Esenciais USC

UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

INFERENCIA ESTATÍSTICA PARAMÉTRICA I

intervalos de confianza e contrastes de hipóteses para unha poboación

María Isabel Borrajo García, Mercedes Conde Amboage e Rosa María Crujeiras Casais (Departamento de Estatística, Análise Matemática e Optimización)

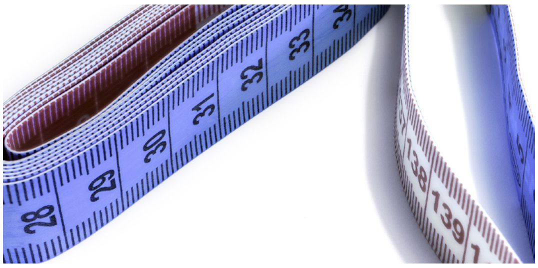
A inferencia estatística é, xunto coa estatística descritiva e cos fundamentos da teoría da probabilidade, un dos piares básicos da metodoloxía estatística clásica. O obxectivo final da inferencia estatística é o de poder tirar conclusións sobre unha poboación de interese a partir da observación dunha parte dela. Dependendo do tipo de conclusións extraídas poderemos clasificar os labores de inferencia en dúas grandes categorías:

- **Estimación:** consiste en aproximar un valor dun certo parámetro ou característica de interese, e pode facerse proporcionando un único valor (estimación puntual) ou proporcionando un rango que conteña o valor de interese cunha certa probabilidade (estimación por intervalos de confianza).
- **Contrastes de hipóteses:** consiste en establecer un procedemento de decisión para unha certa condición imposta (hipótese) sobre a característica de interese.

Para poder desenvolver correctamente as técnicas de inferencia estatística, precisamos coñecer algúns conceptos fundamentais:

- **Poboación**: universo de individuos ao que se refire o estudo que se pretende realizar.
- **Mostra**: subconxunto representativo da poboación. Así se X é a variable de interese, denotaremos por $\{X_1, \ldots, X_n\}$ unha mostra, onde cada variable X_i tomará un valor x_i que se corresponderá coa observación da variable X na unidade i-ésima da mostra.
- **Parámetro**: característica da poboación que caracteriza o comportamento da variable de interese.
- **Estatístico:** calquera función da mostra.
- **Estimador:** estatístico independente dos parámetros poboacionais que se emprega para estimalos, é dicir, todos os estimadores son estatísticos, pero non todos os estatísticos son estimadores (só o son aqueles que teñen por obxectivo estimar un parámetro).

Exemplo 1. Para ilustrar estes conceptos empregaremos un exemplo sinxelo: interésanos saber a altura media do profesorado da Facultade de Matemáticas. Nese caso teriamos como **poboación** todo o profesorado, como **mostra** unha selección deste colectivo a quen lle a súa estatura, que será a **variable** de interese, e como **parámetro** teriamos a media da estatura de todo o profesorado.



Estimación por intervalos de confianza

Consideremos unha mostra aleatoria simple $\{X_1, \ldots, X_n\}$ de X con distribución F_θ descoñecida pero con F cunha certa forma paramétrica (Normal, Binomial, Poisson, ...). A idea é construír un intervalo que conteña o parámetro θ cunha certa probabilidade.

Intervalo de confianza: un **intervalo de confianza** para o parámetro θ cun nivel de confianza de $(1-\alpha)$, con $\alpha \in (0,1)$ é un intervalo aleatorio $(L(X_1, \ldots, X_n), U(X_1, \ldots, X_n))$ tal que $\mathbb{P}(L(X_1, \ldots, X_n) < \theta < U(X_1, \ldots, X_n)) = 1-\alpha$. É dicir, o intervalo contén o verdadeiro valor de θ cunha certa probabilidade. Nótese que, en efecto, os extremos son aleatorios posto que dependen da mostra. No caso de que o que se cumpra sexa $\mathbb{P}(L(X_1, \ldots, X_n) < \theta < U(X_1, \ldots, X_n)) \approx 1-\alpha$, dirase que se trata dun intervalo de confianza aproximado cun nivel de confianza de $(1-\alpha)$, con $\alpha \in (0,1)$.

Intervalo de confianza para unha proporción

Consideremos unha mostra aleatoria simple $\{X_1, \ldots, X_n\}$ de $X \in Ber(p)$, interésanos construír un intervalo de confianza para p, a proporción teórica. Para iso necesitamos recordar que, como consecuencia do teorema central do límite, temos que

$$\frac{\widehat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1),$$

sendo $\widehat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i$ o estimador puntual de p e n o tamaño da mostra. Este tipo de estatísticos cuxa distribución non depende do parámetro de interese denomínanse **pivotes**. Nótese que o denominador do pivote, que se coñece como erro típico e denotaremos por

$$ET(\widehat{p}) = \sqrt{\frac{p(1-p)}{n}},$$

representa o erro da estimación proporcionada por $\widehat{p}$.

Partindo da distribución do pivote, podemos escribir:

$$\begin{aligned}
\mathbb{P}\left(-z_{1-\alpha/2} < \frac{\widehat{p}-p}{ET(\widehat{p})} < z_{1-\alpha/2}\right) \approx 1-\alpha &\iff \mathbb{P}\left(-z_{1-\alpha/2}\cdot ET(\widehat{p}) < \widehat{p}-p < z_{1-\alpha/2}\cdot ET(\widehat{p})\right) \approx 1-\alpha \\
&\iff \mathbb{P}\left(-z_{1-\alpha/2}\cdot ET(\widehat{p}) - \widehat{p} < -p < z_{1-\alpha/2}\cdot ET(\widehat{p}) - \widehat{p}\right) \approx 1-\alpha \\
&\iff \mathbb{P}\left(z_{1-\alpha/2}\cdot ET(\widehat{p}) + \widehat{p} > p > z_{1-\alpha/2}\cdot ET(\widehat{p}) - \widehat{p}\right) \approx 1-\alpha \\
&\iff \mathbb{P}\left(\widehat{p} - z_{1-\alpha/2}\cdot ET(\widehat{p}) < p < \widehat{p} + z_{1-\alpha/2}\cdot ET(\widehat{p})\right) \approx 1-\alpha
\end{aligned}$$

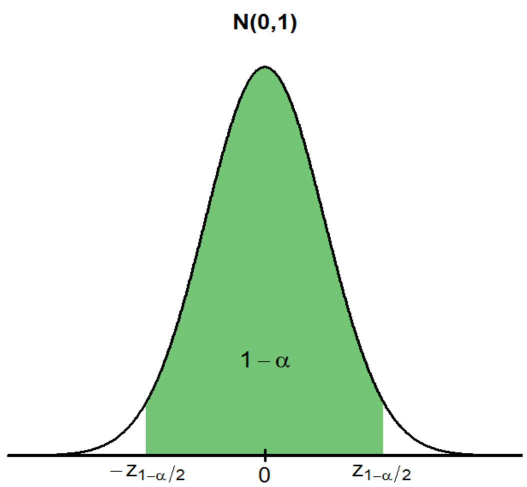
onde $z_{1-\alpha/2}$ denota o cuantil de orde $1-\alpha/2$ dunha distribución Normal estándar. Así teriamos que

$$(\widehat{p} - \text{ cuantil } \cdot ET(\widehat{p}),\, \widehat{p} + \text{ cuantil } \cdot ET(\widehat{p})) = \left(\widehat{p} - z_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}},\, \widehat{p} + z_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right)$$

é un intervalo de confianza para p a un nivel de confianza de $(1-\alpha)100\,\%$.

Como o termo da raíz non se pode obter a partir dunha mostra posto que depende do parámetro teórico, existen diferentes procedementos para solucionar isto, un dos máis habituais é substituír nese termo p por $\widehat{p}$ obtendo como intervalo final:

$$(\widehat{p} - \text{ cuantil } \cdot \widehat{ET(\widehat{p})},\, \widehat{p} + \text{ cuantil } \cdot \widehat{ET(\widehat{p})}) = \left(\widehat{p} - z_{1-\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}},\, \widehat{p} + z_{1-\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}\right).$$



Nesta imaxe da esquerda ilústrase a construción dun intervalo de confianza para o caso no que o estatístico segue unha distribución Normal (tal e como ocorre coa proporción).

Podemos ver como delimitamos unha zona central que abrangue unha probabilidade $(1-\alpha)$, e que deixa polo tanto en cada unha das colas unha probabilidade $\alpha/2$.

É importante destacar que se poderían construír intervalos de confianza de nivel $(1-\alpha)$ non centrados, pero os que se presentan aquí teñen menor lonxitude.

Intervalos de confianza para unha poboación Normal, $N(\mu, \sigma^2)$

Consideremos unha mostra aleatoria simple $\{X_1, \ldots, X_n\}$ de $X \in N(\mu, \sigma^2)$. Neste caso existen dous parámetros, a media e a varianza, e veremos a continuación como se obterían os intervalos de confianza en cada caso.

Intervalo de confianza para a media con varianza poboacional coñecida

Neste caso a media poboacional, μ, é o parámetro de interese e imos supoñer que a varianza poboacional é coñecida, é dicir, σ^2 toma un valor coñecido. Polo tanto, o pivote que precisamos neste escenario é:

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \in N(0,1),$$

sendo $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ o estimador puntual que se coñece como media da mostra. Nótese que neste caso o erro típico vén dado por $ET(\overline{X}) = \frac{\sigma}{\sqrt{n}}$.

Seguindo o mesmo procedemento que para o caso da proporción, chegamos á seguinte expresión do intervalo de confianza de nivel de confianza $(1-\alpha)100\,\%$ para a media con varianza poboacional coñecida:

$$(\overline{X} - \text{ cuantil } \cdot ET(\overline{X})\,,\, \overline{X} + \text{ cuantil } \cdot ET(\overline{X})) = \left(\overline{X} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\,,\, \overline{X} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right),$$

onde $z_{1-\alpha/2}$ denota o cuantil de orde $1-\alpha/2$ dunha distribución Normal estándar.

Intervalo de confianza para a media con varianza poboacional descoñecida

Neste caso a media poboacional, μ, segue a ser o parámetro de interese, pero coa diferenza de que agora a varianza poboacional é descoñecida. Polo tanto, o pivote non pode incluír o valor de σ, xa que non se coñece, e ten que ser substituído por un estimador, que alterará a distribución do pivote:

$$\frac{\overline{X} - \mu}{S/\sqrt{n-1}} \in T_{n-1} \quad \text{ou} \quad \frac{\overline{X} - \mu}{S_c/\sqrt{n}} \in T_{n-1}\,,$$

onde $S^2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$ e $S_c^2 = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$ son estimadores puntuais da varianza, varianza mostral e cuasivarianza, respectivamente; e T_{n-1} denota a distribución T de Student con $n-1$ graos de liberdade.

De maneira análoga ao caso anterior, tendo en conta que neste caso o erro típico vén dado por $S/\sqrt{n-1} = S_c/\sqrt{n}$, chegamos ao intervalo de confianza de nivel de confianza $(1-\alpha)100\,\%$ para a media con varianza poboacional descoñecida:

$$\left(\overline{X} - t_{n-1,1-\alpha/2}\frac{S}{\sqrt{n-1}}\,,\, \overline{X} + t_{n-1,1-\alpha/2}\frac{S}{\sqrt{n-1}}\right) = \left(\overline{X} - t_{n-1,1-\alpha/2}\frac{S_c}{\sqrt{n}}\,,\, \overline{X} + t_{n-1,1-\alpha/2}\frac{S_c}{\sqrt{n}}\right),$$

onde $t_{n-1,1-\alpha/2}$ denota o cuantil de orde $1-\alpha/2$ dunha distribución T de Student con $n-1$ graos de liberdade.

Intervalo de confianza para a varianza

Neste caso, o parámetro de interese é a varianza σ^2, e imos asumir que a media poboacional non é coñecida (situación máis habitual). Entón o pivote que se precisa é:

$$\frac{nS^2}{\sigma^2} \in \chi^2_{n-1} \quad \text{ou} \quad \frac{(n-1)S_c^2}{\sigma^2} \in \chi^2_{n-1},$$

onde χ^2_{n-1} denota a distribución de khi cadrado con $n-1$ graos de liberdade. Neste caso o procedemento que se debe seguir é análogo aos anteriores cun matiz, que a distribución de khi cadrado non é simétrica, polo que o intervalo de confianza resultante tampouco o será:

$$\left(\frac{nS^2}{\chi^2_{n-1,1-\alpha/2}}\,,\, \frac{nS^2}{\chi^2_{n-1,\alpha/2}}\right) = \left(\frac{(n-1)S_c^2}{\chi^2_{n-1,1-\alpha/2}}\,,\, \frac{(n-1)S_c^2}{\chi^2_{n-1,\alpha/2}}\right),$$

onde $\chi^2_{n-1,1-\alpha}$ denota o cuantil de orde $1-\alpha/2$ e $\chi^2_{n-1,\alpha}$ o cuantil de orde $\alpha/2$ nunha distribución de khi cadrado con $n-1$ graos de liberdade.

Intervalos de confianza para unha poboación Normal, $N(\mu, \sigma^2)$ *(continuación)*

Nota. No caso de querer obter un intervalo de confianza de nivel $(1-\alpha)$ para a desviación típica, σ, non existe un pivote nin un método específico, senón que o procedemento consiste en obter un intervalo de confianza dese mesmo nivel para a varianza, σ^2, e aplicarlle a raíz cadrada a cada un dos seus extremos, é dicir, o intervalo sería

$$\left(\sqrt{\frac{nS^2}{\chi^2_{n-1,1-\alpha/2}}}, \sqrt{\frac{nS^2}{\chi^2_{n-1,\alpha/2}}}\right) = \left(\sqrt{\frac{(n-1)S_c^2}{\chi^2_{n-1,1-\alpha/2}}}, \sqrt{\frac{(n-1)S_c^2}{\chi^2_{n-1,\alpha/2}}}\right).$$

Contrastes de hipóteses

Un contraste de hipóteses é un procedemento estatístico mediante o que se investiga a veracidade ou falsidade dunha certa hipótese (conxectura sobre o parámetro de interese). A idea é partir dunha conxectura e baseándonos nas observacións da mostra, ser quen de determinar se existen evidencias para rexeitar esa hipótese. Para formalizar este procedemento precisamos definir unha serie de conceptos:

- **Hipótese nula:** é aquela conxectura que se dá por certa antes de obter a realización mostral, é dicir, é a hipótese que goza de "presunción de inocencia", posto que se asume válida salvo que a mostra proporcione evidencias do contrario. Denótase habitualmente por H_0.
- **Hipótese alternativa:** é, habitualmente, o complementario da hipótese nula, é dicir, aquilo que sucede cando a hipótese nula non é certa. Denótase habitualmente por H_1 ou H_a.
- **Erro de tipo I:** é o erro que se comete cando rexeitamos a hipótese nula sendo certa. É o principal erro que desexamos controlar.
- **Erro de tipo II:** é o erro que se comete cando non rexeitamos a hipótese nula sendo falsa.
- **Nivel de significación:** é a probabilidade do erro de tipo I, é dicir, $\mathbb{P}(\text{rexeitar } H_0 \mid H_0 \text{ é certa})$. Denótase habitualmente por α.

	Decisión	
	Rexeitar H_0	Aceptar H_0
H_0 verdadeira	✗ Erro de tipo I	✓
H_0 falsa	✓	✗ Erro de tipo II

- **Estatístico de contraste:** é unha función da mostra (por ser un estatístico) cunha distribución coñecida baixo H_0 e que serve para medir a discrepancia entre H_0 e a información que proporciona a mostra sobre esta hipótese.
- **Rexión de rexeitamento:** é o conxunto de valores do estatístico de contraste para os que se rexeita a hipótese nula.
- **Valor p ou nivel crítico:** é o menor nivel de significación para o que podemos rexeitar H_0. Polo tanto, permitiranos determinar se podemos rexeitar ou non H_0 xa que:
  - se o valor p é menor que o nivel de significación, valor p < α, entón rexeitamos H_0,
  - se o valor p é maior ou igual que o nivel de significación, valor p > α, entón non temos evidencias para rexeitar H_0.

Para realizar calquera contraste de hipóteses debemos de seguir sempre estes pasos:

1. Determinar a hipótese nula, H_0, e a hipótese alternativa, H_a.
2. Fixar un nivel de significación, sendo $\alpha = 0.1$, $\alpha = 0.05$ ou $\alpha = 0.01$ os máis habituais.
3. Determinar un estatístico de contraste axeitado e establecer a súa distribución baixo H_0.
4. Construír a rexión de rexeitamento.
5. Avaliar o estatístico na mostra e determinar se pertence ou non á rexión de rexeitamento. Alternativamente tamén poderiamos calcular o p-valor ou nivel crítico.
6. Redactar a conclusión do contraste.

A continuación imos considerar e detallar tanto os contrastes para unha proporción (nunha poboación Bernoulli), como os contrastes para a media, μ, e a varianza, σ^2, nunha poboación Normal.

Contrastes para unha proporción

Consideremos de novo unha mostra aleatoria simple $\{X_1, \ldots, X_n\}$ de $X \in Ber(p)$, o noso obxectivo neste caso é valorar unha certa conxectura sobre o parámetro de interese, a proporción p.

En función do tipo de hipótese que se queira establecer existen tres posibles contrastes:

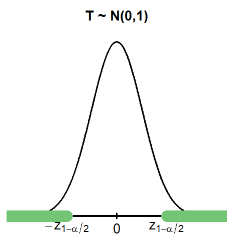
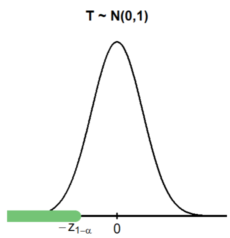
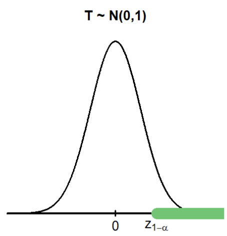
Contraste bilateral	Contraste unilateral á esquerda	Contraste unilateral á dereita
$\left\lVert \begin{array}{l} H_0 : p = p_0 \\ H_a : p \neq p_0 \end{array} \right.$	$\left\lVert \begin{array}{l} H_0 : p \geq p_0 \\ H_a : p < p_0 \end{array} \right.$	$\left\lVert \begin{array}{l} H_0 : p \leq p_0 \\ H_a : p > p_0 \end{array} \right.$

Estatístico de contraste e rexión de rexeitamento

Retomando o pivote que empregamos para o intervalo de confianza para unha proporción, e considerando neste caso a súa distribución baixo a hipótese nula teriamos que:

$$T = \frac{\widehat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim N(0,1),$$

e deste xeito xa poderemos establecer a rexión de rexeitamento en cada contraste:

<table>
<tr><td>$\left(-\infty\,,\,-z_{1-\alpha/2}\right) \cup \left(z_{1-\alpha/2}\,,\,+\infty\right)$</td><td>$(-\infty\,,\,-z_{1-\alpha})$</td><td>$(z_{1-\alpha}\,,\,+\infty)$</td></tr>
<tr><td>
</td><td>
</td><td>
</td></tr>
</table>

Exemplo 2. O Colexio de Odontólogos quere realizar un estudo que lles proporcione información sobre a necesidade ou non de realizar unha campaña de atención bucodental entre as/os adolescentes do Concello de Santiago de Compostela. Para iso acoden a varios centros de educación secundaria para realizar unha enquisa na que se pregunta a 500 mozas/os se acoden habitualmente á/ao dentista, obtendo 440 respostas positivas e 60 negativas.

O Colexio de Odontólogos decidirá realizar a campaña se a proporción de mozas/os que visitan habitualmente á/ao dentista é inferior ao 90 %. Á vista dos datos, hai probas significativas deste feito a un nivel de significación do 1 %? Que decidirá o Colexio de Odontólogos?

1. $\left\lVert \begin{array}{l} H_0 : p \geq 0.9 \\ H_a : p < 0.9 \end{array} \right.$

2. α=0.01

3. $T = \dfrac{\widehat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim N(0,1)$

4. $\text{RR} = (-\infty\,,\,-z_{0.99}) = (-\infty\,,\,-2.33)$

5. $t = \dfrac{\widehat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \dfrac{0.88 - 0.9}{\sqrt{\frac{0.9(1-0.9)}{500}}} = -1.49$ que non pertence á RR.

6. Non se rexeita H_0, é dicir, non existen, segundo estes datos, evidencias estatisticamente significativas de que a porcentaxe de mozas/os que acoden á/ao dentista sexa inferior ao 90 %.

Contrastes para unha poboación Normal, $N(\mu,\sigma^2)$

Consideremos unha mostra aleatoria simple $\{X_1,\ldots,X_n\}$ de $X\in N(\mu,\sigma^2)$, neste caso existen dous parámetros, a media e a varianza, e veremos a continuación como se poderían realizar contrastes de hipóteses en cada caso.

Contraste para a media con varianza poboacional coñecida

Se a media poboacional, μ, é o parámetro de interese, e estamos no contexto de varianza poboacional coñecida, podemos considerar os seguintes contrastes:

Contraste bilateral	Contraste unilateral á esquerda	Contraste unilateral á dereita
$H_0: \mu = \mu_0$ $H_a: \mu \neq \mu_0$	$H_0: \mu \geq \mu_0$ $H_a: \mu < \mu_0$	$H_0: \mu \leq \mu_0$ $H_a: \mu > \mu_0$

Estatístico de contraste e rexión de rexeitamento

Dado que a varianza poboacional é coñecida, retomando o pivote que empregabamos para os intervalos de confianza neste contexto, teriamos o seguinte estatístico de contraste baixo a hipótese nula:

$$\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} \in N(0,1),$$

o que nos permite determinar a rexión de rexeitamento en cada caso:

$\left(-\infty\,,\,-z_{1-\alpha/2}\right)\cup\left(z_{1-\alpha/2}\,,\,+\infty\right)$	$\left(-\infty\,,\,-z_{1-\alpha}\right)$	$\left(z_{1-\alpha}\,,\,+\infty\right)$

Contraste para a media con varianza poboacional descoñecida

Neste caso a media poboacional, μ, segue a ser o parámetro de interese, pero coa diferenza de que agora a varianza poboacional é descoñecida. Os contrastes que podemos considerar seguen a ser:

Contraste bilateral	Contraste unilateral á esquerda	Contraste unilateral á dereita
$H_0: \mu = \mu_0$ $H_a: \mu \neq \mu_0$	$H_0: \mu \geq \mu_0$ $H_a: \mu < \mu_0$	$H_0: \mu \leq \mu_0$ $H_a: \mu > \mu_0$

Estatístico de contraste e rexión de rexeitamento

Agora, como a varianza poboacional é descoñecida, debemos considerar como estatístico de contraste:

$$\frac{\overline{X}-\mu_0}{S/\sqrt{n-1}} \in T_{n-1} \quad \text{ou} \quad \frac{\overline{X}-\mu_0}{S_c/\sqrt{n}} \in T_{n-1}.$$

cuxa distribución baixo a hipótese nula nos permite determinar as rexións de rexeitamento asociadas a cada contraste:

$\left(-\infty\,,\,-t_{n-1,1-\alpha/2}\right)\cup\left(t_{n-1,1-\alpha/2}\,,\,+\infty\right)$	$\left(-\infty\,,\,-t_{n-1,1-\alpha}\right)$	$\left(t_{n-1,1-\alpha}\,,\,+\infty\right)$

Contraste para a varianza

Finalmente, se o parámetro de interese é a varianza poboacional, σ^2, poderiamos considerar os seguintes contrastes:

Contraste bilateral	Contraste unilateral á esquerda	Contraste unilateral á dereita
$H_0: \sigma^2 = \sigma_0^2$ $H_a: \sigma^2 \neq \sigma_0^2$	$H_0: \sigma^2 \geq \sigma_0^2$ $H_a: \sigma^2 < \sigma_0^2$	$H_0: \sigma^2 \leq \sigma_0^2$ $H_a: \sigma^2 > \sigma_0^2$

Estatístico de contraste e rexión de rexeitamento

Supoñendo que a media poboacional non é coñecida, de maneira análoga ao que vimos cos intervalos de confianza, entón o estatístico de contraste vén dado por:

$$\frac{nS^2}{\sigma_0^2} \in \chi^2_{n-1} \quad \text{ou} \quad \frac{(n-1)S_c^2}{\sigma_0^2} \in \chi^2_{n-1},$$

o que nos permite determinar as rexións de rexeitamento en cada caso:

$\left(0\,,\,\chi^2_{n-1,\alpha/2}\right)\cup\left(\chi^2_{n-1,1-\alpha/2}\,,\,+\infty\right)$	$\left(0\,,\,\chi^2_{n-1,\alpha}\right)$	$\left(\chi^2_{n-1,1-\alpha}\,,\,+\infty\right)$

Contrastes para unha poboación Normal, $N(\mu, \sigma^2)$ *(continuación)*

Exemplo 3. Deséxase analizar se un determinado tipo de auga mineral é apta para pacientes hipertensos. Para isto, analízase o contido de sodio (en mg/l) en 10 botellas de distintos lotes deste agua:

49.6 44.7 48.7 48.4 41.6 48.0 41.3 47.6 49.8 49.5

Á vista dos resultados, e asumindo que o contido en sodio segue unha distribución Normal,

a) A auga mineral considérase apta para persoas con hipertensión se, en media, o contido de sodio é inferior a 50 mg/l. Tendo en conta a mostra anterior, existen evidencias para que un profesional da medicina recomende o consumo desta auga ás persoas hipertensas cun nivel de significación do 1 %?

b) Nas tarefas de control de calidade da empresa de envasado, non só terán en conta o contido medio de sodio, senón tamén a variabilidade deste. Considérase que unha desviación típica superior a 3 mg/l é inaceptable para manter o selo de calidade da empresa. En base aos datos recollidos, poderá a empresa manter os seus estándares de calidade? (Realiza este contraste a un nivel de significación do 5 %).

O primeiro que faremos será definir a variable de interese: $X =$ "Contido en sodio dunha botella de auga", que sabemos que segue unha distribución Normal con media e varianza poboacionais descoñecidas, isto é, $X \in N(\mu, \sigma^2)$. A continuación resolveremos o apartado a) seguindo o mesmo esquema de pasos que o levado a cabo no Exemplo 2.

1. $\left|\left| \begin{array}{l} H_0 : \mu \geq 50 \\ H_a : \mu < 50 \end{array}\right.\right.$
2. α=0.01
3. $T = \frac{\overline{X}-\mu_0}{S_c/\sqrt{n}} \in T_{n-1}$
4. $\mathrm{RR} = (-\infty\,,\, -t_{9,0.99}) = (-\infty\,,\, -2.82)$
5. $t = \frac{\overline{x}-\mu_0}{s_c/\sqrt{n}} = \frac{46.92-50}{3.23/\sqrt{10}} = -3.01$ que si pertence á RR.
6. Rexéitase H_0, é dicir, existen evidencias estatisticamente significativas de que o contido medio de sodio é inferior a 50 mg/l, polo queo profesional da medicina pode recomendar o consumo desta auga ás persoas hipertensas.

Agora resolveremos o problema que se presenta no apartado b), que se basea na mesma variable aleatoria xa definida ao comezo do exemplo.

1. $\left|\left| \begin{array}{l} H_0 : \sigma \leq 3 \\ H_a : \sigma > 3 \end{array}\right.\right. \iff \left|\left| \begin{array}{l} H_0 : \sigma^2 \leq 9 \\ H_a : \sigma^2 > 9 \end{array}\right.\right.$
2. $\alpha = 0.05$
3. $T = \frac{(n-1)S_c^2}{\sigma_0^2} \in \chi^2_{n-1}$
4. $\mathrm{RR} = \left(\chi^2_{9,0.95}\,,\, +\infty\right) = (16.92\,,\, +\infty)$
5. $t = \frac{(n-1)s_c^2}{\sigma_0^2} = \frac{9\cdot 10.44}{9} = 10.44$ que si pertence á RR.
6. Non se rexeita H_0, é dicir, non existen evidencias estatisticamente significativas de que a desviación típica do contido de sodio sexa superior a 3 mg/l, polo que asumimos que a empresa pode manter o seu selo de calidade.