

Capítulo 8

Comparación de poblaciones

-
1. Introducción
 2. Comparación de dos medias usando muestras independientes
 3. Comparación de dos medias usando datos emparejados
 4. Comparación de dos proporciones
 5. Comparación de dos varianzas en poblaciones normales
-

⁰Apuntes realizados por Ismael Sánchez. Universidad Carlos III de Madrid.

8.1. Introducción

En este tema estamos interesados en comparar dos poblaciones a partir de la información de dos muestras, una de cada población. Por ejemplo, podríamos estar interesados en comparar dos algoritmos de computación alternativos, para ver cuál es más rápido. Ejecutaríamos entonces el Algoritmo I n_1 veces y el Algoritmo II n_2 veces, y a partir de la información de ambas muestras **inferir** si uno de los dos algoritmos será, por término medio, más rápido en ejecuciones futuras. Otro ejemplo sería la comparación de la resistencia a la rotura de dos aleaciones diferentes. Haríamos entonces un ensayo con n_1 piezas con la Aleación I y n_2 piezas de la Aleación 2, y a partir de la comparación de ambas muestras podríamos **inferir** si hay diferencias significativas que nos hagan decantarnos por una de las dos aleaciones.

En otros problemas, lo que queremos comparar no serán sólo las medias, como en los dos ejemplos anteriores, sino también las varianzas. Un ejemplo típico en el que interesa comparar varianzas es la comparación de instrumentos de medida. Será preferible aquél instrumento de medida con menos variabilidad.

Al igual que se decía para el caso de una sola población, el objetivo final **no** es sacar conclusiones sobre los datos concretos obtenidos en las muestras, sino en los datos que se generen en el futuro. No nos interesa saber si la muestra 1 tiene valores más o menos diferentes a la muestra 2, sino si las poblaciones de las que proceden son diferentes. En ese caso, esperaremos que las diferencias se produzcan también en posteriores muestras, y tomaremos decisiones en función de esas expectativas.

Si las diferencias entre las muestras son pequeñas, podrían perfectamente explicarse por la variabilidad del muestreo. En general, si extraemos dos muestras de una misma población, las muestras no serán idénticas: tendrán diferente media muestral, varianza muestral, etc. Por tanto, el que dos muestras sean diferentes no significa necesariamente que procedan de poblaciones diferentes. Cuando la diferencia entre dos muestras es pequeña, de forma que sea del mismo orden de magnitud que la que se observa en muestras de una misma población, diremos que **la diferencia encontrada no es significativa**. Por el contrario, cuando la diferencia entre dos muestras sea mayor que la que se suele encontrar en muestras de una misma población, diremos que **la diferencia encontrada es significativa**. En ese caso, concluiremos que las poblaciones serán, con mucha probabilidad, diferentes. ¿Cómo establecemos que la diferencia entre las muestras es o no es suficientemente grande para considerarla significativa? Lo haremos mediante pruebas estadísticas similares a las presentadas en temas anteriores.

En general, el problema que nos planteamos en este tema es la comparación de dos poblaciones X_1 y X_2 independientes. La primera población tiene media μ_1 y varianza σ_1^2 mientras que la segunda población tiene media μ_2 y varianza σ_2^2 . La inferencia sobre ambas poblaciones la realizaremos a partir de sendas muestras. La primera muestra será de la población X_1 de tamaño n_1 y elementos X_{11}, \dots, X_{1n_1} , y la segunda muestra será de la población X_2 y de elementos X_{21}, \dots, X_{2n_2} .

Supondremos que las muestras son lo suficientemente grandes (más de 30 elementos en cada muestra) como para poder utilizar el teorema central del límite al hacer comparaciones basadas en medias muestrales. De esta forma, podremos asumir que las medias muestrales se distribuyen según cierta normal aunque las poblaciones no sean normales. Si las muestras no son grandes, supondremos (suposición que habrá que comprobar) que las variables aleatorias X_1 y X_2 son normales, y utilizaremos estadísticos de contraste basados en dicha normalidad.

8.2. Comparación de dos medias usando muestras independientes

8.2.1. Introducción

Queremos hacer inferencia sobre la diferencia de dos medias poblacionales $\mu_1 - \mu_2$ de dos poblaciones X_1 y X_2 a partir de la información de dos muestras. De la población X_1 se extrae una muestra de tamaño n_1 , y de la población X_2 se extrae una muestra de tamaño n_2 . Al no compartir ningún elemento, ambas muestras serán independientes. Si las poblaciones X_1 y X_2 son normales, o en caso de no serlo tenemos muestras grandes, tendremos que las medias muestrales de cada población tienen la siguiente distribución en el muestreo

$$\bar{X}_1 \sim N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right); \bar{X}_2 \sim N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right). \quad (8.1)$$

La inferencia sobre $\mu_1 - \mu_2$, tanto los intervalos de confianza como los contrastes de hipótesis, se hará tratando esta diferencia de parámetros como si fuese un nuevo parámetro $\delta = \mu_1 - \mu_2$, cuyo estimador es $\hat{\delta} = \bar{X}_1 - \bar{X}_2$. Veremos entonces que la mayoría de los resultados se basarán en la aplicación de propiedades que ya hemos visto en temas anteriores, y habrá muy pocos elementos nuevos.

Al estudiar las propiedades de las variables normales, vimos que la combinación lineal de normales es siempre normal. Por tanto, de (8.1) tendremos que, al ser ambas muestras independientes,

$$\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right). \quad (8.2)$$

Estandarizando tenemos entonces que

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1), \quad (8.3)$$

que será el resultado básico para hacer inferencia de la diferencia $\mu_1 - \mu_2$.

8.2.2. Intervalos para la diferencia de medias

De (8.3) se pueden deducir muchos resultados para realizar inferencia. Por ejemplo, usando los mismos argumentos que se utilizaron para construir intervalos de confianza para la media μ , tenemos que, de (8.3),

$$P\left(-z_{\alpha/2} < \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} < z_{\alpha/2}\right) = 1 - \alpha,$$

y operando en el interior del paréntesis, se tiene que

$$P\left((\bar{X}_1 - \bar{X}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right) = 1 - \alpha.$$

Por tanto, un intervalo de confianza de $\mu_1 - \mu_2$ de nivel de confianza $(1 - \alpha)$ será

$$IC(1 - \alpha) : \mu_1 - \mu_2 \in \left\{ (\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right\}. \quad (8.4)$$

La interpretación de este intervalo es similar al caso de un solo parámetro. Es decir,

Si tomásemos infinitos pares de muestras (en cada par, tenemos una muestra de cada población), y para cada par de muestras calculásemos el intervalo (8.4), el 100(1- α) % de esos intervalos contendrá a la diferencia $\mu_1 - \mu_2$.

En la realidad tendremos sólo un par de muestras, una por cada población. Y por tanto tendremos un solo intervalo. No sabremos si ese intervalo contendrá o no al valor de $\mu_1 - \mu_2$. Diremos entonces que la confianza que tenemos en que dicho intervalo contenga el valor de la diferencia $\mu_1 - \mu_2$ es $(1 - \alpha)$.

En el caso particular de que las varianzas de ambas poblaciones sean iguales ($\sigma_1^2 = \sigma_2^2 = \sigma^2$) tendremos

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim N(0, 1). \quad (8.5)$$

y entonces el intervalo de confianza será

$$IC(1 - \alpha) : \mu_1 - \mu_2 \in \left\{ (\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2} \sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right\}. \quad (8.6)$$

Ejemplo 1 Se quiere comparar rodamientos de agujas de 45 mm de diámetro de dos fabricantes distintos: SKF e INA. Una característica de la calidad de los rodamientos es su capacidad estática de carga, que es la carga máxima que puede soportar sin que se produzca deformación. Se toma una muestra de 80 rodamientos del SKF y 67 de INA. Sea X_1 la capacidad estática de carga de los rodamientos de SKF y X_2 la capacidad estática de carga de los rodamientos de INA. La media muestral de la capacidad estática de carga de SKF es $\bar{x}_1 = 1400$ kN (kilonewtons), y la de INA es $\bar{x}_2 = 1380$ kN. Las varianzas de X_1 y de X_2 son iguales y tiene un valor conocido $\sigma^2 = 90^2$ kN². Un intervalo de confianza del 95 % de la diferencia de medias poblacionales será, tomando que $z_{0,025} = 1,96$,

$$\begin{aligned} IC(1 - \alpha) & : \mu_1 - \mu_2 \in \left\{ (1400 - 1380) \pm 1,96 \times 90 \sqrt{\frac{1}{80} + \frac{1}{67}} \right\} \\ & = \{20 \pm 29,2\}. \end{aligned}$$

Como el cero está dentro del intervalo, la diferencia entre ambos fabricantes no resulta significativa (con $\alpha = 0,05$) a la luz de las muestras analizadas

Ejemplo 2 Sea X_1 la duración de un filamento de wolframio en espiral simple, y X_2 la duración usando espiral doble. Se toma una muestra de tamaño 200 de ambos tipos de filamentos teniéndose unas duraciones medias de $\bar{x}_1 = 1100$ horas, y $\bar{x}_2 = 1108$ horas. Las varianzas de las duraciones

de cada tipo de filamento son $\sigma_1^2 = 20^2$ y $\sigma_2^2 = 23^2$. Un intervalo de confianza del 95 % para la diferencia entre ambos tipos de filamentos es

$$\begin{aligned} IC(1 - \alpha) &: \mu_1 - \mu_2 \in \left\{ (1100 - 1108) \pm 1,96 \sqrt{\frac{20^2}{200} + \frac{23^2}{200}} \right\} \\ &= \{-8 \pm 4,2\} = (-12,2; -3,8). \end{aligned}$$

En este ejemplo, la diferencia entre ambos tipos de filamentos sí resulta significativa. Vemos, sin embargo, que la diferencia es muy pequeña. El extremo inferior del intervalo es -12.2 horas de diferencia entre ambos filamentos, y el superior indica una diferencia de menos de 4 horas. Esa diferencia es muy pequeña si la comparamos con la vida de los filamentos, que es del orden del millar de horas. La diferencia es significativa, en el sentido de que los datos muestran evidencia suficiente para pensar que el filamento doble dura más (en promedio) que el simple. Sin embargo, el intervalo de confianza nos revela que aunque haya una diferencia real entre ambos filamentos, es muy pequeña y tal vez sea de poco interés práctico.

Vemos en este segundo ejemplo, que no debemos confundir diferencias estadísticamente significativas, con diferencias de interés práctico. **Una diferencia es significativa si los datos muestran suficiente evidencia de que existen diferencias en las poblaciones, ya sean diferencias grandes o pequeñas.** Mientras que una diferencia será de interés en la práctica si supera una determinada magnitud de acuerdo con las necesidades de los analistas.

El ancho del intervalo de confianza muestra la incertidumbre que tenemos sobre las poblaciones a partir de la información de la muestra. Vemos en las fórmulas (8.4) y (8.6) que el ancho de los intervalos depende del tamaño de las muestras. A menores tamaños muestrales, más anchos serán los intervalos, aumentando el riesgo de que puedan contener al valor 0. En esos casos, concluiríamos que la diferencia no es significativa. Por tanto, si una diferencia entre dos poblaciones es grande, será también fácil que las detectemos en muestras de tamaño pequeño o moderado (es decir, que encontremos diferencias significativas en las muestras). Pero si la diferencia entre las poblaciones es muy pequeña, será difícil detectarla salvo que las muestras sean muy grandes. Los intervalos de confianza de diferencias de medias nos permiten encontrar diferencias significativas y también tener una idea de su magnitud.

Si las varianzas poblacionales son desconocidas, emplearemos un estimador de las varianzas. Veamos a continuación en qué cambia la inferencia dependiendo de si las poblaciones tienen varianzas iguales o diferentes.

Caso 1: varianzas iguales $\sigma_1^2 = \sigma_2^2 = \sigma^2$

Si ambas poblaciones tienen varianzas que podemos suponer iguales pero no conocemos su valor, utilizaremos un estimador conjunto a partir de las varianzas estimadas en cada muestra. Esa igualdad de varianzas será una hipótesis que tendremos que contrastar con los datos. Más adelante dedicaremos una sección a este tipo de contrastes de igualdad de varianzas.

Sea \hat{S}_1^2 la cuasivarianza obtenida con la muestra de tamaño n_1 de la población 1, y \hat{S}_2^2 con la muestra de tamaño n_2 de la población 2. Entonces, si suponemos que ambas poblaciones tienen la misma varianza, construiremos un estimador común, que llamaremos \hat{S}_T^2 mediante la expresión

$$\hat{S}_T^2 = \frac{(n_1 - 1)\hat{S}_1^2 + (n_2 - 1)\hat{S}_2^2}{n_1 + n_2 - 2}. \quad (8.7)$$

Puede demostrarse que si sustituimos σ^2 por el estimador \hat{S}_T^2 en (8.5) tendremos un estadístico T que **para poblaciones normales** se distribuirá como una t de Student de $n_1 + n_2 - 2$ grados de libertad. Es decir

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\hat{S}_T \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2} \quad (8.8)$$

y el intervalo de confianza de la diferencia de medias, **para poblaciones normales**, quedará

$$IC(1 - \alpha) : \mu_1 - \mu_2 \in \left\{ (\bar{x}_1 - \bar{x}_2) \pm t_{n_1+n_2-2; \alpha/2} \hat{S}_T \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right\}. \quad (8.9)$$

Si tenemos muestras grandes, no es necesario que las poblaciones implicadas sean normales para poder construir intervalos de confianza basados en el estadístico T . Además, en el caso de poblaciones normales tenemos también que para muestras grandes, la distribución t de Student será muy parecida a la $N(0,1)$. Por esta razón, para muestras grandes y para cualquier tipo de poblaciones, podemos utilizar un intervalo de confianza basado en la $N(0,1)$. Este **intervalo para muestras grandes** será

$$IC(1 - \alpha) : \mu_1 - \mu_2 \in \left\{ \bar{x}_1 - \bar{x}_2 \pm z_{\alpha/2} \hat{S}_T \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right\}. \quad (8.10)$$

Ejemplo 3 Se quiere elegir entre dos tipos de material textil para construir sistemas de amarre. Para ello se mide la tensión de rotura de varias cintas de prueba utilizando dos tipos de material. Se toman 24 datos usando el material M1, obteniéndose $\bar{x}_1 = 87$ (kg/mm²) y $\hat{s}_1 = 2$. Se toman además 30 datos usando el material M2, obteniéndose $\bar{x}_2 = 75$ y $\hat{s}_2 = 2,3$. Se sabe que las tensiones de rotura se distribuyen como una normal. Se supondrá, además, que las varianzas de ambas poblaciones son iguales.

Si suponemos que las varianzas de ambas poblaciones son iguales, podemos obtener una estimación conjunta de la varianza, que será

$$\hat{s}_T^2 = \frac{23 \times 2^2 + 29 \times 2,1^2}{23 + 29} = 4,2 \Rightarrow \hat{s}_T = 2,06.$$

Si las poblaciones son normales, y puesto que las muestras no son muy grandes, se usará el intervalo (8.9), donde para un nivel de confianza del 95 % se tiene que

$$t_{n_1+n_2-2; \alpha/2} = t_{52; 0,025} = 2,0$$

y el intervalo del 95 % para la diferencia entre las medias de las tensiones de rotura será

$$\begin{aligned} IC(0,95) & : \mu_1 - \mu_2 \in \left\{ 87 - 75 \pm 2 \times 2,06 \times \sqrt{\frac{1}{24} + \frac{1}{30}} \right\} \\ & = (12 \pm 1,13). \end{aligned}$$

Por tanto, los datos muestran evidencia a favor del material M1, siendo el intervalo de confianza del 95 % muy estrecho en torno a las 12 unidades de ventaja respecto al material M2.

Caso 2: $\sigma_1^2 \neq \sigma_2^2$

Si las poblaciones que se comparan a través de dos muestras tienen varianzas diferentes, no podremos utilizar el estimador conjunto (8.7). En ese caso, ya no tenemos un resultado como (8.8). Existe un resultado análogo, pero es aproximado. Este resultado dice que si las poblaciones son normales, entonces, aproximadamente

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\hat{S}_1^2}{n_1} + \frac{\hat{S}_2^2}{n_2}}} \sim t_v$$

$$v = \frac{\left(\frac{\hat{S}_1^2}{n_1} + \frac{\hat{S}_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1} \left(\frac{\hat{S}_1^2}{n_1}\right) + \frac{1}{n_2 - 1} \left(\frac{\hat{S}_2^2}{n_2}\right)}$$

El valor que se obtenga de v no será en general un número entero, por lo que usaremos el entero más próximo. El intervalo de confianza para poblaciones normales será

$$IC(1 - \alpha) : \mu_1 - \mu_2 \in \left\{ (\bar{x}_1 - \bar{x}_2) \pm t_{v;\alpha/2} \sqrt{\frac{\hat{s}_1^2}{n_1} + \frac{\hat{s}_2^2}{n_2}} \right\} \quad (8.11)$$

En cualquier caso, **si los tamaños muestrales n_1 y n_2 son grandes**, siempre podremos utilizar el resultado asintótico

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\hat{S}_1^2}{n_1} + \frac{\hat{S}_2^2}{n_2}}} \sim N(0, 1), \quad (8.12)$$

que es válido incluso si las poblaciones no son normales, con tal que los tamaños muestrales sean mayores de, digamos, **30 datos**. El intervalo de confianza para muestras grandes y varianzas distintas será

$$IC(1 - \alpha) : \mu_1 - \mu_2 \in \left\{ (\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2} \sqrt{\frac{\hat{s}_1^2}{n_1} + \frac{\hat{s}_2^2}{n_2}} \right\}. \quad (8.13)$$

Ejemplo 4 Vamos ahora a calcular los intervalos de confianza del problema anterior, sobre la elección entre dos tipos de material textil para construir sistemas de amarre, sin utilizar la suposición de que las varianzas son iguales. Como los tamaños muestrales no son grandes, no podemos utilizar el intervalo (8.13). No obstante, como las poblaciones son normales podremos utilizar (8.11). El cálculo de los grados de libertad de la distribución t de Student es

$$v = \frac{\left(\frac{2^2}{24} + \frac{2,3^2}{30}\right)^2}{\frac{1}{24 - 1} \left(\frac{2^2}{24}\right)^2 + \frac{1}{30 - 1} \left(\frac{2,3^2}{30}\right)^2} \approx 52.$$

Por tanto, el intervalo de confianza del 95 % será, usando que $t_{52;0,025} = 2,0$,

$$\begin{aligned} IC(1 - \alpha) &: \mu_1 - \mu_2 \in \left\{ 87 - 75 \pm 2\sqrt{\frac{2^2}{24} + \frac{2,3^2}{30}} \right\} \\ &= 12 \pm 1,17, \end{aligned}$$

que es sólo ligeramente más ancho que el calculado anteriormente.

La siguiente tabla resume los intervalos de confianza para la diferencia de medias. Los resultados para muestras grandes son válidos para cualquier distribución. Los resultados para poblaciones normales son válidos para cualquier tamaño muestral.

Intervalos de confianza, $IC(1 - \alpha)$		
Varianzas	Muestras grandes	Poblaciones normales
$\sigma_1^2 = \sigma_2^2$	$\mu_1 - \mu_2 \in \left\{ \bar{x}_1 - \bar{x}_2 \pm z_{\alpha/2} \sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right\}$ $\mu_1 - \mu_2 \in \left\{ \bar{x}_1 - \bar{x}_2 \pm z_{\alpha/2} \hat{s}_T \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right\}$	$\mu_1 - \mu_2 \in \left\{ \bar{x}_1 - \bar{x}_2 \pm z_{\alpha/2} \sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right\}$ $\mu_1 - \mu_2 \in \left\{ \bar{x}_1 - \bar{x}_2 \pm t_{n_1+n_2-2; \alpha/2} \hat{s}_T \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right\}$
$\sigma_1^2 \neq \sigma_2^2$	$\mu_1 - \mu_2 \in \left\{ \bar{x}_1 - \bar{x}_2 \pm z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right\}$ $\mu_1 - \mu_2 \in \left\{ \bar{x}_1 - \bar{x}_2 \pm z_{\alpha/2} \sqrt{\frac{\hat{s}_1^2}{n_1} + \frac{\hat{s}_2^2}{n_2}} \right\}$	$\mu_1 - \mu_2 \in \left\{ \bar{x}_1 - \bar{x}_2 \pm z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right\}$ $\mu_1 - \mu_2 \in \left\{ \bar{x}_1 - \bar{x}_2 \pm t_{v; \alpha/2} \sqrt{\frac{\hat{s}_1^2}{n_1} + \frac{\hat{s}_2^2}{n_2}} \right\}$
$\hat{S}_T^2 = \frac{(n_1 - 1)\hat{S}_1^2 + (n_2 - 1)\hat{S}_2^2}{n_1 + n_2 - 2}$ $v = \frac{\left(\frac{\hat{S}_1^2}{n_1} + \frac{\hat{S}_2^2}{n_2} \right)^2}{\frac{1}{n_1 - 1} \left(\frac{\hat{S}_1^2}{n_1} \right) + \frac{1}{n_2 - 1} \left(\frac{\hat{S}_2^2}{n_2} \right)}$		

Estos intervalos de confianza se pueden utilizar para hacer el contraste $H_0 : \mu_1 = \mu_2$ frente a $H_1 : \mu_1 \neq \mu_2$ pues como ya dijimos hay una equivalencia entre el resultado de un contraste de nivel de significación α y un intervalo de nivel de confianza $(1 - \alpha)$. En la sección siguiente veremos este tipo de contrastes explícitamente.

8.2.3. Contrastes para la diferencia de medias

Estamos interesados en realizar contrastes sobre la diferencia de las medias poblacionales $\mu_1 - \mu_2$. Los contrastes se basan en los mismos argumentos que en temas anteriores. Para ello basta con considerar a $\mu_1 - \mu_2$ como un único parámetro. Podemos interpretar estos contrastes como si fuesen los contrastes de $\delta = \mu_1 - \mu_2$ y contrastaremos si δ es $=, \leq, \text{ ó } \geq$ cierto valor numérico, digamos δ_0 . Podemos escribir los contrastes como

1. $H_0 : \mu_1 - \mu_2 = \delta_0$ frente a $H_1 : \mu_1 - \mu_2 \neq \delta_0$,
2. $H_0 : \mu_1 - \mu_2 \geq \delta_0$ frente a $H_1 : \mu_1 - \mu_2 < \delta_0$
3. $H_0 : \mu_1 - \mu_2 \leq \delta_0$ frente a $H_1 : \mu_1 - \mu_2 > \delta_0$

Lo más frecuente será comparar si las medias son iguales o si una es mayor que la otra. En este tipo de comparaciones se tendrá que $\delta_0 = 0$, quedando mlos contrastes

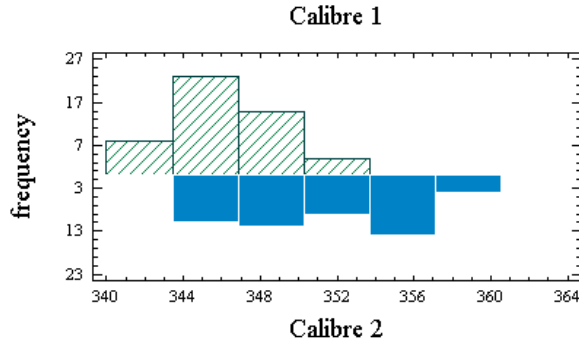
1. $H_0 : \mu_1 = \mu_2$ frente a $H_1 : \mu_1 \neq \mu_2$,
2. $H_0 : \mu_1 \geq \mu_2$ frente a $H_1 : \mu_1 < \mu_2$
3. $H_0 : \mu_1 \leq \mu_2$ frente a $H_1 : \mu_1 > \mu_2$

Por simplicidad, nos concentraremos únicamente en este tipo de contrastes. Los contrastes se basan en las mismas propiedades desarrolladas en la sección anterior, por lo que no las vamos a repetir aquí. Basándonos en las propiedades anteriores se ha construido la siguiente tabla, que resume los estadísticos de contraste a utilizar en cada caso así como las distribuciones de referencia y regiones de rechazo.

Varianzas	Contrastes	Estadísticos de contraste	Distribución de referencia	Región de rechazo
$\sigma_1^2 = \sigma_2^2$	(1)- $H_0 : \mu_1 = \mu_1$; $H_1 : \mu_1 \neq \mu_2$ (2)- $H_0 : \mu_1 \geq \mu_2$; $H_1 : \mu_1 < \mu_2$ (3)- $H_0 : \mu_1 \leq \mu_2$; $H_1 : \mu_1 > \mu_2$	(a) $Z_0 = \frac{X_1 - X_2}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ (b) $T_0 = \frac{\bar{X}_1 - \bar{X}_2}{\hat{S}_T \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$	Muestras grandes $Z_0, T_0 \sim N(0, 1)$ Normalidad (a) $Z_0 \sim N(0, 1)$ (b) $T_0 \sim t_{n_1+n_2-2}$	(1-a) $ z_0 > z_{\alpha/2}$ (2-a) $z_0 < -z_{\alpha}$ (3-a) $z_0 > z_{\alpha}$ (1-b) $ t_0 > z_{\alpha/2}$ ó $t_{n_1+n_2-2; \alpha/2}$ (2-b) $t_0 < -z_{\alpha}$ ó $-t_{n_1+n_2-2; \alpha}$ (3-b) $t_0 > z_{\alpha}$ ó $t_{n_1+n_2-2; \alpha}$
$\sigma_1^2 \neq \sigma_2^2$	(1)- $H_0 : \mu_1 = \mu_1$; $H_1 : \mu_1 \neq \mu_2$ (2)- $H_0 : \mu_1 \geq \mu_2$; $H_1 : \mu_1 < \mu_2$ (3)- $H_0 : \mu_1 \leq \mu_2$; $H_1 : \mu_1 > \mu_2$	(a) $Z_0 = \frac{X_1 - X_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ (b) $T_0 = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\hat{S}_1^2}{n_1} + \frac{\hat{S}_2^2}{n_2}}}$	Muestras grandes $Z_0, T_0 \sim N(0, 1)$ Normalidad (a) $Z_0 \sim N(0, 1)$ (b) $T_0 \sim t_v$	(1-a) $ z_0 > z_{\alpha/2}$ (2-a) $z_0 < -z_{\alpha}$ (3-a) $z_0 > z_{\alpha}$ (1-b) $ t_0 > t_{v; \alpha/2}$ ó $z_{\alpha/2}$ (2-b) $t_0 < -t_{v; \alpha}$ ó $-z_{\alpha}$ (3-b) $t_0 > t_{v; \alpha}$ ó z_{α}

Ejemplo 5 Se quiere comparar la precisión de dos calibres diferentes. Para ello se comparan las mediciones (en mm) realizadas en 100 clavos procedentes del mismo lote de fabricación (fichero longitudclavos.sf3). Se miden 50 clavos con un calibre y los otros 50 con otro calibre distinto. Los clavos se han asignado a cada subgrupo al azar, por lo que hemos de suponer que las medidas de un mismo calibre son una muestra de una misma población. Se desea saber si ambos calibres miden por término medio lo mismo. Aunque los clavos sean todos del mismo tipo serán todos distintos, siendo sus diferencias pequeñas y debidas a pequeñas desviaciones aleatorias del proceso de producción.

La figura siguiente muestra los histogramas de ambas mediciones.



Puede verse en los histogramas que puede haber diferencias significativas de un calibre al otro. Si ambos calibres fuesen idénticos esperaríamos que ambas distribuciones fuesen muy parecidas. Más concretamente, esperaríamos que la media poblacional de las mediciones del primer calibre μ_1 fuese igual a la media poblacional del segundo calibre μ_2 . Nuestro contraste será:

$$H_0 : \mu_1 = \mu_2; H_1 : \mu_1 \neq \mu_2$$

No supondremos que las varianzas son iguales, pues no tenemos ninguna razón a priori para hacer dicha suposición. No debemos confundir la varianza de las longitudes de los clavos en cada muestra con la varianza de las mediciones, que es lo que realmente tenemos. Al ser las dos muestras de clavos de la misma población de clavos, sus varianzas sólo diferirán por el azar de la muestra. Sin embargo, al medirse cada muestra con un calibre diferente, la varianza de las mediciones finales puede ser diferente si los calibres no tienen la misma precisión.

Por tanto, al considerarse varianzas diferentes, el estadístico de contraste será

$$T_0 = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\hat{S}_1^2}{n_1} + \frac{\hat{S}_2^2}{n_2}}}.$$

De los datos se obtiene que

$$\begin{aligned}\bar{x}_1 &= 346,16; \hat{s}_1^2 = 7,40 \\ \bar{x}_2 &= 351,12; \hat{s}_2^2 = 21,90\end{aligned}$$

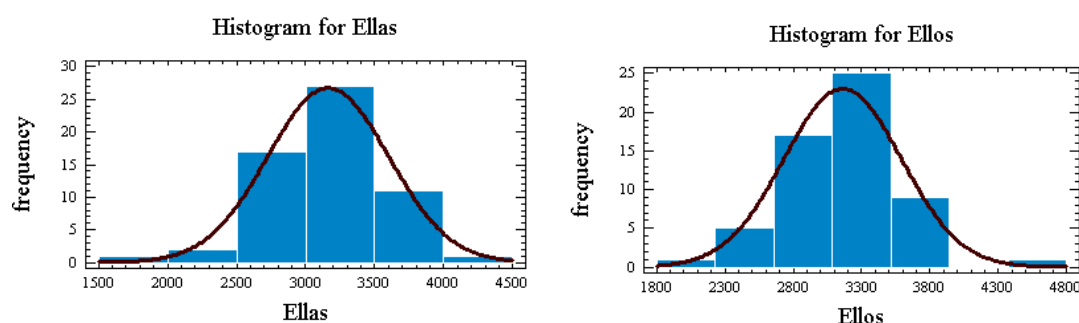
y por tanto

$$t_0 = \frac{346,16 - 351,12}{\sqrt{\frac{7,40}{50} + \frac{21,90}{50}}} = -6,48.$$

Para ver si dicho valor está en la zona de aceptación o rechazo debemos especificar un nivel de significación y decidir qué distribución de referencia es la más apropiada. Usaremos un nivel de significación $\alpha = 0,05$. Como tenemos un número de datos suficientemente grande en cada grupo, usaremos como distribución de referencia la $N(0, 1)$. Los valores críticos son $\pm z_{0,025} = \pm 1,96$ por lo que $t_0 = -6,48$ está en la región de rechazo, al ser $|t_0| > z_{0,025}$. El p-valor obtenido es (calculado con Statgraphics) $-2,67 \times 10^{-7}$ que es realmente pequeño. Por tanto debemos rechazar la hipótesis nula

de igualdad de medias. Aunque la diferencia entre las medias de ambos calibres no es muy grande ($351.12 - 346.16 = 4.96\text{mm}$), la diferencia entre las medias de ambos calibres es muy significativa. Hay una fuerte evidencia de que ambos calibres no miden, por término medio, lo mismo.

Ejemplo 6 El fichero *reciennacidos.sf3* contiene los pesos (en gramos) de las niñas (muestra 1) y niños (muestra 2) nacidos en el hospital San Pedro de Alcántara de Cáceres en la semana santa del año 2002. Supondremos a estos bebés como una muestra representativa de los bebés españoles. Con estos datos queremos saber si el peso medio de los niños es el mismo que de las niñas. A continuación se muestran los histogramas con la normal estimada.



Estos datos sugieren que ambas poblaciones serán normales. Los p -valores de los respectivos contrastes de normalidad basados en el test de la χ^2 son muy elevados, lo que indica que la hipótesis de normalidad es muy razonable. Las varianzas muestrales son también muy parecidas, por lo que supondremos que las varianzas poblacionales también lo son ($\sigma_1^2 = \sigma_2^2 = \sigma^2$). Más adelante veremos un procedimiento para contrastar esta suposición. Por tanto el estadístico de contraste es

$$T_0 = \frac{\bar{X}_1 - \bar{X}_2}{\hat{S}_T \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$\hat{S}_T^2 = \frac{(n_1 - 1)\hat{S}_1^2 + (n_2 - 1)\hat{S}_2^2}{n_1 + n_2 - 2}$$

Los datos muestran que

$$\begin{aligned} n_1 &= 59; \bar{x}_1 = 3163,9; \hat{s}_1^2 = 194004. \\ n_2 &= 58; \bar{x}_2 = 3162.; \hat{s}_2^2 = 185806. \\ \hat{s}_T^2 &= \frac{(59 - 1) \times 194004 + (58 - 1) \times 185806}{59 + 58 - 2} = 189940,6 \\ \hat{s}_T &= 435,82. \\ t_0 &= \frac{3163,9 - 3162}{435,82 \sqrt{\frac{1}{59} + \frac{1}{58}}} = 0,024 \end{aligned}$$

La distribución de referencia, al existir normalidad en las poblaciones, es la t_{115} . El contraste es bilateral por lo que el valor crítico es, usando $\alpha = 0,05$, (Statgraphics) $t_{115;0,025} = 1,98$. Por lo tanto t_0 está en la región de aceptación. El p -valor es 0.98 (Statgraphics), que es realmente alto. Puede

decirse entonces que no podemos rechazar la igualdad de medias en el peso de niños y niñas al nacer. Los tamaños muestrales de cada grupo son suficientemente grandes, por lo que el contraste también se podría haber hecho usando la $N(0,1)$ como distribución de referencia, y se obtendrían resultados similares (el valor crítico sería $z_{0,025} = 1,96$, que es muy similar a $t_{115;0,025} = 1,98$)

8.3. Comparación de dos medias usando datos emparejados

Los datos emparejados son pares de datos de los mismos elementos. Por ejemplo, datos de las mismas personas antes y después de ser sometidas a algún tratamiento, o rendimiento de un motor con el combustible A o el B. Tendremos así una muestra de n elementos sobre los que se toman datos en dos circunstancias distintas. Un ejemplo de datos apareados se puede ver en el fichero **calibretornillos.sf3**, que contiene las mediciones de la longitud de 95 tornillos del mismo tipo. Cada tornillo se ha medido dos veces, una vez con un calibre digital (muy preciso) y una segunda vez con un calibre analógico (menos preciso). Los 95 pares de puntos son por tanto de los mismos tornillos. A este tipo de datos que son dependientes por pares se les llama **datos emparejados, pareados, o apareados**. El objetivo es comprobar con la muestra si ambas mediciones son significativamente diferentes.

El análisis de datos emparejados es muy sencillo y consiste en construir una nueva variable que sea la diferencia entre las dos variables 'emparejadas'. Es decir, construimos

$$Y = X_a - X_d$$

donde X_a es la medición con el calibre analógico y X_d es la medición con el calibre digital. Sea

$$\mu_y = E(Y) = E(X_a - X_d) = E(X_a) - E(X_d) = \mu_a - \mu_d$$

la media poblacional de la diferencia de las mediciones analógica (μ_a) y digital (μ_d). La hipótesis que queremos contrastar es

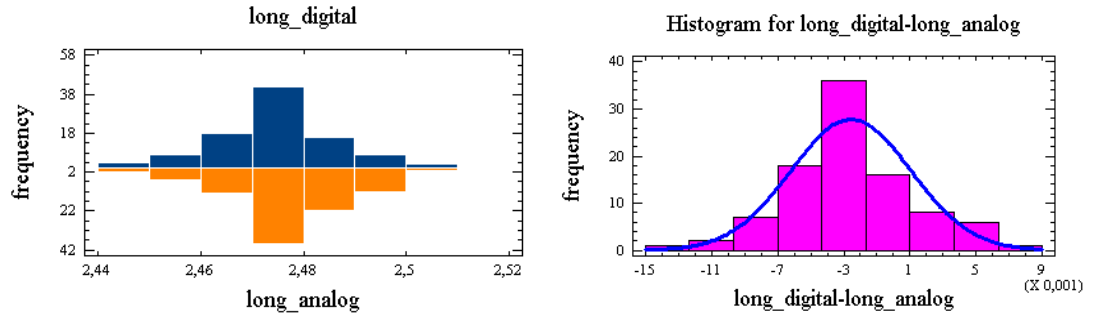
$$H_0 : \mu_y = 0$$

$$H_1 : \mu_y \neq 0.$$

Por lo tanto estamos ante un contraste para una media poblacional como los vistos en temas anteriores. El estadístico de contraste será

$$T_0 = \frac{\bar{y} - 0}{\hat{S}_y / \sqrt{n}},$$

donde \bar{y} es la media muestral de Y y \hat{S}_y es su covarianza. La figura siguiente muestra los histogramas de las 95 mediciones de cada calibre, así como el histograma de las 95 diferencias de ambas mediciones.



El test de normalidad de la Chi-cuadrado realizado sobre la diferencia tiene un p-valor mayor que 5 % para los datos del calibre digital, pero rechaza la normalidad para el calibre analógico y para la diferencia entre ambos. Por tanto, a la hora de analizar esta diferencia no asumiremos normalidad. El tamaño muestral es, no obstante, suficientemente grande para poder utilizar la $N(0, 1)$ como distribución de referencia para el estadístico de contraste.

Con los datos se obtiene

$$\begin{aligned}\bar{y} &= -0,00256; s_y^2 = 0,00364 \\ t_0 &= \frac{\bar{y} - \mu_0}{\hat{s}_y/\sqrt{n}} = \frac{-0,00256 - 0}{0,00364/\sqrt{95}} = -6,8549 \\ z_{0,025} &= 1,96\end{aligned}$$

Como $|t_0| = 6.8549 > 1.96$ se rechaza H_0 y ambos calibres no parecen medir lo mismo. Al menos uno no está calibrado (tal vez el analógico)

8.4. Comparación de dos proporciones

8.4.1. Intervalo de la diferencia de proporciones

La inferencia sobre la diferencia de dos proporciones poblacionales $p_1 - p_2$ se basa en la propiedad de que, si tenemos dos muestras **suficientemente grandes** n_1 y n_2 , en las que hemos contabilizado la proporción muestral de presencia de cierto atributo \hat{p}_1 y \hat{p}_2 , respectivamente, se tiene que

$$\hat{p}_1 \sim N\left(p_1, \frac{p_1 q_1}{n_1}\right); \hat{p}_2 \sim N\left(p_2, \frac{p_2 q_2}{n_2}\right),$$

y por lo tanto

$$\hat{p}_1 - \hat{p}_2 \sim N\left(p_1 - p_2, \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}\right). \quad (8.14)$$

Y estandarizando

$$\frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}} \sim N(0, 1) \quad (8.15)$$

De esta forma, un intervalo de confianza para la diferencia de medias $p_1 - p_2$ será

$$IC(1 - \alpha) : p_1 - p_2 \in \left\{ \hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}} \right\}.$$

8.4.2. Contrastes

Los contrastes en los que estamos interesados son:

1. $H_0 : p_1 = p_2$; frente a $H_1 : p_1 \neq p_2$
2. $H_0 : p_1 \geq p_2$; frente a $H_1 : p_1 < p_2$
3. $H_0 : p_1 \leq p_2$; frente a $H_1 : p_1 > p_2$

Los contrastes tendrán el mismo esquema que los contrastes de diferencia de medias. El estadístico de contraste se construye usando la propiedad (8.15). El estadístico de contraste se construye utilizando la expresión (8.15) anterior, pero suponiendo $p_1 = p_2 = p_0$. Como p_0 es un valor desconocido, tendremos que usar una estimación utilizando la información de ambas muestras. Tendremos entonces el siguiente estadístico de contraste

$$Z_0 = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}_0 \hat{q}_0 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

donde

$$\hat{p}_0 = \frac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n_1 + n_2}; \hat{q}_0 = 1 - \hat{p}_0.$$

La distribución de referencia es siempre la $N(0, 1)$. La siguiente tabla resume este contraste

Contrastes	Estadísticos de contraste	Distribución de referencia	Región de rechazo
(1)- $H_0 : p_1 = p_2; H_1 : p_1 \neq p_2$ (2)- $H_0 : p_1 \geq p_2; H_1 : p_1 < p_2$ (3)- $H_0 : p_1 \leq p_2; H_1 : p_1 > p_2$	$Z_0 = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}_0 \hat{q}_0 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$ con $\hat{p}_0 = \frac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n_1 + n_2}$	$N(0, 1)$	(1) $ z_0 > z_{\alpha/2}$ (2) $z_0 < -z_{\alpha}$ (3) $z_0 > z_{\alpha}$

Ejemplo 7 Se desea saber si el porcentaje de alumnos que aprueban todas las asignaturas de primer curso en primera convocatoria, en la carrera de Ingeniería Industrial, es el mismo en la Carlos III y en la Universidad Politécnica de Madrid. Para ello se analiza una muestra de 100 expedientes, elegidos al azar, de cada universidad. De los 100 expedientes seleccionados de la Carlos III, el número de alumnos que aprobaron todas las asignaturas de primero en primera convocatoria fue de 22 mientras que en la Politécnica fue de 18. ¿Se puede concluir que la probabilidad de aprobar es diferente en ambas universidades? (utiliza un nivel de significación de $\alpha = 0,05$) (sep 01)

Sea p_1 =porcentaje de alumnos aprobados en la Universidad Politécnica, y p_2 el porcentaje en la Carlos III. Sea además, \hat{p}_1 el porcentaje de alumnos aprobados en la Politécnica en una muestra de tamaño n_1 , y \hat{p}_2 y n_2 enm la Carlos III, entonces:

$$\frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}} \sim N(0, 1)$$

Por tanto el contraste

$$H_0 : p_1 = p_2$$

$$H_1 : p_1 \neq p_2$$

utilizará el estadístico de contraste

$$z_0 = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}_0 \hat{q}_0 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

y se rechazará H_0 si $z > z_{\alpha/2}$ o $z < -z_{\alpha/2}$. Con los datos del enunciado se tiene que

$$\begin{aligned}\hat{p}_0 &= \frac{18 + 22}{200} = 0,20 \\ z_0 &= \frac{(0,18 - 0,22)}{\sqrt{0,20 \times 0,80 \left(\frac{1}{100} + \frac{1}{100} \right)}} = -0,71\end{aligned}$$

Como $z_{0,025} = 1,96$ se tiene que $|z_0| < 1,96$ y por tanto la diferencia observada entre ambas universidades no es significativa.

8.5. Comparación de dos varianzas en poblaciones normales

8.5.1. Introducción

En esta sección estamos interesados en comparar las varianzas de dos poblaciones normales a través de las varianzas estimadas con sus muestras. **Esta comparación sólo se realizará para poblaciones normales, por lo que es importante confirmar este aspecto en los datos.** Los resultados que se van a exponer **son muy sensibles a esta hipótesis de normalidad**, de forma que sólo podremos estar seguros de su aplicación si estamos muy seguros de que los datos proceden de poblaciones normales. Esta sensibilidad a pequeñas desviaciones de la normalidad resta interés a este tipo de análisis de comparación de varianzas.

En muchos casos, el interés en realizar la comparación de varianzas es para poder realizar un contraste o intervalo de confianza para la igualdad de medias basado en un estimador común de la varianza, a través de la fórmula (8.7). La inferencia sobre diferencia de medias sólo precisa la normalidad si tenemos muestras pequeñas, y así aplicar la distribución t de Student, a través de la propiedad expresada en la fórmula (8.8). Sin embargo, **la propiedad (8.8) aún es válida aunque los datos se devíen un poco de la normalidad**. Aún más, **la utilización del estimador conjunto (8.7) sigue dando resultados satisfactorios incluso con pequeñas diferencias en las varianzas poblacionales σ_1^2 y σ_2^2** . Por esta razón, muchos autores recomiendan la decisión de aplicar (8.7) en la inferencia sobre la diferencia de medias mediante la comparación directa de las estimaciones \hat{s}_1 y \hat{s}_2 . Es muy popular la utilización de la siguiente regla empírica:

Si la mayor desviación típica estimada no es más del doble que la menor, los procedimientos de comparación de medias basados en (8.7) siguen siendo aproximadamente válidos.

Una vez realizadas estas consideraciones, pasemos a ver cómo es la inferencia de comparación de varianzas en poblaciones normales.

Sea una población $X_1 \sim N(\mu_1, \sigma_1^2)$ de la que se extrae una muestra de tamaño n_1 y se calcula la varianza muestral corregida \hat{S}_1^2 . Y sea una segunda población $X_2 \sim N(\mu_2, \sigma_2^2)$ de la que se extrae una muestra de tamaño n_2 y se calcula la varianza muestral corregida \hat{S}_2^2 . La comparación de las varianzas σ_1^2 y σ_2^2 a partir de los estimadores \hat{S}_1^2 y \hat{S}_2^2 se realiza a partir del siguiente ratio:

$$F = \frac{\hat{S}_1^2 / \sigma_1^2}{\hat{S}_2^2 / \sigma_2^2}. \quad (8.16)$$

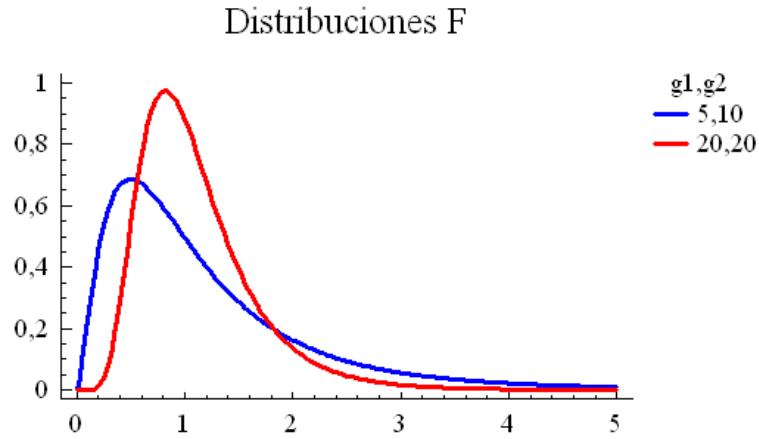
Este ratio F es una variable aleatoria, al ser la división de dos variables aleatorias \hat{S}_1^2 y \hat{S}_2^2 . El carácter aleatorio de F viene por tanto de los diferentes valores que podrían tomar \hat{S}_1^2 y \hat{S}_2^2 si tomásemos muestras diferentes de cada población. En poblaciones normales, la variable aleatoria F es conocida y se denomina distribución F de Fisher, o también distribución F de Snedecor, o simplemente distribución F . Es una distribución que está tabulada y puede encontrarse en la mayoría de los textos de estadística y en los programas de ordenador como el Statgraphics.

La distribución F de Fisher depende de dos parámetros g_1 y g_2 denominados grados de libertad. El valor de g_1 está relacionado con el numerador de (8.16) y toma el valor $g_1 = n_1 - 1$. Por esta razón, a g_1 también se le denomina **grados de libertad del numerador**. El valor de g_2 está relacionado con el denominador de (8.16) y toma el valor $g_2 = n_2 - 1$. Por esta razón, a g_2 también se le denomina **grados de libertad del denominador**. La notación de esta distribución F es F_{g_1, g_2} .

Por tanto, si $X_1 \sim N(\mu_1, \sigma_1^2)$ y $X_2 \sim N(\mu_2, \sigma_2^2)$, entonces

$$F = \frac{\hat{S}_1^2 / \sigma_1^2}{\hat{S}_2^2 / \sigma_2^2} \sim F_{n_1-1, n_2-1}. \quad (8.17)$$

La distribución F_{g_1, g_2} es una variable aleatoria continua no negativa, definida en $[0, \infty)$, y su forma varía dependiendo de los valores g_1 y g_2 . En general, es una distribución unimodal y asimétrica positiva, disminuyendo la asimetría al aumentar los valores de g_1 y g_2 . La siguiente figura muestra dos ejemplos de esta distribución: la $F_{5,10}$ y la $F_{20,20}$.

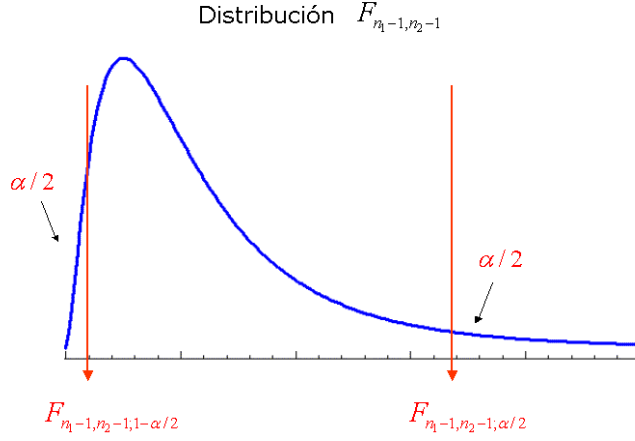


8.5.2. Intervalo de confianza del ratio de varianzas

A partir de (8.17) se puede deducir un intervalo de confianza para el ratio de varianzas σ_1^2 / σ_2^2 . De (8.17) se puede escribir que

$$P\left(F_{n_1-1, n_2-1; 1-\alpha/2} < \frac{\hat{S}_1^2 / \sigma_1^2}{\hat{S}_2^2 / \sigma_2^2} < F_{n_1-1, n_2-1; \alpha/2}\right) = 1 - \alpha, \quad (8.18)$$

donde $F_{n_1-1, n_2-1; 1-\alpha/2}$ y F_{n_1-1, n_2-1} son los valores de la distribución F_{n_1-1, n_2-1} que dejan a izquierda y derecha, respectivamente, un área $\alpha/2$, como ilustra la siguiente figura



Se puede demostrar que

$$F_{n_1-1, n_2-1; 1-\alpha/2} = \frac{1}{F_{n_2-1, n_1-1; \alpha/2}},$$

que es una relación útil cuando queremos buscar dichos valores en tablas, y la tabla sólo muestra valores situados en la cola de la derecha.

Operando en el interior de (8.18) se obtiene

$$P\left(\frac{\hat{S}_1^2}{\hat{S}_2^2} F_{n_1-1, n_2-1; 1-\alpha/2} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{\hat{S}_1^2}{\hat{S}_2^2} F_{n_1-1, n_2-1; \alpha/2}\right) = 1 - \alpha.$$

Por tanto, a partir de la información de una muestra, el intervalo de confianza de nivel $(1 - \alpha)$ del ratio de varianzas σ_1^2/σ_2^2 es

$$IC(1 - \alpha) : \frac{\sigma_1^2}{\sigma_2^2} \in \left\{ \frac{\hat{S}_1^2}{\hat{S}_2^2} F_{n_1-1, n_2-1; 1-\alpha/2}; \frac{\hat{S}_1^2}{\hat{S}_2^2} F_{n_1-1, n_2-1; \alpha/2} \right\}. \quad (8.19)$$

Ejemplo 8 Siguiendo con el Ejemplo 5, en el que se comparaban las mediciones de dos calibres sobre dos muestras independientes de 50 clavos cada una. Si hacemos el test de la chi-cuadrado en ambos grupos de datos, observamos que el p-valor es superior al 5 % en ambas muestras, por lo que asumiremos la hipótesis de normalidad es las poblaciones. Con 50 datos en cada muestra, la distribución F de Fisher será $F_{49,49}$. Para hacer un intervalo de confianza del 95 % tendremos que $F_{49,49;0,975} = 0,57$ y $F_{49,49;0,025} = 1,76$. El intervalo de confianza es

$$IC(0,95) : \frac{\sigma_1^2}{\sigma_2^2} \in \left\{ \frac{7,4}{21,9} 0,57; \frac{7,4}{21,9} 1,76 \right\} = \{0,193; 0,595\}.$$

Por tanto, la diferencia entre la variabilidad de las mediciones de ambos calibres es muy significativa. El intervalo está muy alejado del valor 1, que sería el caso de igualdad de varianzas. Los datos muestran así una gran evidencia a favor de la precisión del calibre 1, que con un 95 % de confianza tiene una variabilidad de entre un 20 % a un 60 % de la del calibre 2 (que es el analógico, de peor calidad que el digital)

8.5.3. Contrastes

Nuestras hipótesis son:

1. $H_0 : \sigma_1^2 = \sigma_2^2$; frente a $H_1 : \sigma_1^2 \neq \sigma_2^2$
2. $H_0 : \sigma_1^2 \geq \sigma_2^2$; frente a $H_1 : \sigma_1^2 < \sigma_2^2$
3. $H_0 : \sigma_1^2 \leq \sigma_2^2$; frente a $H_1 : \sigma_1^2 > \sigma_2^2$

El contraste se basa en la propiedad (8.17) expuesta anteriormente. De esta propiedad se deduce el estadístico de contraste, que será

$$F_0 = \frac{\hat{S}_1^2}{\hat{S}_2^2},$$

y denotaremos por f_0 al valor del estadístico F_0 obtenido en la muestra.

Al igual que sucedía con otros contrastes, es fácil comprobar que la región de rechazo está situada en la zona de la distribución de referencia que señala H_1 . La siguiente tabla muestra los detalles de estos contrastes.

Contrastes	Estadístico de contraste	Distribución de referencia	Región de rechazo
(1)- $H_0 : \sigma_1^2 = \sigma_2^2$; $H_1 : \sigma_1^2 \neq \sigma_2^2$; (2)- $H_0 : \sigma_1^2 \geq \sigma_2^2$; $H_1 : \sigma_1^2 < \sigma_2^2$; (3)- $H_0 : \sigma_1^2 \leq \sigma_2^2$; $H_1 : \sigma_1^2 > \sigma_2^2$	$F_0 = \frac{\hat{S}_1^2}{\hat{S}_2^2}$	$F_0 \sim F_{n_1-1; n_2-1}$	(1) $f_0 > F_{n_1-1; n_2-1; \alpha/2}$ ó $f_0 < F_{n_1-1; n_2-1; 1-\alpha/2}$ con $F_{n_1-1; n_2-1; 1-\alpha/2} = 1/F_{n_2-1; n_1-1; \alpha/2}$ (2) $f_0 < F_{n_1-1; n_2-1; 1-\alpha}$ con $F_{n_1-1; n_2-1; 1-\alpha} = 1/F_{n_2-1; n_1-1; \alpha}$ (3) $f_0 > F_{n_1-1; n_2-1; \alpha}$

Ejemplo 9 Siguiendo con el Ejemplo 6 en el que se mostraba el peso de 59 niños y 58 niñas recién nacidos. Se concluyó entonces que se podía asumir la normalidad de ambas poblaciones. Queremos contrastar la igualdad de varianzas. La hipótesis a contrastar es

$$H_0 : \sigma_V^2 = \sigma_M^2; \quad H_1 : \sigma_V^2 \neq \sigma_M^2$$

El valor del estadístico de contraste es

$$f_0 = \frac{194004}{185806} = 1,04.$$

La distribución de referencia es la $F_{58,57}$, y los dos valores críticos, con $\alpha = 0,05$, que delimitan la región de aceptación son $F_{58,57;0,975} = 0,59$ y $F_{58,57;0,025} = 1,69$. Como f_0 está entre ambos valores críticos, no podemos rechazar H_0 y asumimos que las varianzas de ambas poblaciones son normales.

Recuérdese que también habíamos deducido que las distribuciones eran normales (o, mejor dicho, que los test de la chi-cuadrado mostraban una evidencia suficiente). Como una distribución normal queda totalmente identificada con la media y la varianza, una vez que se concluye que dos normales tienen la misma media y la misma varianza, se concluye que las variables aleatorias son idénticas. Como anteriormente habíamos llegado a la conclusión de que las medias podían también asumirse iguales, llegamos entonces a la conclusión de que los pesos de los recién nacidos son distribuciones idénticas tanto en niños como en niñas.