Estadística

Tema 6. Contraste de hipótesis

María Dolores Frías Domínguez
Jesús Fernández Fernández
Carmen María Sordo

Departamento de Matemática Aplicada y Ciencias de la Computación

TEMA 6: Contraste de hipótesis

Introducción al contraste de hipótesis

Contrastes con una muestra:

- Contraste de una proporción
- Contraste de una media
- Contraste de una varianza

Contrastes con dos muestras:

- Contraste de dos proporciones
- Contraste de dos medias

Contraste de hipótesis

El objetivo del contraste de hipótesis es decidir si una determinada hipótesis o conjetura sobre la **distribución poblacional** estudiada es confirmada o invalidada estadísticamente a partir de las **observaciones de una muestra**, es decir, avalar o rechazar tales informaciones sobre la característica de la población, pero no estimarla.

Ejemplo

La proporción de mujeres en Madrid toma un valor determinado:

$$P = 50.58\%$$

Contraste de hipótesis

El planteamiento general de un problema de contraste es el siguiente:

- Se formula una hipótesis o conjetura acerca de la **población**
- Se trata de ver si esa afirmación se encuentra apoyada por la evidencia experimental que se obtiene a través de una muestra aleatoria.

Una **hipótesis estadística** es una afirmación que se hace con respecto a una o más características desconocidas de una población de interés.

Ejemplo

Se desea contrastar que la proporción de mujeres en Madrid toma un valor determinado:

$$H_0: P = 50.58\%$$

Contraste de hipótesis

La realización de un contraste implica la existencia de dos hipótesis:

La **hipótesis nula H_0** es la que se formula y se quiere contrastar. Es la que el investigador asume como correcta y que no necesita ser probada, es decir, la aceptación de H_0 no implica que ésta sea correcta o que haya sido probada, sino que los datos no han proporcionado evidencia suficiente como para rechazarla.

La **hipótesis alternativa H_1** es la hipótesis opuesta de H_0, de forma que si a partir de la muestra se rechaza H_0 entonces se acepta como cierta H_1.

Ejemplo

Se desea contrastar que la proporción de mujeres en Madrid toma un valor determinado:

$$H_0: P = 50.58\%$$

$$H_1: P \neq 50.58\%$$

Contraste de hipótesis

Las afirmaciones no son todas del mismo tipo, pueden involucrar el valor numérico de algún parámetro, suponiendo la distribución conocida (generalmente la Normal), o la forma funcional no conocida de la distribución de interés a partir de la cual se obtiene la muestra .

1. $H_0 : P = 0.5$
2. $H_0 : \mu = 1.68$
3. $H_0 : F \sim \text{Normal}$

Contraste paramétrico (1, 2)

Contraste no paramétrico (3)

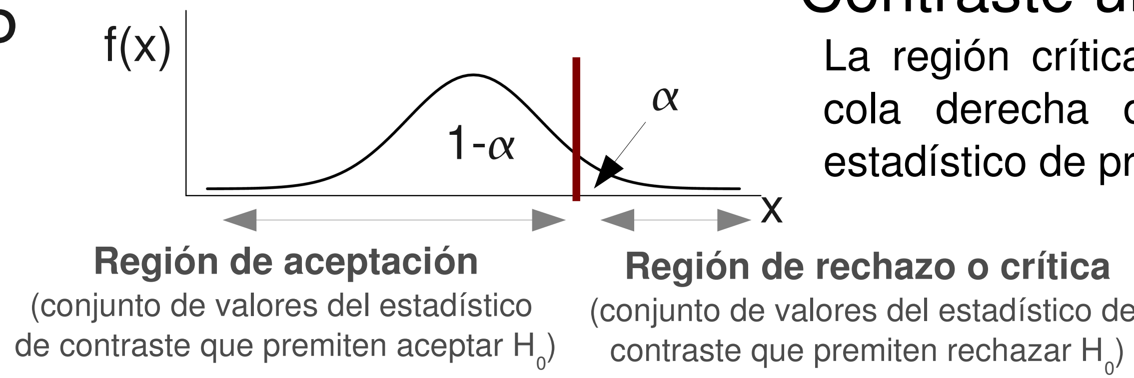
Contraste de hipótesis

Contrastes paramétricos:

Si: $H_0 : \theta = 0.5$,

entonces H_1 puede ser:

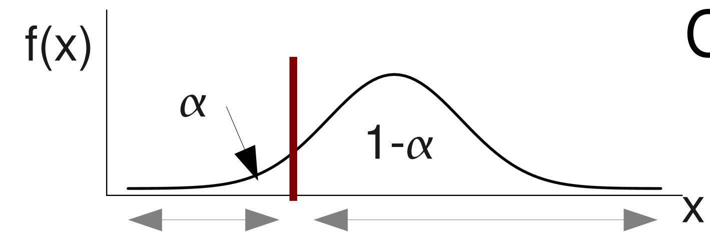
$H_1 : \theta > 0.5$



Contraste unilateral derecho:

La región crítica cae totalmente en la cola derecha de la distribución del estadístico de prueba.

$H_1 : \theta < 0.5$



Contraste unilateral izquierdo

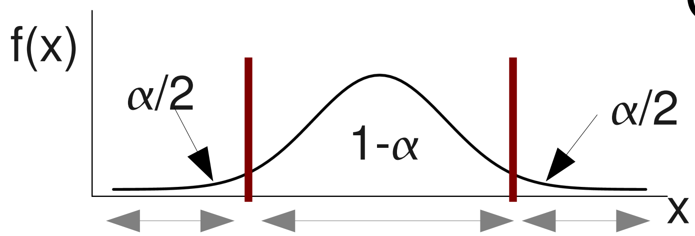
Contraste de hipótesis

Contrastes paramétricos:

Si: $H_0 : \theta = 0.5$,

entonces H_1 puede ser:

$H_1 : \theta \neq 0.5$



Contraste bilateral

El contraste bilateral es equivalente a calcular un **intervalo de confianza** (a un nivel de confianza 1-α) **de θ** y **rechazar H_0** (a un nivel de confianza 1-α) si **θ_0 no está dentro del intervalo de confianza** y aceptarla en caso contrario

Contraste de hipótesis

Hipótesis nula H_0

- La que contrastamos
- Característica de la **población** que asumimos como cierta
- Los datos pueden refutarla
- No debería ser rechazada sin una buena razón.

Hipótesis alternativa H_1

- Característica de la **población** que niega a H_0 (y creemos que es 'mejor').
- Los datos de la muestra pueden mostrar evidencia a favor
- No debería ser aceptada sin una gran evidencia a favor.

Ejemplo

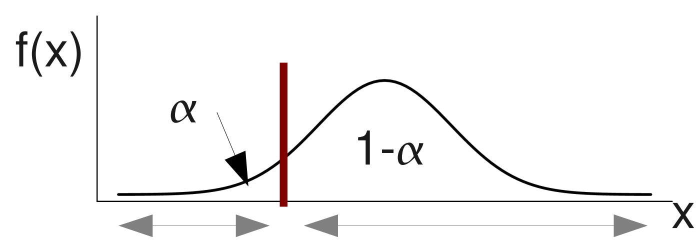
Lo que el investigador desea demostrar es la hipótesis alternativa:

- unas plantas tratadas con hormonas crecen más que el grupo control.
- un nuevo medicamento muestra mejores resultados que el que se prescribe actualmente.

Contraste de hipótesis

En el planteamiento de un contraste de hipótesis debemos realizar los siguientes pasos:

1. Definir las hipótesis a contrastar acerca de la población.
2. Definir un "estadístico de contraste" entre la muestra y H_0, que servirá para definir la acción a emprender (aceptar o rechazar la hipótesis nula). Esta medida debe tener una distribución conocida cuando H_0 es cierta.
3. Tomar una muestra y determinar el valor del estadístico de contraste.
4. Definir el criterio de aceptación o de rechazo. Es decir, el procedimiento parte los posibles valores del estadístico de contraste clasificados en dos subconjuntos o regiones: una región de aceptación de H_0 y una región de rechazo de H_0.
5. Tomar la decisión de aceptar o rechazar H_0 dependiendo de si el estadístico de contraste queda en la región de aceptación o en la región de rechazo.



Región de rechazo o crítica
(conjunto de valores del estadístico de contraste que premiten rechazar H_0)

Región de aceptación
(conjunto de valores del estadístico de contraste que premiten aceptar H_0)

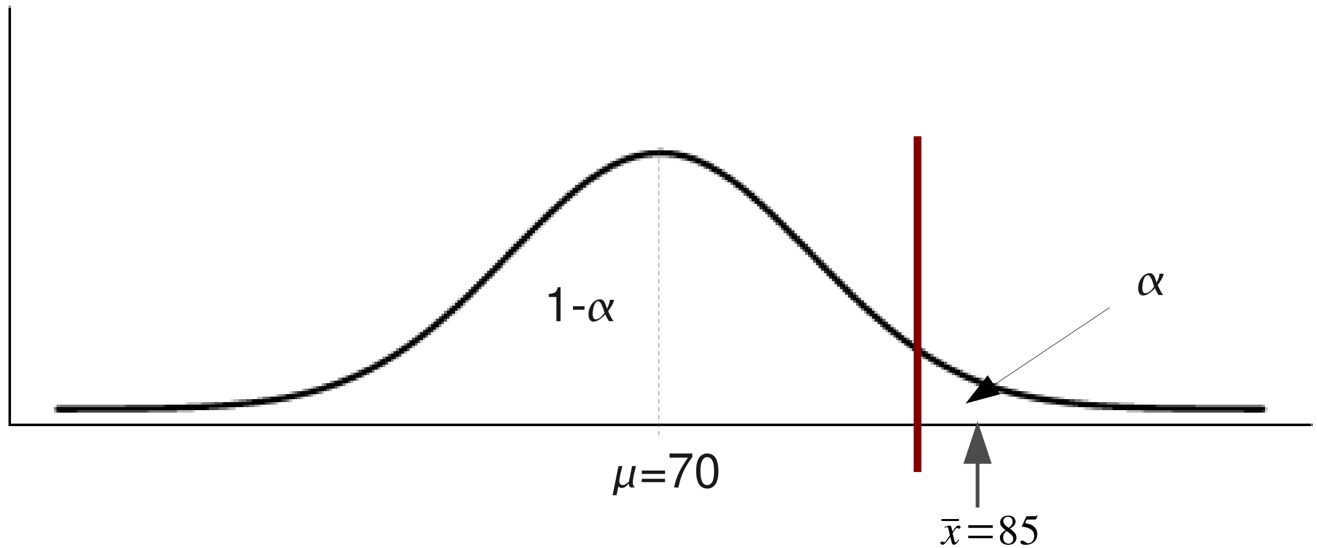
Ejemplo

En la población de no fumadores, el peso medio es 70 kg.

¿Cómo podríamos ‘demostrar’ si los fumadores pesan más?

H_0: $\mu \leq 70$

H_1: $\mu > 70$



el resultado del experimento sería improbable (si H0 fuese cierta), sin embargo ocurrió (lo que nos hace dudar de H0).

H_0 se acepta o rechaza con una confianza 1-α dependiendo de si el estadístico de contraste calculado queda en la región de aceptación o en la región de rechazo.

María Dolores Frías, Jesús Fernández y Carmen María Sordo

Contraste de hipótesis

Introducción al contraste de hipótesis

Contrastes con una muestra:

- Contraste de una proporción
- Contraste de una media
- Contraste de una varianza

Contrastes con dos muestras:

- Contraste de dos proporciones
- Contraste de dos medias

Contraste de una proporción

Se trata de contrastar un valor para la proporción poblacional en base al valor observado en una muestra:

$$H_0 : P = p_0$$
$$H_1 : P \neq p_0$$

(donde p_0 es un valor prefijado para la proporción de la población)

basándonos en la teoría vista en inferencia de una proporción podemos definir un estadístico de contraste (d) que sigue una distribución aproximadamente normal para tamaños muestrales suficientemente grandes:

$$p = \frac{m}{n} \Rightarrow N(P, \frac{P(1-P)}{n})$$

Si H_0 es cierta:

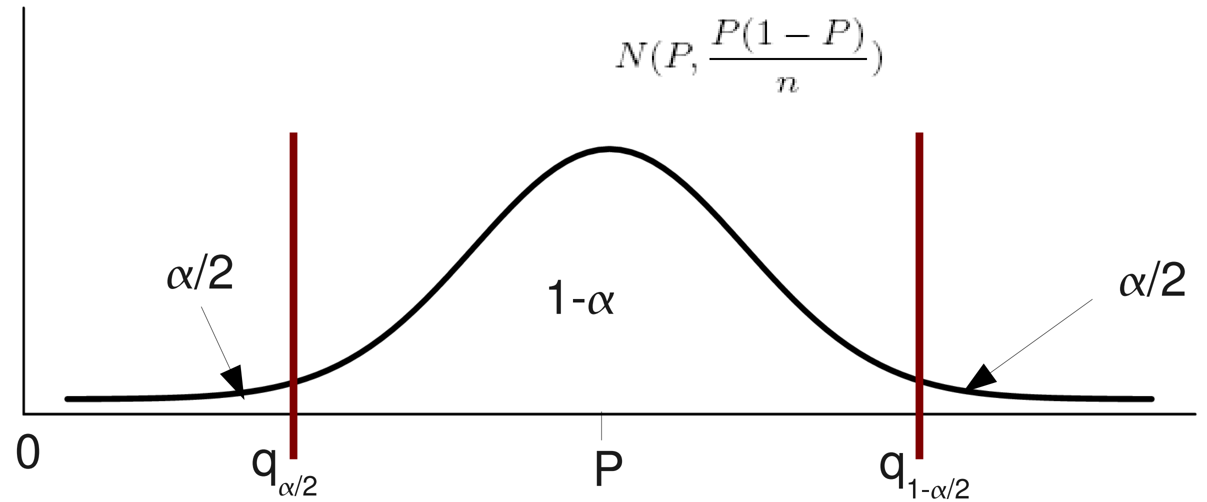
$$p = \frac{m}{n} \Rightarrow N(p_0, \frac{p_0(1-p_0)}{n}) \longrightarrow d = \frac{p - p_0}{\sqrt{p_0\,(1-p_0)/n}}$$

donde d se ajusta a una distribución N(0,1)

Contraste de una proporción

Contraste bilateral

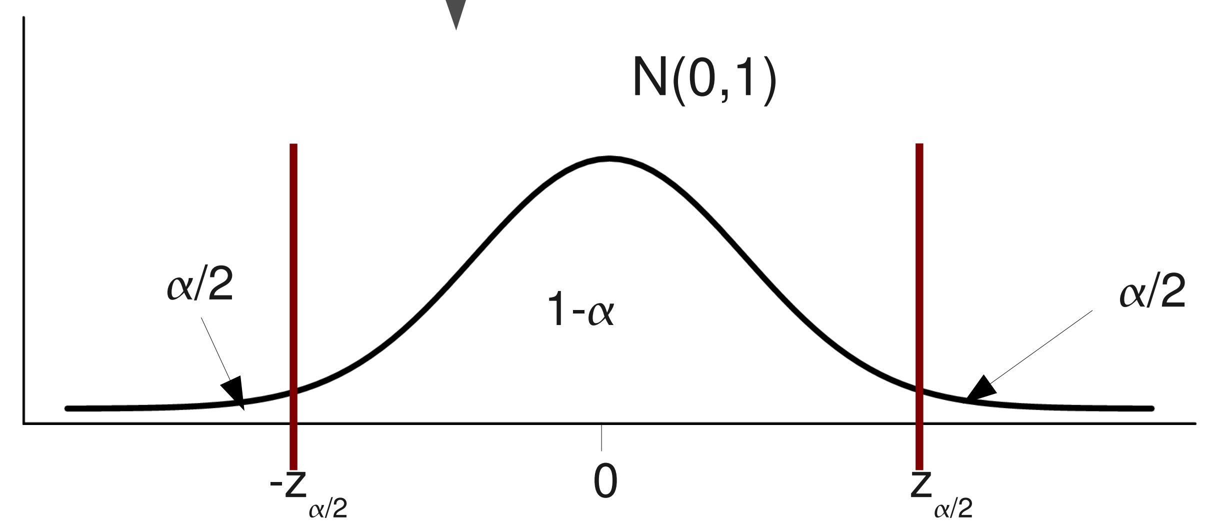
$$H_0 : P = p_0$$
$$H_1 : P \neq p_0$$



$$d = \frac{p - p_0}{\sqrt{p_0\,(1 - p_0)/n}}$$

Fijado α,

si $|d|>z_{\alpha/2}$ se RECHAZA H_0



$$z_{\alpha/2} = F^{-1}_{N(0,1)}(1 - \alpha/2)$$

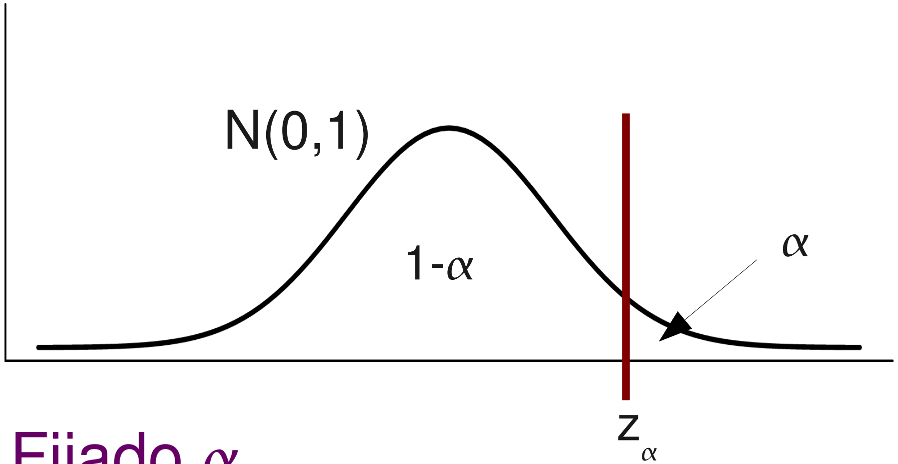
María Dolores Frías, Jesús Fernández y Carmen María Sordo

Contraste de una proporción

Contraste unilateral (una cola):

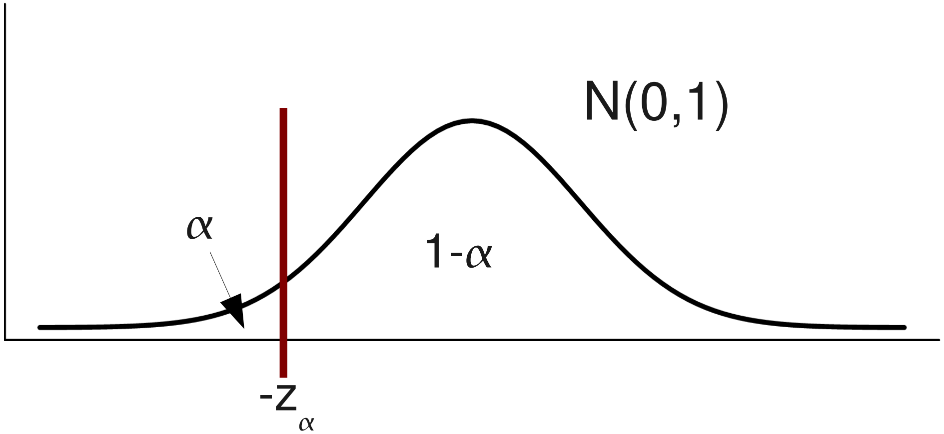
derecho:

$$H_0 : P = p_0 \Leftrightarrow P \leq p_0$$
$$H_1 : P > p_0$$



izquierdo:

$$H_0 : P = p_0 \Leftrightarrow P \geq p_0$$
$$H_1 : P < p_0$$



Fijado α,

si $d>z_\alpha$ se RECHAZA H_0

si $d<-z_\alpha$ se RECHAZA H_0

$$z_\alpha = F^{-1}_{N(0,1)}(1-\alpha)$$

María Dolores Frías, Jesús Fernández y Carmen María Sordo

Ejercicio

Se dispone de una moneda cuyo aspecto no es simétrico y se observa que al lanzarla 1000 veces se obtiene 550 veces “cruz”. ¿Qué podemos decidir sobre la moneda a un nivel del 5%?

Contraste de una media

Se trata de contrastar un valor para la media de una población normal en base al valor observado en una muestra:

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu \neq \mu_0$$

(donde μ_0 es un valor prefijado para la media de la población)

Como con las proporciones, la técnica para hacer contrastes consiste en suponer que H_0 es cierta y averiguar con esta hipótesis quien es la distribución del estadístico de contraste.

Al igual que se hizo con intervalos de confianza, barajaremos dos hipótesis:

- Varianza de la población conocida y n grande.
- Varianza de la población desconocida y n pequeña.

Contraste de una media

Varianza de la población conocida y n grande:

Con las suposiciones hechas, la distribución de la media muestral, $\bar{x} \Rightarrow N(\mu, \sigma^2/n)$

Si H0 es cierta: $\bar{x} \Rightarrow N(\mu_0, \sigma^2/n)$ $\longrightarrow$ $d = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$ $\quad d \Rightarrow N(0,1)$

Contraste bilateral (dos colas):

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu \neq \mu_0$$

Fijado α, si $|d|>z_{\alpha/2}$ se RECHAZA H0

Contraste unilateral por la derecha:

$$H_0 : \mu = \mu_0 \Leftrightarrow \mu <= \mu_0$$
$$H_1 : \mu > \mu_0$$

Fijado α, si $d>z_{\alpha}$ se RECHAZA H0

Contraste unilateral por la izquierda:

$$H_0 : \mu = \mu_0 \Leftrightarrow \mu >= \mu_0$$
$$H_1 : \mu < \mu_0$$

Fijado α, si $d<-z_{\alpha}$ se RECHAZA H0

$$z_\alpha = F^{-1}_{N(0,1)}(1-\alpha)$$

R tip

```
> qnorm(1-α)   # z_α
```

María Dolores Frías, Jesús Fernández y Carmen María Sordo

Contraste de una media

Varianza de la población desconocida y n pequeña:

En este caso, σ^2 se estima a partir de S^2 (estimador insesgado), $d \Rightarrow t(n-1)$

Si H0 es cierta: $$d = \frac{\bar{x} - \mu_0}{S/\sqrt{n}}$$

Contraste bilateral (dos colas):

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu \neq \mu_0$$

Fijado α, si $|d|>t_{n-1;\alpha/2}$ se RECHAZA H0

Contraste unilateral por la derecha:

$$H_0 : \mu = \mu_0 \Leftrightarrow \mu <= \mu_0$$
$$H_1 : \mu > \mu_0$$

Fijado α, si $d>t_{n-1;\alpha}$ se RECHAZA H0

Contraste unilateral por la izquierda:

$$H_0 : \mu = \mu_0 \Leftrightarrow \mu >= \mu_0$$
$$H_1 : \mu < \mu_0$$

Fijado α, si $d<-t_{n-1;\alpha}$ se RECHAZA H0

$$t_{n-1;\alpha} = F^{-1}_{t(n-1)}(1-\alpha)$$

R tip

```
> qt(1-α, n-1)   # t_{n-1,α}
```

Ejercicio

Sabiendo que la altura de los individuos de una población sigue una distribución gaussiana, se desea contrastar si su altura media es diferente de 174 cm con una significación de $\alpha = 0.05$. Para ello nos basamos en un estudio en el que con una muestra de 25 personas se obtuvo una media de 170 cm y una desviación típica de 10 cm.

Contraste de una varianza

Se trata de contrastar un valor para la varianza de una población normal en base al valor observado en una muestra:

$$H_0 : \sigma^2 = \sigma_0^2$$
$$H_1 : \sigma^2 \neq \sigma_0^2$$

(donde σ^2_0 es un valor prefijado para la media de la población)

Como con en los casos anteriores, la técnica para hacer el contraste consiste en suponer que H_0 es cierta y averiguar con esta hipótesis quien es la distribución del estadístico de contraste.

A partir de la teoría vista en el tema de inferencia sabemos que

$$d = \frac{(n-1)\,S^2}{\sigma_0^2} \qquad \text{ó} \qquad d = \frac{n\,S_n^2}{\sigma_0^2}$$

se ajusta a una distribución χ^2(n-1).

La suposición de normalidad de la población de partida en este caso es **muy importante.** El caso de las medias es menos sensible.

Contraste de una varianza

Contraste bilateral (dos colas): $H_0: \sigma^2 = \sigma_0^2$
$H_1: \sigma^2 \neq \sigma_0^2$

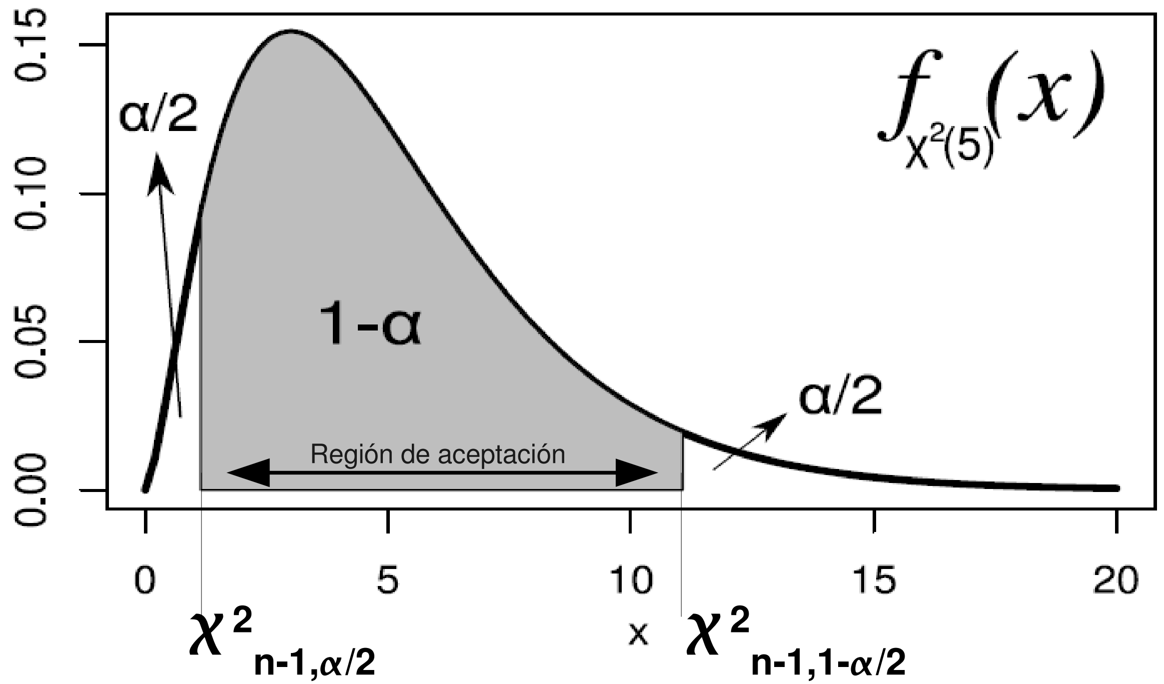
La distribución χ^2 **NO** es simétrica. Por lo tanto $\chi^2_{n-1,1-\alpha/2} \neq -\chi^2_{n-1,\alpha/2}$

Hay que buscar cada valor independientemente en la tabla.

Fijado α,

si d < $\chi^2_{n-1,\alpha/2}$ ó d > $\chi^2_{n-1,1-\alpha/2}$

se RECHAZA H0



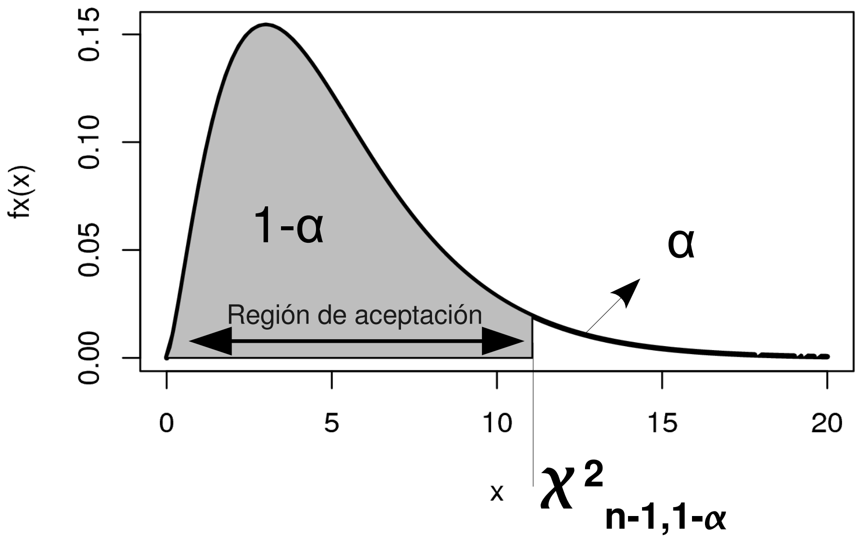
R tip

```
> # χ²n-1,α/2
> qchisq(α/2,n-1)
> # χ²n-1,1-α/2
> qchisq(1-α/2,n-1)
```

$$\chi^2_{n-1;\alpha/2} = F^{-1}_{\chi^2(n-1)}(\alpha/2)$$
$$\chi^2_{n-1;1-\alpha/2} = F^{-1}_{\chi^2(n-1)}(1-\alpha/2)$$

Contraste de una varianza

Contraste unilateral por la derecha:

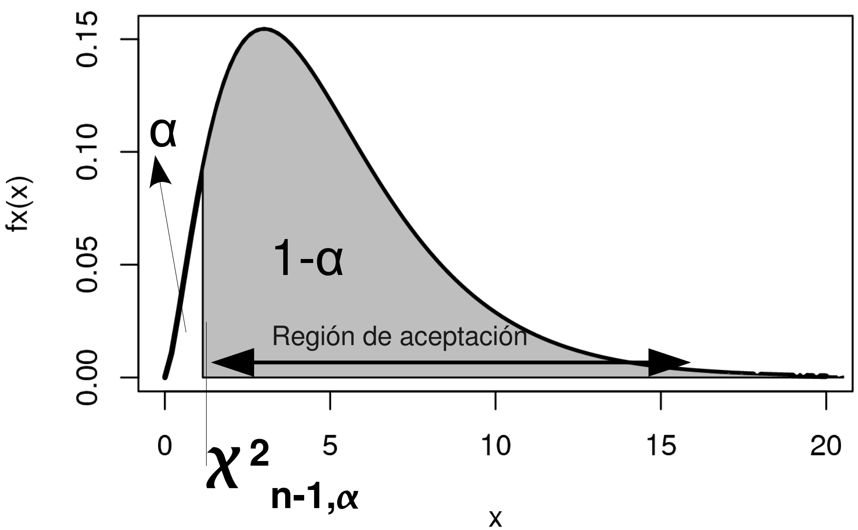


$$H_0: \sigma^2 \leq \sigma_0^2$$
$$H_1: \sigma^2 > \sigma_0^2$$

Fijado α, si d>$\chi^2_{n-1,1-\alpha}$ se RECHAZA H0

$$\chi^2_{n-1;1-\alpha} = F^{-1}_{\chi^2(n-1)}(1-\alpha)$$

Contraste unilateral por la izquierda:



$$H_0: \sigma^2 \geq \sigma_0^2$$
$$H_1: \sigma^2 < \sigma_0^2$$

Fijado α, si d<$\chi^2_{n-1,\alpha}$ se RECHAZA H0

$$\chi^2_{n-1;\alpha} = F^{-1}_{\chi^2(n-1)}(\alpha)$$

Ejercicio

Un fabricante especifica que la desviación estandard del tiempo de acceso al disco duro debe ser como mucho de 0.1 milisegundos. El supervisor de control de calidad seleccionó aleatoriamente 10 discos de esta marca midiendo el tiempo de acceso para cada uno obteniendo los siguientes valores (ms):

7.96 7.90 7.98 8.01 7.97 7.96 8.03 8.02 8.04 8.02

¿Proporciona esta información pruebas suficientes de que la desviación estandard de estos discos es mayor de 0.1? Tomar α=0.05

Ejercicio

Un fabricante especifica que la desviación estandard del tiempo de acceso al disco duro debe ser como mucho de 0.1 milisegundos. El supervisor de control de calidad seleccionó aleatoriamente 10 discos de esta marca midiendo el tiempo de acceso para cada uno obteniendo los siguientes valores (ms):
7.96 7.90 7.98 8.01 7.97 7.96 8.03 8.02 8.04 8.02
¿Proporciona esta información pruebas suficientes de que la desviación estandard de estos discos es mayor de 0.1? Tomar α=0.05

H0: $\sigma^2 \leq 0.01$
H1: $\sigma^2 > 0.01$

Contraste unilateral derecho

Suponiendo que se trata de una población Normal:

$$d=(10-1)0.43^2/0.01=1.66$$

Con α=0.05 y 9 grados de libertad:

$\chi^2_{n,1-\alpha}=16.919$

```
> qchisq(0.95,9)       # χ²n-1,α
```

	Valores de p									
n	0.005	0.010	0.025	0.050	0.100	0.900	0.950	0.975	0.990	0.995
1	$0.0_4 39$	$0.0_3 16$	$0.0_3 98$	$0.0_2 39$	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188

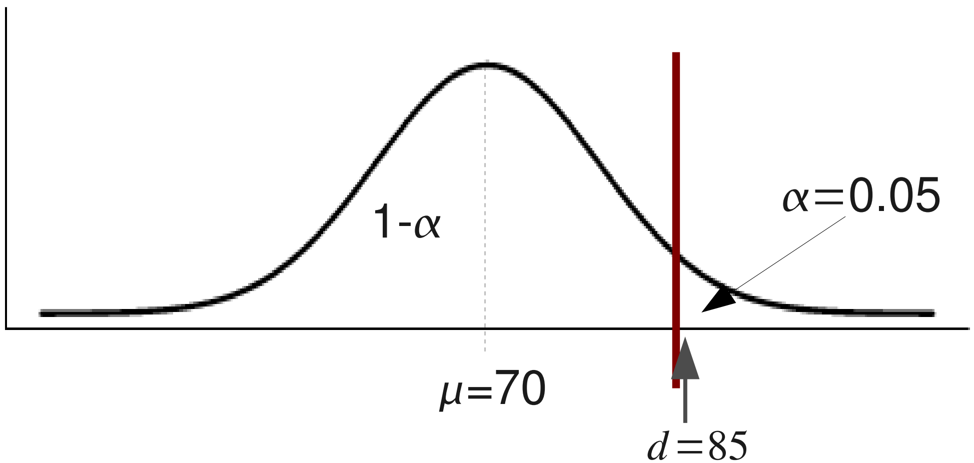
Tabla 5.5: Valores de $\chi^2_{n;p} \equiv F^{-1}_{\chi^2(n)}(p)$. Los subíndices indican el número de repeticiones de un digito. Por ejemplo, $F^{-1}_{\chi^2(1)}(0.005) = 0.0_4 39 = 0.000039$.

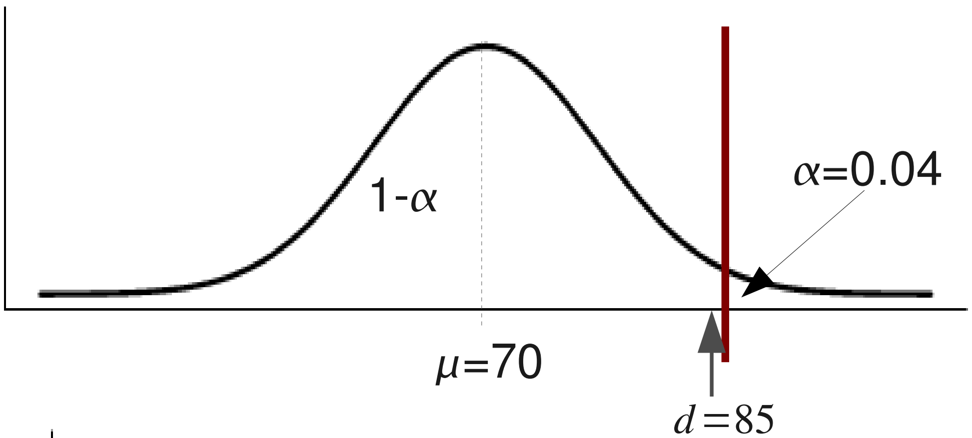
Puesto que el estadístico de contraste es menor que 16.91 (está en la zona de aceptacion), se acepta H0, luego el supervisor puede llegar a la conclusion de que la desviación estandard de la población como mucho es 0.1.

p - valor

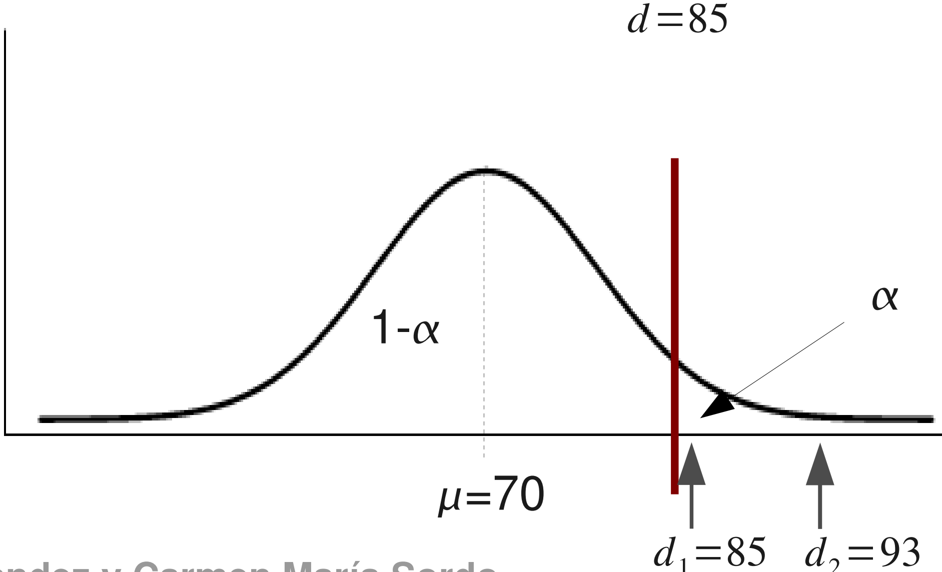
El procedimiento de selección de una región de rechazo mediante el nivel de significación está sujeto a dos críticas principales:

1. El resultado del test puede depender mucho del valor de α, que es arbitrario, siendo posible rechazar H_0 con $\alpha = 0.05$ y aceptarla con $\alpha = 0.04$.



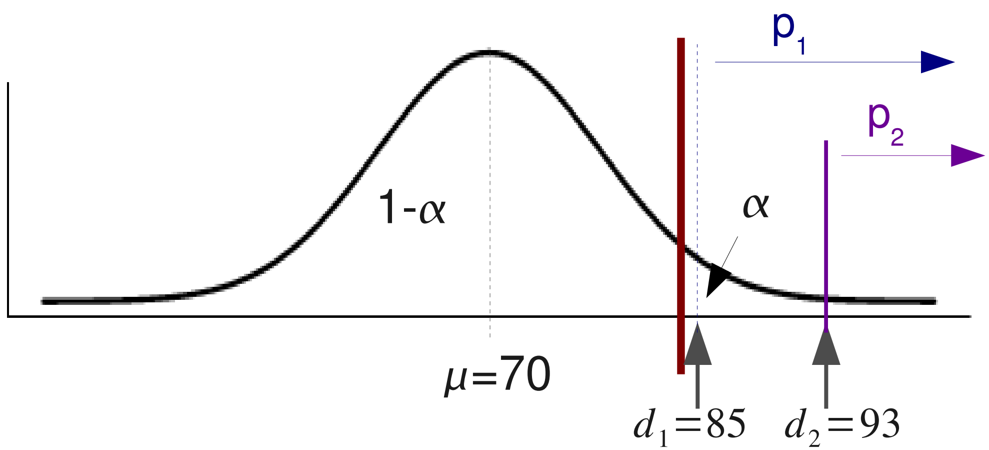


2. Dar sólo el resultado del contraste no permite diferenciar el grado de evidencia que la muestra indica a favor o en contra de H_0. Nótese que valores cercanos o lejanos al valor crítico producen el mismo resultado.



p - valor

Un procedimiento para hacer frente a estas críticas es utilizar en lugar de α, el nivel crítico de un test o **p valor**. Es el nivel de significación más pequeño que conduce al rechado de H_0



Significación asociada a una región crítica que comenzase exactamente en el valor del estadístico obtenido de la muestra.

Si $p \leq \alpha$ se rechaza H_0, caso contrario se acepta H_0.

Cuanto menor sea p menor probabilidad tendrá H_0 de explicar el resultado obtenido (menor es la credibilidad de H_0).

Contraste bilateral:

$$\text{p-valor} = 2\,P(d>|d_0|)$$

Contraste unilateral drcho:

$$\text{p-valor} = P(d>d_0)$$

Contraste unilateral izqdo:

$$\text{p-valor} = P(d<d_0)$$

María Dolores Frías, Jesús Fernández y Carmen María Sordo

Ejercicio

Un estudio ha demostrado que las personas que trabajan con benceno durante más de cinco años tienen una incidencia de leucemia 20 veces mayor que la población en general. En consecuencia la normativa vigente ha bajado el nivel máximo permisible de benceno en el lugar de trabajo de 10 partes por millón (ppm) a 1 ppm.
Se sospecha que una fábrica de artículos de acero, que expone a sus trabajadores diariamente a benceno, no cumple esta normativa. La administración examina 20 muestras de aire, tomadas durante un periodo de un mes, para determinar el contenido de benceno. obteniendo un valor medio de 2.1ppm y una cuasi-desviación típica de 1.7ppm.
¿Determinar si la fábrica de artículos de acero está violando la nueva normativa del gobierno con una confianza del 95%? Determinar el valor crítico (p valor) de la prueba.

Contraste de hipótesis

Introducción al contraste de hipótesis

Contrastes con una muestra:

- Contraste de una proporción
- Contraste de una media
- Contraste de una varianza

Contrastes con dos muestras:

- Contraste de dos proporciones
- Contraste de dos medias

Contraste de dos proporciones

Se trata de contrastar el valor de la proporción poblacional en dos poblaciones distintas en base a dos muestras independientes:

$$H_0: P_1 = P_2 \qquad \longleftrightarrow \qquad H_0: P_1 - P_2 = 0$$
$$H_1: P_1 \neq P_2 \qquad \qquad \qquad H_1: P_1 - P_2 \neq 0$$

Basándonos en el estadístico de contraste siguiente, que sigue una distribución N(0,1) para muestras suficientemente grandes.

$$d = \frac{p_1 - p_2}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$

Fijado α, si $|d|>z_{\alpha/2}$ se RECHAZA H0, contraste bilateral

si $d>z_\alpha$ se RECHAZA H0, contraste unilateral derecho

si $d<-z_\alpha$ se RECHAZA H0, contraste unilateral izquierdo

$$z_{\alpha/2} = F^{-1}_{N(0,1)}(1-\alpha/2) \qquad z_\alpha = F^{-1}_{N(0,1)}(1-\alpha)$$

María Dolores Frías, Jesús Fernández y Carmen María Sordo

Ejercicio

Se está considerando cambiar el procedimento de fabricación de cierta pieza. Para ello, se tomaron 1000 piezas al azar de las producidas según el proceso actual. 75 de ellas resultaron defectuosas. Así mismo, de 1500 piezas tomadas de aquellas fabricadas por el nuevo método, 85 resultaron defectuosas. Determinar, con una confianza del 90 % si el nuevo procedimiento puede considerarse mejor.

Contraste de dos medias

Muestras independientes

Se trata de contrastar la igualdad de las medias de dos poblaciones normales distintas de **varianzas conocidas** en base a dos muestras **independientes** con medias muestrales $\overline{x}_1$ y $\overline{x}_2$ de tamaños respectivos n_1 y n_2

$$H_0 : \mu_1 - \mu_2 = a$$
$$H_1 : \mu_1 - \mu_2 \neq a$$

El estadístico de contraste,

$$d = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

se distribuye según una N(0,1)

Contraste de dos medias

Muestras independientes

Caso 1: Si las varianzas poblacionales son desconocidas y n_1 y n_2 son suficientemente grandes (>15), se pueden sustituir las varianzas poblacionales por las cuasi-varianzas muestrales.

Fijado α, si $|d|>z_{\alpha/2}$ se RECHAZA H0, contraste bilateral

si $d>z_{\alpha}$ se RECHAZA H0, contraste unilateral derecho

si $d<-z_{\alpha}$ se RECHAZA H0, contraste unilateral izquierdo

$$z_{\alpha/2} = F^{-1}_{N(0,1)}(1-\alpha/2) \qquad z_{\alpha} = F^{-1}_{N(0,1)}(1-\alpha)$$

Contraste de dos medias

Muestras independientes

Caso 2: Si las varianzas son desconocidas pero iguales, $\sigma_1 \approx \sigma_2 \approx \sigma$, consideraremos el estadístico de contraste:

$$d = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\hat{S}_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

donde

$$\hat{S}_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

siendo S^2_1 y S^2_2 las cuasi-varianzas muestrales

d se distribuye según una t de Student con n_1+n_2-2 grados de libertad

Fijado α, si $|d|>t_{n1+n2-2;\alpha/2}$ se RECHAZA H0, contraste bilateral

si $d>t_{n1+n2-2;\alpha}$ se RECHAZA H0, contraste unilateral derecho

si $d<-t_{n1+n2-2;\alpha}$ se RECHAZA H0, contraste unilateral izquierdo

$$t_{n_1+n_2-2;\alpha} = F^{-1}_{t(n_1+n_2-2)}(1-\alpha)$$

Ejercicio

Suponga que un centro de datos desea comparar los tiempos de respuesta medios de sus dos unidades de disco de computadora. Sean μ_1 el tiempo de respuesta medio del disco 1 y μ_2 el tiempo de respuesta medio del disco 2. Se selecionaron muestras aleatorias del tiempo de respuesta para cada uno de los discos cuyos valores (en milisegundos) se muestran en la siguiente tabla:

Suponiendo que $\sigma_1 \approx \sigma_2 \approx \sigma$, se pide:

$disco1$	59	73	74	61	92	60	84	54	73	47	102	75	33		
$disco2$	71	63	40	34	38	48	60	75	47	41	44	86	53	68	39

Determinar si hay pruebas suficientes de una diferencia entre los tiempos de respuesta medios de las dos unidades de disco. Considerar un $\alpha = 0{,}05$.

R tip

```
> t1<-c(59,73,74,61,92,60,84,54,73,47,102,75,33)
> t2<-c(71,63,40,34,38,48,60,75,47,41,44,86,53,68,39)
> t.test(t1,t2,var.equal=TRUE,alternative='two.sided',conf.level=.95)
```

Muestras apareadas

Imaginemos el caso de que tomamos una muestra de tamaño n y valores $\{x_1, ..., x_n\}$ y otra muestra $\{y_1, ..., y_n\}$ en la que cada valor de y_i depende de x_i. Estas dos muestras son dependientes y **no se puede** utilizar la teoría mostrada en las transparencias anteriores. Supongamos que, aparte de la dependencia entre miembros de una pareja, los pares son independientes entre sí (i.e. no hay relación entre x_i y x_j).

Ejemplos de este tipo de muestreos son aquellos que involucran medir una cualidad en unos objetos, realizar un proceso y medir otra (o la misma) cualidad en **los mismos** objetos.

Horas perdidas por un trabajador antes y despues de implantar un programa de aprovechamiento del tiempo.

Otro ejemplo pueden ser objetos que han pasado emparejados un proceso aleatorio.

Desgaste de un par de zapatos en los que uno de ellos ha sido tratado con un nuevo polímero protector.

Contraste de dos medias

Muestras apareadas

Entonces, si construimos una nueva variable $D_i = x_i - y_i$, sus elementos serán independientes entre sí y podremos tratarla como una muestra única y el estadístico:

$$d = \frac{\overline{D} - (\mu_x - \mu_y)}{S_D/\sqrt{n}}$$

seguirá una distribución t de Student con n-1 grados de libertad.

Fijado α, si $|d| > t_{n-1;\,\alpha/2}$ se RECHAZA H0, contraste bilateral

si $d > t_{n-1;\,\alpha}$ se RECHAZA H0, contraste unilateral derecho

si $d < -t_{n-1;\,\alpha}$ se RECHAZA H0, contraste unilateral izquierdo

$$t_{n-1;\alpha} = F^{-1}_{t(n-1)}(1-\alpha)$$

Ejercicio

En un estudio se desea saber si la presión sistólica de personas fumadoras cambia cuando las personas son sometidas a un tratamiento para dejar de fumar. Para ello se seleccionan 10 personas a las que se les toma la presión sistólica antes de empezar el tratamiento y tres meses después de dejar de fumar. Con los datos de ambas presiones que se muestran en la siguiente tabla determinar la conclusión del estudio para un nivel de significación del 5%.

Individuo	1	2	3	4	5	6	7	8	9	10
Antes	140	165	160	160	175	190	170	175	155	160
Despues	145	150	150	160	170	175	160	165	145	170