

CE 02 1197

Seguimos en Revista de Educación, un programa de la Universidad Nacional de Educación a distancia en Radio 3 FM de Radio Nacional de España y en el punto 1359 de la Onda Media, y vamos a ocuparnos ahora de Pedagogía Experimental, otra asignatura del curso de Adaptación. Para hablarnos de Diseños Básicos en Investigación Educativa, nos acompañan los profesores de Pedagogía Experimental del curso de Adaptación, José Luis García Llamas y María Ángeles González Galán. Buenas noches.

Buenas noches. En realidad, la parte esta que corresponde a este cuatrimestre de la asignatura se centra toda ella en Diseños Básicos en Investigación Educativa, así que la pregunta que tenemos que hacernos es ¿qué aporta el diseño a la investigación pedagógica? Profesora González Galán. Pues el diseño de investigación es como un marco, un plan de acción general en el que se apuntan los pasos o las acciones que debemos de realizar desde que se formulan las hipótesis hasta que se llega a las conclusiones que confirman nuestras hipótesis de partida, y nos ilustran para llevarlas a la práctica educativa.

Por lo tanto, los objetivos de un diseño, como ya dice Kerlinger en su obra El proceso de la investigación, son fundamentalmente dos, responder a preguntas de investigación y controlar la varianza. Este segundo objetivo se consigue maximizando la varianza sistemática experimental, minimizando la varianza de error, es decir, aquella no sistemática producida por variables extrañas que no podemos controlar, y controlando la varianza sistemática no experimental, es decir, aquella varianza producida por unas variables extrañas que no nos interesa, que contaminen los resultados de nuestra investigación, y que de alguna forma podemos tener controladas para evitar esa contaminación. Si no controláramos esta varianza, estaría engrosada del término de error.

Esta forma de control de varianza se conoce con la abreviatura de sus siglas, que es Maxmincon. Una de las formas de control consiste en actuar sobre las variables extrañas introduciéndolas en el diseño, conformando con ellas una covariable, de esa forma estamos controlando. También mediante el bloqueo o emparejamiento de los sujetos.

También podemos controlar esta varianza manteniendo constantes sus valores o condiciones en todos los grupos experimentales. Y por último, la forma más usual es controlarla a través de la aleatorización. El azar siempre homogeneiza los grupos y nos permite este control.

Y las otras formas que he comentado de control se utilizan cuando el investigador considera que el azar no es suficiente, bien porque no ha podido aleatorizar totalmente las muestras, o bien porque le interesa un control más directo. ¿Cuáles son los factores que condicionan la validez de los diseños, profesora González-Galán? Pues vamos a ver, cuando el investigador elabora un diseño de investigación, busca que éste cumpla dos condiciones básicas, que son la validez interna del diseño y la validez externa. La validez interna se refiere a la configuración técnica del propio diseño y pretende confirmarnos que los cambios que observamos en la

variable dependiente, es decir, aquella en la que vemos los resultados de nuestra investigación, aquella que varía por el efecto de lo que nosotros manipulamos, se debe sólo a la independiente, es decir, a aquella variable que manipulamos.

Sin embargo, a pesar de que nosotros buscamos ese control de las variables extrañas para tener esta validez interna del diseño, hay factores que influyen en ella y que a veces no tenemos que tener en cuenta para ver si realmente el diseño cumple esas condiciones de validez interna necesarias para poderle considerar valioso, poder utilizar esos resultados. Estos factores son algunos como la historia, con la historia nos referimos a la propia historia de los sujetos que van variando a lo largo de la investigación, la maduración de los sujetos, es decir, los sujetos a lo largo de la investigación van creciendo, van adquiriendo experiencias que no tenían al principio, entonces esto puede afectar también. La administración de test o pruebas, a veces el sentirse investigados a través de un instrumento ocasiona unas reacciones imprevistas que pueden dificultar o que pueden afectar a esa validez interna.

La instrumentación, es decir, la utilización de instrumentos de investigación a los que no están habituados. La regresión estadística es un factor que las palabras a lo mejor son poco claras para los alumnos, pero quiere decir que siempre hay una tendencia de los sujetos a volver hacia la media. Igual que cuando hablamos, por ejemplo, de la altura, los hijos de padre y madre muy altos tienden a ser menos altos, vuelven hacia la media de altura de la población, pues en la administración de test ocurre igual, se suele regresar hacia la media los valores.

La selección de los sujetos, es decir, el muestreo, hay veces que dependiendo de cómo esté hecho el muestreo puede afectar también a la validez del diseño. La mortalidad experimental es otro factor a lo largo de una investigación que se alarga en el tiempo, sobre todo en los estudios longitudinales, que son unas investigaciones que se van prolongando en el tiempo y se utilizan varios años para recabar esos resultados finales, puede haber sujetos que vamos perdiendo en el camino, bien por muerte propia del sujeto, bien porque son sujetos que cambian de domicilio, de provincia, y los vamos perdiendo. Entonces, si esta es muy grande, esta mortalidad, puede invalidar el diseño.

Los efectos interactivos de las pruebas, es decir, sujetos que han realizado pruebas similares, test en otras circunstancias, pueden de alguna manera tener una experiencia sobre aquello que otros no tienen, y esto también puede ser una causa de invalidez. Es decir, estos factores afectarían a esa validez interna, es decir, a poder nosotros deducir que los cambios observados en la variable dependiente se deben sólo a la variable independiente. Si se deben en gran parte a alguno de estos factores, el diseño estaría invalidado en cuestiones de validez interna.

La validez externa, sin embargo, se refiere más a cuestiones de generalización, es decir, que los resultados de nuestra investigación podamos hacerlos extensivos a otras muestras en contextos similares, es decir, es la representatividad de nuestro diseño. Entonces, factores que afectan a la validez externa, es decir, a ese poder de generalización de nuestros resultados, son algunos que también afectan a la validez interna que comentábamos antes, como pueden ser

el efecto reactivo o interactivo de las pruebas, que ya he comentado, la interacción entre la selección de los sujetos y la variable experimental, es decir, puede haber muestras sesgadas que nos estén dando unos resultados que luego al hacerlos extensivos a otras poblaciones, no produzcan los mismos resultados o resultados similares. Tenemos también los efectos reactivos de los dispositivos experimentales, situaciones de reacción ante la prueba que experimentan algunos sujetos y que en algunos contextos pueden invalidar su generalización y la interdependencia de tratamientos múltiples.

Normalmente, aunque estos efectos pueden invalidar la generalización de un diseño, suelen ser más importantes los factores que afectan a la validez interna que a la validez externa, es decir, es más importante que controlemos estos efectos que comenté al principio, que nos confirman que los cambios en la variable dependiente se deben a nuestra variable independiente y no a otros factores, que estos otros que afectan a su generalización, aunque también hay que tenerlos en cuenta. De acuerdo, pues si les parece, ahora podemos hablar de lo que se refiere a la resolución práctica de los diseños y ver cuáles son los puntos principales que tenemos que tener en cuenta, profesor García Llamas. Bueno, una vez marcada lo que es la investigación, seleccionado el diseño, la fase siguiente sería la fase más operativa, es decir, nosotros tenemos ya que proceder a la recogida de datos bien mediante la selección de algún instrumento previamente elaborado o bien elaborando un instrumento propio.

Entonces, esta aplicación de los instrumentos a la muestra seleccionada nos va a dar unos datos, va a generar un conjunto de datos que es lo que llamamos el modelo estadístico. Naturalmente, en función de las características, es decir, en función de la calidad que tienen esos datos, nosotros podremos resolver mediante las pruebas oportunas ese modelo estadístico. Para ello tendremos que tener en cuenta si los requisitos que cumple ese modelo son los propios del modelo que llamamos paramétrico, como por ejemplo la independencia de las observaciones, el nivel de medida de intervalo, es decir, que la distancia entre puntuaciones sea comparable, la normalidad en la distribución, que siga el modelo de los datos generado sea el modelo similar al de la curva normal o bien que en algunos casos, como el caso del análisis de varianza, se dé el requisito de la linealidad, o sea, los efectos aditivos.

Si el modelo estadístico reúne estos requisitos quiere decir que se podrá resolver por la vía paramétrica. En caso contrario, es decir, cuando los datos no llegan al nivel de intervalo, sino que son datos nominales u ordinales, entonces tendríamos que resolverlo mediante la vía no paramétrica. En todo caso, es un aspecto importante dentro del diseño pues nos va a ayudar a elegir la prueba más oportuna.

Por lo tanto, lo que interesa al investigador desde esa perspectiva es conseguir una buena calidad de los datos para que la investigación también tenga una buena validez. Aspecto que ya contó antes la compañera. Entonces, desde esa perspectiva, es importante tener en cuenta todos estos criterios de ese modelo.

Además, hay otros que son secundarios, pero que también hay que tener en cuenta. Por

ejemplo, el investigador va a fijar a la hora de aceptar o rechazar las hipótesis qué margen de error lo va a hacer, el nivel de significación, o qué potencia tiene la prueba, o qué tamaño, pero, en general, lo fundamental a la hora de elegir esa prueba van a ser las características del modelo. Y tenemos distintos tipos de diseños, así que, si les parece, comenzamos por comentar, aunque sea de forma breve, las ideas básicas en torno a los diseños de un solo grupo.

Profesora González-Galán. El diseño de un solo grupo, de un grupo único, como viene comentado en las unidades, es la estrategia de investigación que utiliza un solo grupo y una sola variable independiente. Por lo tanto, carece de grupo de control, utiliza un solo grupo, y sin ningún sistema de control no es posible hacer afirmaciones científicas.

Por su propia estructura tampoco es posible el control estadístico. Por eso, a los diseños de un solo grupo se les llama diseños pre-experimentales o pre-científicos, pues los datos que se obtienen con ellos no permiten hacer afirmaciones inequívocas. Son muchos los factores que no controlan y que pueden actuar simultáneamente con la variable independiente, es decir, el problema de la validez interna que comentábamos antes aquí es totalmente incontrolado.

Dentro de los diseños de un solo grupo podemos considerar también varias modalidades. Hay diseños de grupo único con pretés solamente. Esto significa un grupo cada vez, después de someterlo a la acción de algún estímulo o tratamiento que se supone capaz de provocar un cambio.

Tiene una absoluta falta de control y por lo tanto su valor científico es prácticamente nulo. Sin embargo, han sido muy usados y por eso se estudian, porque solamente conociendo la debilidad de estos diseños se puede evitar esas debilidades. En ellos no se controlan para nada las variables extrañas y tienen una estructura muy débil.

También tenemos diseños de grupo único con pretés y postés. Seguimos utilizando un solo grupo pero tienen una prueba previa, después se aplica el tratamiento y posteriormente se vuelve a aplicar una prueba para ver qué cambios han sufrido después de introducido el tratamiento. También son unos diseños muy débiles, puesto que todos los factores en los que comentábamos antes de la validez interna están presentes y no son nada controlados.

No podemos afirmar que los cambios que hemos obtenido al encontrar cambios en esa segunda prueba se deban a la introducción del tratamiento, porque la historia no está controlada, la maduración de los sujetos tampoco, el efecto reactivo de la prueba tampoco, tampoco el instrumento, es decir, todos estos factores que debemos ir diciendo cómo los controlamos aquí no tienen ningún control. Y por último existen los diseños de comparación con un grupo estático, que aunque se consideran diseños de grupo único, puesto que al grupo estático, este que consideraríamos grupo de control, no le tocamos. En realidad es un diseño al que manipulamos es a uno solo, pero lo estamos comparando con otro.

Estos tienen un poquito más de valor, porque ya sí podemos comparar los cambios que ha

sufrido el grupo experimental con otro que no le hemos tocado ni le hemos variado nada, pero que se supone que sería la evolución propia de nuestro grupo si no hubiéramos introducido ese tratamiento experimental. De todas maneras, los controles que aporta este tipo de este grupo estático son también muy débiles. Entonces esto es un poco resumidamente lo que podemos decir de los diseños de un solo grupo que es conveniente conocerlos, pero que realmente no es conveniente utilizarlos en la investigación.

Al hilo de esto yo quería añadir que naturalmente después en la resolución práctica, lo que venimos comentando, también se podrá optar por la vía paramétrica o por la vía no paramétrica según esa calidad de los datos que comentamos. Por lo tanto, hay dos temas donde se analizan las pruebas paramétricas y otro tema donde se analizan las pruebas no paramétricas. Además, habría que añadir también la significación estadística de los coeficientes de correlación.

Es decir, lo que vamos a probar en estos temas es si la relación entre las variables es una relación significativa y no explicable sólo por efecto del azar. O sea, que habría que añadir a esos modelos que ha comentado María Ángeles pues todos estos aspectos relacionados con su resolución ya práctica y con datos. Pasamos a hablar de los diseños de dos grupos.

¿Qué aspectos hemos de tener en cuenta? Profesor García Llamas. Allí hay que tener en cuenta a diferencia de los anteriores en que el control ya es más riguroso. Es decir, ya trabajamos con dos grupos.

Uno de ellos va a ser el grupo experimental que va a ser aquel que vamos a someter a la influencia del tratamiento y otro va a ser el grupo de control. Es decir, son grupos que parten de unas características similares y la diferencia fundamental va a ser el nivel de tratamiento. Entonces, desde esa perspectiva pues podemos garantizar ese control mediante la selección aleatoria de los sujetos, mediante el emparejamiento caso de grupos relacionados o mediante lo que llamamos grupo de control que son las técnicas más rigurosas.

Naturalmente, hay también diferentes modelos. Por ejemplo, está el diseño post C solamente. Es más débil.

El diseño pretest-posttest que además de permitir contrastar las hipótesis nos permite otra serie de controles previos, como por ejemplo saber si antes del tratamiento los grupos ya diferían o no, para ver la influencia de los tratamientos, etc. Y también están los modelos de grupos relacionados o emparejados en este caso. Y podríamos hablar de las pruebas, de las principales pruebas de resolución en uno u otro caso, en el caso de los diseños de dos grupos independientes primero y después de los grupos relacionados, profesor García Llamas.

Sí, a grandes rasgos, ya hemos dicho que los criterios de selección de pruebas naturalmente va a ser el fundamental, va a ser el modelo estadístico, pero también tendremos que tener en cuenta cómo se han seleccionado los sujetos. Si la selección de los sujetos ha sido al azar, nos encontramos con los diseños que llamamos de dos grupos independientes. En ese primer

requisito, tendremos que ver qué modelo estadístico viene generado por los datos.

Si el modelo es el paramétrico, reuniendo los requisitos de normalidad, homogeneidad, etcétera, ahí tendremos tres posibilidades, tres posibles pruebas, aunque nosotros recomendamos dos. Las pruebas principales en dos grupos serían la prueba T, para dos grupos que se utilizaría obligatorio cuando la muestra es menor o igual a treinta. A partir de treinta y hasta cincuenta o sesenta, hay muchos autores que recomiendan utilizar T. Ya cuando son muestras grandes, a partir de cincuenta o sesenta para arriba, se utiliza Z. Hay una alternativa, que sería la prueba de la Nava, pero no es recomendable utilizarla en diseños de dos grupos, puesto que la resolución estadística es bastante más compleja, lleva mucho más aparato estadístico y no sería necesario hacerlo en dos grupos.

Esta prueba es más recomendable, ya lo veremos más adelante, para los diseños de más de dos grupos. Eso sería digamos la vía paramétrica. En la vía no paramétrica hay cuatro pruebas principales.

Hay muchas más, pero nosotros fundamentalmente nos vamos a centrar en cuatro. La primera sería la prueba de chi cuadrado, que es una prueba de discrepancia, donde se analiza esa discrepancia global que existe entre las distribuciones de dos grupos independientes, mediante las frecuencias. Por lo tanto, vamos a trabajar con datos de medida que llamamos nominal, frecuencias.

Después hay una segunda prueba, que es la prueba de la mediana, que sería un nivel de medida intermedio entre la nominal y la ordinal, que se llama medida ordinal Burda. Es decir, lo único que se hace es dicotomizar las puntuaciones por el punto central, que si los alumnos recuerdan, en estadística descriptiva, el punto central de una distribución es la mediana. Por lo tanto, lo que se hace es dicotomizar las puntuaciones y establecer unas tablas de contingencia de dos por dos que se resuelven mediante chi cuadrado.

Aunque las pruebas principales para estos diseños son la prueba U de Mann-Whitney y la prueba de Kolmogorov-Smirnov. Ahí hay que fijarse muy bien porque son unas pruebas que tienen unos comportamientos muy peculiares en función de los tamaños de las muestras. Por ejemplo, la prueba U tiene tres supuestos distintos, lo que llamaríamos muestras muy pequeñas, muestras pequeñas y muestras grandes.

Cada una de ellas se resuelve por distinta vía y además se interpreta de forma distinta. Por lo tanto, tienen que fijarse muy bien en esos casos peculiares. Lo mismo podríamos hablar de la prueba de Kolmogorov.

La prueba de Kolmogorov, pues ocurre lo mismo. También hay que fijarse muy bien si son pruebas bilaterales o unilaterales y el tamaño de las muestras. Es decir, que son pruebas que requieren un cierto esfuerzo y una aproximación a su resolución.

Estas pruebas que nos acaba de comentar se refieren a cuando hablamos de diseños de dos

grupos independientes. Nos quedaría el de dos grupos relacionados, profesor García-Llamas. Efectivamente, en los dos grupos relacionados ahí hay una variable de emparejamiento que se utiliza para formar parejas de sujetos.

Es decir, se igualan antes del tratamiento. Después se someten a tratamiento y se analizan las parejas. Es decir, la diferencia fundamental radica en que la unidad de análisis va a ser la pareja.

Lo que le importa al investigador es ver cómo se comportan aquellos sujetos que antes estaban igualados. Entonces, una vez teniendo claro este criterio general, la resolución también sería lo mismo. La vía paramétrica con las pruebas T y Z el criterio es lo mismo.

Cuando hay 30 parejas o menos de 30 es obligatorio utilizar el estadístico T. Cuando son más recomendable T en un primer momento pero a partir de 50 o 60 parejas ya sería Z. Como alternativa a la Nava menos recomendable aún. Y después, en las pruebas no paramétricas aunque hay dos, fundamentalmente nos interesa la prueba de Wilcoxon. Fíjense muy bien porque es la asignación de rangos en función de la cuantía de las diferencias.

Hay otra prueba que es la prueba de los signos que es una prueba más débil que hay que conocer desde el punto de vista teórico y conocer su funcionamiento y fijarse muy bien porque son pruebas que tienen un funcionamiento distinto. A diferencia de las paramétricas que casi siempre es lo mismo, el tratamiento aquí en las pruebas no paramétricas, hay que fijarse muy bien el tipo de contraste y el tamaño de la muestra porque en función del tamaño pueden ser muestras pequeñas o grandes cuya resolución es distinta y su interpretación también es distinta. Bien, y ocupándonos ahora de los diseños de más de dos grupos, profesor García Llamas, ¿cuáles serían las principales pruebas en el caso de más de dos grupos independientes? En el caso de más de dos grupos independientes, la prueba paramétrica principal sería la prueba de la Nava.

Aquí verán los alumnos que tienen en las unidades didácticas varios temas donde se desarrollan los diferentes pruebas paramétricas. Por ejemplo, la resolución de la Nava, vienen todas las fases de resolución y después tienen que fijarse también el contraste a posteriori. Es decir, en la Nava lo que nos va a permitir es comprobar si la diferencia entre tres o más de tres grupos es significativa.

Si en el conjunto hay diferencias significativas después tendremos que comprobar a posteriori entre qué grupo se generan esas diferencias. Para ello tenemos dos pruebas, la prueba T de Tukey que se emplearía cuando el número de sujetos de los grupos es igual y la prueba S o prueba de Sf cuando el número de sujetos de los grupos es distinto. Esto como opción paramétrica.

En las no paramétricas, pues tenemos muy parecidas a los diseños de los grupos. Está la prueba de Chi cuadrado que es una prueba de análisis de tablas de contingencia, datos nominales. La extensión de la prueba de la mediana, que es muy similar a los dos grupos, lo

que pasa es que en vez de tener dos filas, va a tener tres, cuatro o cinco filas, pero va a tener dos categorías también.

Y la más importante es la prueba de Kruskal-Wallis también llamada análisis de varianza por rangos porque, claro, al trabajar con datos ordinales es un análisis de varianza, pero no sobre las puntuaciones, sino sobre los rangos. Entonces ahí hay que tener en cuenta si son muestras grandes o son muestras pequeñas. ¿Y en el caso de más de dos grupos relacionados? Ahí en el caso de más de dos grupos relacionados, desde el punto de vista teórico, hay un tema de las unidades didácticas de experimental 2 que es el tema 28, que hay que saber teóricamente su planteamiento.

Desde el punto de vista de la resolución práctica lo que nos interesa es el ANAVA de bloques, con un sujeto por casilla, y además el modelo aditivo. Es decir, frente a la ANAVA simple, pues lo que añade es una fuente más de variación, que es la debida al bloqueo. Por supuesto también los contrastes a posteriori mediante Tukey, puesto que los grupos están igualados.

Eso como vía paramétrica. Y en las no paramétricas tenemos también dos pruebas. La prueba Q de Conran, que trabaja con datos de medida nominal, son datos dicotómicos generalmente, y la más importante es la prueba de Friedman, que es también la alternativa no paramétrica al ANAVA de bloques.

Es decir, el ANAVA por rangos o prueba de Friedman, también hay que fijarse según el número de grupos que tiene, según el número de sujetos, etcétera, para diferenciar entre muestras grandes y pequeñas. Esa es la diferencia fundamental que hay que contemplar. Bien, hemos hecho un repaso por todos los aspectos más importantes de esta segunda parte del curso, relacionados con los diseños de investigación.

Los alumnos tienen en la guía didáctica, en la guía del curso, toda la información precisa para desarrollar su estudio. Tienen sus unidades didácticas, tienen también una colección de audiocasetes centrada en estas cuestiones de los diseños experimentales, y quizás en el último minuto alguna recomendación para ese trabajo práctico que tenemos que elaborar en esta segunda parte del curso. Profesora.

Pues únicamente recordar a todos nuestros alumnos que el trabajo de la asignatura es obligatorio, es un requisito sin el cual no podemos ponerles la nota final de la asignatura, que la fecha de entrega mejor es a finales del mes de mayo, primeros de junio, antes de que realicen los exámenes, para que ya se quiten ese problema de la cabeza y puedan dedicarse a los exámenes con tranquilidad, pero que seguimos recibiendo trabajos hasta el 15 de junio, unos días después de finalizada la segunda semana de exámenes. De todas maneras, después de esta fecha, los trabajos que lleguen también los corregiremos, sólo que la nota final del alumno quizás no pueda ser incluida en actas en la convocatoria de junio, sino que ya pasaría a incluirse en la convocatoria de septiembre. Una pequeña puntualización es que en el trabajo no se les olvide nunca al final poner un enjuiciamiento crítico de la investigación que han hecho, no sólo la resolución estadística de un diseño, sino su valoración personal y las conclusiones en

las que han llegado.

Bien, Iván, a recordarles a nuestros alumnos que tienen mucho plazo todavía para poder solucionar cualquier duda y hacer cualquier consulta a sus profesores tutores en los centros asociados y a sus profesores de la sede central, para que en el momento de las pruebas no haya ningún problema. En cualquier caso, unos minutos dedicaremos en alguna revista antes del examen a recordar esas pautas para el ejercicio. Gracias, profesor José Luis García Llamas y profesora María Ángeles González Galán y buenas noches.

Esto ha sido todo por hoy en Revista de Educación. Recuerden que en este curso pueden consultar la programación radiofónica de la UNED en las páginas web de nuestra universidad, la dirección uned.es. Y además, desde el mes de enero pueden ustedes escuchar los programas de radio también a través de esas páginas web de la UNED, uno o dos días después de la emisión. Por supuesto que las copias en audio casete se siguen enviando a todos los centros asociados.

Hasta el próximo jueves. Buenas noches, gracias por su atención y ya saben, tiempo ahora en la UNED para el programa de enseñanza abierta.