PRECISIÓN Y VERACIDAD. EVALUACIÓN DE DATOS ANALÍTICOS
Pruebas de significación

INTRODUCCIÓN

Supongamos una situación en la que se está determinando el contenido en carbonato cálcico de una muestra de un material de referencia certificado que contiene 50,00 mg de $CaCO_3$. Para ello se llevan a cabo 4 determinaciones (n=4). La media que obtenemos es $\bar{x}$ = 49,30 y previamente se ha estimado que la σ de nuestras determinaciones es 0,82.

Deseamos conocer si las medias obtenidas por nosotros son significativamente diferentes del valor verdadero o teórico esperado. De esta manera podremos determinar si nuestras determinaciones llevan asociados un componente de error sistemático (bias) además del error aleatorio inevitable. Dicho de otra forma, queremos conocer si la diferencia entre $\bar{x}$ y μ_0 es superior a la pura aletoriedad que conlleva toda determinación experimental. (Recordemos que $x_i - \mu_0 = (x_i - \bar{x}) + (\bar{x} - \mu_0)$ o bien $x_i - \mu_0 = (x_i - \mu) + (\mu - \mu_0)$)

Para ello se lleva a cabo un test de hipótesis o prueba de significación, que trata de encontrar si hay diferencias estadísticamente significativas entre μ (estimado por $\bar{x}$) y μ_0 (lo cual no significa necesariamente que esas diferencias sean relevantes desde el punto de vista químico)

HIPÓTESIS NULA Y ALTERNATIVA

Como punto de partida tenemos la **hipótesis nula** de que no hay diferencias significativas entre μ y μ_0, lo que se escribe:

H_0: $\mu = \mu_0$

Para el caso en que H_0 no sea cierta necesitamos una **hipótesis alternativa**

H_1: $\mu \neq \mu_0$. (También puede plantearse $\mu > \mu_0$ o $\mu < \mu_0$)

PLANTEAMIENTO DE LAS PRUEBAS

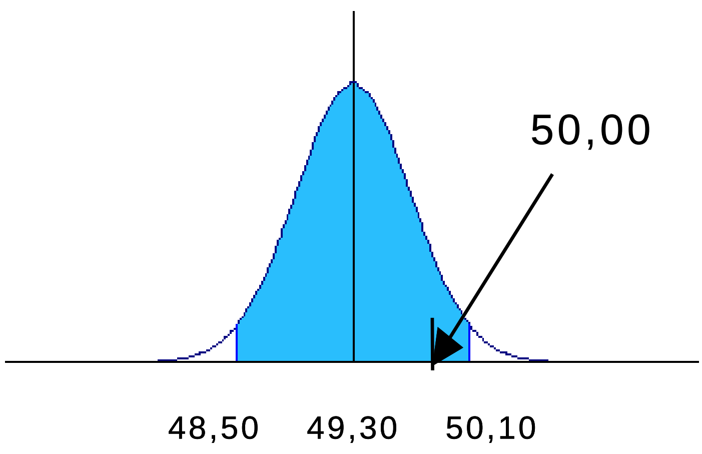
Utilización de intervalos de confianza

Supongamos que tomamos como nivel de confianza el 95% (nos confundiremos una vez de cada 20) o sea un **nivel de significación α = 0,05.**

Puesto que $\bar{x}$ es un estimador de μ, sabemos que $\mu = \bar{x} \pm 1{,}96\sigma/\sqrt{n}$, luego μ = 49,30 $\pm$ 1,96.0,82/$\sqrt{4}$ = 49,30 $\pm$ 0,80, es decir 48,50 $\leq \mu \leq$ 50,10. Es decir el valor verdadero de nuestras determinaciones está dentro del intervalo 48,50 – 50,10. Dado que ese intervalo incluye al valor verdadero μ_0 (que vale 50,00), podemos concluir que $\mu = \mu_0$, y por tanto **aceptamos la hipótesis nula**, de que no hay diferencias significativas entre μ y μ_0. Esto no significa que hayamos probado que H_0 es cierta, sino solo que **no tenemos evidencias para rechazarla**.

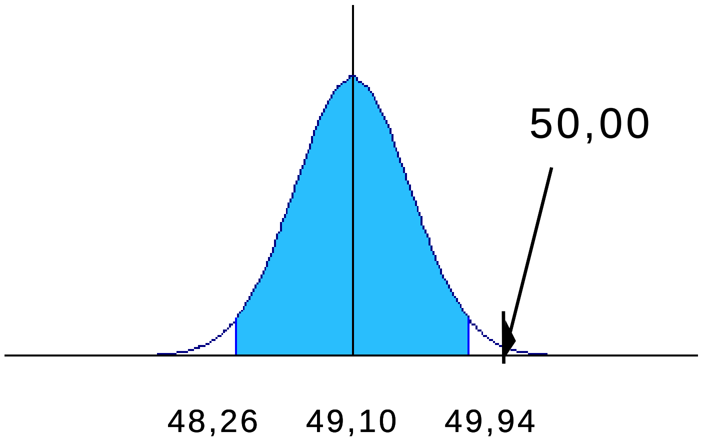
Las etapas del test o prueba de significación tal como lo hemos realizado son

1) Plantear H_0 y H_1
2) Elegir el nivel de significación α
3) Calcular el intervalo de confianza alrededor de $\bar{x}$ a un nivel de 100-α%= 95%
4) Investigar si μ_0 está dentro de ese intervalo
5) Si le respuesta es positiva, se acepta H_0. Si es negativa se rechaza H_0 (y se acepta H_1)



En el ejemplo anterior (Ejemplo 1), la σ de la población de nuestras determinaciones era conocida pero habitualmente ese no es el caso, sino que σ se estima simultáneamente con $\bar{x}$, mediante s. En esas condiciones se debe utilizar la distribución t, pero las etapas de la prueba siguen siendo las mismas. Supongamos, p.e. que se hacen 6 réplicas (n=6), tales que $\bar{x}$ = 49,10 y s = 0,80 (Ejemplo 2)

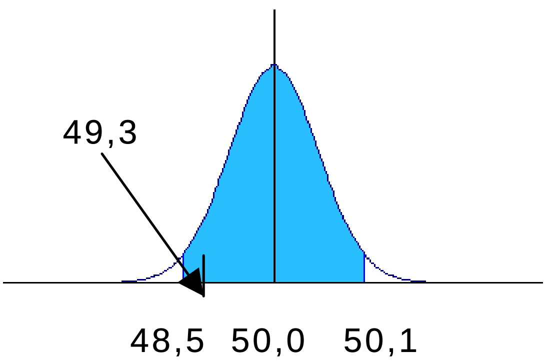
1) H_0: $\mu = \mu_0$; H_1: $\mu \neq \mu_0$ (H_0: μ = 50,0; H_1: $\mu \neq$ 50,0)
2) Nivel de significación α = 0,05
3) Intervalo de confianza alrededor de $\bar{x}$ a un nivel de 100-α%= 95% $\mu = \bar{x} \pm t_{\alpha/2,(n-1)}\,(s/\sqrt{n}) =$ μ = 49,10 $\pm$ 2,57.0,80/$\sqrt{6}$ = 49,10 $\pm$ 0,84, es decir 48,26 $\leq \mu \leq$ 49,94
4) Investigar si μ_0 = 50,0 está dentro de ese intervalo
5) Ya que no lo está se rechaza H_0 (y se acepta H_1), es decir hay diferencias significativas entre μ y μ_0

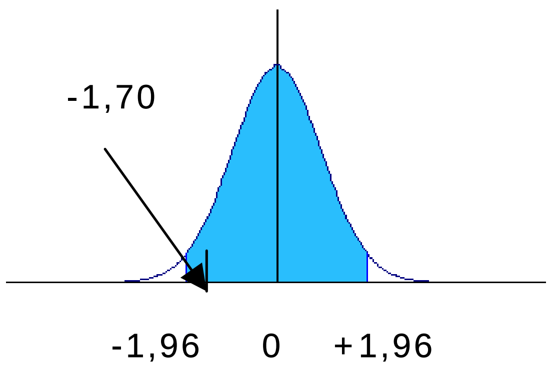


Comparación con valores críticos

Hay otra forma de hacer el test, que parece diferente pero que es exactamente la misma. Veamos El Ejemplo 1 (con σ conocida)

Sabemos que el 95% de todas las medias muestrales cae dentro del intervalo $\mu \pm 1{,}96\sigma/\sqrt{n}$. Si suponemos que H_0 es cierta ($\mu = \mu_0$) también caerán dentro de $\mu_0 \pm 1{,}96\sigma/\sqrt{n}$. En la Figura siguiente aparecen los intervalos de confianza alrededor de μ_0 en unidades originales y en unidades z.





Razonando en unidades z, el 95% de todas las medias compatibles con H_0: $\mu = \mu_0$ estará dentro del intervalo comprendido entre z=-1,96 y z=+1,96, por lo que si expresamos la distancia de $\bar{x}$ a μ_0 en unidades z, cuando **el valor absoluto de ese z experimental (z_{exp}) sea menor que 1,96 $\bar{x}$ estará dentro del intervalo, y la H_0 se aceptará**. Si el valor absoluto de z_{exp} es mayor que 1,96, entonces $\bar{x}$ estará fuera del intervalo, por lo que H_0 se rechazará (y se aceptará H_1).

Por consiguiente 1,96 es un **valor crítico**, ya que representa la máxima distancia en unidades z a que puede estar situada una media $\bar{x}$ del valor esperado μ_0, para poder ser considerada como perteneciente a la población de μ_0 con un 95% de confianza.

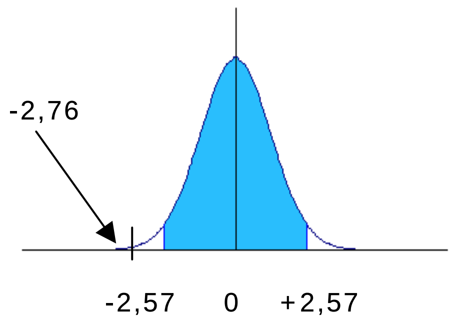
En este caso el valor experimental de |z|, calculado a partir de $|z| = \frac{|\bar{x} - \mu_0|}{\sigma/\sqrt{n}}$, vale $|z|_{exp} = \frac{|49{,}30 - 50{,}0|}{0{,}82/\sqrt{4}} = 1{,}70$, y por tanto está más cerca que 1,96, o lo que es igual $|z_{exp}| < z_{crit}$ por lo que la H_0 se mantiene.

En esta otra modalidad las **etapas del test** son:

1) Plantear H_0 y H_1
2) Elegir el nivel de significación α
3) Determinar el valor crítico del estadístico correspondiente
4) Calcular el valor experimental del estadístico
5) Comparar ambos valores
6) Si $|valor_{exp}| < valor_{crit}$ se acepta H_0. En caso contrario se rechaza H_0(y se acepta H_1)

Si lo aplicamos ahora al Ejemplo 2 (n=6), $\bar{x}$= 49,10 y s = 0,80), la prueba seguiría las siguientes etapas:

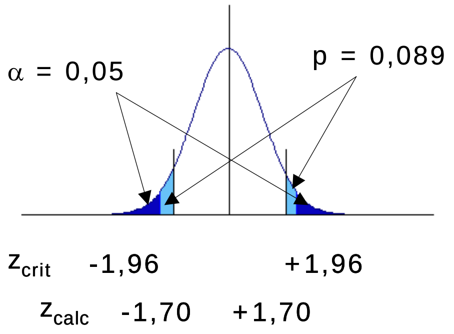
1) H_0: $\mu = \mu_0$; H_1: $\mu \neq \mu_0$ (H_0: $\mu = 50{,}0$; H_1: $\mu \neq 50{,}0$)
2) Nivel de significación $\alpha = 0{,}05$
3) Valor crítico del estadístico correspondiente. Será t con $\alpha = 0{,}05$ y n-1 = 5 g.d.l. De acuerdo con la tabla 3.4, $t_{crit} = 2{,}571$ (Función EXCEL : DIST.T.INV(α;g.d.l.))
4) Valor experimental del estadítico $|t|_{exp} = \frac{|\bar{x} - \mu_0|}{s/\sqrt{n}} = \frac{|49{,}10 - 50{,}0|}{0{,}80/\sqrt{6}} = 2{,}76$
5) $t_{exp} > t_{crit}$
6) La H_0 se rechaza y debe aceptarse la H_1: hay diferencias significativas entre μ y μ_0.



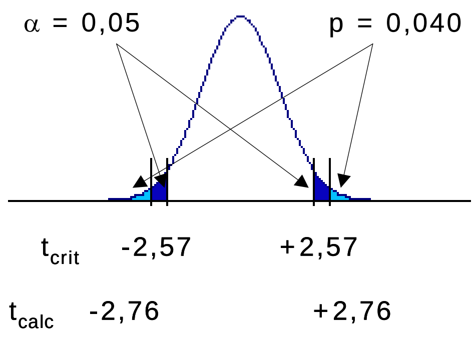
Evidentemente, ambas maneras de hacer la prueba deben originar el mismo resultado.

A la hora de presentar los resultados, además de indicar si existen o no diferencias significativas, se calcula un **nivel de significación a posteriori** (que se denomina **p** para diferenciarlo del nivel de significación α a priori). Ese valor es la probabilidad correspondiente a los valores experimentales de los estadísticos utilizados (z_{exp} o t_{exp}) Para ello se calcula el valor de p para el cual se obtiene un valor de z o de t igual al $|valor|_{exp}$. Para el Ejemplo 1, p= 0,089 (Función de EXCEL (1-DISTR.NORM.ESTAND(1,70))*2). Para el Ejemplo 2, p= 0,040 (Función de EXCEL DISTR.T(2,76;5;2)).

El valor de p representa el nivel de significación hasta el cual se puede mantener la H_0. Siempre que **$p > \alpha$ la H_0 se mantiene**, y si $p < \alpha$ la H_0 se rechaza.



Ejemplo 1



Ejemplo 2

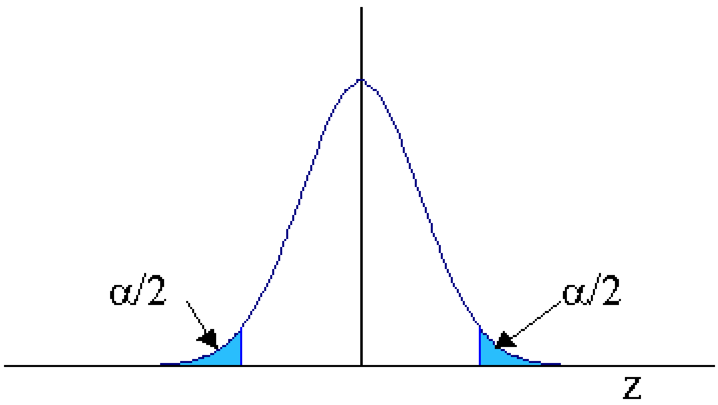
Según se aprecia en la Figura, cuando p sea mayor de α el estadístico experimental será menor que el crítico (por lo que el valor pertenecerá a la distribución de μ_0) y viceversa.

ERRORES DE TIPO I Y II

Error de tipo I o error α

Es la **probabilidad de rechazar la H_0 cuando es cierta**. Es decir es la probabilidad de decir que $\bar{x}$ no pertenece a la población de μ_0, cuando sí que pertenece a ella. **El error de tipo I coincide con el nivel de significación a priori** α.

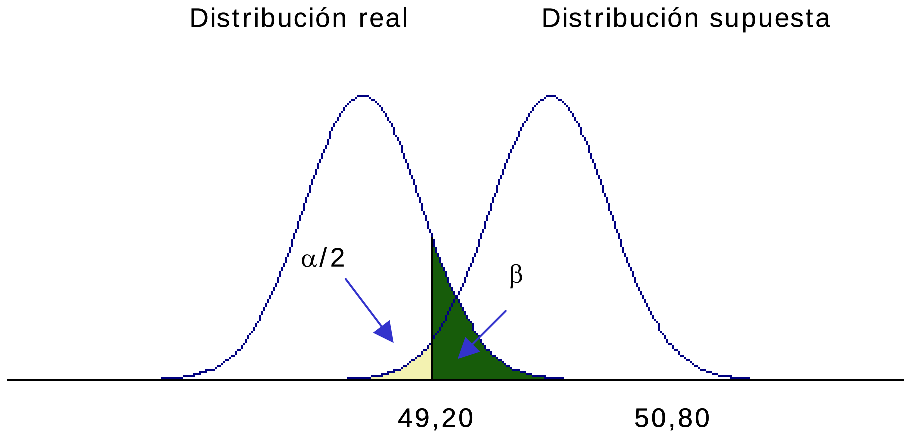
Así para un nivel α = 0,05 (un nivel de confianza del 95 %), cuando examinemos muchas medias determinadas por nosotros, el 5% de las mismas se rechazará a pesar de pertenecer a la población de μ_0.



Error de tipo II o β

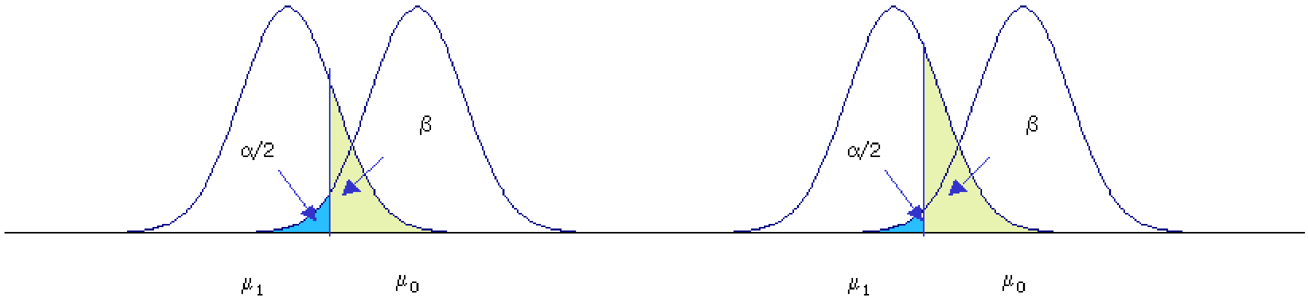
Sea el Ejemplo 1 (con σ conocida) y supongamos que la H_0 es cierta, es decir que $\mu = \mu_0$ =50,00. Los valores de $\bar{x}$ que deberíamos obtener deberían estar (con un 95% de probabilidad) dentro del intervalo $50{,}00 \pm 1{,}96.0{,}82/\sqrt{4} = 50{,}00 \pm 0{,}80$, es decir $49{,}20 \leq \mu \leq 50{,}80$. Cualquier valor de $\bar{x}$ que obtengamos y que esté fuera de ese intervalo hará que rechacemos H_0 y aceptemos H1. Si por ejemplo, obtuviéramos obtenemos $\bar{x}$ = 48,80 rechazaríamos H_0, como rechazaríamos cualquier valor que fuera inferior a 49,20.

Supongamos ahora que el método que utilizamos tiene un **bias de -1,20, desconocido para nosotros**, de manera que la media de la población de las determinaciones es en realidad μ_1 = 48,80. Como **no lo sabemos**, nosotros pensamos que el intervalo es 49,20 - 50,80, cuando de hecho será 48,00 - 49,60 por lo que cuando encontremos un valor de $\bar{x}$ mayor de 49,60 (p.e. 49,80) y aceptemos H_0 (ya que estaría dentro del intervalo 49,20-50,80) cometeremos un error. Dicho error se denomina **error de tipo II o error** β y consiste en **aceptar la H_0 cuando es falsa** (o de rechazar H_1 cuando es cierta). El bias no habría sido detectado.

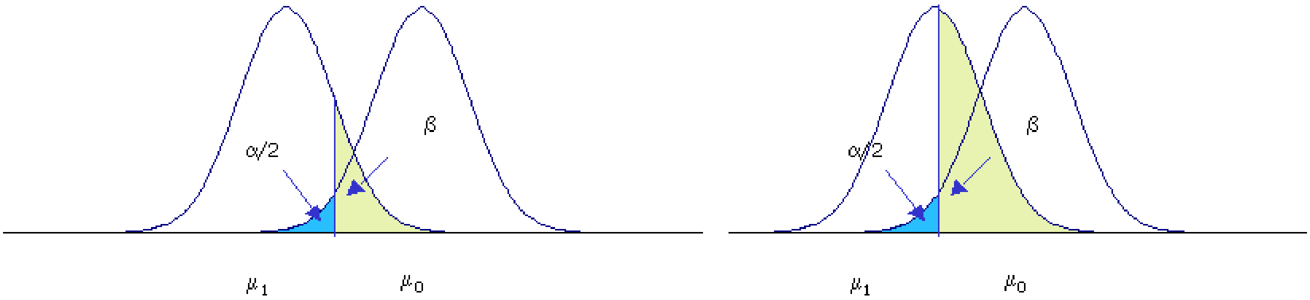


Por tanto P(I) = α y P(II) = β. Sabiendo que hay un bias de -1,20 y que μ = 48,80, para calcular β nos debemos preguntar por la probabilidad de encontrar un valor mayor de 49,20 para la distribución centrada en la distribución real. El $|z|_{exp} = \frac{|49{,}20 - 48{,}80|}{0{,}82/\sqrt{4}} = 0{,}976$ y la fracción de medidas con z > 0,976 es 0,165 (EXCEL =1-DISTR.NORM.ESTAND(0,976)) o sea del 16,5 %. Ese es el valor de β.

Ambos tipos de error están relacionados entre sí: Una disminución de α origina un aumento correspondiente de β. Por ejemplo, si α = 0,01, cuando no haya bias el intervalo será ahora 50,00 ± 2,57.0,82/$\sqrt{4}$ = 50,00 ± 1,05, es decir 48,95 ≤ μ ≤ 51,05. Cuando hay un bias de -1,20 el valor $|z|_{exp} = \frac{|48{,}95 - 48{,}80|}{0{,}82/\sqrt{4}} = 0{,}366$ La fracción de medidas con z > 0,366 es 0,357 (EXCEL =1-DISTR.NORM.ESTAND(0,366)) o sea del 35,7 %. Por tanto, se ha producido un aumento de β.



Para un α dado, cuanto mayor es el bias o diferencia entre μ_1 y μ_0, menor es β. Por el contrario, cuanto menor es el bias, más grande se hace β. Esto es lógico ya que cuanto más pequeño es un bias, más difícil es su detección.



Cuanto mayor es n, menores son α y β, ya que las distribuciones de medias muestrales se hacen más estrechas al disminuir la desviación típica que es $\sigma / \sqrt{n}$.

Se denomina **potencia o poder de una prueba a la probabilidad de rechazar correctamente H_0 cuando es falsa**. Puesto que β es la probabilidad de aceptar H_0 en esas condiciones, la potencia o poder de un test es **1 - β.**

En resumen:

H_0	CIERTA	Se acepta H_0	No hay error. P=1 - α
		Se rechaza H_0	Hay error de tipo I. P(I) = α
	FALSA	Se acepta H_0	Hay error de tipo II. P(II) = β
		Se rechaza H_0	No hay error. P=1 - β

La potencia de una prueba recibe a veces el nombre de sensibilidad, por ejemplo en ensayos clínicos, en los que una enfermedad se detecta en base a la presencia de un determinado marcador. En ese contexto, las consideraciones sobre los errores α y β, son muy importantes. La H_0 es que no hay enfermedad, por lo que α es la probabilidad de un **falso positivo** (decirle al paciente que está enfermo cuando no lo está). Por el contrario, β es la probabilidad de un **falso negativo** (decirle al paciente que está sano cuando en realidad está enfermo).

Ejemplo:

Un fabricante de gamusinos certifica que su contenido medio en luciferina es de 0,300 g con una desviación típica de 0,020 g

1) Se analizan 20 gamusinos y se encuentra un contenido medio de 0,285 g. Evalúe la certeza de la afirmación del fabricante al 95% y al 99%

Es un test de hipótesis de comparación de una media con un valor verdadero

$H_0 : \mu = \mu_0$
$H_1 : \mu \neq \mu_0$ (dos colas)
$\alpha = 0{,}05$
$z_{crit} = 1{,}96$ (Tabla 3.1) (Función EXCEL z_{crit} =|DISTR.NORM.ESTAND.INV(0,025)|)

$$z_{calc} = \left| \frac{0{,}285 - 0{,}300}{0{,}020 / \sqrt{20}} \right| = 3{,}35$$

$z_{calc} > z_{crit}$ Luego H_0 se rechaza

$H_0 : \mu = \mu_0$
$H_1 : \mu \neq \mu_0$ (dos colas)
$\alpha = 0{,}01$
$z_{crit} = 2{,}576$ (Tabla 3.1) (Función EXCEL z_{crit} =|DISTR.NORM.ESTAND.INV(0,005)|)

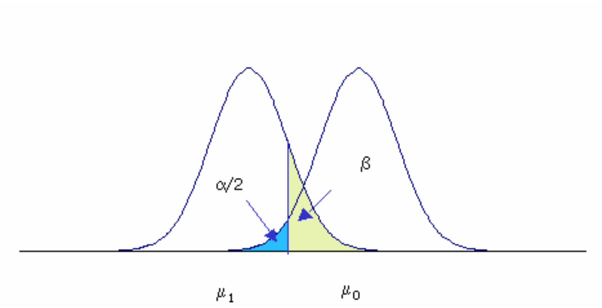
$$z_{calc} = \left| \frac{0{,}285 - 0{,}300}{0{,}020 / \sqrt{20}} \right| = 3{,}35$$

$z_{calc} > z_{crit}$ Luego H_0 se rechaza

2) Determine la probabilidad de cometer un error del tipo I en ambos casos

P(I) coincide con el nivel de significación α, que será del 5% en el caso primero y del 1 % en el caso segundo

3) Si se demuestra que nuestro método analítico tiene un bias de manera que la media de nuestras medias muestrales es 0,290 g, determine la probabilidad de error de tipo II y la potencia del test para tamaños de muestra n=9 y n=27, con un nivel de significación del 5%.



El valor de x desde el que se debe calcular β, coincide con el extremo inferior correspondiente a μ_0, el cual variará dependiendo del tamaño de muestra:

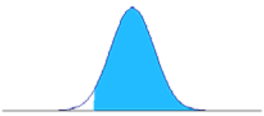
n = 9

El intervalo se extiende a $0{,}300 \pm 1{,}96 \dfrac{0{,}020}{\sqrt{9}} = 0{,}300 \pm 0{,}013 = 0{,}287 - 0{,}313$

Luego hay que buscar de z correspondiente 0,287 alrededor de la distribución de nuestras medias muestrales (que es 0,290 ya que tenemos un bias)

$$z = \left| \frac{0{,}287 - 0{,}290}{0{,}020 / \sqrt{9}} \right| = -0{,}45$$

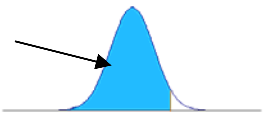
Necesitamos calcular



-0,45

La Tabla 3.3 nos da

P=0,674



0,45

Luego el error β tiene una probabilidad P(II) = 0,674 = 67,4 % y la Potencia = 32,6 % (Función EXCEL=DISTR.NORM.ESTAND(0,45)= 0,674)

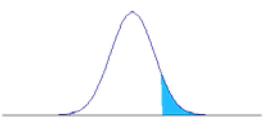
n = 27

El intervalo se extiende a $0{,}300 \pm 1{,}96 \dfrac{0{,}020}{\sqrt{27}} = 0{,}300 \pm 0{,}008 = 0{,}292 - 0{,}308$

Luego hay que buscar de z correspondiente 0,292 alrededor de la distribución de nuestras medias muestrales (que es 0,290 ya que tenemos un bias)

$$z = \left| \frac{0{,}292 - 0{,}290}{0{,}020 / \sqrt{27}} \right| = 0{,}520$$

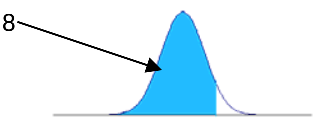
Necesitamos calcular



0,520

La Tabla 3.3 nos da:

P=0,698



0,520

Lo que nos interesa es 1- 0,698 = 0,302
(Función EXCEL=DISTR.NORM.ESTAND(0,520)= 0,698)

Luego el error β tiene una probabilidad P(II) = 0,302 = 30,2 % y la Potencia = 69,8 %

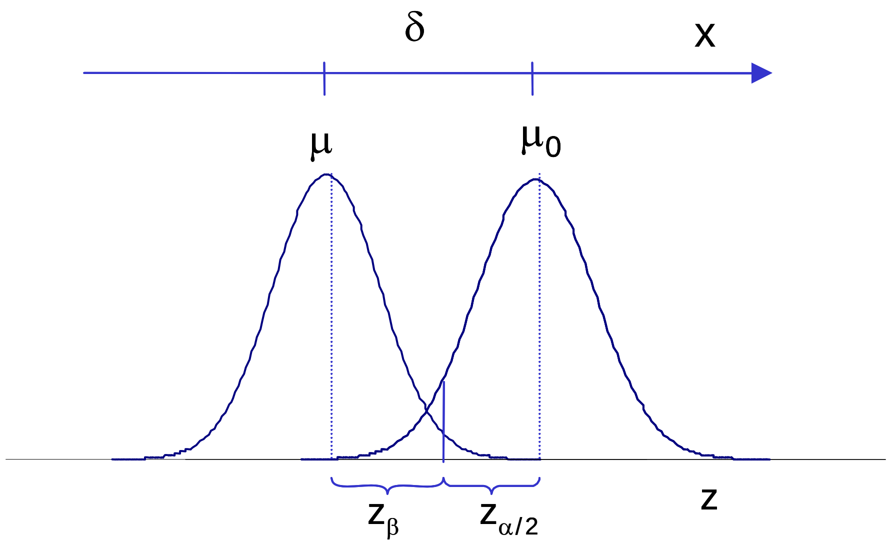
Importancia del tamaño de muestra

El intervalo de confianza de cualquier media puede disminuirse simplemente aumentando el tamaño de la muestra.

$$\mu = \bar{x} \pm 1{,}96\sigma / \sqrt{n}$$

Así aumentando cuatro veces el tamaño, el intervalo se reduce a la mitad, y esto es manipulable.

Hasta ahora hemos visto como se puede calcular el valor β una vez conocidos todos los demás valores: α, n, μ, σ y el bias $|\mu - \mu_0|$. Otra posibilidad más interesante es el poder calcular el **tamaño n de la muestra** que debe ser analizado para llevar a cabo nuestra prueba con la **suficiente confianza**, es decir para poder detectar una diferencia (bias) $\delta = |\mu - \mu_0|$ mediante un procedimiento que tiene una precisión σ (o s), de manera que haya no más de un α% de probabilidades de decidir que hay diferencia (bias) cuando no la hay, y al mismo tiempo una probabilidad de no más que β% de no detectar la diferencia (bias) cuando sí que existe.



Ya que la distancia $|\mu - \mu_0|$ en unidades z es $z_{\alpha/2}+z_\beta$, al transformarla en las unidades originales, resulta al reordenar:

$$n \geq \left[\left(z_{\alpha/2} + z_\beta\right)\sigma/\delta\right]^2$$

También pueden utilizarse gráficos publicados por la ISO.

PRUEBA O TEST DE UNA O DOS COLAS

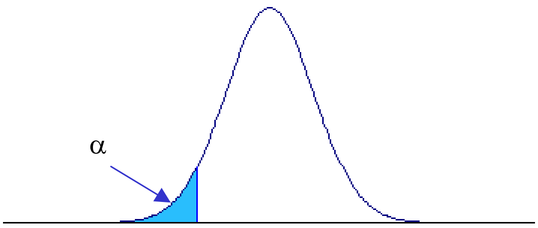
En el contexto de los ejemplos anteriores, tan malo era que el resultado del análisis fuera mayor ($\mu > \mu_0$) o menor ($\mu < \mu_0$) que el valor certificado. Por ello la hipótesis alternativa era siempre $\mu \neq \mu_0$. El test se llama de **dos colas pues se exploran ambos lados** de la distribución.

Pero hay ocasiones en que estamos solo interesados en saber si el valor hallado es mayor o menor que la referencia. Por ejemplo al comprobar la calidad de una materia prima, o al estudiar la estabilidad de un compuesto. El test es de **una cola pues solo se explora un lado** de una distribución.

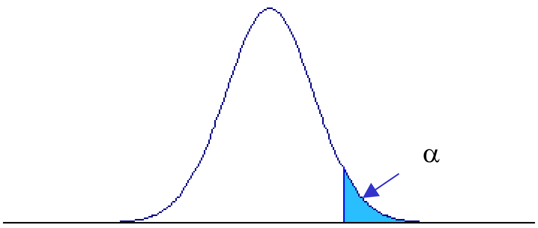
En resumen:

H_0: $\mu = \mu_0$ Siempre

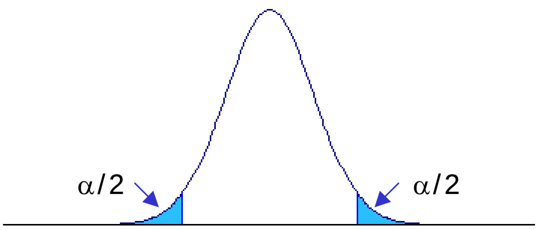
H_1: $\mu \neq \mu_0$ Test de dos colas
$\mu > \mu_0$ Test de una cola
$\mu < \mu_0$ Test de una cola



Una cola



Una cola



Dos colas

PRUEBA χ^2 DE NORMALIDAD

Además del prueba rankit que únicamente nos da una impresión visual, existen otras pruebas para evaluar la normalidad de una determinada población, que nos dan una estimación cuantitativa de la confianza del resultado.

El esquema que se sigue es el habitual de las pruebas de significación, pero el estadístico utilizado es χ^2:

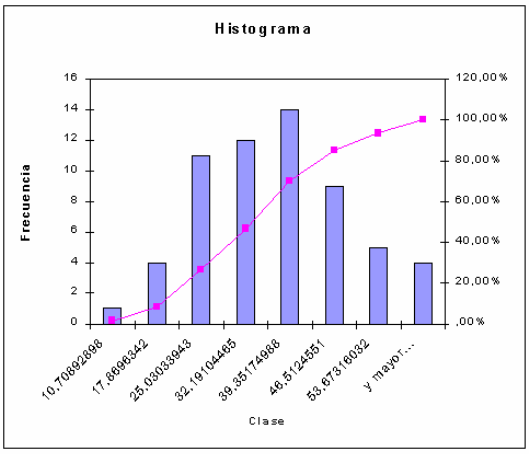
1) H_0: la muestra pertenece a una población normal
   H_1 : la muestra pertenece a una población que no es normal
2) Se elige el nivel de significación α, habitualmente 0,05
3) Se determina el valor crítico del estadístico χ^2_{exp}
4) Se calcula χ^2_{crit} el valor experimental del estadístico
5) Se compara χ^2_{exp} con χ^2_{crit}
6) Si $\chi^2_{exp} < \chi^2_{crit}$ se acepta H_0. En caso contrario se rechaza H_0(y se acepta H_1)

El procedimiento práctico parte de un histograma, es decir de un reparto de la muestra en clases de idéntico intervalo, y tiene las siguientes etapas:

1) Se calculan la media y la desviación típica
2) Se calcula el valor z del límite superior de cada clase
3) Se calcula la Frecuencia relativa acumulativa esperada para cada z (Tabla 3.3 o EXCEL =DISTR.NORM.ESTAND(z))
4) Se calculan las Frecuencias relativas esperadas por diferencia entre cada dos acumulativas
5) Se calcula la Frecuencia esperada (E_i) multiplicando lo anterior por el nº de datos
6) Se ve la Frecuencia observada (O_i) a partir de los datos del histograma en cada clase
7) Se calculan los $(O_i-E_i)^2/E_i$ para cada clase y se suman: SUMA
8) Se compara esa suma con el χ^2 crítico con el nivel de significación requerido (α = 0,05) y nº de g.d.l. = k -3 siendo k el nº de clases utilizadas.
(Tabla, o en EXCEL PRUEBA.CHI.INV(0,05;k-3))
9) Si SUMA < χ^2 Crítico, se mantiene H0
Si SUMA > χ^2 Crítico, se rechaza H0 y se acepta H1

Por ejemplo, sea el siguiente histograma:

Clase	*Frecuencia*	*% acumulado*
10,708929	1	1,67%
17,8696342	4	8,33%
25,0303394	11	26,67%
32,1910447	12	46,67%
39,3517499	14	70,00%
46,5124551	9	85,00%
53,6731603	5	93,33%
y mayor...	4	100,00%



Resultados del Ejemplo (Con EXCEL)

Inferior	*Superior*	*Frec (O)*	*z superior*	*Frec. Relat. acu espe*	*Frec. Relat. espe*	*Frec. espe (E)*	*$(O_i-E_i)^2/E_i$*
7,16070522	10,708929	1	-1,88198267	0,02991912	0,02991912	1,79514714	0,35220454
10,708929	17,8696342	4	-1,28987047	0,09854788	0,06862876	4,11772581	0,00336578
17,8696342	25,0303394	11	-0,69775828	0,24266411	0,14411623	8,64697394	0,64030859
25,0303394	32,1910447	12	-0,10564608	0,45793153	0,21526741	12,9160447	0,06496865
32,1910447	39,3517499	14	0,48646612	0,68668164	0,22875011	13,7250066	0,00550975
39,3517499	46,5124551	9	1,07857832	0,85961209	0,17293045	10,3758271	0,18243367
46,5124551	53,6731603	5	1,67069051	0,95260861	0,09299652	5,57979128	0,06024561
53,6731603	60,8338655	4	2,26280271	0,98817611	0,0355675	2,13404992	1,63153152

Media	33,4686749	χ^2 Exp	2,94056811
desviación	12,0934939	χ^2 Críti	11,0704826
n	60		

La distribución es Normal

RESUMEN DE PRUEBAS DE SIGNIFICACION ESTADÍSTICA
(TESTS DE HIPOTESIS)

PRUEBAS PARA EVALUAR LA EXACTITUD (DETECCION DE ERROR SISTEMATICO)			
Comparación de una media experimental con un valor conocido			
H_0: $\bar{x} = \mu_o$ H_1: $\bar{x} \neq \mu_o$ (test de 2 colas)			
σ conocida	$z = \dfrac{\lvert\bar{x} - \mu_o\rvert}{\sigma/\sqrt{n}}$	$z_{crit}\left(1-\frac{\alpha}{2}\right)$	
σ desconocida n>30	$z = \dfrac{\lvert\bar{x} - \mu_o\rvert}{s/\sqrt{n}}$	$z_{crit}\left(1-\frac{\alpha}{2}\right)$	
σ desconocida n<30	$t = \dfrac{\lvert\bar{x} - \mu_o\rvert}{s/\sqrt{n}}$	$t_{crit}\left(1-\frac{\alpha}{2},\nu\right)$	ν=n-1
Comparación de dos medias			
H_0: $\bar{x}_1 = \bar{x}_2$ (ó $\bar{x}_1 - \bar{x}_2 = 0$) H_1: $\bar{x}_1 \neq \bar{x}_2$ (ó $\bar{x}_1 - \bar{x}_2 \neq 0$) (test de 2 colas)			
σ_1,σ_2 conocidas	$z = \dfrac{\lvert\bar{x}_1 - \bar{x}_2\rvert}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$	$z_{crit}\left(1-\frac{\alpha}{2}\right)$	
σ_1,σ_2 desconocidas n_1,n_2>30	$z = \dfrac{\lvert\bar{x}_1 - \bar{x}_2\rvert}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$	$z_{crit}\left(1-\frac{\alpha}{2}\right)$	
σ_1,σ_2 desconocidas s_1,s_2 comparables n_1,n_2<30	$t = \dfrac{\lvert\bar{x}_1 - \bar{x}_2\rvert}{\sqrt{s^2\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$	$t_{crit}\left(1-\frac{\alpha}{2},\nu\right)$	$\nu = (n_1 - 1) + (n_2 - 1) = n_1 + n_2 - 2$ $s^2 = \dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)}$
σ_1,σ_2 desconocidas s_1,s_2 no comparables n_1,n_2<30	$t = \dfrac{\lvert\bar{x}_1 - \bar{x}_2\rvert}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$	$t_{crit}\left(1-\frac{\alpha}{2},\nu\right)$ Test de Cochran: $t^*_{crit} = \dfrac{t_1 \dfrac{s_1^2}{n_1} + t_2 \dfrac{s_2^2}{n_2}}{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}$ $t_1 = t_{crit(\upsilon_1 = n_1 - 1)}$ y $t_2 = t_{crit(\upsilon_2 = n_2 - 1)}$	$\dfrac{1}{\nu} = \dfrac{1}{(n_1 - 1)}\left(\dfrac{\dfrac{s_1^2}{n_1}}{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}\right) + \dfrac{1}{(n_2 - 1)}\left(\dfrac{\dfrac{s_2^2}{n_2}}{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}\right)$

Comparación de pares de valores (medidas apareadas)			
$d_i = x_{1i} - x_{2i}$	$H_0:\ \mu_d = 0$	$H_1:\ \mu_d \neq 0$	(test de 2 colas)
n>30	$z = \dfrac{\lvert\bar{d} - \mu_d\rvert}{s_d/\sqrt{n}} = \dfrac{\lvert\bar{d}\rvert}{s_d/\sqrt{n}}$	$z_{crit\left(1-\frac{\alpha}{2}\right)}$	
n<30	$t = \dfrac{\lvert\bar{d} - \mu_d\rvert}{s_d/\sqrt{n}} = \dfrac{\lvert\bar{d}\rvert}{s_d/\sqrt{n}}$	$t_{crit\left(1-\frac{\alpha}{2},\nu\right)}$	ν=n-1

Comparación de pares de valores cuando s depende de C y ésta varía en un amplio rango Comparación de dos métodos por regresión lineal			
$x_{2i} = a + bx_{1i}$	$H_0:\ a = 0\ ,\ b = 1$	$H_1:\ a \neq 0 \quad b \neq 1$	(test de 2 colas)
	$t = \dfrac{\lvert a - 0\rvert}{s_a}$ $t = \dfrac{\lvert b - 1\rvert}{s_b}$	$t_{crit\left(1-\frac{\alpha}{2},\nu\right)}$	ν=n-2

PRUEBAS PARA EVALUAR LA PRECISION			
Comparación de dos varianzas: Prueba F de Fischer			
$H_0:\ \sigma_1^2 = \sigma_2^2$	$H_1:\ \sigma_1^2 > \sigma_2^2$	(test de 1 cola)	
$s_1^2 \geq s_2^2$	$F = \dfrac{s_1^2}{s_2^2}$	$F_{crit(1-\alpha,\nu_1,\nu_2)}$	$\nu_1 = n_1 - 1$ $\nu_2 = n_2 - 1$

PRUEBAS DE RECHAZO DE VALORES DISCREPANTES ("OUTLIERS")			
Prueba Q de Dixon			
H_0: el valor sospechoso $\in N(\mu,\sigma^2)$	H_1: el valor sospechoso $\notin N(\mu,\sigma^2)$	(test de 2 colas)	
$Q = \dfrac{\lvert x_{sospechoso} - x_{próximo}\rvert}{x_{mayor} - x_{menor}}$	$Q_{crit\left(1-\frac{\alpha}{2},\nu\right)}$	ν=n	

<table>
<tr><th colspan="3">TESTS DE NORMALIDAD</th></tr>
<tr><th colspan="3">Prueba chi-cuadrado (χ^2)</th></tr>
<tr><td colspan="3">H_0: $X \in N(\mu,\sigma^2)$ H_1: $X \notin N(\mu,\sigma^2)$ (test de 2 colas)</td></tr>
<tr><td>$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$</td><td>$\chi^2_{crit(1-\alpha,\nu)}$</td><td>$\nu = k - 1 - h$
k, número de clases
h, número de parámetros de la distribución normal (2: μ y σ^2)</td></tr>
<tr><td colspan="3">Calcular la media y la desviación estándar de los datos experimentales
Dividir los datos experimentales en k clases, siendo $k \approx n/5$
Calcular las frecuencias O_i observadas en cada clase ($\Sigma O_i = n$).
Transformar los valores límite superiores (LS) de cada clase en valores z: $z = \frac{LS - \bar{x}}{s}$
Calcular la frecuencia acumulada para cada valor de z, y la frecuencia relativa de cada clase.
Calcular las frecuencias esperadas, E_i: $E_i = nf_i$, siendo f_i la frecuencia relativa de cada clase
Calcular para cada clase el valor $(O_i - E_i)^2/E_i$
Calcular χ^2_{exp} como suma de los valores anteriores</td></tr>
</table>

BIBLIOGRAFÍA

Massart D.L., Vandeginste B.G.M., Buydens L.M.C., De Jong S., Lewi P.J. and Smeyers-Verbeke J., *Handbook of Chemometrics and Qualimetrics*. Elsevier, Amsterdam, 1997

EJERCICIOS PROPUESTOS

1.- Se llevan a cabo determinaciones de un material de referencia internacional de concentración conocida en Pb (μ_0 = 340 μg) para validar un nuevo método de análisis. Se obtienen los siguientes resultados:
0,380; 0,346; 0,291; 0,278; 0,404; 0,331; 0,409; 0,285; 0,361; 0,268; 0,306; 0,243; 0,316; 0,223; 0,229
¿Es diferente la media obtenida del valor certificado con un nivel de significación α=0,05?

2.- Se desea conocer si un nuevo reactivo orgánico es estable o no en disolución. Para ello se prepara una disolución que contiene 250,0 mg del mismo. Al cabo de siete días se analizan alícuotas de la disolución encontrándose los siguientes resultados:
$\bar{x} = 247{,}1$; s= 3,5 ; n=9
¿Ha disminuido la cantidad de reactivo de forma significativa?

3.- Se desea comparar dos procedimientos de digestión previos a la determinación de nitrógeno en harina. Para ello se analiza una misma muestra mediante los dos procedimientos obteniendo los siguientes resultados:

Procedimiento 1: $\bar{x}_1 = 2{,}05\,g/100\,g$; $s_1^2 = 0{,}050$; $n_1 = 30$

Procedimiento 2: $\bar{x}_2 = 2{,}21\,g/100\,g$; $s_2^2 = 0{,}040$; $n_2 = 32$

1) Se desea conocer si ambos procedimientos originan o no resultados comparables
2) Se sospecha que el procedimiento 1 origina pérdidas por volatilización ¿es cierto?

Utilice en todos los casos un nivel de significación α=0,05
Repita el estudio pero con n_1=8 y n_2=7 y el resto de los datos idénticos
Repita con los nuevos datos

Procedimiento 1: $\bar{x}_1 = 2{,}05\,g/100\,g$; $s_1^2 = 0{,}050$; $n_1 = 9$

Procedimiento 2: $\bar{x}_2 = 2{,}21\,g/100\,g$; $s_2^2 = 0{,}010$; $n_2 = 8$

4.- Para validar los resultados de un laboratorio se procedió a un ejercicio de inter-comparación en el cual se analizaron 10 muestras diferentes que se analizaron también en un laboratorio externo y acreditado (la calidad de sus resultados está asegurada). A partir de los resultados indique hay diferencias significativas entre ambos laboratorios, con un nivel de significación α=0,05.

Muestra	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Laboratorio	5,1	4,7	5,8	3,7	4,5	5,3	6,0	5,1	5,7	4,8
Laboratorio. Ref.	5,3	4,9	5,8	3,8	4,4	5,4	5,9	5,0	5,9	4,7

5.- Se desea validar un método voltamperométrico basado en la redisolución anódica (ASV) para la determinación de Pb en zumos de frutas naturales. Para ello se utiliza como referencia el método basado en AAS con cámara de grafito (ETAAS). Se preparan una serie de muestras dopadas con cantidades variables de Pb, que se determinan mediante los dos procedimientos. Se obtienen los siguientes resultados. ¿Se puede decir que ambos métodos originan los mismos resultados?

Muestra	**ASV**	**ETAAS**
1	35	35
2	70	75
3	80	75
4	80	80
5	120	125
6	200	205
7	220	205
8	200	215
9	250	240
10	330	350

6.- Se desea validar un procedimiento para la determinación de un estimulante prohibido por la U.C.I. en orina. Para ello se dopan muestras de orina de un sujeto que no ha consumido el estimulante, con cantidades conocidas del mismo, obteniéndose los resultados de la tabla. Indicar si el método propuesto está libre o no de bias (con un nivel de confianza del 95%)

Analito puesto (μg)	**Analito hallado(μg)**
0	0,5
1	1,3
3	3,5
10	10,3
30	33,9
100	104,1
300	316,1
1000	1037,0

7.- Comparación de varianzas
Aplíquese la prueba F al ejemplo de la determinación de Nitrógeno en harina en los dos siguientes casos:

1) Procedimiento 1: $\bar{x}_1 = 2{,}05\,g/100\,g$; $s_1^2 = 0{,}050$; $n_1 = 8$

   Procedimiento 2: $\bar{x}_2 = 2{,}21\,g/100\,g$; $s_2^2 = 0{,}040$; $n_2 = 7$

2) Procedimiento 1: $\bar{x}_1 = 2{,}05\,g/100\,g$; $s_1^2 = 0{,}050$; $n_1 = 9$

   Procedimiento 2: $\bar{x}_2 = 2{,}21\,g/100\,g$; $s_2^2 = 0{,}010$; $n_2 = 8$