FF 17 33

Carrera Ciencias de la Educación. Asignatura Pedagogía Experimental II, III y Adaptación. Título La técnica del análisis de la varianza.

Autor Ramón Pérez Juste. Grabación 7 de marzo de 1984. Emisión 17 de marzo de 1984.

Entre los grandes objetivos del Departamento de Pedagogía Experimental y Orientación en el área de la pedagogía experimental, se encuentran los del logro de capacidad de los alumnos para leer críticamente obras en que la metodología experimental sea utilizada, confiriendo valor o no a los resultados en función de dicha valoración crítica, y el de la capacitación para la planificación y realización de investigaciones con metodología experimental. Pues bien, en un sentido amplio, cualquier investigación requiere un plan, una estructura, un diseño, encaminado a dar respuesta a las preguntas derivadas del problema, o, si se prefiere, a poner a prueba la o las hipótesis avanzadas como tentativa de solución al mismo. Cuando la metodología de la investigación es la experimental, uno de sus momentos clave consiste en la utilización de la estadística en la prueba de las hipótesis.

Se trata de aplicar a los datos específicamente recogidos unas pruebas de significación. Esto es, unas pruebas que ayuden a decidir ante la alternativa, hipótesis de nulidad, hipótesis alternativa o sustantiva del investigador. Esto es, a si las diferencias apreciadas entre los diversos grupos son nulas, o lo que es lo mismo, pueden explicarse simplemente por azar, o por el contrario, si van más allá de tales límites, con lo que cabe pensar en alguna fuente productora.

Fuente que podríamos considerar como causa si el diseño tuviera validez interna. Esto es, si han sido formados los grupos en forma adecuada y durante el tiempo en que se ha desarrollado la investigación se han controlado las posibles fuentes de invalidez, con lo que se podría descartar, como hipótesis rivales, todas aquellas alternativas de la variable independiente. Las pruebas de significación disponibles son muy variadas, cada una de ellas con sus propias características y peculiaridades.

En esencia, se centran en la comparación de valores entre grupos, ya sea en el experimental y el de control, ya los diversos grupos actuando cada uno en función de control de los demás. Así, se pueden comparar valores positivos o negativos, rangos atribuidos a las puntuaciones de ambos grupos, frecuencias en las diversas celdillas, rachas o series de datos favorables o desfavorables, o, en el caso más sencillo y conocido, puntuaciones representativas de grupos, como puede ser la mediana o la media. Estas comparaciones dan lugar a diferencias cuyo valor ha de ser referido a algún otro, que, según los planteamientos anteriores, debe ser alguna estimación de esos mismos valores por puro azar.

Las pruebas se resuelven con la adjudicación a tales valores de una determinada probabilidad, con arreglo a la cual se decidirá el investigador a rechazar o no la hipótesis de nulidad,

aceptando o no, en consecuencia, su hipótesis alternativa. Pues bien, uno de los valores que pueden ser utilizados en las pruebas de significación puede ser la variabilidad o dispersión de las puntuaciones de los sujetos. Como sabemos, los sujetos, como consecuencia de la enorme variedad de factores que inciden en su origen y desarrollo, tienden a diferir unos de otros.

La mayoría de las variables medidas presentan, junto a una tendencia central, valor como la media, una mayor o menor dispersión o concentración en torno al mismo. A la variabilidad en cualquier variable contribuyen otras que desde ese punto de vista pueden ser consideradas fuentes de variabilidad, fuentes de varianza. Basta que tomemos, por ejemplo, una muestra de mil sujetos que acuden a un lugar lo suficientemente atractivo como para recibir personas de todas las edades y condiciones, tanto en sexo como clase social y nacionalidad.

Esto podría ser, por ejemplo, el caso de un estadio durante las Olimpiadas en un momento de una prueba de gran importancia. Si tomáramos esas mil personas hipotéticas, cabe pensar de inmediato una gran dispersión, por ejemplo, en la variable talla. Pero a la vez, y si decidiéramos quedarnos con mil sujetos de una única nacionalidad, presumiblemente el rango o amplitud total de la serie disminuiría.

Lo mismo ocurriría si lo hacemos quedándonos con sujetos de un único sexo o de edades comprendidas, por ejemplo, pues entre 12 y 20 años o de una clase determinada. Y todavía esto sería más patente si aplicáramos sucesivamente estas variables fuente de varianza, por ejemplo, sujetos suecos de sexo femenino entre 18 y 35 años de clase media. Naturalmente, el rango de la serie seguirá disminuyendo si somos capaces de seguir encontrando otras nuevas fuentes productoras de varianza.

La variabilidad o varianza debida a cada una de tales fuentes puede ser denominada varianza sistemática, mientras que aquella otra no identificada por desconocimiento o porque no interesa al investigador en ese caso, puede contraponerse a aquella como varianza aleatoria. Pues bien, uno de los métodos de análisis de datos más utilizados en la investigación experimental es justamente el del análisis de la varianza, debido a Ronald Fisher. Su esencia consiste en descomponer la variación total de los datos recogidos en dos o más partes.

Aquella parte presumiblemente debida a los efectos de la variable independiente, tratamiento experimental, y aquella otra que, al no ser identificada en su procedencia, se considera varianza de error, varianza aleatoria o casual. La varianza experimental puede ser única o incluir varias partes, tantas como variables independientes y sus correspondientes interacciones. Con todo, conviene hacer una aclaración previa, y es que en el análisis de la varianza no descomponemos la varianza en partes, sino la suma de cuadrados, que siendo una medida de variación o dispersión, tiene la propiedad de ser siempre aditiva, esto es, de que la suma de sus partes es igual a la suma total.

Con posterioridad, en el momento de decidir sobre la significación de las diferencias, se trabajará con las medias de cuadrados, que en realidad son varianzas, con la peculiaridad de que sus denominadores no son el número de sujetos de los grupos, sino los grados de libertad

correspondientes. Este hecho se debe a que estas varianzas son estimaciones insesgadas de la varianza de la población, aspecto clave, como veremos más adelante, en la decisión sobre la significación de diferencias. El análisis de la varianza puede realizarse mediante pruebas paramétricas y no paramétricas, como en el caso de la prueba de Friedman o la de Kruskal-Wallis, ambas aptas para mediciones ordinales.

La elección se basa en el cumplimiento o no de los supuestos del análisis paramétrico, más perfecto y potente que el no paramétrico. Por otra parte, el análisis paramétrico admite el de tipo simple, o en un sentido, cuando se trabaja con una sola variable independiente, y el llamado análisis factorial de la varianza, donde se examinan los efectos de dos o más variables independientes que actúan simultáneamente sobre la variable dependiente. Las variables independientes aquí reciben la denominación de efectos principales, que pueden ser fijos cuando el investigador decide los valores de la variable independiente directamente, por ejemplo, cuando escoge dos o tres o cuatro métodos, los que a él le interesan, aleatorios, si llega a ellos mediante una selección al azar, por ejemplo, cuando se trata de decidir qué momento es el óptimo para aprender determinada tarea, y, habiendo realizado previamente una serie larga de posibles tiempos, selecciona al azar aquellos que va a contrastar, y mixtos, cuando los correspondientes a una variable los decide él con base a sus necesidades o intereses, y los de la otra los toma al azar.

La importancia de este punto es grande, porque en cada caso el término de error en la prueba F, en la prueba del análisis de la varianza, es diferente, y de utilizarlo o de seleccionarlo mal se podrían llegar a conclusiones equivocadas. En el caso factorial, una situación problemática se da cuando los N, el número de casos de las casillas, son desiguales, especialmente si no fueran proporcionales, lo que es muy frecuente precisamente en la investigación no estrictamente experimental. Realmente ambas situaciones tienen una solución, aunque problemática, según dice Glase-Stanley, por lo que Carlinger se muestra partidario de resolver estos casos mediante la regresión múltiple.

Cabe además distinguir el análisis de la varianza cuando los sujetos son independientes de cuando son correlacionados. La técnica paramétrica difiere al identificar como varianza sistemática la debida a la correlación. En el caso no paramétrico, por el contrario, se utiliza un modelo específico, el modelo de Friedman para los casos correlacionados, y el de Kruskal-Wallis para los independientes.

En realidad, esta situación se da siempre que el investigador decide tratar de investigar o de identificar fuentes de varianza para reducir el término de error. Siempre que empareje sujetos con arreglo a una variable independiente relevante o siempre que construya bloques más homogéneos, habrá conseguido esta meta. Interesa que destaquemos ahora algunos puntos que fundamentan la técnica del análisis de la varianza.

Las dos partes fundamentales en que puede descomponerse o analizarse la variación de los datos experimentales, es decir, aquellos recogidos para probar la hipótesis, son aquellas cuya

fuente de variación ha sido identificada y aquella otra de origen desconocido, aunque se pudieran encontrar fuentes productoras de alguna porción de dicha variación. En la investigación experimental, el investigador manipulará la variable independiente justamente con la intención de comprobar que es capaz de producir efectos, de dar lugar a cambios, respecto de otras situaciones en que esto no ocurre o en que ocurre con determinada intensidad menor o en forma diferente, es decir, en que la variable independiente no actúa. Consiguientemente, antes de ejercer su influjo en la variable independiente, si el experimento está bien planeado, los grupos de sujetos que intervienen no diferirán significativamente, sino justamente dentro de los límites esperables como consecuencia de los efectos del azar.

Pero, tras el tratamiento experimental, uno o varios de esos grupos habrán experimentado cambios que les harán distintos de aquel o aquellos que no lo hayan recibido o lo hayan hecho en grado menor o en forma diferente. El caso es que los grupos tomados como unidad variarán tanto más fuertemente cuanto más eficaz haya sido el tratamiento o variable independiente. Como vemos, la variable independiente se convierte en una fuente de variación, es decir, da origen a varianza que llamamos sistemática, varianza que, por razones obvias, recibe el nombre de varianza entre grupos y se representa por la media de cuadrados entre grupos.

Ahora bien, los grupos varían también dentro de sí mismos al margen de los efectos de la presencia o ausencia de la variable independiente. Parte de esa variación es puramente casual, parte tendrá origen conocible, pero por diversas razones, como la complicación del diseño, la artificialidad o la dificultad de medición, el investigador no la identifica. Tal variación pasa a englobar lo que se considera variación debida al error o varianza de error o, si queremos, aleatoria o casual, entendida como varianza no experimental.

Su denominación es la de variabilidad dentro de los grupos y se representa por la media cuadrática o la media de cuadrados dentro de los grupos. Un principio fundamental del análisis de la varianza es justamente que ambas medias de cuadrados, la media de cuadrados entre grupos y la media de cuadrados dentro de los grupos, son estimaciones insesgadas del parámetro, esto es, de la varianza de la población, estimaciones insesgadas justamente porque no hemos dividido por el número de casos, sino por los grados de libertad. Consiguientemente, no debe diferir significativamente la una de la otra, sino dentro de los márgenes justamente del azar, salvo que algo exterior haya acontecido que haya hecho variar fuertemente a una mientras no ha alterado la otra.

Y esto es justamente lo que ocurre con el tratamiento experimental. Al hacer que uno o varios de los grupos, el grupo experimental o los grupos experimentales, reciban los efectos de la variable independiente, cambiarán elevándose, descendiendo sus puntuaciones respecto al de control, con lo que la varianza entre grupos aumentará. La media de cuadrados dentro de los grupos, por el contrario, únicamente tiene una fuente de variación.

El error o todo lo que no es varianza sistemática experimental, la media cuadrática entre grupos, incluye junto a ésta las variaciones producidas por la variable independiente. Si ésta ha

sido eficaz, su valor respecto de la varianza de la población irá más allá de los límites de las fluctuaciones del azar. En caso contrario, una y otra estarán muy próximas y cercanas a la varianza de la población, y desde luego dentro de las fluctuaciones del azar.

El análisis de la varianza pretende comprobar si la variación sistemática experimental es suficientemente mayor que la de error. Lógicamente, los valores no son absolutos. Tener unos 70 metros de altura puede ser mucho, poco o muy poco según nos refiramos a pigmeos, a niños o niñas de 14 a 16 años o a suecos.

Por ello es necesario referir la variación experimental a algo, y ese algo es la variación atribuida al error o a la casualidad. Es necesario que aquella sea suficientemente más elevada que ésta para atrevernos a afirmar que es significativa. No olvidemos el principio fundamental de la investigación experimental.

El investigador es una persona enormemente atrevida a la hora de formular sus hipótesis, pero muy precavida a la hora de probarlos. Ahora, el problema es cuánto mayor debe ser esa varianza experimental frente a la de error. Snedecor es el autor de unas tablas que nos indican los valores de f, razón entre ambas medias de cuadrados, que permiten afirmar la significación para los diferentes grados de libertad y niveles de significación.

Si la hipótesis de nulidad fuera cierta, ambas medias de cuadrados tenderían a ser iguales y su cociente muy próximo a la unidad. Por contra, si la hipótesis de nulidad fuera falsa, el valor de la media cuadrática entre grupos será sensiblemente superior al de la existente dentro de los grupos, dando lugar a un cociente cuya probabilidad de ocurrencia conforme a la hipótesis de nulidad tendría una probabilidad máxima igual al valor del nivel de significación o alfa decidido. Lógicamente, el valor de f será tanto más elevado cuanto mayor sea la media cuadrática o medio de cuadrados entre grupos y cuanto menor sea la que se da dentro de los grupos.

De ahí el principio de Karlinger para la finalidad del diseño entendido como maximización de la varianza experimental y, en este caso lo que nos interesa, minimizar la varianza de error. La varianza de error puede minimizarse por varios procedimientos. Cabe utilizar aquellos grupos correlacionados o la técnica del bloqueo en los que los sujetos o son los mismos o se han emparejado con respecto a alguna variable relevante.

Pero también se podrá hacer utilizando, en este caso, el efecto de la correlación que al identificar una nueva fuente de varianza sistemática dentro de la varianza total disminuirá el término de error. También cabe, por el contrario, utilizar o introducir dentro del diseño la variable que nos interesa controlar como edad, sexo, clase social, nivel académico, etc. con lo que los grupos serían más homogéneos pero, en uno y otro caso, las posibilidades de generalización de los resultados quedarían disminuidas reduciéndose a situaciones similares a las investigadas.

Cabe también utilizar diseños factoriales en que esas variables quedan incorporadas al diseño de forma que, sin reducir las posibilidades de generalización, la varianza debida a ellas sea

identificada y convertida en sistemática con lo que el término de error disminuirá y, además, se podrán determinar las posibles interacciones entre las variables. Después de haber considerado estos puntos referidos a la fundamentación del análisis de la varianza, nos interesa referirnos ahora a sus aportaciones. El análisis de la varianza concluye en una prueba de significación de diferencias, prueba F, que, en su sentido más elemental, permite comprobar si diversos niveles de una misma variable independiente dan lugar a diferencias significativas.

En el caso de dos niveles del tratamiento experimental, su potencia y sus efectos son sustancialmente iguales a los alcanzados en una prueba como T. Ahora bien, en el análisis de la varianza se permite determinar, además, de una sola vez si existen diferencias entre los efectos de tres o más niveles de la variable independiente. Además, y esto es más importante, posibilita determinar el efecto simultáneo de dos o más variables independientes sobre una variable dependiente, con la ventaja adicional de atribuir a cada una de aquellas su influencia relativa como proporción o tanto por ciento. En concreto, estamos hablando del tema de los componentes de la varianza.

Si de la media de cuadrados entre grupos restamos la media de cuadrados dentro de los grupos y dividimos por el número de casos de cada grupo, habremos localizado el componente de la varianza debido a la variable independiente. Ahora, dividiendo este componente por la suma del mismo más la media de cuadrados dentro de los grupos, podremos determinar el tanto por ciento del peso que la variable independiente tiene sobre la varianza total. Lógicamente, dividiendo la media de cuadrados dentro de los grupos por ambos valores, el componente de la variable independiente más dicha media de cuadrados dentro de los grupos, habremos tenido, multiplicando por cien, cuál es el componente o el peso que la varianza de error tiene en la varianza total.

Ahora bien, con todo, el gran valor del análisis de la varianza, sobre todo en el caso de las ciencias humanas, en que los fenómenos son muy complejos y las variables actúan en forma interactiva, es precisamente la posibilidad de determinar la existencia o no de interacciones significativas de una modalidad ordinal o no, y de su peso o contribución a la varianza total. No debemos olvidar que justamente en el ámbito de las ciencias humanas lo normal es que las variables incidan unas a la vez que otras o unas interactuando con otras. Pues bien, el análisis de la varianza va a permitir detectar justamente la existencia de interacción y su valor relativo.

Un problema adicional al que estamos considerando se refiere a las comparaciones posteriores al análisis de la varianza. No deberíamos olvidar que en el caso de este tipo de análisis, se pueden contrastar simultáneamente los efectos de dos o más variables independientes o de una variable independiente en tres, cuatro o n grupos en definitiva. Naturalmente, si los grupos son dos y la diferencia fuera significativa, ya se sabe a favor de qué grupos se produce.

Pero si hay más de dos grupos, será necesario proseguir el análisis para localizar las diferencias significativas. En algunos textos antiguos se recomendaba utilizar la prueba T para todos los pares de medias. Sin embargo, hoy en día sabemos que no es adecuado este procedimiento

estadístico porque se puede demostrar que el valor de alfa, el valor del nivel de significación, es en realidad bastante más elevado del que pretende el investigador, con lo que corre el riesgo de rechazar hipótesis de nulidad no rechazables.

En cambio de esta prueba T, se recomienda utilizar otras como la de Tukey en el caso de que los valores de n para los distintos grupos sean iguales, o el DSF válido para todas las ocasiones. Terminaremos esta exposición haciendo una breve referencia a los supuestos en que se basa el análisis de la varianza. El análisis de la varianza paramétrico en el cual fundamentalmente nos estamos centrando, hemos dicho con anterioridad que tiene un mayor nivel de potencia que el de las pruebas no paramétricas.

Ese mayor nivel de potencia se traduce en la posibilidad de llegar a rechazar hipótesis de nulidad que con un análisis no paramétrico quedarían como no rechazadas. Sin embargo, justamente porque posee esta mayor potencia y además porque tiene unas aplicaciones muy superiores a partir de sus datos, es una prueba más exigente en requisitos. Estos requisitos se pueden centrar fundamentalmente en los siguientes, en el supuesto de homogeneidad de las varianzas, en el de normalidad de la distribución y en el de la independencia de los valores seleccionados para las muestras.

Las posturas, llamémosle exigentes, como las de Siegel, dicen que cuando estos presupuestos no se pueden mantener el investigador debería ser lo suficientemente honrado como para decidir, optar por un procedimiento no paramétrico, aunque sea menos potente, y renunciar a la prueba paramétrica. Esta polémica que se ha producido entre los defensores de las pruebas paramétricas y no paramétricas ha llevado a que algunos investigadores hayan realizado ciertos trabajos tratando de evaluar hasta donde fuera posible el efecto que el incumplimiento de los supuestos del análisis de la varianza tiene sobre los resultados. En concreto, sobre el supuesto de homogeneidad, Vox nos informa del efecto que se produce.

Si el valor de n es igual en el caso de todas las celdillas, parece ser que, según estas investigaciones, el efecto no es preocupante. Ahora, en el caso de n desigual se pueden producir dos fallos graves ambos. Por ejemplo, el aumento de la probabilidad del error de tipo 1 cuando se seleccionan menos sujetos de poblaciones con mayor varianza.

Y por contra, el aumento de la probabilidad de tipo 2, de error de tipo 2, cuando se seleccionan más sujetos de poblaciones con mayor varianza. Naturalmente, después de estos estudios, se han ido desarrollando una serie de pruebas que pretenden decidir justamente sobre la homogeneidad o no de las mismas para ver si es conveniente o no la aplicación del análisis de la varianza. Entre ellas, nosotros solemos recomendar concretamente las pruebas como la de Harley, puesto que otras como la de Bartlett exigen cálculos más laboriosos.

También a veces utilizamos la prueba de Cochrane. Esos estudios a los que nos hemos referido también se han centrado en el tema de la normalidad. El hecho de que las poblaciones de las que se obtienen las muestras se distribuyan con arreglo al modelo normal o gaussiano.

Pues bien, los estudios que se han realizado vienen a decir que el incumplimiento del supuesto de normalidad no produce graves efectos sobre el nivel de significación. No así en el caso anterior, donde un alfa, por ejemplo, teórico, un nivel de significación teórico del 5%, según algunos estudios podría llegar al 19%. Por último, se trata del supuesto de independencia.

También aquí los estudios han resaltado la gravedad que se supone o que se produce sobre los efectos cuando variables que no son independientes, que tienen algún grado de correlación o de asociación, se tratan como si fueran independientes. En realidad, la técnica estadística está lo suficientemente desarrollada como para que en el supuesto de datos que estén asociados o correlacionados, se puedan utilizar técnicas diferentes a las propias del análisis parcial.