Capítulo 7

Inferencia en una población normal

1. Inferencia en muestras pequeñas
2. Inferencia con la distribución t de Student
3. Inferencia sobre μ
4. Inferencia sobre σ^2

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

7.1. Inferencia en muestras pequeñas

En este tema estamos interesados en hacer inferencia sobre los parámetros de una variable aleatoria normal $X \sim N(\mu, \sigma^2)$. Es decir, sobre su media μ y sobre su varianza σ^2. En el tema anterior se introdujeron los elementos para realizar inferencia para la media μ de una población cualquiera **en muestras grandes**. El principio fundamental era que para una muestra $X_1, X_2, ..., X_n$ de una variable aleatoria X, la media muestral

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

tiene una distribución muestral que se aproxima asintóticamente a la normal (es decir, a mayor tamaño muestral n, mayor parecido a la normal) de la forma

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

La justificación de este resultado se encuentra en la aplicación del teorema central del límite. En la práctica, tamaños muestrales en torno a 30 observaciones pueden ser suficientes para que podamos realizar intervalos de confianza y contrastes sobre μ basados en la media muestral y su aproximación a la normal. Cuando el tamaño muestral es pequeño, el teorema central del límite ya no se cumple, y la distribución de la variable aleatoria $\bar{X}$ en el muestreo depende de la distribución de la variable X que estamos analizando. En estos casos, los intervalos de confianza que construyamos siguiendo la formulación del tema anterior ya no tendrán el nivel de confianza que deseamos, ni los contrastes tendrán el nivel de significación o el p-valor que nos salga en los cálculos, al estar basados en propiedades estadísticas que ya no se cumplen.

En este tema nos ocuparemos de la inferencia cuando la variable de interés X es normal, y que será de especial interés en muestras pequeñas. Como ya se ha mencionado en temas anteriores, **las variables aleatorias normales verifican que su combinación lineal produce variables aleatorias normales, para cualquier número de variables que combinemos**. De esta forma tenemos que, para cualquier tamaño muestral grande o pequeño, si $X \sim N(\mu, \sigma^2)$,

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right),$$

y por tanto

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1), \qquad (7.1)$$

para cualquier n. Cuando σ^2 es desconocida, ha de utilizarse un estimador. En este tema seguiremos utilizando como estimador de σ^2 el estimador insesgado

$$\hat{S}^2 = \frac{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2}{n-1}. \qquad (7.2)$$

Si sustituimos σ^2 por $\hat{S}^2$ en (7.1) obtenemos el estadístico T siguiente:

$$T = \frac{\bar{X} - \mu}{\hat{S}/\sqrt{n}}. \qquad (7.3)$$

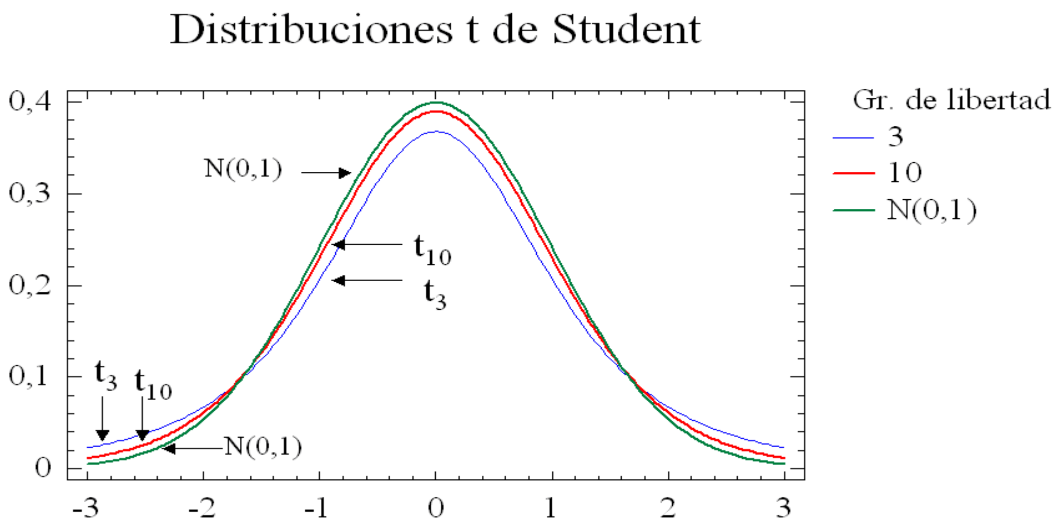
En el tema anterior, también acudíamos a este estadístico T para hacer inferencia. Hay una diferencia importante entre los estadísticos Z y T que hace que **en muestras pequeñas** sus propiedades estadísticas sean diferentes. En Z sólo interviene una variable aleatoria, que es la media muestral $\bar{X}$. Al ser $\bar{X}$ normal y estandarizarse con sus verdaderos parámetros, obtenemos que Z es la normal estándar. En T hay, sin embargo, dos variables aleatorias, $\bar{X}$ en el numerador y $\hat{S}$ en el denominador. Se puede demostrar que con muestras grandes, el componente aleatorio que aporta $\hat{S}$ en las propiedades estadísticas de T puede despreciarse. Por esta razón, en el tema anterior utilizamos que, **para muestras grandes**

$$T \sim N(0,1). \tag{7.4}$$

Para muestras pequeñas, la distribución muestral de T viene influenciada tanto por $\bar{X}$ como por $\hat{S}$, por lo que la aproximación a la normal que se usa en (7.4) será muy imprecisa. La distribución muestral de T cuando X es normal es conocida y se denomina distribución t de Student. En la siguiente sección se describe brevemente esta distribución.

7.2. Inferencia con la distribución t de Student

La distribución t de Student es una variable aleatoria continua, simétrica, de media cero, y de perfil muy parecido a la normal estándar. Depende de un parámetro g que se denomina grados de libertad. Su notación habitual es t_g. La figura siguiente muestra dos ejemplos de distribución t_g con $g = 3$ y $g = 10$ junto con la distribución $N(0,1)$.



En este gráfico puede verse que cuanto mayor es el número de grados de libertad, más parecido hay entre la distribución t_g y $N(0,1)$. Puede demostrarse que efectivamente la función de densidad de t_g tiende hacia la normal a medida que aumentan los grados de libertad. Para $g = \infty$ la distribución t_g es idéntica a la $N(0,1)$, pero a efectos prácticos, para $g > 30$ ambas distribuciones proporcionan probabilidades similares. La principal diferencia entre ambas distribuciones es que con pocos grados de libertad, la distribución t_g tiene la zona de las colas más ancha que la $N(0,1)$. **Esta diferencia es muy importante, pues en estas zonas de las colas donde nos interesará**

calcular probabilidades; tanto para la construcción de intervalos de confianza como en contrastes. Esta distribución está tabulada y puede encontrarse en la mayoría de los manuales de estadística.

Puede demostrarse que si $X \sim N(\mu, \sigma^2)$,

$$T = \frac{\bar{X} - \mu}{\hat{S}/\sqrt{n}} \sim t_{n-1}, \tag{7.5}$$

donde n es el tamaño de la muestra. **Para tamaños muestrales pequeños, tendremos g reducidos, y por tanto mayores diferencias entre t_{n-1} y $N(0,1)$. Será entonces más preciso utilizar la distribución t_{n-1} en aquellos lugares en los que al hacer inferencia para una población normal (intervalos y contrastes) usemos el estadístico T.** Para tamaños muestrales grandes será indiferente usar la distribución exacta t_{n-1} que se muestra en (7.5) que la aproximación asintótica a la $N(0,1)$ que se usaba en el tema anterior.

7.3. Inferencia sobre μ

7.3.1. Intervalos de confianza

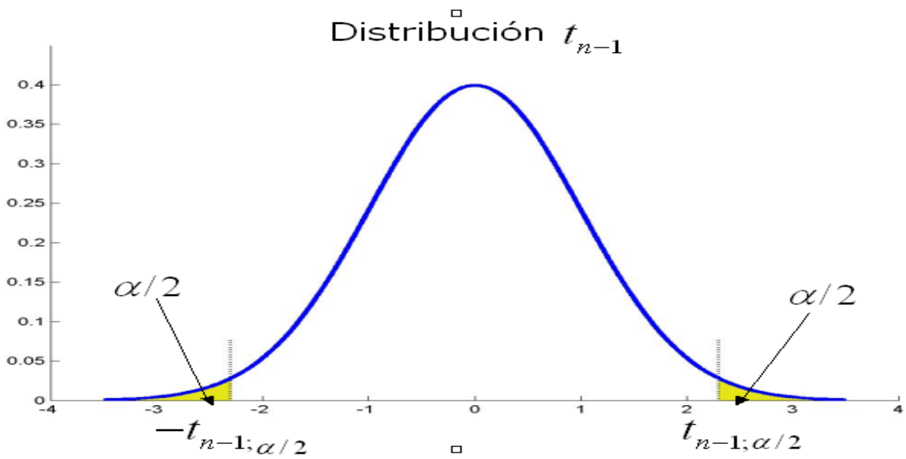
En el tema anterior se dedujo el intervalo de confianza para μ **para muestras grandes**, válido para cualquier distribución de X. Este intervalo, de nivel de confianza $(1-\alpha)$ es

$$\mu \in \left\{ \bar{x} \pm z_{\alpha/2} \frac{\hat{s}}{\sqrt{n}} \right\}. \tag{7.6}$$

En el caso $X \sim N(\mu, \sigma^2)$, un intervalo más preciso, sobre todo con muestras pequeñas, se obtiene reemplazando los valores de la normal estándar $z_{\alpha/2}$ por los de la distribución t_{n-1}. El razonamiento es el mismo que el que se siguió en el tema anterior. De (7.5) se tiene que

$$P\left(-t_{n-1;\alpha/2} < T < t_{n-1;\alpha/2}\right) = 1 - \alpha$$

donde $t_{n-1;\alpha/2}$ es el valor de la distribución t_{n-1} que deja el área $\alpha/2$ a la derecha, como se muestra en la siguiente figura



Por tanto, se tiene que

$$P\left(-t_{n-1;\alpha/2} < \frac{\bar{X}-\mu}{\hat{S}/\sqrt{n}} < t_{n-1;\alpha/2}\right) = 1-\alpha$$

y operando en el interior del paréntesis

$$P\left(\bar{X} - t_{n-1;\alpha/2}\frac{\hat{S}}{\sqrt{n}} < \mu < \bar{X} + t_{n-1;\alpha/2}\frac{\hat{S}}{\sqrt{n}}\right) = 1-\alpha.$$

Por lo tanto, un intervalo de nivel de confianza $(1-\alpha)$ para la media μ de una población normal a partir de la información que suministra una muestra de tamaño n es

$$IC(1-\alpha) : \mu \in \left\{\bar{x} \pm t_{n-1;\alpha/2}\frac{\hat{s}}{\sqrt{n}}\right\}. \tag{7.7}$$

En la práctica, si los datos proceden de una normal, deberemos utilizar siempre la distribución t_{n-1} en los intervalos. De esta forma aseguraremos que el nivel de confianza real es $(1-\alpha)$. Los intervalos de confianza en (7.6) son intervalos asintóticos, y en la práctica sólo podremos estar seguros de que el nivel de confianza real es $(1-\alpha)$ si el tamaño muestral es muy grande. Sin embargo, el intervalo (7.7) está hecho a la medida de una población normal y el tamaño muestral n, y por eso siempre tienen el nivel de confianza $(1-\alpha)$. Por esta razón se dice que los intervalos (7.7) son **exactos**. Para poder aplicar (7.7) debemos asegurarnos que nuestros datos se ajustan suficientemente a la normal. Para saber si los datos de la muestra proceden de una normal podemos hacer un histograma de los mismos o incluso algún test de bondad de ajuste como el test de la chi-cuadrado que se vio en temas anteriores.

Ejemplo 1 *En una explotación minera las rocas excavadas se someten a un análisis químico para determinar su contenido de Cadmio (densidad). Después de analizar 25 rocas se obtiene que* $\bar{x} = 9{,}77$ *y* $\hat{s} = 3{,}164$. *Se sabe de anteriores análisis que el contenido de Cadmio sigue una distribución normal. Se quiere construir un intervalo de confianza al 95 % para el contenido medio de Cadmio en las rocas de esa mina. El tamaño de la muestra* $n = 25$ *es muy pequeño para que los intervalos asintóticos (7.6) sean válidos. Al ser la variable de interés una normal, podemos emplear el intervalo exacto (7.7). El intervalo de nivel de confianza* $0{,}95$ *será*

$$\begin{aligned} 0{,}95 &= P\left\{-t_{n-1,\alpha/2} < \frac{\bar{X}-\mu_0}{\hat{S}/\sqrt{n}} < t_{n-1,\alpha/2}\right\} \\ &= P\left\{\bar{X} - t_{n-1,\alpha/2}\frac{\hat{S}}{\sqrt{n}} < \mu < \bar{X} + t_{n-1,\alpha/2}\frac{\hat{S}}{\sqrt{n}}\right\} \end{aligned}$$

Luego el intervalo es de la forma

$$IC(1-\alpha) : \mu \in \left\{\bar{x} \pm t_{n-1,\alpha/2}\frac{\hat{s}}{\sqrt{n}}\right\}.$$

Usando $\alpha = 0{,}05$ *tenemos que, según las tablas de la* t_{24}, $t_{24,0,025} = 2{,}06$. *El intervalo para el contenido medio de cadmio de las rocas que se extraigan de la mina es*

$$IC(0{,}95) : \mu \in \left\{9{,}77 \pm 2{,}06\frac{3{,}164}{\sqrt{25}}\right\} = (8{,}5, 11{,}1).$$

A la vista de este resultado, los técniccos de la mina pueden tomar una decisión acerca de la conveniencia de seguir haciendo prospección en dicha mina o por el contrario deben descartar su explotación.

7.3.2. Contrastes de hipótesis

Se quiere contrastar alguna de las siguientes hipótesis:

1. $\mathrm{H}_0 : \mu = \mu_0$; frente a $\mathrm{H}_1 : \mu \neq \mu_0$,
2. $\mathrm{H}_0 : \mu \geq \mu_0$; frente a $\mathrm{H}_1 : \mu < \mu_0$,
3. $\mathrm{H}_0 : \mu \leq \mu_0$; frente a $\mathrm{H}_1 : \mu > \mu_0$.

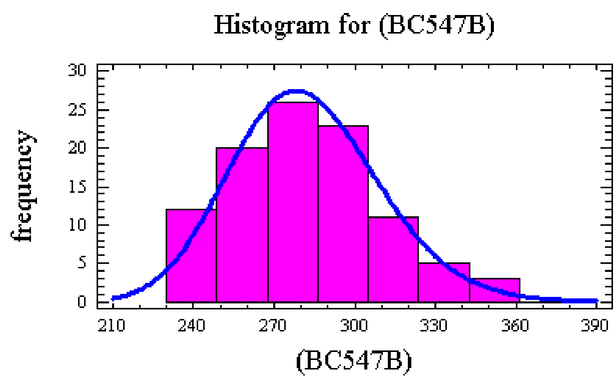
La forma de realizar los contrastes para poblaciones normales es la misma que la que se mencionó en el tema anterior, con la única diferencia de la distribución de referencia del estadístico T, que de acuerdo con (7.5) será la distribución t_{n-1} en lugar de la aproximación a la normal estándar. La siguiente tabla resume los detalles de estos contrastes.

Contrastes	Estadísticos de contraste	Distribución de referencia	Región de rechazo
(1)-$\mathrm{H}_0 : \mu = \mu_0$; $\mathrm{H}_1 : \mu \neq \mu_0$ (2)-$\mathrm{H}_0 : \mu \geq \mu_0$; $\mathrm{H}_1 : \mu < \mu_0$ (3)-$\mathrm{H}_0 : \mu \leq \mu_0$; $\mathrm{H}_1 : \mu > \mu_0$	$(a)\ Z_0 = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$ $(b)\ T_0 = \dfrac{\bar{X} - \mu_0}{\hat{S}/\sqrt{n}}$	$(a)\ Z_0 \sim N(0,1)$ $(b)\ T_0 \sim t_{n-1}$	(1-a) $\lvert z_0 \rvert > z_{\alpha/2}$ (2-a) $z_0 < -z_\alpha$ (3-a) $z_0 > z_\alpha$ (1-b) $\lvert t_0 \rvert > t_{n-1;\alpha/2}$ (2-b) $t_0 < -t_{n-1;\alpha}$ (3-b) $t_0 > t_{n-1;\alpha}$

Ejemplo 2 *Con los datos de la muestra de transistores BC547B mencionados en el tema anterior deseamos contrastar si se mantiene el valor nominal* $\mu = 290$ *como media de la distribucion poblacional de valores* β*, es decir,*

$$\begin{aligned} H_0 &: \mu = 290 \\ H_1 &: \mu \neq 290 \end{aligned}$$

Para hacer el contraste se toma una muestra de $n = 100$ *observaciones y se obtiene la media muestral* $\bar{x}$ *y la cuasivarianza* $\hat{s}^2$*. El histograma de este conjunto de datos junto con la normal* $N(\bar{x}, \hat{s}^2)$ *sobreimpresa es el siguiente*



Esta figura sugiere que loss datos podrían proceder de una distribución normal. El p-valor del contraste chi-cuadrado es mayor que 5 %, lo que refuerza la bondad del ajuste de la normal a nuestros datos. Consideramos entonces, con un p-valor<0.05, que es aceptable la normalidad de la población de valores β. Por tanto realizaremos el contraste usando como distribución de referencia para el estadístico de contraste la distribución t_{n-1}. Los datos muestran que

$$\begin{aligned} \bar{x} &= 282{,}3; \hat{s} = 27{,}57; \\ t_0 &= \frac{\bar{x} - \mu_0}{\hat{s}/\sqrt{n}} = \frac{282{,}3 - 290}{27{,}69/10} = -2{,}78. \end{aligned}$$

Como es un contraste bilateral necesitamos dos valores críticos. Como la distribución de referencia t_{n-1} es simétrica de media cero, ambos valores críticos serán iguales pero de signo contrario. Uusando un nivel de significación $\alpha = 0{,}05$, y la distribución de referencia t de Student con $n - 1 = 99$ grados de libertad se tiene que $t_{99;0{,}025} = 1{,}984$. Por tanto, como $|t_0| = 2{,}78 > 1{,}984$ rechazamos H_0. Rechazamos, con un nivel de significación del 5 %, que la ganancia media de los transistores se siga manteniendo en elvalor 290.

7.4. Inferencia sobre σ^2

7.4.1. Estimación

En esta sección simplemente recordaremos que hemos visto dos estimadores para σ^2 : la varianza muestral

$$S^2 = \frac{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2}{n},$$

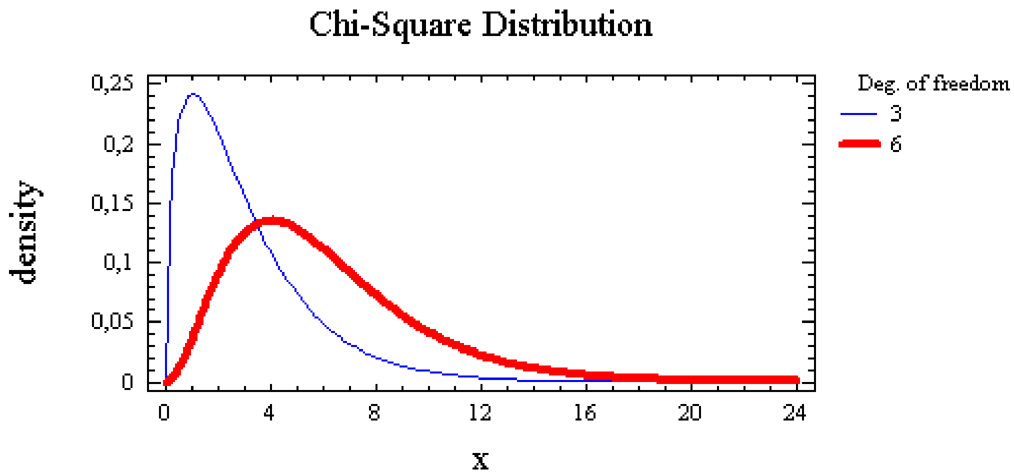
que es un estimadorde σ^2 sesgado, y la cuasivarianza

$$\hat{S}^2 = \frac{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2}{n-1},$$

que es insesgado. Para poblaciones normales, la distribución muestral de ambos estimadores está relacionada con la distribución llamada chi-cuadrado. Utilizaremos esa distribución para poder hacer intervalos de confianza y contrastes sobre σ^2. A continuación vamos a describir brevemente esta distribución.

7.4.2. La distribución χ_g^2

Antes de presentar resultados sobre la inferencia relacionada con σ^2 en poblaciones normales, presentaremos una variable aleatoria denominada chi-cuadrado, y que se denota por χ_g^2. La distribución chi-cuadrado es una distribución que depende del parámetro g que se denomina grados de libertad ($g = 1, 2, ...$). La distribución χ_g^2 va de 0 a ∞ y es asimétrica positiva. Su asimetría disminuye al aumentar los grados de libertad. La figura siguiente muestra la función de densidad de la χ_3^2 y la χ_6^2.



Las medidas características de la χ_g^2 son

$$\begin{aligned} E(\chi_g^2) &= g, \\ \mathrm{Var}(\chi_g^2) &= 2g. \end{aligned}$$

Esta distribución está también tabulada y puede encontrarse en la mayoría de los textos de estadística.

La distribución muestral de los estimadores de σ^2, la varianza y la cuasivarianza muestral, en poblaciones normales están relacionadas con esta distribución. Puede demostrarse que

$$\begin{aligned} \frac{(n-1)\hat{S}^2}{\sigma^2} &\sim \chi_{n-1}^2; \\ \frac{nS^2}{\sigma^2} &\sim \chi_{n-1}^2. \end{aligned} \tag{7.8}$$

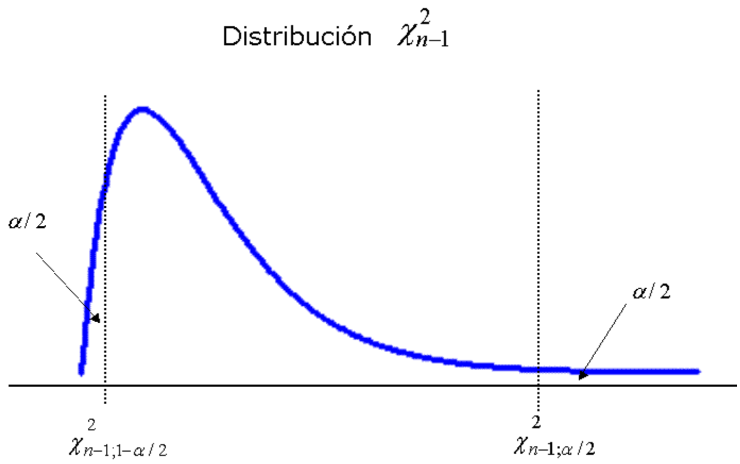
donde n es el tamaño de la muestra.

7.4.3. Intervalos de confianza para σ^2

Para construir los intervalos de confianza para σ^2 en una población normal vamos a seguir el mismo razonamiento que el utilizado para deducir los intervalos de μ. De (7.8) puede deducirse que

$$P\left(\chi_{n-1;1-\alpha/2}^2 < \frac{(n-1)\hat{S}^2}{\sigma^2} < \chi_{n-1;\alpha/2}^2\right) = 1-\alpha. \tag{7.9}$$

donde $\chi^2_{n-1;\alpha/2}$ es el valor de la distribución χ^2_{n-1} que deja el área $\alpha/2$ a la derecha. La figura siguiente ilustra estos valores $\chi^2_{n-1;\alpha/2}$ y $\chi^2_{n-1;1-\alpha/2}$.



Operando en el interior del paréntesis de (7.9) se obtiene que

$$P\left(\frac{(n-1)\hat{s}^2}{\chi^2_{n-1;\alpha/2}} < \sigma^2 < \frac{(n-1)\hat{s}^2}{\chi^2_{n-1;1-\alpha/2}}\right),$$

o bien, para el caso del estimador S^2,

$$P\left(\frac{nS^2}{\chi^2_{n-1;\alpha/2}} < \sigma^2 < \frac{nS^2}{\chi^2_{n-1;1-\alpha/2}}\right).$$

Por tanto, un intervalo de confianza de nivel de confianza $(1-\alpha)$ para el parámetro σ^2 será

$$IC(1-\alpha): \sigma^2 \in \left(\frac{(n-1)\hat{s}^2}{\chi^2_{n-1;\alpha/2}}, \frac{(n-1)\hat{s}^2}{\chi^2_{n-1;1-\alpha/2}}\right); \quad (7.10)$$

o bien, si utilizamos el estimador S^2,

$$IC(1-\alpha): \sigma^2 \in \left(\frac{ns^2}{\chi^2_{n-1;\alpha/2}}, \frac{ns^2}{\chi^2_{n-1;1-\alpha/2}}\right). \quad (7.11)$$

A diferencia de los intervalos de confianza para μ, los intervalos para σ^2 no son simétricos alrededor de las estimaciones $\hat{s}^2$ o s^2.

Ejemplo 3 *Continuando con el ejemplo 1 anterior sobre el contenido de cadmio en rocas, queremos construir un intervalo de confianza al 99 % para σ^2. Como el estimador utilizado para σ^2 es* *la cuasivarianza muestral $\hat{S}^2$, se tiene que*

$$0{,}99 = P\left\{\chi^2_{n-1,\alpha/2} < \frac{(n-1)\hat{S}^2}{\sigma^2} < \chi^2_{n-1,1-\alpha/2}\right\} = P\left\{\frac{(n-1)\hat{S}^2}{\chi^2_{n-1,1-\alpha/2}} < \sigma^2 < \frac{(n-1)\hat{S}^2}{\chi^2_{n-1,\alpha/2}}\right\}$$

Como $\alpha = 0{,}01$, tenemos que según las tablas de la chi cuadrado: $\chi^2_{24,0,995} = 9{,}89$, $\chi^2_{24,0,005} = 45{,}6$. El intervalo es:

$$IC\,(0{,}99) : \sigma^2 \in (5{,}27, 24{,}29)$$

7.4.4. Contraste de hipótesis

Se quiere contrastar las siguientes hipótesis.

1. $\mathrm{H}_0 : \sigma^2 = \sigma_0^2$; $\mathrm{H}_1 : \sigma^2 \neq \sigma_0^2$
2. $\mathrm{H}_0 : \sigma^2 \geq \sigma_0^2$; $\mathrm{H}_1 : \sigma^2 < \sigma_0^2$
3. $\mathrm{H}_0 : \sigma^2 \leq \sigma_0^2$; $\mathrm{H}_1 : \sigma^2 > \sigma_0^2$

donde σ_0^2 es un valor numérico concreto. Los contrastes para σ^2 en poblaciones normales siguen las mismas reglas que en los contrastes vistos para μ. El contraste se basa en las siguientes propiedades de la varianza muestral en poblaciones normales $X \sim N(\mu, \sigma^2)$ que ya se han mencionado anteriormente:

$$\frac{(n-1)\hat{S}^2}{\sigma^2} \sim \chi^2_{n-1}, \tag{7.12a}$$

$$\frac{nS^2}{\sigma^2} \sim \chi^2_{n-1}. \tag{7.12b}$$

El estadístico de contraste que resuma la información necesaria para realizar un contraste se basará en (7.12), pero sustituyendo σ^2 por σ_0^2. El estadístico de contraste es por tanto:

$$X_0^2 = \frac{(n-1)\hat{S}^2}{\sigma_0^2}; \tag{7.13}$$

$$X_0^2 = \frac{nS^2}{\sigma_0^2}. \tag{7.14}$$

Ejemplo 4 *Volviendo a los datos sobre los transistores BC547B mencionados anteriormente, teníamos el objetivo de comprobar si la media no había cambiado, así como comprobar si la varianza no había aumentado. Podemos ahora contrastar este segundo punto. Los datos históricos decían que $\sigma_0^2 = 760$. Por tanto el contraste es*

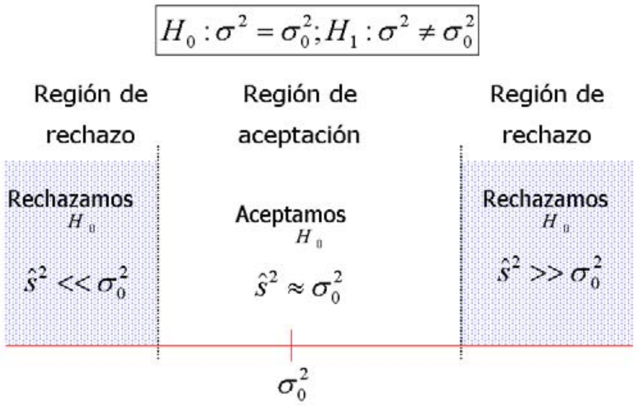
$$H_0 : \sigma^2 \leq 760; H_1 : \sigma^2 > 760.$$

Asumiendo a la vista del test de la chi-cuadrado que los datos son normales podemos realizar el contraste presentado más arriba. De los datos se obtiene

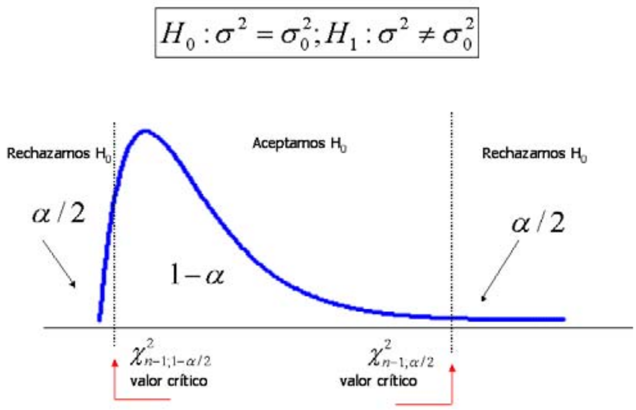
$$x_0^2 = \frac{(n-1)\hat{s}^2}{\sigma_0^2} = \frac{99 \times 766{,}85}{760} = 99{,}89.$$

Al realizar un contraste de hipótesis, aceptaremos la hipótesis nula salvo que los datos arrojen mucha evidencia en contra. Por tanto, rechazaremos la hipótesis nula cuando el valor del estimador de σ^2 que usemos haga lo que especifique la hipótesis alternativa de forma muy acusada.

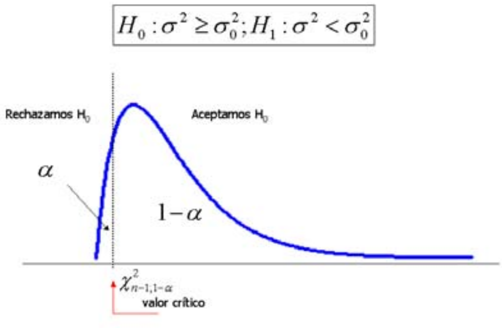
En el caso del contraste con alternativa bilateral $\mathrm{H}_0 : \sigma^2 = \sigma_0^2$; $\mathrm{H}_1 : \sigma^2 \neq \sigma_0^2$, rechazaremos H_0 cuando $\hat{s}^2$ (o s^2) tenga un valor $\hat{s}^2 >> \sigma_0^2$ o $\hat{s}^2 << \sigma_0^2$, como se ilustra en la siguiente figura

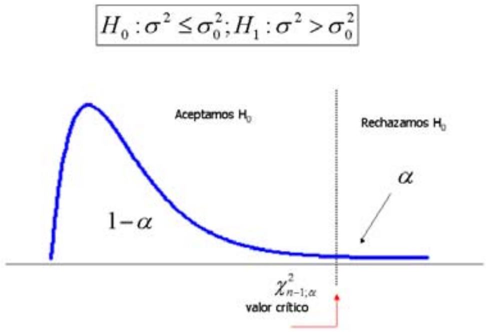


Puede verse en (7.13) que cuando $\hat{s}^2 >> \sigma_0^2$, el estadístico de contraste X_0^2 tendrá también un valor alto, y tenderá a estar en la cola de la derecha de la distribución de referencia, mientras que cuando $\hat{s}^2 << \sigma_0^2$, el estadístico X_0^2 estará en la zona de la izquierda de la distribución de referencia. La región de rechazo, de área igual al nivel de significación α, estará a ambos extremos de la distribución χ^2_{n-1}, como se ilustra en la siguiente figura.



Análogamente, en el caso de un contraste con alternativa unilateral, la región de rechazo estará sólo a un lado de la distribución. En el caso del contraste $\mathrm{H}_0 : \sigma^2 \geq \sigma_0^2$; $\mathrm{H}_1 : \sigma^2 < \sigma_0^2$, rechazaremos cuando $\hat{s}^2 << \sigma_0^2$, o análogamente, cuando X_0^2 tenga un valor muy bajo. Finalmente, en el caso del contraste $\mathrm{H}_0 : \sigma^2 \leq \sigma_0^2$; $\mathrm{H}_1 : \sigma^2 > \sigma_0^2$, rechazaremos cuando $\hat{s}^2 >> \sigma_0^2$, lo que dará un valor del estadístico de contraste en la cola de la derecha de la distribución χ^2_{n-1}. Puede verse por tanto, que la región de rechazo está allá donde señala H_1. Las siguientes figuras muestran las regiones de rechazo en estos dos contrastes.





La siguiente tabla resume las características de estos contrastes

Contrastes	Estadísticos de contraste	Distribución de referencia	Región de rechazo
(1)-$H_0 : \sigma^2 = \sigma_0^2$; $H_1 : \sigma^2 \neq \sigma_0^2$ (2)-$H_0 : \sigma^2 \geq \sigma_0^2$; $H_1 : \sigma^2 < \sigma_0^2$ (3)-$H_0 : \sigma^2 \leq \sigma_0^2$; $H_1 : \sigma^2 > \sigma_0^2$	$X_0^2 = \dfrac{(n-1)\hat{S}^2}{\sigma_0^2}$ $X_0^2 = \dfrac{nS^2}{\sigma_0^2}$	$X_0^2 \sim \chi^2_{n-1}$	(1) $x_0^2 > \chi^2_{n-1;\alpha/2}$ ó $x_0^2 < \chi^2_{n-1;1-\alpha/2}$ (2) $x_0^2 < \chi^2_{n-1;1-\alpha}$ (3) $x_0^2 > \chi^2_{n-1;\alpha}$

Ejemplo 5 *Volviendo a los datos sobre los transistores BC547B, rechazaremos H_0 si $x_0^2 > \chi^2_{99;0,05}$. Como $\chi^2_{99;0,05} = 123{,}2$ y $x_0^2 = 99{,}89$, no podemos rechazar la hipótesis nula, con un nivel de significación de $\alpha = 0{,}05$, de que el proceso no ha aumentado su variabilidad. Por tanto, aunque $\hat{s}^2 = 766{,}85 > \sigma_0^2$, la diferencia no es significativa, y es es perfectamente explicable por la variabilidad debida a la muestra. El p-valor de este contraste será la probabilidad $P(\chi^2_{99} > 99{,}89) = 0{,}456$ que es muy elevada, por lo que x_0^2 está bastante dentro de la región de aceptación. Aceptamos la hipótesis nula con bastante seguridad.*

Ejemplo 6 *Un fabricante de aparatos de precisión garantiza que la desviación típica de las medidas que pueden efectuarse con el tipo de balanza que comercializa es $\sigma \leq 5$ unidades. Para comprobar dicha afirmación se pesa un objeto en 100 básculas de dicho tipo y se obtiene una varianza muestral de s^2 =26.243. Si sabemos que la distribucion de los pesos siguen una normal, realizar un contraste que permita tomar una decisión respecto a aceptar o no la información suministrada por el fabricante ($\alpha = 0{,}05$).*

Lo que queremos contrastar es

$$H_0 : \sigma^2 \leq \sigma_0^2 (= 25); H_1 : \sigma^2 > 25.$$

Se rechaza H_0 si $s^2 >> \sigma_0^2$; más concretamente, si

$$\frac{ns^2}{\sigma_0^2} > \chi^2_{n-1;\alpha}$$

De los datos se tiene que $n = 100$, $s^2 = 26{,}243$. Por tanto

$$\begin{aligned} x_0^2 &= \frac{ns^2}{\sigma_0^2} = \frac{100 \times 26{,}243}{25} = 104{,}972, \\ \chi^2_{99,0,05} &= 123{,}2 \text{ (\textit{Statgraphics})}. \end{aligned}$$

Como $X_0 < \chi^2_{99,0,05}$ *no tenemos evidencia suficiente (con un nivel del 5 %) para sospechar del fabricante. Es decir, si la población tiene* $\sigma^2 \leq 25$ *no es raro encontrar que en una muestra de tamaño* $n = 100$ *tengamos* $s^2 = 26{,}243$. *Entra dentro de la variabilidad muestral que se encuentra por azar al tener muestras de tamaño 100.*