

# 12

## Inferencia estatística. Intervalo de confianza e contraste de hipótese

Proxéctase a creación dun centro comercial nunha cidade e quérese saber o poder adquisitivo dos habitantes, dato que influirá no tamaño do centro comercial. Establécese unha variable aleatoria  $X$  que indica o salario mensual das persoas maiores de idade na cidade. Suponse que  $X$  segue unha distribución normal, pero ignórase cal é a súa media,  $\mu$ , e a súa desviación típica,  $\sigma$ , para dispoñer dun modelo completamente determinado da distribución de  $X$  co que poder facer cálculos de probabilidade do poder adquisitivo, predicións de consumo, etc.

O modo de coñecer, ou mellor dito, de axustar os parámetros,  $\mu$  e  $\sigma$ , da distribución da variable aleatoria  $X$  é extraendo unha mostra. Estimar un parámetro consiste en obter unha aproximación do seu valor baseándose nos datos da mostra.

Hai tres cousas que se poden facer para axustar o valor do parámetro dunha poboación: a) estimar o parámetro descoñecido por un número que se obtén en función dos datos da mostra (isto chámase estimación por punto ou puntual); b) determinar un intervalo ao cal é moi probable que pertenza o parámetro (isto chámase estimación por intervalo de confianza); c) aventurar, a xeito de hipótese, un número como valor do parámetro buscado e contrastar con datos da mostra se estes apoian ou rexeitan a hipótese formulada.

<b>1. INTERVALO DE CONFIANZA PARA A MEDIA.....</b>	<b>2</b>
1.1. Tamaño da mostra.....	4
1.2. Intervalo de confianza para a media dunha distribución normal con $\sigma$ descoñecida.....	5
<b>2. INTERVALO DE CONFIANZA PARA A PROPORCIÓN.....</b>	<b>6</b>
<b>3. CONTRASTE DE HIPÓTESE PARA A MEDIA.....</b>	<b>8</b>
3.1. Contraste bilateral.....	9
3.2. Contraste unilateral pola esquerda.....	11
3.3. Contraste unilateral pola dereita.....	12
<b>4. CONTRASTE DE HIPÓTESE PARA A PROPORCIÓN.....</b>	<b>13</b>

## 12.1 Intervalo de confianza para a media

É moi improbable que a media poboacional,  $\mu$ , coincida co valor medio dunha mostra  $\bar{X}$  (estimación puntual); non obstante, é mais interesante determinar un intervalo centrado na media mostral, de modo que o valor descoñecido de  $\mu$  se atope no devandito intervalo cunha probabilidade tan grande como desexemos. A esta probabilidade chamámola nivel de confianza e simbolizámola por  $1 - \alpha$ , mentres que a  $\alpha$  denominámolo nivel de significación ou nivel de risco.

Concretando, temos que buscar un número  $c$ , de modo que  $\mu$  pertenza ao intervalo

$(\bar{X} - c, \bar{X} + c)$  con probabilidade  $1 - \alpha$ ; é dicir, que cumpra que

$$P[\bar{X} - c < \mu < \bar{X} + c] = 1 - \alpha$$

Despexando  $c$  nas desigualdades do corchete anterior,

$$\bar{X} - c < \mu \Rightarrow \bar{X} - \mu < c$$

$$\mu < \bar{X} + c \Rightarrow -c < \bar{X} - \mu$$

podemos escribir a probabilidade así:

$$P[-c < \bar{X} - \mu < c] = 1 - \alpha,$$

E empregando o valor absoluto queda:

$$P[|\bar{X} - \mu| < c] = 1 - \alpha$$

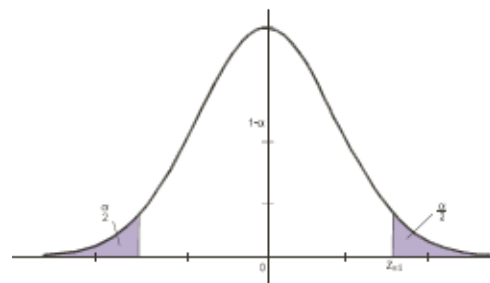
Supoñendo que coñecemos  $\sigma$ , dado  $\bar{X}$  que se distribúe segundo una  $N(\mu, \frac{\sigma}{\sqrt{n}})$ , ao dividir a desigualdade por  $\frac{\sigma}{\sqrt{n}}$  tipificamos a variable  $\bar{X}$ ,

$$P\left[\frac{|\bar{X} - \mu|}{\sigma/\sqrt{n}} < \frac{c}{\sigma/\sqrt{n}}\right] = P\left[|Z| < \frac{c}{\sigma/\sqrt{n}}\right] = 1 - \alpha$$

Onde  $Z$  é a  $N(0,1)$ . Buscamos nas táboas o valor de  $\frac{c}{\sigma/\sqrt{n}}$  correspondente á probabilidade  $1 - \alpha$  e chamamos

a ese valor  $z_{\alpha/2}$ , despexando  $c$  obtemos  $c = \frac{z_{\alpha/2} \cdot \sigma}{\sqrt{n}}$

¿Que é  $z_{\alpha/2}$ ? É o valor dunha abscisa da  $N(0,1)$  que deixa á súa dereita unha área de probabilidade  $\alpha/2$ .



En consecuencia, o intervalo buscado é  $\left(\bar{X} - \frac{z_{\alpha/2} \cdot \sigma}{\sqrt{n}}, \bar{X} + \frac{z_{\alpha/2} \cdot \sigma}{\sqrt{n}}\right)$

que obviamente depende da probabilidade elixida. Trátase, polo tanto, dun intervalo centrado na media da mostra tomada,  $\bar{X}$ , e de radio  $\frac{z_{\alpha/2} \cdot \sigma}{\sqrt{n}}$ . Este raio do intervalo de confianza denomínase *error máximo*, e é a máxima diferenza que pode existir entre  $\mu$  e a media da mostra elixida,  $\bar{X}$ , para un nivel de confianza  $1-\alpha$ .

Simbolízase por  $E = \frac{z_{\alpha/2} \cdot \sigma}{\sqrt{n}}$

Da simple observación desta fórmula vemos que ao aumentar o tamaño da mostra,  $n$ , diminúe o erro  $E$ ; mentres que ao aumentar o nivel de confianza, aumenta  $z_{\alpha/2}$  e, polo tanto, o erro cometido.

## Exemplos

1. Nun laboratorio obtivéronse 6 estimacións de PH dunha solución cos resultados seguintes

7,91 7,94 7,90 7,93 7,89 7,91

Suponse que a poboación de todas as determinacións do pH da solución ten unha distribución normal de media descoñecida e desviación típica 0,02.

- Determinéase un intervalo de confianza ao 98% para a media de todas as determinacións do pH da mesma solución obtidas co mesmo método.
- Co mesmo nivel de confianza anterior, cal debe ser o tamaño mínimo da mostra para que a amplitude do intervalo de confianza sexa como máximo 0,01?

### Solución:

- a) Determinamos o intervalo de confianza ao 98% para a media. Sabemos que

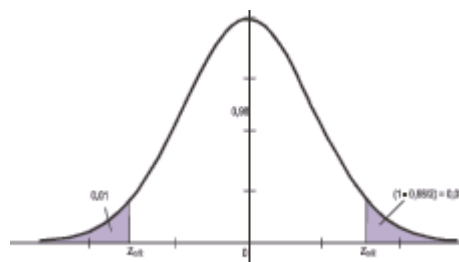
$$\text{nivel de confianza} = 1-\alpha = 98\% = 0,98$$

Temos que achar  $c$  tal que  $P[|\bar{X} - \mu| < c] = 0,98$ . Ademais coñecemos que se distribúe segundo unha normal  $N(\mu, 0,02/\sqrt{6})$ . Tipificamos  $\bar{X}$ ,

$$P\left[\frac{|\bar{X} - \mu|}{0,02/\sqrt{6}} < \frac{c}{0,02/\sqrt{6}}\right] = P\left[|Z| < \frac{c}{0,02/\sqrt{6}}\right] = 0,98$$

Chamamos  $z_{\alpha/2} = \frac{c}{\sigma/\sqrt{n}}$ , e buscamos nas táboas da  $N(0,1)$  o valor de  $z_{\alpha/2}$  que corresponde á abscisa da  $N(0,1)$  que vemos na figura

e que calculamos da mesma figura sabendo que  $P[Z < z_{\alpha/2}] = 0,01 + 0,98 = 0,99$



Nas táboas da  $N(0,1)$  lemos  $z_{\alpha/2} = 2,33$ ,  $2,33 = \frac{c}{0,02/\sqrt{6}}$ ,

Despexando  $c$  obtemos  $c = \frac{2,33 \cdot 0,02}{\sqrt{6}}$

Como a media da mostra é  $\bar{X} = \frac{7,91+7,94+7,90+7,93+7,89+7,91}{6} = 7,91$ ,

podemos escribir o intervalo buscado

$$\left(\bar{X} - \frac{z_{\alpha/2} \cdot \sigma}{\sqrt{n}}, \bar{X} + \frac{z_{\alpha/2} \cdot \sigma}{\sqrt{n}}\right) = (7,91 - 0,019, 7,91 + 0,019) = (7,891, 7,929).$$

Isto significa que o 98% das mostras de tamaño 6 teñen unha media mostral que difire ou dista da media poboacional menos que 0,019.

## 1.1. Tamaño da mostra

Outro problema relacionado co intervalo de confianza é: cal debe ser o tamaño da mostra para que cun nivel de confianza determinado 90%, 95%, 99% ou outro, a diferenza entre a media mostral e a media poboacional sexa menor que un valor fixo? Ou, o que é o mesmo, cal debe ser o tamaño da mostra para que cun nivel de confianza establecido o erro máximo cometido sexa menor que unha cantidade dada?

Nos problemas do intervalo de confianza dannos o nivel de confianza, o tamaño da mostra e pídennos achar a diferenza (o erro) entre a media poboacional e a media mostral. Nos problemas de tamaño da mostra dannos o nivel de confianza, a diferenza máxima entre  $\bar{X}$  e  $\mu$ , é dicir, o erro máximo admitido temos que achar o tamaño da mostra. Na práctica, unicamente temos que despxear  $n$  na fórmula do erro máximo

$$E = \frac{z_{\alpha/2} \cdot \sigma}{\sqrt{n}} \Rightarrow \sqrt{n} = \frac{z_{\alpha/2} \cdot \sigma}{E} \Rightarrow n = \left(\frac{z_{\alpha/2} \cdot \sigma}{E}\right)^2$$

## Exemplos

2. Agora resolvemos o apartado b) do problema do exemplo 1.

b) Co mesmo nivel de confianza anterior, cal debe ser o tamaño mínimo da mostra para que a amplitude do intervalo de confianza sexa como máximo 0,01?

**Solución:**

$$\text{Na fórmula do erro máximo despxamos } n: E = \frac{z_{\alpha/2} \cdot \sigma}{\sqrt{n}} \Rightarrow \sqrt{n} = \frac{z_{\alpha/2} \cdot \sigma}{E} \Rightarrow n = \left(\frac{z_{\alpha/2} \cdot \sigma}{E}\right)^2$$

Recordamos que  $\sigma = 0,02$  e que  $1 - \alpha = 98\% = 0,98$  polo que  $z_{\alpha/2} = 2,33$ , como vimos no apartado a), e ademais queremos achar o tamaño da mostra,  $n$ , para que como máximo o erro sexa  $E = 0,01$ , é dicir,

$$\frac{z_{\alpha/2} \cdot \sigma}{\sqrt{n}} < 0,01 \Rightarrow \sqrt{n} > \frac{z_{\alpha/2} \cdot \sigma}{0,01}. \text{ Substituíndo na formula,}$$

$$n = \left(\frac{z_{\alpha/2} \cdot \sigma}{E}\right)^2, \text{ resulta } n = \left(\frac{2,33 \cdot 0,02}{0,01}\right)^2 = 21,715$$

Como ha de ser  $n > 21$ , entón o tamaño mínimo da mostra é  $n = 22$ .

3. O tempo de vida dunha clase de depuradoras de auga utilizadas nunha planta industrial distribúese normalmente cunha desviación típica de 2000 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) Áchese un intervalo de confianza ao 99% para a vida media das depuradoras.

b) Calcúlese o tamaño mínimo que debería ter a mostra, no caso de admitir un erro máximo de 500 horas, cun grao de confianza do 95%.

### Solución:

A variable  $X$  é o tempo de vida das depuradoras que se distribúe segundo unha  $N(\mu, 2000)$ . Imos resolver este problema dun xeito máis directo que no exemplo 1.

a) O intervalo de confianza para a media vén dado pola expresión:

$$\left( \bar{X} - \frac{z_{\alpha/2} \cdot \sigma}{\sqrt{n}}, \bar{X} + \frac{z_{\alpha/2} \cdot \sigma}{\sqrt{n}} \right)$$

Imos calcular cada un dos ingredientes da expresión anterior.

Como  $1-\alpha = 99\% = 0,99$ , entón de  $P[Z < z_{\alpha/2}] = 0,005 + 0,99 = 0,995$  e obtemos nas táboas que  $z_{\alpha/2} = 2,575$ ; ademais,  $\sigma = 2000$  horas, o tamaño da mostra  $n = 9$  e

$$\bar{X} = \frac{9,5+10+7,5+10,5+16,5+10+12+32+10}{9} = 14, \text{ en miles de horas, } \bar{X} = 14000 \text{ horas.}$$

Substituíndo na expresión do intervalo de confianza queda

$$\left( \bar{X} - \frac{z_{\alpha/2} \cdot \sigma}{\sqrt{n}}, \bar{X} + \frac{z_{\alpha/2} \cdot \sigma}{\sqrt{n}} \right) = \left( 14000 - \frac{2,575 \cdot 2000}{\sqrt{9}}, 14000 + \frac{2,575 \cdot 2000}{\sqrt{9}} \right) =$$

$$= (14000 - 1716,67, 14000 + 1716,67) = (12283,33, 15716,67).$$

Isto significa que o 99% das mostras de tamaño 9 teñen unha media mostral que difire ou dista da media poboacional menos que 1716,66 horas.

b) Na fórmula do erro máximo desexamos n:

$$E = \frac{z_{\alpha/2} \cdot \sigma}{\sqrt{n}} \Rightarrow \sqrt{n} = \frac{P[Z < z_{\alpha/2}] \cdot \sigma}{E} \Rightarrow n = \left( \frac{z_{\alpha/2} \cdot \sigma}{E} \right)^2.$$

Recordamos que  $\sigma = 2000$  horas, agora o grao de confianza ou nivel de confianza é  $1-\alpha = 95\% = 0,95$  polo que de  $P[Z < z_{\alpha/2}] = 0,025 + 0,95 = 0,975$ , obtemos que  $z_{\alpha/2} = 1,96$ .

Queremos achar o tamaño da mostra,  $n$ , para que o erro máximo admitido sexa  $E = 500$  horas. Substituíndo na última fórmula,

$$n = \left( \frac{z_{\alpha/2} \cdot \sigma}{E} \right)^2, \text{ resulta } n = \left( \frac{1,96 \cdot 2000}{500} \right)^2 = 61,46$$

O tamaño mínimo da mostra é  $n = 62$ .

## 1.2. Intervalo de confianza para a media dunha distribución normal con $\sigma$ descoñecida

No desenvolvemento anterior supuxemos que a desviación típica  $\sigma$  é coñecida, claro que alguén pode pensar ¿como é posible coñecer  $\sigma$  sen coñecer  $\mu$ .?

Cando se descoñece o valor da desviación típica,  $\sigma$ , substitúese pola desviación típica mostral,  $S$ , que nalgúñas calculadoras áchase coa tecla  $\sigma_{n-1}$ . O malo é que se empregamos a desviación típica mostral,  $S$ , a tipificación da variable  $\bar{X}$

$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$

Non segue unha distribución normal  $N(0,1)$ , senón unha distribución  $t$  de Student que non é obxectivo deste curso, aínda que tamén é certo que a  $t$  de Student se aproxima á  $N(0,1)$  cando  $n$  é grande.

## 2. Intervalo de confianza para a proporción

Sabemos que se unha poboación ten unha proporción poboacional  $p$  dunha determinada característica, entón a variable aleatoria,  $\hat{p}$  das proporcións mostrais, cando  $n$  é grande ou  $n \cdot p > 5$ , tende a unha distribución normal de media  $\mu_{\hat{p}} = p$  e desviación típica  $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$ ; é dicir, a distribución de  $\hat{p}$  aproxímase moito a unha  $N(p, \sqrt{\frac{p(1-p)}{n}})$ . Aínda que non hai unanimidade, considérase  $n$  grande cando  $n \geq 30$ .

Coñecido o valor da proporción  $\hat{p}$  dunha mostra, podemos determinar un intervalo para  $p$ , a proporción poboacional da característica, do mesmo modo que fixemos no caso do intervalo de confianza para a media. Buscamos un intervalo no que o valor descoñecido de  $p$  se atope cunha probabilidade ou nivel de confianza  $1 - \alpha$ , onde  $\alpha$  indica o nivel de significación.

Non obstante, en vez de seguir o mesmo camiño que no intervalo de confianza para a media imos facer outra cousa. Pódese demostrar que se  $n$ , tamaño da mostra, é grande, o cociente

$$\frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

segue aproximadamente unha distribución normal  $N(0,1)$ ; é dicir, que

$$P[|Z| < z_{\alpha/2}] = P\left[\frac{|\hat{p} - p|}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} < z_{\alpha/2}\right] = 1 - \alpha$$

Da desigualdade do segundo corchete

$$\frac{|\hat{p} - p|}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} < z_{\alpha/2}$$

Obtemos  $|\hat{p} - p| < z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$  ou  $-z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < \hat{p} - p < z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

Multiplicando por  $-1$  as desigualdades anteriores cambian de sentido

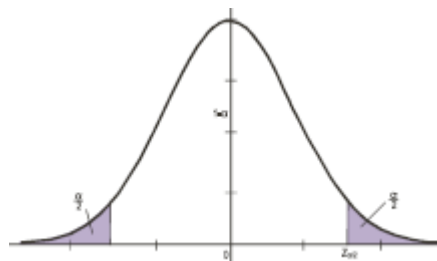
$$z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} > -\hat{p} + p > -z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

E sumando  $\hat{p}$  a todas as desigualdades chégase a  $\hat{p} + z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} > p > \hat{p} - z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

que, cambiando a orde, tamén se pode escribir  $\hat{p} - z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

Logo o intervalo de confianza para a proporción cando  $n$  é grande é aproximadamente

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



Onde  $z_{\alpha/2}$ , como no caso do intervalo de confianza para a media, é o valor dunha abscisa da  $N(0, 1)$  que deixa á súa dereita unha área de probabilidade  $\alpha/2$

Trátase dun intervalo centrado na proporción elixida,  $\hat{p}$ , e de raio  $z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ . A este raio, como no caso do intervalo para a media, denomínase *erro máximo*; e é o erro que se comete ao estimar  $p$  mediante  $\hat{p}$ , e tamén se simboliza por

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

Como no apartado anterior, agora estamos en condicións de achar o tamaño da mostra para a proporción. Coñecemos o nivel de confianza, a diferenza máxima entre  $\hat{p}$  e  $p$ , é dicir, o erro máximo admitido, e queremos determinar o tamaño da mostra. Isto conséguese despegando  $n$  na fórmula:

$$E = z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \Rightarrow E^2 = (z_{\alpha/2})^2 \frac{\hat{p}(1-\hat{p})}{n} \Rightarrow n = (z_{\alpha/2})^2 \frac{\hat{p}(1-\hat{p})}{E^2}$$

## Exemplos

4. Nunha comunidade autónoma faise unha enquisa para coñecer se os contribuíntes están a favor da implantación dun novo imposto. Interrógase a unha mostra de 600 contribuíntes e o resultado é favorable só en 225 casos.
- Cun nivel de significación do 5% establece un intervalo de confianza para a proporción de contribuíntes favorables ao imposto.
  - Cal ha de ser o tamaño da mostra para que a diferenza entre a proporción poboacional  $p$  e a proporción mostrai  $\hat{p}$  sexa menor que 0,02, co mesmo nivel de confianza?

### Solución:

- a) O intervalo de confianza para a proporción poboacional  $p$  é

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

Calculamos os ingredientes do intervalo de confianza:  $\alpha = 5\% = 0,05$ ,  $1-\alpha = 0,95$ ; ademais

$n = 600 \geq 30$ ,  $\hat{p} = 225/600 = 0,375$  e  $z_{\alpha/2}$  calculámolo das táboas, sabendo que  $P[Z < z_{\alpha/2}] = 0,95 + 0,025 = 0,975$ , e resulta  $z_{\alpha/2} = 1,96$ . O intervalo buscado será:

$$\left( 0,375 - 1,96 \cdot \sqrt{\frac{0,375 \cdot 0,625}{600}}, 0,375 + 1,96 \cdot \sqrt{\frac{0,375 \cdot 0,625}{600}} \right) \Rightarrow$$

$$(0,375 - 0,039, 0,375 + 0,039) = (0,336, 0,414)$$

Isto significa que o intervalo de confianza 95% das mostras de tamaño 600 contén a proporción poboacional buscada; tan só o 5% restante non contén á devandita proporción. Á vista do intervalo obtido non parece que os cidadáns estean moi entusiasmados co imposto.

- b) Debemos calcula  $n$  na fórmula  $E = z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$



Onde  $E = 0,02$ ,  $z_{\alpha/2} = 1,96$ , calculámolo antes, porque o nivel de confianza  $1-\alpha = 0,95$ , é o mesmo e  $\hat{p} = 0,375$ . Substituíndo e despxandp n, obtemos

$$0,02 = 1,96 \cdot \sqrt{\frac{0,375 \cdot 0,625}{n}} \Rightarrow 0,02^2 = 1,96^2 \cdot \frac{0,375 \cdot 0,625}{n} \Rightarrow n = 1,96^2 \cdot \frac{0,375 \cdot 0,625}{0,02^2} = 2250,94$$

Logo o tamaño mínimo da mostra é  $n = 2251$ .

### 3. Contraste de hipótese para a media

Outra forma de abordar o problema de estimar a media dunha poboación consiste en formular unha hipótese sobre a media e despois usar a información que proporciona a mostra para confirmar ou rexeitar a devandita hipótese. Esta técnica de inferencia estatística coñécese co nome de contraste de hipótese ou test de hipótese. E, por suposto, suporemos que estamos ante unha variable aleatoria  $X$  que segue unha distribución normal de desviación típica,  $\sigma$ , coñecida.

O proceso do contraste de hipótese realízase nos 4 pasos que indicamos a continuación.

1. Establecer a hipótese que provisionalmente se considera verdadeira. Simbolizamos esta hipótese por  $H_0$ , e consiste en que  $\mu$ , media poboacional, teña un valor  $\mu_0$ ; é dicir,

$$H_0 : \mu = \mu_0$$

Esta  $H_0$  denomínase *hipótese nula* porque se parte do suposto de que a diferenza entre o valor verdadeiro da media e o seu valor hipotético é cero.

A *hipótese alternativa*, simbolizada por  $H_1$ , poderá ser de tres tipos diferentes en función da natureza do problema:

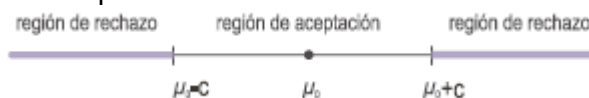
$$H_1 : \mu \neq \mu_0 \text{ (contraste bilateral)}$$

$$H_1 : \mu > \mu_0 \text{ (contraste unilateral pola dereita)}$$

$$H_1 : \mu < \mu_0 \text{ (contraste unilateral pola esquerda)}$$

2. Fixar o nivel de significación  $\alpha$ , que indica a probabilidade de rexeitar  $H_0$  aíndo sendo verdadeira, ou establecer o nivel de confianza  $1-\alpha$ , que indica a probabilidade de aceptar  $H_0$  cando é certa. Os valores de  $\alpha$  máis común son 5%, 1% e 10% (ou 0,05, 0,01 e 0,1) e os de  $1-\alpha$ , 95%, 99% e 90% (ou 0,95, 0,99 e 0,90).
3. Determinar a rexión de aceptación para o nivel de significación  $\alpha$  ou nivel de confianza  $1-\alpha$ , de tal maneira que a probabilidade de aceptar  $H_0$  cando sexa certa coincida con  $1-\alpha$ . Logo a rexión de aceptación será o intervalo  $(\mu_0 - c, \mu_0 + c)$  no que o número  $c$  verifica que :  
$$P[|\bar{X} - \mu_0| < c] = 1-\alpha$$

Gráficamente a rexión de aceptación e a rexión de rexeitamento serían as da figura.



4. Extráese unha mostra e calcúlase a media mostral,  $\bar{X}$ ; a continuación compróbase se cae dentro ou fóra da rexión de aceptación. Se cae dentro acéptase  $H_0$  e se non, rexéitase.



### 3.1 Contraste bilateral

Aparece cando a hipótese alternativamente é simplemente distinta da hipótese nula.

**Paso 1.** Neste caso a hipótese nula e alternativa son:  $H_0 : \mu = \mu_0$

$$H_1 : \mu \neq \mu_0$$

**Paso 2.** Fixamos un nivel de significación  $\alpha$  ou un nivel de confianza  $1-\alpha$ .

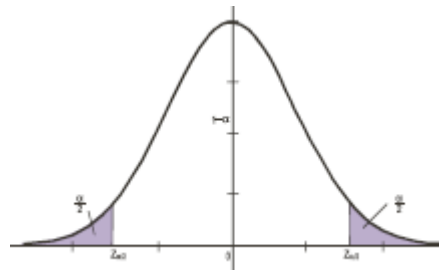
**Paso 3.** Determinación da rexión de aceptación ( $\mu_0 - c$ ,  $\mu_0 + c$ ). O valor  $c$  na rexión de aceptación, para un nivel de significación  $\alpha$  ou un nivel de confianza  $1-\alpha$ , calcúlase da probabilidade

$$P[|\bar{X} - \mu_0| < c] = 1-\alpha$$

Dividindo a desigualdade do corchete por  $\sigma/\sqrt{n}$  tipificamos  $\bar{X}$ ,

$$P\left[\frac{|\bar{X} - \mu_0|}{\sigma/\sqrt{n}} < \frac{c}{\sigma/\sqrt{n}}\right] = 1-\alpha,$$

$$P\left[|Z| < \frac{c}{\sigma/\sqrt{n}}\right] = 1-\alpha,$$



Buscamos nas táboas un valor,  $z_{\alpha/2}$ , da  $N(0, 1)$  tal que

$$P[|Z| < z_{\alpha/2}] = 1-\alpha,$$

Como  $z_{\alpha/2} = \frac{c}{\sigma/\sqrt{n}}$ , entón  $c = \frac{z_{\alpha/2} \cdot \sigma}{\sqrt{n}}$  e a rexión de aceptación será:

$$\left(\mu_0 - \frac{z_{\alpha/2} \cdot \sigma}{\sqrt{n}}, \mu_0 + \frac{z_{\alpha/2} \cdot \sigma}{\sqrt{n}}\right)$$

**Paso 4.** Extráese unha mostra, calcúlase a súa media  $\bar{X}$  e compróbase se cae dentro ou fóra da rexión anterior, se cae dentro aceptamos  $H_0$  e se non, rexeitase.

### Exemplos

5. Os diámetros dunhas pezas cilíndricas producidas por unha máquina seguen unha distribución normal de media descoñecida e desviación típica  $\sigma = 2$  mm. Tómase unha mostra de 25 pezas e obtense un diámetro medio de 36 mm. Pódese afirmar cun nivel de significación de 0,01 que a media do diámetro destas pezas é 40 mm?

**Solución:**

**Paso 1.** A hipótese nula é  $H_0 : \mu = 40$  e a hipótese alternativa,  $H_1 : \mu \neq 40$ . Estamos ante un contraste bilateral.

**Paso 2.** Fixamos un nivel de significación  $\alpha = 0,01$  ou un nivel de confianza  $1-\alpha = 0,99$ .

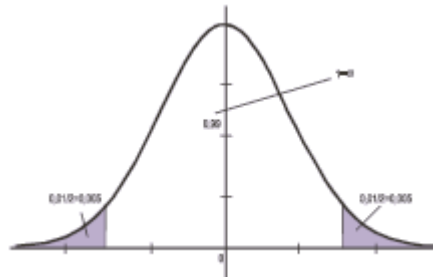
**Paso 3.** Determinamos a rexión de aceptación:

$$\left( \mu_0 - \frac{z_{\alpha/2} \cdot \sigma}{\sqrt{n}}, \mu_0 + \frac{z_{\alpha/2} \cdot \sigma}{\sqrt{n}} \right)$$

Como  $\alpha = 0,01$ ,  $\alpha/2 = 0,005$  e o tamaño da mostra  $n = 25$ , segundo o gráfico e as fórmulas

$$P[|Z| < z_{\alpha/2}] = 1 - \alpha$$

$$P[Z < z_{\alpha/2}] = 0,99 + 0,005 = 0,995$$



Obtemos o valor de  $z_{\alpha/2}$  nas táboas; resultando  $z_{\alpha/2} = 2,575$ ; en consecuencia, a rexión de aceptación será:

$$\left( 40 - \frac{2,575 \cdot 2}{\sqrt{25}}, 40 + \frac{2,575 \cdot 2}{\sqrt{25}} \right) = (38,97, 41,03)$$

**Paso 4.** Para unha mostra de tamaño  $n = 25$  resultou que  $\bar{X} = 36$  mm. É evidente que 36 non pertence ao intervalo (38,97, 41,03), logo rexeitamos  $H_0$ , a media da poboación non é 40 mm.

6. Quérese comprobar se unha máquina destinada á unchedura de botellas de auga mineral sufriu un desaxuste. Unha mostra aleatoria de 10 botellas saídas desa máquina proporciona os seguintes datos:

0,49 0,52 0,51 0,48 0,53 0,55 0,49 0,50 0,52 0,49

Supoñendo que a cantidade de líquido que a máquina deposita en cada envase segue unha distribución normal de media 0,5 l e unha desviación típica de 0,02 l, deséxase contrastar se o contido medio das botellas saídas desa máquina é 0,5 l cun nivel de significación de 5%. Pídese:

- Formular a hipótese nula e alternativa.
- Determinar a rexión crítica do contraste.
- Realizar o contraste.

**Solución:**

- a) **Paso 1.** A hipótese nula e a alternativa son:

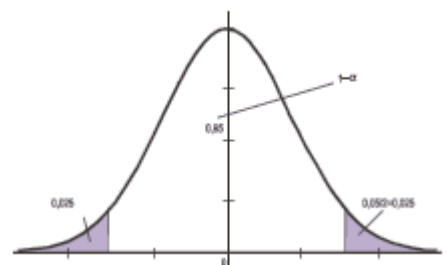
$$H_0 : \mu = 0,5 \text{ l}$$

$$H_1 : \mu \neq 0,5 \text{ l}$$

**Paso 2.** Fixamos o nivel de significación  $\alpha = 5\% = 0,05$  e, por conseguinte, o nivel de confianza  $1 - \alpha = 1 - 0,05 = 0,95$ .

- b) **Paso 3.** Determinación da rexión de aceptación cando  $\sigma = 0,02$  l e o tamaño da mostra  $n = 10$ ,

$$\left( 0,5 - \frac{z_{\alpha/2} \cdot \sigma}{\sqrt{n}}, 0,5 + \frac{z_{\alpha/2} \cdot \sigma}{\sqrt{n}} \right) = \left( 0,5 - \frac{z_{\alpha/2} \cdot 0,02}{\sqrt{10}}, 0,5 + \frac{z_{\alpha/2} \cdot 0,02}{\sqrt{10}} \right).$$



Como  $\alpha = 5\% = 0,05$ ,  $\alpha/2 = 0,025$ , segundo o gráfico vemos que

$$P[|Z| < z_{\alpha/2}] = 1 - 0,05 = 0,95$$

$$P[Z < z_{\alpha/2}] = 0,95 + 0,025 = 0,975$$

Buscando nas táboas o valor de  $z_{\alpha/2}$  obtemos que  $z_{\alpha/2} = 1,96$ ; en consecuencia, a rexión de aceptación será:

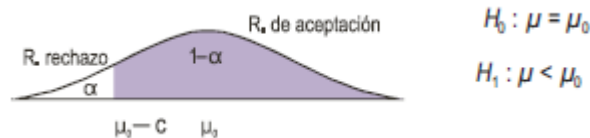
$$\left(0,5 - \frac{z_{\alpha/2} \cdot \sigma}{\sqrt{n}}, 0,5 + \frac{z_{\alpha/2} \cdot \sigma}{\sqrt{n}}\right) = \left(0,5 - \frac{1,96 \cdot 0,02}{\sqrt{10}}, 0,5 + \frac{1,96 \cdot 0,02}{\sqrt{10}}\right) = (0,5 - 0,012, 0,5 + 0,012) = (0,488, 0,512).$$

A rexión crítica ou rexión de rexeitamento está formada polo complementario da rexión de aceptación, é dicir, os intervalos  $(-\infty, 0,488)$  e  $(0,512, \infty)$

c) **Paso 4.** Calculamos  $\bar{X}$ ,  $\bar{X} = \frac{0,49 + 0,52 + \dots + 0,49}{10} = 0,508$ . Como 0,508 está dentro do intervalo  $(0,488, 0,512)$ , aceptamos a hipótese nula.

### 3.2 Contraste unilateral pola esquerda

Hai problemas en que nos interesa decidir se a hipótese alternativa é menor que o valor da hipótese nula. Estamos entón ante un contraste unilateral pola esquerda, cuxas hipóteses son:



Neste caso a rexión de aceptación é o intervalo  $(\mu_0 - c, \infty)$  e a rexión crítica ou rexión de rexeitamento  $(-\infty, \mu_0 - c)$  está situada á esquerda da distribución das medias mostrais.

Para calcular  $c$  partimos de  $P[\bar{X} > \mu_0 - c] = 1 - \alpha$ .

Restando  $\mu_0$  e dividindo por  $\sigma/\sqrt{n}$  tipificamos  $\bar{X}$ :

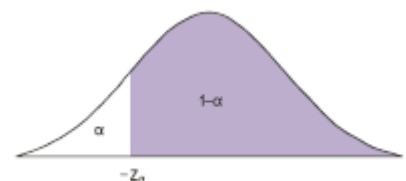
$$P\left[\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} > \frac{\mu_0 - c - \mu_0}{\sigma/\sqrt{n}}\right] = 1 - \alpha,$$

$$P\left[Z > \frac{-c}{\sigma/\sqrt{n}}\right] = 1 - \alpha \quad \text{ou} \quad P[Z > -z_\alpha] = 1 - \alpha$$

Buscamos nas táboas un valor,  $-z_\alpha$ , da  $N(0, 1)$  tal que  $P[Z > -z_\alpha] = P[Z < z_\alpha] = 1 - \alpha$ .

Entón, como  $z_\alpha = \frac{c}{\sigma/\sqrt{n}} \Rightarrow c = \frac{z_\alpha \cdot \sigma}{\sqrt{n}}$ , a rexión de aceptación é:

$$\left(\mu_0 - \frac{z_\alpha \cdot \sigma}{\sqrt{n}}, \infty\right)$$



## Exemplo

7. A vida media das lámpadas de 60 w dunha determinada marca está garantida polo fabricante en 800 horas, cunha desviación típica de 120 horas. Escóllese unha mostra de 50 lámpadas e despois de deixalas acendidas ininterrompidamente atópase que teñen unha vida media de 750 horas. Haberá que rexeitar a afirmación do fabricante cun nivel de confianza de 95%?

### Solución:

Seguiremos o mesmo procedemento de resolución que no contraste bilateral.

**Paso 1.** As hipóteses son  $H_0 : \mu = 800$

$$H_1 : \mu < 800.$$

Trátase dun contraste unilateral pola esquerda

**Paso 2.** Fixamos o nivel de significación  $1-\alpha = 95\% = 0,95$ ,  $\alpha = 0,05$

**Paso 3.** Determinamos a rexión de aceptación. Da probabilidade

$P[Z > -z_\alpha] = P[Z < z_\alpha] = 0,95$  obtemos nas táboas 1,645. Entón a rexión de aceptación é

$$\left(\mu_0 - \frac{z_\alpha \cdot \sigma}{\sqrt{n}}, \infty\right) = \left(800 - 1,645 \cdot \frac{120}{\sqrt{50}}, \infty\right) = (800 - 27,9, \infty) = (772,08, \infty)$$

**Paso 4.** Para aceptar  $H_0$  ten que acontecer que  $\bar{X}$  caia no intervalo  $(772,08, \infty)$ .

Evidentemente  $\bar{X} = 750$  non pertence ao intervalo  $(772,08, \infty)$ . Logo rexeítase a hipótese nula. Non é certo que as lampadas teñan unha vida media de 800 horas.

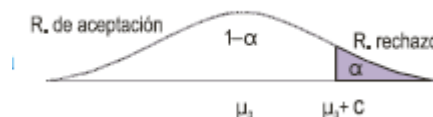
## 3.2 Contraste unilateral pola dereita

Noutros casos debemos contrastar se a hipótese alternativa é maior que a hipótese nula. Estamos entón ante un contraste unilateral pola dereita cuxas hipóteses son

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu > \mu_0$$

Neste caso a rexión aceptación é o intervalo  $(-\infty, \mu_0 + c)$  e a rexión crítica ou rexión de rexeitamento  $(\mu_0 + c, \infty)$  está situada á dereita da distribución das medias mostral.



Para calcular  $c$  partimos de  $P[\bar{X} < \mu_0 + c] = 1-\alpha$ . Restando  $\mu_0$  e dividindo por  $\sigma/\sqrt{n}$

tipificamos  $\bar{X}$  :

$$P\left[\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} < \frac{\mu_0+c-\mu_0}{\sigma/\sqrt{n}}\right] = 1-\alpha,$$

$$P\left[Z < \frac{c}{\sigma/\sqrt{n}}\right] = 1-\alpha \quad \text{ou} \quad P[Z < z_\alpha] = 1-\alpha$$



Buscamos nas táboas un valor un valor,  $z_\alpha$ , da  $N(0, 1)$  tal que  $P[Z < z_\alpha] = 1-\alpha$ .

Entón, como  $z_\alpha = \frac{c}{\sigma/\sqrt{n}} \Rightarrow c = \frac{z_\alpha \cdot \sigma}{\sqrt{n}}$ , a rexión de aceptación é:  $\left(-\infty, \mu_0 + \frac{z_\alpha \cdot \sigma}{\sqrt{n}}\right)$

## Exemplos

8. Un servizo telefónico de atención ao cliente asegura que o tempo medio de espera antes de ser atendidos é de 5 minutos e desviación típica 0,6 minutos. Tómase unha mostra de 36 chamadas e botan unha media de espera de 5,2 minutos. Existen razón para crer, cun nivel de significación de 0,05, que o tempo de espera é maior que 5 minutos?

**Solución:**

**Paso 1.** As hipótesis son  $H_0: \mu = 5$

$H_1: \mu > 5$

Trátase dun contraste unilateral pola dereita

**Paso 2.** Fixamos o nivel de significación  $\alpha = 0,05$

**Paso 3.** Determinamos de  $P[Z \leq z_\alpha] = 0,95$ , a rexión de aceptación e nas táboas

obtemos  $z_\alpha = 1,645$ .

Entón a rexión de aceptación é

$$\left(-\infty, \mu_0 + \frac{z_\alpha \cdot \sigma}{\sqrt{n}}\right) = \left(-\infty, 5 + 1,645 \cdot \frac{0,6}{\sqrt{36}}\right) = (-\infty, 5,1645).$$

**Paso 4.** Para aceptar  $H_0$  ten que acontecer que  $\bar{X}$  caia no intervalo  $(-\infty, 5,1645)$ .

Como  $\bar{X} = 5,2$  non pertence ao intervalo  $(-\infty, 5,1645)$ , entón rexeítase a hipótese nula. Non é certo que o tempo medio de espera sexa de 5 minutos.

## 4. Contraste de hipótese para a proporción

O contraste de hipótese para unha proporción ou porcentaxe baséase nos mesmos principios do contraste de hipótese para a media; tan só temos que recordar que se unha poboación ten unha proporción poboacional  $p$  dunha determinada característica, entón a variable aleatoria  $\hat{p}$ , das proporcións mostrais, cando  $n \geq 30$ , aproxímase a unha distribución normal

$$N\left(p, \sqrt{\frac{p(1-p)}{n}}\right).$$

Se poñemos en vez de  $p$ , proporción poboacional, o valor de  $p_0$ , hipótese que queremos

contrastar, entón debe aproximarse a unha normal  $N\left(p_0, \sqrt{\frac{p_0(1-p_0)}{n}}\right)$

O proceso realízase nos 4 pasos que formulamos no contraste da media.

1. Establecer a hipótese que provisionalmente se considera verdadeira,  $H_0$ , que  $p$ , media poboacional, teña un valor  $p_0$ ; é dicir,  $H_0: p = p_0$

A hipótese complementaria da hipótese nula é a hipótese alternativa,  $H_1$ , que pode ser de tres tipos diferentes:

$H_1: p \neq p_0$  (contraste bilateral)

$H_1: p < p_0$  (contraste unilateral pola esquerda)

$H_1: p > p_0$  (contraste unilateral pola dereita)

2. Fixar o nivel de significación  $\alpha$ , que indica a probabilidade de rexeitar  $H_0$  aínda sendo verdadeira, ou establecer o nivel de confianza  $1-\alpha$ , que indica a probabilidade de aceptar  $H_0$  cando é certa.
3. Determinar a rexión de aceptación para o nivel de significación  $\alpha$  supón determinar un intervalo  $(p_0-c, p_0+c)$  ao que pertenza a proporción mostral con probabilidade  $1-\alpha$ . Logo  $c$  é un número que cumpre que:  $P[|\hat{p} - p_0| < c] = 1-\alpha$

Dividindo a desigualdade por  $\sqrt{\frac{p_0(1-p_0)}{n}}$ , tipificamos  $\hat{p}$ ; e chamando  $z_{\alpha/2} = \frac{c}{\sqrt{\frac{p_0(1-p_0)}{n}}}$ ,

$$\text{quédanos : } P\left[\frac{|\hat{p}-p_0|}{\sqrt{\frac{p_0(1-p_0)}{n}}} < z_{\alpha/2}\right] = 1-\alpha.$$

Da desigualdade do corchete obtemos  $|\hat{p} - p_0| < z_{\alpha/2} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$   
que conduce, cando o contraste é bilateral, á rexión de aceptación buscada:

$$\left(p_0 - z_{\alpha/2} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}, p_0 + z_{\alpha/2} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}\right)$$

Nos contrastes unilaterales as rexións de aceptación son:

- Pola esquerda  $\left(p_0 - z_{\alpha/2} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}, \infty\right)$
  - Pola dereita  $\left(-\infty, p_0 + z_{\alpha/2} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}\right)$
4. Extráese unha mostra e calcúlase a proporción mostral,  $\hat{p}$ ; a continuación compróbase se cae dentro ou fóra da rexión de aceptación. Se cae dentro acéptase  $H_0$  e se non, como é moi improbable que a proporción mostral obtida siga a distribución normal das proporcións mostrais, rexéitase.

## Exemplos

9. Un comerciante de produtos informáticos asegura que o 45% dos fogares de certa cidade posúen ordenador. Extráese unha mostra de 300 fogares e resulta que 120 posúen ordenador. Pódese rexeitar a afirmación do comerciante cun nivel de significación do 5%

### Solución:

**Paso 1.** As hipótesis do contraste son:  $H_0 : p = 0,45$  e  $H_1 : p \neq 0,45$  (contraste bilateral)

**Paso 2.** O nivel de significación é  $\alpha = 0,05$  e  $1-\alpha = 0,95$

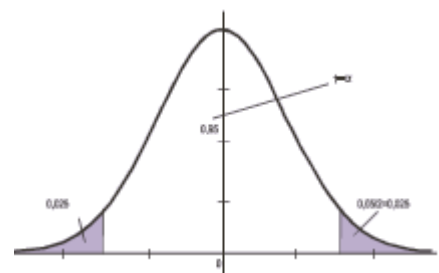
**Paso 3.** Determinamos a rexión de aceptación:

$$\left(p_0 - z_{\alpha/2} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}, p_0 + z_{\alpha/2} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}\right)$$

Como  $\alpha = 5\% = 0,05$ ,  $\alpha/2 = 0,025$ , segundo o gráfico vemos que

$$P[|Z| \leq z_{\alpha}] = 1-0,05 = 0,95$$

$$P[Z < z_{\alpha}] = 0,95 + 0,025 = 0,975$$



Buscando nas táboas o valor de  $z_{\alpha/2}$  obtemos que  $z_{\alpha/2} = 1,96$ ; en consecuencia, como  $n = 300$  e  $p_0 = 0,45$ , a rexión de aceptación será:

$$\left( 0,45 - 1,96 \cdot \sqrt{\frac{0,45 \cdot 0,55}{300}}, 0,45 + 1,96 \cdot \sqrt{\frac{0,45 \cdot 0,55}{300}} \right) = (0,394, 0,506).$$

**Paso 4.** Da mostra  $\hat{p} = 120/300 = 0,4$ , e ademais cae dentro do intervalo  $(0,394, 0,506)$ ; polo tanto, aceptamos a afirmación do comerciante.

10. A experiencia de anteriores eleccións mostran que certo partido obtén o 15% dos votos nunha cidade. Pódese aceptar esta afirmación, para un nivel de significación de 0,1, se na última enquisa só 205 persoas, entre 1500, se mostraron favorables ao devandito partido?

**Solución:**

**Paso 1.** As hipóteses do contraste son:

$H_0 : p = 0,15$  e  $H_1 : p < 0,15$  (contraste unilateral pola esquerda)

**Paso 2.** O nivel de significación é  $\alpha = 0,1$  e  $1-\alpha = 0,9$

**Paso 3.** Determinamos a rexión de aceptación:

$$\left( p_0 - z_{\alpha} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}, \infty \right) \text{ de } P[Z > -z_{\alpha}] = P[Z < z_{\alpha}] = 0,9.$$

Obtemos nas táboas  $z_{\alpha} = 1,28$  e como  $p_0 = 0,15$  e  $n = 1500$ , a rexión de aceptación é

$$\left( 0,15 - 1,28 \cdot \sqrt{\frac{0,15 \cdot 0,85}{1500}}, \infty \right) = (0,138, \infty)$$

**Paso 4.** Da mostra  $\hat{p} = 205/1500 = 0,136$ , e como vemos non cae dentro do intervalo  $(0,138, \infty)$ . Polo tanto, rexeitamos a hipótese nula.

11. Un fabricante de quentadores de gas afirma que como máximo o 2% dos quentadores de gas que comercializa teñen unha avaría durante o período de garantía. Non obstante, nunha mostra de 400 quentadores atopouse que 12 deles tiveron unha avaría durante o primeiro ano de garantía. Pódese aceptar a afirmación do fabricante cun nivel de significación do 0,05?

**Solución:**

**Paso 1.** As hipóteses do contraste son:

$H_0 : p = 0,02$  e  $H_1 : p > 0,02$  (contraste unilateral pola dereita).

**Paso 2.** O nivel de significación é  $\alpha = 0,05$  e  $1-\alpha = 0,95$

**Paso 3.** Determinamos a rexión de aceptación:

De  $P[Z < z_{\alpha}] = 0,95$ , obtemos nas táboas  $z_{\alpha} = 1,645$  e como  $p_0 = 0,02$  e  $n = 400$ , a rexión de aceptación é

$$\left( -\infty, 0,02 + 1,645 \cdot \sqrt{\frac{0,02 \cdot 0,98}{400}} \right) = (-\infty, 0,031)$$

**Paso 4.** Na mostra  $\hat{p} = 12/400 = 0,03$ , e como vemos cae dentro do intervalo  $(-\infty, 0,031)$ . Polo tanto, aceptamos a hipótese nula.