

11

Inferencia estatística. Mostraxe. Distribucións mostrais

Na Estatística distínguense dúas partes perfectamente diferenciadas. Unha delas coñécese co nome de Estatística Descritiva e ten como obxectivo a recollida, organización e análise de datos, obtendo a partir deles uns valores chamados parámetros, que os identifican, e ademais permiten facer comparacións con outros conxuntos de datos e establecer relacións entre eles.

Outra parte da Estadística, chamada Estatística Inferencial, trata de elaborar conclusións sobre unha poboación a partir dos datos recollidos dunha parte dela chamada mostra. A representatividade da mostra será de suma importancia para a fiabilidade das conclusións.

As aplicacións de Estatística Inferencial abranguen campos moi diversos. En Medicina empréganse para investigar os resultados de tratamentos con novos fármacos. En Socioloxía úsanse para facer enquisas de opinión aos resignados contribuíntes. En Industria para mellorar a calidade dos produtos cos métodos de control de calidade e coñecer o grao de aceptación dos consumidores.

1. VARIABLES ALEATORIAS	2
1.1. Variables aleatorias discretas. Distribución binomial	2
1.2. Variable aleatoria continua. A distribución normal	3
2. MOSTRAXE	6
2.1. Tipos de mostraxe.	7
2.2. Parámetros e estatísticos.	8
3. DISTRIBUCIÓN DAS MEDIAS MOSTRAIS	8
4. DISTRIBUCIÓN DE PROPORCIÓNS MOSTRAIS	10

11.1 Variables aleatorias

Unha variable cuxos valores se determinan sobre os resultados dun experimento aleatorio chámase unha variable aleatoria. Por exemplo, nun colexio de 300 alumnos, eliximos un alumno ao chou (experimento aleatorio), anotamos a súa idade (variable aleatoria discreta), medimos a súa estatura (variable aleatoria continua), rexistramos o número de irmáns que ten (variable aleatoria discreta) e o seu peso (variable aleatoria continua).

Os valores dunha variable aleatoria discreta son números enteiros positivos; os valores dunha variable aleatoria continua son números reais comprendidos nun intervalo real. As variables aleatorias simbolízanse por unha letra maiúscula como X, Y ou Z.

1.1. Variables aleatorias discretas. Distribución binomial

Nas variables aleatorias discretas a cada valor x da variable X asóciase a probabilidade de que ocorra, $P[X = x]$ e a esta asociación chámase lei de probabilidade. Unha distribución de probabilidade dunha variable aleatoria discreta é semellante a unha distribución de frecuencias dunha variable estatística, só que en vez de frecuencias relativas temos probabilidades.

O curso pasado estudamos as variables aleatorias discretas que seguen a distribución binomial. Supoñamos un experimento aleatorio que se poida repetir indefinidamente e que en cada proba só teña dous resultados: éxito (E) e fallo (F). experimentos deste tipo son: tirar unha moeda, onde únicamente sae cara ou cruz; tirar un dado e observar se sae 5 ou non; etc.

Suporemos que p é a probabilidade de éxito en cada proba e, polo tanto, $1-p$ será a probabilidade de fallo en cada proba. Se o experimento se repite n veces, ao anotar os resultados obtemos unha palabra de lonxitude n formada polas letras E e F

E E F F E F E E F F F F E

Cantas palabras deste tipo conteñen x veces a letra E? Isto é equivalente a dicir: se temos n caixas, de cantas maneiras distintas podemos situar x letras E, unha por caixa? Ou, de cantas maneiras podemos elixir x caixas entre n dadas? Esta última pregunta ten unha resposta coñecida, e é o número $C_{n,x}$ ou $\binom{n}{x}$.

Se agora nos preguntamos cal é a probabilidade de obter x éxitos en n probas, ou cal é a probabilidade do suceso

$$A = \overset{x \text{ veces}}{E} \overset{n-x \text{ veces}}{F} \dots \overset{x \text{ veces}}{E} \overset{n-x \text{ veces}}{F} \dots \overset{x \text{ veces}}{E} \overset{n-x \text{ veces}}{F} \dots ?$$

Como as probas son independentes, a probabilidade non varía dunha a outra proba, entón:

$$P(A) = \overset{x \text{ veces}}{P(E)} \dots \overset{x \text{ veces}}{P(E)} \cdot \overset{n-x \text{ veces}}{P(F)} \dots \overset{n-x \text{ veces}}{P(F)}$$

$$P(A) = \overset{x \text{ veces}}{p} \dots \overset{x \text{ veces}}{p} \cdot \overset{n-x \text{ veces}}{(1-p)} \dots \overset{n-x \text{ veces}}{(1-p)} = p^x \cdot (1-p)^{n-x}$$

Ao haber $\binom{n}{x}$ palabras de lonxitude n con x letras E e ter cada palabra unha probabilidade de $p^x \cdot (1-p)^{n-x}$ entón definimos unha función de probabilidade para a variable aleatoria X , que conta o número de éxitos en n probas, así:

$$P[X = x] = \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x}$$

Esta función recibe o nome de función de probabilidade dunha distribución binomial de n probas con probabilidade de éxito p , simbólicamente $B(n, p)$

Exemplo

1. Nunha cidade, o 40% dos alumnos que ascenden a bacharelato ten suspensa alguna materia de 4º de ESO. Elíxense 6 alumnos de 1º de Bacharelato ao chou. Cal é a probabilidade de que a metade deles teña alguna materia suspensa de ESO?

Solución:

Trátase de distribución binomial:

1º) en cada proba hai dous únicos resultados: ter alguna suspensa ou non;

2º) o resultado de cada proba é independente do anterior;

3º) a probabilidade de atopar un alumno con algún suspenso é constante $p = 0,4$.

É, polo tanto, unha distribución binomial de parámetros $n = 6$ e $p = 0,4$, $B(6; 0,4)$.

A metade de 6 é 3, a probabilidade buscada é

$$P[X = 3] = \binom{6}{3} \cdot 0,4^3 \cdot 0,6^3 = 0,2765$$

1.2. Variable aleatoria continua. A distribución normal

As variables aleatorias continuas caracterízanse por que a probabilidade atribuída a cada valor x da variable aleatoria X é cero. Como se atribúen entón probabilidades? Pois, en vez de asignar probabilidades a cada valor determinado de X , faise a intervalos de valores da variable, así:

$$P[a \leq X \leq b] = \int_a^b f(x).$$

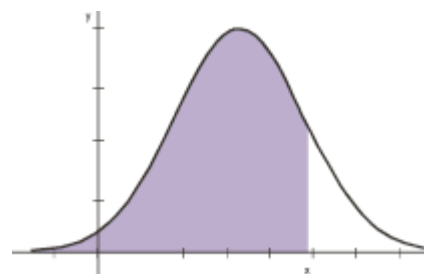
Deste modo, calcular probabilidades equivale a calcular áreas como a da rexión sombreada da figura.

Unha variable aleatoria continua, X , dice que está normalmente distribuída ou que segue unha distribución normal de media μ e desviación típica σ , e simbolízase por $N(\mu, \sigma)$, se a función $f(x)$ da integral é:

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



Se X está normalmente distribuída, a probabilidade de que X tome un valor menor ou igual que x , $P[X \leq x]$, é a área da rexión sombreada na figura, sabendo que a área baixo toda a curva é 1.



O cálculo desa área faise mediante unha integral definida, pero estas integrais están tabuladas para $N(0,1)$. Entón para achar $P[X \leq x]$, con $X, N(\mu, \sigma)$, transformamos a variable X noutra, que simbolizamos por Z , que sexa $N(0, 1)$. En realidade, consiste en cambiar a variable X por Z , onde

$$Z = \frac{X - \mu}{\sigma}$$

Esta transformación chámase tipificación da variable, e cúmprese que

$$P[X \leq x] = P\left[\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right] = P\left[Z \leq \frac{x - \mu}{\sigma}\right]$$

Nos exemplos recordamos algúns casos que se poden presentar no manexo das táboas da $N(0, 1)$, que aparecen ao final da unidade.

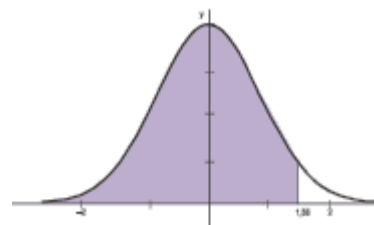
Exemplos

2. Nunha distribución normal $N(0,1)$, achar: a) $P[Z \leq 1,58]$; b) $P[Z \geq 0,46]$; c) $P[Z \leq -1,79]$; d) $P[Z \geq -1,79]$; e) $P[-0,89 \leq Z \leq 1,98]$

Solución:

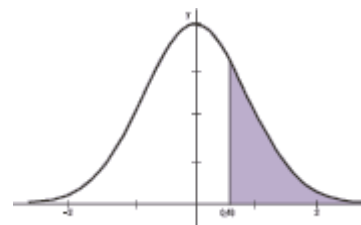
- a) $P[Z \leq 1,58] = 0,9429$

O número 0,9429 aparece na táboa na intersección da fila que empeza por 1,5 e a columna que encabeza 0,08; significa que o 94,29% dos valores de Z están comprendidos entre $-\infty$ e 1,58.



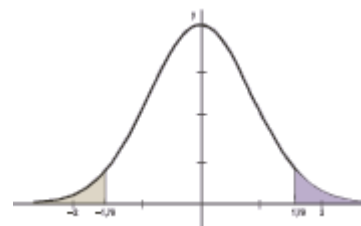
- b) $P[Z \geq 0,46] = 1 - P[Z \leq 0,46] = 1 - 0,6772 = 0,3228$

Na gráfica vemos que a probabilidade buscada corresponde á área sombreada e é igual á área total, 1, menos a área de $P[Z \leq 0,46]$



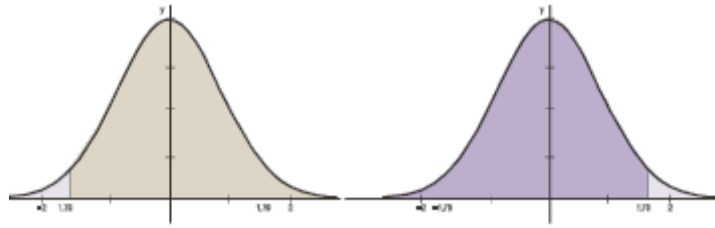
- c) $P[Z \leq -1,79] = P[Z \geq 1,79] = 1 - P[Z \leq 1,79] = 1 - 0,9633 = 0,0367$

Na gráfica observamos que por simetría a área $P[Z \leq -1,79]$ é igual que $P[Z \geq 1,79]$

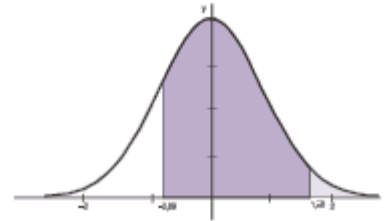


- d) $P[Z \geq -1,79] = P[Z \leq 1,79] = 0,9633$

Por simetría a área de $P[Z \geq -1,79]$ é igual que $P[Z \leq 1,79]$



$$\begin{aligned} \text{e) } P[-0,89 \leq Z \leq 1,98] &= P[Z \leq 1,98] - P[Z \geq -0,89] = \\ &= P[Z \leq 1,98] - (1 - P[Z \leq 0,89]) = \\ &= 0,9761 - (1 - 0,8133) = 0,7894 \end{aligned}$$



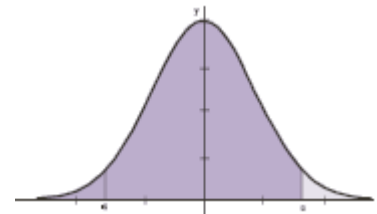
3. Nunha distribución $N(0,1)$ calcula o valor de c que cumpre a igualdade
 $P[-c \leq Z \leq c] = P[|Z| \leq c] = 0,95$

Solución:

Segundo a figura a área sombreada á dereita de c é

$$\frac{1-0,95}{2} = \frac{0,05}{2} = 0,025$$

Logo da igualdade $P[Z \leq c] = 0,95 + 0,025 = 0,975$
obtemos nas táboas que o valor de c que
corresponde a 0,975 é 1,9 na columna 0,06, é dicir, $c = 1,96$.



4. Se X é unha variable aleatoria que segue unha distribución $N(60, 12)$, calcula
 $P[X < 65]$

Solución:

$$P[X < 65] = P\left[\frac{X-60}{12} \leq \frac{65-60}{12}\right] = P\left[Z \leq \frac{5}{12}\right] = P[Z \leq 0,42] = 0,6628$$

5. Sabendo que X é unha variable aleatoria que segue unha distribución $N(10, 3)$ acha o
valor de x en $P[X < x] = 0,9761$

Solución:

$$\text{Como } P[X < x] = 0,9761 = P\left[\frac{X-10}{3} \leq \frac{x-10}{3}\right] = P[Z < z];$$

De $P[Z < z] = 0,9761$, descubrimos o valor de z con axuda das táboas e resulta que a
0,9761 lle corresponde 1,98, logo $z = 1,98$. Dado que $z = \frac{x-10}{3} = 1,98$ despxendo x
obtemos $x = 3 \cdot 1,98 + 10 = 15,94$

6. As estaturas de 500 alumnos dun colexio distribúense normalmente con media 148
cm e desviación típica 12 cm. Calcular cantos alumnos non alcanzan os 160 cm e
cantos hai cuxo talle está comprendido entre os 140 e os 160 cm.

Solución:

As estaturas distribúense segundo unha $N(148,12)$.

En primeiro lugar, pídennos $P[X < 160]$ e $P[X < 160] = P\left[\frac{X-148}{12} \leq \frac{160-148}{12}\right] =$
 $= P[Z < 1] = 0,8413$, é dicir, o 84,13% dos alumnos non chega aos 160 cm. Como o 84,13% de 500 é $0,8413 \cdot 500 = 420,65$, truncando a parte decimal, podemos dicir que 420 alumnos non chegan aos 160 cm de altura.

En segundo lugar, $P[140 \leq X \leq 160] = P\left[\frac{140-148}{12} \leq \frac{X-148}{12} \leq \frac{160-148}{12}\right] =$
 $= P[-0,66 \leq Z \leq 1] = P[Z \leq 1] - (1 - P[Z \leq 0,66]) = 0,8413 - (1 - 0,7454) = 0,5867$

É dicir, o 58,67% ten unha altura no intervalo $[140, 160]$ e como $0,5867 \cdot 500 =$

$= 293,35$ hai 293 alumnos cuxo talle está comprendido entre os 140 e os 160 cm.

7. O 45% dos habitantes dun municipio son contrarios a un proxecto da alcaldía e o resto son partidarios. Tómasse unha mostra de 64 habitantes. Cal é a probabilidade de que máis da metade dos individuos da mostra sexan contrarios ao proxecto do alcalde?

Solución:

O número x dos habitantes contrarios ao proxecto en mostras de 64 individuos distribúese segundo unha binomial $B(64, 0,45)$; é dicir, temos que calcular

$P[X > 32]$. Este cálculo obríganos a determinar 32 números combinatorios do tipo $\binom{64}{x}$, pero este monumental traballo pode abreviarse recordando que unha binomial, X , pode aproximarse por unha normal, Y , cando n é grande ou $n \cdot p > 5$ e $n \cdot (1-p) > 5$.

Neste caso a binomial $B(n, p)$ pode aproximarse pola normal $N(n \cdot p, \sqrt{n \cdot p \cdot (1-p)})$.

Logo a binomial $B(64, 0,45)$ é moi parecida á normal

$N(64 \cdot 0,45; \sqrt{64 \cdot 0,45 \cdot (1-0,45)}) = N(28,8; 3,98)$. Polo tanto, recordando que ao pasar dunha distribución discreta a unha continua engadimos e restamos un factor de corrección de 0,5, $P[a \leq X \leq b] \approx P[a - 0,5 \leq Y \leq b + 0,5]$, resulta $P[X > 32] \approx$

$\approx P[Y > 31,5] = P\left[Z > \frac{31,5-28,8}{3,98}\right] = P[Z > 0,68] = 1 - P[Z < 0,68] = 0,2483$.

11.2 Mostraxe

Cando se quere estudar unha característica dunha poboación, en moitas ocasións, resulta imposible analizar todos os individuos. Hai moitas causas desta imposibilidade: unha poboación moi numerosa, que faría custoso e interminable o estudo; ou moi dispersa; tamén é importante a natureza dos individuos porque o estudo pode supoñer a súa destrución. Por estas razóns resulta indicado elixir un subconxunto de individuos da poboación para facer o estudo. A este subconxunto chámase **mostra** e denominamos tamaño da mostra ao número de individuos que a compoñen.

Para estender os resultados obtidos da mostra ao total da poboación, é necesario que este sexa representativa. Agora, a representatividade da mostra depende do tipo de mostraxe empregada. Imos describir algúns tipos de mostraxe

2.1. Tipos de mostraxe

Mencionaremos os tipos de mostraxe máis habituais.

Mostraxe aleatoria simple. Realízase este tipo de mostraxe cando cada membro da poboación ten a mesma probabilidade de ser escollido na mostra.

Mostraxe aleatoria sistemática. Consiste en ordenar numericamente todos os individuos da poboación, elixir, ao chou, un deles e a partir del escoller sistematicamente de k en k os restantes individuos, ata completar a mostra.

Mostraxe aleatoria estratificada. Consiste en dividir previamente a poboación en grupos homoxéneos ou estratos, nos que os individuos comparten algunha característica común, e elixir mostraxes aleatorias simples en cada estrato.

Cando hai k estratos cada un con diferentes poboacións: N_1, N_2, \dots, N_k entón para conformar unha mostra de tamaño n tomamos n_1, n_2, \dots, n_k individuos en cada estrato de modo que

$$n = n_1 + n_2 + \dots + n_k,$$

e ademais cada un dos números n_1, n_2, \dots, n_k ha de ser proporcional aos tamaños dos estratos: N_1, N_2, \dots, N_k

Aínda hai outro tipo de mostraxe. A mostraxe aleatoria por conglomerados consiste en dividir previamente a poboación en subconxuntos ou conglomerados. Elíxense a continuación ao chou algúns destes conglomerados. A mostra confórmase elixindo mostraxes aleatorias simples nos conglomerados escollidos

Exemplo

8. Nun colexio hai 1500 alumnos: 600 en Primaria, 500 na ESO e 400 en Bacharelato. Quérese extraer unha mostra de 50 alumnos para realizar un estudo. Como se seleccionará a devandita mostra?

Solución:

Hai tres estratos con poboacións diferentes entre todos os alumnos do colexio:

N_1 = Primaria, N_2 = ESO e N_3 = Bacharelato,

$$N_1 + N_2 + \dots + N_k = 600 + 500 + 400$$

Extraemos tres mostraxes n_1, n_2 e n_3 , unha de cada estrato, de maneira que

$$n_1 + n_2 + n_3 = 50,$$

e ademais n_1, n_2 e n_3 sexan proporcionais a 600, 500, e 400 respectivamente. É dicir,

$$\frac{n_1}{600} = \frac{50}{1500} \Rightarrow n_1 = 50 \cdot \frac{600}{1500} = 20;$$

$$\frac{n_2}{500} = \frac{50}{1500} \Rightarrow n_2 = 50 \cdot \frac{500}{1500} = 16,6, \text{ redondeando, } 17;$$

$$\frac{n_3}{400} = \frac{50}{1500} \Rightarrow n_3 = 50 \cdot \frac{400}{1500} = 13,3, \text{ redondeando, } 13.$$

A mostra estaría formada por 20 de primaria, 17 de ESO e 13 de Bacharelato.

2.2. Parámetros e estatísticos

Chamamos **estimación** ao procedemento polo cal os resultados da mostra permiten deducir resultados relativos ao total da poboación.

O valor descoñecido dunha poboación, que estimamos a partir dunha mostra, chámase **parámetro poboacional**. Os parámetros que se adoitan estimar son a media, a proporción ou porcentaxe, a desviación típica, a varianza, etc. Por exemplo, o salario medio dos madrileños é un parámetro de toda a poboación de Madrid, mentres que o salario medio dunha mostra dos madrileños, que é un parámetro da mostra, é un **estatístico**.

Os símbolos dos parámetros da poboación son: a media, μ (lese mu), a desviación típica, σ (lese sigma), a proporción ou porcentaxe, p , mentres que para os parámetros da mostra ou estatísticos se utilizan os símbolos: \bar{x} , media mostral, $e\hat{p}$ para a proporción ou porcentaxe.

Cando tomamos os parámetros da poboación polos estatísticos correspondentes ás mostras estamos facendo unha estimación por punto do parámetro da poboación; nestas estimacións a marxe de erro pode ser grande e, polo tanto, o grao de fiabilidade pequeno. Na unidade didáctica próxima veremos que resulta máis fiable establecer un intervalo no que se atope o parámetro poboacional buscado. Este tipo de estimación chámase estimación por intervalos.

11.3 Distribución das medias mostrais

Estamos interesados, por exemplo, en coñecer a media dos gastos mensuais en alimentación dos fogares dun barrio de Madrid. Temos, pois, unha poboación: os fogares dun barrio particular e unha variable aleatoria, X , que asigna a cada fogar a cantidade de diñeiro dedicada mensualmente á alimentación. Queremos achar μ , valor medio desta variable aleatoria.

Podemos facer unha estimación de μ a partir da media obtida dunha mostra aleatoria da poboación, \bar{x} . É evidente que este valor \bar{x} pode variar se tomamos outra mostra. Algo nos di que a media \bar{x} , dunha mostra determinada, non corresponde exactamente coa media μ da poboación.

¿Existe algunha relación entre a media das mostras \bar{x} , e a media da poboación μ ?

Si existe. Se tomamos unha mostra de tamaño n da poboación e calculamos a media mostral, \bar{x}_1 , a continuación eliximos outra mostra de tamaño n e calculamos a súa media, obtemos outro número, \bar{x}_2 ; procedendo do mesmo xeito ata elixir k mostras diferentes obtemos unha serie de k medias mostrais

$$\bar{x}_1, \bar{x}_2, \bar{x}_3, \dots, \bar{x}_k$$

Considerando os valores das medias mostrais como unha variable aleatoria que simbolizaremos por \bar{X} , imos estudar esta variable e a súa relación coa variable X dos gastos mensuais en alimentación da nosa poboación.

A esta nova variable aleatoria \bar{X} chámase variable das medias mostrais e á distribución dos valores de \bar{X} sobre o conxunto das mostras de tamaño n chámasele distribución das medias mostrais. Es ta variable \bar{X} ten unha media $\mu_{\bar{X}}$ e unha desviación típica que simbolizaremos por $\sigma_{\bar{X}}$. ¿Que relación existe entre $\mu_{\bar{X}}$ e $\sigma_{\bar{X}}$, media e desviación típica de toda a poboación?

Pódense demostrar as afirmacións que seguen.

1ª) Se a variable X segue unha distribución normal, $N(\mu, \sigma)$, entón a variable aleatoria das medias mostrais, \bar{X} , segue tamén unha distribución normal $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$. É dicir, ten a mesma media que X , $\mu_{\bar{X}} = \mu$, e a súa desviación típica é menor, $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$

2ª) Se a variable X segue unha distribución descoñecida ou non é normal, e o tamaño da mostra $n \geq 30$, entón \bar{X} segue tamén unha distribución normal $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$

Na demostración da segunda afirmación emprégase o chamado Teorema Central de Límite, un dos resultados máis importantes en estatística. Este teorema pon de relevo a importancia da distribución normal, que aparece asociada a calquera distribución, tanto se consideramos as súas medias mostrais como a suma independente desa distribución varias veces

Resumindo :

$$\text{Se } X \text{ é } N(\mu, \sigma) \Rightarrow \bar{X} \text{ é } N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

$$\text{Se } X \text{ non é normal ou é descoñecido} \Rightarrow \bar{X} \text{ é } N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \text{ cando } n \geq 30.$$

¿Que acontece cando $n < 30$? Ben, cando as mostras son pequenas, as cousas complícanse un pouco máis, non moito, pero non é un obxectivo deste curso.

Exemplos

9. Nun servizo de atención ao cliente, o tempo de espera ata recibir atención é unha variable aleatoria normal de media 10 minutos e desviación típica 2 minutos. Tómanse mostras aleatorias do tempo de espera dos clientes que chegan en día concreto. Pídese:
- Cal é a probabilidade de que o tempo medio de espera dunha mostra de 25 clientes non supere os 9 minutos?
 - Cal é a probabilidade da media mostral, se toman mostras aleatorias de 64 clientes? Especificar os seus parámetros.

Solución:

- a) As mostras de calquera tamaño n , maior ou menor que 30, dunha poboación $N(\mu, \sigma)$ distribúense segundo a normal $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$. Neste caso, $n = 25$ e X é unha normal $N(10, 2)$; polo que a distribución de \bar{X} das medias mostrais é unha normal $N(10, 2/\sqrt{25})$ ou $N(10, 2/5)$. En consecuencia, temos que calcular:
- $$P[\bar{X} < 9] = P\left[\frac{\bar{X}-10}{2/5} < \frac{9-10}{2/5}\right] = P[Z < -2,5] = 1 - P[Z < 2,5] = 1 - 0,9938 = 0,0062$$
- b) Se $n = 64$ a variable aleatoria das medias mostrais distribúense segundo a normal $N(10, 2/\sqrt{64})$ ou $N(10, 0,25)$, é dicir, con $\mu = 10$ e $\sigma = 0,25$.
10. Admítese que o perímetro cranial de certa especie animal segue unha distribución normal con $\sigma = 12,8$ cm. Toma unha mostra de 10 individuos ao chou. ¿Cal é a probabilidade de que a media da mostra, $\bar{\mu}$, difira de μ , media poboacional, 4,5 cm?

Solución:

Como X se distribúe segundo unha normal $N(\mu, 12,8)$, \bar{X} distribúese segundo a normal $N(\mu, 12,8/\sqrt{10})$, independentemente do tamaño da mostra. Temos que

calcular $P[|\bar{x} - \mu| \geq 4,5]$. Dividindo por $\frac{\sigma}{\sqrt{n}} = \frac{12,8}{10}$ tipificamos a variable e temos

$$P\left[\frac{|\bar{x} - \mu|}{12,8/\sqrt{10}} > \frac{4,5}{12,8/\sqrt{10}}\right] = P\left[|Z| > \frac{4,5}{12,8/\sqrt{10}}\right] = P[|Z| > 1,11] = 2 \cdot P[Z < -1,11] = \\ = 2 [1 - 0,8865] = 2 \cdot 0,1135 = 0,227$$

11. Rexístrouse o peso dos recém nados dunha maternidade durante un ano e observouse que se distribúen con media $\mu = 3250$ g e desviación típica $\sigma = 250$ g. ¿Cal é a probabilidade de que a media dunha mostra de 100 acabados de nacer sexa superior a 3300 g?

Solución:

Descoñecemos a natureza da distribución da variable X , que nos dá os pesos dos recém nados, pero como o tamaño da mostra é $n = 100$, maior que 30, entón a variable \bar{X} segue unha distribución normal $N(3250, \frac{250}{\sqrt{100}}) = N(3250, 25)$. Temos que calcular:

$$P[\bar{X} > 3300] = P\left[\frac{\bar{X} - 3250}{25} > \frac{3300 - 3250}{25}\right] = P[Z > 2] = 0,0228.$$

12. Nun exame tipo test de 100 preguntas, cualificado a punto por pregunta correcta, as cualificacións distribúense con media $\mu = 65$ puntos e desviación típica $\sigma = 14$ puntos. Calcula a probabilidade de que unha mostra de 50 exames elixidos ao chou teña nota media superior a 70 puntos.

Solución:

Descoñecemos a natureza da distribución da variable X , que nos dá as cualificacións dos exames, pero como o tamaño da mostra é $n = 50$, maior que 30, entón a variable segue unha distribución normal $N(65, 14/\sqrt{50}) = N(65, 1,97)$. Temos que calcular:

$$P[\bar{X} > 70] = P\left[\frac{\bar{X} - 65}{1,97} > \frac{70 - 65}{1,97}\right] = P[Z > 2,53] = 1 - P[Z < 2,53] = 0,0057$$

11.4 Distribución de proporcións mostrais

Se ao facer o estudo dunha característica dunha poboación atopamos que a variable únicamente pode tomar dous valores: éxito ou fracaso, a poboación que tratamos de estudar segue unha distribución binomial, pero cando o número de probas é grande esta binomial pódese aproximar por unha normal.

Consideremos entón unha poboación numerosa de N individuos; por exemplo, todos os profesores de matemáticas da Comunidade de Madrid e unha característica: *ser seguidor do Atlético de Madrid*; interroguemos cada profesor se é, ou non, seguidor do Atlético. Sexa n_A o número de respostas afirmativas, entón o cociente

$$p = \frac{n_A}{N}$$

dáanos a proporción poboacional ou porcentaxe poboacional de seguidores do Atlético entre os profesores de matemáticas madrileños. Empeños ambos os dous, tanto seguir ao Atlético como ensinar matemáticas, de xente sufrida.

Imaxinemos que da mencionada poboación tomamos unha mostra de tamaño n , chamémoslle M_1 , e calculemos a proporción de atléticos: $p_1 = \frac{n_1}{n}$. Se a continuación obtemos todas as mostras posibles, de tamaño n , M_2, M_3, M_4, \dots , entón veremos que as proporcións de atléticos

$p_2 = \frac{n_2}{n}, p_3 = \frac{n_3}{n}, p_4 = \frac{n_4}{n}, \dots$ variarán dunha mostra a outra.

Todas estas proporcións p_1, p_2, p_3, \dots son valores dunha nova variable aleatoria que chamaremos proporción mostral e simbolizaremos por \hat{p} . Esta variable á súa vez terá unha media $\mu_{\hat{p}}$ e unha desviación típica $\sigma_{\hat{p}}$. ¿Que relación teñen a media e a desviación típica da proporción mostral coa proporción poboacional p ?

Ben, pois pódese demostrar a seguinte afirmación:

Se unha poboación numerosa ten unha proporción poboacional p dunha determinada característica, entón a variable aleatoria \hat{p} , das proporcións mostrais extraídas desa poboación, cando o tamaño da mostra é suficientemente grande $n \geq 30$, aproxímase a unha

distribución normal de media $\mu_{\hat{p}} = p$ e desviación típica $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$; é dicir, \hat{p} distribúese case segundo $N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$

Exemplos

13. Nunha empresa fuma o 40% dos seus traballadores. Se eliximos unha mostra de 30 persoas, cal é a probabilidade de que a proporción de fumadores na mostra sexa maior que o 50%

Solución:

A mostra é de tamaño $n = 30$ e a proporción poboacional de fumadores é $p = 0,4$,

logo \hat{p} é unha normal $N\left(0,4, \sqrt{\frac{0,4 \cdot 0,6}{30}}\right) = N(0,4; 0,089)$. Temos que calcular:

$$P[\hat{p} > 0,5] = P\left[\frac{\hat{p}-0,4}{0,089} > \frac{0,5-0,4}{0,089}\right] = P\left[Z > \frac{0,5-0,4}{0,089}\right] = P[Z > 1,12] = 1 - P[Z < 1,12] = 0,1314.$$

14. Nunhas eleccións municipais a lista do alcalde saíu co 35% dos votos. Se antes das eleccións se fixese unha sondaxe cunha mostra de 400 veciños, cal sería, de manterse a mesma intención de voto, a probabilidade de obter menos do 30% dos votos para a citada lista?

Solución:

Tamaño da mostra $n = 400$ e $p = 0,35$, logo \hat{p} segue unha distribución normal

$N\left(0,35; \sqrt{\frac{0,35 \cdot 0,65}{400}}\right) = N(0,35; 0,023)$. Temos que calcular;

$$P[\hat{p} < 0,3] = P\left[\frac{\hat{p}-0,35}{0,023} > \frac{0,3-0,35}{0,023}\right] = P\left[Z < \frac{-0,05}{0,023}\right] = P[Z < -2,08] = 1 - P[Z < 2,08] = 0,0188.$$

Táboa da distribución normal, $N(0,1)$

Z	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
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	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	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	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	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	0,7995	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	0,9996	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
4,0	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000