INFERENCIA ESTATÍSTICA PARAMÉTRICA II

comparación de poboacións

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 poder extraer conclusións sobre unha ou máis poboacións de interese a partir da observación dunha ou varias mostras. Neste Esencial trataremos a problemática da comparación de dúas poboacións, abranguendo tanto a construción de intervalos de confianza como a realización de contrastes de hipóteses.

A comparación de dúas poboacións pode facerse a través dos parámetros que as caracterizan. Para abordar esta cuestión debemos, en primeiro lugar, realizar a tarefa de observación das mostras. Atendendo á posible relación entre as dúas mostras, podemos diferenciar entre:

- **mostras independentes:** as observacións dunha das mostras non condicionan en ningún caso as observacións da outra. Isto acontece habitualmente cando se mide a mesma variable en dous grupos diferentes de individuos.
- **mostras emparelladas:** existe unha relación ou influencia entre as observacións dunha mostra e as observacións da outra. Isto acontece habitualmente cando se mide unha variable nun único grupo de individuos en momentos ou situacións diferentes.

Exemplo 1. Se quixésemos comparar o nivel medio de colesterol nun grupo de homes e noutro de mulleres, estariamos ante un exemplo de mostras independentes. Pola contra, se o noso obxectivo fose comparar o nivel medio de colesterol dun grupo de persoas antes e despois de someterse a un certo tratamento, estariamos ante un exemplo de mostras emparelladas.

Neste contexto de comparación de poboacións, de xeito análogo a como procedemos cunha única poboación (ver Esencial *Inferencia estatística paramétrica I*), consideraremos tanto o caso de ter poboacións Bernoulli e estar interesados na comparación de proporcións, como ter poboacións Normais e estar interesados tanto na comparación de medias como na comparación de varianzas.

Comparación de proporcións

Consideremos dúas mostras aleatorias simples $\{X_1, \ldots, X_{n_X}\}$ de $X \in Ber(p_X)$ e $\{Y_1, \ldots, Y_{n_Y}\}$ de $Y \in Ber(p_Y)$. O que nos interesa é comparar p_X e p_Y, polo que estudaremos a súa diferenza, $p_X - p_Y$.

Intervalo de confianza para a diferenza de proporcións

Interésanos construír un intervalo de confianza para $p_X - p_Y$, a diferenza de proporcións teóricas. Para iso necesitamos recordar que, como consecuencia da aplicación do teorema central do límite, temos que

$$\frac{(\widehat{p_X} - \widehat{p_Y}) - (p_X - p_Y)}{\sqrt{\widehat{p}\,(1-\widehat{p})\left(\frac{1}{n_X} + \frac{1}{n_Y}\right)}} \sim N(0,1),$$

onde $\widehat{p} = \dfrac{n_X \widehat{p_X} + n_Y \widehat{p_Y}}{n_X + n_Y}$ é un estimador puntual da proporción poboacional común se $p = p_X = p_Y$.

Este tipo de estatísticos, nos que a distribución non depende do parámetro de interese, denomínanse pivotes. E, partindo desa distribución, ao igual que se fixo no Esencial *Inferencia estatística paramétrica I* para unha poboación, chegamos á seguinte expresión para o intervalo de confianza aproximado de nivel de confianza $(1-\alpha)\,100\,\%$ para a diferenza de proporcións:

$$\left((\widehat{p_X} - \widehat{p_Y}) - z_{1-\alpha/2}\sqrt{\widehat{p}\,(1-\widehat{p})\left(\frac{1}{n_X} + \frac{1}{n_Y}\right)},\ (\widehat{p_X} - \widehat{p_Y}) + z_{1-\alpha/2}\sqrt{\widehat{p}\,(1-\widehat{p})\left(\frac{1}{n_X} + \frac{1}{n_Y}\right)} \right),$$

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

Contraste de hipóteses para a diferenza de proporcións

Agora o que queremos é contrastar algunha característica sobre $p_X - p_Y$. De novo, tal e como se detalla no Esencial *Inferencia estatística paramétrica I*, poderemos establecer tres tipos de contrastes:

Contraste bilateral	Contraste unilateral á esquerda	Contraste unilateral á dereita
$\left\| \begin{array}{l} H_0 : p_X - p_Y = 0 \\ H_a : p_X - p_Y \neq 0 \end{array} \right.$	$\left\| \begin{array}{l} H_0 : p_X - p_Y \geq 0 \\ H_a : p_X - p_Y < 0 \end{array} \right.$	$\left\| \begin{array}{l} H_0 : p_X - p_Y \leq 0 \\ H_a : p_X - p_Y > 0 \end{array} \right.$

Estatístico de contraste e rexión de rexeitamento

O pivote hérdase dos intervalos considerando, neste caso, a súa distribución baixo a hipótese nula que viría dada por:

$$\frac{(\widehat{p_X} - \widehat{p_Y}) - (p_{X,0} - p_{Y,0})}{\sqrt{\widehat{p}\,(1-\widehat{p})\left(\frac{1}{n_X} + \frac{1}{n_Y}\right)}} = \frac{\widehat{p_X} - \widehat{p_Y}}{\sqrt{\widehat{p}\,(1-\widehat{p})\left(\frac{1}{n_X} + \frac{1}{n_Y}\right)}} \sim N(0,1),$$

onde recordemos que $\widehat{p} = \dfrac{n_X\widehat{p_X} + n_Y\widehat{p_Y}}{n_X + n_Y}$ e $p_{X,0}$ e $p_{Y,0}$ son os valores das proporcións teóricas que se queren contrastar. Deste xeito xa poderemos establecer a rexión de rexeitamento en cada contraste:

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

Nota. Expoñemos o caso máis habitual no que o segundo termo das ecuacións que aparecen nas hipóteses é sempre cero (é dicir, $p_{X,0} - p_{Y,0} = 0$), realmente poderiamos propoñer un contraste (de calquera das tres tipoloxías) no que se valorase a diferenza das proporcións teóricas contra un certo valor $p_{X,0} - p_{Y,0}$, non necesariamente nulo.

Comparación de dúas medias en mostras emparelladas

Consideremos unha mostra aleatoria simple $\{X_1, \ldots, X_n\}$ de $X \in N(\mu_X, \sigma_X^2)$, e outra mostra aleatoria simple $\{Y_1, \ldots, Y_n\}$ de $Y \in N(\mu_Y, \sigma_Y^2)$. Asumiremos ademais que se trata de mostras emparelladas, polo tanto, hai unha certa relación de dependencia entre ambas as dúas. A idea para comparar a diferenza de medias $\mu_X - \mu_Y$ consiste en considerar unha nova mostra coas diferenzas $D_1 = X_1 - Y_1, D_2 = X_2 - Y_2, \ldots, D_n = X_n - Y_n$ que tamén seguen unha distribución Normal de media $\mu_X - \mu_Y$. Deste xeito, o problema orixinal redúcese a realizar intervalos de confianza e contrastes de hipóteses sobre a media dunha distribución Normal, presentados no Esencial *Inferencia estatística paramétrica I*.

Intervalo de confianza para a diferenza de medias en mostras emparelladas

O noso obxectivo será construír un intervalo de confianza para $\mu_X - \mu_Y$, é dicir, a diferenza das medias poboacionais. Entón, tendo en conta que $D = X - Y$ denota unha variable Normal con media $\mu_D = \mu_X - \mu_Y$ e varianza descoñecida, empregaremos o pivote

$$\frac{\overline{D} - \mu_D}{S_D/\sqrt{n-1}} = \frac{\overline{D} - \mu_D}{S_{c,D}/\sqrt{n}} \in T_{n-1}$$

onde $\overline{D} = \frac{1}{n}\sum_{i=1}^{n} D_i$ denota o estimador puntual de $\mu_D = \mu_X - \mu_Y$. Ademais, $S_D^2 = \frac{1}{n}\sum_{i=1}^{n}\left(D_i - \overline{D}\right)^2$ e $S_{c,D}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(D_i - \overline{D}\right)^2$ son estimadores puntuais da varianza, varianza mostral e cuasivarianza, respectivamente. Finalmente, T_{n-1} denota a distribución T de Student con $n-1$ graos de liberdade, e partindo desa distribución, chegamos á seguinte expresión para o intervalo de confianza de nivel de confianza $(1-\alpha)\,100\,\%$ para a diferenza de medias en mostras emparelladas:

$$\left(\overline{D} - t_{n-1,1-\alpha/2}\,\frac{S_D}{\sqrt{n-1}}\,,\,\overline{D} + t_{n-1,1-\alpha/2}\,\frac{S_D}{\sqrt{n-1}}\right) = \left(\overline{D} - t_{n-1,1-\alpha/2}\,\frac{S_{c,D}}{\sqrt{n}}\,,\,\overline{D} + t_{n-1,1-\alpha/2}\,\frac{S_{c,D}}{\sqrt{n}}\right),$$

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.

Contraste de hipóteses para a diferenza de medias en mostras emparelladas

Agora o que queremos é contrastar algunha característica sobre o parámetro $\mu_D = \mu_X - \mu_Y$. De novo, tal e como se detalla no Esencial *Inferencia estatística paramétrica I* para o caso dunha variable Normal con varianza descoñecida, poderemos establecer tres tipos de contrastes:

Contraste bilateral	Contraste unilateral á esquerda	Contraste unilateral á dereita
$\left\Vert \begin{array}{l} H_0 : \mu_D = \mu_X - \mu_Y = 0 \\ H_a : \mu_D = \mu_X - \mu_Y \neq 0 \end{array}\right.$	$\left\Vert \begin{array}{l} H_0 : \mu_D = \mu_X - \mu_Y \geq 0 \\ H_a : \mu_D = \mu_X - \mu_Y < 0 \end{array}\right.$	$\left\Vert \begin{array}{l} H_0 : \mu_D = \mu_X - \mu_Y \leq 0 \\ H_a : \mu_D = \mu_X - \mu_Y > 0 \end{array}\right.$

Estatístico de contrastes e rexión de rexeitamento

O pivote hérdase dos intervalos considerando, neste caso, a súa distribución baixo a hipótese nula que viría dada por:

$$\frac{\overline{D}}{S_D/\sqrt{n-1}} = \frac{\overline{D}}{S_{c,D}/\sqrt{n}} \in T_{n-1}$$

e grazas ao devandito estatístico xa poderemos establecer a rexión de rexeitamento en cada contraste:

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

Nota. Expoñemos o caso máis habitual no que o segundo termo das ecuacións que aparecen nas hipóteses é sempre cero, e polo tanto, queremos comprobar se as medias nas dúas poboacións son iguais (é dicir, $\mu_{X,0} - \mu_{Y,0} = 0$). Realmente, poderiamos propoñer un contraste (de calquera das tres tipoloxías) no que se valorase a diferenza das medias poboacionais contra un certo valor $\mu_{X,0} - \mu_{Y,0}$, non necesariamente nulo. Isto aplícase tamén a mostras independentes posto que os contrastes que se poden formular son os mesmos, pero cambian os estatísticos de contraste.

Comparación de dúas medias en mostras independentes

Consideremos unha mostra aleatoria simple $\{X_1, \ldots, X_{n_X}\}$ de $X \in N(\mu_X, \sigma_X^2)$, e unha segunda mostra aleatoria simple $\{Y_1, \ldots, Y_{n_Y}\}$ de $Y \in N(\mu_Y, \sigma_Y^2)$ independentes entre si. Neste caso existen dous parámetros de interese que son a media e a varianza. Centrarémonos, en primeiro lugar, na comparación de dúas medias, é dicir, construiremos intervalos de confianza e propoñeremos contrastes de hipóteses para $\mu_X - \mu_Y$.

Intervalos de confianza para a diferenza de medias

Varianzas poboacionais coñecidas

Nun primeiro caso asumiremos que as varianzas poboacionais son coñecidas, é dicir, σ_X^2 e σ_Y^2 toman valores coñecidos. Polo tanto, o pivote que precisamos neste escenario é:

$$\frac{\left(\overline{X} - \overline{Y}\right) - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}} \in N(0,1),$$

sendo $\overline{X} = \frac{1}{n_X}\sum_{i=1}^{n_X} X_i$ o estimador puntual de μ_X e $\overline{Y} = \frac{1}{n_Y}\sum_{i=1}^{n_Y} Y_i$ o estimador puntual de μ_Y.

Seguindo o mesmo procedemento que para o caso da proporción no Esencial *Inferencia estatística paramétrica I*, chegamos á seguinte expresión do intervalo de confianza de nivel de confianza $(1-\alpha)\,100\,\%$ para a diferenza de medias con varianzas poboacionais coñecidas:

$$\left(\left(\overline{X} - \overline{Y}\right) - z_{1-\alpha/2}\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}\,,\, \left(\overline{X} - \overline{Y}\right) + z_{1-\alpha/2}\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}} \right),$$

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

Intervalos de confianza para a diferenza de medias (continuación)

Varianzas poboacionais descoñecidas pero asumidas iguais

Neste segundo caso, asumiremos que descoñecemos as varianzas poboacionais, pero temos información para poder asumilas iguais, é dicir, $\sigma^2 = \sigma_X^2 = \sigma_Y^2$. Nestas circunstancias, en lugar de intentar estimar a varianza de cada poboación cos datos da mostra correspondente, farémolo conxuntamente, e propoñemos como estimador conxunto da varianza, σ^2, o seguinte:

$$S_p^2 = \frac{(n_X - 1)\, S_{c,X}^2 + (n_Y - 1)\, S_{c,Y}^2}{n_X + n_Y - 2},$$

onde $S_{c,X}^2 = \frac{1}{n_X - 1} \sum_{i=1}^{n_X} \left(X_i - \overline{X}\right)^2$ e $S_{c,Y}^2 = \frac{1}{n_Y - 1} \sum_{i=1}^{n_Y} \left(Y_i - \overline{Y}\right)^2$ son as cuasivarianzas de X e de Y, respectivamente. Así o pivote sería:

$$\frac{\left(\overline{X} - \overline{Y}\right) - (\mu_X - \mu_Y)}{S_p \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}} \in T_{n_X + n_Y - 2},$$

e o intervalo que se constrúe a partir del viría dado por:

$$\left(\left(\overline{X} - \overline{Y}\right) - t_{n_X + n_Y - 2, 1-\alpha/2}\; S_p \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}} \,,\, \left(\overline{X} - \overline{Y}\right) + t_{n_X + n_Y - 2, 1-\alpha/2}\; S_p \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}} \right),$$

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

Varianzas poboacionais descoñecidas e distintas

Finalmente, asumiremos que descoñecemos as varianzas poboacionais e ademais non temos información para poder asumir que son iguais, é dicir, traballaremos coa condición de que as varianzas son descoñecidas e distintas. Isto deriva na necesidade de estimar cada unha delas por separado e obter como pivote:

$$\frac{\left(\overline{X} - \overline{Y}\right) - (\mu_X - \mu_Y)}{\sqrt{\frac{S_{c,X}}{n_X} + \frac{S_{c,Y}}{n_Y}}} \sim T_\gamma$$

onde γ denota o enteiro máis próximo a $\frac{\left(S_{c,X}^2/n_X + S_{c,Y}^2/n_Y\right)^2}{\frac{\left(S_{c,X}^2/n_X\right)^2}{n_X - 1} + \frac{\left(S_{c,Y}^2/n_Y\right)^2}{n_Y - 1}}$. E así, obteriamos o intervalo de confianza aproximado:

$$\left(\left(\overline{X} - \overline{Y}\right) - t_{\gamma, 1-\alpha/2} \sqrt{\frac{S_{c,X}^2}{n_X} + \frac{S_{c,Y}^2}{n_Y}} \,,\, \left(\overline{X} - \overline{Y}\right) + t_{\gamma, 1-\alpha/2} \sqrt{\frac{S_{c,X}^2}{n_X} + \frac{S_{c,Y}^2}{n_Y}} \right)$$

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

Contraste de hipóteses para a diferenza de medias

Varianzas poboacionais coñecidas

Neste caso a diferenza de medias, $\mu_X - \mu_Y$, segue a ser o parámetro de interese sobre o que queremos contrastar unha certa característica. Entón podemos formular os seguintes contrastes:

Contraste bilateral	Contraste unilateral á esquerda	Contraste unilateral á dereita
$\left\| \begin{array}{l} H_0 : \mu_X - \mu_Y = 0 \\ H_a : \mu_X - \mu_Y \neq 0 \end{array} \right.$	$\left\| \begin{array}{l} H_0 : \mu_X - \mu_Y \geq 0 \\ H_a : \mu_X - \mu_Y < 0 \end{array} \right.$	$\left\| \begin{array}{l} H_0 : \mu_X - \mu_Y \leq 0 \\ H_a : \mu_X - \mu_Y > 0 \end{array} \right.$

Estatístico de contraste e rexión de rexeitamento

Dado que estamos no contexto de varianzas poboacionais coñecidas, retomando o pivote que empregabamos para os intervalos de confianza e considerando a súa distribución baixo a hipótese nula, temos:

$$\frac{\overline{X} - \overline{Y}}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}} \in N(0,1),$$

e o devandito pivote permítenos establecer as seguintes rexións de rexeitamento:

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

Contraste de hipotéses para a diferenza de medias (continuación)

Varianzas poboacionais descoñecidas pero asumidas iguais

No caso de que descoñezamos as varianzas poboacionais, pero teñamos información para poder asumilas iguais, é dicir $\sigma^2 = \sigma_X^2 = \sigma_Y^2$, tamén podemos establecer os seguintes contrastes:

Contraste bilateral	Contraste unilateral á esquerda	Contraste unilateral á dereita
$\left\| \begin{array}{l} H_0 : \mu_X - \mu_Y = 0 \\ H_a : \mu_X - \mu_Y \neq 0 \end{array} \right.$	$\left\| \begin{array}{l} H_0 : \mu_X - \mu_Y \geq 0 \\ H_a : \mu_X - \mu_Y < 0 \end{array} \right.$	$\left\| \begin{array}{l} H_0 : \mu_X - \mu_Y \leq 0 \\ H_a : \mu_X - \mu_Y > 0 \end{array} \right.$

Estatístico de contraste e rexión de rexeitamento

Dado que podemos asumir que as varianzas de ambas as poboacións son descoñecidas pero iguais, do mesmo xeito que se fixo para o intervalo de confianza, temos que a distribución do estatístico baixo a hipótese nula será:

$$\frac{\overline{X} - \overline{Y}}{S_p \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}} \in T_{n_X + n_Y - 2}$$

onde $S_p^2 = \dfrac{(n_X - 1) S_{cX}^2 + (n_Y - 1) S_{cY}^2}{n_X + n_Y - 2}$ é a estimación conxunta da varianza poboacional σ^2. Deste xeito, a rexión de rexeitamento asociada a cada contraste vén dada por:

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

Varianzas poboacionais descoñecidas e distintas

Por último, asumiremos que as varianzas poboacionais son descoñecidas e distintas. Neste caso tamén podemos realizar os seguintes contrastes:

Contraste bilateral	Contraste unilateral á esquerda	Contraste unilateral á dereita
$\left\| \begin{array}{l} H_0 : \mu_X - \mu_Y = 0 \\ H_a : \mu_X - \mu_Y \neq 0 \end{array} \right.$	$\left\| \begin{array}{l} H_0 : \mu_X - \mu_Y \geq 0 \\ H_a : \mu_X - \mu_Y < 0 \end{array} \right.$	$\left\| \begin{array}{l} H_0 : \mu_X - \mu_Y \leq 0 \\ H_a : \mu_X - \mu_Y > 0 \end{array} \right.$

Estatístico de contraste e rexión de rexeitamento

De maneira análoga ao presentado para os intervalos de confianza, neste caso o pivote que se obtén é:

$$\frac{\overline{X} - \overline{Y}}{\sqrt{\frac{S_{c,X}^2}{n_X} + \frac{S_{c,Y}^2}{n_Y}}} \sim T_\gamma$$

sendo γ o enteiro máis próximo a $\dfrac{\left(S_{c,X}^2/n_X + S_{c,Y}^2/n_Y\right)^2}{\frac{\left(S_{c,X}^2/n_X\right)^2}{n_X-1} + \frac{\left(S_{c,Y}^2/n_Y\right)^2}{n_Y-1}}$, e así podemos establecer as seguintes rexións de rexeitamento:

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

Exemplo 2. Para comprobar se a tolerancia á glicosa en varóns sans varía coa idade realizouse un test oral de glicosa a dúas mostras de doentes sans, unha de mozos e outra de adultos. O test consistiu en medir o nivel de glicosa en sangue aos 60 minutos da inxestión de 100 gramos de glicosa. Os resultados obtidos foron os seguintes:

	Tamaño da mostra	Media da mostra	Cuasidesviación típica
Mozos	$n_X = 41$	$\overline{x} = 146.53$ mg/dl	$s_{c,X} = 7.87$ mg/dl
Adultos	$n_Y = 50$	$\overline{y} = 184.90$ mg/dl	$s_{c,Y} = 10.72$ mg/dl

Supoñendo que o nivel de glicosa en ambos os grupos segue unha distribución Normal e fixado un nivel de significación $\alpha = 5\,\%$, existen evidencias estatisticamente significativas de que o nivel medio de glicosa difire en ambos os grupos?

1. $\left\| \begin{array}{l} H_0 : \mu_X - \mu_Y = 0 \\ H_a : \mu_X - \mu_Y \neq 0 \end{array} \right.$

2. $\alpha = 0.05$

3. $T = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\frac{S_{cX}^2}{n_X} + \frac{S_{cY}^2}{n_Y}}} \sim T_{88}$ dado que $\dfrac{\left(S_{c,X}^2/n_X + S_{c,Y}^2/n_Y\right)^2}{\frac{\left(S_{c,X}^2/n_X\right)^2}{n_X-1} + \frac{\left(S_{c,Y}^2/n_Y\right)^2}{n_Y-1}} = \dfrac{(7.87^2/41 + 10.72^2/50)^2}{\frac{(7.87^2/41)^2}{40} + \frac{(10.72^2/50)^2}{49}} = 88.01$

Nótese que estamos traballando con mostras independentes e varianzas descoñecidas. Ademais, neste caso, como os graos de liberdade da T de Student son maiores que 30 pode admitirse unha aproximación pola N(0,1).

4. $RR = \left(-\infty\, ,\, -t_{88,1-0.05/2}\right) \cup \left(t_{88,1-0.05/2}\, ,\, +\infty\right) = (-\infty\, ,\, -1.96) \cup (1.96\, ,\, +\infty)$

5. $t = \dfrac{\overline{x} - \overline{y}}{\sqrt{\frac{s_{c,X}^2}{n_X} + \frac{s_{c,Y}^2}{n_Y}}} = \dfrac{146.53 - 184.90}{\sqrt{\frac{7.87^2}{41} + \frac{10.72^2}{50}}} = -19.66$

que pertence á RR.

6. Rexéitase H_0, é dicir, existen evidencias estatisticamente significativas de que o nivel medio de glicosa difire entre varóns adultos e mozos.

Comparación de dúas varianzas

Consideremos unha mostra aleatoria simple $\{X_1, \ldots, X_{n_X}\}$ de $X \in N(\mu_X, \sigma_X^2)$, e unha segunda mostra aleatoria simple $\{Y_1, \ldots, Y_{n_Y}\}$ de $Y \in N(\mu_Y, \sigma_Y^2)$ independentes entre si. Como xa comentamos antes, neste caso existen dous parámetros, a media e a varianza. Agora centrarémonos na comparación de dúas varianzas, é dicir, construiremos intervalos de confianza e proporemos contrastes de hipóteses para σ_X^2/σ_Y^2.

Intervalos de confianza para o cociente de varianzas

Estamos interesados no cociente de varianzas, polo que intuitivamente intentaremos construír un pivote a partir do cociente de estimadores desas varianzas. Entón, a partir do teorema de Fisher obtemos que

$$T = \frac{S_{c,X}^2}{S_{c,Y}^2}\frac{\sigma_Y^2}{\sigma_X^2} \in F_{n_X-1,n_Y-1},$$

onde F_{n_X-1,n_Y-1} denota a distribución F de Snedecor con $n_X - 1$ e $n_Y - 1$ graos de liberdade. Seguindo o mesmo razoamento que para a construción de calquera outro intervalo, tendo en conta que neste caso non estamos a traballar cunha distribución simétrica como a Normal ou a T de Student, chegamos a:

$$\left(\frac{1}{f_{n_X-1,n_Y-1,1-\alpha/2}}\frac{S_{c,X}^2}{S_{c,Y}^2}\,,\ \frac{1}{f_{n_X-1,n_Y-1,\alpha/2}}\frac{S_{c,X}^2}{S_{c,Y}^2}\right),$$

onde $f_{n_X-1,n_Y-1,1-\alpha/2}$ e $f_{n_X-1,n_Y-1,\alpha/2}$ denotan o cuantil de orde $1-\alpha/2$ e $\alpha/2$ dunha F de Snedecor con $n_X - 1$ e $n_Y - 1$ graos de liberdade.

Contraste de hipóteses para o cociente de varianzas

Agora o que queremos é contrastar algunha característica sobre σ_X^2/σ_Y^2, polo que podemos considerar os seguintes contrastes de hipóteses:

Contraste bilateral	Contraste unilateral á esquerda	Contraste unilateral á dereita
$H_0 : \sigma_X^2 = \sigma_Y^2$ ou $\frac{\sigma_X^2}{\sigma_Y^2} = 1$ $H_a : \sigma_X^2 \neq \sigma_Y^2$ ou $\frac{\sigma_X^2}{\sigma_Y^2} \neq 1$	$H_0 : \sigma_X^2 \geq \sigma_Y^2$ ou $\frac{\sigma_X^2}{\sigma_Y^2} \geq 1$ $H_a : \sigma_X^2 < \sigma_Y^2$ ou $\frac{\sigma_X^2}{\sigma_Y^2} < 1$	$H_0 : \sigma_X^2 \leq \sigma_Y^2$ ou $\frac{\sigma_X^2}{\sigma_Y^2} \leq 1$ $H_a : \sigma_X^2 > \sigma_Y^2$ ou $\frac{\sigma_X^2}{\sigma_Y^2} > 1$

Estatístico de contraste e rexión de rexeitamento

O pivote hérdase dos intervalos de confianza, e considerando a súa distribución baixo a hipótese nula teriamos o seguinte estatístico de contraste:

$$T = \frac{S_{c,X}^2}{S_{c,Y}^2} \in F_{n_X-1,n_Y-1},$$

que nos permite calcular a rexión de rexeitamento asociada a cada tipo de contraste.

$\left(0\,,\ f_{n_X-1,n_Y-1,\alpha/2}\right) \cup$ $\left(f_{n_X-1,n_Y-1,1-\alpha/2}\,,\ +\infty\right)$	$\left(0,\ f_{n_X-1,n_Y-1,\alpha}\right)$	$\left(f_{n_X-1,n_Y-1,1-\alpha}\,,\ +\infty\right)$

Nota. Expoñemos o caso máis habitual no que o segundo termo das ecuacións que aparecen nas hipóteses é sempre un $\left(\sigma_{X,0}^2/\sigma_{Y,0}^2 = 1\right)$. Realmente poderiamos propoñer un contraste (de calquera das tres tipoloxías) no que se valorase o cociente das varianzas poboacionais contra un certo valor, non necesariamente a unidade.