

Capítulo 17

Bases para el razonamiento en Estadística inferencial

*Alejandra Moreno Altamirano
Antonio R. Villa Romero*

Introducción

A fin de conocer, explicar y mejorar las condiciones de salud de una población, además de la Salud Pública y la Epidemiología, se aplica la Estadística para la cuantificación de la distribución, frecuencia y determinantes de la salud y enfermedad en dichas poblaciones.

En específico, la inferencia estadística es de gran utilidad, ya que a partir del estudio de una muestra seleccionada es posible llegar a conclusiones respecto de la población de donde proviene la muestra. Así que, dentro de ciertos límites totalmente cuantificables, es posible generalizar (inferir) que los resultados encontrados en una muestra pueden ser aplicados al universo.

Además de conocer las condiciones de salud a nivel poblacional, la inferencia estadística también es útil para la toma de decisiones en el campo clínico-administrativo o de políticas de salud.

Un aspecto fundamental es que la muestra estudiada sea representativa de la población que se desea conocer; por tanto, para extrapolar sus resultados debe reunir las siguientes características:

1. Los sujetos deben ser seleccionados en forma aleatoria.
2. Todas las personas deben tener una probabilidad conocida de ser seleccionadas para entrar en la muestra.
3. Cada persona debe ser elegida de modo independiente de las demás.

En la muestra seleccionada se miden las variables de interés y se elabora la presentación resumida de los datos en forma de estadísticos: medias, diferencia entre dos medias, proporciones, diferencia entre dos proporciones, varianzas, etc. Lo anterior es parte de la definición de los estimadores (calculados en la muestra) cuya utilidad radica en inferir los parámetros (en el universo): esto es la media, la proporción, etc., de tipo poblacional.

Así, por ejemplo, si se desea determinar la prevalencia de obesidad en escolares de la ciudad de México, es factible estimar el parámetro a partir del estudio de una muestra

representativa. El estadístico a estimar será una proporción (como se sabe, la prevalencia es una proporción) calculada de la siguiente forma:

$$\frac{(\text{Total de escolares con obesidad en la muestra})}{(\text{Total de escolares estudiados en la muestra})}$$

Se espera que el resultado obtenido pueda ser aplicado a todos los escolares de la ciudad de México, en caso de que la muestra estudiada haya garantizado los supuestos señalados previamente, es decir, sujetos seleccionados en forma aleatoria, con probabilidad conocida de selección y elegidos de manera independiente.

Alternativas para la inferencia estadística

La inferencia estadística se logra a partir de:

- a) Estimación puntual o de intervalos.
- b) Pruebas de hipótesis.

En el ejemplo anterior, la prevalencia de sobrepeso y obesidad estimada pudiera ser 15.3%. Como resulta evidente, es un solo valor; es decir, se trata de una información puntual a la que se le denomina “**estimador puntual**”. Este estimador no proporciona información de la variabilidad inherente; se ignora cuán cerca está el estimador (muestra) del parámetro (universo). Es probable y esperable que se localice cerca de la verdadera proporción poblacional si la muestra es lo bastante grande.

Intervalo de confianza

La estimación por intervalo consta de dos valores numéricos que definen un rango dentro del cual se pretende que se encuentre el parámetro de interés con un cierto grado de confianza. A este intervalo se le conoce como “**intervalo de confianza**” (IC) y por lo general se expresa con un nivel de 95% (IC_{95%}).

A fin de estimar el intervalo de confianza se debe conocer la distribución de la variable, la cual puede adoptar diferentes formas. La forma más típica en variables de tipo continuo que se distribuyen en forma normal es la distribución Z. El grado de confianza deseado se indica por $1 - \alpha$ y por lo general se da en porcentaje $(1 - \alpha)\%$. A menudo se fija en el 95% correspondiente a un valor de $\alpha = 0.05$, el cual representa el grado de confianza o probabilidad que se establece *a priori* de que el intervalo de confianza obtenido contenga el valor verdadero del parámetro en la población. Cuanto más nivel de confianza se desee, menor será el valor de α y más elevado el número necesario de sujetos.

Se considera el nivel de confianza y no de probabilidad, ya que una vez extraída la muestra, el intervalo de confianza puede o no incluir al valor verdadero del parámetro. Es un hecho que si se repite el proceso con muchas muestras sería factible afirmar que el $(1 - \alpha)\%$ de los intervalos así construidos contendría al valor verdadero del parámetro. Los valores que se suelen utilizar para construir el nivel de confianza son 90, 95 y 99%.

La **fórmula general para calcular intervalos de confianza** es la siguiente:

$$\text{Estimador puntual} \pm (\text{coeficiente de confiabilidad}) \times (\text{error estándar})$$

En cuanto a la **interpretación de los intervalos de confianza**, se dice que existe un $100(1 - \alpha)\%$ de confianza de que el intervalo calculado contiene el parámetro verdadero de la población. En otras palabras, significa que si se eligieran 100 muestras aleatorias de la población y se calculara en cada una de ellas el $IC_{95\%}$, aproximadamente 95 de ellos contendrían el verdadero parámetro poblacional y 5 no lo contendrían.

Recuadro 17-1 Ejemplo 1.

Datos obtenidos:

Estimador puntual = media de edad de escolares con sobrepeso u obesidad = 9.0 años

Coefficiente de confiabilidad = valor z en prueba de 2 colas para 95% de confianza = 1.96

Error estándar = 0.9

Cálculo de $IC_{95\%}$:

De acuerdo con la fórmula general:

$$9.0 \pm 1.96 (0.9)$$

$$9.0 \pm 1.764$$

$$IC_{95\%} = 7.236 - 10.764$$

Interpretación:

Con un 95% de confianza se puede asegurar que la media de edad en escolares con sobrepeso u obesidad se encuentra entre 7.2 y 10.8 años.

Prueba de hipótesis

Otra forma de llegar a las mismas conclusiones a las que se llega utilizando intervalos de confianza es por medio de la prueba de hipótesis. Antes de revisar el tema es necesario distinguir dos tipos de hipótesis:

1. **Hipótesis de investigación.** Es la conjetura o suposición que motiva el estudio y que está sustentada en la experiencia o el conocimiento que se tiene del objeto estudiado.
2. **Hipótesis estadística.** Es una proposición sobre los parámetros de una o más poblaciones.

Las hipótesis de investigación pueden ser de cuatro tipos:

- a) Descriptivas, cuando se pretende describir alguna característica en la población a partir del valor de la o las variables.

- b) De correlación, cuando se establecen asociaciones, predicciones o explicaciones no causales.
- c) Diferenciales, son las que establecen diferencias y, a su vez, son de dos tipos:
 - Direccionales o de una cola, las cuales plantean magnitud de la diferencia entre grupos en una dirección.
 - No direccionales o de dos colas, que establecen diferencias entre variables sin especificar magnitud o dirección de la diferencia.
- d) De causalidad, cuando se establece un sentido de relación causa-efecto.

Hipótesis nula e hipótesis alterna

La hipótesis estadística define a través de dos planteamientos (hipótesis nula e hipótesis alterna) el contraste que debe ser comprobado a través de una prueba estadística (prueba de hipótesis). En términos generales, cabe señalar que la prueba de hipótesis estadística va en el sentido de rechazar la hipótesis nula.

Hipótesis nula (H_0). Sustenta que las dos muestras a comparar han sido extraídas de la misma población, por tanto, asume la igualdad de los parámetros de ambas poblaciones. Así, los valores encontrados en ellas son iguales (es decir, son equivalentes), y cualquier diferencia observada entre ellas se considera como un hecho casual resultado únicamente del error de muestreo. De modo que la diferencia que existe entre las dos muestras no representa una diferencia real, sino una diferencia que se atribuye a un error de muestreo y que se expresa en términos generales como: “La media de la población 1 es igual a la media de la población 2” y se simboliza de la siguiente forma: $\mu_1 = \mu_2$.

Hipótesis alterna (H_1). Se trata de la afirmación contraria a la H_0 y es la hipótesis que va en el sentido lógico que busca el investigador. Esto es, pretende establecer una diferencia en la comparación de dos muestras. Por ejemplo, la media de años de los escolares con sobrepeso u obesidad es diferente en la población urbana que en la rural: “La media de la población 1 (urbana) es diferente a la media de la población 2 (rural)” y se simboliza de la siguiente forma: $\mu_1 \neq \mu_2$.

En otras palabras, si existe suficiente poder estadístico (expresado en significancia estadística), es posible rechazar la hipótesis nula, la cual señala que no existe diferencia en la edad media de los escolares con sobrepeso u obesidad entre aquellos de área urbana *vs* área rural. De modo que se acepta la hipótesis alterna, la cual establece que sí existe diferencia.

Al igual que en la inferencia estadística, se espera que la muestra cumpla los criterios antes descritos de aleatoriedad e independencia, ya que en ella se probará la hipótesis nula. Esto es que a partir de la información obtenida de las variables medidas en la muestra se debe tener información suficiente para tomar una decisión correcta con respecto a la población. Sin embargo, ni la prueba de hipótesis ni la inferencia estadística conducen a la demostración de una hipótesis en sí, tan sólo indican si la hipótesis alterna es respaldada o no con los datos disponibles.

Como ya se señaló, la decisión de rechazar o no la hipótesis nula se establece a partir de un estadístico de prueba que es el mecanismo por medio del cual se contrasta la hipótesis y depende de su nivel de significancia estadística el hecho de que pueda o no ser rechazada.

La selección de la prueba de hipótesis estadística depende de la **escala de medición** en que estén medidas las variables estudiadas, ya sea cuantitativa (continua o discreta) o cualitativa (nominal u ordinal) y en función de lo que se pretende **comparar**, como puede ser: promedios, varianzas, proporciones, medida de riesgo, etc. Con base en lo anterior, la prueba seleccionada podrá ser paramétrica o no paramétrica, según si requiere o no del cumplimiento de distribución normal de los valores de sus observaciones.

Entre las pruebas paramétricas se encuentran la prueba Z, la prueba t-Student (para muestras independientes o para muestras pareadas), la prueba F y la prueba de correlación de Pearson y su aplicación depende del cumplimiento de ciertas condiciones relacionadas al modelo matemático que las define o el parámetro de interés.

Entre las pruebas no paramétricas se encuentra la chi-cuadrada, la prueba exacta de Fisher, la prueba de Mann-Whitney, la prueba de Kruskal-Wallis, entre otras. Se aplican en función de si se tienen muestras independientes (no relacionadas) o pareadas (relacionadas) y según la escala de medición.

Cada una de estas estadísticas tiene su propia distribución y todos los valores posibles que pueda asumir son puntos representados sobre el eje horizontal de una gráfica. En el caso de la distribución normal estos valores reciben el nombre de valores z .

En este eje se hace una división para establecer la zona de rechazo y la conforman aquellos valores que tienen menor probabilidad de ocurrir si la hipótesis nula es verdadera, para lo cual se tiene que fijar el nivel de significancia del estadístico de prueba al que se designa con la letra griega α . El valor más ampliamente usado es de 0.05, que implica el rechazo de la hipótesis nula con la probabilidad de cometer un error debido al muestreo del 5%. También se puede establecer un nivel de confianza más estricto como el de 0.01 y aquí la probabilidad de equivocarse al rechazar la hipótesis nula es del 1%.

Región de rechazo

En términos generales, la fórmula para calcular cualquier prueba estadística es la siguiente:

$$\text{Estadística de prueba} = \frac{(\text{estadística relevante} - \text{parámetro supuesto})}{(\text{error estándar de la estadística relevante})}$$

Para el estadístico z sería de la siguiente forma:

$$z = (X - \mu_0) / [\delta / \sqrt{n}]$$

Donde: μ es el valor supuesto de la media de una población.

Regla de decisión

Con los datos obtenidos en la muestra estudiada se calcula la prueba estadística y su resultado es un valor particular observado en esa muestra, el cual sirve para tomar las decisiones de rechazar o no la hipótesis nula.

El valor obtenido en la muestra se compara con los valores de la distribución de prueba de hipótesis (en las tablas estadísticas específicas) para conocer si el valor calculado cae en la zona de rechazo. La hipótesis nula se rechaza en favor de la hipótesis alterna cuando

el valor calculado está en la zona de rechazo, pero si ese valor no se encuentra en esa zona entonces no se puede rechazar la hipótesis nula, por lo que habría que aceptar que no hay diferencias. Así que las dos conclusiones posibles de un análisis por prueba de hipótesis estadística son: **rechazar H_0** o **no rechazar H_0** .

A pesar de lo anterior, una hipótesis estadística nunca se considera “probada”, incluso si es rechazada por los datos. Por ejemplo, la hipótesis de que dos poblaciones tienen la misma media puede ser rechazada fácilmente cuando ambas difieren mucho y se analizan muestras lo bastante grandes de ambas poblaciones, pero no podrá ser “demostrada” la diferencia mediante muestreo, puesto que siempre cabe la posibilidad de que las medias difieran en una cantidad lo suficientemente pequeña para que no pueda ser detectada, aunque la muestra sea muy grande.

Como ya se mencionó, el procedimiento de prueba de hipótesis depende de la información obtenida en la muestra. Si esta información es consistente con la hipótesis alterna, se concluye que ésta es verdadera; sin embargo, si esta información es inconsistente con la hipótesis alterna, se concluye que ésta es falsa. Nunca puede conocerse con certidumbre, a menos que fuera posible examinar a toda la población; por tanto, es necesario la mayoría de la veces desarrollar un procedimiento de prueba de hipótesis teniendo en cuenta la probabilidad de llegar a una conclusión equivocada (cometer un error).

Tipos de error en pruebas de hipótesis

El **error tipo I** o también conocido como **error α** , es el error que se comete cuando se rechaza una H_0 que es verdadera y el **error tipo II** o **error β** se define como la aceptación de la hipótesis nula cuando ésta es falsa.

Se llama poder estadístico de la prueba ($1 - \beta$) a la probabilidad de rechazar la hipótesis nula H_0 cuando efectivamente existe una asociación o diferencia en la población. El cálculo del poder estadístico del estudio permite identificar un efecto de una magnitud determinada si realmente dicho efecto existe. En general, antes de realizar el estudio se determina la magnitud del error α que se está dispuesto a asumir, normalmente 0.05. La probabilidad de aceptar la hipótesis nula cuando es verdadera, está dada por el valor de $1 - \alpha$; en otras palabras, esto constituye el nivel de significancia estadística (por lo general 95%).

Por tanto, al probar cualquier hipótesis estadística, existen cuatro situaciones diferentes que determinan si la decisión final es correcta o errónea, mismas que se muestran en el cuadro 17-1.

Cuadro 17-1. Cuatro situaciones que definen si la decisión final es correcta o errónea al probar cualquier hipótesis estadística.

Decisión	H_0 es verdadera	H_0 es falsa
Aceptar H_0	No hay error $1 - \alpha$ (nivel significancia)	Error tipo II o β
Rechazar H_0	Error tipo I o α	No hay error $1 - \beta$ (poder estadístico)

En general, el nivel de significancia o la región crítica de rechazo usado es de 0.05, en correspondencia al valor α . El valor para β convencional es de 0.20, lo cual representa un poder estadístico de 80% ($1 - 0.20$) que es aceptable en muchos estudios clínicos y epidemiológicos. Tanto el valor α como β son determinantes en la mayoría de las fórmulas para calcular tamaño de muestra. En esas fórmulas se pide declarar el valor z para un valor $\alpha = 0.05$ (1.96 que debe leerse en prueba de dos colas) y de $\beta = 0.20$ (0.84 que debe leerse en prueba de una cola). El primer valor ($1 - \alpha = 0.95$) representa el nivel de confianza de la muestra y el segundo ($1 - \beta = 0.80$) representa el poder estadístico de la muestra (o probabilidad de encontrar diferencias estadísticas).

Significancia estadística

Como ya ha sido señalado, el nivel de confianza está determinado por el nivel de aceptación del error α . La diferencia $1 - \alpha$ expresada en porcentaje representa el nivel de confianza del estudio.

La decisión de rechazar o no la hipótesis nula depende de la probabilidad de error α que sea obtenida después de realizar la prueba de hipótesis estadística y de consultar el **valor p** en las tablas. El valor p representa la **probabilidad de cometer el error α** . Si el valor p es menor al valor α (convencionalmente < 0.05), se decide rechazar la hipótesis nula. Si el valor p es mayor al valor α , no se puede rechazar la hipótesis nula y se dice que el estudio falló en rechazar la hipótesis nula. Otra forma de interpretar lo anterior con respecto al valor p consiste en estimar la probabilidad de que las diferencias observadas sean debidas al azar. Se acepta como máximo de esta probabilidad de ocurrencia del azar el valor correspondiente al error α (por lo general 5%).

Significancia clínica-biológica vs significancia estadística

Hay dos elementos que deben ponderarse, pues son relevantes en el momento de decidir la importancia de los resultados de un estudio en salud pública.

La significancia estadística representa la estimación de probabilidades de cometer errores (α y β) ante la decisión de rechazar o no la hipótesis nula. Si bien existen estimaciones puntuales de ello (valor p) y de rango (intervalos de confianza), las cuales son precisas en términos cuantitativos, dependen en gran medida del tamaño de muestra del estudio y del procedimiento de muestreo llevado a cabo.

No obstante, la decisión de la **significancia** de los hallazgos debe obedecer a un balance entre la relevancia obtenida mediante el procedimiento de prueba de hipótesis estadística y el peso de los resultados ante una traducción clínica-biológica de los mismos (figura 17-1).

En otras palabras, es posible tener significancia estadística y no tener significancia clínico-biológica, y viceversa (recuadro 17-2).

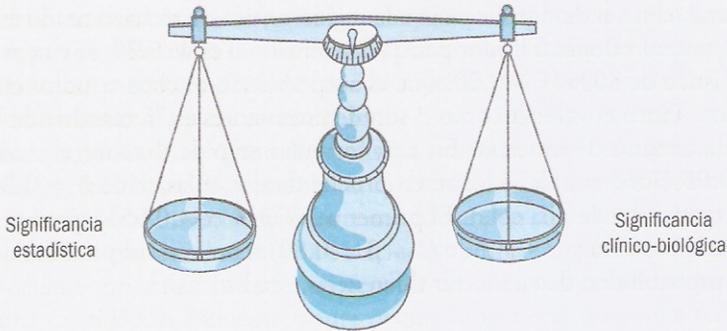


Figura 17-1 Equilibrio entre las significancias estadística y clínico-biológica.

Recuadro 17-2 Ejemplos 2 y 3.

Diferencia en las medias de niveles de colesterol total en escolares del sexo masculino vs femenino.

μ_1 = media de colesterol total en escolares de sexo masculino

μ_2 = media de colesterol total en escolares de sexo femenino

$$H_0: \mu_1 = \mu_2$$

Se realiza el estudio en una gran muestra $n = 100\,000$ (50 000 niños y 50 000 niñas).
Se encuentra lo siguiente:

$$\mu_1 = 150.05 \text{ mg/100 ml}$$

$$\mu_2 = 150.04 \text{ mg/100 ml}$$

Se hace la prueba de hipótesis por medio de prueba t-Student y se encuentra un valor de $p = 0.0001$.

Se rechaza la hipótesis nula.

Se concluye que el nivel de colesterol en niños es estadísticamente diferente al de las niñas, con un alto valor de significancia.

La pregunta que sigue es: ¿cuán significativa, desde el punto de vista clínico-biológico, es esa diferencia de una centésima de mg/100 ml?

Conclusión: grandes tamaños de muestra llevan a diferencias altamente significativas.

Ejemplo 3.

Diferencia en las medias de circunferencia de cintura en escolares de zona urbana vs rural.

μ_1 = media de cintura en escolares de zona urbana

μ_2 = media de cintura en escolares de zona rural

$$H_0: \mu_1 = \mu_2$$

Se realiza el estudio en una muestra $n = 100$ (50 escolares de zona urbana y 50 de zona rural). Se encuentra lo siguiente:

$$\mu_1 = 78 \text{ cm}$$

$$\mu_2 = 68 \text{ cm}$$

Se hace la prueba de hipótesis por medio de prueba Mann-Whitney y se encuentra un valor de $p = 0.06$.

No se rechaza la hipótesis nula.

Se concluye que la media de cintura en escolares de zona urbana no es diferente a la de escolares de zona rural.

La pregunta que sigue es: ¿la diferencia de 10 cm de cintura, mayor en escolares de zona urbana que de zona rural es significativa desde el punto de vista clínico-biológica, aunque desde el punto de vista estadístico la significancia haya sido limítrofe?

Conclusión. Pequeños tamaños de muestra fallan en encontrar diferencias significativas.

Bibliografía

- Asociación de Medicina Interna de México. *Temas de Medicina Interna. Epidemiología Clínica*. Director Huésped Dr. Juan José Calva Mercado. Vol. II Núm. 2, Interamericana McGraw-Hill, México DF. 1994.
- Bhattacharyya GK, Johnson RA. *Statistical Concepts and Methods*. John Wiley & Sons. 1977.
- Daniel W. *Bioestadística: Bases para la aplicación a las Ciencias de la Salud*. 4a Edición, Editorial Limusa Wiley, México. 2002.
- Moreno-Altamirano L, Güemez SC, López-Moreno S. *Factores de riesgo en la comunidad. Elementos para la salud colectiva*. UNAM. 1990.
- Pagano M, Gauvreau K. *Fundamentos de Bioestadística*. International Thomson Editores. 2001.