FF 17 23

Todo tiene la señal. Carrera, Ciencias de la Educación. Asignatura, Pedagogía Experimental II.

Título, El diseño de dos grupos, factores de validez. Autor, Carmen Jiménez Fernández. Fecha de grabación, 4 de marzo de 1983.

Fecha de emisión, 17 de marzo de 1983. Dado que este guión se emite cuando ya los alumnos han realizado la primera prueba personal y tienen constancia de la calificación obtenida, vamos a dedicar unos minutos a comentar los aciertos y fallos de la misma. Espero y deseo que estos comentarios contribuyan a aprovechar más racionalmente el tiempo dedicado al estudio de la disciplina y a situarse ante la realización del examen con una metodología más eficaz.

Pueden preguntar cuanto deseen, pero establezcamos un orden. Primero, nos referiremos a la prueba objetiva. En segundo lugar, a la resolución del ejercicio práctico o problema.

Y finalmente, a lo que es propiamente el título de este guión, el diseño de dos grupos como paradigma de investigación experimental. Con este esquema como referencia, pueden intervenir cuanto deseen. La prueba objetiva era del tipo verdadero o falso.

¿Cómo se corrigen estas pruebas? Pues algunos de nosotros, y yo en particular, no sabría si era mejor responder a todas las preguntas o solo a las que realmente se sabía. Responder por azar, y perdone mi franqueza, beneficia o perjudica la calificación. Como saben, las pruebas objetivas de lección múltiple se corrigen aplicando la fórmula de número de respuestas correctas menos el número de respuestas 5 restas divididas las 5 restas por el número de lecciones posibles menos 1. En el caso de las pruebas objetivas de verdadero o falso, las elecciones son solo 2, o acierto o error.

Luego, al dividir los errores por 2 menos 1, queda 1, lo que significa que se corrigen contando el número de aciertos y restándole el número de errores. Así, si un alumno tiene 20 aciertos y 0 errores, su puntuación final es 20 puntos, si cada pregunta vale 1 punto. Pero si tiene 20 aciertos y 4 errores, su puntuación final es de 16 puntos al restarle los 4 errores cometidos.

¿Por qué se procede así? Sencillamente para eliminar la influencia de un factor que usted ha citado en su pregunta, el azar. En efecto, supongamos que un alumno que no sabe nada de una prueba objetiva de verdadero o falso se pone a realizarla. Por azar, aproximadamente acertará la mitad de las veces y fallará la otra mitad.

Entonces, al tener tantos aciertos como errores, su puntuación final es 0. En cambio, el alumno que conoce y domina su contenido tendrá un número de aciertos muy superior al de errores. Quiero aclarar desde aquí que he dulcificado la calificación final de todos ustedes. Entonces, ¿se deben solo contestar las que se saben y no las que se ignoran o se dudan? No es así.

Bueno, ante cualquier prueba objetiva, pero sobre todo en las pruebas objetivas de verdadero

o falso y en las de lección múltiple, el procedimiento recomendado es el siguiente. Primero, ir leyendo cada una de las preguntas y responder con la primera respuesta que se cree cierta. Suele ser la verdadera.

Segundo, las que no se saben realmente, pasarlas en esta primera lectura o, en todo caso, ponerles una pequeña marca para volver sobre ellas. Tercero, las dudosas, saltarlas también y marcarlas al margen. Cuarto, cuando se han resuelto todas las cuestiones del examen que se dominan bien, se vuelve primero sobre las dudosas y, por último, sobre las que se cree no saber.

A veces, una lectura más detenida orienta la respuesta. Hay en ocasiones que el alumno piensa si me pidiera esto, pues sería verdadero, y si me pidiera esto otro, pues sería falso, etc. Al leer una pregunta, ¿qué aconseja que se haga? Elaborar buenos exámenes y buenas pruebas objetivas no estaría fácil.

Puede haber alguna pregunta realmente confusa, pero sólo alguna. Cuando el alumno tiene la certeza de que es así, puede indicar a pie de página que ha respondido como lo ha hecho por entenderla de determinada manera y que habría respondido de esta otra forma si quisiera preguntarse con ella este otro significado. Todo profesor normal entiende y valora esta aclaración.

Lo que no puede ser norma es el hacer aclaraciones sobre cada pregunta, aclaraciones no procedentes la mayoría de las veces. Hay que responder a lo que se pide. Un concepto puede tener más de una aplicación.

Responder si la aplicación contenida en la pregunta es correcta o incorrecta es la norma, ejemplo de este tipo de error. Una pregunta puede ser la siguiente. ¿Una variable continua es aquella que cambia fácilmente de valor? La respuesta correcta es verdadero.

Una aclaración improcedente y no válida sería añadirle previamente la inteligencia o la talla, etc., etc., es una variable continua. Si no tienen más dudas sobre la prueba objetiva, podemos pasar a comentar el ejercicio práctico. Yo personalmente he suspendido la prueba, aunque sé que justamente me despisté y no supe entender lo que se pedía en el problema.

¿Sucedió esto mismo a otros compañeros? Sí, sí fallaron otros compañeros suyos, aunque son muchos más los que la han superado que los que la han suspendido. Si se lee con objetividad y calma, pedía dos cosas. Primero, probar la hipótesis de normalidad de la distribución, prueba de bondad de ajuste.

Y segundo, justificar la elección de la prueba estadística. ¿En qué datos habría que apoyarse para elegir la prueba estadística adecuada y justificar así su elección? Fundamentalmente en dos. El primero, y más sencillo porque lo indicaba el propio enunciado del problema, en el interrogante planteado en primer lugar.

Recuerden que se daban una serie de datos brutos obtenidos en una muestra grande de

alumnos. Se pedía literalmente, prueba de la hipótesis de normalidad de la distribución. Segundo, al leer este enunciado, el alumno debe de saber que se trata de una prueba de bondad de ajuste.

Estudiar si la distribución de frecuencias de la muestra obedece o se adapta a la distribución de frecuencias de la curva normal. Ahora, y apoyándose en sus conocimientos sobre el tema, sabe que puede elegir entre dos pruebas, y prueba de Kolmogorov-Smirnov de una muestra. Dado el tamaño de la muestra, las dos son adecuadas y conducen a la misma conclusión.

Pero si un alumno hubiera hecho muchos intervalos, a lo mejor hubiera tenido que proceder a reagruparlos para poder aplicar G cuadrado por ser insuficientes en alguno de ellos las frecuencias esperadas o teóricas. Si a ello añadimos que es además menos potente que la prueba de Kolmogorov, ya tendría contestada también la segunda pregunta del problema. Bien, pues después de corregir todos los ejercicios, ¿cuáles han sido los errores más frecuentes cometidos en el problema? Bien, lo reduciría a dos grupos o categorías.

Primero, errores conceptuales o de interpretación. Segundo, errores o confusión entre los conceptos estadísticos y su manipulación práctica, entiéndase cálculo. Ejemplos de los primeros podrían ser los siguientes.

Confundir el concepto teórico de distribución muestrar, recuérdese el primer guión, con una distribución empírica de frecuencias. Más de un alumno, al leer lo de «pruebe la hipótesis de normalidad», ha querido probar la significación estadística de la media o de la diferencia entre pares de medias o entre dos proporciones, etcétera, etcétera. Y es que recordaban que la distribución muestrar de estos estadísticos es normal cuando N es grande, como era el caso.

Luego, acarreaban otras confusiones al verse obligados a «inventar» parámetros para probar esas hipótesis. Con frecuencia, no sabían interpretar el resultado obtenido que sí era interpretable. Ejemplos del segundo tipo de errores, menos dolorosos pero muy peligrosos por ser bastante generalizados, son no saber realizar un agrupamiento de intervalos o no saber manejar la tabla de la curva normal, herramientas necesarias para este y otros problemas.

Errores que cabalgan entre las dos categorías pueden ser definir correctamente el tamaño de la muestra y, a continuación, tomarlo como el valor del parámetro, definir y calcular la amplitud total correctamente y luego no saber operar con ella, etc. En resumen, aconsejo tres cosas. Una, claridad conceptual.

Dos, cierta soltura en el cálculo estadístico. Y tres, leer con calma los exámenes sin buscarle tres patas al gato. No van a sorprender, sino a estimular y evaluar el aprendizaje.

Y tras estos comentarios, pasemos al tercer punto, estudio de los diseños de dos grupos y de su validez. Estos diseños son ya diseños experimentales y no diseños pre-experimentales, como los del grupo único. Se podría decir que el objetivo de estos diseños es estudiar temas complejos mediante paradigmas o estructuras simples.

Y me explico. Estos diseños manipulan o miden una sola variable independiente, tipo de método o sexo, por ejemplo, para ver si esta variable puede explicar la variación observada al medir a los dos grupos en la variable dependiente, actitudes, por ejemplo. El diseño, como estructura, persigue mantener a ambos grupos en igualdad de condiciones extrañas para observar los efectos diferenciales, si los hubiere, de la variable independiente.

Supone implícitamente que puede existir una relación clara entre el tipo de método o el sexo y las actitudes medidas. Es más, desearía afirmar, aunque no lo hace el investigador, que la causa de determinadas actitudes es el aprender con uno u otro método o el pertenecer a uno u otro sexo. En nuestras ciencias, este diseño, por más estrictamente experimental que sea, nunca podrá probarlo anterior, al menos para variables complejas.

¿Por qué? Pues simplemente porque no hay una relación causa-efecto entre las variables estudiadas. Y es que, aunque el método o el sexo influya en la creación de determinadas actitudes, éstas dependen también de otras variables como puede ser el credo religioso, la edad, la inteligencia, etcétera, etcétera. Éstas y otras variables contribuyen también, bien paralelamente, bien de modo interdependiente, a forjar determinadas actitudes.

Es cierto que no todas las variables contribuirán en igual grado a explicar una variable dependiente, actitudes en nuestro ejemplo. Pero el diseño de dos grupos, al ser incapaz de manejar más de una variable independiente, no puede identificar nada más que un tipo de varianza. La varianza debida a otras variables independientes o a su interacción queda fuera de su alcance.

Entonces, ¿por qué se estudia no se emplean tanto? Al menos eso creo yo. Por varias razones. Una, que han sido y son instrumentos muy valiosos en ciencias más exactas que las nuestras.

Y dos, que aún para estudiar la conducta del hombre pueden arrojar y arrojan evidencias interesantes que orientan la práctica y la investigación educativas. En efecto, desde ambos puntos de vista más vale cierto conocimiento que la ignorancia total. Diseños más complejos se aplican a raíz de los resultados obtenidos con diseños más simples como son estos.

Entonces, ¿qué aportan realmente que no aporten los diseños de grupo único? Bueno, aportan un elemento decisivo para la ciencia, el control. Estos diseños permiten, en teoría, insisto, en teoría, igualar los dos grupos en todos los elementos o al menos en aquellos elementos o características que pueden ser relevantes para el efecto medido. Es decir, mediante diversas técnicas el investigador puede considerar que sus grupos son iguales en términos estadísticos.

Si tras establecer este control introduce una variable independiente que varía de modo distinto en cada uno de los dos grupos, si ha mantenido dicho control experimental puede presumir que si hay efectos distintos en cada grupo estos se deben a la variable independiente. Prácticamente afirma que la causa es la variable independiente aunque no puede precisar si interactuaban o no otras variables. Control quiere decir grupo de control.

Es decir, que hay dos grupos. Uno que es el grupo experimental y otro que es el grupo de control. Sí.

Quiere decir eso fundamentalmente pero también quiere decir algo más. El control se refiere, como hemos dicho, aunque sin precisarlo, a los siguientes puntos. Primero, a la capacidad del investigador para variar o medir la variable independiente, método o sexo en nuestro caso.

Dos, a la medida de la variable dependiente, actitudes. Es decir, control de las condiciones e instrumentos de medida. Tres, control en la forma de elegir a los sujetos, muestreo, y de asignarlos a los grupos y tratamientos.

Y cuatro, control de variables extrañas o, como hemos dicho, constancia de las condiciones durante el proceso de la investigación. Este diseño puede lograr el control de los tres últimos aspectos. Al tener dos grupos, si los elige, bien al azar, bien por alguna técnica de equiparación, puede suponer o comprobar que al iniciarse la investigación los grupos no difieren entre sí en términos científicos, al menos en variables relevantes.

Durante el proceso, si ambos grupos trabajan en condiciones similares, salvo en la modalidad de la variable independiente, cada grupo actúa como control del otro. Es decir, tanto si el grupo experimental recibe una cantidad X de variable independiente y el grupo de control una cantidad 0 de la misma, cuanto si reciben cantidades distintas de la variable independiente, cabe esperar efectos distintos en la medida de la variable dependiente, debidos precisamente a la variable independiente al estar controladas las variables extrañas. Como se recordará, en los diseños de un solo grupo el problema radica en saber si el efecto observado se debe realmente a la variable independiente.

Puede deberse, pero no es posible evidenciarlo porque existen otras explicaciones rivales, entiéndase variables extrañas, que pueden explicar dicho efecto, aunque en un estudio a lo mejor no lo explican. ¿Significa lo anterior que estos diseños tienen validez interna y externa? Sí y no. Es decir, pueden tener validez interna y externa, o sobre todo interna o especialmente validez externa.

Depende de cada diseño y de la posibilidad de aplicar el control en todas y cada una de las etapas. Ambos tipos de validez no van necesariamente unidas, pero tampoco son necesariamente contrapuestas, aunque pueden serlo. Dado que hoy no disponemos de más tiempo, digamos por último lo siguiente.

Desde el punto de vista teórico, este paradigma de investigación, con sus variantes, puede lograr ambos tipos de validez, lo que no quiere decir que la alcance. Ante cada investigación concreta habrá que ver qué variables extrañas son relevantes, cómo pueden ser controladas y si se han controlado. Es analizar críticamente el modo de elegir a los sujetos y de formar los grupos, de aplicar los tratamientos experimentales y de recoger sus resultados, y de discriminar entre aquellas variables extrañas que al interactuar con otras impiden generalizar a poblaciones no sometidas a algunas de las condiciones de la investigación.

El estudio obtenido de ejemplos de diseños y de las variables que controlan es sumamente interesante y como tal lo recomiendo a todos ustedes como forma de aprendizaje, aunque sea con ejemplos ficticios.