

CE 02 1122

Buenas noches, en Revista de Educación nos centraremos hoy en dos materias del curso de Adaptación, Pedagogía Experimental y Pedagogía Social y Sociología de la Educación. Dentro de un momento repasaremos los aspectos más importantes que hay que tener en cuenta al estudiar los diseños de investigación, así que como se trata de un asunto complejo y de orientaciones bastante puntuales para que ustedes puedan preparar esta parte de la asignatura de Pedagogía Experimental Adaptación, es conveniente que tengan a mano lápiz y papel para poder tomar alguna nota, incluso si es posible que graben este espacio, porque además repasaremos más tarde el programa de estudio de Pedagogía Social y Sociología de la Educación. Cada semana elaboran este programa Revista de Educación, profesores de la Facultad de Educación de la UNED y personal técnico de su Centro de Diseño y Producción de Medios Audiovisuales.

Entramos ya en materia con Pedagogía Experimental Adaptación, vamos a hablarles de los diseños de investigación en educación, tal como avanzábamos hace un momento. Esta es la parte que corresponde al segundo cuatrimestre de esta asignatura del curso de Adaptación. Nos acompañan los profesores José Luis García Llamas y Catalina Martínez Mediano.

Buenas noches. Comenzamos por aclarar cuáles son los bloques de contenido de esta segunda prueba presencial. Sí, a diferencia del primer cuatrimestre, diremos que en este segundo cuatrimestre la materia se centra en torno al estudio de los diseños de investigación.

Por ello, vamos a dividir la asignatura, aunque son tres unidades didácticas, en realidad hablamos de cuatro grandes bloques de contenidos que configuran estos contenidos. Así, en primer lugar abordaremos los aspectos generales relacionados con los diseños de investigación en general, para después centrarnos en los tres bloques siguientes en el análisis de los diseños de un grupo, los diseños de dos grupos y los diseños de más de dos grupos con las correspondientes pruebas estadísticas de resolución estadística. Pues a partir de este marco que nos ha hecho el profesor García Llamas, de los contenidos que corresponden a este cuatrimestre, iniciamos el repaso por los aspectos generales relacionados con el diseño de investigación.

Profesora Martínez Mediano. Bien, pues vamos a hablar de esto, del concepto de diseño de investigación, de diseño de investigación experimental. La investigación experimental consiste en modificar deliberadamente y de manera controlada las condiciones que determinan un hecho y observar e interpretar los cambios que ocurren en este último.

Kerlinger define el concepto de diseño experimental como el plan, la estructura y la estrategia de investigación concebidos para obtener respuestas a las preguntas de investigación y controlar la varianza. El plan es el esbozo general del proyecto de investigación, que incluye desde la formulación de la pregunta y de las hipótesis hasta el análisis de los datos. La estructura se refiere al esquema o papel que van a jugar las variables de estudio.

La estrategia se refiere a los métodos de recogida y análisis de datos. Bien, los objetivos del diseño de investigación son tres. El primer objetivo es responder a las preguntas de investigación.

Toda investigación busca la solución a un problema y éste se operativiza formulando la hipótesis como tentativa de acción. Podría decirse que hay tantos diseños como posibilidades tenga la hipótesis de ser puesta a prueba. Entonces, no habrá un único diseño ideal sino tantos como los problemas que se planteen.

El diseño sugiere las direcciones de la observación y del análisis. El segundo objetivo es asegurar la validez interna de la investigación. En una situación experimental, el experimentador debe preguntarse si la variación sufrida en la medición de la variable dependiente fue realmente provocada por la variable experimental activa por el tratamiento para asegurar esa validez interna del experimento.

El tercer objetivo de los diseños de investigación es la validez externa de la investigación. Tiene que ver con el poder de generalización de los resultados y para conseguir este objetivo está íntimamente relacionado con la definición de las poblaciones presentes en el experimento y la creación de muestras representativas. Podemos pasar ahora a hablar de los criterios para seleccionar el diseño de investigación.

Pues en función de la línea de investigación, en el grado de control, en la técnica de control y en el número de variables dependientes. Así, según la línea de investigación, tenemos diseños estadísticos y no estadísticos, según el grado de control, diseños pre-experimentales, cuasi-experimentales y diseños experimentales. Según las técnicas de control de las variables, tenemos que se puede realizar el control de las variables mediante la técnica de aleatorización, teniendo diseños de dos o más grupos seleccionados al azar.

También se puede controlar las variables mediante la equivalencia inicial en los valores de una determinada variable asociados a la variable independiente que se trata de estudiar. Otra técnica de control de las variables es la técnica de bloqueo, que consiste en dividir la muestra a estudiar en grupos atendiendo alguna característica denominada variable de bloqueo. Y ya por último, otro criterio de clasificación de los diseños de investigación es atendiendo al número de variables independientes objeto de estudio.

Así, tenemos diseños univariados, bivariados, multivariados. En estos aspectos generales que tenemos que tener en cuenta a la hora de trabajar con el diseño de investigación, nos ha hablado de los objetivos y de los criterios para seleccionar el diseño de la investigación. Pero desde un punto de vista más técnico, ¿cuáles serían los factores que condicionan la validez de los diseños? Profesora Martínez Mediano.

Bien, pues en este punto vamos a hablar fundamentalmente de la validez interna y externa de los diseños. El objetivo de la validez interna es asegurar que los valores obtenidos en la variable dependiente se deben a los efectos de la variable independiente o tratamiento. La variable

dependiente es el resultado final del experimento.

Entonces, la validez interna, su problema, lo que le preocupa es controlar si una tercera o terceras variables pueden ser la causa de tales efectos por introducirse entre la variable independiente y la dependiente. Así, por ejemplo, la validez de constructo contribuye a asegurar la validez interna. Por ejemplo, definiciones inadecuadas o incompletas de variables en los instrumentos de medida.

Otro aspecto que contribuye a la validez interna es la validez estadística y ésta depende de la potencia de las pruebas para el contraste de hipótesis de los supuestos del modelo estadístico con el que estemos trabajando, paramétrico o no paramétrico, de la calidad de las medidas, su fiabilidad y validez, de la aplicación de los tratamientos, del contexto, etc. También la validez interna se ve amenazada por factores tales como la historia, la maduración, si durante el tratamiento y la medición de los resultados ha pasado mucho tiempo y puede deberse los cambios no sólo al tratamiento sino a la maduración natural de los sujetos, la instrumentación, cómo están construidas las pruebas de medida, la selección y asignación de los sujetos y de los tratamientos, la mortalidad experimental, es decir, la reducción de los sujetos de un principio al final, etc. Y otro aspecto también relacionado con la validez de los diseños es la validez externa.

Por supuesto, la validez fundamental es la interna en los diseños. La validez externa ya hemos mencionado que tiene que ver con la posibilidad de generalizar los hallazgos a otros sujetos, a otras poblaciones, a otros contextos, a otras situaciones. Y ya en lo que se refiere a la resolución práctica de los diseños, ¿cuáles son los aspectos principales que tenemos que tener en cuenta, profesora Martínez Mediano? Bien, pues esto tiene que ver con los modelos estadísticos.

Son de dos tipos, los modelos paramétricos y los no paramétricos. Para utilizar las pruebas estadísticas paramétricas se han de cumplir unos supuestos de este modelo. Que las características que se estudian en la muestra existan en la población, que se distribuyan siguiendo el modelo de la curva normal, que el estadístico muestral sea una estimación del parámetro, de modo que estas exigencias del supuesto paramétrico deben controlarse adecuadamente.

Las pruebas paramétricas son más poderosas que las no paramétricas para rechazar hipótesis nula cuando deban ser rechazadas. Y, por otro lado, las pruebas estadísticas de los otros modelos, modelos no paramétricos, no parten de supuestos acerca de la población o son muy débiles, operan con datos de medida ordinal e incluso nominal. Y los dos supuestos asociados a la mayoría de estas pruebas son independencia de las observaciones y la continuidad básica de la variable.

Bien, pues una vez que tenemos claros esos aspectos generales que hemos de tener en cuenta a la hora de trabajar con diseños de investigación, vamos ya a desglosar o a tratar cada uno de los otros apartados que nos mencionaba al principio de esta conversación el profesor José Luis García Llamas. Comenzamos por los diseños de un solo grupo. Profesor García Llamas, ¿cuáles

son esos aspectos más destacables en relación con este tipo de diseños de un solo grupo? Bueno, debemos partir de la base teniendo en cuenta estos criterios que ha apuntado la compañera, que fundamentalmente estos diseños de un solo grupo se les denomina también diseños pre-experimentales o precientíficos, porque en realidad el grado de control sobre las variables extrañas o las variables independientes es un control muy bajo, muy pobre.

Por lo tanto, desde esa perspectiva hay que contemplar estos diseños. Entonces, ahí dentro del temario de la asignatura de pedagogía experimental del curso de adaptación, hay un tema que se dedica a presentar los diferentes modelos u opciones que hay dentro de los diseños de un grupo. Es decir, tenemos en primer lugar el que podemos llamar como un diseño base, que es aquel que utiliza, digamos, un grupo único solo con medida post-test, que es el propio de lo que llamamos la investigación es post-facto, es decir, recoger datos cuando los fenómenos que se analizan ya han acontecido.

También tenemos un diseño de grupo único pre-test post-test, que el mismo grupo sirve como control de sí mismo, es decir, se analiza antes del tratamiento y con posterioridad. Es también un diseño interesante. Y existe un tercer modelo que es distinto, que a los alumnos lo confunden con diseños de dos grupos, que es el que llamamos diseño de comparación con un grupo estático o grupo testigo.

En realidad, este grupo testigo aporta datos para la comparación, pero no forma parte de lo que es la propuesta de investigación. Por lo tanto, es un diseño de grupo único. En cuanto a la resolución práctica de estos diseños, se le dedican dos temas dentro del temario.

Según las opciones que han comentado anteriormente, puede haber una vía paramétrica cuando los datos que aporta el instrumento de recogida de datos tienen una buena calidad, que sería la opción paramétrica, o cuando no reúnen esos requisitos básicos, que sería la no paramétrica. En el primer caso o la primera opción, fundamentalmente nos interesa que el alumno domine fundamentalmente los contrastes referidos a la media del grupo. Es decir, lo que tratamos de comprobar en estos casos es si la media del grupo experimental difiere de la media de la población, y ahí tienen que tener en cuenta fundamentalmente dos pruebas.

La prueba T para muestras pequeñas, o sea, estamos hablando de muestras pequeñas cuando el número de sujetos del grupo experimental es menor o igual a 30, y la prueba Z cuando ya son muestras grandes, es decir, el contraste entre esa media y la de la población. También es interesante dentro de ese tema que el alumno conozca la significatividad estadística de los coeficientes. En la primera parte del curso estudiaron el análisis de diferentes coeficientes de correlación entre variables medidas dentro de un mismo grupo.

Aquí lo que se trata es de establecer si esa relación establecida de forma cuantitativa se puede considerar desde un punto de vista estadístico que tiene la suficiente significatividad. Es decir, si la relación entre las variables no es explicable por efecto del azar, sino porque hay una influencia directa entre una y otra variable, y ahí fundamentalmente nos interesa el coeficiente de correlación de Pearson. En cuanto a la otra opción, la opción no paramétrica, lo vamos a

centrar fundamentalmente en el estudio del estadístico o prueba de chi cuadrado que emplea datos de medida nominal frecuencias y cuya utilidad principal, al menos en este tema, se refiere a probar si un conjunto de datos cumple la condición de normalidad.

Es decir, si el modelo generado por los datos empíricos se asemeja al modelo de curva normal, de tal forma que este estadístico no es significativo, por lo tanto las discrepancias entre los modelos no son significativas y se cumple el requisito de normalidad. También se presenta la prueba de Kolmogorov-Smirnov, es otra prueba de discrepancia que, a diferencia de chi cuadrado, esta prueba lo que busca es el punto donde se produce la máxima discrepancia entre una distribución teórica de la población y una distribución empírica de la muestra, buscando ese punto, aunque hemos de reconocer que esta prueba tiene mucha más importancia, como veremos a continuación, en los diseños de dos grupos. También conviene que conozcan la prueba de McNemar, que es una prueba de diseño pretest-posttest o prueba de significación de cambios.

Es una prueba muy sencilla de aplicar y, mismo en línea con el tema anterior, pues serían también los coeficientes de correlación, la significatividad tanto del coeficiente de correlación de Spearman o Ordinal y el coeficiente de contingencia o coeficiente de datos nominales. Pasamos entonces a hablar de los diseños de dos grupos. Respecto a los diseños de dos grupos, ¿qué aspectos tenemos que tener en cuenta, profesor? Fundamentalmente yo creo que aquí entramos ya en lo que podríamos denominar diseños experimentales o cuasi-experimentales.

Es decir, ya hay un grado de control bastante amplio que viene garantizado. Desde tres puntos de vista se puede garantizar ese control. En primer lugar, mediante la selección aleatoria de los sujetos, es decir, el azar garantiza de alguna forma que existe un control sobre esos procesos.

También se puede alcanzar ese control mediante lo que llamamos el emparejamiento, es decir, formar parejas de sujetos que puntúan de forma similar en aquella variable que podemos controlar. Ahí tendremos la opción de los diseños de dos grupos relacionados y también se puede garantizar ese control mediante lo que llamamos grupo de control. Aquí sí que estamos hablando ya de dos grupos, por lo tanto hay un grupo que se somete a la influencia de la variable experimental o variable independiente y hay otro grupo que no recibe esa influencia, pero que parten de características similares.

Por lo tanto, al final de ese proceso, o lo que llamamos post-test o post-tratamiento, si se producen diferencias, probablemente esas diferencias sean achacables a la actuación que ha tenido el investigador sobre la variable independiente o, podemos llamar, variable experimental. Entonces, siguiendo con esta terminología, ¿cuáles serían las principales pruebas de resolución de los diseños de dos grupos independientes, profesor García Llamas? Bien, lógicamente, en base a esos modelos generales de los diseños de dos grupos, hay dos vías de resolución. En este caso, vamos a abordar lo que son dos grupos independientes.

Según las características de los datos, nosotros debemos diferenciar entre lo que

denominamos pruebas paramétricas, que son lógicamente más exigentes que las otras, y las pruebas no paramétricas. En el caso de las primeras, lo mismo que los diseños de un grupo, tendríamos que tener en cuenta el tamaño de la muestra para elegir entre t o z . Es decir, elegiremos t en aquellos casos en que al menos uno de los dos grupos sea menor o igual a 30, sería obligatorio. Si los dos grupos están por encima de 30, pero con valores próximos a 30, es más recomendable, no obligatorio, recomendable t frente a z . Y ya cuando las muestras son a partir de 50, 60, para arriba, pues mejor el estadístico t . Perdón, el estadístico z . También hay que tener en cuenta otros aspectos importantes.

Es decir, que la elección del estadístico va a depender básicamente del tamaño de la muestra, pero también hay que tener en cuenta esas otras condiciones. Por ejemplo, si conocemos las varianzas de la población, por lo tanto, no hay que estimarlas de los estadísticos. Si las desconocemos, pero se supone que son iguales entre sí y al parámetro, este caso vendría o tendría que venir expresamente enunciado en el problema o el caso más común, que serían aquellos casos en que se desconocen estas varianzas y deben ser estimadas a partir de los estadísticos, es decir, a partir de los datos de las muestras.

En cuanto a la otra opción o vía, la vía no paramétrica, presenta una mayor complejidad. Aquí fundamentalmente hay cuatro pruebas principales que vamos a reseñar brevemente, aunque sí que les recomendamos que, sobre todo, las dos últimas que vamos a citar son las más importantes. La primera se refiere a la prueba de chi cuadrado, que trabaja con datos de medida nominal, o sea, con frecuencias.

Lo que se trata es de analizar la discrepancia entre los valores empíricos de uno y otro grupo, de tal forma que buscar si esa máxima discrepancia es o no significativa. La segunda prueba es la prueba de la mediana, que trabaja con datos de medida ordinal burda, que es un paso intermedio entre lo que podemos llamar estrictamente nominal y la ordinal. Es decir, hay una dicotomización de las puntuaciones, se establecen dos categorías en base a la mediana como puntuación que ocupa el lugar central, y lo que se busca es si hay unas discrepancias significativas.

De hecho, esta prueba de la mediana utiliza como estadístico de contraste el estadístico chi cuadrado, naturalmente, como nos sale una tabla de dos grupos por dos categorías en base a la mediana, es una tabla dos por dos, por lo tanto, los grados de libertad que tiene es uno. Entonces, es una prueba que se utiliza chi cuadrado como estadístico de contraste y se interpreta como en el caso anterior. Estas dos pruebas, si quieren, son secundarias.

Es interesante desde el punto de vista teórico, pero desde el punto de vista práctico, el estudio se debe centrar fundamentalmente en la prueba U de Mann-Whitney y en la prueba de Kolmogorov-Smirnov a la que ya hemos hecho alusión. En cuanto a la prueba U , es una de las más potentes dentro de los diseños de dos grupos independientes. Yo creo que dentro del programa de la asignatura quizá sea la que encierra una mayor dificultad, no en cuanto a la prueba en sí, sino que hay que tener en cuenta las circunstancias en las que se aplica esta

prueba, de tal forma que los tres supuestos básicos que pueden confluir en esta prueba nos llevan por una vía de resolución distinta y por una interpretación distinta.

Por lo tanto, hay que fijarse muy bien si nos encontramos ante lo que denominamos muestras muy pequeñas, muestras pequeñas y muestras grandes. Es importante analizarlo porque desde esa perspectiva la resolución cambia y, por supuesto, la interpretación. En cuanto a la otra prueba, también importante, es la prueba de Kolmogorov-Smirnov.

Esta prueba se basa, como ya apuntamos anteriormente, en el punto donde se produce la máxima discrepancia en las distribuciones acumulativas de frecuencias. En el caso de muestras pequeñas serían las frecuencias absolutas y en el caso de muestras grandes serían las frecuencias relativas. Es decir, hay que analizar ese punto de máxima discrepancia e interpretarlo.

También tengan en cuenta que esta prueba tiene unos supuestos distintos según que el contraste al que hagamos alusión se refiera a un contraste unilateral, es decir, lo que llamamos un contraste de una cola, un contraste direccional donde el investigador postula que un grupo obtendrá mejores resultados que otro o el contraste bilateral donde se postula que habrá diferencias pero el investigador, al menos a priori, no puede aventurar cuál sería la tendencia de esas diferencias. Entonces, según uno y otro caso, se interpreta de forma diferente. Nos ha hablado de esas pruebas para el caso de la resolución en los diseños de dos grupos independientes. Tenemos que hablar también de cuáles son las pruebas para el caso de que los dos grupos estén relacionados.

En este caso partimos de la base de que nos encontramos dentro del modelo que llamamos modelo de grupos emparejados o relacionados, es decir, lo que hace el investigador en este caso es controlar alguna de las variables extrañas que piensa pueden contaminar los resultados. En base a las puntuaciones a esa variable extraña forma parejas de sujetos que son asignados a tratamientos distintos, de tal forma que garantiza un control inicial sobre esa variable extraña. En base a eso, los datos del posttest pueden tener lo mismo que antes, distinta categoría y la opción de resolución sería la vía paramétrica y la no paramétrica.

En el caso de la vía paramétrica tenemos o nos encontramos con las mismas pruebas, es decir, la prueba T y la prueba Z. Hay que tener en cuenta que estas pruebas, el criterio de elección, como estamos hablando de grupos emparejados, va a depender la elección de la prueba del número de parejas, es decir, que es obligatorio utilizar el estadístico T cuando el número de parejas es menor o igual a 30. En casos superiores, 40, 35, incluso hasta 50, es más recomendable T que Z y a partir de 50 o 60 para arriba es conveniente utilizar Z. Naturalmente aquí tendría que advertirles que además de estas pruebas que hemos comentado T o Z, lo mismo antes que ahora, también en las unidades didácticas presenta el ANAVA o análisis de varianza, pero no es recomendable utilizarlo en los diseños de dos grupos. Después cuando comentemos más de dos grupos veremos cómo tiene más utilidad.

Por lo tanto, aunque dentro del programa de la asignatura viene una referencia al ANAVA, no lo

estudian ahora, sino cuando llegue a esa tercera unidad. En cuanto a la opción no paramétrica, ahí tenemos dos pruebas básicas, digamos. La primera sería la prueba de los signos, que trabaja con datos de medida ordinal burda, es decir, es una medida muy pobre de los datos del postés, donde el investigador lo único que contempla es si en cada pareja que forma parte de esa muestra de análisis se han producido ganancias o pérdidas.

Es decir, si el componente de la pareja que ha ido al grupo experimental ha obtenido mejores resultados que el del grupo de control, hay un signo positivo, ha habido ganancias. Si ocurre la inversa, ha habido pérdida. Entonces es una prueba muy sencilla que se basa precisamente en las ganancias o pérdidas en las parejas.

Pero la prueba principal es la prueba T de Wilkosen, T mayúscula. Es decir, en este caso ya no sólo se tiene en cuenta la ganancia de una pérdida, sino la cuantía de esa ganancia o pérdida, que naturalmente después se transforma en rangos o posiciones, y se trabaja sobre esos rangos. Hay que tener en cuenta dos supuestos básicos, muestras pequeñas y grandes, que funcionan y se interpretan de forma distinta.

Hablamos ahora de esas pruebas, de las principales pruebas que tenemos que aplicar en el caso de los diseños de más de dos grupos independientes, en este caso profesor García Llamas.