

Bioestadística y Epidemiología

ANÁLISIS DE VARIANZAJORGE DAGNINO S.¹

- El ANOVA es un conjunto de técnicas estadísticas de gran utilidad y ductilidad. Es útil cuando hay más de dos grupos que necesitan ser comparados, cuando hay mediciones repetidas en más de dos ocasiones, cuando los sujetos pueden variar en una o más características que afectan el resultado y se necesita ajustar su efecto o cuando se desea analizar simultáneamente el efecto de dos o más tratamientos diferentes.
- La forma más simple es el llamado ANOVA de una vía o factor, cuando existe una sola variable independiente para clasificar a los sujetos y dos o más niveles (que definen los grupos) de ella.
- Las otras formas de ANOVA (de 2 o más factores o de medidas repetidas) son extensiones basadas en el mismo raciocinio.
- El lector debe tener una comprensión del raciocinio global y la manera de presentar los resultados del ANOVA para que este resulte inteligible. El uso de las formas más elaboradas requiere de la asistencia de un estadístico profesional.

El análisis de la varianza (ANOVA, ANalysis Of VAriance, según la terminología inglesa) es uno de los aspectos más interesantes dentro del tema de las pruebas de hipótesis, por el ingenio desplegado en su desarrollo y, quizás, por las variadas formas que puede tomar. El ingenioso fue Ronald Aylmer Fisher (1890-1962), un genetista que fue uno de los estadísticos más influyentes del siglo XX. El tema puede llegar a ser muy complejo; existen extensos textos dedicados íntegramente al ANOVA, por lo que este artículo en ningún caso pretende cubrir todos sus aspectos y sólo debe ser considerado como una introducción. El enfoque, como el de los te-

mas anteriormente tratados, es más bien conceptual y de explicación de la terminología. Los objetivos perseguidos son que sea útil para comprender mejor los resultados de un trabajo en el que se usó un ANOVA y descubrir posibles errores en el uso de la técnica, para identificar situaciones en las cuales se debió haber usado y no se hizo y, quizás, usar adecuadamente la técnica aplicando un programa computacional. Esto es particularmente importante pues estos programas son bastante fáciles de usar pero la validez de los resultados depende esencialmente de una apreciación correcta de la estructura de los datos recolectados.

No existe un solo ANOVA sino que el ANOVA comprende una serie de técnicas cuya aplicación particular depende del diseño experimental. Para sus usos más complejos, es mejor pedir tempranamente ayuda a un estadístico profesional para el diseño y análisis de los datos y no lamentar después los errores que se pudieron haber evitado.

El ANOVA permite analizar la variación en una variable de respuesta (variable continua aleatoria) medida en circunstancias definidas por factores discretos (variables de clasificación). Se usa un ANOVA en cuatro situaciones:

- 1) Cuando hay más de dos grupos que necesitan ser comparados. El ANOVA también puede ser usado para comparar solamente dos grupos; de hecho, el test t de Student es un caso especial de ANOVA de una vía.
- 2) Cuando hay mediciones repetidas en más de dos ocasiones o cuando hay dos o más grupos en quienes se hacen mediciones repetidas en dos ocasiones.
- 3) Cuando los sujetos pueden variar en una o más características que afectan el resultado y se necesita ajustar su efecto.
- 4) Cuando se desea analizar simultáneamente el efecto de dos tratamientos diferentes, cuando el efecto de cada uno por separado y su posible interacción es importante.

¹ Profesor Titular
División de Anestesiología. Pontificia Universidad Católica de Chile.

El raciocinio detrás del ANOVA se basa en ideas y conceptos presentados en artículos anteriores. En las pruebas de hipótesis, se vio que para comparar dos grupos se pueden usar pruebas como la *t* de Student o sus equivalentes no paramétricos, pero cuando hay varios grupos el problema se complica. La tentación obvia, y también factible, es la de hacer la comparación entre cada par usando repetidamente el test *t* de Student. Sin embargo, este es el error más frecuente en la literatura biomédica en lo que al uso de pruebas de hipótesis se refiere. Ello porque las comparaciones múltiples alteran el valor de α en la globalidad de los datos, aunque no en cada comparación, y estos deben ser considerados en su conjunto pues el diseño experimental así lo exige desde su planteamiento. Cuando hay más de dos grupos es indispensable usar primero una prueba que permita analizar los datos en una sola instancia para ver si hay alguna diferencia. Solo si esta existe se puede proceder a estudiar el problema de las diferencias entre los grupos individuales, tema de otro artículo, el de las comparaciones múltiples.

ANOVA DE UNA VÍA

En el ANOVA, el término factor se refiere a la variable que determina los grupos del estudio, o sea, la variable independiente o predictora. El número de grupos definido por un factor se conoce como el número de niveles del factor. En medicina, estos corresponden habitualmente a los tratamientos que se comparan. Cuando hay un solo factor de clasificación de los datos, se habla de un **ANOVA de un factor, de una vía o de un sentido**. Por ejemplo, un trabajo que investigue si hay asociación entre el anestésico usado y la producción de carboxihemoglobina luego de una hora de exposición puede ser analizado con un ANOVA de un factor o vía (anestésico halogenado usado), con tres niveles (sevoflorano, isoflurano, desflurano) con respecto a una variable dependiente: nivel máximo de carboxihemoglobina. El raciocinio detrás del ANOVA se basa en la hipótesis nula que no hay diferencias entre los niveles de carboxihemoglobina lo que implica que las muestras provienen de una misma población, con una media y varianza determinadas. Si la hipótesis nula es verdadera, se espera que la variabilidad de cada grupo por separado sea similar a la variabilidad global, considerando los grupos en conjunto. En cambio, si la hipótesis nula no es verdadera, la variabilidad global es proporcionalmente mayor que la estimada dentro de los grupos. Esencialmente, en un set de datos cualesquiera, se

compara la variabilidad entre individuos dentro de cada grupo con aquella que proviene de cualquier diferencia sistemática entre los grupos. En jerga estadística, se calcula la variabilidad total de un set de datos (diferencia al cuadrado de cada valor individual con la media global) y la de los componentes de las diversas fuentes de variabilidad. Para ello, se calcula la varianza de las medias de los grupos (varianza ENTRE, “between”) y la varianza dentro de cada grupo (varianza DENTRO; “within”) también llamada ERROR o RESIDUAL. Se denomina así pues, luego de explicar la parte de la variabilidad total que corresponde al factor que interesa analizar, queda esta variabilidad no explicada y que puede tener diversos orígenes como, por ejemplo, error en las mediciones (y de aquí su nombre) o en la variabilidad individual. Si los grupos provienen de una misma población, las varianzas calculadas serán muy similares entre sí, por lo que su razón será cercana a 1. Esta razón es un número denominado *F* (en honor a Fisher), un estadístico como la *t* de Student o la *r* de Pearson, cuya distribución es conocida cuando la hipótesis nula es verdadera.

$$F = \frac{S_{\text{entre}}}{S_{\text{dentro}}}$$

Por el contrario, si la media de uno o más de los grupos se apartan del resto, la varianza **ENTRE** los grupos será desproporcionadamente mayor, mayor cuanto más se aparten las medias el valor de *F* diferirá progresivamente de 1 y la probabilidad que la hipótesis nula sea verdadera, el valor de *p*, es menor. (Figura 1).

La secuencia de los cálculos y el resultado de ellos puede resumirse así: la variabilidad total es la suma de los cuadrados de las diferencias de cada una de las observaciones con la media global (Tabla 1). Este total se divide en varianza **DENTRO** de los grupos, que es la suma de los cuadrados de la diferencia entre cada observación y la media del grupo respectivo, y la variabilidad **ENTRE** los grupos, calculada como la suma de los cuadrados de las diferencias entre la media de cada grupo y la media global. Cada una de estas variabilidades es reducida a una varianza estimada, conocida como media cuadrática (*mean sum of squares*), dividiéndola por los respectivos grados de libertad. Es la razón entre estas medias cuadráticas la que se usa para calcular el valor de *F*.

Los grados de libertad tienden a confundir. En su forma más simple, representan la manera en que

el número de muestras y el tamaño de estas entran en las fórmulas matemáticas usadas para calcular un determinado estadístico y, por lo tanto, para calcular las tablas con una hipótesis nula verdadera contra la cual se compara el valor obtenido del estadístico. Para la varianza, son uno menos que el número de observaciones: grupos -1 y la suma del número de observaciones en cada grupo -1. Los grados de libertad del error (dentro) son iguales a la diferencia entre los dos anteriores.

Los resultados del ANOVA pueden, y deben ser resumidos, aunque raramente se cumple, como se observa en la Tabla 1. Si no se presenta la tabla en los resultados, por lo menos debiera citarse los grados de libertad, el valor de F y la p correspondiente.

La media cuadrática dentro de los grupos también se llama residual o error pues es lo que resta luego de sacar el efecto de grupos: mide la variación aleatoria luego de restar el efecto sistemático. Tiene la misma connotación estadística que los residuales en la regresión lineal.

Los programas estadísticos computacionales arrojan el valor de p exacto para un determinado valor de F. Sin embargo, para una mejor comprensión del punto, vale la pena que el lector consulte una tabla de F, en el apéndice de cualquier libro de bioestadística, las que muestran, para los respectivos grados de libertad, los valores que delimitan el 5% (o el 1%) del área bajo la curva de los posibles valores de F si la hipótesis nula fuese verdadera. Valores sobre estos límites hacen cada vez más improbable que la hipótesis nula sea verdadera y se acepta entonces la hipótesis alternativa.

Si el resultado del ANOVA arroja que es poco probable que la hipótesis nula sea verdadera, o sea, el valor de F calculado es suficientemente grande y se asocia a un valor de $P < \alpha$ (u otro nivel α elegido), entonces se acepta la hipótesis alternativa: existen diferencias entre los grupos. Sin embargo, queda el problema de averiguar cuál, o cuáles, son los grupos distintos. Este problema será visto en el artículo sobre comparaciones múltiples.

Se pueden calcular **intervalos de confianza** para la media de cada grupo de la manera habitual excepto que el error estándar que se usa es aquel

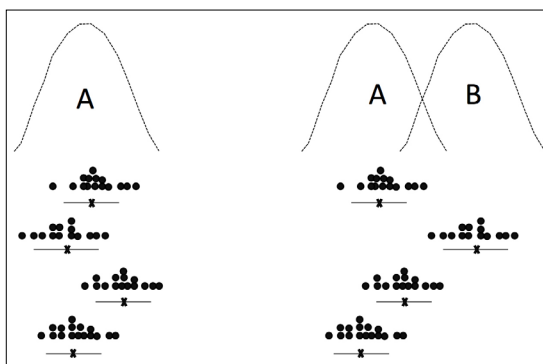


Figura 1. La figura pretende explicar gráficamente el raciocinio del ANOVA y el cálculo de F. A la izquierda hay cuatro muestras que provienen de una misma población A, representadas por sus observaciones individuales y resumidas cada una por su media y su variabilidad DENTRO de cada muestra. Como las muestras provienen de la misma población la variabilidad ENTRE las muestras es similar a la DENTRO por lo que el valor de F será cercano a 1. En la parte derecha, la segunda de las muestras proviene de otra población (B); la variabilidad DENTRO, en el denominador del cálculo de F, sigue siendo la misma, pero es evidente que la variabilidad ENTRE, en el numerador, es mayor lo que hace aumentar correspondientemente el valor de F y hace disminuir el valor de p.

basado en la desviación estándar residual. Esta está representada por los cuadrados de medias dentro de los grupos.

En un ANOVA, hay varias presunciones que deben cumplirse:

- 1) Las observaciones son independientes y constituyen una muestra aleatoria; el valor de una no se relaciona, de ninguna manera, con otra.
- 2) Las muestras provienen de poblaciones con una distribución normal con la misma desviación estándar. Esto no debe ser asumido sino que verificado a través de un gráfico que muestre la dispersión de los valores en cada grupo, un histograma de frecuencias, y de un gráfico normal de los residuales. Por definición, la distribución de estos debe ser normal si la población de la cual viene la muestra es normal.

Tabla 1. Resumen de los resultados de un ANOVA de una vía en que se comparan 3 grupos con 23 sujetos en total

Fuente de variación	GL	Σ de los cuadrados	Media cuadrática	F	p
Entre grupos	2	15516	7758	3,49	0,05
Dentro de los grupos (error o residual)	20	44460	2223		
Total	22	59976			

- 3) La varianza es igual en cada grupo (homocedasticidad).
- 4) La variable dependiente debe ser medida en una escala de intervalo.
- 5) Cuando hay más de una variable independiente, los efectos deben ser aditivos. Esta es una pre-supunción de los llamados modelos lineales simples de los cuales ANOVA y la regresión lineal son ejemplos.

OTROS DISEÑOS DE ANOVA

Muchas veces se desea estudiar el efecto simultáneo de dos o más variables o factores. El procedimiento es una extensión del ya visto y se denomina, cuando son dos los factores, **ANOVA de 2 vías**. La varianza total en este caso también se divide en varianza ENTRE los grupos y la varianza DENTRO o residual. Sin embargo, la varianza DENTRO se parte en dos para dar cuenta del efecto separado o combinado de los dos factores independientes. Por ejemplo, evaluar el efecto del índice de masa corporal (IMC) y de la posición del paciente (horizontal o semisentado) en la velocidad de desarrollo de hipoxemia en tres grupos de 20 pacientes con peso normal, sobrepeso u obesos. En este caso pueden establecerse tres preguntas:

- ¿Cuál es el efecto del IMC, independiente de la posición?
- ¿Cuál es el efecto de la variable posición, independiente del IMC?
- ¿Cuál es el efecto combinado o la interacción entre IMC y posición?

En este caso, una tabla del resultado del ANOVA tomará una forma similar a esta:

Los números se han ajustado para dar el valor exacto de p en relación con los grados de libertad involucrados en el cálculo de F (Entre/Dentro).

Un ANOVA de dos vías requiere el mismo número de observaciones en cada celda de la clasificación cruzada de los dos factores. Esto es frecuen-

temente un problema en estudios observacionales que raramente cumplen con este requisito. En estos casos el análisis apropiado es una regresión múltiple.

En resumen, ANOVA es un conjunto de técnicas usadas en relación con un diseño experimental particular. Para determinar qué tipo de ANOVA es apropiado, debemos conocer las relaciones entre los factores y las unidades experimentales (generalmente los individuos, pero puede ser una institución o una célula) a las cuales se aplican dichos factores. Así, el ANOVA puede dividirse según si los efectos son fijos, aleatorios o mixtos. Los primeros se refieren a aquellos que han sido elegidos de antemano por el investigador: determinados minutos, dosis o drogas. Son aleatorios, en cambio, cuando la elección se ha hecho al azar dentro de un conjunto de infinitos elementos posibles, usado en el análisis del error relacionado a métodos de medición. El ANOVA de una vía puede ser de efectos fijos o de efectos aleatorios; el ANOVA de dos vías, además puede ser de tipo mixto. Otros diseños son ANOVA factorial (cuando hay varios factores que pueden influir sobre el resultado) o ANOVA para medidas repetidas (cuando el grupo de estudio es sometido a condiciones diferentes, frecuentemente a lo largo del tiempo); esto no incluye a las mediciones seriadas donde las condiciones no cambian. Otros que se pueden mencionar son el ANCOVA (análisis de covarianza, una especie de mezcla entre ANOVA y regresión lineal) y MANOVA (ANOVA multivariado cuando hay más de una variable dependiente). Aun cuando difieren en cálculos y detalles, todas estas técnicas se basan en la misma estructura de comparar las varianzas dentro y entre los grupos.

La alternativa no paramétrica para ANOVA de una vía es el test de Kruskal-Wallis, que se usa con más de dos muestras independientes, y el test de Friedman cuando son dos o más muestras pareadas. No hay alternativas no paramétricas para el ANOVA de dos vías.

	GL	Σ de los cuadrados	Media cuadrática	F	p
ENTRE IMC	2 (3-1)	20320	10160	5,02	0,01
ENTRE Posición	1 (2-1)	14431	14431	7,13	0,01
ENTRE Interacción (3-1) (2-1)	2	6419	3,17	0,05	
Error o residual (o DENTRO)	54	109296	2024		
Total	59	156885			

REFERENCIAS

1. Altman DG, Bland JM. Statistics Notes: Comparing several groups using analysis of variance. *BMJ* 1996; 312: 1472-1473.
2. Altman DG. *Practical Statistics for Medical Research*. London: Chapman & Hall, 1991.
3. Armitage P, Berry G. *Estadística para la investigación biomédica*. 3a ed. Barcelona: Harcourt Brace, 1997.
4. Bland M. *An Introduction to Medical Statistics*. 3rd Ed, Oxford: OUP, 2006.
5. Evans SJW. Uses and abuses of analysis of variance. *Br J Clin Pharmacol* 1983; 15: 629-648.
6. Glantz SA. *Primer of Biostatistics*. 3a edición, New York: McGraw-Hill, 1992.
7. Larson MG. Analysis of Variance. *Circulation* 2008; 117: 115-121.
8. Ludbrook J. Repeated measurements and multiple comparisons in cardiovascular research. *Cardiovasc Res* 1994; 28: 303-311.
9. Matthews JNS, Altman DG, Campbell MJ, Royston P. Analysis of serial measurements in medical research. *Br Med J* 1990; 300: 230-235.
10. Pandit JJ. The analysis of variance in anaesthetic research: statistics, biography and history. *Anaesthesia* 2010; 65: 1212-1220.
11. Portney LG, Watkins MP. *Foundations of Clinical Research. Applications to practice*. 2nd ed. Upper Saddle River: Prentice-Hall, 2000.

Correspondencia a:
Dr. Jorge Dagnino S.
jdagnino@med.puc.cl