

EVALUACIÓN DE DIFICULTADES EN EL ANÁLISIS DE VARIANZA ELEMENTAL POR ESTUDIANTES DE PSICOLOGÍA

Assessing psychology students' difficulties in elementary variance analysis

Vera, O.D.^a, Díaz, C.^b, Batanero, C.^c y López-Martín, M. M.^c

^aUniversidad Nacional de Quilmes, Argentina, ^bUniversidad de Huelva, ^cUniversidad de Granada.

osmar.vera@unq.edu.ar/overal17@gmail.com

Resumen

En este trabajo presentamos un estudio de evaluación de las dificultades de comprensión de algunas componentes del análisis de varianza. Utilizamos las respuestas de una muestra de 224 estudiantes de Psicología, después de finalizar un curso de análisis de datos a un cuestionario. Analizamos la selección de un modelo de análisis de varianza, la comprensión de los supuestos de aplicación y del modelo lineal asociado, los cálculos necesarios para completar una tabla de análisis de varianza y la interpretación del análisis. Los resultados obtenidos, proporcionan información en un área donde hay poca investigación previa.

Palabras clave: *Estudiantes de Psicología; análisis de varianza elemental; evaluación.*

Abstract

In this paper we present a study where we assess the difficulties in understanding some elements of variance analysis using a questionnaire administered to a sample of 224 Psychology students after taking a data analysis course. We analyse the selection of a variance analysis model, the understanding of assumptions and of the associated linear model, the computations involved and the interpretation of results. These results provide information in an area where little prior research is available.

Keywords: *Psychology students; elementary variance analysis; assessment*

INTRODUCCIÓN

La inferencia estadística juega un papel destacado en las ciencias humanas, incluyendo la Psicología, donde la investigación se basa en los datos obtenidos de muestras de las que se infieren conclusiones para la población de las que se han tomado. Sin embargo, el uso y la interpretación de la estadística en el campo de la Psicología no siempre es apropiada, como se muestra en diversos estudios incluidos en trabajos de survey (ej., Batanero, y Díaz, 2006; Harlow, Mulaik, y Steiger, 1997).

También se han encontrado dificultades de comprensión de la inferencia en estudiantes universitarios, tal como lo muestran numerosos estudios (ej., Castro-Sotos; Vanhoof; Van Den Nororgat y Onghena, 2007, Harradine, Batanero y Rossman, 2011; Krauss y Wassner, 2002 o Vallecillos, 1994). La mayoría de estas dificultades se relacionan con la comprensión del nivel de significación en un contraste de hipótesis, concepto que es esencial para determinar si la presencia de un factor es significativa en el análisis de varianza. Existen otros puntos necesarios para una correcta comprensión y aplicación del método; sin embargo en ese sentido la investigación previa disponible es escasa.

El doble objetivo de este estudio ha sido evaluar la comprensión, tanto de alguno de los conceptos básicos, como de los procedimientos que intervienen en el análisis de la varianza en estudiantes

Vera, O. D., Díaz, C., Batanero, C. y López Martín, M. M. (2016). Evaluación de dificultades en el análisis de varianza elemental por estudiantes de Psicología. En J. A. Macías, A. Jiménez, J. L. González, M. T. Sánchez, P. Hernández, C. Fernández, F. J. Ruiz, T. Fernández y A. Berciano (Eds.), *Investigación en Educación Matemática XX* (pp. 539-548). Málaga: SEIEM.

españoles de Psicología. El estudio se llevó a cabo después de completar un curso de inferencia que incluía ese tópico. Con el fin de alcanzar el objetivo definido anteriormente, se han analizado las respuestas a un pequeño cuestionario que incluyen los siguientes puntos: a) selección de un modelo particular, b) comprensión del modelo lineal asociado, c) supuestos necesarios para aplicar el método, d) cálculos involucrados y e) interpretación de los resultados. En lo que sigue describimos los antecedentes de esta investigación y el método empleado; se finaliza con la discusión de los resultados y algunas implicaciones para mejorar la enseñanza de los tópicos.

ANTECEDENTES

Entre los pocos estudios relacionados con la enseñanza y el aprendizaje del análisis de varianza, podemos citar el trabajo de Rubin y Rosebery (1990), quienes implementaron una enseñanza experimental con estudiantes universitarios, con objeto de estudiar las dificultades en la interpretación de algunas ideas básicas del diseño experimental, que tiene conceptos comunes con el análisis de varianza. Los resultados sugieren que los estudiantes no distinguen fácilmente la diferencia entre variables dependientes, independientes y extrañas; además de no comprender el rol de la aleatorización.

El concepto de interacción adquiere cierta importancia en el análisis de varianza multifactorial, puesto que la interpretación incorrecta de la interacción de un factor en los diferentes niveles de otro conduce a conclusiones erróneas. Rosnow y Rosenthal (1991) señalaron que la interacción es el concepto poco comprendido en el campo de la Psicología, además, en 2007, Green observa, con estudiantes universitarios, ciertas dificultades en la interpretación de dicho concepto. Por otro lado, son varios los autores que han analizado la interpretación que se realiza de la interacción en artículos publicados en prestigiosas revistas de investigación, encontrando un alto porcentaje de errores relativos a la interpretación. Por ejemplo, Zukerman, Hodgins, Zuckerman y Rosenthal (1993) establecen que aproximadamente un tercio de los artículos tienen errores en la interpretación realizada, mientras que Umesh, Peterson, McCann-Nelson y Vaidyanathan (1996) y Pardo, Garrido, Ruiz y San Martín (2007) señalan que el porcentaje asciende aproximadamente al 75%.

MÉTODO

Nos basamos en las respuestas a un cuestionario realizado por 224 estudiantes de segundo curso de la Licenciatura en Psicología de la Universidad de Huelva, España. Estos estudiantes completaron el cuestionario después de finalizar dos cursos de análisis de datos (un año lectivo para cada uno) que incluían estadística descriptiva y probabilidad (primer año) y estadística inferencial (intervalos de confianza, contraste de hipótesis y análisis de varianza) en el segundo año.

Tabla 15. Contenidos evaluados por ítem

| Contenidos de Análisis de Varianza | Ítem | | | | | | | | | |
|--|------|----|----|----|----|----|----|----|----|-----|
| | I1 | I2 | I3 | I4 | I5 | I6 | I7 | I8 | I9 | I10 |
| Selección de modelos en análisis de varianza | X | X | | | | | | | | |
| Reconocimiento de los supuestos del modelo | | | X | X | | | | | | |
| Cálculos usados en análisis de varianza | | | | | X | X | X | | | |
| Interpretación de resultados | | | | | | | | X | X | X |

El cuestionario consta de 10 ítems de opción múltiple (ver Apéndice). La confección de los ítems está basada tanto en el contenido de análisis de varianza enseñado como en el incluido en los libros de texto usados por los estudiantes. La selección de los ítems se realizó siguiendo los criterios de calidad habituales que sugieren Wang y Osterlind (2013), es decir nos aseguramos de la congruencia entre cada ítem y el objetivo del cuestionario mediante un análisis a priori y el juicio de expertos. Para ello se seleccionaron 10 profesores universitarios que imparten el tema e investigadores en educación estadística, todos ellos doctores y con tradición investigadora. Se realizó una prueba piloto de cada ítem del cuestionario con muestras de alrededor de 50 estudiantes

del mismo curso, el año anterior. De este modo nos aseguramos de obtener información objetiva referente a los ítems: dificultades, legibilidad y ajuste al tiempo previsto. A partir de la información obtenida de la prueba piloto, se obtuvieron valores suficientemente elevados del coeficiente de fiabilidad del cuestionario; Alfa de Crombach ($\alpha=0,79$) y del coeficiente de consistencia interna de los ítems del cuestionario, basado en el Análisis Factorial de los ítems, theta de Carmines ($\Theta=0,85$). La Tabla 1 recoge los tópicos evaluados en los diferentes ítems.

El cuestionario fue completado como parte de la evaluación final del curso. Con el fin de evitar las preguntas contestadas al azar, se penalizó las respuestas incorrectas (como es habitual en la asignatura). Por otro lado, las preguntas dejadas en blanco no puntúan (ni en negativo ni en positivo).

RESULTADOS Y DISCUSIÓN

A continuación presentamos y discutimos los resultados, comentando conjuntamente los ítems que evalúan el mismo contenido. Comenzamos presentando una síntesis global de los resultados en la Tabla 2, donde se marcan con negrita las respuestas correctas. Observamos una variedad de índices de dificultad (porcentaje de respuestas correctas); desde 26,8 % a 72,3%. En todo caso, de los 10 ítems propuestos se tiene que en 7 de ellos el porcentaje de respuestas correctas es superior al 50% lo que indica que no resultaron excesivamente difíciles.

Tabla 2. Porcentaje de respuestas por distractor en cada ítem ($n=224$)

| Respuesta | Í1 | Í2 | Í3 | Í4 | Í5 | Í6 | Í7 | Í8 | Í9 | Í10 |
|--------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| a | 23.7 | 4.9 | 11.2 | 5.8 | 57.1 | 3.6 | 3.1 | 33.0 | 4.9 | 13.8 |
| b | 7.6 | 9.8 | 4.0 | 12.5 | 7.1 | 4.0 | 72.3 | 6.3 | 9.4 | 26.8 |
| c | 50.9 | 63.8 | 43.3 | 69.2 | 21.9 | 72.3 | 3.1 | 12.5 | 74.6 | 16.5 |
| No respuesta | 17.9 | 21.4 | 41.5 | 12.5 | 13.8 | 20.1 | 21.4 | 48.2 | 11.2 | 42.9 |

Selección de modelos en análisis de varianza

Los dos primeros ítems evalúan la selección del modelo. Puesto que hay diferentes tipos de análisis de varianza, el primer paso en el análisis sería determinar, a partir de la lectura de un enunciado, cuál es el modelo apropiado. En el primer ítem evaluamos la comprensión de los estudiantes de la diferencia entre las situaciones donde puede utilizarse la prueba t de diferencia de medias y aquellas en que se debe usar el análisis de varianza. Puesto que en el enunciado del ítem una profesora aleatoriamente divide la población en tres grupos y aplica a cada uno diferentes técnicas que quiere comparar, se debe usar el análisis de varianza de un factor completamente aleatorizado (la respuesta correcta es la c). Aquellos que eligen uno de los otros dos distractores, confunden el modelo de análisis de varianza con la prueba t .

Algo más de la mitad de los estudiantes (50,9%) interpretan correctamente la situación que se describe y eligen el modelo más adecuado que resuelve el problema (ver Tabla 2). A pesar de que el 23,7% de los estudiantes fueron capaces de diferenciar en el contexto del problema las muestras independientes de las relacionadas, no eligieron el procedimiento adecuado, ya que la prueba t solo es válida para comparar dos muestras. Esta dificultad también ha sido encontrada por Rubin and Rosebery (1990). El 7,6% de los sujetos seleccionan incorrectamente la prueba t , y adicionalmente confunden independencia con muestras relacionadas. Finalmente, aproximadamente el 18% no responde la unidad de medida.

En el segundo ítem investigamos si el estudiante comprende la diferencia entre factor y nivel de un factor y si discrimina situaciones donde un investigador aplicaría un modelo de análisis de varianza de dos factores para analizar sus datos. La respuesta correcta es la c), en el estudio hay dos variables, cada una con dos o más niveles. Los estudiantes que eligen la opción b) confunden factor con nivel del factor ya que en el estudio se ha incluido una variable con dos factores. Por otro lado,

los que seleccionan el distractor a) no diferencian el modelo de análisis de varianza de dos factores y el análisis multivariante unifactorial, puesto que se muestran dos variables dependientes. Sin embargo, cabe señalar que, de los alumnos que respondieron el ítem (78.6%), lo resolvieron de forma correcta el 63,8%.

Comprensión de los supuestos para la aplicación del modelo de análisis de varianza

Dos ítems evalúan esta comprensión, ítem 3 e ítem 4. El ítem 3 se aplica con el objetivo de evaluar la comprensión de los estudiantes de los supuestos requeridos para usar modelos de análisis de varianza. La respuesta correcta es la c), que fue seleccionada por 43,3% de los estudiantes (ver Tabla 2) lo cual denota un nivel moderado-bajo de dificultad. El 11,2% elige la opción a), olvidando que para aplicar este modelo se requiere que las varianzas tomadas para cada nivel deben ser estadísticamente iguales (homocedasticidad). La violación de este supuesto conduciría a conclusiones erróneas en aquellos casos donde no se trabaje con un modelo de efectos fijos y muestras de igual tamaño (Montgomery, 2005). Por otro lado, el 4% de los estudiantes seleccionan la opción b), en cumplimiento del supuesto de normalidad. Hay que señalar el alto porcentaje de estudiantes que no contestaron a esta pregunta (41,5%). No hemos encontrado investigaciones que evalúen la comprensión por parte de los estudiantes de los supuestos en las pruebas de análisis de varianza.

El ítem 4 está enfocado en evaluar la comprensión de la descomposición de la variabilidad total en un modelo de análisis de varianza de dos factores en las variabilidades por factor principal e interacción, junto con la del error. Esta descomposición es fundamental en el modelo y contribuye a explicar por qué llamamos análisis de varianza a un método donde el objetivo principal es comparar medias. De la información recogida en la Tabla 2, se sigue que el 69,2% de los estudiantes respondió correctamente al ítem; el 5,8% no consideran la variabilidad de ambos factores principales y la interacción (respuesta a)); el 12,5% confunden la descomposición de la variabilidad total con aquella que se corresponde para el modelo de análisis de varianza de medidas repetidas (respuesta c)). Un 18,3% de la muestra (respuestas en ambos distractores sumados) no asociaron la interacción con el modelo completo de análisis de varianza de dos factores, error señalado por varios autores (Rosnow y Rosenthal, 1991; Pardo, Garrido, Ruiz y San Martín, 2007). Finalmente, un 12,5% no responde a esta pregunta.

Cálculo en análisis de varianza.

Una vez elegido un modelo de análisis de varianza es necesario realizar una serie de cálculos para estimar los diferentes componentes de la tabla de análisis de varianza (ANOVA) y realizar el contraste de hipótesis requerido para la toma de decisiones. Los ítems 5, 6 y 7 evalúan la comprensión en relación con dichos cálculos.

El ítem 5 evalúa la comprensión de la definición del estadístico F como un cociente de sumas de cuadrados. Los estudiantes han estudiado previamente la distribución F en otros tópicos, por ejemplo en relación con la prueba de igualdad de varianzas. La respuesta correcta es la a), que fue seleccionada por el 57,1%, por los que observamos que la dificultad del ítem ha sido solamente moderada. Estos estudiantes comprendieron que el estadístico F es usado para comparar la variabilidad entre grupos con la variabilidad aleatoria; en el caso en que ambas varianzas sean iguales o cercanas, la diferencia entre grupos podría deberse a la variabilidad aleatoria y no al efecto de algún factor.

La opción b) fue seleccionada solamente por el 7,1% de los estudiantes, en este caso, la varianza entre grupos fue confundida con la varianza entre sujetos; esta confusión no ayudaría para la comparación de la diferencia entre grupos; en este sentido a la no comprensión del método de cálculo se agregaría la de la lógica del análisis de varianza. Aquellos estudiantes que han elegido la opción c) (21,9%) piensan que en el análisis de varianza de medidas repetidas se debe examinar la

interacción entre las variables sujeto-grupo. No encontramos investigaciones previas para comparar nuestros resultados.

Para resolver los ítems 6 y 7 los estudiantes completan una tabla ANOVA y, a continuación se les pregunta en relación al cálculo de los valores que completaron en la tabla. Los resultados de ambos ítems sugieren una buena comprensión de los cálculos que involucran el procedimiento del análisis de varianza. Tampoco hemos encontrado investigaciones previas sobre este tipo de ítems.

La respuesta correcta al ítem 6 (relacionado con la obtención de la suma de cuadrados es la opción c), y fue elegida por el 72,3% de los estudiantes (Tabla 2), de donde resulta, tal como lo advertíamos, que la pregunta tuvo una fácil resolución. El 3,6% de los estudiantes eligió la opción a) que da el valor de F observado para ese factor, confundiéndolo con la suma de cuadrados. Otro 4% confunde el valor de sumatorio cuadrado para el factor B con el de la media cuadrática para el otro factor. Un 20,1% no responde la pregunta, no recordando el concepto de suma de cuadrados.

En el ítem 7 se requiere el valor de la media de cuadrados de un factor. También encontramos un porcentaje alto, 72,3% de aciertos (opción b)), indicando que una alta proporción de estudiantes comprendieron la lógica para el cálculo de la media de cuadrados. Aquellos que se decidieron por la opción a) (3,1%) muestran confusión entre la media de cuadrados del factor y la media de cuadrados de la interacción; mientras que los que seleccionaron la opción c) (3,1%) tienen dificultades en diferenciar entre ese valor y el F observado para el factor B. Por último señalar que la mayoría de los estudiantes que no respondieron (21,4%), dejaron en blanco el ítem 6, sugiriendo que habían olvidado el cálculo de la tabla de análisis de varianza.

Interpretación de resultados

Finalmente, los tres últimos ítems evalúan la comprensión de los resultados del análisis de la varianza y la interpretación de los mismos en relación al contexto y la pregunta de investigación.

En el ítem 8 nosotros evaluamos en que porcentaje, los estudiantes realizan una buena interpretación de los resultados de la tabla de análisis de varianza y la decisión que toman acerca de una hipótesis sobre los efectos de los factores involucrados en la situación. Con la finalidad de responder a la pregunta, los estudiantes deben tener completa la tabla ANOVA usada en las preguntas anteriores, que dará tres valores para F empíricos, todos estadísticamente significativos. Los estudiantes deben interpretar correctamente cuando un resultado es significativo; interpretación que resultó difícil en las investigaciones llevadas a cabo por Vallecillos (1994) y Lecoutre (1999) y Cañadas (2012).

Dado que el p -valor del factor A, está próximo a 0 y sería poco probable que el resultado no se debiera a la influencia de ese factor sobre la respuesta, entonces la respuesta correcta es la a). Esta unidad de medida resultó la más difícil para los estudiantes, ya que solo la pudo contestar correctamente el 33% de la muestra, y un alto porcentaje de no respuesta (48.2%). La selección de las opciones b) o c) pone de manifiesto la no correcta comprensión de la significación estadística tanto del factor B como de la interacción. A pesar de que esos distractores han sido seleccionados por pocos estudiantes, el ítem aún fue difícil de acuerdo con los resultados que se tienen de otras investigaciones donde se les pregunto a los estudiantes por la interpretación de los resultados de análisis estadísticos (Vallecillos, 1994). Consideramos que ello es debido a que alrededor del 20% no completó o no entendió bien la tabla de análisis de varianza, como vimos en los ítems 6 y 7. No obstante, el porcentaje de interpretaciones correctas sigue siendo bajo.

Mediante el ítem 9, se pretende evaluar la interpretación que pueden realizar los estudiantes a partir de la información recogida en las salidas que ofrece un software estadístico, en concreto, SPSS. Los estudiantes habían realizado prácticas de análisis de varianza con SPSS y habían anteriormente interpretado tablas parecidas. El porcentaje de respuestas correctas fue del 74,6%; lo que sugiere que el grupo posee una capacidad buena de interpretación de este tipo de salida del software. Un

4,9% elige como correcto el distractor (a), contraste de dos medias independientes, sin tener en cuenta que se quiere estudiar si influyen o no dos variables independientes sobre una tercera (“tipo de trabajo” y “flexibilidad en horario” sobre la “satisfacción laboral”), además de estudiar si interactúan. También un 9,4%, elige equivocadamente el modelo medidas repetidas, el cual no sería correcto por la independencia de las dos variables presentadas. También es bajo el porcentaje de no respuesta (11,2%). No hay antecedentes para compararlo con otras investigaciones.

Al igual que el ítem 9, el ítem 10 evalúa la interpretación de las salidas del software SPSS centrada en el rechazo de hipótesis nula. Se evidencia dificultad, ya que sólo hubo un 26,8% de estudiantes que responde correctamente (ver Tabla 2). Un 13,8% eligen, de forma incorrecta, el distractor (a), creyendo que el tipo de trabajo incide significativamente en la satisfacción laboral. Por otro lado, a pesar de que en la tabla de análisis de varianza del enunciado la significación es de 0,855, un 16,5% elige el distractor (c) creyendo que existe interacción entre los dos factores. Destacamos por un lado el alto porcentaje de preguntas en blanco, 42,9%, y por otro, la escasez de investigaciones empíricas de este tipo para que puedan ser comparados los resultados obtenidos. No obstante, en este ítem se debe interpretar un nivel de significación, concepto en el que los alumnos presentaron muchos errores en la investigación de Vallecillos (1994).

Al comparar el porcentaje de éxito en los tres ítems relacionados con la interpretación, vemos que el más sencillo fue el ítem 9, consistente en identificar el modelo de análisis a partir de la tabla de análisis de varianza; los otros dos ítems requieren interpretar el nivel de significación y además en el ítem 8 hay que construir la tabla.

CONCLUSIONES E IMPLICACIONES PARA LA ENSEÑANZA DEL ANALISIS DE VARIANZA

El aprendizaje mostrado en las respuestas de los estudiantes en esta investigación fue razonable, a pesar que nuestra muestra consiste exclusivamente en estudiantes de Psicología, quienes tienen poca experiencia en el uso de modelos y herramientas matemáticas. A pesar del alto grado de no respuesta en algunos ítems, se muestra una buena competencia en la mayor parte de los puntos evaluados. En nuestra opinión, los resultados satisfactorios se deben al hecho de que, antes de la evaluación, los estudiantes realizaron dos cursos de estadística (Análisis de Datos en Psicología I y II), y el cuestionario se pasó como parte de la evaluación de la asignatura. En el primer curso estudiaron las bases para la aplicación de pruebas de hipótesis y en el segundo tuvieron la oportunidad de aplicarlas utilizando métodos de comparación de dos muestras (sus medias, proporciones y varianzas) y pruebas para varias muestras (análisis de varianzas).

No obstante, se han detectado errores en las respuestas que persisten después de completar el ciclo de dos años de estudios. Por ejemplo, los resultados obtenidos en el ítem 8 corroboran los obtenidos por Vallecillos (1994) y Olivo (2008), quienes claramente han observado el nivel de dificultad entre la interpretación de un estudio estadístico, aunque el estudiante domine el cálculo necesario en la resolución de los ítems.

También nuestros resultados confirman lo expuesto por Vallecillos (1994) y Díaz, Batanero y Wilhelmi (2008), que indican que el p -valor y el nivel de significación son conceptos que no resultan suficientemente claros para los estudiantes, incluso para aquellos que sean capaces de tomar la decisión correcta sobre la hipótesis que se debe rechazar.

Las principales implicaciones que deducimos para la enseñanza del tema son las siguientes:

- Sería importante hacer hincapié en los supuestos de aplicación del análisis de varianzas y la consecuencia de la violación de los mismos. Igualmente se debe mejorar la capacidad de interpretar resultados estadísticos por los futuros licenciados. Dicha interpretación será necesaria para una correcta comprensión de los resultados de sus estudios y de los recogidos en investigaciones publicadas.

- Es necesario revisar la enseñanza de la inferencia estadística para reforzar la comprensión del concepto de p-valor, como se indica en Vera, Díaz y Batanero (2011) y también de otros objetos matemáticos necesarios para comprender el contraste de hipótesis y el análisis de varianza. Para ello sería importante comenzar introduciendo estos objetos informalmente desde la enseñanza media, siguiendo la tendencia actual denominada *inferencia informal* (ver Batanero, 2015; Gil y Ben-Zvi, 2011; Wild, Pfannkuch, Regan y Horton, 2011).

Nuestra investigación aporta alguna información en un tema que ha sido poco analizado, pero, no obstante, es un estudio exploratorio. Finalizamos animando a otros investigadores para continuar sobre esta línea de investigación con el fin de mejorar el aprendizaje de estos conceptos por parte de los estudiantes.

REFERENCIAS

- Batanero, C. (2015). Aproximación informal al contraste de hipótesis. En J.M. Contreras (Ed.), *II Jornadas Virtuales de Didáctica de la Estadística, la Probabilidad y la Combinatoria* (pp. 125-144). Granada: SEIEM.
- Batanero, C. y Díaz, C. (2006). Methodological and didactical controversies around statistical inference. *Actes du 36ièmes Journées de la Société Française de Statistique* [CD-ROM]. Paris: Société Française de Statistique.
- Cañadas, G. (2012). *Comprensión intuitiva y aprendizaje formal de las tablas de contingencia en alumnos de psicología*. Tesis Doctoral. Universidad de Granada.
- Castro Sotos, A. E., Vanhoof, S., Van Den Nororgate, W. y Onghena, P. (2007). Student's misconceptions of statistical inference: A review of the empirical evidence form research on statistical education. *Educational Research Review*, 2(2), 98-113.
- Green, K. E. (2007). Assessing understanding of the concept of interaction in analysis of variance. Trabajo presentado en la *IASE/ISI-Satellite Conference on Assessing Student Learning in Statistics*. Guimarães, Portugal: International Statistical Institute.
- Gil, E. y Ben-Zvi, D. (2011). Explanations and context in the emergence of students' informal inferential reasoning. *Mathematical Thinking and Learning*, 13, 87-108.
- Harlow, L. L., Mulaik, S. A. y Steiger, J. H. (1997). *What if there were no significance tests?* Mahwah, NJ: Lawrence Erlbaum.
- Harradine, A., Batanero, C. y Rossman, A. (2011). Students and teachers' knowledge of sampling and inference. En C. Batanero, G. Burrill y C. Readings (Eds.), *Teaching statistics in school mathematics. Challenges for teaching and teacher education* New York: Springer, 2011, pp. 235-246.
- Krauss, S. y Wassner, K. (2002). How significance tests should be presented to avoid the typical misinterpretations. En B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics*. Cape Town, South Africa: International Association for Statistics Education. Online: www.stat.auckland.ac.nz/~iase/publications.
- Lecoutre, B. (1999). Beyond the significance test controversy: Prime time for Bayes? *Bulletin of the International Statistical Institute: Proceedings of the Fifty-second Session of the International Statistical Institute* (Tomo 58, Libro 2) (pp. 205 – 208). Helsinki: International Statistical Institute.
- Montgomery, D. (2005). *Diseño y análisis de experimentos*. México: Limusa.
- Olivo, E. (2008). *Significado de los intervalos de confianza para los estudiantes de ingeniería en México*. Universidad de Granada.

- Pardo, A., Garrido, J., Ruiz, M.A. y San Martín, R. (2007). La interacción entre factores en el análisis de la varianza: error de interacción. *Psicothema* 19 (2), 343-349.
- Rosnow, R. L. y Rosenthal, R. (1991). If you're looking at the cell means, you're not looking at *only* the interaction (unless all main effects are zero). *Psychological Bulletin*, 110, 574-576.
- Rubin, A. y Rosebery, A. S. (1990). Teachers' misunderstandings in statistical reasoning; evidence from a field test of innovative materials. En A. Hawkins (Ed.) *Training teachers to teach Statistics* (pp, 72-89) Voorburg, The Netherlands: ISI.
- Umesh, U. N., Peterson, R. A., McCann-Nelson, M. y Vaidyanathan, R. (1996). Type IV error in marketing research: The investigation of ANOVA interactions. *Journal of the Academy of Marketing Science*, 24, 17-26.
- Vallecillos, A. (1994). *Estudio teórico experimental de errores y concepciones sobre el contraste de hipótesis en estudiantes universitarios*. Tesis doctoral. Universidad de Granada.
- Vera, O. D., Díaz, C. y Batanero, C. (2011). Dificultades en la formulación de hipótesis estadísticas por estudiantes de Psicología. *Unión*, 27. Recuperado de: www.fisem.org/web/union/.
- Wang, Z. y Osterlind, S. J. (2013). Classical test theory. En T. Teo (Ed.), *Handbook of quantitative methods for educational research* (pp. 31-44). Rotterdam: Sense Publishers.
- Wild, C. J., Pfannkuch, M., Regan, M., & Horton, N. J. (2011). Towards more accessible conceptions of statistical inference. *Journal of the Royal Statistical Society: Series A* 174(2), 247 – 295.
- Zuckerman, M., Hodgins, H. S., Zuckerman, A. y Rosenthal, R. (1993). Contemporary issues in the analysis of data: A survey of 551 psychologists. *Psychological Science*, 4,49-52.

APÉNDICE: CUESTIONARIO

Ítem 1. Para mejorar la psicomotricidad de los niños de primaria, una maestra cree que ayudarán unas nuevas actividades físicas. La maestra divide aleatoriamente su grupo de trabajo en tres partes iguales. A cada grupo le aplica un tipo de ejercicio diferente, pues desea saber qué tipo de ejercicios le dará mejores resultados. De las técnicas estadísticas que siguen, ¿cuál debería aplicar la maestra para comprobar si los métodos que aplica son diferentes?

- Contraste de hipótesis t sobre medias independientes.
- Contraste de hipótesis t sobre medias relacionadas.
- Análisis de varianza de un factor completamente aleatorizado.**

Ítem 2. Un investigador utilizará un análisis de varianza de dos factores, efectos fijos y completamente aleatorizado cuando:

- En el estudio haya dos variables dependientes.
- En el estudio haya una variable independiente, con dos niveles seleccionados al azar.
- En el estudio haya dos variables independientes, cada una con dos o más niveles.**

Ítem 3. Los supuestos de aplicación del análisis de varianza de dos factores, efectos fijos y completamente aleatorizado son:

- Independencia de las observaciones, normalidad de las distribuciones, y aditividad
- Independencia de las observaciones, igualdad de varianza y aditividad.

c. **Independencia de las observaciones, normalidad de las distribuciones e igualdad de varianzas.**

Ítem 4. El análisis de varianza de dos factores, con efectos fijos, descompone la variabilidad total en los siguientes componentes:

- a. Variabilidad total = V. entre grupos + V. error
- b. Variabilidad total = V. entre grupos + V. entre sujetos + V. error
- c. **Variabilidad total = V. factor A + V. factor B + V. interacción + V. error.**

Ítem 5. Si en un Análisis de varianza de un factor, y medidas repetidas encuentro que la F empírica u observada toma un valor de 8,16 esto quiere decir que:

- a. **CM entregrupos / CM error = 8,16**
- b. CM entresujetos / CM error = 8.16
- c. CM entregrupos / CM intrasujetos = 8,16

A partir de la información de la Tabla 1 se quiere estudiar el efecto de ciertas variables motivacionales sobre el rendimiento en tareas de logro. Se manipularon dos variables: “tipo de entrenamiento motivacional” (A1: instrumental; A2: atribucional y A3: control) y “clima de clase” (B1: cooperativo; B2: competitivo y B3: individual). Se seleccionaron a 45 sujetos y se dividieron en grupos para cada condición experimental.

Tabla 1

| Fuente de variación | SC | GL | CM | F |
|---------------------|-----|----|-------|------|
| Factor A | 70 | | | |
| Factor B | | | 20 | |
| Interacción AB | | | | 3,91 |
| Error | 46 | | 1,278 | |
| Total | 176 | 44 | | |

Ítem 6. El valor del sumatorio cuadrado para el factor B (Tabla 1) es:

- a. 15,65
- b. 35
- c. 40

Ítem 7. El valor de la media cuadrática para el factor A (ver Tabla 1) es:

- a. 5
- b. **35**
- c. 15,65

Ítem 8. Una de las conclusiones del estudio sería (alfa = 0,05) (ver Tabla 1)

- a. **Hay efecto del factor A (“entrenamiento”) sobre el rendimiento en tareas de logro**
- b. No hay efecto del factor B (“clima de clase”) sobre el rendimiento en tareas de logro
- c. No hay interacción de los factores.

Se ha decidido estudiar si hay efecto del “tipo de trabajo” (A1: trabajos de cualificación baja; A2: trabajos de cualificación media ó A3: trabajos de cualificación alta) sobre la “satisfacción laboral”.

También se ha tenido en cuenta, al mismo tiempo, si la “flexibilidad en el horario” (B1: si; B2: no) influye en la “satisfacción laboral”. Los resultados se muestran en la Tabla 2.

Tabla 2. Pruebas de los efectos inter-sujetos

| Fuente | Suma de cuadrados | GL | Media cuadrática | F | Significación |
|----------------------|-------------------|-----|------------------|---------|---------------|
| Modelo corregido | 465,614(a) | 5 | 93,123 | ,852 | ,516 |
| Intersección | 17192,763 | 1 | 17192,763 | 157,355 | ,000 |
| Tipo de trabajo | 49,120 | 2 | 24,560 | ,225 | ,799 |
| Flexibilidad horaria | 270,878 | 1 | 270,878 | 2,479 | ,008 |
| Tipo * Flexibilidad | 34,383 | 2 | 17,191 | ,157 | ,855 |
| Error | 12564,981 | 115 | 109,261 | | |
| Total | 33298,000 | 121 | | | |
| Total corregida | 13030,595 | 120 | | | |

Ítem 9. El tipo de análisis aplicado en este estudio (Tabla 2) es:

- Contraste de dos medias independientes.
- Análisis de varianza de un factor, con medidas repetidas.
- Análisis de varianza de dos factores completamente aleatorizado.**

Ítem 10. Entre las posibles conclusiones del estudio se encuentra, considerando un $\alpha=0,05$ (ver Tabla 2):

- Hay diferencias estadísticamente significativas en la “satisfacción laboral” en función del “tipo de trabajo”.
- Hay efecto del factor “flexibilidad horaria” sobre la “satisfacción laboral”.**
- Hay interacción entre la “flexibilidad horaria” y “tipo de trabajo”.