
Breves conceptos sobre el Anova

El Anova es una metodología de análisis estadístico para cuantificar y determinar las variaciones en poblaciones debidas a factores que permitan definir en ellas distintos estratos. Es ampliamente conocida por numerosos profesionales aunque no necesariamente comprendida de un modo adecuado por los no especialistas.

Para comprender con mayor claridad de que se trata con exactitud es necesario adentrarse más en los implícitos involucrados en su concepción.

Para comenzar el nombre de esta técnica deriva del acrónimo de ANalysis Of VAriance, es decir análisis de varianza. Fue introducida por el estadístico y genetista Sir Ronald Aylmer Fisher en 1925 y busca estimar y comparar las varianzas dentro de una población con las varianzas de subgrupos que pueden establecerse dentro de las mismas.

Quizá una aproximación elemental, aunque burda que nos permita adentrarnos en la información que busca proporcionar el anova sea considerar número de aciertos que tiene un grupo de tiradores de escopeta. Si se considera más acertado cuanto menos área en la diana abarquen los disparos, el área promedio de un grupo de tiradores es un indicativo grupal, pero si sabemos que algunos de los tiradores utilizan escopetas que aseguran muy poca dispersión de la munición mientras el resto lo hace con armas convencionales, casi no hace falta análisis estadístico para saber que la media del grupo estará influida en forma desigual por ambos sub promedios, lo que hace el anova es indicarnos con criterios estadísticos que los tiradores que utilizan escopetas de precisión contribuyen de un modo significativo al comportamiento del promedio grupal.

Este ejemplo es evidente, pero cuando hay poblaciones con una característica común que puede ser subdividida no siempre es evidente que hay una contribución estadísticamente desigual a la dispersión del grupo. A fin de aclarar los conceptos involucrados en el análisis de varianza nos vamos a servir de una serie de ejemplo de poblaciones:

- *Un conjunto de 120 de vehículos con tres años de uso y que no han presentado fallas.*
- *Ciento quince niños que egresan de primaria a final de un año escolar con un promedio de 18 puntos sobre 20.*
- *Sesenta y tres pacientes que han presentado mejora tras la aplicación de fármacos.*
- *Cuarenta y siete aspirantes a secretarias son capaces de escribir más de cincuenta palabras por minuto.*
- *Treinta y cinco formulaciones experimentales de una pintura resisten 4 horas en una cámara de envejecimiento acelerado sin perder su propiedades cromáticas.*
- *Veinticinco personas de un poblado tienen más de 100 años.*
- *El conjunto de 120 clientes que son atendidos en más de quince minutos en una estación de servicio.*

La característica básica de estas poblaciones es que son conjuntos con una propiedad en común. El investigador puede estar interesado en las causas que definen esa propiedad común. Si se dispone de datos puede ser de interés observar que si bien hay una propiedad en común, esta puede haber sido influida o generada por otras variables, en los mismos ejemplos:

- *La marca de los autos, el tiempo semanal de uso, el kilometraje recorrido cada mes o el tipo de lubricante que utilizan.*
- *La maestra del último año de los niños, el uso o no de determinado plan pedagógico, el aula, la disponibilidad de biblioteca o la disciplina de la escuela (porque no se ha dicho que provengan de una misma escuela).*
- *La marca del fármaco común, el uso de otro fármaco adicional, la ausencia o presencia de determinado tipo de alimento en la dieta o la localidad geográfica.*
- *La academia de la que egresan las secretarias, el tiempo de uso de computadoras ó el aprendizaje con métodos alternos.*
- *Presencia de compuestos comunes en la fórmula, cantidad de compuestos, presencia de dos compuestos comunes, tiempo de mezclado o condiciones particulares e identificadas de mezcla.*
- *Hábitos de dieta, hábitos de sueño, composición química del agua de la zona o conductas observadas y notorias.*
- *Persona que atiende, tipo de surtidor, surtidor específico o ubicación del surtidor.*

Estas subdivisiones llevan a preguntarse si la característica común se deberá a uno de estos factores internos, a una combinación de ellos o si por el contrario esta estratificación no tiene importancia en la conformación de la característica común. Al hacer esta última pregunta estamos planteando las bases para el análisis de varianza, que busca, siempre y cuando de disponga de los datos cuantitativos para estas características, identificar si hay una contribución significativa de alguno de los elementos clasificadores.



Responder esto puede tener implicaciones económicas y en otros casos nos pueden servir para identificar y normalizar mejores prácticas

En el lenguaje técnico la población en estudio (cada uno de los siete ejemplos anteriores) se dividen en bloques o estratos (cada una de las características internas en los grupos del ejemplo) también denominados niveles, grupos o tratamientos.

La pregunta planteada es la hipótesis básica de trabajo en Anova y mediante ella se desea saber si la dispersión de los datos se debe a causas atribuibles a alguno de los bloques o no, en este último caso correspondiendo a una distribución normal. El estadístico que mide la dispersión de los datos es la varianza y cuando la varianza de los grupos o estratos es igual a la de la población se dice que hay homocedasticidad y los factores medidos en cada uno de los bloques de las poblaciones no contribuyen significativamente a generar la diferencia o dispersión en caso contrario uno de los factores tiene una contribución significativa a la dispersión de datos y es por tanto causa de atención para el dueño del proceso.

El análisis de varianza comienza por dividir los datos de la población en los diferentes grupos. Se procede a calcular las sumas de cuadrados dentro de grupos, entre grupos y totales, SDE, SCE y SCT respectivamente.

Se continua calculando el número de grados de libertad dentro de grupos, entre grupos y totales, GLD, GLE y GLT respectivamente.

Posteriormente se calculan los cuadrados medios entre grupos CME y dentro de grupos como razón entre SCE y GLE y SCD y GLD respectivamente.

Se prosigue calculando el estadístico de contraste F como el cociente de CME y CMD.

La información final se suele presentar en una tabla como la siguiente:

Variabilidad	Suma de Cuadrados	Grados de Libertad	Cuadrado medio	Estadístico de Contraste
Entre Grupos	SCE	GLE	CME	F
Dentro de Grupos	SCD	GLD	CMD	
Total	SCT	GLT		

El análisis de varianza se basa en el hecho matemáticamente establecido de que hay una diferencia entre los grupos sólo si la varianza inter-grupos es mayor que la varianza intra-grupo.

Finalmente nos referimos a una tabla (en manuales estadísticos) que muestra qué valores puede alcanzar el coeficiente F cuando sólo actúa el azar. Si el F obtenido en el ejercicio es mayor que el valor de la tabla, hay una diferencia entre los grupos que es significativa.

El anova es una de las metodologías de análisis que se estudian con profundidad en los cursos de diseños experimentales.