Tema 3. Comparaciones de dos poblaciones

Contenidos

- Hipótesis para la diferencia entre las medias de dos poblaciones: muestras pareadas
- Hipótesis para la diferencia entre las medias de dos poblaciones: muestras independientes
  - Dos poblaciones normales con varianzas iguales (y desconocidas)
  - Dos poblaciones normales con varianzas conocidas
  - Dos poblaciones no normales y muestras grandes
    - Dos poblaciones Bernoulli
- Hipótesis para la razón de las varianzas de dos poblaciones: muestras independientes

Tema 3. Comparaciones de dos poblaciones

Objetivos de aprendizaje

Al finalizar este tema, debieras ser capaz de:

- Llevar a cabo un contraste de hipótesis para la diferencia de las medias y para el cociente de las varianzas de dos poblaciones
- Construir intervalos de confianza para la diferencia o el cociente
- Diferenciar las situaciones en las que un contraste basado en muestras pareadas es adecuado, de aquellas en las que se debe aplicar un contraste basado en muestras independientes
- Calcular la potencia de un contraste y la probabilidad de un error de Tipo II

Tema 3. Comparaciones de dos poblaciones

Referencias

- Newbold, P. "Estadística para administración y economía"
  - Capítulo 9 (9.6-9.9)
- Ross, S. "Introducción a la Estadística"
  - Capítulo 10

Introducción

En este tema estudiamos el caso en el que en lugar de disponer de una muestra aleatoria, tenemos dos muestras aleatorias de dos poblaciones, y estamos interesados en contrastar:

- la diferencia entre las medias de las dos poblaciones
  - en el caso de muestras pareadas
  - y en el caso de muestras independientes
- el cociente entre las varianzas de las dos poblaciones
  - en el caso de muestras independientes

Emplearemos los procedimientos introducidos en los Temas 1 y 2 para construir intervalos de confianza y realizar contrastes de hipótesis para las diferencias o cocientes de los parámetros de las poblaciones indicados.

Contrastes para la diferencia entre dos medias: muestras pareadas

Ejemplo: Se ha llevado a cabo un estudio sobre la relación entre la actividad cerebral mientras se ven anuncios en televisión y la capacidad de la persona para recordar dichos anuncios. Se han mostrado anuncios de dos marcas para diez productos a las personas en la muestra. Para cada anuncio se ha medido la capacidad de cada persona para recordarlo pasadas 24 h., y a cada anuncio de un producto se le han asignado las etiquetas "recuerdo fuerte" o "recuerdo débil". La siguiente tabla muestra un índice de la actividad cerebral de las personas que han visto estos anuncios en el estudio.

producto: i	1	2	3	4	5	6	7	8	9	10
recuerdo fuerte: x_i	137	135	83	125	47	46	114	157	57	144
recuerdo débil: y_i	53	114	81	86	34	66	89	113	88	111
dif.: $d_i = x_i - y_i$	84	21	2	39	13	-20	25	44	-31	33

Contrastes para la diferencia entre dos medias: muestras pareadas

- Sea X una población con media μ_X e Y otra población con media μ_Y.
- Disponemos de una muestra aleatoria de n observaciones pareadas de ambas poblaciones, $\{(X_i, Y_i)\}$. Denotaremos como

$$d_1 = x_1 - y_1, d_2 = x_2 - y_2, \ldots, d_n = x_n - y_n$$

las n diferencias de valores con media $\bar{d}$ y cuasi desviación típica s_d.
  - Supondremos que la población de las diferencias sigue una distribución normal.
- Contraste bilateral $H_0: \mu_X - \mu_Y = D_0$ frente a $H_1: \mu_X - \mu_Y \neq D_0$
  - El estadístico del contraste es

$$T = \frac{\bar{D} - D_0}{s_D/\sqrt{n}} \sim_{H_0} t_{n-1}$$

  - La región de rechazo (a un nivel de significación α) es:

$$\mathrm{RR}_\alpha = \{t : t < -t_{n-1;\alpha/2} \text{ o } t > t_{n-1;\alpha/2}\}$$

Contrastes para la diferencia entre dos medias: muestras pareadas

Ejemplo: cont.

Población:

$D =$ "diferencia entre recuerdo fuerte y débil"

$D \sim N(\mu_X - \mu_Y, \sigma_D^2)$

MAS: $n = 10$

Muestra: $\bar{d} = \frac{210}{10} = 21$

$s_d^2 = \frac{142022 - 10(21)^2}{10-1} = 1088$

Objetivo: contrastar

$$\boxed{H_0: \ \mu_X - \mu_Y \leq \overbrace{0}^{D_0} \text{ frente a } H_1: \ \mu_X - \mu_Y > 0}$$

(Contraste unilateral)

Estadístico del contraste:

$$T = \frac{\bar{D} - D_0}{s_D/\sqrt{n}} \sim t_{n-1}$$

Valor observado del estadístico:

$D_0 = 0 \qquad n = 10$

$\bar{d} = 21 \qquad s_d = \sqrt{1088} = 32.98$

$$t = \frac{\bar{d} - D_0}{s_d/\sqrt{n}} = \frac{21}{32.98/\sqrt{10}} = 2.014$$

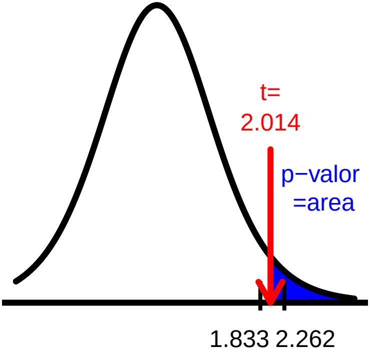
Contrastes para la diferencia entre dos medias: muestras pareadas

Ejemplo: cont.

$$\text{p-valor} = P(T \geq 2.014)$$
$$\in (0.025, 0.05) \text{ ya que}$$
$$\overbrace{1.833}^{t_{9;0.05}} < 2.014 < \overbrace{2.262}^{t_{9;0.025}}$$

Por tanto, como p-valor $< \alpha = 0.05$, rechazamos la hipótesis nula a este nivel.

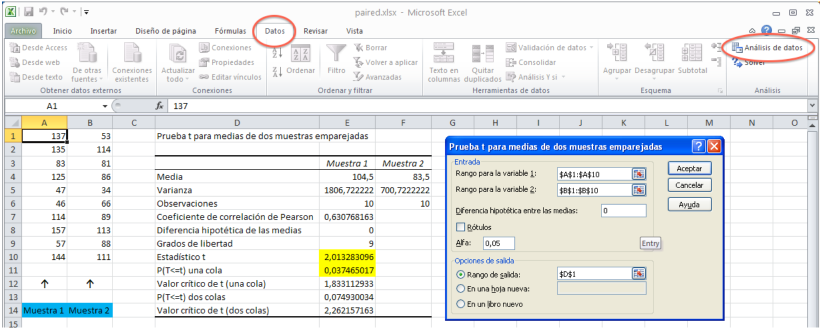
Densidad t_{n-1}



Conclusión: La evidencia muestral apoya que en promedio la actividad cerebral es mayor para el grupo con recuerdo fuerte que para el grupo con recuerdo débil. Si la actividad cerebral promedio fuese igual para ambos grupos, la probabilidad de obtener muestras tan extremas o más que la observada estaría entre 0.025 y 0.05 (un valor bajo).

Contrastes para la diferencia entre dos medias: muestras pareadas

Ejemplo: cont. En Excel: Ir al menu “Datos”, submenu “Análisis de datos”, seleccionar la opción: “Prueba t para medias de dos muestras emparejadas”
Columnas A y B (datos), en amarillo se muestran el valor observado del estadístico y el p-valor.



Contraste bilateral para la diferencia entre dos medias via IC: muestras pareadas

Ejemplo: cont. Construir un intervalo de confianza al 95% para $\mu_X - \mu_Y$.

$$\begin{aligned}
\mathrm{IC}_{0.95}(\mu_X - \mu_Y) &= \left(\bar{d} - t_{n-1;0.025}\frac{s_d}{\sqrt{n}}, \bar{d} + t_{n-1;0.025}\frac{s_d}{\sqrt{n}}\right) \\
&= \left(21 - 2.262\frac{32.98}{\sqrt{10}}, 21 + 2.262\frac{32.98}{\sqrt{10}}\right) \\
&= (-2.59, 44.59)
\end{aligned}$$

Como el valor 0 pertenece a este intervalo, no podemos rechazar la hipótesis nula de la igualdad de las medias de las dos poblaciones a un nivel de significación de $\alpha = 0.05$.

Contrastes para la diferencia entre dos medias: muestras normales independientes, varianzas de poblaciones iguales

- Sea X una población con media μ_X y varianza σ_X^2, e Y otra población con media μ_Y y varianza σ_Y^2
  - ambas distribuidas normalmente
  - con varianzas poblacionales desconocidas, pero iguales $\sigma^2 = \sigma_X^2 = \sigma_Y^2$
- Muestras aleatorias de n_1 observaciones de X y n_2 observaciones de Y, independientes.
- Contraste bilateral $H_0: \ \mu_X - \mu_Y = D_0$ frente a $H_1: \ \mu_X - \mu_Y \neq D_0$
  - El estadístico del contraste es

$$T = \frac{\bar{X} - \bar{Y} - D_0}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim_{H_0} t_{n_1+n_2-2}$$

  donde el estimador de la varianza común para las dos poblaciones es

$$s_p^2 = \frac{(n_1 - 1)s_X^2 + (n_2 - 1)s_Y^2}{n_1 + n_2 - 2}$$

  Nota: grados de libertad $= n_1 + n_2 - 2$ (número de observaciones de las muestras menos dos - por tener que estimar μ_X y μ_Y)
  - La región de rechazo (para un nivel de significación α) es:

$$\mathrm{RR}_\alpha = \{t : t < -t_{n_1+n_2-2;\alpha/2} \text{ o } t > t_{n_1+n_2-2;\alpha/2}\}$$

Contrastes para la diferencia entre dos medias: muestras normales independientes, varianzas de poblaciones iguales

Ejemplo: 9.8 (Newbold) Se estudia el efecto que la presencia de un moderador puede tener en el número de ideas generadas en un grupo de trabajo. Se observan grupos de cuatro personas, con y sin moderador. En una muestra aleatoria de cuatro grupos con moderador el número promedio de ideas generadas por grupo fue 78.0, con cuasi desviación típica muestral de 24.4. Para una muestra independiente de cuatro grupos sin moderador el promedio de ideas generadas fue 63.5, y su cuasi desviación típica fue 20.2. Suponiendo que distribuciones normales con varianzas iguales, contraste la hipótesis nula (para $\alpha = 0.1$) de igualdad de medias, frente a la alternativa de que la media de la población es mayor para grupos con moderador.

Población 1:
$X =$ "número de ideas en grupos con moderador"
$X \sim N(\mu_X, \sigma_X^2)$

MAS: $n_1 = 4$

Muestra: $\bar{x} = 78.0$
$s_x = 24.4$

Población 2:
$Y =$ "número de ideas en grupos sin moderador"
$X \sim N(\mu_Y, \sigma_Y^2)$

MAS: $n_2 = 4$

Muestra: $\bar{y} = 63.5$
$s_y = 20.2$

Suponemos muestras normales independientes y $\boxed{\sigma_X^2 = \sigma_Y^2 = \sigma^2}$

Contrastes para la diferencia entre dos medias: muestras normales independientes, varianzas de poblaciones iguales

Ejemplo: 9.8 (Newbold cont.)

Objetivo: contrastar

$$\boxed{H_0 : \mu_X - \mu_Y = \overbrace{0}^{D_0}}$$

frente a

$$\boxed{H_1 : \mu_X - \mu_Y > 0}$$

(Test unilateral)

Estadístico del contraste:

$$T = \frac{\bar{X} - \bar{Y}}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim_{H_0} t_{n_1+n_2-2}$$

Valor observado del estadístico:

$$D_0 = 0 \qquad n_1 = 4 \; n_2 = 4$$

$$\bar{x} = 78.0 \; s_x = 24.4 \qquad \bar{y} = 63.5 \; s_y = 20.2$$

$$\begin{aligned} s_p^2 &= \frac{(n_1 - 1)s_x^2 + (n_2 - 1)s_y^2}{n_1 + n_2 - 2} \\ &= \frac{(4-1)24.4^2 + (4-1)20.2^2}{4+4-2} \\ &= 501.7 \end{aligned}$$

$$\begin{aligned} s_p &= \sqrt{501.7} = 22.4 \\ t &= \frac{\bar{x} - \bar{y}}{s_p \sqrt{1/n_1 + 1/n_2}} \\ &= \frac{78.0 - 63.5}{22.4\sqrt{1/4 + 1/4}} = 0.915 \end{aligned}$$

Región de rechazo:

$$\mathrm{RR}_{0.1} = \{t : t > \overbrace{t_{6;0.1}}^{1.440}\}$$

Como $t = 0.915 \notin RR_{0.1}$, no podemos rechazar la hipótesis nula a un nivel del 10%.

Conclusión: Los datos muestrales no contienen suficiente evidencia para pensar que en promedio se generan más ideas en grupos con moderador. Pero para tamaños muestrales tan pequeños el contraste tiene potencia baja y serían necesarias diferencias muy grandes entre las medias de las poblaciones para rechazar la hipótesis nula.

Contrastes para la diferencia entre dos medias: muestras normales independientes, varianzas de poblaciones iguales

Ejemplo: 9.8 (Newbold cont.) Construya un intervalo de confianza al 99% para $\mu_X - \mu_Y$.

$$\begin{aligned} \mathrm{IC}_{0.99}(\mu_X - \mu_Y) &= \left(\bar{x} - \bar{y} \mp t_{n_1+n_2-2;0.005} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right) \\ &= \left(78.0 - 63.5 \mp 3.707 \cdot 22.4 \sqrt{\frac{1}{4} + \frac{1}{4}}\right) \\ &= (-44.22, 73.22) \end{aligned}$$

Como el valor 0 pertenece a este intervalo, no podemos rechazar la hipótesis nula de igualdad de las medias de las dos poblaciones a un nivel de significación $\alpha = 0.01$.

Contrastes para la diferencia entre dos medias: muestras grandes independientes o dos poblaciones normales con varianzas conocidas

- Sea X una población con media μ_X y varianza σ_X^2, e Y otra población con media μ_Y y varianza σ_Y^2.
- Muestras aleatorias de n_1 observaciones de X y n_2 observaciones de Y, independientes, y
  - Bien tanto n_1 como n_2 son grandes y σ_X^2 y σ_Y^2 son desconocidas,
  - O X e Y siguen distribuciones normales y σ_X^2 y σ_Y^2 son conocidas
- Contraste bilateral $H_0: \ \mu_X - \mu_Y = D_0$ frente a $H_1: \ \mu_X - \mu_Y \neq D_0$
  - El estadístico del contraste es:
    - **Bien**
$$Z = \frac{\bar{X} - \bar{Y} - D_0}{\sqrt{\frac{s_X^2}{n_1} + \frac{s_Y^2}{n_2}}} \sim_{H_0,\ aprox.} N(0,1)$$
    - **O**
$$Z = \frac{\bar{X} - \bar{Y} - D_0}{\sqrt{\frac{\sigma_X^2}{n_1} + \frac{\sigma_Y^2}{n_2}}} \sim_{H_0} N(0,1)$$
  - La región de rechazo (para un nivel de significación α) es:
$$\mathrm{RR}_\alpha = \{z : z < -z_{\alpha/2} \text{ o } z > z_{\alpha/2}\}$$

Contrastes para la diferencia entre dos medias: muestras grandes independientes o dos poblaciones normales con varianzas conocidas

Ejemplo: 9.7 (Newbold) Se realiza un estudio entre auditores sobre la actividad de las mujeres en su profesión. A los encuestados se les pide que den su opinión con un valor entre uno (muy en desacuerdo) y cinco (muy de acuerdo) sobre la afirmación "En auditoría se asignan los mismos trabajos a las mujeres y a los hombres". De una muestra de 186 auditores (varones) se obtuvo una respuesta promedio de 4.059 con una cuasi desviación típica de 0.839. Para una muestra independiente de 172 mujeres auditoras la respuesta promedio fue de 3.680 con una cuasi desviación típica de 0.966. Contraste la hipótesis nula (para $\alpha = 0.0001$) de que las medias de las dos poblaciones son iguales, frente a la alternativa de que la media de la población es mayor para auditores varones.

Población 1:
$X =$ "respuesta de un auditor varón"
$X \sim \ \mu_X, \sigma_X^2$

MAS: $n_1 = 186$

Muestra: $\bar{x} = 4.059$
$s_x = 0.839$

Población 2:
$Y =$ "respuesta de una mujer auditora"
$X \sim \ \mu_Y, \sigma_Y^2$

MAS: $n_2 = 172$

Muestra: $\bar{y} = 3.680$
$s_y = 0.966$

Contrastes para la diferencia entre dos medias: muestras grandes independientes o dos poblaciones normales con varianzas conocidas

Ejemplo: 9.7 (Newbold cont.)

Objetivo: contrastar

$$\boxed{H_0 : \mu_X - \mu_Y = \overbrace{0}^{D_0}}$$

frente a

$$\boxed{H_1 : \mu_X - \mu_Y > 0}$$

(Contraste unilateral)

Estadístico del contraste:

$$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_X^2}{n_1} + \frac{s_Y^2}{n_2}}} \sim_{H_0,\ aprox.} N(0, 1)$$

Valor observado del estadístico:

$$D_0 = 0 \qquad n_1 = 186 \; n_2 = 172$$

$$\bar{x} = 4.059 \; s_x = 0.839 \qquad \bar{y} = 3.680 \; s_y = 0.966$$

$$\begin{aligned} z &= \frac{\bar{x} - \bar{y}}{\sqrt{s_x^2/n_1 + s_y^2/n_2}} \\ &= \frac{4.059 - 3.680}{\sqrt{0.839^2/186 + 0.966^2/172}} = 3.95 \end{aligned}$$

Región de rechazo:

$$RR_{0.0001} = \{z : z > \overbrace{z_{0.0001}}^{3.75}\}$$

Como $z = 3.95 \in RR_{0.0001}$, rechazamos la hipótesis nula a un nivel del 0.01%.

Conclusión: Los datos contienen una evidencia muy fuerte en favor de que la respuesta promedio entre los varones es mayor que entre las mujeres - esto es, en promedio los varones están mas convencidos que las mujeres de que se asignan los mismos trabajos a las mujeres que a los hombres.

Contrastes para la diferencia entre dos medias: muestras grandes independientes o dos poblaciones normales con varianzas conocidas

Ejemplo: 9.7 (Newbold) Construya un intervalo de confianza al 95% para $\mu_X - \mu_Y$.

$$
\begin{aligned}
IC_{0.95}(\mu_X - \mu_Y) &= \left(\bar{x} - \bar{y} \mp z_{0.025}\sqrt{\frac{s_x^2}{n_1} + \frac{s_y^2}{n_2}}\right) \\
&= \left(4.059 - 3.680 \mp 1.96\sqrt{0.839^2/186 + 0.966^2/172}\right) \\
&= (0.19, 0.57)
\end{aligned}
$$

Como el valor 0 no pertenece a este intervalo, podemos rechazar la hipótesis nula de igualdad de las dos medias poblacionales a un nivel de significación $\alpha = 0.05$.

Contrastes para la diferencia entre dos proporciones: muestras grandes independientes

- Sea $X \sim \text{Bernoulli}(p_X)$ y sea $Y \sim \text{Bernoulli}(p_Y)$, donde p_X y p_Y son dos proporciones poblacionales para los individuos que cumplan una propiedad de interés.
- Muestras aleatorias de n_1 observaciones de X y n_2 observaciones de Y, independientes, y
  - tanto n_1 como n_2 son grandes
- Contraste bilateral $H_0: \ p_X - p_Y = 0$ frente a $H_1: \ p_X - p_Y \neq 0$
  - El estadístico del contraste es:

$$Z = \frac{\hat{p}_X - \hat{p}_Y}{\sqrt{\hat{p}_0(1-\hat{p}_0)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim_{H_0, \ aprox.} N(0,1),$$

    donde

$$\hat{p}_0 = \frac{n_1 \hat{p}_X + n_2 \hat{p}_Y}{n_1 + n_2}$$

  - La región de rechazo (para un nivel de significación α) es:

$$\text{RR}_\alpha = \{z : z < -z_{\alpha/2} \text{ or } z > z_{\alpha/2}\}$$

Contrastes para la diferencia entre dos proporciones: muestras grandes independientes

Ejemplo: 9.9 (Newbold) En Investigación de Mercados, es importante conseguir un porcentaje de respuestas elevado para las encuestas. Para mejorar este porcentaje se puede incluir una pregunta inicial de motivación que aumente el interés del encuestado por completarlo. Se han enviado cuestionarios con pregunta de motivación sobre la mejora los espacios de ocio en una ciudad, a una muestra de 250 hogares, obteniendo 101 respuestas. Otros cuestionarios idénticos sin pregunta de motivación se han enviado a otra muestra independiente de 250 hogares, obteniendo 75 respuestas. Contraste la hipótesis nula de que las dos proporciones poblacionales sean iguales, frente a la alternativa de que la tasa de respuestas sea más elevada cuando se incluye pregunta de motivación.

Población 1:
$X = 1$ si una persona completa el cuestionario con pregunta de motivación, y 0 en caso contrario
$X \sim Bernoulli(p_X)$

MAS: $n_1 = 250$

Muestra: $\hat{p}_x = \frac{101}{250} = 0.404$

Población 2:
$Y = 1$ si una persona completa el cuestionario sin pregunta de motivación, y 0 en caso contrario
$Y \sim Bernoulli(p_Y)$

MAS: $n_2 = 250$

Muestra: $\hat{p}_y = \frac{75}{250} = 0.300$

Contrastes para la diferencia entre dos proporciones: muestras grandes independientes

Ejemplo: 9.9 (Newbold cont.)

Objetivo: contrastar

$$\boxed{H_0 : \; p_X = p_Y}$$

frente a

$$\boxed{H_1 : \; p_X > p_Y}$$

(Contraste unilateral)

Estadístico del contraste:

$$Z = \frac{\hat{p}_X - \hat{p}_Y}{\sqrt{\hat{p}_0(1-\hat{p}_0)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \sim_{H_0,\ \textit{aprox.}} N(0,1)$$

Valor observado del estadístico:

$$n_1 = 250 \qquad n_2 = 250$$

$$\hat{p}_x = 0.404 \qquad \hat{p}_y = 0.300$$

$$\begin{aligned} \hat{p}_0 &= \frac{n_1\hat{p}_x + n_2\hat{p}_y}{n_1 + n_2} \\ &= \frac{250(0.404) + (250)(0.300)}{250 + 250} \\ &= 0.352 \end{aligned}$$

$$\begin{aligned} z &= \frac{\hat{p}_x - \hat{p}_y}{\sqrt{\hat{p}_0(1-\hat{p}_0)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \\ &= \frac{0.404 - 0.300}{\sqrt{0.352(1-0.352)\left(\frac{1}{250}+\frac{1}{250}\right)}} = 2.43 \end{aligned}$$

p-valor $= P(Z \geq z) = P(Z \geq 2.43) = 0.0075$

Como el p-value es muy pequeño, podemos rechazar la hipótesis nula a cualquier nivel de significación mayor que 0.0075.

Conclusión: Los datos muestrales contienen una fuerte evidencia de que al incluir una pregunta de motivación se obtiene una tasa de respuesta más elevada que cuando no se incluye.

Contrastes para la diferencia entre dos proporciones: muestras grandes independientes

Ejemplo: 9.9 (Newbold cont.) Construya un intervalo de confianza al 95% para $p_X - p_Y$.

$$
\begin{aligned}
IC_{0.95}(p_X - p_Y) &= \left(\hat{p}_x - \hat{p}_y \mp z_{0.025}\sqrt{\hat{p}_0(1-\hat{p}_0)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}\right) \\
&= \left(0.404 - 0.300 \mp 1.96\sqrt{0.352(1-0.352)\left(\frac{1}{250}+\frac{1}{250}\right)}\right) \\
&= (0.1877, 0.0203)
\end{aligned}
$$

Como el valor 0 no pertenece a este intervalo, podemos rechazar la hipótesis nula de igualdad de las proporciones de las dos poblaciones para un nivel de significación $\alpha = 0.05$.

Contrastes para el cociente de varianzas: muestras normales

- Sea X una población con media μ_X y varianza σ_X^2, e Y otra población con media μ_Y y varianza σ_Y^2,
  - ambas distribuidas normalmente
- Muestras aleatorias de n_1 observaciones de X y n_2 observaciones de Y, independientes.
- Contraste bilateral $H_0: \ \sigma_X^2 = \sigma_Y^2 \ (= \sigma^2)$ frente a $H_1: \ \sigma_X^2 \neq \sigma_Y^2$
  - El estadístico del contraste es

$$F = \frac{s_X^2}{s_Y^2} \sim_{H_0} F_{n_1-1, n_2-1}$$

  - La región de rechazo (para un nivel de significación α) es:

$$\mathrm{RR}_\alpha = \{f : f < F_{n_1-1,n_2-1;1-\alpha/2} \text{ o } f > F_{n_1-1,n_2-1;\alpha/2}\}$$

La distribución F

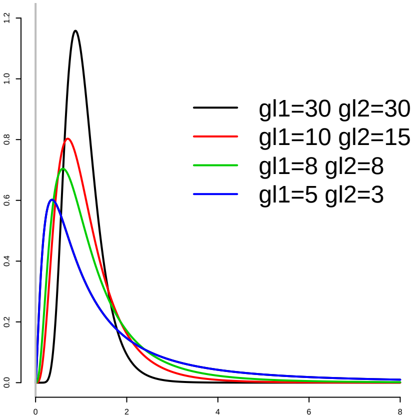
$X_1, X_2, \ldots, X_n$ y $Y_1, Y_2, Y_3, \ldots, Y_m$ son dos conjuntos de variables aleatorias independientes, con distribución $N(0,1)$. La variable aleatoria (cociente de dos v.a.s chi-cuadrado normalizadas)

$$F = \frac{\frac{1}{n}\sum_{i=1}^{n} X_i^2}{\frac{1}{m}\sum_{i=1}^{m} Y_i^2}$$

sigue una distribución $F_{n,m}$ con n y m grados de libertad. Para el resultado de la transparencia anterior:

$$\frac{s_X^2}{s_Y^2} =_{H_0} \frac{\frac{1}{n_1-1}\overbrace{\frac{(n_1-1)s_X^2}{\sigma^2}}^{\chi^2_{n_1-1}}}{\frac{1}{n_2-1}\underbrace{\frac{(n_2-1)s_Y^2}{\sigma^2}}_{\chi^2_{n_2-1}}} \sim F_{n_1-1,n_2-1}$$

Densidades F



Contrastes para el cociente de varianzas: muestras normales

Ejemplo: 9.10 (Newbold) Para una muestra de 17 bonos industriales emitidos recientemente con calificación AAA, la cuasi varianza de sus vencimientos (en años al cuadrado) fue de 123.35. Para otra muestra independiente de 11 bonos industriales emitidos con calificación CCC, la cuasi varianza de sus vencimientos fue de 8.02. Si se denotan las correspondientes varianzas poblacionales como σ_X^2 y σ_Y^2, lleve a cabo un contraste bilateral para compararlas al 5%.

Población 1:
X vencimiento de bonos AAA (en años)
$X \sim N(\mu_X, \sigma_X^2)$

MAS: $n_1 = 17$

Muestra: $s_x^2 = 123.35$

Población 2:
Y vencimiento de bonos CCC (en años)
$Y \sim N(\mu_Y, \sigma_Y^2)$

MAS: $n_2 = 11$

Muestra: $s_y^2 = 8.02$

Contrastes para el cociente de varianzas: muestras normales

Ejemplo: 9.10 (Newbold cont.)

Objetivo: contrastar

$$\boxed{H_0 : \sigma_X^2 = \sigma_Y^2}$$

frente a

$$\boxed{H_1 : \sigma_X^2 \neq \sigma_Y^2}$$

(Contraste bilateral)

Estadístico del contraste:
$F = \frac{s_X^2}{s_Y^2} \sim_{H_0} F_{n_1-1,n_2-1}$
Valor observado del estadístico:

$$n_1 = 17 \qquad n_2 = 11$$

$$s_x^2 = 123.35 \qquad s_y^2 = 8.02$$

$$f = \frac{123.35}{8.02} = 15.38$$

Región de rechazo:

$$\begin{aligned} RR_{0.10} &= \{f : f < \overbrace{F_{16,10;1-0.05}}^{0.402}\} \\ &\cup \{f : f > \underbrace{F_{16,10;0.05}}_{2.83}\} \end{aligned}$$

Nota: el cuantil $F_{16,10;0.05} = 2.83$ aparece en la tabla de la F, pero no $F_{16,10;1-0.05}$. Para calcularlo podemos emplear la propiedad de esta distribución $\boxed{F_{n,m;\alpha} = \frac{1}{F_{m,n;1-\alpha}}}$

Obtenemos

$$F_{16,10;1-0.05} = \frac{1}{F_{10,16;0.05}} = \frac{1}{2.49} = 0.402$$

Vemos que $f = 15.38 \in RR_{0.10}$.
Conclusión: Existe una fuerte evidencia de que las dos varianzas poblacionales son distintas.

Contraste bilateral para el cociente de varianzas mediante intervalos de confianza

Ejemplo: 9.10 (Newbold cont.) Construya un intervalo de confianza al 90% para el cociente de las varianzas.

$$
\begin{aligned}
\mathrm{IC}_{0.90}\left(\frac{\sigma_X^2}{\sigma_Y^2}\right) &= \left(\frac{s_x^2}{s_y^2}\frac{1}{F_{n_1-1,n_2-1;0.05}}, \frac{s_x^2}{s_y^2}\frac{1}{F_{n_1-1,n_2-1;1-0.05}}\right) \\
&= \left(\frac{123.35}{8.02}\frac{1}{2.83}, \frac{123.35}{8.02}\frac{1}{0.402}\right) \\
&= (5.43, 38.26)
\end{aligned}
$$

Como era de esperar por el resultado anterior, el valor 1 no pertenece a este intervalo, y podemos rechazar la hipótesis nula de que las dos varianzas poblacionales sean iguales, para un nivel de significación $\alpha = 0.1$.

Estadísticos pivotales

Parámetro	Hipótesis	Estadístico del contraste
$\mu_X - \mu_Y = D_0$	Diferencias normales Muestras pareadas	$\frac{\bar{D} - D_0}{s_D / \sqrt{n}} \sim t_{n-1}$
	Pobls. normales Varianzas iguales	$\frac{\bar{X} - \bar{Y} - D_0}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim_{H_0} t_{n_1+n_2-2}$
	Pobls. normales Vars. conocidas	$\frac{\bar{X} - \bar{Y} - D_0}{\sqrt{\frac{\sigma_X^2}{n_1} + \frac{\sigma_Y^2}{n_2}}} \sim_{H_0} N(0,1)$
	Pobls. no normales Vars. desconocidas Muestras grandes	$\frac{\bar{X} - \bar{Y} - D_0}{\sqrt{\frac{s_X^2}{n_1} + \frac{s_Y^2}{n_2}}} \sim_{H_0} \; aprox \; N(0,1)$
$p_X - p_Y = 0$	Pobls. Bernoulli Muestras grandes	$\frac{\hat{p}_X - \hat{p}_Y}{\sqrt{\hat{p}_0(1-\hat{p}_0)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim_{H_0} \; aprox \; N(0,1)$
$\sigma_X^2 / \sigma_Y^2 = 1$	Pobls. normales	$\frac{s_X^2}{s_Y^2} \sim_{H_0} F_{n_1-1,n_2-1}$

Pregunta: ¿Como definirías RR_α para contrastes unilaterales?