



# 10

## Pruebas de contraste de hipótesis. Estimación puntual y por intervalos

---

Ágata Carreño Serra

### 10.1. Introducción

La mayoría de las investigaciones realizadas en el ámbito médico-clínico, comportan estudios comparativos entre dos o más muestras comparando, la mayoría de ellos, el efecto producido por terapias o tratamientos. En este último caso, algunos estudios enmascaran el placebo de fármaco activo, aunque esta práctica ha sido objeto de amplio debate por plantear dudas sobre su ética. La finalidad de estas investigaciones es contestar preguntas tales como: ¿es igual el tratamiento A al tratamiento B? ¿Cuál es la efectividad del tratamiento? En estos casos es cuando necesitamos evaluar si las diferencias que se obtienen a partir de una muestra, se deben a factores distintos al azar y están directamente relacionadas con la administración de un tratamiento u otro.

Para conocer en qué se basan este tipo de estudios, deberemos introducir conceptos como las pruebas de hipótesis y los errores asociados a ellas. Además, veremos que esta probabilidad puede ser calculada a partir de pruebas estadísticas paramétricas, en las que se supone la normalidad de los datos, y las no paramétricas, usadas en condiciones no idóneas de normalidad.

133



## 10.2. Pruebas de contraste de hipótesis

En la situación que hemos comentado a modo de introducción, la respuesta al tratamiento se determinará en base a una medida numérica, sea el descenso de la TAS o mediante un porcentaje de respuesta al tratamiento. Por tanto, las respuestas obtenidas en cada grupo de tratamiento podremos decir que constituyen una estimación de la efectividad de los mismos.

Podemos intuir que aunque la efectividad de los dos tratamientos fuera teóricamente la misma, por el simple hecho de tratar con muestras, la diferencia observada no sería exactamente el valor cero, sino que estaría muy próxima a él, siendo bastante improbable obtener valores lejanos a ese valor (aunque no imposible).

El principio básico que hace referencia al contraste de hipótesis pretende contraponer dos suposiciones contrarias (hipótesis) formuladas al principio de todo estudio: una a favor de la igualdad, bien sea de respuesta de tratamientos o hacia un valor conocido, y otra que establece la desigualdad entre la efectividad real de los tratamientos estudiados o entre el valor que le suponemos al inicio.

En realidad, el propósito de la prueba de hipótesis es determinar si el valor supuesto (hipotético) de un parámetro poblacional, como la media de la población, debe aceptarse como verosímil en base a evidencias muestrales. Si el valor observado de una estadística muestral (por ejemplo, la media), se acerca al valor paramétrico supuesto y difiere muy poco de él, el valor hipotético no se rechaza. Si por el contrario el valor obtenido muestralmente difiere en gran medida de lo que podemos atribuir al azar, la hipótesis formulada al inicio se rechazará por inverosímil.

### 10.2.1. Estimación puntual. Definición de estadístico

Un estadístico es un valor que se obtiene de la muestra y que representa o estima a su parámetro poblacional. Por tanto, es un instrumento mediante el cual podremos estimar parámetros, utilizando lo que llamamos inferencia o estadística inferencial. La estimación puntual hace referencia al cálculo de valores que apuntan hacia el verdadero valor poblacional, como por ejemplo: estimación de una media o de una prevalencia. De esta forma, un buen estimador debe ser:

- o *Insesgado*: Que el valor del parámetro coincida con el valor promedio del estimador. Esta propiedad la tienen la mayoría de los estimadores usados en la práctica.





- o *Consistente*: Que el valor de la muestra se acerque al valor del parámetro al aumentar el tamaño de la muestra.
- o *Suficiente*: Que el estimador use toda la información que la muestra contiene respecto al parámetro de interés.
- o *Eficiente*: Que el estimador tenga menor variabilidad que otro posible.

Por ejemplo, la media muestral es un buen estimador de la media poblacional, porque su valor apunta al verdadero valor promedio en la población. Otros estimadores puntuales son, la proporción muestral para estimar proporciones poblacionales y la desviación estándar en la muestra para estimar la poblacional. En estos ejemplos, la estimación puntual permanece igual, pero como asumimos cierto error por el hecho de elegir una muestra y no otra, debemos acotar el error que cometemos y ello se realiza mediante el *intervalo de confianza*, puesto que la estimación puntual es insuficiente. El **intervalo de confianza** se puede definir como el intervalo de longitud mínima tal que contiene el verdadero valor del parámetro poblacional con una probabilidad igual a  $1-\alpha$ . A efectos prácticos esto significa que si seleccionamos 100 muestras distintas de una misma población y calculamos el intervalo de confianza del 95%, el estimador obtenido en 95 de estas muestras estará contenido en dicho intervalo.

### 10.2.2. Hipótesis nula e hipótesis alternativa

En un contraste de hipótesis se denomina hipótesis nula ( $H_0$ ) a la que considera que ambos tratamientos son iguales y si, en el supuesto de que sea cierta, la probabilidad de que se observe una diferencia tan grande o mayor que la obtenida en nuestro estudio es muy baja (usualmente el valor crítico es 0,05) se rechaza dicha hipótesis y se acepta la contraria, denominada hipótesis alternativa, que establece que ambos tratamientos son diferentes. La nomenclatura utilizada en un contraste de hipótesis suele ser la siguiente:

$$\left\{ \begin{array}{l} H_0 \text{ (Hipótesis nula): Igualdad de tratamientos} \\ H_a \text{ (Hipótesis alternativa): Diferencia de tratamientos.} \end{array} \right. \quad \text{ó también} \quad \left\{ \begin{array}{l} H_0: \text{ Respuesta igual a valor conocido} \\ H_a: \text{ Respuesta distinta al valor conocido} \end{array} \right.$$

Por tanto, aceptaremos la hipótesis nula, si las diferencias obtenidas experimentalmente se deben sólo al azar; mientras que la rechazaremos y aceptaremos la hipótesis alternativa, si consideramos que las diferencias obtenidas no se deben únicamente al azar, si no que la administración de un tratamiento u otro ha provocado respuestas sensiblemente distintas entre ellos. Según sean los estadísticos que se desean contrastar, la expresión de la hipótesis nula toma diver-





sas formas, por ejemplo:

	Hipótesis nula	Contraste
Comparación de dos medias muestrales $X_1$ y $X_2$	$\bar{x}_1 = \bar{x}_2$	$\bar{x}_1 - \bar{x}_2 = 0$
Comparación de la media muestral $X_1$ con media poblacional $\mu$	$\bar{x}_1 = \mu$	$\bar{x}_1 - \mu = 0$
Comparación de dos proporciones muestrales $p_1$ y $p_2$	$p_1 = p_2$	$p_1 - p_2 = 0$
Comparación de una proporción muestral $p_1$ con proporción poblacional $p$	$p_1 = p$	$p_1 - p = 0$

**Tabla 4.** Planteamiento de la Hipótesis nula según tipos de contraste de hipótesis

Los ejemplos que hemos puesto al inicio del capítulo, referentes a las preguntas ¿Es igual el tratamiento A al tratamiento B? ¿Cuál es la efectividad del tratamiento? Aparecen a continuación en formato de contraste de hipótesis:

$$\begin{cases} H_0 : \mu_A = \mu_B \text{ (Igualdad de tratamientos)} \\ H_a : \mu_A \neq \mu_B \text{ (Diferencia de tratamientos)} \end{cases} \quad \text{ó también} \quad \begin{cases} H_0 : \mu_A = 25 \text{ (Efectividad igual a 25 mmHg)} \\ H_a : \mu_A \neq 25 \text{ (Efectividad distinta de 25 mmHg)} \end{cases}$$

Aunque estos conceptos parecen muy teóricos, en realidad son muy intuitivos. Las pruebas de hipótesis son formulaciones de lo que queremos contrastar y, en base a resultados probabilísticos (asociados a la popularmente conocida «p» o «p-value»), aceptaremos la hipótesis nula o la rechazaremos. Puesto que trabajamos con muestras y no con la población entera, debemos asumir que podremos cometer un cierto error en nuestras decisiones. Estos errores asociados a la aceptación y rechazo de la hipótesis nula son los errores alfa y beta.

### 10.2.3. Error alfa y error beta

En los contrastes de hipótesis se pueden dar dos tipos de errores en función de lo que nos aporta nuestra muestra y lo que objetivamente está ocurriendo en la realidad. Estos errores (o probabilidades) reciben el nombre de alfa y beta y se definen tal como se indica a continuación:

- o El **error alfa** o tipo I, es el que se comete al rechazar la hipótesis nula  $H_0$  siendo cierta. Es decir, aceptamos que existen diferencias entre tratamientos cuando en realidad no las hay.
- o El **error beta** o tipo II, es el error que se comete al aceptar la hipótesis





nula  $H_0$  siendo falsa. Es decir, existe una diferencia real entre tratamientos pero no se ha podido rechazar la hipótesis nula.

Como lo que ocurre en la realidad es desconocido, lo único que podemos acotar es la probabilidad de equivocarnos. La siguiente tabla muestra los errores alfa y beta en función del resultado de la prueba y lo que realmente ocurre en la realidad.

Decisión	Realidad	
	$H_0$ verdadera	$H_0$ falsa
Aceptamos $H_0$	Sin error	Error beta o tipo II
Rechazamos $H_0$	Error alfa o tipo I	Sin error

**Tabla 5.** Errores asociados a las pruebas de hipótesis

El error alfa debe ser fijado a priori por los responsables del estudio, puesto que es el que marca el nivel de rechazo de la hipótesis nula. Este valor se sitúa generalmente en 0,05 y es el valor con el que compararemos nuestra «p», obtenida mediante el estadístico de contraste adecuado a nuestros objetivos. Avanzándonos un poco más en las pruebas de hipótesis, cuando hablamos de una «p», «p-value» o «nivel de significación» con un valor cercano a cero, estamos indicando que la prueba realizada (t de Student, Prueba F, prueba Z) supera el umbral probabilístico previamente fijado de 0,05 como tope mínimo para considerar la igualdad entre los parámetros. Como hemos dicho, el nivel de significación  $\alpha$  lo marcamos nosotros de manera que, cuanto mayor sea, más fácil será aceptar la hipótesis alternativa cuando en realidad es falsa.

Si al valor  $\alpha$  se le llama también grado de significación estadística, a su complementario (1-alfa) se le llama nivel de confianza, o probabilidad de que las diferencias que se detecten no se deban al azar. Por convenio, suele utilizarse un valor de 0,05, lo que significa que 5 de cada 100 veces detectaríamos diferencias entre tratamientos cuando realmente serían iguales, aunque también es usual un valor de 0,01.

Teniendo en cuenta que en la gran mayoría de los casos el nivel alfa se fija a 0,05 en un contraste de hipótesis aceptaremos una hipótesis o la otra según la siguiente regla de decisión:





Si **p es menor de 0,05**, se admite que la probabilidad de que las diferencias encontradas se deban al azar son demasiado pequeñas, por lo que rechazaremos la hipótesis nula y aceptaremos la alternativa.

Si **p es mayor o igual a 0,05**, la probabilidad de que las diferencias encontradas se deban al azar son demasiado grandes para aceptar la hipótesis alternativa y, por tanto, se acepta la hipótesis nula: las diferencias encontradas están dentro de las que consideramos debidas al azar.

El error beta se ha definido como la probabilidad de rechazar la hipótesis nula cuando ésta es falsa, es decir, la probabilidad de detectar diferencias entre tratamientos cuando realmente existen diferencias. Asociado al error beta se ha definido la **potencia estadística** de un test como  $1 - \beta$ . Al igual que el nivel de significación, el error beta o poder estadístico debe definirse antes de iniciar el estudio para estimar el tamaño muestral necesario para poder evaluar el objetivo del estudio.

Debido a que los dos errores definidos son imposibles de controlar a la vez, vamos a fijarnos solamente en el error alfa o nivel de significación y nos informa de la probabilidad que tenemos de estar equivocados si aceptamos la hipótesis alternativa. Este error es el que nos interesa ya que la hipótesis alternativa (existen diferencias) que estamos interesados en probar, no queremos aceptarla si en realidad no es cierta, es decir, si aceptamos la hipótesis alternativa queremos equivocarnos con un margen de error muy pequeño.

#### 10.2.4. Estimación por intervalos. Región crítica y región de aceptación

Como hemos visto, asociado al valor del estadístico calculado para evaluar las hipótesis, existe un intervalo o rango de valores dentro del cuál aceptamos la hipótesis nula, rechazando la alternativa. Ese intervalo o región se denomina **región de aceptación**, y será mayor o menor dependiendo del nivel de confianza que precisemos,  $1 - \alpha$ . La región que quede fuera de la región de aceptación indica que en este caso los cambios no se pueden atribuir al azar, y por tanto hemos de rechazar  $H_0$  y aceptar  $H_1$ . Tal región se llama **región crítica o de rechazo**. Llegados a este punto, hemos de distinguir entre dos tipos de contraste o test, que determinan la región de aceptación y la región de rechazo.

En una distribución Normal, cerca del 68% de los valores de un parámetro, por ejemplo la media poblacional, están incluidos dentro del intervalo abarcado por la media muestral  $\pm 1$  desviación estándar (DE); más del 95% están dentro de la media  $\pm 2$  DE y más del 99% entre la media  $\pm 3$  DE. El verdadero valor poblacional estará situado dentro de estos valores con dicha probabilidad.





Al plantearnos el contraste de hipótesis, si desconocemos en qué dirección puede ser falsa  $H_0$ , entonces  $H_a$  es simplemente la negación ( $\neq$  o tratamientos distintos) y decimos entonces que el contraste es **bilateral**. Si por el contrario, conocemos que un tratamiento, si tiene efectos, puede mejorar la respuesta pero nunca empeorarla, entonces estamos hablando de un contraste **unilateral**, dado que la hipótesis alternativa recoge diferencias entre dos parámetros pero en un único sentido. A continuación los veremos por separado.

#### 10.2.4.1. Contraste bilateral (o de dos colas)

En este caso la región de rechazo o región crítica está formada por dos áreas disjuntas. Dicho caso se presenta cuando la hipótesis nula es del tipo  $H_0 : \mu = k$  (o bien  $H_0 : p = k$ ) y la hipótesis alternativa, por tanto, es del tipo  $H_1 : \mu \neq k$  (o bien  $H_1 : p \neq k$ ). La región crítica para un cierto nivel  $\alpha$  sería, en la  $N(0;1)$ :

El valor  $z(\alpha/2)$  para  $\alpha$  igual a 0,05 corresponde al valor 1,96, muchas veces aproximado a 2, que es el que corresponde al valor de la curva normal que deja una probabilidad a cada lado de la curva de  $\alpha/2$ , es decir, deja a cada lado 0,025.

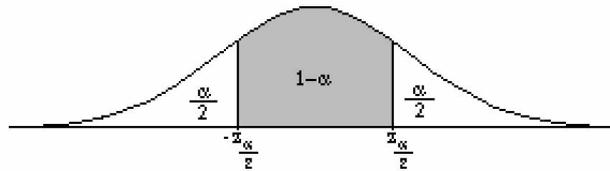


Figura 39. Intervalo  $-z(\alpha/2)$ ,  $z(\alpha/2)$ . Región de aceptación y de rechazo.

Fijémonos en que el nivel de significación  $\alpha$  se concentra en dos partes (o colas) simétricas respecto de la media. La región de aceptación en este caso no es más que el correspondiente intervalo de probabilidad para  $x$  o  $p$ , es decir:

$$\left( \mu - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \mu + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right) \quad \text{ó} \quad \left( p - z_{\alpha/2} \sqrt{\frac{pq}{n}}, p + z_{\alpha/2} \sqrt{\frac{pq}{n}} \right)$$

#### 10.2.4.2. Contraste unilateral (o de una cola)

En este caso la región crítica está formada por una sola área. Como se observa en las figuras, el nivel de significación  $\alpha$  se concentra sólo en una parte o cola. Este caso se presenta cuando la hipótesis nula es del tipo  $H_0 : \mu \geq k$  (o bien  $H_0 : p \geq k$ ) y la hipótesis alternativa, por tanto, es del tipo  $H_1 : \mu < k$  (o bien  $H_1 : p < k$ ). También puede aparecer en sentido contrario, es decir,  $H_0 : \mu \leq k$  (o bien  $H_0$





:  $p \leq k$ ) y la hipótesis alternativa,  $H_1 : \mu > k$  (o bien  $H_1 : p > k$ ). A nivel de confianza  $1 - \alpha$ , las regiones serán, en la  $N(0;1)$ :



**Figura 40.** Nivel de aceptación de los contrastes unilaterales

En estos casos el valor  $z(\alpha)$  corresponde al valor 1,64 ó -1,64 según el sentido de la diferencia, que es el que corresponde al valor de la curva normal que deja una probabilidad por encima o por debajo de 0,05, constituyendo esta región la región de rechazo y el resto la región de aceptación y engloba el 95% del área de la curva.

Los valores más conocidos de  $\alpha$ , corresponden a puntos de la distribución de referencia que engloban un área por debajo de la curva correspondiente a  $1 - \alpha$ . Si el estadístico de contraste sigue una distribución normal estandarizada  $z$ , estos valores son los que aparecen a continuación:

Contrastes bilaterales		Contrastes unilaterales	
Valores de Z	Área bajo la curva	Valores de Z	Área bajo la curva
Entre -1.645 y 1.645	0.90	Entre $-\infty$ y 1.28	0.90
Entre -1.96 y 1.96	0.95	Entre $-\infty$ y 1.645	0.95
Entre -2.575 y 2.575	0.99	Entre $-\infty$ y 2.33	0.99

**Tabla 6.** Valores Z y área bajo la curva asociados a los contrastes bilaterales y unilaterales

### 10.2.5. Algunos contrastes paramétricos más importantes

Hemos dado mucha importancia a los errores alfa y beta, sin embargo, debe tenerse claro que el cálculo más complicado, corresponde al estadístico de contraste adecuado para solucionar nuestras hipótesis. La elección de la prueba estadística consistirá en definir una medida que permita comparar los resultados obtenidos en nuestra muestra con los resultados teóricos según la hipótesis planteada. La distribución de esta medida, cuando la hipótesis nula es cierta, deberá aproximarse a la de alguna distribución conocida. Este valor del estadístico es el que debe «situarse» en su función de distribución para establecer el área bajo la curva y conocer la probabilidad asociada a ese valor. Esa probabilidad es la que llamare-





mos p-valor, por lo que si el valor de nuestro estadístico está muy lejano a los valores normales de la distribución, dejará un área muy grande bajo la curva y una probabilidad muy pequeña de obtener valores mayores a él.

La mayoría de los paquetes estadísticos tienen incorporados los cálculos necesarios para los estadísticos adecuados y nos ofrecen su p-valor asociado; sin embargo, a continuación se muestra una pequeña introducción a los estadísticos que se necesitarían para poder los contrastes de hipótesis más importantes. Dicha información nos permitirá entender un poco más el objetivo de los contrastes de hipótesis.

o **Contrastes sobre la media.**

A partir de una muestra extraída de una población  $X$  normal con media  $\mu$  y varianza  $\sigma^2$  desconocidas, se desea contrastar la hipótesis nula  $H_0: \mu = \mu_0$ . El estadístico de contraste es  $d_1$ ,  $S$  es la desviación típica muestral corregida y  $\bar{x}$  es un estimador de la media poblacional  $\mu$ . Si  $H_0$  es cierta, el estadístico  $d_1$  sigue una distribución  $t$  de Student con  $n-1$  grados de libertad.

$$d_1 = \frac{\bar{X} - \mu_0}{\hat{S}} \sqrt{n},$$

o **Contrastes sobre la varianza.**

Si partimos de una muestra extraída de una población  $X$  normal con varianza  $\sigma^2$  y se desea contrastar  $H_0: \sigma^2 = \sigma_0^2$ , siendo  $\sigma_0^2$  un valor preestablecido de la varianza. El estadístico de contraste es  $d_2$  donde  $\hat{S}^2$  es el estimador habitual de la varianza  $\sigma^2$ . Si  $H_0$  es cierta  $d_2$  sigue una distribución  $F$  de Snedecor.

$$d_2 = \frac{(n-1) \hat{S}^2}{\sigma_0^2}$$

o **Contrastes sobre la igualdad de varianzas.**

Se desea contrastar si dos varianzas son iguales, suponiendo que provienen de dos poblaciones normales.  $H_0: \sigma_x^2 = \sigma_y^2 \leftrightarrow (\sigma_x^2 / \sigma_y^2) = 1$ . El estadístico de contraste es  $d_3$  y si  $H_0$  es cierto  $d_3 \sim F_{n-1, m-1}$

$$d_3 = \frac{\hat{S}_x^2}{\hat{S}_y^2}$$

o **Contrastes sobre la diferencia de medias, muestras independientes e igualdad de varianzas.**

Se supone que  $\sigma_x^2 = \sigma_y^2 = \sigma^2$ . Se desea contrastar  $H_0: \mu_x = \mu_y \leftrightarrow (\mu_x - \mu_y) = 0$ . El estadístico de contraste es  $d_4$  y si  $H_0$  es cierto se verifica que  $d_4 \sim t_{n+m-2}$

$$d_4 = \frac{\bar{X} - \bar{Y}}{\hat{S}_r \sqrt{\frac{1}{n} + \frac{1}{m}}}$$





siendo

$$\hat{S}_r^2 = \frac{(n-1)\hat{S}_x^2 + (m-1)\hat{S}_y^2}{n + m - 2}$$

o **Contrastes sobre la diferencia de medias, muestras independientes y varianzas desiguales.**

Se supone que  $\sigma_x^2 \neq \sigma_y^2$ . Se desea contrastar  $H_0: \mu_x = \mu_y \leftrightarrow (\mu_x - \mu_y) = 0$ . El estadístico de contraste que se utiliza es  $d_5$  y si  $H_0$  es cierto se verifica que  $d_5 \sim t_g$ , siendo  $g = n + m - 2$  + un término de corrección

$$d_5 = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\hat{S}_x^2}{n} + \frac{\hat{S}_y^2}{m}}}$$

o **Contrastes sobre la diferencia de medias, muestreo apareado.**

Tenemos dos muestras aleatorias simples de igual tamaño muestral y son obtenidas al realizar dos observaciones  $X_i$  e  $Y_i$  sobre el mismo individuo. Por la naturaleza del muestreo apareado las dos muestras son dependientes. Para eliminar este problema se estudia la variable diferencia  $Z = Y - X$ , por tanto, a partir de las dos muestras iniciales se calcula la muestra de diferencias,  $Z_i = X_i - Y_i$ . Para contrastar la hipótesis  $H_0: \mu_x = \mu_y \leftrightarrow (\mu_x - \mu_y) = 0 \leftrightarrow \mu_z = 0$ . Se utiliza el estadístico de contraste  $d_6$  y si  $H_0$  es cierto  $d_6 \sim t_{n-1}$

$$d_6 = \frac{\bar{Z}}{\hat{S}_z} \sqrt{n}.$$

### 10.2.6. Etapas de la prueba de hipótesis

Los procedimientos seguidos en las pruebas de hipótesis se encuentran fijados y se dan por etapas que facilitan su comprensión. En primer lugar, debe enunciarse la hipótesis nula y alternativa y determinarse el valor  $\alpha$  para poder identificar el valor  $z(\alpha/2)$  que separa las regiones de aceptación o rechazo. En segundo lugar, debemos determinar la distribución del parámetro muestral. Y, por último, calcular el estadístico y aplicar el test; en función de si el estadístico cae en la región crítica o en la región de aceptación, entonces, se tomará una decisión con respecto a una de las dos hipótesis.

### 10.2.7. Ejemplos de utilización

En los próximos capítulos veremos los tests necesarios y los estadísticos a calcular cuando se nos plantea un contraste de hipótesis utilizando el paquete





estadístico SPSS. En este apartado hemos preferido elegir dos ejemplos que ilustren la forma manual de cálculo, sin recurrir al paquete estadístico, con el objetivo de conceptualizar mejor el contraste de hipótesis.

### 10.2.7.1. Estimación por intervalo de una proporción

Imaginemos que deseamos estimar la prevalencia del tabaquismo en los pacientes hospitalarios. Para ello, se revisan las historias clínicas de 150 pacientes, observando que 45 eran fumadores. ¿Cuál es la prevalencia del tabaquismo?.

La proporción de pacientes fumadores es  $p = 45 / 150 = 0,3$  (30%), por lo que la proporción de pacientes no fumadores es  $q = 1 - p = 1 - 0,3 = 0,7$  (70%).

Si decimos que la prevalencia del tabaquismo es del 30%, estamos incurriendo en el error de no proporcionar el intervalo de confianza correspondiente a esa proporción y que nos indicará el rango de valores reales que puede adoptar nuestra prevalencia. Una prevalencia del 30% puede obtenerse de una muestra de 30 casos o de un estudio poblacional con miles de pacientes. Cuanto mayor sea el tamaño muestral, mayor será la precisión de nuestra estimación y esto se refleja con el intervalo de confianza, que será menor. Para obtener el intervalo de confianza para proporciones, deberemos aplicar la siguiente fórmula:

$$IC(p) = p \pm Z_{\alpha/2} \sqrt{\frac{p \cdot q}{n}}$$

donde  $n$  es el tamaño muestral,  $p$  corresponde a la proporción obtenida en nuestra muestra, como hemos dicho anteriormente  $q$  corresponde a  $1-p$  y, finalmente,  $z_{\alpha/2}$  es el valor crítico de la distribución normal que deja una probabilidad  $1-\alpha$  bajo la curva y corresponde al valor 1,96 en el caso de definir un nivel de confianza del 95% (referenciado en la Tabla 6 de este capítulo o bien mediante los valores tabulados de la distribución normal). De esta forma, el cálculo del intervalo de confianza del 95% se obtiene como:

$$IC\ 95\%(p) = 0,3 \pm 1,96 \sqrt{\frac{0,3 \times 0,7}{150}} = 0,3 \pm 1,96 \sqrt{0,0014} = 0,3 \pm 0,073 \Rightarrow 0,227 \leq p \leq 0,373$$

Por lo tanto, a partir de los datos obtenidos en la muestra podemos afirmar que la prevalencia de tabaquismo hospitalario está entre el 22,7% y el 37,3% con una confianza del 95%.

### 10.2.7.2. Estimación por intervalos de dos medias

Un estudio pretende comparar la efectividad de dos tratamientos A y B en la





hipertensión arterial, suponiendo que existen diferencias entre ellos pero sin conocer el sentido de éstas. De esta forma, se plantea el siguiente contraste de hipótesis:

$$H_0 : \mu_A = \mu_B$$

$$H_a : \mu_A \neq \mu_B$$

Para seguir una consistencia en todos los ejemplos, fijaremos el nivel de error alfa en 0,05. En el tratamiento A tenemos una muestra de pacientes ( $n_A=15$ ) en los que se ha observado una reducción media de las cifras de TAS de  $x_A=18$  mmHg con una desviación estándar ( $s_A$ ) de 2,3; mientras que con el tratamiento B, se ha obtenido una muestra de 20 pacientes ( $n_B=20$ ) que han disminuido en promedio  $x_B=15,25$  mmHg de la TAS con una desviación ( $s_B$ ) de 1,92

En primer lugar, calcularemos el estadístico de referencia, que corresponderá a una  $t$  de Student con  $(n_A-1) + (n_B-1)$  grados de libertad, es decir  $t \sim t_{33}$ . La fórmula a aplicar aparece a continuación:

$$t = \frac{x_A - x_B}{\sqrt{\frac{s^2}{n_A} + \frac{s^2}{n_B}}} \quad \text{donde} \quad s^2 = \frac{(n_A-1)s_A^2 + (n_B-1)s_B^2}{n_A + n_B - 2}$$

Calculando las distintas fórmulas, se obtiene que:

$$S^2 = \frac{(14)2,3^2 + (19)1,92^2}{15 + 20 - 2} = \frac{74,00 + 69,75}{33} = 4,356 \Rightarrow t = \frac{18,00 - 15,25}{\sqrt{\frac{4,356}{15} + \frac{4,356}{20}}} = \frac{2,75}{\sqrt{0,508}} = 3,86$$

En la tabulación de la distribución  $t$  de Student (ver Anexo), encontraríamos que el valor de esta distribución con 33 grados de libertad asociada al valor  $\alpha=0,05$  es el  $\pm 2,04$ . Como el valor que nosotros hemos encontrado (3,86) excede el valor teórico (2,04), rechazamos la hipótesis nula y concluimos que existen diferencias entre ambos tratamientos. Otra forma sería buscar en las tablas el valor calculado 3,86 en las tablas para conocer la probabilidad que tiene asociada y, si ésta es menor a 0,05, rechazar la hipótesis nula, tal y como se realiza en las herramientas de análisis, donde generalmente se realiza el cálculo de la probabilidad asociada al valor del estadístico muestral, el  $p$ -valor.

Mediante el test de hipótesis se han observado diferencias y, mediante la muestra, el tratamiento A ha obtenido mejores resultados. Si además quisiéramos conocer la estimación de la diferencia de valores, es decir, la estimación del efecto del tratamiento deberíamos calcular el intervalo de confianza de la diferencia de medias. El intervalo de confianza sería:

$$IC\ 95\%(x_A - x_B) = (x_A - x_B) \pm t_{\alpha/2, s_{x_A - x_B}} \Rightarrow IC\ 95\%(x_A - x_B) = 2,75 \pm 2,04 \cdot \sqrt{\frac{4,356}{15} + \frac{4,356}{20}} = 2,75 \pm 1,04$$





Por lo tanto, la diferencia observada entre el tratamiento A y el tratamiento B es de 2,75 y podemos afirmar que está entre 1,71 y 3,79 mmHg a favor del tratamiento A con una confianza del 95%. Esta conclusión no indica que la efectividad de los tratamientos sea clínicamente distinta, si no que las diferencias observadas están cuantificadas y sólo un clínico puede determinar cuál es la diferencia clínicamente relevante.

### 10.3. Tipos de pruebas estadísticas

Es importante establecer el objetivo o propósito de la prueba para la variable o variables seleccionadas y las limitaciones que pueden tener en cuanto a algunos supuestos que deben cumplirse, esto suele ocurrir en las denominadas pruebas paramétricas. Si después de considerar esto, la prueba no se considera robusta, entonces es más conveniente buscar una prueba de la estadística no paramétrica y que resulte más confiable.

#### 10.3.1. Pruebas paramétricas

Para poder aplicar pruebas paramétricas se requieren una serie de supuestos, además de asegurar de que la muestra que se toma debe haber sido seleccionada de forma aleatoria o probabilística. En las pruebas paramétricas de muestra pequeña, se requiere el supuesto de que las muestras sean extraídas de una población con distribución normal y cuando se trata de dos o más muestras también se requiere la igualdad de varianzas, que puede evaluarse mediante una prueba específica. Existen pruebas estadísticas por medio de las cuales se podría comprobar esto, sin embargo suele no dársele importancia a esto y se pasa por alto.

Las muestras de tamaño grande, tienen menos supuestos, puesto que se pueden aplicar sin saber o comprobar si la población o poblaciones eran normales, estas pruebas se dice que son robustas, porque no es necesario que se cumpla dicho supuesto (ver Teorema Central del Límite). Se dispone de muchas pruebas estadísticas no paramétricas que tienen una aplicación semejante a las paramétricas de muestra pequeña en las que se tienen menos supuestos. Se suele utilizar media y desviación estándar.

#### 10.3.2. Pruebas no paramétricas

En estas técnicas, solamente se necesitan conocimientos elementales de matemáticas, pues los métodos son relativamente más sencillos que en las pruebas paramétricas. Existe toda una teoría que no se basa en la distribución de la función de probabilidades de las variables a estudiar. En muchos casos, resulta





muy útil y su enfoque es bastante claro. Entre ellas destacan la prueba Chi-Cuadrado ( $\chi^2$ ) y otras pruebas para evaluar diferencias de medias para variables que no siguen la ley normal.

Una limitación que tienen estas pruebas es que no son aplicables a casos en los que se desean manejar muchas variables al mismo tiempo, para estos casos, sí se requeriría una prueba paramétrica. Lo que sí se requiere y en general es el supuesto que se debe cumplir en la mayoría de las pruebas no paramétricas para confiar en ellas, es que la muestra haya sido seleccionada en forma probabilística. Cuando se aplican pruebas no paramétricas las medidas resumen que se utilizan suelen ser la mediana y los cuartiles.

Las pruebas que se mencionarán en los siguientes capítulos son las que se podrían necesitar con mayor frecuencia, se mencionarán sus principales características y aplicaciones, además de la prueba paramétrica a la que podrían sustituir. Toda prueba no paramétrica tiene una equivalente paramétrica, no siendo al contrario, del todo cierto.

#### **10.4. Cálculo del tamaño de la muestra**

El tamaño de la muestra es el número de individuos que la componen. Es un factor esencial en las estimaciones y en las pruebas de significación. En las estimaciones, el tamaño de la muestra determina el intervalo de confianza y el error de la estimación; en las pruebas de significación, con los mismos resultados se puede aceptar o rechazar la hipótesis nula según sea el tamaño de la muestra. Por eso, es muy importante conocer el número mínimo de individuos necesario para una estimación antes de realizar la recogida de datos. Esta se realiza con una muestra de tamaño igual o superior al determinado previamente, puesto que en muchos estudios, existe un porcentaje de pérdidas que debe ser tenido en cuenta.

Existe una interdependencia entre el grado de significación ( $p$  ó  $\alpha$ ), el poder estadístico ( $1-\beta$ ), el número de individuos estudiados (tamaño muestral) y la magnitud de la diferencia observada, de tal forma que conociendo tres de estos parámetros, se puede calcular el cuarto. Así, antes de iniciar un estudio, podremos determinar el número de individuos necesarios para detectar una diferencia determinada, fijando a priori un nivel de significación y el poder estadístico deseado.

#### **10.5. Consideraciones importantes**

El grado de significación estadística no es una medida de la fuerza de la aso-





ciación, no mide si un tratamiento es más eficaz o mejor que otro; simplemente nos da la probabilidad de que los resultados obtenidos sean fruto de la casualidad o el azar. La  $p$  tampoco mide la importancia clínica o la relevancia de las diferencias observadas puesto que un estudio puede demostrar diferencias muy significativas entre las variables y carecer de importancia clínica dado que el nivel de significación disminuye aumentando el tamaño de muestra aún manteniendo las diferencias a detectar. Por ejemplo, si un fármaco A reduce la presión arterial 10 mmHg y otro B la reduce 9 mmHg y existen diferencias estadísticamente significativas entre ambos ( $p < 0,05$ ), ello no significa que deba usarse el fármaco A antes que el B, sino que hay que considerar el dudoso beneficio clínico que pueda reportar el reducir la presión arterial 1 mmHg más, los efectos secundarios, la seguridad o el coste económico. El nivel de significación nos da la importancia o significación estadística de las diferencias pero nunca su relevancia clínica.

Todas las pruebas de significación estadística (o pruebas de hipótesis) intentan rechazar o no la hipótesis nula, calculando la probabilidad de que los resultados sean debidos al azar; nos dan, por tanto el grado de significación estadística  $p$ . Las pruebas de significación estadística son métodos de contraste de hipótesis utilizados para valorar el efecto del azar en una investigación. Mediante los tests de hipótesis podemos calcular cómo de probable es que las diferencias observadas en una investigación se deban al azar.

Actualmente las pruebas de hipótesis reciben algunas críticas por varios motivos. En primer lugar, se desconoce la magnitud de la diferencia que se observa y por tanto, no se puede definir la relevancia clínica. En segundo lugar, damos como significativo un resultado con una  $p = 0,045$  y sin embargo aceptamos la hipótesis nula con una  $p = 0,05$ . Y finalmente, con un tamaño de muestra elevado, cualquier resultado puede cobrar significación estadística.

En contraposición, un intervalo de confianza es un recorrido de valores, basados en una muestra tomada de una población, en el que cabe esperar que se encuentre el verdadero valor de un parámetro poblacional con cierto grado de confianza. En otras palabras, se puede tener gran confianza en que el intervalo resultante abarca el valor verdadero, pues dicho intervalo se ha obtenido por un método que casi siempre acierta.

Un intervalo de confianza posee la ventaja de que se puede calcular para cualquier valor. Si se desea determinar si es verdadera la diferencia observada entre dos grupos, se calcula el intervalo de confianza de 95% de la diferencia entre sus respectivas medias. Si el intervalo abarca el valor cero, no se puede descartar que no haya una diferencia; si no lo abarca, la probabilidad de que se esté observando una diferencia que en realidad no existe se considera remota.





La precisión de los resultados guarda relación con el tamaño muestral y con la variabilidad de los datos, de tal manera que cuanto más grande la muestra, más se acercarán los resultados al verdadero valor poblacional y más estrecho será el intervalo de confianza. Asimismo, mientras más grande sea la desviación estándar de los datos, menos precisos serán los resultados y más amplio el intervalo de confianza.

