mostra de tamaño n.

Calcular o valor de n para que, cun nivel de confianza do 95%, se consiga un erro na estimación inferior a 5 horas.

27 S A media das estaturas dunha mostra aleatoria de 400 persoas dunha cidade é 1,75 m. Sábese que a estatura das persoas desa cidade é unha variable aleatoria que segue unha distribución normal con varianza $\sigma^2 = 0{,}16$ m^2.

a) Constrúe un intervalo, dun 95% de confianza, para a media das estaturas da poboación.

b) ¿Cal sería o mínimo tamaño mostral necesario para que poida dicirse que a verdadeira media das estatura está a menos de 2 cm da media mostral, cunha confianza do 90%?

28 S As medidas dos diámetros dunha mostra ó azar de 200 rodamentos de bólas deron unha media de 2 cm e unha desviación típica de 0,1 cm. Calcula os intervalos de confianza do 68,26%, 95,44% e 99,73% para o diámetro medio de todos os rodamentos.

29 S O peso, en kg, dos mozos entre 16 e 20 anos dunha certa cidade é unha variable aleatoria, x, que segue unha distribución normal con $\sigma^2 = 25$.

a) Se consideramos mostras de 25 mozos, ¿cal é a distribución que ten a variable aleatoria media mostral?

b) Se se desexa que a media da mostra non difira en máis de 1 kg da media da poboación, con probabilidade 0,95, ¿cantos mozos se deberían tomar na mostra?

30 S Unha variable aleatoria, x, ten unha distribución normal, sendo a súa desviación típica igual a 3.

a) Se se consideran mostras de tamaño 16, ¿que distribución segue a variable aleatoria media mostral?

b) Se se desexa que a media da mostra non difira en máis dunha unidade da media da poboación, con probabilidade 0,99, ¿cantos elementos, como mínimo, se deberían tomar na mostra?

31 S O tempo de vida dunha clase de depuradoras de auga utilizadas nunha planta industrial distribúese normalmente, cunha desviación típica de 2 000 horas. Nun ensaio realizado cunha mostra aleatoria de 9 depuradoras, obtivéronse os seguintes tempos de vida en miles de horas:

9,5 10 7,5 10,5 16,5 10 12 32 18

a) Calcula un intervalo de confianza ó 99% para a vida media das depuradoras.

b) ¿Cal é o erro máximo que se comete coa estimación anterior para a media?

c) Calcula o tamaño mínimo que debería ter a mostra, no caso de admitir un erro máximo de 500 horas, cun grao de confianza do 95%.

32 S Ó medir o tempo de reacción, un psicólogo sabe que a desviación típica do mesmo é de 0,5 segundos. ¿Cal será o número de medidas que deberá facer para que sexa do 95% a confianza de que o erro da estimación da media non excederá de 0,05 segundos?

33 Ó medir o diámetro das chumaceiras producidas por unha empresa, estímase que a desviación típica de dito diámetro é de 0,05 cm. Fixéronse 121 medicións.

¿Pódese afirmar, co 99% de confianza, que o erro na estimación da media non excederá a 0,01 cm?

34 S As vendas mensuais dunha tenda de electrodomésticos distribúense segundo unha lei normal, con desviación típica 900 €. Nun estudio estatístico das vendas realizadas nos últimos nove meses, encontrouse un intervalo de confianza para a media mensual das vendas, con extremos de 4 663 € e 5 839 €.

a) ¿Cal foi a media das vendas nestes nove meses?

b) ¿Cal é o nivel de confianza para este intervalo?

35 Suponse que os gastos correntes por empregado dos distintos departamentos dunha empresa seguen unha distribución normal con desviación típica 500 €. Dos datos dispoñibles para 16 departamentos, obtívose o seguinte intervalo de confianza para estimar a media do gasto corrente por empregado da empresa: (1 928,125; 2 571,875)

¿Cal é o nivel de confianza, $1 - \alpha$, co que se fixo a estimación?

CUESTIÓNS TEÓRICAS

36 Cunha mostra de 500 individuos estimamos, cun nivel de confianza do 90%, que a estatura media dos soldados dun certo cuartel está entre 174,3 cm e 175,1 cm.

a) Se a desviación típica da poboación era descoñecida, indaga a media, $\bar{x}$, e a desviación típica, s, da mostra.

b) ¿Cal sería o intervalo se a mostra fora de tamaño a cuarta parte (500 : 4 = 125) e mantiveramos o nivel de confianza?

37 S Supoñamos que, a partir dunha mostra aleatoria de tamaño $n = 25$, se calculou o intervalo de confianza para a media dunha poboación normal, obténdose unha amplitude igual a ± 4. Se o tamaño da mostra fora $n = 100$, permanecendo invariables todos os demais valores que interveñen no cálculo, ¿cal sería a amplitude do intervalo?

38 A partir dunha mostra aleatoria, estimamos o peso dos touros dunha manda mediante o intervalo (443, 528).

¿Cal é a media da mostra obtida?

39 Mediante unha mostra de 100 individuos estimamos a estatura dun colectivo de persoas cun nivel de confianza do 95%. O erro máximo admisible obtido é $E = 1,274$. ¿Cal é a desviación típica da mostra obtida?

13 INFERENCIA ESTATÍSTICA: ESTIMACIÓN DUNHA PROPORCIÓN

Na unidade anterior aprendemos a estimar a media dunha poboación a partir da media mostral, coa axuda da distribución normal. Cabe sinalar que isto só é posible cando a mostra de que se dispón é suficientemente grande. Para mostras pequenas a curva normal falla e hai que recorrer a outra distribución, que non estudiaremos neste curso, chamada *t de Student* (Student era o pseudónimo que utilizaba o seu inventor, **Gosset,** un químico que traballaba na fábrica de cervexa Guinness. Recorreu a un pseudónimo porque a súa empresa non lle permitía publicar resultados de investigacións científicas utilizando o seu propio nome).

Pois ben, hai outros parámetros que, ocasionalmente, deben ser estimados mediante mostras. Nesta unidade dedicarémonos a un deles: a *proporción* de individuos dun colectivo que posúe unha certa calidade (ou, o que é equivalente, a *probabilidade* de que ocorra un certo suceso).

A descrición das probabilidades dos distintos valores dunha proporción realízase coa axuda da distribución binomial e esta, á súa vez, pode ser substituída en certos casos pola normal. De modo que, tamén nesta unidade, volveremos facer uso da distribución normal para realizar estimacións.

PARA EMPEZAR, REFLEXIONA E RESOLVE

¿Cantas caras cabe esperar?

Para entender estes parágrafos, é necesario que domines a distribución binomial que aprendiches no curso pasado. Se non é así, repásaa nas páxinas 300 e 301.

Lanzamos 100 moedas correctas. ¿Cantas caras cabe esperar que saian? Como sabes, o número de caras segue unha distribución binomial con $n = 100$ e $p = \frac{1}{2}$. É dicir, x é $B\left(100, \frac{1}{2}\right)$ e, por tanto, aproxímase a unha normal de media $\mu = 100 \cdot \frac{1}{2} = 50$

e desviación típica $\sigma = \sqrt{100 \cdot \frac{1}{2} \cdot \frac{1}{2}} = \sqrt{25} = 5$.

En resumo, x é aproximadamente $N(50, 5)$.

O intervalo característico para esta distribución, correspondente a unha probabilidade do 90%, é:

$$50 \pm 1{,}645 \cdot 5 = (41{,}775;\ 58{,}225)$$

Isto significa que no 90% dos casos en que tiremos 100 moedas, o número de caras que obteremos será maior que 41 e menor que 59. Calquera outro resultado será un "caso raro" (chamando "casos raros" a ese 10% de casos extremos).

- Repite o razoamento anterior para descubrir *cántas caras cabe esperar* se lanzamos 100 moedas e consideramos "casos raros" ó 5% dos casos extremos.

Un saco de feixóns

Temos un saco con 10 000 feixóns. Deles, 9 500 son brancos e 500 negros. Están ben mesturados.

Extraemos 600 feixóns.

¿Cantos feixóns negros *cabe esperar* que haxa entre eles?

- Resolve o problema anterior considerando "casos raros" só o 1% dos casos extremos. Para iso:

  a) Indaga a proporción, p, de feixóns negros no saco.

  b) Considera a distribución $B(600, p)$ e calcula a súa media $\mu = 600p$ e a súa desviación típica $\sigma = \sqrt{600 \cdot p(1-p)}$.

  c) Considera a distribución $N(\mu, \sigma)$ e calcula o seu intervalo característico correspondente a unha probabilidade do 99%.

  d) Decide, como consecuencia do resultado anterior, entre qué valores se encontra o número de feixóns que *cabe esperar*.

Peixes nun pantano

Para estimar o número de peixes que hai nun pantano, procédese do seguinte modo:

- Péscase unha certa cantidade deles, por exemplo, 349, márcanse e devólvense ó pantano. (Para marcalos existen unha tintas indelebles que son resistentes á auga).

- Ó cabo de varios días, vólvese pescar outro montón e indágase qué proporción deles están marcados.

  Supoñamos que nesta segunda pesca se obtiveron 514 peixes, dos cales hai 37 marcados.

- Cos datos anteriores, descubre cántos peixes hai, aproximadamente, no pantano.

Este problema resolverase de forma completa (mediante un intervalo de confianza) ó terminar a unidade. Agora só se pretende que deas un valor aproximado.

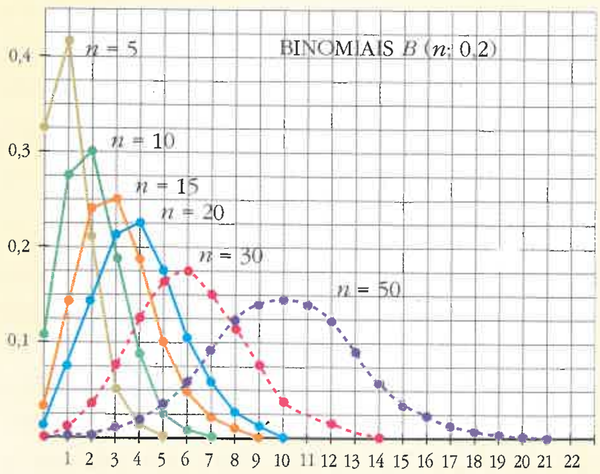
NESTA UNIDADE VERÁS

O itinerario que seguimos na unidade anterior para estimar a media dunha poboación a partir da media dunha mostra, seguirémolo nesta unidade para **estimar proporcións** nunha poboación: a proporción de daltónicos entre as persoas, a proporción de parafusos defectuosos nunha producción ou a probabilidade de que ocorra un suceso (a probabilidade é a proporción de veces que ocorrería tal suceso se a experiencia se repetira indefinidamente).

- Comezaremos revisando as ferramentas básicas. Neste caso, a **distribución binomial,** a súa aproximación á normal e, por tanto, a posibilidade de obter intervalos característicos correspondentes a unha certa probabilidade.

- Veremos que, para unha certa proporción na poboación, as **proporcións mostrais** (as proporcións nas distintas mostras) distribúense de forma normal, cuns parámetros perfectamente determinados.

- Finalmente, pasaremos dos intervalos característicos para as proporcións mostrais a intervalos de confianza para a proporción poboacional. Desta forma obtemos o procedemento para a inferencia dunha proporción.



13.1 DISTRIBUCIÓN BINOMIAL. REPASO DE TÉCNICAS BÁSICAS PARA A MOSTRAXE

Se nunha experiencia aleatoria destacamos un suceso A e prestamos atención, exclusivamente, a se ocorre A ou o seu contrario, A', trátase dunha **experiencia dicotómica.** Ó suceso A adoita chamárselle éxito e á súa probabilidade, p. A probabilidade do seu contrario é q.

É dicir, $P[A] = p$, $P[A'] = 1 - p = q$.

EXEMPLO

Se o 3% das persoas son daltónicas e lle chamamos A ó suceso "unha persoa é daltónica", tense $P[A] = 0{,}03$ e $P[A'] = 0{,}97$.

Distribución binomial

Repítese n veces unha mesma experiencia dicotómica. Preguntámonos polo número, x, de éxitos. A variable x é unha variable discreta e pode tomar o valores 1, 2, 3, ..., n.

A distribución de probabilidade da variable x chámase **distribución binomial** $B(n, p)$.

A probabilidade de que x tome o valor k é:

$$P[x = k] = \binom{n}{k} p^k q^{n-k}$$

Os parámetros desta distribución son: $\mu = np$, $\sigma = \sqrt{npq}$

NOTACIÓN

$p = P[A]$ é a probabilidade de éxito en cada unha das experiencias.

n é o número de veces que se repite a experiencia.

Se no exemplo da marxe tomamos ó azar 7 persoas e nos preguntamos polo número delas que son daltónicas, entón trátase dunha distribución binomial con $n = 7$ e $p = 0{,}03$. É dicir, x é $B(7; 0{,}03)$.

$$P[2 \text{ daltónicas}] = P[x = 2] = \binom{7}{2} 0{,}03^2 \cdot 0{,}97^5 = 0{,}01623$$

Os parámetros da distribución son:

$$\mu = np = 7 \cdot 0{,}03 = 0{,}21, \quad \sigma = \sqrt{npq} = \sqrt{7 \cdot 0{,}03 \cdot 0{,}97} = 0{,}4513$$

A distribución binomial aproxímase á normal

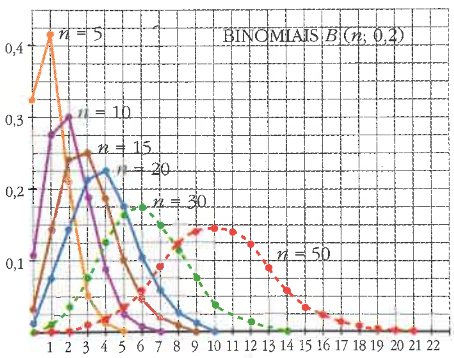
Unha distribución binomial parécese a unha normal tanto máis canto maior é o producto np (ou nq se $q < p$).

Cando np e nq son ambos maiores que 3, a aproximación é bastante boa. E se superan a 5, a aproximación é case perfecta.

Naturalmente, a curva normal á cal se aproxima a $B(n, p)$ ten a mesma media $\mu = np$ e a mesma desviación típica $\sigma = \sqrt{npq}$:

$$B(n, p) \approx N(np, \sqrt{npq})$$

Na aproximación de ambas as distribucións hai que ter en conta que a binomial é discreta e a normal, continua.



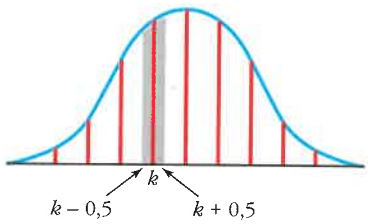
Se no mesmo exemplo en vez de 7 tomamos 200 persoas, o número de daltónicas, x, é unha distribución binomial $B(200; 0{,}03)$ cos parámetros son $\mu = 200 \cdot 0{,}03 = 6$ e $\sigma = \sqrt{200 \cdot 0{,}03 \cdot 0{,}97} = 2{,}41$.

Aproxímase a unha normal cos mesmos parámetros: $N(6; 2{,}41)$

Cálculo de probabilidades nunha binomial mediante a aproximación á normal

LEMBRA

Se x é $B(n, p)$ e x' é $N(np, \sqrt{npq})$, entón:
$P[x = k] = P[k - 0{,}5 < x' < k + 0{,}5]$



A área da parte gris, de base 1, é aproximadamente igual á lonxitude da barra vermella.

A variable x distribúese $B(n, p)$, $\mu = np$, $\sigma = \sqrt{npq}$.

A variable x' distribúese $N(np, \sqrt{npq})$.

Se $np \geq 5$ e $nq \geq 5$, estas dúas distribucións son case idénticas salvo que x é discreta (toma valores 0, 1, 2, ..., n) mentres que x' é continua. Esta diferencia apréciase cando calculamos probabilidades, que debemos indagalas do seguinte modo:

$$P[x = k] = P[k - 0{,}5 < x' < k + 0{,}5] \quad \text{(ver figura na marxe)}$$

Como x é discreta e toma valores puntuais: 0, 1, 2, ..., k, ..., ó pasar a x' continua, asóciaselle un intervalo unidade centrado no punto correspondente:

$$k \to [k - 0{,}5;\ k + 5] \quad \begin{cases} P[a \leq x < b] = P[a - 0{,}5 < x' < b - 0{,}5] \\ P[a < x \leq b] = P[a + 0{,}5 < x' < b + 0{,}5] \end{cases}$$

Esta é a formulación correcta, como a aprendiches o curso pasado. Non obstante, para simplificar o proceso que se segue nos próximos apartados (xa de por si algo complicado), déixase de lado esta diferencia entre as variables x binomial e x' normal e calcularanse as probabilidades de intervalos de x directamente na curva normal.

EXERCICIOS RESOLTOS

1. *Unha máquina fabrica parafusos. O 5% deles son defectuosos. Empaquétanse en caixas de 400. Calcula a probabilidade de que nunha caixa haxa máis de 30 defectuosos.*

x: número de parafusos defectuosos nunha caixa de 400.

x é binomial, con $n = 400$ e $p = 0{,}05$: $B(400;\ 0{,}05)$

Os seus parámetros son: $\mu = 400 \cdot 0{,}05 = 20$; $\sigma = \sqrt{400 \cdot 0{,}05 \cdot 0{,}95} = 4{,}36$

A distribución x é moi parecida a unha normal: x' é $N(20;\ 4{,}36)$

$$x\ B(400;\ 0{,}05) \xrightarrow{(*)} x\ N(20;\ 4{,}36)$$

$$P[x > 30] = P\left[z > \frac{30 - 20}{4{,}36}\right] = P[z > 2{,}29] = 1 - 0{,}9890 = 0{,}011$$

(*) O correcto sería proceder así

$$\left.\begin{array}{l} x\ B(400;\ 0{,}05) \\ \quad\downarrow \\ x'\ N(20;\ 4{,}36) \end{array}\right\} \begin{array}{l} P[x > 30] = P[x' \geq 30{,}5] = P\left[z \geq \dfrac{30{,}5 - 20}{4{,}36}\right] = \\ = P[z \geq 2{,}41] = 1 - \phi(2{,}41) = 1 - 0{,}9920 = 0{,}008 \end{array}$$

Sen embargo, xa advertimos que imos simplificar o proceso identificando, sen máis, x con x'.

EXERCICIOS PROPOSTOS

1. A variable x é binomial, con $n = 1\,200$ e $p = 0{,}0008$.

a) Calcula a probabilidade de que x sexa maior que 100.

b) Determina o intervalo característico para unha probabilidade do 95%.

2. Se temos un dado correcto e o lanzamos 50 veces:

a) ¿Cal é a probabilidade de que "o 1" saia máis de 10 veces?

b) ¿Cal é a probabilidade de que saia "múltiplo de 3" ó menos 20 veces?

13.2 DISTRIBUCIÓN DAS PROPORCIÓNS MOSTRAIS

TEN EN CONTA

Nunha certa poboación, a proporción de individuos que posúe unha certa característica é p. Consideramos todas as posibles mostras de tamaño n que se poden extraer desa poboación. En cada unha das mostras haberá unha proporción pr de individuos con esa característica. ¿Como se distribúen todos os posibles valores de pr? É o que se estudia nesta páxina.

Estudio dun exemplo

O 15% dos mozos de 18 a 25 anos son miopes (este dato non é real). Propoñémonos elixir ó azar 40 mozos e preguntámonos qué proporción, pr, de miopes haberá nesa mostra.

Para cada individuo da mostra (aínda non extraída) a probabilidade de ser miope é $p = 0{,}15$.

Como na mostra hai $n = 40$ individuos, o número x de miopes segue unha distribución binomial $B(40;\ 0{,}15)$.

Posto que $np = 40 \cdot 0{,}15 = 6$ é maior que 5, $B(n, p) \sim N(np, \sqrt{npq})$. Por tanto, o número x de miopes en cada mostra aproxímase a unha distribución normal con:

$$\mu = 40 \cdot 0{,}15 = 6 \quad \text{e} \quad \sigma = \sqrt{40 \cdot 0{,}15 \cdot 0{,}85} = 2{,}26$$

x = "nº de miopes nunha mostra" é: $N(6;\ 2{,}26)$

A proporción de miopes nunha mostra é:

$$pr = \frac{\text{número de miopes na mostra}}{40} = \frac{x}{40}$$

Por tanto, pr é $N\left(\frac{6}{40}, \frac{2{,}26}{40}\right) = N(0{,}15;\ 0{,}0565)$.

Vimos, pois, que a proporción de miopes nunha mostra aínda non extraída, de tamaño 40, segue unha distribución normal de media $pr = p = 0{,}15$ e desviación típica $\sqrt{\frac{pq}{n}} = 0{,}0565$.

NOTA IMPORTANTE

A proporción pr é unha variable discreta. Para calcular probabilidades de pr (e intervalos característicos) deberían facerse as correccións correspondentes ó paso a unha continua. Sen embargo, para non complicar o proceso, trataremos pr como se fora continua.

Distribución das proporcións mostrais. Se nunha poboación a proporción de individuos que posúe unha certa característica C é p, a proporción, pr, de individuos con dita característica nas mostras de tamaño n segue unha distribución normal de media p e desviación típica $\sqrt{\frac{pq}{n}}$. É dicir, pr é $N\left(p, \sqrt{\frac{pq}{n}}\right)$.

Para xustificalo, repitamos o razoamento do exemplo en forma xeral.

x = número de individuos da mostra que teñen a característica C.

x é $B(n, p)$

Se $np \geq 5$ e $nq \geq 5$, entón x é $N(np, \sqrt{npq})$

pr = proporción de individuos da mostra que teñen a característica C.

Como $pr = \frac{\text{número de individuos coa característica C}}{\text{número de individuos da mostra}} = \frac{x}{n}$, a distribución de pr será como a de x, pero cos parámetros media e desviación típica divididos por n. Por tanto:

$$pr \text{ é } N\left(\frac{np}{n}, \frac{\sqrt{npq}}{n}\right) = N\left(p, \sqrt{\frac{pq}{n}}\right)$$

EXERCICIOS RESOLTOS

1. *Unha máquina produce parafusos. Sábese que o 5% deles son defectuosos. Empaquétanse en caixas de 400.*

a) ¿Como se distribúe a proporción pr de parafusos defectuosos nas caixas?

b) Encontra un intervalo no cal se encontre o 90% das proporcións de parafusos defectuosos.

c) Encontra un intervalo no cal se encontre o 99% das proporcións de parafusos defectuosos nas caixas.

a) A totalidade dos parafusos producidos pola máquina é a poboación.

A proporción de parafusos defectuosos na poboación é $p = 0{,}05$.

Cada caixa é unha mostra de 400 elementos: $n = 400$

Como acabamos de ver na páxina anterior, a proporción de parafusos defectuosos nas caixas segue unha distribución normal.

Media $p = 0{,}05$. Desviación típica $\sqrt{\dfrac{pq}{n}} = \sqrt{\dfrac{0{,}05 \cdot 0{,}95}{400}} = 0{,}011$

b) Unha probabilidade do 90% significa:

$$1 - \alpha = 0{,}9 \rightarrow \alpha/2 = 0{,}05 \rightarrow z_{\alpha/2} = 1{,}645$$

O intervalo correspondente é:

$$(0{,}05 \pm 1{,}645 \cdot 0{,}011) = (0{,}05 \pm 0{,}018) = (0{,}032;\ \ 0{,}068)$$

Isto quere dicir que o 90% das caixas teñen unha proporción de parafusos defectuosos comprendida entre 0,032 e 0,068.

c) O 99% significa: $1 - \alpha = 0{,}99 \rightarrow \alpha/2 = 0{,}005 \rightarrow z_{\alpha/2} = 2{,}575$

$$z_{\alpha/2} \cdot \sigma = 2{,}575 \cdot 0{,}011 = 0{,}028$$

O intervalo correspondente é:

$$(0{,}05 - 0{,}028;\ \ 0{,}05 + 0{,}028) = (0{,}022;\ \ 0{,}078)$$

2. *Supoñamos que o 15% dos mozos entre 18 e 25 anos son miopes.*

a) ¿Como se distribúe a proporción pr de mozos miopes en mostras de 40 individuos?

b) Calcula o intervalo característico das proporcións mostrais correspondente ó 80%.

a) $p = 0{,}15$; $n = 40$; media $= p = 0{,}15$; desviación típica $= \sqrt{\dfrac{pq}{n}} = 0{,}0565$

pr é normal $N(0{,}15;\ \ 0{,}0565)$.

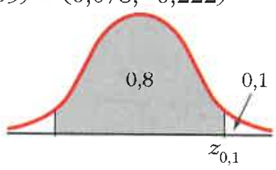
b) Comecemos por calcular o valor crítico $z_{\alpha/2}$ correspondente a $1 - \alpha = 0{,}8$:

$$P[z > z_{0,1}] = 0{,}1 \rightarrow P[z < z_{0,1}] = 0{,}9 \rightarrow z_{0,1} = 1{,}28$$

O valor crítico correspondente a $1 - \alpha = 0{,}8$ é $z_{\alpha/2} = 1{,}28$

Por tanto, o intervalo será $(0{,}15 \pm 1{,}28 \cdot 0{,}0565) = (0{,}078;\ \ 0{,}222)$

É dicir, o 80% das mostras de 40 mozos desas idades conteñen unha proporción de miopes comprendida entre 0,078 e 0,222.



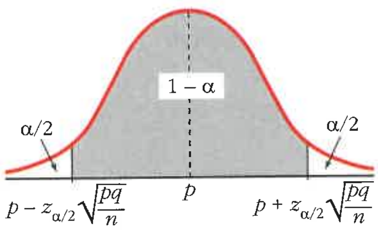
EXERCICIOS PROPOSTOS

1. Como sabemos, nun dado correcto a proporción de veces que sae o 5 é $1/6 = 0{,}1\widehat{6}$. Calcula cada un dos intervalos característicos correspondentes ó 90%, 95% e 99% para a "proporción de cincos", en series de 100 lanzamentos dun dado correcto.

13.3 INTERVALO DE CONFIANZA PARA UNHA PROPORCIÓN OU UNHA PROBABILIDADE

PROPORCIÓN ↔ PROBABILIDADE

Dicir que a **probabilidade** dun suceso é p é o mesmo que dicir que na poboación que resulta de repetir esa experiencia aleatoria unha infinidade de veces, a **proporción** de ocasións na que se dá o suceso é p. Por tanto, *unha probabilidade pode ser considerada como unha proporción nunha poboación infinita.*



Deséxase estimar a proporción, p, de individuos cunha certa característica que hai nunha poboación. Para iso recórrese a unha mostra de tamaño n, na que se obtén unha proporción mostral pr.

O **intervalo de confianza** de p cun nivel de confianza $(1-\alpha)\cdot 100\%$ é:

$$\left(pr - z_{\alpha/2}\cdot\sqrt{\frac{pr(1-pr)}{n}},\ pr + z_{\alpha/2}\cdot\sqrt{\frac{pr(1-pr)}{n}}\right)$$

Demostración

A proporción, pr, en mostras de tamaño n, distribúese segundo $N\left(p, \sqrt{\frac{pq}{n}}\right)$.

O intervalo característico de pr para unha probabilidade $1-\alpha$ é:

$$\left(p - z_{\alpha/2}\cdot\sqrt{\frac{pq}{n}},\ p + z_{\alpha/2}\cdot\sqrt{\frac{pq}{n}}\right)$$

É dicir:

$$P\left[|pr - p| < z_{\alpha/2}\cdot\sqrt{\frac{pq}{n}}\right] = 1-\alpha$$

Por tanto: $P\left[p \in \left(pr - z_{\alpha/2}\cdot\sqrt{\frac{pq}{n}},\ pr + z_{\alpha/2}\cdot\sqrt{\frac{pq}{n}}\right)\right] = 1-\alpha$

A igualdade anterior pretende servir para estimar o valor de p mediante un intervalo.

O erro máximo admisible, $E = z_{\alpha/2}\cdot\sqrt{\frac{pq}{n}}$, ten o grave inconveniente de que está dado en función de p. Por tanto, unha vez extraída a mostra e obtida a proporción mostral, pr, debemos estimar os valores de p e q así: $\hat{p} = pr$, $\hat{q} = 1 - pr$

Deste modo, o erro máximo admisible (**cota de erro**) para a estimación de p é:

$$E = z_{\alpha/2}\cdot\sqrt{\frac{pr(1-pr)}{n}}$$

Obtense así o **intervalo de confianza** de p.

No razoamento anterior hai que engadir que tanto $p\cdot n$ como $q\cdot n$ teñen que ser maiores ou iguais que 5 para que, realmente, pr se distribúa normal.

Ademais, para poder estimar p por pr na construción do erro máximo admisible, é necesario que a mostra sexa "grande" $(n \geq 30)$.

0,1 h: → 0,05 → 0,95 → 1,645

EXERCICIOS RESOLTOS

1. *Tomada unha mostra de 300 persoas maiores de 15 anos nunha gran cidade, encontrouse que 104 delas lían o xornal regularmente. Determina, cun nivel de confianza do 90%, un intervalo para estimar a proporción de lectores de xornais entre os maiores de 15 anos.*

Nivel de confianza do 90% $(1 - \alpha = 0{,}9) \rightarrow \alpha = 0{,}1 \rightarrow z_{\alpha/2} = 1{,}645$

A proporción mostral é $pr = \dfrac{104}{300} = 0{,}347$.

O erro máximo admisible (**cota de erro**) é:

$$E = z_{\alpha/2} \cdot \sqrt{\frac{pr(1-pr)}{n}} = 1{,}645 \cdot \sqrt{\frac{0{,}347 \cdot 0{,}653}{300}} = 0{,}045$$

O intervalo pedido é: $(0{,}347 - 0{,}045;\ 0{,}347 + 0{,}045) = (0{,}302;\ 0{,}392)$

Conclusión: Cun nivel de confianza do 90%, a proporción de lectores de xornais, no colectivo total, está entre 0,302 e 0,392.

2. *Á vista do resultado do problema anterior, preténdese repetir a experiencia para conseguir unha cota de erro de 0,01 co mesmo nivel de confianza do 90%. ¿Cantos individuos debe ter a mostra?*

Na expresión da cota do erro, $E = z_{\alpha/2} \cdot \sqrt{\dfrac{pr(1-pr)}{n}}$, temos que calcular n e coñecemos $z_{\alpha/2} = 1{,}645$, $E = 0{,}01$.

O valor de pr non podemos obtelo a partir da mostra, pois aínda non se extraeu (non esquezamos que estamos calculando o tamaño que debe ter). Tomaremos como referencia os resultados previos (problema anterior) e valoraremos $pr = 0{,}347$.

$$0{,}01 = 1{,}645 \cdot \sqrt{\frac{0{,}347 \cdot 0{,}653}{n}} \rightarrow n = \left(\frac{1{,}645}{0{,}01}\right)^2 \cdot 0{,}347 \cdot 0{,}653 = 6\,131{,}6$$

Haberá que tomar unha mostra de 6 132 persoas. Nela calculamos a proporción, pr_0, dos que len o xornal habitualmente. Con pr_0 formamos o intervalo $(pr_0 - 0{,}01;\ pr_0 + 0{,}01)$ dentro do cal se estima, cunha confianza do 90%, que estará a proporción p buscada.

3. *A partir dunha mostra de 100 individuos estimouse unha proporción mediante o intervalo de confianza (0,17; 0,25).*

¿Cal é o nivel de confianza co que se fixo a estimación?

A partir do intervalo obtemos:

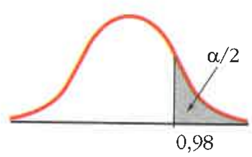
pr é o punto medio do intervalo: $pr = 0{,}21$

E é a metade da lonxitude do intervalo: $E = 0{,}04$

$$E = z_{\alpha/2} \cdot \sqrt{\frac{pr(1-pr)}{n}} \rightarrow 0{,}04 = z_{\alpha/2} \cdot \sqrt{\frac{0{,}21 \cdot 0{,}79}{100}} \rightarrow z_{\alpha/2} = 0{,}98$$

$$P[z > z_{\alpha/2}] = \frac{\alpha}{2} \rightarrow P[z > 0{,}98] = 1 - \phi(0{,}98) = 0{,}1635$$

$$\frac{\alpha}{2} = 0{,}1635 \rightarrow \alpha = 0{,}3270 \rightarrow 1 - \alpha = 0{,}6730$$



A estimación de p mediante (0,17; 0,25) realizouse cun nivel de confianza do 67,30%.

EXERCICIOS PROPOSTOS

1. Lanzouse un dado 400 veces e obtívose 72 veces o valor 4. Estima o valor da probabilidade $P[4]$ cun nivel de confianza do 90%.

2. ¿Cantas veces temos que lanzar un dado, que supoñemos levemente incorrecto, para estimar a probabilidade de "6" cun erro menor que 0,002 e un nivel de confianza do 95%?

EXERCICIOS E PROBLEMAS RESOLTOS

1. Distribución das proporcións mostrais

Nunha empresa que fabrica microcircuítos comprobouse que o 4% destes son defectuosos. Un cliente vai comprar un paquete de 500 microcircuítos procedentes da fábrica. Determina:

a) O número esperado de microcircuítos non defectuosos nun paquete de 500.

b) A distribución da proporción de microcircuítos defectuosos nas caixas de 500 microcircuítos.

c) A probabilidade de que o número de microcircuítos defectuosos (nun paquete de 500) estea entre 20 e 30.

a) A probabilidade de que un microcircuíto sexa defectuoso é $p = 0{,}04$. O número de microcircuítos defectuosos nun paquete de 500 segue unha distribución binomial $B(500;\ 0{,}04)$. O número esperado de microcircuítos defectuosos é $500 \cdot 0{,}04 = 20$. Por tanto, o número esperado de microcircuítos non defectuosos é: $500 - 20 = 480$

b) A distribución das proporcións de microcircuítos defectuosos en mostras de tamaño $n = 500$ é unha normal de media $p = 0{,}04$ e desviación típica $\sqrt{\dfrac{pq}{n}} = \sqrt{\dfrac{0{,}04 \cdot 0{,}96}{500}} = 0{,}00876$.

É dicir, pr é $N(0{,}04;\ 0{,}00876)$.

c) Dicir que o número de circuítos defectuosos nun paquete de 500 estea entre 20 e 30 é o mesmo que dicir que a súa proporción está entre 20/500 e 30/500. É dicir, que $pr \in (0{,}04;\ 0{,}06)$.

$$P[0{,}04 < pr < 0{,}06] = P\left[\frac{0{,}04 - 0{,}04}{0{,}00876} < z < \frac{0{,}06 - 0{,}04}{0{,}00876}\right] =$$

$$= P[0 < z < 2{,}28] = \phi(2{,}28) - 0{,}5 =$$

$$= 0{,}9887 - 0{,}5 = 0{,}4887$$

2. Estimación dunha probabilidade

Lanzouse 100 veces unha moeda obténdose 62 caras. Estima a probabilidade de "cara" mediante intervalos de confianza:

a) do 90%.

b) do 95%.

c) do 99%.

Consideramos unha poboación infinita consistente nunha serie reiterada de lanzamentos con esa moeda. A proporción, p, de caras nesa poboación é a probabilidade de "cara" para esa moeda.

Dispoñemos dunha mostra de 100 individuos coa cal obtemos o parámetro $pr = 0{,}62$ a partir do cal estimaremos p.

A desviación típica é: $s = \sqrt{\dfrac{pr(1-pr)}{n}} = \sqrt{\dfrac{0{,}62 \cdot 0{,}38}{100}} = 0{,}0485$

a) $1 - \alpha = 0{,}90 \rightarrow z_{\alpha/2} = 1{,}645$

A cota de erro é: $E = 1{,}645 \cdot 0{,}0485 = 0{,}080$

Co que o intervalo de confianza correspondente a un nivel de confianza do 90% queda:

$(0{,}62 \pm 0{,}08) = (0{,}54;\ 0{,}70)$

b) $1 - \alpha = 0{,}95 \rightarrow z_{\alpha/2} = 1{,}96$

A cota de erro é: $E = 1{,}96 \cdot 0{,}0485 = 0{,}095$

Intervalo de confianza:

$(0{,}62 \pm 0{,}095) = (0{,}525;\ 0{,}715)$

c) $1 - \alpha = 0{,}99 \rightarrow z_{\alpha/2} = 2{,}575$

A cota de erro é: $E = 2{,}575 \cdot 0{,}0485 = 0{,}125$

Intervalo de confianza: $(0{,}62 \pm 0{,}125) = (0{,}495;\ 0{,}745)$

3. Tamaño da mostra para estimar unha proporción

Baseándonos na experiencia do problema anterior, pretendemos estimar a probabilidade de "cara" cun erro menor que 0,002 e un nivel de confianza do 95%.

¿Cantas veces teremos que lanzar a moeda?

Na expresión da cota de erro, $E = z_{\alpha/2} \cdot \sqrt{\dfrac{pr(1-pr)}{n}}$, coñecemos:

$E = 0{,}002$, $z_{\alpha/2} = 1{,}96$ (correspondente a $1 - \alpha = 0{,}95$), $pr = 0{,}62$ tomado da experiencia anterior.

De aí podemos despexar n:

$$n = \frac{{z_{\alpha/2}}^2 \cdot pr(1-pr)}{0{,}002^2} = \frac{1{,}96^2 \cdot 0{,}62 \cdot 0{,}38}{0{,}000004} = 226\,270{,}24$$

Teriamos que lanzar a moeda 226 271 veces para estimar a probabilidade con menos de dúas milésimas de erro e cun nivel de confianza do 95%.

Se a proporción de caras obtida nesa mostra é pr_0, aseguraríamos cunha confianza do 95% que a probabilidade de "cara" estaría no intervalo $(pr_0 - 0{,}002;\ pr_0 + 0{,}002)$.

4. Estimación dunha proporción (resolución do problema inicial)

Para estimar o número de peixes que hai nun pantano procédese do seguinte modo: péscase con rede unha certa cantidade deles, 349, márcanse (hai unhas tintas indelebles que resisten á auga) e devólvense ó pantano. Ó cabo de varios días vólvese pescar outro montón deles e obsérvase qué proporción están marcados. Nesta segunda pesca obtivéronse 514 peixes, dos cales hai 37 marcados.

a) Calcula un intervalo de confianza, ó 90%, para a proporción de peixes marcados no pantano.

b) Calcula un intervalo de confianza, ó 90%, para o total de peixes do pantano.

a) A mostra consta de 514 peixes. Deles, hai 37 marcados.

A proporción de peixes marcados na mostra é: $pr = \dfrac{37}{514} = 0{,}072$.

Para calcular un intervalo de confianza para a proporción p na poboación, a partir da proporción pr obtida na mostra, o máximo erro admisible é:

$$E = z_{\alpha/2} \cdot \sqrt{\frac{pr(1-pr)}{n}} = 1{,}645 \cdot \sqrt{\frac{0{,}72 \cdot 0{,}928}{514}} = 0{,}019$$

Por tanto, o intervalo de confianza para p, ó 90%, é:

$(0{,}072 \pm 0{,}019) = (0{,}053;\ 0{,}091)$

b) Para determinar o intervalo de confianza para o número total, N, de peixes no pantano, temos en conta que a proporción de peixes marcados é $p = \dfrac{349}{N}$.

$$\left.\begin{array}{l} 0{,}053 = \dfrac{349}{N_1} \ \rightarrow\ N_1 = \dfrac{349}{0{,}053} \approx 6\,585 \\[2ex] 0{,}091 = \dfrac{349}{N_2} \ \rightarrow\ N_2 = \dfrac{349}{0{,}091} \approx 3\,835 \end{array}\right\}$$

Observa que aquí invértense os extremos do intervalo:

Temos un nivel de confianza do 90% de que o número de peixes do pantano está no intervalo [3 835, 6 585].

EXERCICIOS E PROBLEMAS PROPOSTOS

PARA PRACTICAR

Distribución das proporcións mostrais. Intervalos característicos

1 Indaga cómo se distribúen as proporcións mostrais, *pr,* para as poboacións e as mostras que se describen a continuación:

	a)	b)	c)	d)	e)	f)
PROPORCIÓN, p, NA POBOACIÓN	0,5	0,6	0,8	0,1	0,05	0,15
TAMAÑO, n, DA MOSTRA	10	20	30	50	100	100

2 Calcula os intervalos característicos para as proporcións mostrais do exercicio anterior, correspondentes ás probabilidades que, en cada caso, se indican:

a) 90% b) 95% c) 99%

d) 95% e) 99% f) 80%

3 **S** Catro de cada dez habitantes dunha determinada poboación le habitualmente o xornal Z.

Determina o intervalo característico para a proporción de habitantes desa poboación que len o xornal Z, en mostras de tamaño 49, correspondente ó 95%.

4 Nun saco mesturamos feixóns brancos e feixóns pintos na relación de 14 brancos por cada pinto.

Extraemos un puñado de 100 feixóns.

a) ¿Cal é a probabilidade de que a proporción de feixóns pintos estea comprendida entre 0,05 e 0,1?

b) Obtén un intervalo no cal se encontre o 99% das proporcións das mostras de tamaño 100.

5 Nunha localidade de 6 000 habitantes, a proporción de menores de 16 anos é de 1 500/6 000.

a) ¿Cal é a distribución da proporción de menores de 16 anos en mostras de 50 habitantes de dita poboación?

b) Calcula a probabilidade de que, nunha mostra de 50 habitantes, haxa entre 15 e 20 menores de 16 anos.

6 O 42% dos habitantes dun municipio é contrario á xestión do alcalde e o resto son partidarios deste. Se se toma unha mostra de 64 individuos, ¿cal é a probabilidade de que gañen os que se opoñen ó alcalde?

7 A probabilidade de que un bebé sexa home é 0,515. Se naceron 184 bebés, ¿cal é a probabilidade de que haxa 100 homes ou máis?

Calcula o intervalo característico correspondente ó 95% para a proporción de homes en mostras de 184 bebés.

Intervalos de confianza

8 Realizóuselles unha enquisa a 350 familias preguntando se posuían ordenador na casa, encontrándose que 75 delas posuíano. Estima a proporción real das familias que dispoñen de ordenador cun nivel de confianza do 95%.

9 **S** Selecciónase aleatoriamente unha mostra de 600 persoas nunha cidade e pregúntaselles se consideran que o tráfico nela é aceptablemente fluído. Responden afirmativamente 250 persoas.

¿Cal é o intervalo de confianza da proporción de cidadáns desa cidade que consideran aceptable a fluidez do tráfico, cun nivel de confianza do 90%?

PARA RESOLVER

10 Sabemos que ó lanzar ó chan 100 chinchetas, no 95% dos casos, a proporción delas que quedan coa punta cara arriba está no intervalo (0,1216; 0,2784). Calcula a probabilidade p de que unha desas chinchetas caia coa punta cara arriba e comproba que a amplitude do intervalo dado é correcta.

11 **S** De 120 alumnos, a proporción de que teñan dous ou máis irmáns é de 48/120. Indica os parámetros da distribución á que se axustarían as mostras de tamaño 30.

12 ¿De que tamaño convén tomar a mostra dunha liña de produción para ter unha confianza do 95% de que a proporción estimada non difire da verdadeira en máis dun 4%? Sábese, por estudios previos, que a proporción de obxectos defectuosos é da orde do 0,05.

13 S Deséxase estimar a proporción, p, de individuos daltónicos dunha poboación a través da porcentaxe observada nunha mostra aleatoria de individuos, de tamaño n.

a) Se a porcentaxe de individuos daltónicos na mostra é igual ó 30%, calcula o valor de n para que, cun nivel de confianza de 0,95, o erro cometido na estimación sexa inferior ó 3,1%.

b) Se o tamaño da mostra é de 64 individuos, e a porcentaxe de individuos daltónicos na mostra é do 35%, determina, usando un nivel de significación do 1%, o correspondente intervalo de confianza para a proporción de daltónicos da poboación.

14 Nunha mostra de 100 rótulos publicitarios obsérvase que aparecen 6 defectuosos.

a) Estima a proporción real de rótulos defectuosos, cun nivel de confianza do 99%.

b) ¿Cal é o erro máximo cometido ó facer a estimación anterior?

c) ¿De que tamaño teriamos que coller a mostra, cun nivel de confianza do 99%, para obter un erro inferior a 0,05?

15 S Tomada ó azar unha mostra de 60 estudiantes dunha universidade, encontrouse que un tercio falaban o idioma inglés.

a) Determina, un nivel de confianza do 90%, un intervalo para estimar a proporción de estudiantes que falan o idioma inglés entre os estudiantes desa universidade.

b) Á vista do resultado anterior preténdese repetir a experiencia para conseguir unha cota de erro de 0,01 co mesmo nivel de confianza do 90%. ¿Cantos individuos debe ter a mostra?

16 S Para estimar a proporción de habitantes dunha determinada cidade que posúen ordenador persoal, quérese utilizar unha mostra aleatoria de tamaño n. Calcula o valor mínimo de n para garantir que, cun nivel de confianza do 95%, o erro na estimación non sexa superior ó 2%.

☛ *Como se descoñece a proporción, tense que tomar o caso máis desfavorable, que será 0,5.*

17 S Nunha enquisa realizada a 800 persoas elixidas ó azar do censo electoral, 240 declaran a súa intención de votar ó partido A.

a) Estima, cun nivel de confianza do 95,45%, entre qué valores se encontra a intención de voto ó susodito partido en todo o censo.

b) Discute, razoadamente, o efecto que tería sobre o intervalo de confianza o aumento, ou diminución, do nivel de confianza.

18 S Unha recente enquisa, realizada nun certo país sobre unha mostra aleatoria de 800 persoas, destaca o dato de que 300 delas son analfabetas. Para estimar a proporción de analfabetos do país obtivemos o seguinte intervalo de confianza: (0,3414; 0,4086)

¿Cal é o nivel de confianza co que se fixo a estimación?

CUESTIÓNS TEÓRICAS

19 A partir dunha mostra de tamaño 400 estímase a proporción de individuos que len o xornal nunha gran cidade. Obtense unha cota de erro de 0,0392 cun nivel de confianza do 95%.

a) ¿Poderiamos, coa mesma mostra, mellorar o nivel de confianza na estimación? ¿Que lle ocorrería á cota de erro?

b) ¿Saberías calcular a proporción, pr, obtida na mostra?

PARA PROFUNDAR

20 a) Un fabricante de medicamentos afirma que certa medicina cura unha enfermidade do sangue no 80% dos casos. Os inspectores de sanidade utilizan o medicamento nunha mostra de 100 pacientes e deciden aceptar dita afirmación se se curan 75 ou máis.

Se o que afirma o fabricante é realmente certo, ¿cal é a probabilidade de que os inspectores rexeiten dita afirmación?

b) Se na mostra se curan 60 individuos, cunha confianza do 95%, ¿cal é o erro máximo cometido ó estimar que a porcentaxe de efectividade do medicamento é do 60%?

14 INFERENCIA ESTATÍSTICA: CONTRASTES DE HIPÓTESES

En 1710, o médico inglés **John Arbuthnot** estudiou o sexo das criaturas nacidas nunha certa localidade durante os 82 anos anteriores e advertiu que a proporción de homes foi sempre superior á de mulleres. Con isto rebateu a crenza de que é igualmente probable que naza un home ou unha muller, argumentando do seguinte modo: "O resultado non pode ser casual, xa que, facendo corresponder HOME e MULLER a CARA e CRUZ, dunha moeda, é absurdo pensar que exista tal exceso de homes".

Aínda que formulado de forma matematicamente insuficiente, pode considerarse este o primeiro test de hipóteses da historia. (Arbuthnot, ademais de médico, foi o creador do personaxe de *John Bull*, que personifica o pobo inglés do mesmo modo que o *tío Sam* personifica o estadounidense).

Nun test de hipóteses emítese unha afirmación estatística (relativa ó valor dun parámetro dunha poboación) e mediante unha mostra estúdiase se dita afirmación (hipótese) é compatible co resultado da experiencia (contraste).

Os test de hipóteses foron creados por **Neyman** e **E. Pearson** cara a 1940 e desenvolvidos posteriormente por **Abraham Wald**.

PARA EMPEZAR, REFLEXIONA E RESOLVE

Máquina empaquetadora

O fabricante dunha máquina empaquetadora afirma que, se se regula para que empaquete sacos con 100 kg, os pesos dos sacos envasados por ela distribuiranse $N(100,2)$.

a) Para probala, o posible comprador efectúa un empaquetado que resulta ter 101 kg. Á vista deste resultado, ¿debería desconfiar da afirmación do fabricante?

b) Para probar a máquina, empaquétanse 50 sacos. O peso medio dos 50 é 101 kg. Con este resultado, ¿cres que se debería rexeitar o que afirma o fabricante?

Pilas que duran e duran...

Unha fabricante de pilas afirma que a duración media das súas pilas, funcionando ininterrompidamente, é de 53 horas como mínimo e a súa desviación típica, de 4 horas.

Di se rexeitarías ou non dita afirmación en cada un dos seguintes casos:

a) Se unha pila dura 48 horas.

b) Se a media das duracións de 100 pilas é 50 horas.

c) Se a media das duracións de 100 pilas é 56 horas.

¿Moedas falsas?

Reflexionemos sobre cada unha das seguintes experiencias:

a) Lanzamos unha moeda 10 veces e obtemos 6 caras.

b) Lanzamos unha moeda 100 veces e obtemos 60 caras.

c) Lanzamos unha moeda 1 000 veces e obtemos 600 caras.

- ¿Podemos deducir dalgunha delas que a moeda é incorrecta? ¿Con cal delas chegamos a esa conclusión con máis seguridade? (Responde intuitivamente).

A graxa no leite

Os fabricantes dunha determinada marca de leite afirman que o contido de materias graxas, por termo medio, é do 12% ou menos. A desviación típica é 2,2%.

Para estudiar se é certa ou non a afirmación dos fabricantes, tómanse 50 envases e mídese a porcentaxe de graxa que hai no leite de cada un deles, obténdose unha media de 12,6%.

- Reflexionando sobre o resultado anterior, ¿cres que se debe rexeitar a hipótese feita pola empresa distribuidora de que o contido de graxas por termo medio non supera o 12%, ou ben non hai motivo suficiente para rexeitala?

NESTA UNIDADE VERÁS

A **estatística inferencial** ten por obxecto o desenvolvemento de técnicas que permiten coñecer ou comprobar o valor dos parámetros dunha poboación a partir dos datos obtidos dunha mostra. Os resultados que se obteñen con estes métodos teñen un certo grao de incerteza que se mide en termos de probabilidade.

$$\boxed{\text{MOSTRA}} \xrightarrow[\text{Inferimos}]{} \boxed{\text{POBOACIÓN}}$$

A estatística inferencial ten dúas grandes ramas:

Estatística inductiva, que ten por obxecto estimar os parámetros da poboación. Dela ocupámonos nas dúas unidades anteriores: a media (unidade 12) e unha proporción (unidade 13).

O valor do parámetro (media, proporción) estímase mediante un intervalo de confianza e cun certo nivel de confianza.

Estatística hipotético-deductiva, que ten por obxecto comprobar, mediante métodos matemáticos, **hipóteses** realizadas sobre o valor dalgún parámetro da poboación a partir dunha mostra aleatoria extraída dela. As catro experiencias descritas nestas dúas páxinas son dese tipo:

$$\boxed{\text{HIPÓTESE SOBRE UN PARÁMETRO DA POBOACIÓN}} \xrightarrow[\text{MOSTRA}]{} \boxed{\text{ACEPTAMOS OU REXEITAMOS A HIPÓTESE}}$$

Nesta unidade dedicarémonos a aprender os pasos necesarios para **tomar decisións** sobre hipóteses estatísticas sinxelas relativas a valores da media ou dunha proporción da poboación.

A esta rama da estatística inferencial chámaselle, tamén, teoría da decisión.

14.1 HIPÓTESES ESTATÍSTICAS

Comecemos analizando dous casos concretos.

CASO 1

Temos un dado que supoñemos correcto. Lanzámolo 100 veces e obtemos 25 "CINCOS". ¿Podemos dar por válida a suposición (o dado é correcto) ou debemos rectificala en vista dos resultados?

Neste exemplo dubídase sobre se o parámetro $P[5] = p$ toma o valor de 1/6. Para saír de dúbidas pódese realizar, a partir do resultado dos 100 lanzamentos, **un test estatístico**.

Un **test estatístico** é un procedemento para, a partir dunha mostra aleatoria e significativa, extraer conclusións que permitan aceptar ou rexeitar unha hipótese previamente emitida sobre o valor dun parámetro descoñecido desa poboación.

CASO 2

Hai cinco anos realizóuselle unha proba de coñecementos á totalidade dos soldados dunha quinta. O resultado foi unha media $\mu = 102$ puntos e unha desviación típica $\sigma = 11$. Este ano pasóuselle o mesmo test a unha mostra de 400 soldados e a media foi $\bar{x} = 101$ puntos.

¿Podemos supoñer que non houbo cambio nos coñecementos dos soldados nestes cinco anos e que, por tanto, a diferencia observada é froito do azar?

En ambos os exemplos hai unha **hipótese de partida** e uns **resultados,** obtidos a partir dunha mostra, que difiren da hipótese. E preguntámonos se a diferencia é atribuíble ó azar.

	HIPÓTESE	RESULTADO A PARTIR DA MOSTRA	INTERROGANTE
CASO 1: DADO	O dado é correcto. A probabilidade de "CINCO" é: $p = 1/6 = 0{,}167$	Proporción de "CINCOS" na mostra: $pr = 0{,}25$	A diferencia observada, ¿pode atribuírse ó azar? ¿Podemos supoñer, razoablemente, que a mostra foi extraída da poboación sobre a que fixemos a hipótese?
CASO 2: SOLDADOS	Actualmente tamén é: $\mu = 102$	Media da mostra: $\bar{x} = 101$	

A hipótese emitida desígnase por H_0 e chámase **hipótese nula.** A hipótese contraria desígnase por H_1 e chámase **hipótese alternativa**.

CASO 1 (DADO): H_0: $p = 0{,}167$, H_1: $p \neq 0{,}167$

CASO 2 (SOLDADOS): H_0: $\mu = 102$, H_1: $\mu \neq 102$

NOTA: Dicir que nun dado a probabilidade de que saia "5" é 1/6 é o mesmo que dicir que na poboación resultante de lanzar o dado unha infinidade de veces a proporción de "5" é 1/6. Por tanto, compárase o valor da proporción de "CINCOS" na poboación coa proporción de "CINCOS" na mostra.

Contraste de hipóteses

Para dilucidar se unha hipótese estatística é ou non certa danse unha serie de pasos que imos estudiar. Empezaremos resolvendo os dous exemplos presentados na páxina anterior.

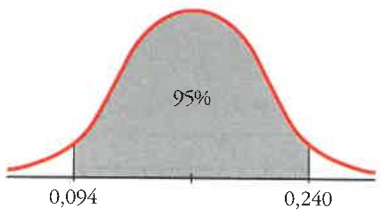
CASO 1: *¿Podemos supoñer que é correcto un dado que, ó lanzalo 100 veces, dá 25 "CINCOS"?*

A hipótese emitida (o dado é correcto) concrétase dicindo que:

$$p = P[5] = 1/6 = 0{,}167$$

Se a hipótese fora certa, entón as proporcións, pr, de "CINCOS" nas mostras de tamaño 100 seguirían unha distribución normal:

$$N\left(p,\ \sqrt{\frac{pq}{n}}\right) = N\left(0{,}167;\ \sqrt{\frac{0{,}167 \cdot 0{,}833}{100}}\right) = N(0{,}167;\ 0{,}037)$$



En tal caso, o 95% das proporcións mostrais de "CINCOS" estarían no intervalo característico correspondente a $1 - \alpha = 0{,}95$:

$$0{,}167 \pm 1{,}96 \cdot 0{,}037 = (0{,}094;\ 0{,}240)$$

(0,094; 0,240) chámase **zona de aceptación.**

Posto que a proporción obtida na mostra, $pr = 0{,}25$, queda fóra do intervalo, consideramos que sería moi improbable que a mostra procedera desta distribución (é dicir, é moi improbable que cun dado correcto se obteñan vintecinco "CINCOS" en 100 tiradas). Por tanto, **rexéitase a hipótese** cun **nivel de significación** do 5%.

NOMENCLATURA

O **nivel de significación** dunha hipótese, α, é o valor complementario do **nivel de confianza** dunha estimación, $1 - \alpha$.

CASO 2: *¿Podemos supoñer que a variable "coñecemento dos soldados" ten os parámetros $\mu = 102$, $\sigma = 11$, tendo en conta que a media de 400 deles é $\bar{x} = 101$?*

Se a hipótese é certa, a poboación de partida ten $\mu = 102$, $\sigma = 11$.

As medias das mostras de tamaño $n = 400$ distribúense así:

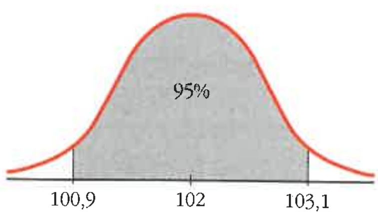
$$\bar{x} \text{ é } N\left(102,\ \frac{11}{\sqrt{400}}\right) = N(102;\ 0{,}55)$$

Obtemos o seguinte intervalo característico ó 95%:

$$\alpha = 0{,}05 \quad \to \quad z_{\alpha/2} = 1{,}96$$

INTERVALO: $(102 \pm 1{,}96 \cdot 0{,}55) = (100{,}9;\ 103{,}1)$

O 95% das mostras de tamaño 400 teñen a súa media neste intervalo (zona de aceptación para $\alpha = 0{,}05$).



$\bar{x} = 101$ está no intervalo. Por tanto, acéptase a hipótese cun **nivel de significación** $\alpha = 0{,}05$, pois só rexeitamos como improbables o 5% dos casos máis extremos.

Método para contrastes de hipóteses estatísticas

1º **Enunciación.** Enúnciase a hipótese nula, H_0. Consiste en atribuírlle un valor a un parámetro de certa poboación. (Neste curso só emitiremos hipóteses sobre a media, μ, e sobre a proporción, p).

2º **Deducción de conclusións.** Se a hipótese nula fora certa, tal parámetro da mostra distribuiríase de forma coñecida. En consecuencia:

— Elíxese un **nivel de significación.** Son os máis comúns $\alpha = 0{,}10$; $\alpha = 0{,}05$; $\alpha = 0{,}01$.

— Constrúese a **zona de aceptación.** É o intervalo fóra do cal só se encontran o $\alpha \cdot 100\%$ dos "casos máis raros".

3º **Verificación.** Extráese unha mostra, o seu tamaño decidímolo no paso anterior, e dela obténse o correspondente parámetro.

4º **Decisión.** Se o valor do parámetro mostral cae dentro da zona de aceptación, acéptase a hipótese cun nivel de significación α. Senón, rexéitase.

EXERCICIOS RESOLTOS

1. *Explicita os catro pasos anteriores na resolución dos dous problemas estudiados na páxina anterior.*

CASO 1

1º H_0: $p = 0{,}167$; H_1: $p \neq 0{,}167$

2º As proporcións mostrais, pr, distribuiríanse:

$$N(0{,}167;\ 0{,}037)$$

Nivel de significación: $\alpha = 0{,}05$

Zona de aceptación:

$$(0{,}094;\ 0{,}240)$$

3º Extráese a mostra e calcúlase o valor do parámetro:

$$pr = \frac{25}{100} = 0{,}25$$

4º 0,25 non está na zona de aceptación. Por tanto, rexéitase a hipótese.

O dado non é correcto.

CASO 2

1º H_0: $\mu = 102$; H_1: $\mu \neq 102$

2º As medias mostrais, $\bar{x}$, distribuiríanse:

$$N(102;\ 0{,}55)$$

Nivel de significación: $\alpha = 0{,}05$

Zona de aceptación:

$$(100{,}9;\ 103{,}1)$$

3º Extráese a mostra e calcúlase o valor do parámetro:

$$\bar{x} = 101$$

4º 101 si está na zona de aceptación. Acéptase a hipótese.

Os coñecementos dos soldados son os mesmos que hai cinco anos.

EXERCICIOS PROPOSTOS

1. Repite, paso a paso, o CASO 1 para un nivel de significación $\alpha = 0{,}01$.

2. Repite, paso a paso, o CASO 2 para un nivel de significación $\alpha = 0{,}10$.

14.2 CONTRASTES DE HIPÓTESES PARA A MEDIA

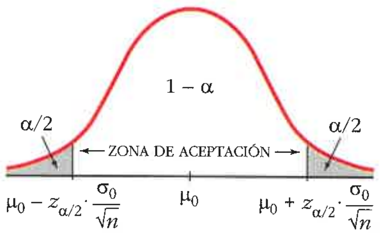
Imos sistematizar os pasos que se dan para realizar contrastes de hipóteses sobre a media da poboación, distinguindo os casos en que a hipótese nula é do tipo $\mu = \mu_0$ daqueles outros nos que na hipótese nula se aceptan desigualdades: $\mu \geq \mu_0$ ou ben $\mu \leq \mu_0$.

Contraste bilateral: $\mu = \mu_0$

1º paso. Hipótese H_0: $\mu = \mu_0$, H_1: $\mu \neq \mu_0$

A hipótese nula consiste en atribuírlle un certo valor á media da poboación. Este método é o que se desenvolve no CASO 2 das páxinas anteriores.

2º paso. Deducción de conclusións. Obtención da zona de aceptación



VALORES CRÍTICOS MÁIS USUAIS			
α	0,10	0,05	0,01
$z_{\alpha/2}$	1,645	1,96	2,575

Ben cando $n \geq 30$ ou ben para calquera valor de n se a poboación de partida é normal, as medias mostrais distribúense $N(\mu_0, \sigma_0/\sqrt{n})$.

Por tanto, a zona de aceptación para un nivel de significación α é o intervalo característico correspondente:

$$\left(\mu_0 - z_{\alpha/2} \cdot \frac{\sigma_0}{\sqrt{n}},\ \mu_0 + z_{\alpha/2} \cdot \frac{\sigma_0}{\sqrt{n}}\right)$$

3º e 4º pasos. Verificación e decisión

Extráese a mostra, calcúlase $\bar{x}$ e compróbase se está dentro ou fóra da zona de aceptación.

EXERCICIOS RESOLTOS

1. *Crese que o cociente intelectual medio dos estudiantes dunha universidade é 113, cunha desviación típica de 7. Para contrastar a hipótese, extráese unha mostra de 180 estudiantes e obtense neles un cociente intelectual medio de 115.*

¿Podemos aceptar a hipótese cun nivel de significación do 5%?

1º paso. H_0: $\mu = 113$, H_1: $\mu \neq 113$

2º paso. Posto que o tamaño da mostra, $n = 180$, é maior que 30, as medias mostrais distribuiríanse así (se a hipótese fora certa), $N\left(113, \frac{7}{\sqrt{180}}\right) = N(113;\ 0{,}52)$. Construímos o intervalo de aceptación:

$\alpha = 0{,}05 \rightarrow z_{\alpha/2} = 1{,}96$. Intervalo: $113 \pm 1{,}96 \cdot 0{,}52 = (111{,}98;\ 114{,}02)$

3º paso. Nunha experiencia real, sería agora cando se extraera a mostra e se obtería nela o valor da media $\bar{x} = 115$.

4º paso. Posto que 115 non pertence á zona de aceptación, *rexeitamos a hipótese*. É dicir, non podemos dar por bo que o C.I. medio dos estudiantes desa universidade sexa 113.

EXERCICIOS PROPOSTOS

1. a) Nunha poboación para a cal é $\sigma = 29$, contrasta a hipótese de que $\mu = 347$, cun nivel de significación do 1%, mediante unha mostra de 200 individuos na que se obtén $\bar{x} = 352$.

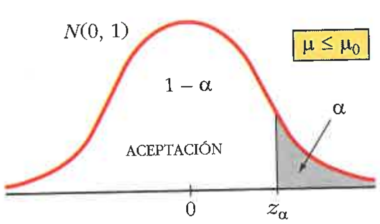
b) Repite o contraste para $\alpha = 10\%$.

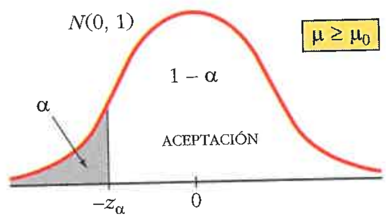
Contraste unilateral: $\mu \leq \mu_0$ o $\mu \geq \mu_0$

1º paso. Hipótese $H_0(\mu \leq \mu_0$ ou ben $\mu \geq \mu_0)$

As correspondentes hipóteses alternativas son $\mu > \mu_0$ ou, respectivamente, $\mu < \mu_0$.

2º paso. Zona de aceptación





Nestes casos, toda a cola (intervalo de non aceptación ou de rexeitamento) está nun dos extremos da distribución. Os **valores críticos** da táboa da esquerda obtéñense directamente da táboa $N(0, 1)$.

A zona de aceptación é:

HIPÓTESE	ZONA DE ACEPTACIÓN	H. ALTERNATIVA
$\mu \geq \mu_0$	$\left(\mu_0 - z_\alpha \cdot \dfrac{\sigma_0}{\sqrt{n}}, +\infty\right)$	$\mu < \mu_0$
$\mu \leq \mu_0$	$\left(-\infty, \mu_0 + z_\alpha \cdot \dfrac{\sigma_0}{\sqrt{n}}\right)$	$\mu > \mu_0$

3º e 4º pasos. Verificación e decisión

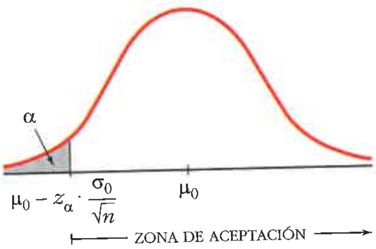
Extráese a mostra, calcúlase $\bar{x}$ e compróbase se está dentro ou fóra da zona de aceptación.

ATENCIÓN

Nos contrastes unilaterais os valores críticos son z_α.

Estes son os máis usuais:

VALORES CRÍTICOS			
α	0,10	0,05	0,01
z_α	1,28	1,645	2,33



EXERCICIOS RESOLTOS

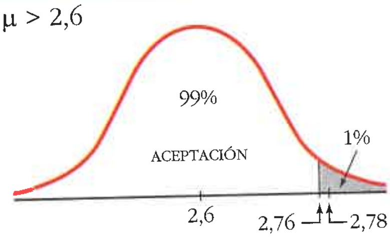
1. *O peso dos polos dunha granxa é normal con media 2,6 kg e desviación típica 0,5. Experiméntase un novo tipo de alimentación con 50 crías. Cando se fan adultos, pésanse e obtense unha media de 2,78 kg. Imos contrastar a hipótese de que o peso medio da poboación non aumenta, cun nivel de significación do 1%.*

H_0: $\mu \leq 2{,}6$. Hipótese alternativa H_1: $\mu > 2{,}6$

$\alpha = 0{,}01 \rightarrow z_\alpha = 2{,}33$

Zona de aceptación:

$$\left(-\infty;\ 2{,}6 + 2{,}33 \cdot \frac{0{,}5}{\sqrt{50}}\right) = (-\infty;\ 2{,}76)$$



O valor obtido mediante a mostra, 2,78 kg, queda fóra da zona de aceptación.

Por tanto, rexéitase H_0 e acéptase H_1 (a poboación aumentará de peso co novo tipo de alimentación) cun nivel de significación de 0,01.

EXERCICIOS PROPOSTOS

2. Nunha poboación para a cal é $\sigma = 29$, contrasta a hipótese de que $\mu \leq 347$ cun nivel de significación do 1%, mediante unha mostra de 200 individuos na que se obtén $\bar{x} = 352$.

14.3 CONTRASTES DE HIPÓTESES PARA A PROPORCIÓN

TEN EN CONTA

p pode ser unha **proporción** ou unha **probabilidade**.

Contraste bilateral: $p = p_0$

A hipótese nula consiste en atribuírlle un valor á proporción de individuos que teñen unha certa característica ou á probabilidade dun suceso nunha experiencia aleatoria.

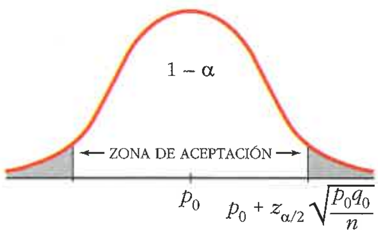
1º paso. Hipótese. H_0: $p = p_0$; H_1: $p \neq p_0$

2º paso. Zona de aceptación

Sabemos que se $p = p_0$ entón as proporcións mostrais, pr, distribúense $N\left(p_0, \sqrt{\dfrac{p_0 q_0}{n}}\right)$.

Por tanto, a zona de aceptación para un nivel de significación α é o intervalo característico correspondente:

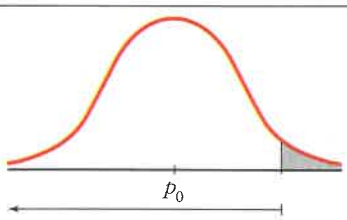
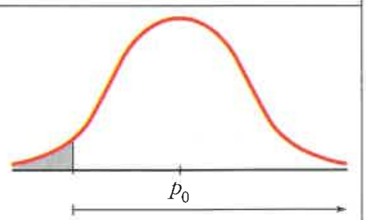
$$\left(p_0 - z_{\alpha/2} \cdot \sqrt{\frac{p_0 q_0}{n}},\ p_0 + z_{\alpha/2} \cdot \sqrt{\frac{p_0 q_0}{n}}\right)$$



3º e 4º pasos. Verificación e decisión

Calcúlase a proporción, pr, na mostra e compróbase se está ou non na zona de aceptación. (Véxase o CASO 1 das páxinas 312 e 313).

Contraste unilateral: $p \leq p_0$ o $p \geq p_0$

1º PASO	HIPÓTESE NULA	H_0: $p \leq p_0$	H_0: $p \geq p_0$
	HIPÓTESE ALTERNATIVA	H_1: $p > p_0$	H_1: $p < p_0$
2º PASO ZONA DE ACEPTACIÓN		 $\left(-\infty,\ p_0 + z_\alpha \cdot \sqrt{\dfrac{p_0 q_0}{n}}\right)$	 $\left(p_0 - z_\alpha \cdot \sqrt{\dfrac{p_0 q_0}{n}},\ +\infty\right)$

Os pasos **3º** e **4º,** como sempre, consisten en extraer a mostra, calcular o parámetro pr nela e comprobar se pertence ou non á zona de aceptación.

Recordemos os **valores críticos** correspondentes ós valores de α máis utilizados:

NIVEL DE SIGNIFICACIÓN, α		0,10	0,05	0,01
VALOR CRÍTICO	Contraste bilateral: $z_{\alpha/2}$	1,645	1,96	2,575
	Contraste unilateral: z_α	1,28	1,645	2,33

EXERCICIOS RESOLTOS

1. *A unhas eleccións preséntanse tres partidos, A, B e C. Un comentarista político afirma que os electores se reparten do seguinte modo:*

A favor de A, o 40%; a favor de B, o 40% ou máis; a favor de C, o 40% ou menos.

Preténdese contrastar estas hipóteses mediante unha mostra de 250 electores.

a) Calcula a zona de aceptación de cada unha delas, cun nivel de significación do 5%.

b) Unha vez extraída a mostra, obtéñense 132 a favor de A, 88 a favor de B e 30 a favor de C.

Toma decisións respecto ás tres hipóteses.

a)

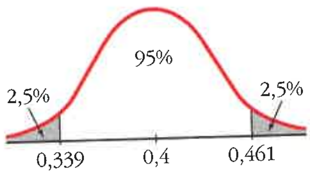
		Para A	Para B	Para C
1º paso.	HIPÓTESE	H_0: $p = 0{,}4$	H_0: $p \geq 0{,}4$	H_0: $p \leq 0{,}4$
	HIPÓTESE ALTERNATIVA	H_1: $p \neq 0{,}4$	H_1: $p < 0{,}4$	H_1: $p > 0{,}4$

2º paso. Para determinar as zonas de aceptación, temos en conta que, nos tres casos, as proporcións mostrais, *pr*, distribuiríanse:

$$N\left(0{,}4;\ \sqrt{\frac{0{,}4 \cdot 0{,}6}{250}}\right) = N(0{,}4;\ 0{,}031)$$

As respectivas zonas de aceptación son:

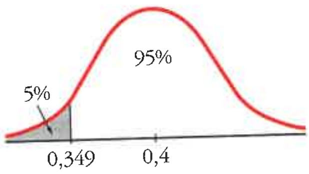
A:



Como é un contraste bilateral (dúas "colas"), o valor característico correspondente a $\alpha = 0{,}05$ é $z_{\alpha/2} = 1{,}96$.

$(0{,}4 \pm 1{,}96 \cdot 0{,}031) = (0{,}339;\ 0{,}461)$

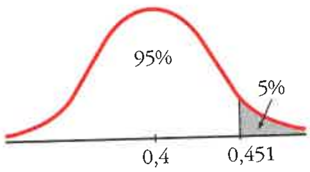
B:



Como é un contraste unilateral (unha "cola"), o valor característico correspondente a $\alpha = 0{,}05$ é $z_\alpha = 1{,}645$.

$(0{,}4 - 1{,}645 \cdot 0{,}031;\ +\infty) = (0{,}349;\ +\infty)$

C:



Analogamente: $\alpha = 0{,}05 \rightarrow z_\alpha = 1{,}645$

$(-\infty,\ 0{,}4 + 1{,}645 \cdot 0{,}031) = (-\infty;\ 0{,}451)$

b) **3º e 4º pasos**

	A	B	C
PROPORCIÓNS OBTIDAS	$132/250 = 0{,}528$	$88/250 = 0{,}352$	$30/250 = 0{,}12$
ZONAS DE ACEPTACIÓN	$(0{,}339;\ 0{,}461)$	$(0{,}349;\ +\infty)$	$(-\infty;\ 0{,}451)$
DECISIÓN	Rexéitase	Acéptase	Acéptase

Pode resultar chocante que aceptemos a hipótese correspondente ó partido B ("a proporción dos seus votantes supera ou iguala o 40%) cando obtivemos unha proporción inferior (35,2%).

Obsérvese que tamén teriamos que aceptar a hipótese seguinte para B: "a proporción de votantes é inferior ou igual ó 40%.". Isto é debido a que só rexeitariamos como improbables o 5% dos posibles resultados. Por iso, en lugar de dicir "acéptase a hipótese", sería máis correcto dicir "non hai motivo para rexeitar a hipótese".

EXERCICIOS PROPOSTOS

1. Respecto a un certo dado, A opina que $P[6] = 0{,}15$, B opina que $P[6] \leq 0{,}15$ e C opina que $P[6] \geq 0{,}15$. Contrasta as tres hipóteses cun nivel de significación de 0,10, sabendo que se lanzou o dado 1 000 veces e se obtivo 183 veces o "6".

14.4 POSIBLES ERROS NO CONTRASTE DE HIPÓTESES

Ó aplicar un test estatístico, podemos cometer dous tipos de erros:

ERRO DE TIPO I. Comételse cando a hipótese nula é verdadeira e, como consecuencia do contraste, rexéitase.

ERRO DE TIPO II. Comételse cando a hipótese nula é falsa e, como consecuencia do contraste, acéptase.

Naturalmente, ó aplicar un test ignoramos se cometemos erro ou non o cometemos. O que si podemos é intentar avaliar a probabilidade de cometer erro dun ou outro tipo e deseñar o experimento de modo que ditas probabilidades de erro se reduzan todo o posible.

Analicemos estas probabilidades sobre un exemplo:

As estaturas das alumnas de COU eran, en 1990, de media 167 cm e desviación típica 7 cm. Emitimos a hipótese de que as actuais alumnas de 2° de Bacharelato teñen a mesma media H_0: $\mu = 167$. *Imos contrastar a hipótese mediante unha mostra de tamaño* $n = 60$ *e cun nivel de significación* $\alpha = 0,10$.

Obtense a zona de aceptación $\left(167 \pm 1,645 \cdot \frac{7}{\sqrt{60}}\right) = (165,51;\ 168,49)$.

Se ó extraer a mostra obtemos $\bar{x} = 168,72$ rexeitamos a hipótese. Pero podemos estar equivocados. É dicir, podemos cometer un erro de tipo I.

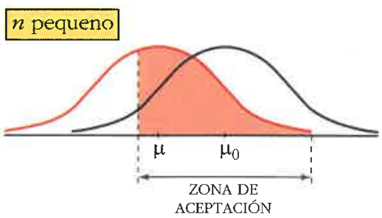
Se ó extraer a mostra se obtivera $\bar{x} = 168,12$ aceptariamos a hipótese. Se estamos equivocados, estariamos cometendo un erro de tipo II.

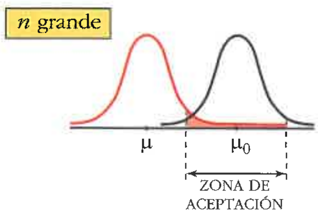
Probabilidade de cometer erro dun tipo ou outro

A probabilidade de cometer **ERRO DE TIPO I** é precisamente α, o nivel de significación, pois se a hipótese é verdadeira, expoñémonos a rexeitar o $\alpha \cdot 100\%$ das medias mostrais. Esta probabilidade non depende do tamaño da mostra.

A probabilidade de cometer **ERRO DE TIPO II** depende do verdadeiro valor de μ e do tamaño da mostra. Se μ é o verdadeiro valor da media e μ_0 o que lle atribuímos mediante a hipótese (H_0: $\mu = \mu_0$), na realidade son distintos, xa que supoñemos que estamos cometendo erro de tipo II. A verdadeira distribución das medias mostrais vén dada polas curvas vermellas (media μ).

As curvas negras son as supostas distribucións (media μ_0). Sobre elas constrúense os intervalos de aceptación. A área vermella dános, en cada caso, a proporción de mostras para as cales se aceptaría a hipótese $\mu = \mu_0$ e, por tanto, a probabilidade de cometer un erro de tipo II. É claro que para mostras grandes esta probabilidade é moito menor.





ERRO TIPO I. Rexeitar H_0 sendo verdadeira	A súa probabilidade é α. Non depende de n.
ERRO TIPO II. Aceptar H_0 sendo falsa	A súa probabilidade depende do verdadeiro valor do parámetro. Faise tanto menor canto maior sexa n.

1. Test de hipóteses sobre unha probabilidade (bilateral)

Reflexionamos sobre cada unha das seguintes experiencias:

a) Lanzamos unha moeda 10 veces e obtemos 6 caras.

b) Lanzamos unha moeda 100 veces e obtemos 60 caras.

c) Lanzamos unha moeda 1 000 veces e obtemos 600 caras.

¿Podemos deducir dalgunha delas que a moeda é incorrecta?

(Problema proposto, para ser resolto intuitivamente, nas páxinas iniciais da unidade: páxina 311).

1º paso: Hipótese

En calquera das tres experiencias, queremos contrastar:

$$H_0\colon p = \frac{1}{2} \text{ (é dicir, “a moeda é correcta”) fronte a } H_1\colon p \neq \frac{1}{2}$$

2º paso: Dedución de conclusións. Obtención da zona de aceptación

Posto que en todos os casos $np = n \cdot \frac{1}{2} \geq 5$, a distribución de proporcións mostrais sería:

$$N\left(\frac{1}{2}, \sqrt{\frac{(1/2)\cdot(1/2)}{n}}\right) = N\left(0{,}5; \frac{0{,}5}{\sqrt{n}}\right)$$

As zonas de aceptación, para un nivel de significación $\alpha = 0{,}01$, son:

$$\left(0{,}5 \pm 2{,}575 \cdot \frac{0{,}5}{\sqrt{n}}\right)$$

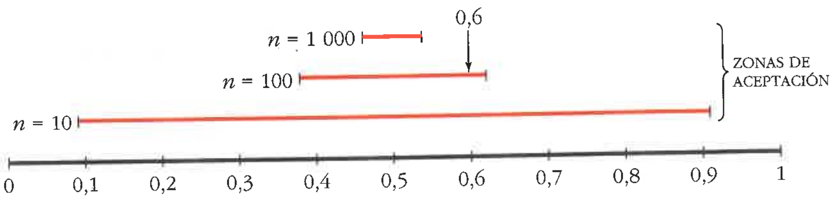
n	10	100	1 000
ZONA DE ACEPTACIÓN	(0,5 ± 0,41) (0,09; 0,91)	(0,5 ± 0,13) (0,37; 0,63)	(0,5 ± 0,04) (0,46; 0,54)

3º paso: Verificación

Agora é cando se supón que se extrae a mostra e se calcula a proporción de caras.

Nos tres casos obtívose $pr = 0{,}6$ $\left(\frac{6}{10} \text{ ou } \frac{60}{100} \text{ ou } \frac{600}{1\,000}\right)$

4º paso: Decisión



O valor de pr obtido nas mostras ($pr = 0{,}6$) cae dentro da zona de aceptación para $n = 10$ e $n = 100$, pero cae fóra da zona de aceptación para $n = 1\,000$.

Só rexeitamos a hipótese de que a moeda sexa correcta no caso c).

Observemos que o resultado desta proba é acorde coa intuición: se obtemos 6 caras en 10 tiradas, parécenos un resultado do máis razoable. Seguramente admitiriamos como non sorprendente calquera resultado entre 2 e 8 caras en 10 tiradas.

En 100 lanzamentos xa nos custa máis admitir certas desviacións sobre a proporción $\frac{1}{2}$.

E en 1 000, aínda sendo tan permisivos como o é o tomar $\alpha = 0{,}01$, non resulta admisible considerar que, por azar, o 60% dos lanzamentos dunha moeda correcta sexan caras. ¡A moeda ten que ser incorrecta!

2. Test de hipóteses para a proporción (unilateral)

Nas últimas votacións, hai un ano, o 53% dos votantes dunha localidade estaban a favor do alcalde. Acaba de realizárselles unha enquisa a 360 persoas elixidas ó azar e 176 delas estaban a favor do alcalde.

¿Pódese afirmar, cun nivel de confianza do 90%, que o alcalde non perde popularidade?

1º paso: Hipótese

Queremos contrastar:

H_0: $p \geq 0{,}53$

(A proporción de votantes favorables ó alcalde é igual ou supera a das últimas eleccións).

fronte a H_1: $p < 0{,}53$

2º paso: Zona de aceptación

$$\left(p_0 - z_\alpha \cdot \sqrt{\frac{p_0 q_0}{n}}, +\infty\right)$$

$p_0 = 0{,}53$; $q_0 = 1 - 0{,}53 = 0{,}47$; $n = 360$

Para $\alpha = 0{,}09 \to z_\alpha = 1{,}28$

$$\left(0{,}53 - 1{,}28 \cdot \sqrt{\frac{0{,}53 \cdot 0{,}47}{360}}, +\infty\right) = (0{,}496; +\infty)$$

3º paso: Verificación

Na mostra obtense unha proporción $pr = \frac{176}{360} = 0{,}489$.

4º paso: Decisión

A proporción de partidarios do alcalde na mostra cae fóra da zona de aceptación. Por tanto, rexéitase a hipótese nula: non podemos considerar que o alcalde non perda popularidade. É dicir, admitimos que perde popularidade.

PARA PRACTICAR

Contraste de hipóteses para a media

1 Realiza en cada caso o contraste de hipóteses coas condicións que se dan a continuación (en todos os casos supoñemos que a poboación de partida é normal):

	H_0	σ	α	n	$\bar{x}$
a)	$\mu = 12$	$\sigma = 1{,}5$	$\alpha = 0{,}01$	10	11
b)	$\mu = 1{,}45$	$\sigma = 0{,}24$	$\alpha = 0{,}05$	16	1,6
c)	$\mu \leq 11$	$\sigma = 4{,}6$	$\alpha = 0{,}05$	100	12
d)	$\mu \geq 15$	$\sigma = 1$	$\alpha = 0{,}1$	150	14,5

2 S Un fabricante de lámpadas eléctricas está ensaiando un novo método de produción que se considerará aceptable se as lámpadas obtidas por este método dan lugar a unha poboación normal de duración media 2 400 horas, cunha desviación típica igual a 300.

Tómase unha mostra de 100 lámpadas producidas por este método e esta mostra dá unha duración media de 2 320 horas. ¿Pode aceptarse a hipótese de validez do novo proceso de fabricación cun risco igual ou menor ó 5%?

3 S Sábese por experiencia que o tempo obtido polos participantes olímpicos da proba de 100 metros, na modalidade décathlon, é unha variable aleatoria que segue unha distribución normal con media 12 segundos e desviación típica 1,5 segundos. Para contrastar, cun nivel de significación do 5%, se non variou o tempo medio na última Olimpíada, extraeuse unha mostra aleatoria de 10 participantes e anotouse o tempo obtido por cada un, cos resultados seguintes, en segundos:

13	12	11	10	11
11	9	10	12	11

a) ¿Cales son a hipótese nula e a alternativa do contraste?

b) Determina a rexión crítica.

c) Realiza o contraste.

d) Explica no contexto do problema, en qué consisten cada un dos erros de tipo I e II.

4 S Comprobouse que o tempo de espera (en minutos) ata ser atendido, en certo servicio de urxencias, segue un modelo normal de probabilidade.

A partir dunha mostra de 100 persoas que foron atendidas en dito servicio, calculouse un tempo medio de espera de 14,25 minutos e unha desviación típica de 2,5 minutos.

a) ¿Poderiamos afirmar, cun nivel de significación do 5% ($\alpha = 0{,}05$), que o tempo medio de espera, nese servicio de urxencias, non é de 15 minutos?

b) ¿Que poderiamos concluír se o nivel de significación fose do 0,1% ($\alpha = 0{,}001$)?

c) ¿Existe contradicción en ambas as situacións?

Xustifica as respostas.

5 S A duración das lámpadas de 100 watts que fabrica unha empresa segue unha distribución normal cunha desviación típica de 120 horas. A súa vida media está garantida durante un mínimo de 800 horas.

Escóllese ó azar unha mostra de 50 lámpadas dun lote e, despois de comprobalas, obtense unha vida media de 750 horas. Cun nivel de significación de 0,01, ¿habería que rexeitar o lote por non cumprir a garantía?

6 Unha empresa asegura que unhas determinadas pastillas de xabón duran máis de 11 días. Para comprobalo, realízase unha enquisa en 100 casos. Estas son as respostas:

DURACIÓN (días)	5 a 9	10 a 14	15 a 19	20 a 24
RESPOSTAS	24	46	19	11

¿Pódese dar como válida a afirmación da empresa, para un nivel de significación de $\alpha = 0{,}05$?

7 S Unha enquisa, realizada a 64 empregados unha fábrica, concluíu que o tempo medio de duración dun emprego nela era de 6,5 anos, cunha desviación típica de 4.

¿Serve esta información para aceptar, cun nivel de significación do 5%, que o tempo medio de emprego nesa fábrica é menor ou igual que 6?

Xustifica adecuadamente a resposta.

8 S A Concellería de Xuventude manexa o dato de que a idade á que os fillos se independizan é unha variable normal con media 29 anos e desviación típica 3 anos. Aínda que a desviación típica non presenta dúbidas, si se sospeita que a media descendeu, sobre todo pola política de axuda ó emprego que levou a cabo o concello. Así, dun estudio recente sobre 100 mozos que se acaban de independizar, obtívose unha media de 28,1 anos de idade.

a) Cun nivel de significación do 1%, ¿pode defenderse que a idade media non diminuíu, fronte a que si o fixo como parecen indicar os datos? Presenta o contraste ou test de hipóteses e resólveo.

b) Explica, neste problema, en qué consisten cada un dos erros do tipo I e II.

Contr. de hipóteses para a proporción

9 Realiza en cada caso o test de hipóteses coas condicións que se indican:

	H_0	α	n	pr
a)	$p = 0,5$	0,01	1 000	0,508
b)	$p \leq 0,6$	0,05	600	0,61
c)	$p \geq 0,3$	0,1	200	0,25

10 Un dentista afirma que o 40% dos nenos de 10 anos presentan indicios de carie dental. Tomada unha mostra de 100 nenos, observouse que 30 presentaban indicios de carie.

Utilizando a aproximación normal, comproba, a nivel de significación do 5%, se o resultado permite rexeitar a afirmación do dentista.

11 Unha empresa de productos farmacéuticos afirma na súa publicidade que un dos seus medicamentos reduce considerablemente os síntomas da alerxia primaveral no 90% da poboación.

Unha asociación de consumidores experimentou dito fármaco nunha mostra de 200 socios, obtendo o resultado indicado na publicidade en 170 persoas.

Determina se a asociación de consumidores pode considerar que a afirmación da empresa é estatisticamente correcta ó nivel de significación de 0,05.

12 S Afírmase que, nunha determinada cidade, ó menos o 30% das familias posúen ordenador. Tómase unha mostra aleatoria de 200 familias da cidade e resulta que 50 posúen ordenador.

A un nivel de significación de 0,05, ¿hai suficiente evidencia para refutar a afirmación?

13 S O 42% dos escolares adoitan perder ó menos un día de clase a causa de gripes e catarros. Sen embargo, un estudio sobre 1 000 escolares revela que no último curso houbo 450 en tales circunstancias. As autoridades sanitarias defenden que a porcentaxe do 42% para toda a poboación de escolares se mantivo.

a) Contrasta, cun nivel de significación do 5%, a hipótese defendida polas autoridades sanitarias, fronte a que a porcentaxe aumentou como parecen indicar os datos, explicando claramente a qué conclusión se chega.

b) ¿Como se chama a probabilidade de concluír erroneamente que o tanto por cento se mantivo?

PARA PROFUNDAR

14 Nun test de hipóteses para estudiar se o cociente intelectual medio dos estudiantes dunha universidade é 113, seleccionamos unha mostra aleatoria de 180 estudiantes, obtendo unha media de 115. A zona de aceptación obtida foi o intervalo (111,98; 114,02) e sabemos que a desviación típica é $\sigma = 7$. Por tanto, rexeitamos a hipótese.

¿Cal é a probabilidade de que rexeitaramos a hipótese, cando en realidade era verdadeira? ¿Como se chama este tipo de erro?

15 Nunha determinada provincia, a nota media en matemáticas dos alumnos de 2º de Bacharelato do curso pasado foi de 5,8, cunha desviación típica de 2,3 puntos. Cun nivel de significación de 0,05 e supoñendo que a desviación típica segue sendo a mesma, queremos contrastar a hipótese de que a media non variou. Para iso, imos extraer unha mostra aleatoria de tamaño 100. Así, a zona de aceptación será o intervalo (5,35; 6,25). Se ó final a media real fora de 5 puntos, ¿cal é a probabilidade de obter unha media mostral que nos leve a cometer un erro de tipo II (é dicir, aceptar H_0 sendo falsa)?

PARA MELLORAR A TÚA FORMACIÓN, ORGANIZA AS TÚAS IDEAS EN ESTATÍSTICA

MOSTRAXES ESTATÍSTICAS

- Mostraxe aleatoria.
- Distribución das medias mostrais (Teorema Central do Límite).
- Distribución das proporcións mostrais.

CÁLCULO DE PROBABILIDADES

- Sucesos. Probabilidade.
- Lei de Laplace.
- Experiencias compostas.
- Probabilidade total.
- Fórmula de Bayes.
- Distribución normal. Intervalos característicos.
- Distribución binomial.

INFERENCIA ESTATÍSTICA

- Estimación de μ.
- Estimación de p.
- Test de hipóteses sobre μ.
- Test de hipóteses sobre p.

EXPERIENCIAS ALEATORIAS. SUCESOS

EXPERIENCIAS ALEATORIAS. SUCESOS		
serven para	**conceptos asociados**	**é especialmente importante**
■ Estudiar e describir fenómenos relacionados co azar. ■ Analizar as probabilidades e as súas leis. ■ Calcular probabilidades de sucesos aleatorios.	■ Sucesos aleatorios. ■ Lei dos grandes números. ■ Propiedades da probabilidade. ■ Lei de Laplace. ■ Probabilidade condicionada. ■ Probabilidade total. ■ Probabilidade "a posteriori".	■ Manexar as operacións con sucesos e as propiedades das súas probabilidades. ■ Utilizar con destreza o diagrama en árbore para describir experiencias compostas e calcular as súas probabilidades.

Consellos útiles

- Son especialmente importantes as seguintes propiedades dos sucesos e as probabilidades:

  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

  $P(A') = 1 - P(A)$

  $(A \cup B)' = A' \cap B', \quad (A \cap B)' = A' \cup B'$

  $P(A/B) = P(A \cap B)/P(B)$ ou a equivalente: $P(A \cap B) = P(B) \cdot P(A/B) = P(A) \cdot P(B/A)$

  A e B independentes $\Leftrightarrow P(A \cap B) = P(A) \cdot P(B)$

  Procura entendelas e aprendelas. (Véxanse os exercicios 2 e 3 da páxina 254; 12 e 13 da páxina 258; 39, 40, 41, 42 e 43 da páxina 261).

- Utiliza o diagrama en árbore para calcular a probabilidade total ou probabilidades "a posteriori" en experiencias compostas. O seu manexo destro permite ver con claridade o proceso e prescindir das fórmulas ou ratificar o resultado obtido por medio delas. (Véxanse os exercicios resoltos 1 e 2 e o proposto 1 da páxina 253; o 5 da páxina 255; 19, 20 e 26 da páxina 259; 32 e 33 da páxina 260).

DISTRIBUCIÓNS BINOMIAL E NORMAL

DISTRIBUCIÓNS BINOMIAL E NORMAL		
serven para	**técnicas asociadas**	**é especialmente importante**
■ Describir situacións aleatorias moi frecuentes nas ciencias e na natureza. ■ Facer inferencias fiables para a poboación a partir dunha mostra.	■ Manexo das táboas da distribución $N(0, 1)$. ■ Relación da distribución binomial coa normal. ■ Construción de intervalos característicos.	■ O cálculo de probabilidades nunha distribución normal calquera. ■ A obtención do intervalo característico, asociado a unha probabilidade, nunha distribución $N(\mu, \sigma)$.

Consellos útiles

As distribucións normal e binomial son instrumentos imprescindibles para manexarte con soltura no traballo con mostras, fundamentalmente para a inferencia. Por iso debes, ademais de entender o que fas, adquirir gran destreza no seu uso:

- Calcula probabilidades nunha distribución $N(0,1)$, facendo uso das táboas, visualizando sobre a gráfica o que conseguiches. (Véxanse páxinas 276 e 277).
- Acostúmate a automatizar o proceso de tipificación da variable para calcular probabilidades nunha distribución normal calquera, $N(\mu, \sigma)$.

  $P[h < x < k] = P\left[\frac{h-\mu}{\sigma} < z < \frac{k-\mu}{\sigma}\right]$. Isto calcúlase nas táboas (páxina 278).

- Familiarízate coa nomenclatura dos intervalos característicos: valores críticos, $z_{\alpha/2}$, probabilidade, $1-\alpha$, ... e intenta memorizar os valores críticos correspondentes ás probabilidades 0,9; 0,95; 0,99. (Véxase páxina 279).

  Aprende a calcular destramente intervalos característicos para distribucións normais $N(0, 1)$ e $N(\mu, \sigma)$. (Véxase páxina 280).

- Distingue en qué casos unha binomial está próxima a unha normal e calcula os seus parámetros. (Véxanse páxinas 300 e 301).

MOSTRAS ESTATÍSTICAS

MOSTRAS ESTATÍSTICAS		
serven para	**conceptos e técnicas asociados**	**é especialmente importante**
■ En xeral, indagar características dunha poboación analizando un reducido número dos seus elementos (mostras). ■ En particular, inferir o valor dos parámetros da poboación a partir dos correspondentes parámetros dunha mostra.	■ Mostraxe: aleatoria simple, sistemática e estratificada. ■ Mostraxe mediante números aleatorios. ■ Distribución das medias mostrais: teorema central do límite. Consecuencias. ■ Distribución das proporcións mostrais: consecuencias.	■ Coñecidos os parámetros dunha poboación, definir a distribución das medias mostrais, $\bar{x}$, e calcular intervalos característicos. ■ Coñecida a probabilidade $P[C]$ na poboación, definir a distribución da proporción de C nunha mostra, pr, e calcular intervalos característicos.

Consellos útiles

- Practica a descrición da distribución das medias, $\bar{x}$, das mostras de tamaño n extraídas dunha poboación de parámetros coñecidos μ e σ:

  Se $n \geq 30$ ou a poboación é normal, entón $\bar{x}$ é $N(\mu, \sigma/\sqrt{n})$.

- Reflexiona sobre o feito de que a desviación típica das medias mostrais se pode facer tan pequena como se queira ó aumentar o tamaño, n, da mostra.

- Calcula os intervalos característicos para $\bar{x}$ en poboacións concretas e para mostras de tamaño determinado. (Véxanse páxinas 281 e 282).

- Practica a descrición das proporcións, pr, (de individuos cunha certa característica C) das mostras de tamaño n extraídas dunha poboación na cal $P[C] = p$ é coñecida:

$$\text{Se } np \geq 5 \text{ e } nq \geq 5, \text{ entón } pr \text{ é } N\left(p, \sqrt{\frac{pq}{n}}\right)$$

- Reflexiona sobre o feito de que a desviación típica de pr se pode facer tan pequena como se queira ó aumentar o tamaño n da mostra.

- Obtén intervalos característicos para pr en poboacións concretas e para mostras de tamaño determinado. (Véxanse páxinas 302 e 303).

INFERENCIA ESTATÍSTICA

INFERENCIA ESTATÍSTICA		
serve para	**técnicas asociadas**	**é especialmente importante**
■ Estimar, mediante un intervalo, o valor da media, μ, ou da proporción, p, de individuos cunha característica da poboación, a partir dunha mostra. ■ Efectuar hipóteses estatísticas sobre o valor de μ ou de p e contrastalas a partir dos resultados dunha mostra.	■ Obtención do intervalo de confianza para μ ou para p, cun certo nivel de confianza. ■ Pasos para establecer unha hipótese estatística. Obtención da zona de aceptación.	■ Comprender, aprender e utilizar con destreza as expresións $E_{\bar{x}} = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$ $E_{pr} = z_{\alpha/2} \cdot \sqrt{\frac{pr(1-pr)}{n}}$ ■ Obter as zonas de aceptación nos contrastes de hipóteses, distinguindo os unilaterais dos bilaterais.

Consellos útiles

- Os intervalos característicos para $\bar{x}$ $(\mu \pm z_{\alpha/2} \cdot \sigma/\sqrt{n})$ obtéñense a partir dos parámetros da poboación. Os intervalos de confianza para μ $(\bar{x} \pm z_{\alpha/2} \cdot \sigma/\sqrt{n})$ obtéñense a partir da media $\bar{x}$ dunha mostra concreta. Debes entender as súas similitudes e as súas diferencias. Analogamente para p e pr. (Véxanse páxinas 286 e 287).

- Aprende a automatiza a utilización das expresións $E = z_{\alpha/2} \cdot \sigma/\sqrt{n}$ para μ (e a correspondente a p) e utilízaas con destreza non só para calcular E, senón tamén para obter o tamaño da mostra, n, ou o nivel de confianza, $1 - \alpha$.

- No deseño dunha hipótese estatística estámoslle supoñendo un valor a un parámetro (μ ou p). Todo o estudio faise dando por bo dito valor e, só ó final, o valor obtido na mostra servirá para decidir se a hipótese é boa ou non o é.

  Aprende e automatiza os catro pasos que hai que dar. Distingue entre contrastes bilaterais e unilaterais. Recorda que nestes os valores críticos son distintos, pois só hai unha cola.

PROBA DE AUTOAVALIACIÓN

OPCIÓN A

1 Nunha academia de Huelva con 80 estudiantes impártanse tres idiomas: inglés (I), francés (F) e alemán (A). Hai 48 estudiantes matriculados en I, 20 en F e 12 en A. Por outra parte, as probabilidades de que un estudiante da academia viva en Huelva son 0,4 para os de I; 0,6 para os de F e 0,8 para os de A. Elíxese ó azar un estudiante da academia:

a) ¿Cal é a probabilidade de que viva en Huelva?

b) Se a persoa elixida vive en Huelva, ¿cal é a probabilidade de que estudie francés?

2 O nivel de protrombina nunha poboación normal é de 200 mg/100 m*l* de plasma, cunha desviación típica de 4.

a) Se consideramos mostras de 50 individuos, ¿cal é a distribución da variable aleatoria media mostral, $\bar{x}$?

b) Calcula a probabilidade de que, nunha mostra de 50, o nivel medio de protrombina sexa maior de 201 mg/100 m*l*.

3 O peso dun determinado colectivo dunha poboación segue unha distribución normal cunha desviación típica de 8 kg. Nun estudio realizado cunha mostra aleatoria de 81 persoas dese colectivo, obtívose un peso medio de 65 kg.

a) Obtén un intervalo de confianza ó 99% para o peso medio do colectivo.

b) Calcula o tamaño mínimo que debería ter a mostra no caso de admitir un erro máximo de 1,5 kg, cun grao de confianza do 95%.

OPCIÓN B

1 Extráese unha carta ó azar dunha baralla española de 40 cartas e, sen devolvela ó mazo, sácase outra carta, tamén ó azar. Calcula as probabilidades de cada un dos seguintes sucesos:

a) A = "ó menos unha das dúas cartas é de OUROS".

b) B = "as dúas cartas son do mesmo pao".

c) $A \cap B$ d) $A' \cup B'$ e) $A \cup B$

2 Nunha determinada rexión hai catro lugares próximos: L1 con 6 500 habitantes, L2 con 3 000, L3 con 8 000 e L4 con 5 000. Vaise realizar un estudio na rexión baseado nunha mostra de 450 persoas.

a) ¿Que tipo de mostraxe se deberá realizar se queremos que estean representados os catro lugares?

b) ¿Cantas persoas de cada lugar se deben seleccionar, atendendo a razóns de proporcionalidade?

c) ¿Como seleccionarías as persoas de cada lugar?

3 Nun estudio realizado nunha cidade hai cinco anos sobre o número de horas semanais dedicadas a practicar algún deporte (en mozos de entre 15 e 20 anos), obtívose unha media de 7,2 horas e unha desviación típica de 3. Para ver se a media variou, elixiuse unha mostra de 80 mozos, obtendo unha media mostral de 7,6 horas (supoñemos $\sigma = 3$). Contrasta, cun nivel de significación do 5%, a hipótese de que o tempo medio segue sendo o mesmo, fronte á hipótese de que variou.