# 13

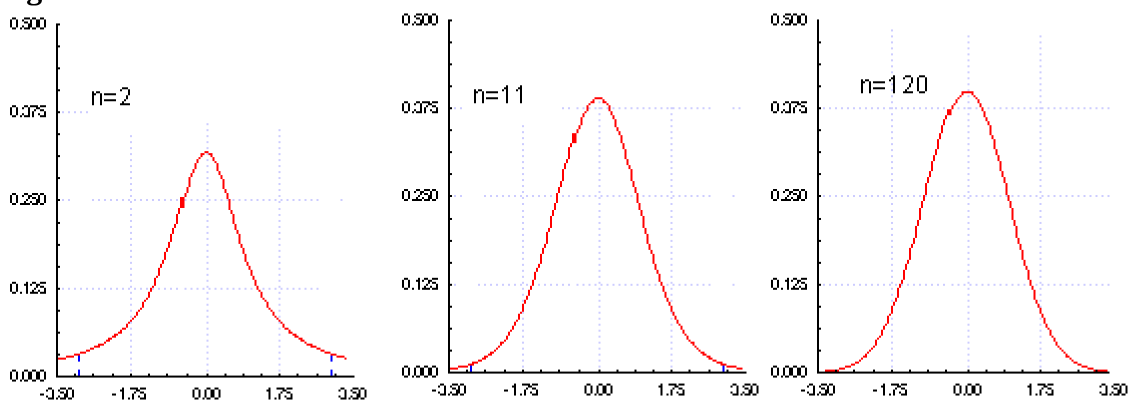
# Teoría de pequeñas muestras

En este capítulo se presentan tres nuevos modelos estadísticos: el llamado *t* de Student, el modelo de la *Chi-cuadrado* ( $\chi^2$ ) y el modelo *F* de Fisher. Los tres no requieren ya más del supuesto de un tamaño muestral grande. Ahora con dos o más mediciones se puede trabajar; por eso se usa la expresión *Teoría de pequeñas muestras* para este tema. El empleo de cualquiera de ellos es enteramente similar al visto en el capítulo anterior. Cambia la manera de calcular el estadígrafo de comparación y su respectiva tabla de valores críticos de la distribución muestral. Mientras que el modelo de la *t* se aplica a medias y proporciones, los dos últimos se usan para el estudio de las desviaciones o dispersiones. También se la llama Teoría Exacta del Muestreo, pues ahora no hay que efectuar la aproximación $DS^2 \cong \sigma^2$ ya que el valor muestral viene en la fórmula de cálculo del estadígrafo de comparación, en lugar del poblacional. Eso hace que no sea necesario efectuar una estimación y se tiene una mayor exactitud que con la gaussiana. Es importante destacar que los tres modelos son válidos tanto para pequeñas como para grandes muestras. Esto amplía el campo de aplicación del modelo de Gauss. Además, al no tener que hacer tantas pruebas disminuye el costo y se gana en tiempo. Todas estas ventajas tienen una contrapartida: se pierde un poco de precisión pues, como se verá, el intervalo de confianza se hace más grande para un mismo caso. Estos modelos se prefieren al de Gauss porque sus ventajas valen la pena, al precio de perder un poco de precisión. Se mostrará su empleo tanto para el caso de una sola muestra de mediciones como para la comparación de dos muestras o grupos de mediciones.

## 13.1 El modelo de Student

Sea un estadígrafo t calculado para la media con la relación: $t = \dfrac{(\bar{x} - \mu)}{DS/\sqrt{n}}$

**Figura 13.1: La distribución de Student**



Si de una población normal, o aproximadamente normal, se extraen muestras aleatorias e independientes y a cada una se le calcula dicho estadígrafo usando los valores muestrales de la media y el desvío estándar, entonces se obtiene una distribución muestral *t* que viene dada por la fórmula de Student. En realidad, fue obtenida por R. A. Fisher y la bautizó *Student* en honor a W. S. Gosset, quien usaba ese seudónimo para poder publicar sus trabajos en la revista *Biometrika.* Esta función matemática tiene un parámetro que la define en forma unívoca: el número de grados de libertad $\upsilon = n$-1 (donde *n* es el tamaño muestral). El concepto matemático de $\upsilon$ está relacionado con la cantidad de observaciones independientes que se hagan y se calcula con el tamaño muestral *n,* menos la cantidad *k* de parámetros poblacionales que deban ser estimados a través de ellas. O sea: $\upsilon = n-k$. Si se observa la ecuación superior, se ve que el único parámetro poblacional que figura es $\mu$, por lo tanto k = 1 y así resulta $\upsilon = n-1$. Cuando el tamaño muestral es mayor que 30 la distribución de Student se aproxima mucho a la de Gauss, en el límite ambas son iguales. Es decir que la función Student tiende asintóticamente a la función de Gauss.

Para cada grado de libertad hay una tabla de valores que pueden obtenerse variando el nivel de significación, parecida a la de Gauss. Sería muy engorroso tener una hoja con la tabla para cada grado de libertad. Esto se soluciona de dos formas: una es usando computadoras para resolver los cálculos (programas estadísticos como Mini-Tab, SPSS, Statistica, Excel, etc.). La otra y más común, es preparar una tabla donde en cada fila se coloquen encolumnados los valores críticos más usuales para cada valor de grados de libertad. Como interesan únicamente los valores pequeños, se listan correlativamente de 1 a 30 y luego algunos como 40, 60, 120 e $\infty$. Este último tendrá los valores vistos para la normal. Así, en una sola hoja se presentan los valores útiles para el empleo de este modelo, como se muestran en el Tabla 5 del Anexo con las tablas estadísticas.

La distribución de Student, al igual que la de Gauss, es simétrica respecto al origen de coordenadas y se extiende desde $-\infty$ hasta $+\infty$. Pero a diferencia de la normal, puede adoptar diferentes formas dependiendo del número de grados de libertad. Por ejemplo, la que tiene un solo grado de libertad (n = 2 y $\upsilon = 1$), se desvía marcadamente de la normal, como se puede ver en la Figura 13.1 anterior. Luego, a medida que los grados van aumentando, se acerca cada vez más, hasta igualarla en el infinito. Se puede ver esto en las tablas y en la Figura 13.1. Los valores críticos de la Tabla Student, para una confianza del 95 % y dos colas, para 1, 5, 10, 30 y $\infty$ grados de libertad son 12,71; 2,57; 2,23; 2,04 y 1,96 respectivamente. Estos valores críticos se denotan con sus dos parámetros así: $t_{\alpha\,;\,\upsilon} = t_{0,05\,;\,\infty} = 1{,}96 = z_\alpha$.

Los intervalos de confianza para esta distribución se arman en forma análoga a la vista para el caso de Gauss. Con la única diferencia en cómo se calcula el valor crítico $t_{\alpha;\upsilon}$ en lugar de $z_\alpha$.

$$\boxed{\mu \in (\mu_e \pm t_{\alpha;\upsilon}\ SE(e)) = (\mu_e \pm t_{\alpha;\upsilon}\ \sigma_e)}$$ *Intervalo de confianza con Student*

De nuevo, el par de valores ( $\mu_e$ ; $\sigma_e$ ) se saca de la Tabla 4, con la salvedad que ahora no se usa más la aproximación DS $\cong \sigma$; pues en el cálculo de *t* se emplea DS directamente. Esto hace que el modelo sea más exacto que el de Gauss. Generalmente, este modelo se aplica al caso de la media, proporciones y sus diferencias o sumas. Para una estimación con 30 o más grados de libertad, se pueden usar tanto el modelo de Gauss, como el de Student. El intervalo es casi igual, salvo que en este último el valor crítico es mayor. En efecto, si se tienen 31 muestras, t = 2,09, mientras que z = 1,96. Esto hace mayores a los intervalos obtenidos con Student que sus equivalentes gaussianos. Por eso, se dice que el modelo Student tiene menor *precisión* que el de Gauss.

La teoría de decisiones se usa en forma análoga, empleando los intervalos de confianza visto más arriba. Pero para poder aplicar este modelo se deben tener en cuenta los requisitos siguientes:

1) *Las muestras fueron extraídas de una población normal o aproximadamente normal.*
2) *La selección de las muestras se hizo en forma aleatoria.*
3) *Las muestras son independientes entre sí.*

Si alguno de ellos no se cumple, las conclusiones que se obtengan no son válidas. Los supuestos se pueden resumir así: *para poder usar Student, se deben tener muestras normales, aleatorias e independientes.* Notar que el error estándar de estimación es SE ($e$) = $\sigma_e$.

Los casos más frecuentes en la práctica son:

# 13.1.1 Student para medias muestrales

En este caso $e = \bar{x}$ luego: $\mu_e = \mu$ y SE ($e$) = $\sigma_e$ = DS / $\sqrt{n}$ . Por lo tanto el valor de comparación se calcula con:

$$\boxed{t = \frac{(\bar{x} - \mu)}{DS/\sqrt{n}}}$$

*Ejemplo 1*) Se desea saber si un instrumento de medición cualquiera está calibrado, desde el punto de vista de la exactitud. Para ello se consigue un valor patrón y se lo mide 10 veces (por ejemplo: una pesa patrón para una balanza, un suero control para un método clínico, etc.). Suponiendo que el resultado de estas mediciones arroja una media de 52,9 y un desvío de 3, usando un patrón de valor 50, se debe determinar si el instrumento está calibrado y la estimación de su error sistemático, si es que se prueba su existencia (no se usan unidades para generalizar este ejemplo).

Ho : $\mu = 50$ el instrumento está calibrado en exactitud
$H_1$ : $\mu \neq 50$ no está calibrado. Hay un error sistemático

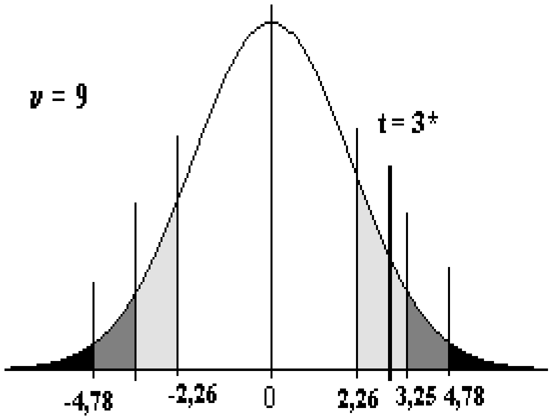
Se trata de un ensayo de dos colas donde hay $\nu = 10 - 1 = 9$ grados de libertad. De la Tabla 4 se obtienen los valores críticos para el 95% de $t_{0,05;\,9} = 2,262$, para el 99% de $t_{0,01;\,9} = 3,25$ y para un nivel del 99,9% es $t_{0,001;\,9} = 4,781$. Lo que permite establecer las zonas de aceptación y rechazo:

$$t = \frac{(52,9 - 50)}{3/\sqrt{10}} = 3^* \ (p = 0,00135)$$

$\mu = 50 \notin$ 95% CI(50,8 ; 55,1) ; o bien
$\bar{x} = 52,9 \notin$ 95% CI(47,9 ; 52,1)

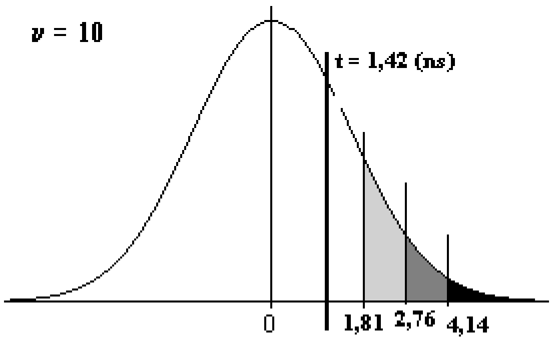
Dibujando las zonas con los valores críticos, el valor de t cae en la de rechazo para el 95% y no alcanza para las otras. La conclusión es que se ha probado la existencia de un error sistemático con una confianza del 95%. Y se estima con:

$$ES \in (\bar{x} - \mu) = 2,90$$



*Ejemplo 2)* Se midió colesterol total a 11 pacientes varones adultos escogidos al azar los resultados obtenidos arrojan una media de 235 mg/dl y un desvío estándar de 35 mg/dl. Ensayar la hipótesis de que se mantienen por debajo del valor límite de referencia ( 220 mg/dl ).

$H_0$ : $\mu \leq 220$ mg/dl
$H_1$ : $\mu > 220$ mg/dl




El valor de Student para una sola cola es:

$$t = \frac{(235 - 220)}{35/\sqrt{11}} = 1{,}42 \text{ (ns)} \quad (p = 0{,}093)$$

Valor no significativo pues $t_{0,05;\,10} = 1{,}81$
$\mu = 220$ cae dentro de 95% CI (216 ; 254)

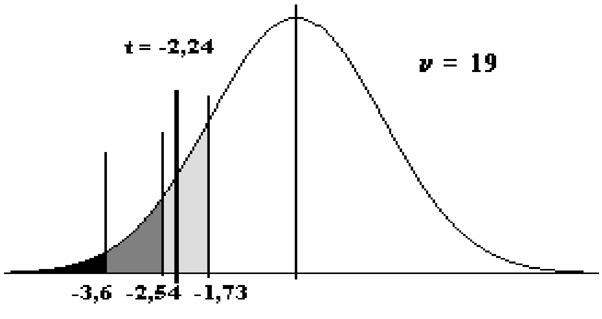
Para el caso de una cola, el valor de tablas para el 95% debe ser el que está en la Tabla 4 para el 90% en dos colas. La idea es que el 10% en dos colas significa el 5% en cada una, por la simetría de la curva de Student. Luego, para $\nu = 10$, el límite para el 95% será t = 1,812 en una cola y t = 2,228 para dos colas. En la figura de más arriba se han marcado los límites del 99% y del 99,9% para una sola cola, a los efectos didácticos. La conclusión es que no puede rechazar la hipótesis nula, por lo que debe considerarse un colesterol total admisible desde el punto de vista clínico, por estar por debajo del límite de referencia.

# 13.1.2 Student para proporciones

En este caso $e = P$ y $\mu_p = \mu = \pi$ luego con SE $(e) = \sigma_\pi = \sqrt{\pi.(1-\pi)/n}$ se puede obtener el valor del estadígrafo de comparación con la relación:

$$\boxed{t = \frac{(P - \pi)}{\sqrt{\pi.(1-\pi)/n}}}$$

*Ejemplo1)* Un analgésico de plaza, afirma en su propaganda que alivia el dolor en el 90% de los casos antes de la primer hora luego de su ingesta. Para validar esa información, se hace un experimento en 20 individuos con cefalea. Se observa que fue efectivo en 15 de ellos.
$H_0$ : $\mu \geq 0{,}9$ La afirmación es correcta
$H_!$ : $\mu < 0{,}9$ La afirmación es falsa




Es un ensayo de una sola cola (a izquierda)
El porcentaje de éxitos es: P = 15 / 20 = 0,75.
Con $\mu = 0{,}9$ y $\sigma_p = \sqrt{0{,}9 \,.\, 0{,}1 / 20} = 0{,}067$

$$t = \frac{(0{,}75 - 0{,}9)}{0{,}067} = -\,2{,}24^* \quad (p = 0{,}019)$$

$\mu = 0{,}9$ cae fuera del CI 95% (- ∞ ; 0,78 )

De tablas: $t_{0,999;\,19} = -\,3{,}579$ $t_{0,99;\,19} = -\,2{,}539$
y $t_{0,95;\,19} = -\,1{,}729$

El resultado obtenido es significativo (t = - 2,24 *). Pero la evidencia no alcanza para rechazar la hipótesis a los niveles del 99% y 99,9%. Se la rechaza al nivel de 95% únicamente. Si bien no es tan terminante, se puede afirmar que la aseveración es falsa con un 95% de confianza.

# 13.2 Student para dos muestras independientes

El modelo de Student también se puede usar cuando se desean comparar dos muestras entre sí, para detectar si hay diferencia significativa entre ellas, debido a algún factor analizado. En primer lugar se analizará el caso de dos muestras independientes como: aplicar dos tipos de remedios a dos grupos de pacientes escogidos al azar, o las mediciones repetidas de una misma magnitud, etc. El otro caso, cuando las muestras no son independientes sino apareadas, se verá en el próximo tema. Una vez más, los supuestos para poder aplicar este modelo se resumen en: *para poder comparar con Student, las dos muestras deben ser normales, aleatorias e independientes.*

Se sacan muestras aleatorias e independientes, de dos poblaciones normales. La idea es averiguar si ambas muestras provienen de la misma población o de poblaciones diferentes. Con eso se puede ver si el efecto de los "tratamientos" aplicados a las muestras es *apreciable,* en cuyo caso las muestras parecerán provenir de diferentes poblaciones. Se usa en los casos donde se compara el efecto de una droga aplicada a un grupo de pacientes, contra otro grupo al cual se le suministra un placebo. También para comparar dos técnicas clínicas y detectar si hay diferencias, por ejemplo: dos marcas comerciales de plaza, dos instrumentos de medición, dos individuos, dos técnicas diferentes (la nueva contra la vieja), dos protocolos, etc. Con estas comparaciones se pueden realizar muchos controles internos en el laboratorio para hacer calibraciones, medir eficacia, etc. Hay una limitación: *solo se pueden comparar dos muestras entre sí a la vez y no más*. Para el caso de tener más de dos muestras, se recurre a los modelos de ANOVA como se verá más adelante.

## 13.2.1 Comparación de medias

Para estos casos, el valor de Student para validaciones de medias se calcula con:

$$\boxed{t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{DS_1^2}{n_1}\right) + \left(\frac{DS_2^2}{n_2}\right)}}}$$

El cual se contrasta con $t_{\alpha;\upsilon}$ donde $\upsilon = n_1 + n_2 - 2$ grados de libertad. Hay casos particulares como (a) las muestras son de igual tamaño y (b) son *homocedásticas* (tienen igual varianza). En ambos casos se simplifican las fórmulas de cálculo.

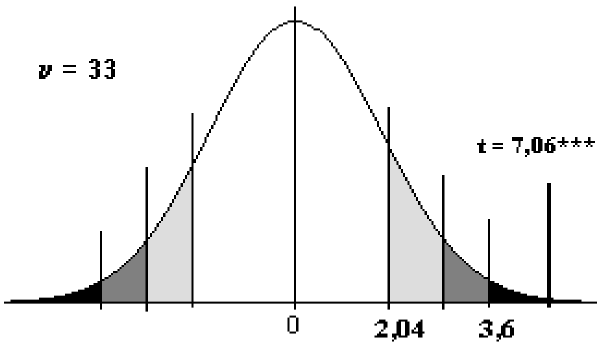
*Ejemplo 1)* Se aplica un medicamento a 15 pacientes que padecen cierta enfermedad, escogidos al azar, y un placebo a 20 pacientes. En el primer grupo, la desaparición del estado febril se observa a las 19 horas de tratamiento en promedio (con un desvío de 2 hs.). En el grupo control, la mejoría se observa en promedio a las 25 horas con un desvío de 3 horas. Decidir si el medicamento modifica el tiempo de curación.

$H_0$ : $\mu_1 = \mu_2$ : el medicamento es inocuo
$H_1$ : $\mu_1 \neq \mu_2$ : el medicamento produce efecto

Es un ensayo de dos colas donde los valores críticos se buscan en la Tabla 5 interpolando entre 30 y 40 grados de libertad.

$$t = \frac{(25-19)-(0)}{\sqrt{\left(\frac{9}{20}\right)+\left(\frac{4}{15}\right)}} = 7{,}06^{***}$$

$\mu_1 - \mu_2 = 0$ cae fuera de CI 99,9% (3,1; 8,9).



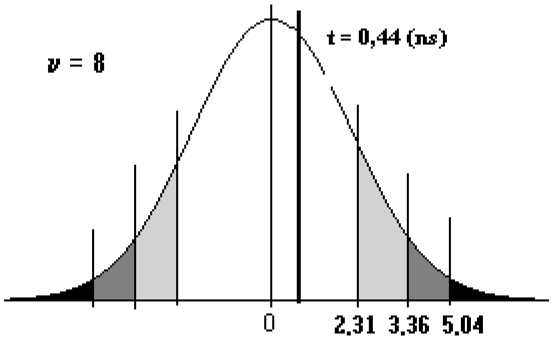

Como el valor hallado de t es mucho más grande que el valor crítico de tablas para 33 grados de libertad: $t_{\alpha\,;\,\upsilon} = t_{0,999;\,33} = 3{,}44$ (ensayo de dos colas y un 99,9% de confianza), la conclusión es: se obtuvieron resultados altamente significativos ( t = 7,06 *** ) como para rechazar la hipótesis nula. Se tiene una prueba científica del efecto del medicamento.

*Ejemplo 2)* Se desea verificar si hay diferencia en las mediciones a través de dos métodos clínicos diferentes. Se toma una muestra de suero lo suficientemente grande como para obtener 10 alícuotas. Se distribuyen al azar 5 alícuotas para cada método. Efectuadas las mediciones, con el primero se tuvo una media de 85 mg/dl con un desvío de 8 mg/dl. Mientras que con el segundo se tuvo una media de 83 mg/dl con un desvío de 6 mg/dl.

$H_0$ : $\mu_1 = \mu_2$ : no hay diferencia.
$H_1$ : $\mu_1 \neq \mu_2$ : hay diferencia

$$t = \frac{(85-83)-(0)}{\sqrt{\left(\frac{64}{5}\right)+\left(\frac{36}{5}\right)}} = 0{,}44 \text{ (ns)}$$

No se puede rechazar la Ho. Se concluye que no hay diferencia entre ambos métodos.
$\mu_1 - \mu_2 = 0$ cae dentro de 95% CI(-8,3 ; 12,3)




# 13.2.2 Comparación de proporciones

Para estos casos, el valor de Student para validaciones de proporciones se calcula la misma fórmula, pero reemplazando los valores esperados con:

$$\mu_{1\text{-}2} = (\pi_1 - \pi_2) \qquad \text{y} \qquad \sigma^2_{1\text{-}2} = [\pi_1 (1-\pi_1)/ n_1] + [\pi_2 (1-\pi_2) / n_2]$$

Entonces el valor de comparación del modelo Student para este caso es:

$$\boxed{t = \frac{(P_1 - P_2) - (\pi_1 - \pi_2)}{\sqrt{(\pi_1 (1-\pi_1)/n_1) + (\pi_2 (1-\pi_2)/n_2)}}}$$

Contrastando con el valor de tablas dado por $t_{\alpha;\upsilon}$; con $\upsilon = n_1 + n_2 - 2$ grados de libertad.

*Ejemplo*) Se escogen al azar dos grupos formados por 20 individuos cada uno, entre los que padecen cierta alergia. Se administra una droga curativa al primer grupo y se observa una mejoría en 15 de los casos. Al segundo grupo se le administra un placebo y mejoran 13 de ellos. Ensayar la hipótesis que la droga sirve para curar ese tipo de alergia. Se emplean las hipótesis siguientes:

$H_0$ : $\mu_{1-2} = 0$ las diferencias observadas se deben al azar.
$H_1$ : $\mu_{1-2} \neq 0$ la droga produce efecto.

Si se supone que ambas muestras fueron extraídas de la misma población, y por lo tanto no hay diferencias entre las muestras observadas ($H_0$) $\mu_{1-2} = 0$, eso significa que el porcentaje de curados en dicha población será $\pi = \pi_1 = \pi_2$ y habrá que estimarlo con los datos muestrales, calculando la proporción ponderada con: $p$ = ( total de curados en las muestras / total muestral ) = (15+13) / 40 = 0,7. Entonces, sacando factor común en la fórmula de la varianza, esta resulta:

$$SE^2(\pi) = \pi\,(1-\pi).[\,(1/n_1) + (1/n_2)\,] = \pi\,(1-\pi)\,[2\,/\,n] = (0{,}7\ .\ 0{,}3)\,(2/20) = 0{,}021$$

Y es SE($\pi$) = 0,145; de los datos del problema surgen $P_1$ =15/20 = 0,75 y $P_2$ = 13/20 = 0, 65

$$t = (\ 0{,}75 - 0{,}65\ )\ /\ (0{,}021)^{1/2} = 0{,}69 < t_{0,95\ ;\ 38} = 2{,}02.$$ $\mu_{1-2} = 0$ cae dentro de 95% CI(-0,19 ; +0,39)

Un resultado no significativo. Las diferencias observadas no se deben a la droga sino al azar.

# 13.2.3 Test de equivalencia biológica

Hay ocasiones donde la Ho no busca establecer si hay o no diferencia entre dos muestras, como las del ejemplo anterior, sino que se trata de establecer si un método clínico o tratamiento nuevo es lo suficientemente bueno como para reemplaza al que se venía usando hasta entonces, el método viejo. Las ventajas de este nuevo método pueden ser: un costo menor, más rápido, menos dañino o peligroso para el paciente, etc. La cuestión básica aquí es ver si, en promedio, la diferencia entre ambos es menor que un cierto valor límite para la magnitud estudiada. Es decir que tal diferencia no implique una inferioridad del nuevo método, desde un punto de vista clínico. Para estos casos la Ho : La diferencia entre ambos promedios es mayor o igual al valor aceptable y la alternativa es H1 : Esta diferencia de medias es menor al valor crítico; en cuyo caso ambos métodos pueden ser considerados clínicamente equivalentes. La idea es que, si se rechaza la Ho se puede usar el método nuevo en lugar del viejo y aprovechar las ventajas que este posee. Pero la decisión se basa más en consideraciones médicas que estadísticas. Entonces, si se trata de magnitudes continuas, se puede usar el test de Student para comparar la diferencia de las dos medias contra el valor crítico $\delta$ o máximo aceptable desde el punto de vista clínico. El planteo se hace así: Ho : $\mu_V - \mu_N = \Delta \geq \delta$. Donde $\mu_V$ es el valor poblacional que se obtiene con el método viejo y $\mu_N$ con el método nuevo, $\Delta$ es la diferencia real entre ambos métodos y $\delta$ es la diferencia máxima admisible entre ambos métodos. De esta manera, cuando Ho pueda rechazarse se tendrá evidencia suficiente como para efectuar el reemplazo, esto es cuando H1 : $\mu_V - \mu_N = \Delta < \delta$.

Se trata de un ensayo de una sola cola. Pero cuando se trate de ver si en valor absoluto la diferencia entre ambos métodos no supere a un cierto valor $\delta$, porque aquí no interesa tanto que sea menor, sino que también interesa que no sea mayor (dependiendo de la magnitud clínica analizada); entonces la Ho será : $\mu_V - \mu_N = \Delta = \delta$ y el ensayo será de dos colas. Análogo al visto en el punto anterior. Para ilustrar este procedimiento se usará un ejemplo tomado de la obra de Armitage (pág. 320).

*Ejemplo*) Sea el índice cardíaco CI (respuesta cardiaca normalizada para la superficie del cuerpo) el cual se mide con un procedimiento invasivo como es el colocar un catéter en el corazón del paciente llamado Termo-dilución (el método viejo) y la unidad de medición son litros por minuto tomado por $m^2$ de superficie del cuerpo humano. Se ha propuesto una nueva manera de medir esa magnitud con una técnica no invasiva, llamada el método de la Bioimpedancia, en la cual se le adosa un instrumento al cuerpo de paciente en forma externa, y mide en forma eléctrica el valor del CI usando una escala adecuada (el método nuevo). El criterio clínico de aceptación es: el nuevo método se considerará equivalente al viejo cuando, en promedio, el valor obtenido difiera en un 20% respecto al promedio aceptado de 2,75 l / min. / $m^2$ para el método del catéter. Esto significa que el 20% de tal valor es $\delta = 0{,}55$. Luego el planteo se hace así:

Ho : $|\mu_V - \mu_N| = |\Delta| \geq \delta = 0{,}55$ o lo que es lo mismo $(\mu_V - \mu_N) = \delta = 0{,}55$

H1 : $|\mu_V - \mu_N| = |\Delta| < \delta = 0{,}55$ cuyo equivalente es $(\mu_V - \mu_N) = \delta \neq 0{,}55$

Se toma una muestra de N = 96 individuos a los cuales se le aplica el método nuevo, los valores encontrados fueron un promedio de 2,68 l / min. / $m^2$, y un desvío estándar de 0,26 l / min. / $m^2$ luego será:

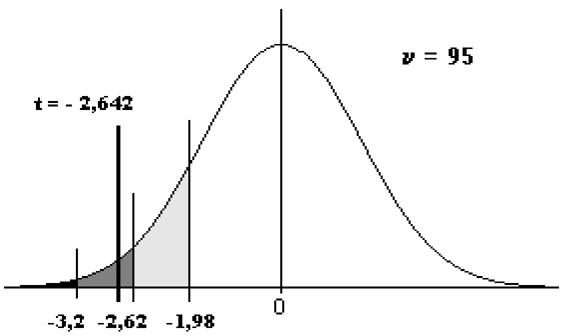
$$t = \frac{(\bar{x} - \mu)}{DS/\sqrt{n}} = (2{,}68 - 2{,}75) / (0{,}26 / \sqrt{96}) = -0{,}07 / 0{,}0265 = -2{,}642^{**} > t_{0{,}99\,;\,95} = -2{,}62$$

Lo que indica que hay evidencia significativa como para rechazar la Ho.

Otra manera de ver lo mismo es cuando se usa el valor límite de confianza, para la diferencia en valor absoluto, la cual es: 0,07. Para un 95% de confianza el error de estimación se calcula con:

$$t_{0{,}95\,;\,95} \,.\, (DS / \sqrt{N}) = 0{,}053$$

Luego 0,07 + 0,053 = 0,123 < 0,55 con lo que se puede rechazar la Ho.




La conclusión final es que se puede usar el método nuevo en lugar del viejo, con una gran ventaja para el paciente, pues ahora ya no tendrá que ser cateterizado para efectuarle su medición del índice cardíaco. A este procedimiento estadístico aparecido en los últimos años en Medicina se lo conoce también con el nombre de *test de equivalencias médicas o biológicas*.

# 13.3 Student para dos muestras apareadas

El modelo de Student se puede usar para el caso especial de muestras *apareadas*, esto es, cuando se le efectúan dos tratamientos a la misma muestra; por ejemplo, del tipo *antes–después* donde al mismo individuo se lo mide dos veces para ver el efecto del tratamiento realizado, o el caso de método nuevo contra el método viejo, donde al mismo grupo de pacientes se le hacen dos mediciones a cada uno, la del método de rutina habitual y una extra con el nuevo método a probar para decidirse entre ambos. La idea básica es como sigue: se sacan n muestras aleatorias e independientes de una población normal. A cada muestra se le aplican dos " tratamientos" A y B diferentes y lo que interesa detectar es si producen algún efecto apreciable. Este caso es muy diferente al anterior (13.2), si bien las muestras son independientes entre sí, los tratamientos no lo son, porque a un mismo individuo se le aplican ambos tratamientos. Entonces, la misma persona aparecerá dos veces en los resultados: uno en el grupo A y el otro en el grupo B. El truco para resolver este problema de la independencia es trabajar con la diferencia de los resultados de cada par de mediciones efectuadas: $d = x_A - x_B$. Luego se tendrán n diferencias $d_1$; $d_2$; $d_3 ... d_n$, que son independientes entre sí, puesto que cada valor $d_i$ corresponde a un solo individuo. Luego, se le aplica el modelo Student para una sola muestra, ensayando la hipótesis de que no hay diferencias entre ambos grupos. O sea, efectuando la hipótesis: $H_0 : \mu_d = 0$ resultará:

$$t = \frac{\bar{d} - 0}{DS_d / \sqrt{n}}$$

La hipótesis alternativa implica un efecto diferente para cada grupo $H_1 : \mu_d \neq 0$. Si se prueba que el valor esperado del promedio de las diferencias es diferente de cero, entonces el tratamiento aplicado produce un efecto demostrable. Para aclarar estas ideas se presenta el siguiente caso:

*Ejemplo)* Se escogen 5 pacientes al azar, del grupo que concurre diariamente al Laboratorio de Análisis Clínicos a efectuarse una determinación de Uremia. Las muestras extraídas se miden con el procedimiento habitual y además con una nueva técnica clínica que se desea probar. Ver si hay diferencia entre ambas técnicas. Los resultados expresados en g/l fueron:

| Pacientes | Vieja | Nueva | Diferencias |
|---|---|---|---|
| 1 | 0,38 | 0,33 | 0,05 |
| 2 | 0,54 | 0,45 | 0,09 |
| 3 | 0,22 | 0,15 | 0,07 |
| 4 | 0,11 | 0,09 | 0,02 |
| 5 | 0,23 | 0,22 | 0,01 |

Promedio = 0,048

Desvío estándar = 0,033

Con los valores de las diferencias se calculan $\bar{d} = 0{,}048$ g/l y $DS_d = 0{,}033$ g/l. Luego:

$$t = \frac{0{,}048}{0{,}033 / \sqrt{5}} = 3{,}25^* > t_{0{,}95\,;\,4} = 2{,}776.$$ O bien, el valor 0 cae fuera del CI 95% (0,01; 0,09)

Se obtuvo evidencia significativa t = 3,25 * que hay diferencia entre ambas técnicas.

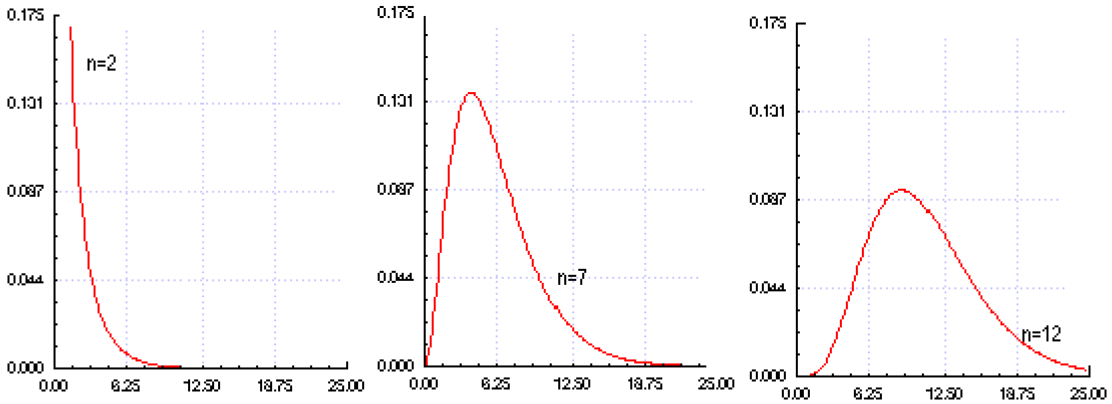
# 13.4 El modelo de Chi-cuadrado

La función Chi-cuadrado es igual a la función normal elevada al cuadrado. Esto es, el producto de dos distribuciones de Gauss es una distribución de Chi-cuadrado. Si de una población normal, o aproximadamente normal, se extraen muestras aleatorias e independientes, y se le calcula el estadígrafo $\chi^2$ usando el valor muestral de la varianza y el poblacional con:

$$\boxed{\chi^2 = ( n - 1 ) \ DS^2 \ / \ \sigma^2}$$

La distribución muestral de $\chi^2$ viene dada por la fórmula de K.R. Pearson. Esta función matemática está caracterizada por el valor del número de grados de libertad $\upsilon = n - 1$ (donde n es el tamaño muestral). Al igual que la Student, el valor total del área bajo la curva es igual a la unidad, pero la diferencia principal es que esta no es simétrica respecto al origen, sino que se extiende desde 0 hasta $+ \infty$ porque no puede ser negativa. A medida que los grados de libertad aumentan, la curva cambia de forma y sus valores se han tabulado en el anexo de tablas estadísticas (Tabla 6), donde se muestran los valores del área bajo la curva, para los principales valores de $\chi^2$ , a la derecha de éste. O sea, se muestra la zona de rechazo para diferentes niveles de significación y de grados de libertad, lo cuales varían entre 1 y 100. Más allá, conviene usar directamente la función de Gauss.

**Figura 13.2: La distribución de Chi-cuadrado**




Para cada grado de libertad hay una tabla de valores que pueden obtenerse variando el nivel de significación, parecida a la de Gauss. El problema de calcular los valores críticos, para un nivel de confianza dado, se resuelve de dos maneras: usando computadoras para resolver los cálculos, y la otra más común, usando tablas resumidas en una sola hoja como la que se muestra en la Tabla 6, en forma análoga a la vista para el modelo de Student.

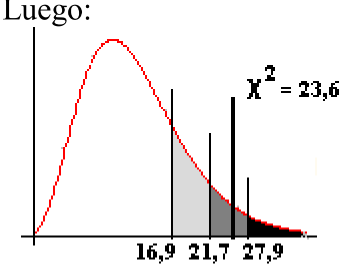
La distribución de $\chi^2$ se usa principalmente para analizar dispersiones. Se compara la dispersión muestral expresada a través de sus cuadrados medios (MS) contra la dispersión poblacional cuantificada a través de la varianza ($\sigma^2$). El valor $MS = ( n - 1 ) DS^2 = \nu \ DS^2$ es otra forma de mostrar la precisión del sistema de medición. El uso más difundido en Bioquímica es para controlar la dispersión de la técnica de Análisis Clínicos empleada en el Laboratorio; se compara la obtenida en forma experimental, contra un valor considerado como aceptable en los libros de texto denominado la *dispersión máxima admisible* ($\sigma_{máx}$). Usualmente, se toma el CV%, como se mostró en el ejemplo de varianzas visto en el capítulo anterior. Existen otros criterios, como el de Thonks, que usa un error relativo admisible máximo, y se calcula como un cuarto del rango de los valores normales de referencia, dividido por el valor medio de dicho intervalo (referido a la magnitud clínica en cuestión y expresado en porcentajes). Todo esto se puede ver con mayor detalle en la Tabla 24.1 del Capítulo 24. También se emplea a este modelo para realizar la llamada *prueba de chi-cuadrado* en las comparaciones de frecuencias observadas contra las frecuencias esperadas, con datos de recuento. Más adelante se desarrolla mejor este tema, lo mismo que su uso para testear la independencia de dos o más factores en una Tabla de Contingencia.

En la industria farmacéutica se la usa para analizar la dispersión de los componentes de los productos terminados. Todo remedio fabricado debe cumplir estrictas normas de calidad, generalmente referidas al contenido en peso de sus principales componentes. Se usan dos límites: el superior e inferior, dentro de los cuales se los debe mantener controlados. Este rango de valores define la dispersión máxima admisible y lo ideal es que la dispersión de los productos terminados sea bastante inferior a dicho rango. Ese control de la dispersión es muy similar al explicado más arriba, para los bioquímicos. Para ilustrar estas ideas se presenta un ejemplo en Control de Calidad de equipos de Laboratorio. Pero antes se debe remarcar que para poder aplicar este modelo, se deben tener en cuenta los requisitos siguientes:

1. *Las muestras fueron extraídas de una población normal o aproximadamente normal.*
2. *La selección de las muestras se hizo en forma aleatoria.*
3. *Las muestras son independientes entre sí.*

*Ejemplo 1*) Un bioquímico sospecha que su micro-centrífuga no mantiene constante su velocidad mientras trabaja, lo cual le da una variabilidad indeseada en sus determinaciones. Para controlarla, consigue un tacómetro regulado y mide cada minuto la velocidad durante 10 minutos. Los resultados fueron: una velocidad promedio en las 10 mediciones de 3098 rpm con un desvío de 100,4 rpm. Testear para un error relativo máximo del 2% o menos, si la centrífuga es estable.

El desvío estándar aceptable es: $\sigma_{máx}$ = 2% de 3098 rpm = 62 rpm. Luego:




$H_0 : \sigma_{máx} \leq$ 62 rpm: la micro centrífuga es estable
$H_1 : \sigma_{máx} >$ 62 rpm: la micro centrífuga no es estable

$\chi^2 = (n-1)\ DS^2 / \sigma^2$
$\chi^2 = (10-1)\ (100{,}4)^2 / (62)^2 = 23{,}6^{**}$

De la Tabla de valores críticos surge: $\chi^2_{0,99\ ;\ 9} = 21{,}666$ y $\chi^2_{0,991\ ;\ 9} = 27{,}877$. Por lo tanto, el bioquímico ha encontrado una muy fuerte evidencia que la velocidad del equipo oscila en forma indeseada, tal como sospechaba. Y deberá ajustarlo si desea disminuir la variabilidad de sus mediciones. Los resultados fueron muy significativos $\chi^2 = 23{,}6^{**}$.

*Ejemplo 2*) Un farmacéutico Jefe del Dpto. Control de Calidad en una industria alimenticia, descubre que en su proceso de producción el contenido de ciclamato en su línea de mermeladas dietéticas varía en forma indeseada. Sospechando que se trata de una falla en el dosificador, decide tomar 10 muestras seguidas del mismo. Encuentra un promedio de 20 gramos con un desvío de 8 gramos. Si en su protocolo de fabricación la variación máxima permitida es del 3%, determinar si el dosificador debe ser corregido.

El desvío estándar aceptable es: $\sigma_{máx}$ = 3% de 20 g = 6 g. Luego:

$H_0$ : $\sigma_{máx} \leq$ 6 g.: el dosificador funciona correctamente
$H_1$ : $\sigma_{máx} >$ 6 g.: el dosificador debe ser cambiado

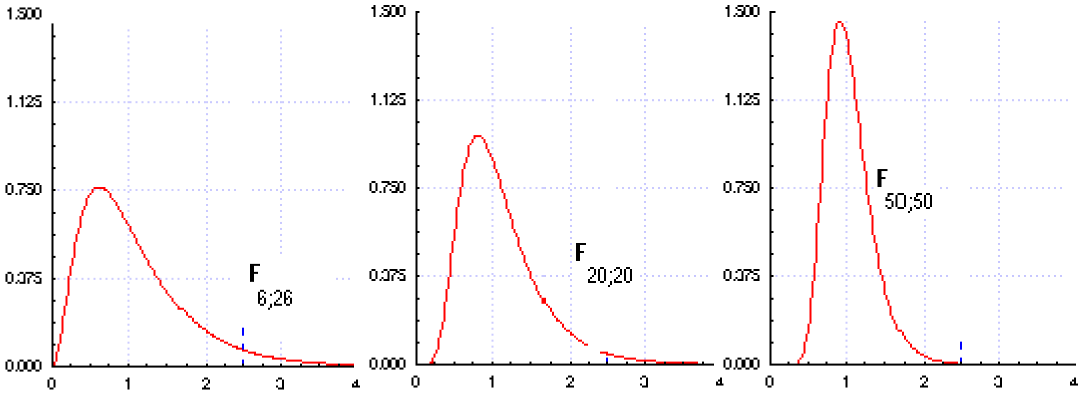
$\chi^2 = ( n - 1 )\ DS^2\ /\ \sigma^2 = (10 - 1)\ .\ (8)^2\ /\ (6)^2 = 16$ (no significativo)

De la Tabla de valores críticos surge: $\chi^2_{0,95\ ;\ 9} = 16,9$. Por lo tanto, el farmacéutico no ha encontrado evidencia que respalde sus sospechas. Sin embargo, el valor hallado es muy cercano al crítico, por lo que le convendría hacer más pruebas.

# 13.5 El modelo de Fisher

Si de dos poblaciones normales, o aproximadamente normales, se extraen dos muestras aleatorias e independientes, y a cada una se le calcula su respectiva varianza, el cociente de ambos valores $F = DS^2_1 / DS^2_2$ (con F > 1, esto es, siempre se coloca el más grande como numerador) tendrá una distribución de Fisher, cuyos valores críticos fueron obtenidos por W. Snedecor y se muestran en la Tabla 7 del anexo. Esta tabla se caracteriza por tener dos grados de libertad: el correspondiente al numerador $\upsilon_1 = n_1 - 1$ y el del denominador $\upsilon_2 = n_2 - 1$. Programas de computación permiten calcular los valores críticos respectivos. En otra forma se puede usar una tabla de doble entrada como la Tabla 7; su forma se puede ver a continuación:

**Figura 13.3: La distribución de Fisher.**

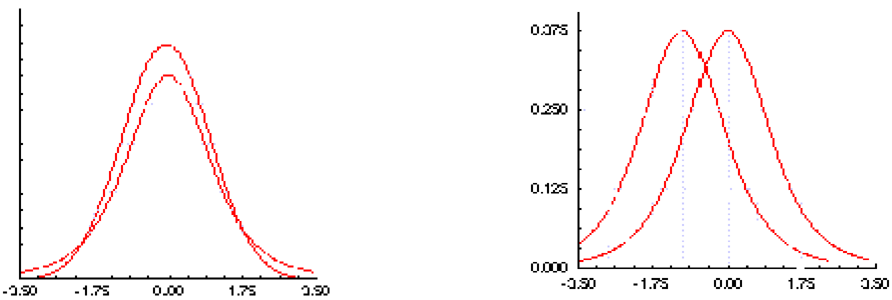



En las Tabla 7 se presenta una hoja para cada nivel de confianza, se eligen los más apropiados como: 95% ; 97,5% ; 99% ; 99,5% y 99,9%. Como siempre, el área total bajo la curva es la unidad y se extiende desde 0 a + ∞. La forma es muy parecida a la Chi-cuadrado. En la Figura 13.3 se muestran tres casos, con diferentes grados de libertad, y se marca el valor de F = 2,5 con una línea punteada vertical.

El principal uso de esta función es el Análisis de Varianza, que se verá más adelante, y es para cuando se necesita comparar más de dos medias muestrales a la vez. En estos casos la idea es detectar si el efecto de uno o más tratamientos afecta a las muestras testeadas. En cambio, cuando se tiene el caso de dos muestras, la idea es testear si hay homoscedasticidad en las dos poblaciones en estudio. Una vez verificado este supuesto, se puede avanzar más verificando si hay diferencia entre las medias muestrales, y así verificar si ambas muestras tienen igual media y varianza, porque eso significa que en realidad provienen de la misma población normal. Eso probaría que no hay efecto de un tratamiento si se lo compara con un placebo, o que dos técnicas de laboratorio son equivalentes. Si el experimento no verifica esto, entonces se deberá elegir el caso que presente menor varianza, para tener menor variabilidad en las mediciones. En Genética se puede verificar si una generación de crías es más variable en un carácter que la de sus padres. En Sistemática se puede testear si dos poblaciones locales tienen la misma variabilidad. En Bioquímica y Farmacia el uso más frecuente es comparar el error casual de mediciones de laboratorio, al introducir algún efecto o cambiar el método de medición.

En el caso de testear si dos técnicas de laboratorio tienen igual dispersión, o bien, para elegir aquella con mayor precisión, conviene pensar el problema como la incidencia de un factor en estudio en lugar de dos técnicas totalmente diferentes entre sí. Por ejemplo, se trata de una misma práctica, pero se usan dos espectrofotómetros diferentes, y se trata de determinar si la modificación de la varianza se debe al uso de un aparato diferente. El factor acá sería: tipo de espectros. También se puede estudiar la incidencia del factor humano, realizando las mismas mediciones a dos personas diferentes. De esa forma se puede imaginar que las dos muestras provienen de diferentes poblaciones, o que el efecto del factor analizado no es despreciable cuando se rechaza la hipótesis nula. En la Figura 13.4 se muestra el caso de dos poblaciones. En el caso (a) ambas poblaciones tienen la misma media, pero por efecto del error casual sus varianzas son diferentes. Si esta diferencia es significativa, resulta evidenciada por el Modelo de Fisher que permite la comparación de ambas.

**Figura 13.4: Dos poblaciones con**




**(a) Igual media, distinta varianza** **(b) Distinta media, igual varianza**

En el caso (b) hay un error sistemático que desplaza la media, pero sus varianzas permanecen iguales. Es lo mismo que sumar una constante a todos los valores; ocurre un desplazamiento hacia la derecha. Student se usa para detectar esto cuando se hace el test de comparación de dos medias independientes.

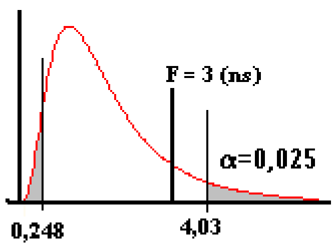
Como se verá más adelante, se puede construir todo un bagaje de métodos para efectuar un Control de Calidad interno en un laboratorio de medición clínica. Por ahora, basta decir que se puede controlar la exactitud con los modelos de Student y la precisión con los de Chi-cuadrado y Fisher. Con esto se pueden comenzar a controlar y calibrar los sistemas de medición. Las limitaciones de todo esto son dos: la primera es que se puede estudiar el efecto del factor analizado en solo dos muestras y no en más de dos. La segunda es que si la calidad se entiende como exactitud y precisión, solo se pueden emplear estos modelos para magnitudes de tipo cuantitativas como las de la Química Clínica, pero no en magnitudes cualitativas como las usuales en Microbiología, Bacteriología, Micología, etc. En magnitudes cuantitativas, por "calidad " se entiende precisión y exactitud, en lugar de la capacidad de una prueba clínica para diagnosticar. Sin embargo, a pesar de estas limitaciones sigue siendo una herramienta sencilla y poderosa de control.

Para poder aplicar este modelo se deben tener en cuenta los requisitos siguientes:

1. *Las muestras fueron extraídas de una población normal o aproximadamente normal.*
2. *La selección de las muestras se hizo en forma aleatoria.*
3. *Las muestras son independientes entre sí.*

*Ejemplo*) El jefe de un laboratorio se encuentra con una técnica de medición fuera del control estadístico. Para investigar las causas decide investigar si el factor humano tiene incidencia, y toma una muestra de suero cualquiera la divide en 20 alícuotas. Luego elige 10 de ellas al azar y se las entrega al laboratorista 1 para que haga las determinaciones; las restantes las encomienda al laboratorista 2 para que las mida. Los resultados obtenidos son: $DS^2_1 = 2,4$ es la varianza obtenida por el laborista, 1 y $DS^2_2 = 0,8$ para el otro. Decidir si hay diferencia en dispersión entre ambos.

$H_0 : \sigma^2_1 = \sigma^2_2$: no hay diferencia y el factor humano no incide
$H_1 : \sigma^2_1 \neq \sigma^2_2$: hay diferencia entre ambas personas




Se calcula el estadígrafo de comparación de Fisher

$F = DS^2_1 / DS^2_2 = 2,4 / 0,8 = 3$

Como se trata de un ensayo de dos colas, para un nivel del 95% de confianza, se busca en las tablas para: $\upsilon_1 = \upsilon_2 = n_1 - 1 = 9$ grados de libertad, mientras que $\alpha = 0,025$ para el límite inferior y $\alpha = 0,975$ para el superior. Estos valores son:

$F_{0,975; (9,9)} = 4,03$.

Luego, para calcular el valor no tabulado $\alpha = 0,025$ se aprovecha una propiedad que tiene la función F usando la inversa como sigue:

$F_{0,025; (9,9)} = 1 / F_{0,975; (9,9)} = 1/ 4,03 = 0,248$

Como el valor hallado F = 3 cae dentro de la zona de aceptación, no hay evidencia significativa como para decir que el factor humano tiene incidencia en la dispersión de las mediciones.

# 13.6 Cuadro resumen

**Test de hipótesis para las medias muestrales.**

Se puede siempre usar el modelo Student. El de Gauss cuando la muestra sea grande (n > 30)

Muestra vs. Población

$\sigma^2$ conocido $\quad Z = \dfrac{(\bar{x} - \mu)}{\sigma/\sqrt{n}} \quad$ vs. $Z\alpha$

$\sigma^2$ desconocido $\quad t = \dfrac{(\bar{x} - \mu)}{DS/\sqrt{n}} \quad$ vs. $t_{\alpha;\upsilon}$ con $\upsilon = n - 1$

*Muestra vs. Muestra*:

<u>Caso 1</u>: Muestras independientes con $\sigma_1^2$ y $\sigma_2^2$ son desconocidos

Si $\sigma_1^2 \neq \sigma_2^2$ es $\quad t = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\left(\dfrac{DS_1^2}{n_1}\right) + \left(\dfrac{DS_2^2}{n_2}\right)}}$ vs. $t_{\alpha;\upsilon}$ con $\upsilon = \dfrac{\left[(DS_1^2/n_1) + (DS_2^2/n_2)\right]^2}{\dfrac{(DS_1^2/n_1)^2}{n_1 - 1} + \dfrac{(DS_2^2/n_2)^2}{n_2 - 1}}$

Si $\sigma_1^2 = \sigma_2^2$ es $\quad t = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}\sqrt{\dfrac{(n_1 - 1)DS_1^2 + (n_2 - 1)DS_2^2}{n_1 + n_2 - 2}}}$ vs. $t_{\alpha;\upsilon}$ con $\upsilon = n_1 + n_2 - 2$

<u>Caso 2</u>: Muestras independientes con $\sigma_1^2$ y $\sigma_2^2$ son conocidos

Si $\sigma_1^2 \neq \sigma_2^2$ es $\quad Z = [\,(\bar{x}_1 - \bar{x}_2) - (\mu\,\bar{x}_1 - \bar{x}_2)\,]\ /\ \sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$ vs. $Z\alpha$

Si $\sigma^2 = \sigma_1^2 = \sigma_2^2$ es $\quad Z = [\,(\bar{x}_1 - \bar{x}_2) - (\mu\,\bar{x}_1 - \bar{x}_2)\,]\ /\ \sigma\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}$ vs. $Z\alpha$

<u>Caso 3</u>: Las muestras son apareadas y se suponen de una misma población con $DS_d \approx \sigma^2$

Se toma $d_i = X_1 i - X_2 i$ entonces $t = \dfrac{\bar{d} - 0}{DS_d / \sqrt{n}}$ vs. $t_{\alpha;\upsilon}$ con $\upsilon = n - 1$

**Test de hipótesis para proporciones**

Se puede siempre usar el modelo Student. El de Gauss cuando la muestra sea grande (n > 25)

*Muestra versus población:* $\pi$ conocido o estimado con $p = r / n$

Gauss: $Z = \dfrac{p - \pi}{\sqrt{\pi.(1 - \pi) / n}}$ o $Z = \dfrac{r - n.\pi}{\sqrt{n.\pi.(1 - \pi)}}$ vs. $Z\alpha$

Student: $t = \dfrac{(P - \pi)}{\sqrt{\pi.(1 - \pi) / n}}$ o $t = \dfrac{(r - n.\pi)}{\sqrt{n.\pi.(1 - \pi)}}$ vs. $t_{\alpha;\upsilon}$ con $\upsilon = n - 1$

*Muestra vs. Muestra*: $\pi_1$ y $\pi_2$ conocidos

$$t = \frac{(P_1 - P_2) - (\pi_1 - \pi_2)}{\sqrt{(\pi_1(1 - \pi_1) / n_1) + (\pi_2(1 - \pi_2) / n_2)}} \quad \text{vs.} \quad t_{\alpha;\upsilon}; \text{ con } \upsilon = n_1 + n_2 - 2$$

*Ambas muestras provienen de la misma población*: $\pi_1 = \pi_2 = \pi$

Se estima $\pi = [(p_1 n_1) + (p_2 n_2)] / (n_1 + n_2)$ y se reemplaza en

$$t = \frac{(P - \pi)}{\sqrt{\pi.(1 - \pi) / n}} \quad \text{o} \quad t = \frac{(r - n.\pi)}{\sqrt{n.\pi.(1 - \pi)}} \quad \text{vs.} \quad t_{\alpha;\upsilon} \text{ con } \upsilon = n - 1$$

**Test de hipótesis para varianzas**

*Muestra vs. Población*: $\chi^2 = (n - 1)\ DS^2 / \sigma^2$ vs. $\chi^2_{\alpha;\nu}$ con $\upsilon = n_1 - 1$

*Muestra vs. Muestra*: $F = DS^2_1 / DS^2_2$ vs. $F_{\alpha;\upsilon 1;\upsilon 2}$

con $DS^2 > DS^2_2$ y $\upsilon_1 = n_1 - 1$ ; $\upsilon_2 = n_2 - 1$

# 13.7 Significación clínica versus estadística

Antes de continuar es necesario realizar una reflexión sobre lo visto hasta ahora, en particular sobre las diferencias conceptuales entre la significación estadística tratada hasta aquí, y lo que se necesita en Medicina que es la significación clínica de los resultados. En 1986/88 Gardner y Altman presentaron un informe en el *BMJ* donde señalaban el peligro del avance inusitado de la estadística en los estudios clínicos. En ese entonces se había llegado a un punto donde, de los resultados obtenidos se informaba el valor (P-value) de probabilidad encontrado en el resumen (por ejemplo: p = 0,027), y a veces se olvidaban de poner lo más importante que eran los resultados hallados (por ejemplo, si comparaban dos muestras en lugar de poner en el "abstract" el valor de la diferencia hallada, ponían el valor de p). La recomendación fue indicar con claridad todos los datos posibles como ser:

"La diferencia entre las medias muestrales de presión sistólica de un grupo de 100 pacientes diabéticos y otros 100 no diabéticos, fue de 6 mm Hg, con un intervalo de confianza del 95% entre 1,1 y 10,9 mm Hg. El estadístico t = 2,4 resultó significativo con 198 grados de libertad, con un valor asociado de p = 0,02."

O resumido con: "promedio 6 mm Hg, 95% CI (1,1 ; 10,9), t = 2.4 con $\nu$ = 198 y p = 0,02"

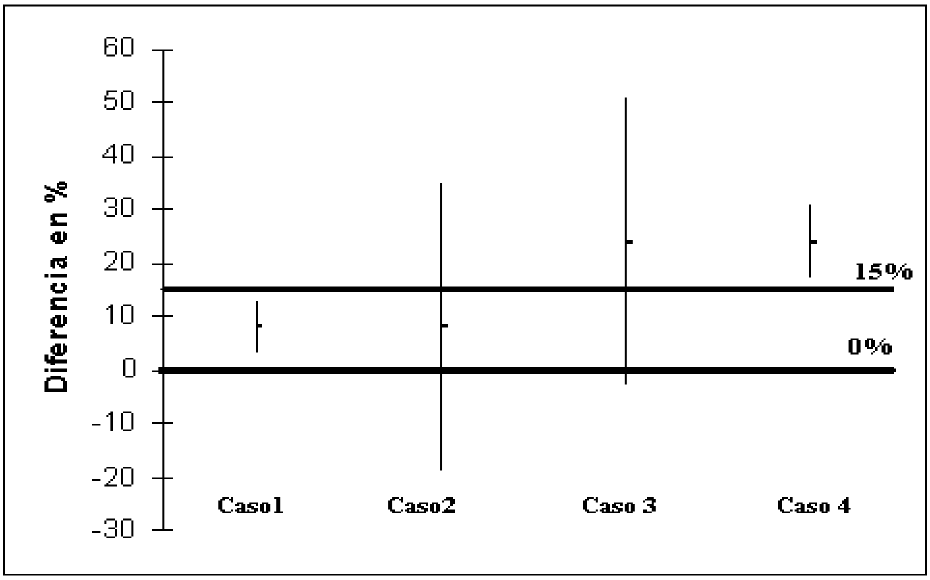
A partir de ese entonces, esta es la forma recomendada para presentar la *significación estadística* en los resultados de las investigaciones clínicas. Sin embargo, falta una cuestión vital para el lector de estos temas: ¿ Esa diferencia hallada que implica clínicamente hablando ? Una primera respuesta es la deducción de que los pacientes diabéticos tienen una presión sistólica más alta que los no diabéticos. Pero la principal cuestión clínica es: ¿ Son hipertensos o no ? Y esta pregunta no se puede deducir de la información presentada más arriba. Falta la investigación de la *significación clínica* de los resultados.

En el año 1991, Braitman pone de manifiesto esta situación en un informe publicado en los Anales de Medicina Interna de USA, presentando un ejemplo sencillo que se muestra a continuación:

En una investigación sobre los resultados de un nuevo tratamiento para un cáncer, se trata de comparar la mejoría de los pacientes a los cuales se les aplicó el nuevo tratamiento, con otro grupo a los cuales se les aplicó el tratamiento viejo. En un primer estudio se tomaron 800 pacientes en cada grupo y se observó que 480 individuos respondieron al tratamiento nuevo, mientras que 416 lo hicieron con el viejo. Entonces, el 60% de los pacientes (480/800) respondieron a la nueva droga, mientras que el 52% (416/800) respondieron con la vieja. Se trata de una comparación entre dos proporciones estadísticamente hablando, y su significación se puede calcular con el modelo de comparación de dos muestras independientes de Student o Gauss. Si se efectúan los cálculos se puede ver que hay una diferencia altamente significativa (p = 0,001), para la diferencia de proporciones hallada del 8%, que tiene un 95% CI (3 ; 13)%. Esto significa, que hay prueba de que la diferencia entre ambos tratamientos existe y que está entre el 3% y el 13% (su mejor estimación es el 8%). Luego se hizo un estudio semejante en otro lugar, pero solo se utilizaron 50 pacientes en total, y los resultados obtenidos fueron: 15 mejorías con la nueva y 13 con la vieja. Si se aplica ahora el mismo tratamiento estadístico para los datos, hay que emplear el modelo Student porque las muestras son pequeñas, y se halla un porcentaje de mejoría del 60% para la nueva (15/25) y un 52% (13/25) para la vieja. Nuevamente la diferencia hallada es del 8%, pero esta vez los resultados no son significativos (p = 0,57) y no se puede hablar de una esta vez los resultados no son significativos (p = 0,57) y no se puede hablar de una diferencia real entre ambos tratamientos, el intervalo de confianza de las diferencias es: 95% CI (-19 ; 35)%. Como el valor 0% (Ho) cae dentro del intervalo, no hay prueba de que el nuevo es mejor que el viejo, es decir no se puede rechazar la hipótesis nula. Se repitió el estudio nuevamente para 50 pacientes y los resultados fueron: Nuevo 60% (15/25) y viejo 36% (9/25), entonces ahora la diferencia entre las proporciones es del 24% con un 95% CI (-3 ; 51)% y el valor p = 0,09 tampoco es significativo. En un lugar diferente, se repite nuevamente la comparación en 800 pacientes y los resultados se muestran en la última línea del cuadro siguiente:

| | **Porcentaje que responden Al tratamiento** | | **Valor P** | **Significación** | **Diferencia entre los métodos en %** | |
|---|---|---|---|---|---|---|
| **Nº** | **NUEVO** | **VIEJO** | (p-value) | **Estadística** | **Puntual** | **95% CI** |
| **1** | 480/800=**60%** | 416/800=**52%** | 0,001 | **SI** | **8%** | 3% a 13% |
| **2** | 15/25 = **60%** | 13/25 = **52%** | 0,57 | **NO** | **8%** | - 19% a 35% |
| **3** | 15/25 = **60%** | 9/25 = **36%** | 0,09 | **NO** | **24%** | - 3% a 51% |
| **4** | 240/400=**60%** | 144/400=**36%** | <0,001 | **SI** | **24%** | 17% a 31% |

Si se comparan estos cuatro estudios, se puede ver que dos de ellos son estadísticamente significativos y los otros con menor tamaño muestral no lo son. Se consulta a un panel de expertos que estudian los costos de ambos tratamientos, la mejoría real en los pacientes, el tiempo empleado, la posible remisión de la enfermedad, el riesgo, etc.; quienes concluyen que: Para que valga la pena el cambio de tratamiento, el porcentaje de pacientes mejorados no puede ser menor del 15% (criterio clínico). Ahora, con ese valor crítico mínimo admisible se comparan los cuatro casos:




Se concluye que solo en el cuarto caso hay una diferencia clínica y estadística como para cambiar de tratamiento. La *diferencia significativa clínica* no existe en el primer caso, y es dudosa en los casos 2 y 3. Notar, que *no es lo mismo hablar de significación clínica que estadística*.

# 13.8 Problemas propuestos

1) Marcar la respuesta correcta a cada una de las afirmaciones siguientes, o completar la frase:

| | | |
|---|---|---|
| 1) El modelo de Student se usa para testear medias y varianzas. | **V** | **F** |
| 2) El modelo de Chi cuadrado se usa para testear proporciones. | **V** | **F** |
| 3) Las ventajas de las pequeñas muestras respecto a las grandes muestras son: ........................ | **...............** | |
| 4) El modelo de Fisher sirve para comparar dos varianzas. | **V** | **F** |
| 5) Las desventajas de las pequeñas muestras frente a las grandes son: .................................... | ................ | |
| 6) Para emplear modelos de muestras pequeñas se necesitan por lo menos 2 valores. | **V** | **F** |
| 7) Con el modelo Student se pueden hacer validaciones de: .................................................... | ................ | |
| 8) A la teoría de pequeñas muestras se la llama también teoría exacta del muestreo. | **V** | **F** |
| 9) Los modelos de muestras pequeñas se pueden aplicar a las grandes, pero no el viceversa. | **V** | **F** |
| 10) La Chi cuadrado y la Fisher sirven para estudiar dispersiones. | **V** | **F** |
| 11) Student sirve para estudiar exactitud. | **V** | **F** |
| 12) Con Student se puede trabajar con 1 y con ..... muestras con las medias y ......................... | .............. | |
| 13) No es lo mismo comparar dos muestras independientes que apareadas. | **V** | **F** |
| 14) La única diferencia entre Gauss y Student en los intervalos de confianza es el valor Z. | **V** | **F** |
| 15) La curva de Student es simétrica respecto al origen, pero Chi cuadrado y Fisher no. | **V** | **F** |
| 16) Cuando al mismo individuo se lo mide dos veces, se dice que hay apareamiento. | **V** | **F** |
| 17) Las muestras apareadas sirven para estudiar casos del tipo: antes-después. | **V** | **F** |
| 18) Estos tres modelos permiten comparar hasta dos muestras. | **V** | **F** |
| 19) Las muestras deben provenir de poblaciones normales. | **V** | **F** |
| 20) Las muestras pueden escogerse de cualquier manera no tendenciosa. | **V** | **F** |
| 21) Si las muestras no son independientes, igual se las puede comprar con la Chi cuadrado. | **V** | **F** |
| 22) Las distribuciones de Chi cuadrado y Fisher varían desde cero a infinito. | **V** | **F** |
| 23) Los grados de libertad siempre valen N-1. | **V** | **F** |
| 24) Para dos muestras, los grados de libertad valen la suma de los tamaños menos dos. | **V** | **F** |
| 25) Cuando en la Ho hay un signo de mayor, la zona de rechazo queda a la izquierda. | **V** | **F** |
| 26) El signo de desigualdad de la H1 define el lugar del rechazo (izquierda o derecha). | **V** | **F** |
| 27) En las comparaciones de dos muestras conviene usar una igualdad en Ho. | **V** | **F** |
| 28) El error casual se investiga con Chi cuadrado o Fisher. | **V** | **F** |
| 29) El error sistemático se investiga con el modelo Student y un patrón. | **V** | **F** |
| 30) En Fisher el denominador siempre debe ser mayor que el numerador. | **V** | **F** |
| 31) La propiedad inversa permite obtener probabilidades chicas en una Fisher. | **V** | **F** |
| 32) Explicar cómo se hace para controlar la exactitud de un método. | ........... | |
| 33) Lo mismo pero respecto a la precisión. | ........... | |
| 34) Explicar el significado del error máximo admisible. | ............ | |
| 35) La tabla de Fisher se caracteriza por tener dos grados de libertad. | **V** | **F** |
| 36) Las tablas de Student y Chi cuadrado tienen un solo grado de libertad. | **V** | **F** |

2) La vida útil de un medicamento A se midió con 20 muestras y dio una media de 140 días, con un desvío de 10 días. El medicamento B se determinó con 10 muestras y se obtuvo una media de 120 días con un desvío de 80 días. Decidir si hay diferencias entre ellos desde el punto de vista de la exactitud y de la precisión.

3) Una máquina produce grageas con un diámetro de 0,5 cm según su promedio histórico. Para comprobar si se mantiene en buenas condiciones se toma una muestra de 10 grageas, se las mide y se obtiene una media de 0,53 cm con un desvío de 0,03 cm. Ensayar la hipótesis que está bien.

4) Dos tipos de soluciones químicas A y B fueron ensayadas para medirle su pH. Se tomaron 6 muestras de A que dieron una media de 7,48 con un desvío de 0,02. De la B se tomaron 5 muestras y se obtuvo una media de 7,32 con un desvío de 0,03. Con esta información decidir si ambas muestras tienen el mismo pH.

5) La especificación de un medicamento requiere un 12% de cierto principio activo. Se analizaron 20 muestras del medicamento y se detectó un promedio de 12,5% del componente, con un desvío de 0,3%. Decidir si el producto cumple las especificaciones requeridas.

6) Para el problema anterior, se eligen 18 muestras de un segundo lote de producción y resulta una media de 12,1% con un desvío de 0,09. Se pide averiguar:

a) si este lote cumple las especificaciones requeridas;
b) si hay diferencias entre las medias de ambos;
c) si hay diferencias entre las varianzas de los dos lotes.

7) En una fábrica de medicamentos un farmacéutico sugiere cambiar una de las principales máquinas del proceso productivo, para reducir los tiempos de producción y bajar los costos. Un estudio económico financiero encuentra que tal cambio resulta si el tiempo se reduce en por lo menos un 10%. Se hicieron 5 pruebas con el nuevo equipo y resultó una disminución promedio del tiempo de fabricación del 10,5% con un desvío del 0,3%. Averiguar si resulta conveniente la adquisición de la nueva máquina.

8) Se tienen dos poblaciones cuyas varianzas son desiguales: $\sigma_1^2 \neq \sigma_2^2$ y desconocidas. Por ello el investigador tiene que estimarlas con los valores muestrales $DS_1^2 \approx \sigma_1^2$ y $DS_2^2 \approx \sigma_2^2$. Elige al azar 25 individuos de la primera población y obtiene un promedio de 14 con un desvío de 2, mientras que con 26 muestras de la segunda obtiene una media de 18 y un desvío de 3. Con esta información se debe efectuar una comparación de medias con el modelo de Student. Calcular en primer lugar el número de grados de libertad y luego efectuar el test

9) Calcular como resultaría el problema anterior si se supiera que las varianzas poblacionales son iguales o se pudiera hacer semejante supuesto.

10) En un hospital se desea saber si hay diferencia entre dos tratamientos diferentes. Para tratar de saber la efectividad de los mismos para disminuir el dolor causado por la enfermedad. En el primer caso, cuando el paciente es tratado en la sala de emergencias se le aplica una mezcla de un IV hidroclorato de ranitidina para tratar la dispepsia. Se mide el dolor antes del tratamiento y 45 minutos después de ser aplicado. Al dolor se lo clasifica en una escala de 1 a 10. El segundo tratamiento es darle una mezcla de lidocaina viscosa con un antiácido. Se eligieron 28 pacientes y el primer tratamiento se le aplicó 15 de ellos. La asignación de cada tratamiento fue hecha en forma aleatoria. Los resultados fueron: una media de 3,87 para el primero y de 4,08 para el segundo. Como no hay datos reales del desvío estándar en los ratings de dolor, el investigador decide usar un dato conocido: en ausencia de tratamiento el desvío es $\sigma = 1,73$. Luego decide usar ese dato para estimar el desvío de las diferencias muestrales $\sigma_\Delta$ con la relación:

$\sigma_\Delta = 1,73 \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}$ Decidir si hay diferencia entre los dos tratamientos