

Estrategias visuales para la optimización del aprendizaje de la prueba de significación estadística*

Vicente Manzano Arrondo**, Hassan Fazeli Khalili y Francisco Javier Pérez Santamaría

Universidad de Sevilla

Resumen: La prueba de significación estadística es una de las herramientas más importantes en la generación de conocimiento científico. Como consecuencia de ello, los tests de significación tienen un peso específico en el currículum de los estudios de psicología. No obstante, la lógica del proceso es muy difícil para los estudiantes. Luego es necesario encontrar estrategias alternativas que optimicen el aprendizaje acerca de los test de significación. En este sentido hay un relativo acuerdo en considerar que la animación de imágenes incrementa el aprendizaje de los conceptos estadísticos. En este trabajo, se prueban dos procedimientos concretos, basados en la animación de imágenes. Los resultados muestran algunas consecuencias de interés para explicar el test de significación estadística en el aula universitaria, desde una perspectiva exploratoria.

Palabras clave: Test de significación estadística, animación por ordenador, clases con vídeo, enseñanza en la universidad.

Title: Strategies to optimize learning about statistical significance test.

Abstract: Statistic significance test is one of more important tool in scientific knowledge generation. As a result of that, significance tests are a specific weight in curriculum of psychology studies. However, logic process is very difficult for students. So, is necessary to find alternative strategies to optimize learning about significance tests. In this respect there is a relative accord in considering that animation of images increment statistics concepts learning. In this work, two concrete procedures are tested, based in computer images animation. Results show some interesting elements for explaining statistic significance test in university classroom, from a exploratory perspective.

Key words: Statistic significance test, computer animation, teaching with video, teaching in university.

Acerca de la prueba de significación

La prueba de significación estadística consiste, básicamente, en tomar una decisión acerca de si se mantiene o no una hipótesis estadística, según la credibilidad que se otorgue a ésta, en función del grado de coherencia que las observaciones guarden con la hipótesis. El proceso implica (Pérez Santamaría, Manzano y Fazeli, 1998):

1. Enunciado de una hipótesis, según la cual una determinada función poblacional cuenta con un valor concreto. La discrepancia existente entre ese valor supuesto y las observaciones realizadas en la muestra, es admisible como fruto de las fluctuaciones propias del azar (pues la muestra es aleatoria). Esta negación de la discrepancia, puede justificar su denominación

como Hipótesis Nula (H_0), si bien en su origen la *nullidad* de la hipótesis se justifica en la intención de rechazar o *anular* su enunciado (Borges, 1997).

2. Cálculo de una medida numérica que exprese un grado de credibilidad para H_0 . En clase y con cálculos de papel y lápiz, esta medida suele ser una distancia estandarizada (Z_0). Con el ordenador, esta medida suele ser un valor de probabilidad: $p(O/H_0)$ (Manzano, 1997a; Pérez Santamaría y otros, 1998).
3. Decisión sobre si se rechaza o mantiene la hipótesis nula. Para ello, previamente se ha identificado una medida numérica que marca el umbral para tomar la decisión. Así, se escoge un valor de probabilidad (α) o una distancia estandarizada (Z_α o $Z_{\alpha/2} = Z_\alpha$). Con ello, se rechaza la hipótesis si lo observado se juzga poco probable considerando cierta H_0 [$p(O/H_0) < \alpha$ o $Z_0 > Z_\alpha$] y se mantiene en caso contrario.
4. Por último, el investigador concluye: traduce la decisión tomada en el paso anterior, en los mismos términos en que se ha enunciado la hipótesis estadística y, posteriormente, con el mismo lenguaje utilizado para enunciar la hipó-

* **Nota:** La realización del trabajo ha sido posible gracias a la subvención otorgada por el Instituto de Ciencias de la Educación de la Universidad de Sevilla.

** **Dirección para correspondencia:** Vicente Manzano Arrondo. Departamento de Psicología Experimental. Facultad de Psicología, Universidad de Sevilla. Avda. San Francisco Javier, s/n, 41005 Sevilla (España). E-mail: vmanzano@cica.es

tesis de investigación.

Este proceso se diferencia claramente del llamado *contraste de hipótesis*, donde aparece una alternativa al valor marcado en H_0 , H_1 con un valor o rango de valores concretos (Batista y Arnau, 1988).

Tanto el contraste de hipótesis como la prueba de significación, son artilugios estadísticos de gran trascendencia en la investigación psicológica en particular (Borges, 1997) como en las ciencias sociales en general (Cortina y Dunlap, 1997), por lo que debería ocupar una atención muy especial en la enseñanza de esta disciplina (Manzano, 1997a). Pero, aún considerando esta trascendencia, se trata de un cuerpo de conocimiento frecuentemente mal entendido, incluso en investigadores y personas que han recibido una formación estadística (Vallecillos, 1996). Consideremos, además, la resistencia que suele observarse en los investigadores no metodólogos a la hora de comprender y aplicar conceptos y técnicas estadísticas (Boulangier, 1971; Solanas y Sierra, 1992; Manzano, 1998).

A la vista de esta situación, sería deseable ensayar diferentes estrategias de docencia, para optimizar el aprendizaje del proceso de decisión estadística mediante la prueba de significación. En este trabajo, se presenta una estrategia concreta con animación de imágenes por ordenador, en comparación con la clase tradicional de pizarra.

La docencia universitaria mediante la animación con ordenador

Algunos conceptos estadísticos, especialmente aquéllos que requieren un elevado nivel de abstracción, facilitan la aparición de problemas de comprensión en los alumnos. La resolución de estos problemas ha pasado por la utilización de diversas estrategias. Una de ellas ha consistido en asociar la explicación de los conceptos estadísticos con imágenes gráficas, con el objetivo de facilitar la *visualización* de los conceptos, obteniendo mejoras en la comprensión (Embe y Engebretsen, 1996). Cuando estos conceptos implican elementos especialmente abstractos, los gráficos dinámicos permiten optimizar el aprendizaje

(Marasinghe, Meeker, Cook y Shin, 1996).

Quizá los conceptos de más difícil comprensión por parte de los alumnos son los relacionados con el manejo de probabilidades y los aspectos derivados de éstas como las distribuciones muestrales y la inferencia estadística (Vallecillos, 1996; Loosen, 1997), lo que favorece la generación de productos y estrategias de presentación visual (Goel, Peruggia y Baoshe, 1997; Siegel, 1997).

La dirección actual se centra en la creación de productos multimedia que están empezando a estar disponibles para los profesores o educadores en general, que imparten conocimientos sobre estadística o análisis de datos. Estos materiales pueden combinar textos, sonidos, imágenes en movimiento, representaciones gráficas... Aspectos éstos que hacen especialmente atractiva la iniciativa de incluir tales recursos en la explicación de conceptos estadísticos en la Web (Webster y Todd, 1998). Velleman y Moore (1996) no obstante describen, además de algunas ventajas derivadas del uso de estas tecnologías, algunos de sus inconvenientes y realizan una llamada de atención acerca de la importancia de la variable *factor humano* para comprender el efecto que la utilización multimedia puede tener en la enseñanza, aspecto ya señalado con respecto a cualquier investigación que implique calidad de la enseñanza (Fazeli et al., 1991).

El vertiginoso ritmo de investigaciones que se realizan durante estos últimos años con respecto a la utilización de nuevas tecnologías para la enseñanza de la estadística, favorece la aparición de tópicos, de pasiones desaforadas a favor o en contra de la implantación de tales recursos y del establecimiento de mitos que, no por ser parcialmente ciertos (Hawkins, 1996) deberían desaconsejar una actitud reflexiva y con abundante soporte empírico.

De todo ello, los resultados obtenidos en algunos de los estudios mencionados parecen aconsejar la utilización de estrategias de enseñanza de la estadística que recurran a la exposición de imágenes en movimiento, especialmente en los contenidos más cercanos a la inferencia. El objetivo de esta investigación fue poner a prueba dos herramientas que implican animación de

imágenes: el ordenador y el vídeo. El tema utilizado para la exposición fue la introducción a la teoría de la decisión estadística a través de la presentación de la prueba de significación de la hipótesis nula.

Método

Sujetos

Alumnos de la asignatura "Análisis de datos en psicología I" de la Facultad de Psicología en la Universidad de Sevilla, en el curso académico 1998/99, pertenecientes a cuatro grupos de prácticas. Sin considerar a quienes repetían la asignatura, se contabilizó un total de 118 alumnos. Dado que la asignación aleatoria de los sujetos a los grupos no es factible, se realizó un control previo de conocimientos en análisis de datos en la primera clase del curso.

Instrumentos

Para llevar a cabo la experiencia se utilizó el siguiente material:

1. Un programa de ordenador, elaborado en lenguaje PLACI (Manzano, 1997b) que contiene un suceso de imágenes animadas en torno a la lógica de la prueba de significación: enunciado de un problema (¿Han cambiado las actitudes xenófobas en la población?); búsqueda de una solución (elección del instrumento de medida de las actitudes xenófobas); aplicación de un esquema lógico (analogía con el sistema judicial: todo el mundo es inocente mientras no se demuestre lo contrario); recogida de datos (muestreo y distribución muestral); categorización de los posibles resultados en muy o poco probables; decisión en torno a la hipótesis nula; conclusiones; generalización a otros problemas.
2. Una película de vídeo en Super VHS que contiene la animación de imágenes anterior, a la vez que sonido de fondo y la voz de narración.
3. Un cañón de imágenes conectado a un PC con salida de vídeo.
4. Un cuestionario común para todos los alumnos en el que se pretende medir el grado de aprendizaje alcanzado durante la sesión. Se elabó

boró estructurado en cuatro bloques:

- Aspectos relacionados con el grupo de pertenencia y opiniones sobre la sesión.
- 3 preguntas de opciones V-F sobre la hipótesis nula y el proceso de decisión.
- 4 preguntas con cuatro opciones de respuesta sobre diferentes aspectos de la prueba de significación.
- Planteamiento de un problema de decisión donde el alumno debía enunciar la hipótesis nula, expresar el resultado obtenido como muy o poco probable según aquella, tomar la decisión de rechazarla o mantenerla y, por último, redactar brevemente la conclusión.

El anexo incluye un extracto de este cuestionario, constituido por los ítems de los bloques 21 y 31.

Procedimiento

De forma aleatoria, los cuatro grupos de prácticas fueron asignados a dos condiciones experimentales, resultando 58 alumnos en los grupos de ordenador (a quienes se impartió la clase mediante la animación por ordenador) y 60 alumnos en los grupos de vídeo (a quienes se impartió la clase mediante la proyección de la cinta de vídeo).

En todos los grupos, el profesor presentaba la clase con los objetivos de aprendizaje propuestos, apagaba las luces y procedía al desarrollo de la actividad.

En los grupos de vídeo, la labor del docente consistió en introducir la cinta en el lector de vídeo y controlar el volumen.

En los grupos de ordenador, el docente utilizó la animación de las imágenes por ordenador como un apoyo a su exposición magistral. El mismo discurso utilizado en estos grupos es al que se ha recurrido también como texto de narración para la situación de vídeo.

Dos profesores participaron en la experiencia. Cada uno de ellos contaba con un grupo asignado a la condición de vídeo y otro a la condición de animación por ordenador. El guión utilizado para la clase magistral de ordenador fue consensuado y ensayado por ambos docentes pa-

ra garantizar comportamientos equivalentes en clase. El aula utilizada fue la misma en los cuatro grupos de alumnos.

El contenido de la exposición en clase consistió en una introducción a la prueba de significación estadística, donde interesó entrar en la lógica del procedimiento, justificando el proceso en su conjunto y en sus fases. En esta primera aproximación no se trataron aspectos tales como: situaciones en las que la hipótesis nula tiene un valor "no nulo", justificación del nivel de significación o identidad del grado de significación. Son éstos aspectos de peso importante en el proceso de decisión (Borges, 1997; Manzano, 1997a) que no fueron suministrados a los alumnos hasta una sesión posterior.

Resultados

En la prueba inicial de conocimientos previos en Análisis de Datos se obtuvo una media de 4,37 puntos para la condición "video" y de 5,29 puntos para "animación por ordenador". La Figura 1 muestra el valor t asociado, de cuantía orientativa, bajo el supuesto de que se estuviera trabajando con muestras aleatorias.

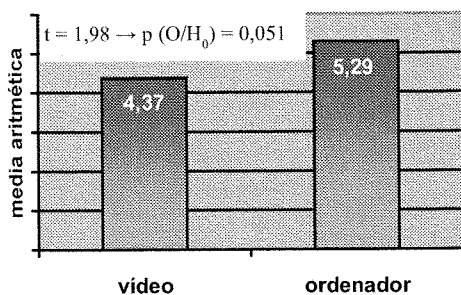


Figura 1: Comparación del rendimiento previo de los grupos de video y ordenador.

En la Tabla 1 se encuentran los porcentajes de acierto para cada condición, con respecto a los ítems de contenido (preguntas 4 a 10 del cuestionario de evaluación que figura en el anexo) y el valor del grado de significación asociado. Esta misma información, unificada en formato gráfico, se suministra en la Figura 2. Se

observa que a pesar de que ninguno de las pruebas de significación realizadas arroja una cuantía suficientemente pequeña² para $p(O/H_0)$, la tendencia es constante: En la Figura 2 resulta particularmente evidente cómo el rendimiento de los alumnos bajo la sesión de ordenador es ligeramente inferior al de los alumnos bajo la sesión de video, en todas las pruebas.

En ambos productos se ha incluido una variable generada a posteriori (*con*): existencia o no de error importante en las preguntas sobre el proceso lógico de la prueba de significación, correspondientes al cuarto bloque del cuestionario. Por *error importante* nos referimos a la existencia de incoherencia en la lógica del proceso, en alguno de los siguientes sentidos:

1. El alumno obtiene un resultado que aconseja rechazar H_0 , pero la mantiene.
2. El alumno obtiene un resultado que aconseja mantener H_0 , pero la rechaza.
3. El alumno rechaza la hipótesis nula, pero concluye en el mismo sentido que ésta.
4. El alumno mantiene la hipótesis nula, pero concluye en sentido contrario a ésta.

Por último, se generó una puntuación global para cada sujeto. Mediante el sistema de considerar los aciertos menos la ponderación de los errores, el bloque de las preguntas 4 a 6 puede suministrar una puntuación en $\{-3, +3\}$, mientras que el bloque de las preguntas 7 a 10 genera resultados localizables en el intervalo $\{-4/3, +4\}$, por lo que se generó la siguiente transformación para obtener finalmente puntuaciones acotadas en $(0, 10)$:

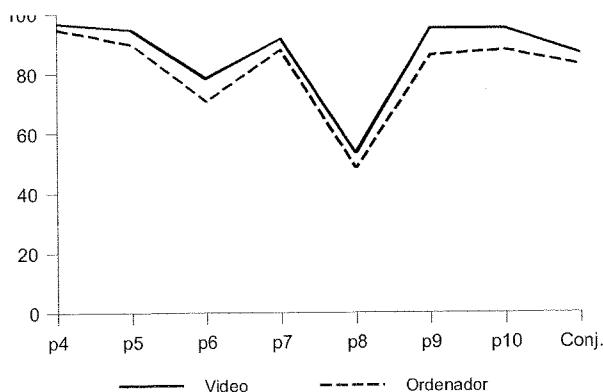
$$x_i = \frac{15}{17} \left(x_i + \frac{13}{3} \right)$$

donde x_i resulta de sumar las puntuaciones de los bloques 4-6 y 7-10.

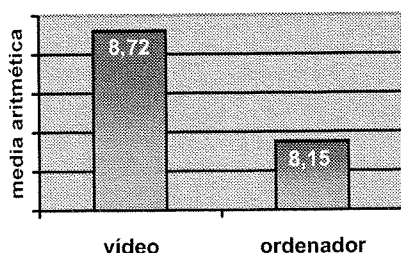
² Hemos utilizado el referente global $\alpha=0.05$. Para la inferencia simultánea de la tabla 1 el nivel de contraste debería ser $\alpha = 1 - (1 - 0.05)^{18} = 0.0064$ (por ejemplo, Fernández y Mayor, 1994). Esta circunstancia resalta aun más la distancia relativa de $p(O/H_0)$.

Tabla 1: Porcentajes de acierto por tratamiento.

	preg4	preg5	preg6	preg 7	preg 8	preg 9	preg10	Conj.
Vídeo	96,7	94,9	78,3	91,7	53,3	95,0	95,0	86,7
Ordenador	94,8	89,7	70,7	87,9	48,3	86,2	87,9	82,8
$p(O/H_0)$	,625	,291	,346	,507	,587	,105	,174	,560

**Figura 2:** Porcentajes de acierto por preguntas, según el tratamiento.

La Figura 3 muestra los resultados según el tratamiento, observándose coherentemente una tendencia no significativa en favor de los grupos bajo la condición de vídeo.

**Figura 3:** Comparación del rendimiento final de los grupos de vídeo y ordenador.

Conclusiones

Como hemos señalado, existe un relativo consenso en considerar beneficiosa la utilización de

imágenes en movimiento para la explicación de conceptos estadísticos de cierto nivel de abstracción. En nuestro planteamiento inicial, se ha pretendido establecer distinciones entre dos procedimientos basados en imágenes animadas, cuando se utilizan para explicar la prueba de significación estadística a un alumnado universitario.

Los resultados señalan tímidamente hacia un mejor rendimiento en la utilización de una cinta de vídeo frente a la animación mediante el ordenador en el aula. Esta tendencia no se encuentra apoyada en la significación estadística de los resultados sino en la monotonía o constancia de los pequeños efectos observados en todas las ocasiones de prueba y en el desequilibrio de partida en los grupos naturales. Obsérvese que, si bien el rendimiento previo fue ligeramente inferior en los grupos de vídeo (con una variable pre-test que suponemos relacionada con el rendimiento posterior), tras la aplicación del tratamiento la tendencia se invierte. Tal conclusión, no obstante, es débil, debido a:

- La mencionada ausencia de significación estadística.
- La circunstancia, importante, de que los grupos originales son naturales y que, si bien la asignación de los tratamientos a aquéllos tuvo lugar mediante una estrategia aleatoria, carecemos de seguridad suficiente como para garantizar la igualdad entre ellos.
- Tanto la cinta de vídeo como la situación de animación con el ordenador son *particulares*. No es posible deducir ni garantizar en absoluto que estos eventos son representativos de sus respectivos conjuntos posibles de origen ("todas las cintas de vídeo" y "todas las animaciones por ordenador").

A nuestro parecer, esta investigación con una clara naturaleza exploratoria, no permite establecer diferencias entre los dos procedimientos

puestos a prueba. Tal conclusión requeriría de una mayor contundencia en los resultados empíricos. Por otro lado, la variable de tratamiento es más un conglomerado que un elemento. Procede aislar los eventos más simples que la constituyen. Así, existen varias diferencias en ambos procedimientos:

- En la clase con la animación por ordenador, el papel del profesor es más activo, más sujeto a la espontaneidad.
- La cinta de vídeo incluye elementos tales como sonido de fondo y una voz de narración que no coincide necesariamente con la del profesor habitual.
- La clase realizada con el vídeo tiene mayor grado de "diferencia" con respecto a las sesiones habituales donde el docente realiza clases magistrales.

Referencias

- Batista, J.M. y Arnau, J. (1988). *La prueba de significación: un procedimiento revalorizado*. Barcelona: Universidad de Barcelona.
- Borges, A. (1997). Algunos problemas frecuentes en la interpretación de los contrastes de hipótesis estadísticas en psicología. *Iberpsicología*, 2, 3.8
- Boulangier, G. (1971). *La investigación en ciencias humanas*. Madrid: MAROVA.
- Cortina, J.M. y Dunlap, W.P. (1997) On the Logic Purpose of Significance Testing. *Psychological Methods*, 2, 161-172.
- Embse, C.V. y Engebretsen, A. (1996). Visual Representations of Mean and Standard Deviation. En Alan Bishop, *International Handbook of Mathematics Education*. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Fazeli, H., Granado, M.C., López Herrador, M., Manzano, V. y Sánchez García, E.F. (1991). Tratamiento del factor humano en las investigaciones sobre calidad de la enseñanza universitaria. En *Actas del primer congreso internacional sobre la calidad de la enseñanza universitaria*. Cádiz: Instituto de Ciencias de la Educación, págs. 87-97.
- Fernández García, F. y Mayor, J. (1994) *Muestreo en poblaciones finitas: curso básico*. Barcelona: PPU.
- Goel, P.K., Peruggia, M. y Baosche, A. (1997). Computer-aided Teaching of Probabilistic Modeling for Biological Phenomena. *The American Statistician*, 51, 2, 164-169.
- Hawkins, A. (1997). Myth-Conceptions! En J.B. Garfield y Gail Burrill *Research on the Role of Technology in Teaching and Learning Statistics*. Voorburg, The Netherlands: International Statistical Institute, págs. 1-14.
- Loosen, F. (1997). A Concrete Strategy for Teaching Hypothesis Testing. *The American Statistician*, 51, 2, 158-163.
- Manzano, V. (1997a). Usos y abusos de error tipo I. *Psicología*, 18, 2, 153-169.
- Manzano, V. (1997b). Aplicaciones del lenguaje PIACI en la docencia e investigación de la psicología. V Congreso de Metodología de las Ciencias Humanas y Sociales. Sevilla: Facultad de Psicología.
- Manzano, V. (1998). La calidad del muestreo en las investigaciones sociales. *REMA*, 3, 1, 16-29.
- Marasinghe, Meeker, Cook y Shin (1996) Using Graphics and Simulation to Teach Statistical Concepts. *The American Statistician*, 50, 4, 342-351.
- Pérez Santamaría, F.J., Manzano, V. y Fazeli, H. (1998). *Análisis de datos en Psicología*. Madrid: Pirámide.
- Solanas, A. y Sierra, V. (1992). Bootstrap: fundamentos e introducción a sus aplicaciones. *Anuario de Psicología*, 55, 143-154.
- Siegel, M.H. (1997). The TI-83 makes the CLT come to life! *The Statistics Teacher Network*, 45, 6-7.
- Vallecillos, A. (1996). *Inferencia estadística y enseñanza: un análisis didáctico del contraste de hipótesis estadísticas*. Granada: Comares.
- Velleman, P.F. y Moore, D.S. (1996) Multimedia for Teaching Statistics: Promises and Pitfalls. *The American Statistician*, 50, 3, 217-225.
- Webster, R. y Todd, R. (1998). Interactive Demonstrations for Statistics Education on the World Wide Web. *Journal of Statistics Education*, 6, 3, (<http://www.stat.ncsu.edu/info/jse/v6n3/west.html>)

(Artículo recibido: 30-8-1999, aceptado: 6-8-2001)

Anexo: Extracto del cuestionario de evaluación

Bloque 2:

- p4: Mantener la hipótesis nula implica mantener que hubo cambio.
- p5: En la decisión estadística el planteamiento de la hipótesis nula implica, entre otras, una negación inicial, por ejemplo, de cualquier variación en la población o de relación entre variables.
- p6: En un estudio para averiguar la diferencia del grado de motivación entre los hombres y las mujeres por los estudios universitarios, si se decide rechazar la hipótesis nula, la conclusión sería que ambos sexos tienen el mismo grado de motivación.

Bloque 3:

- p7: En la decisión estadística, si en la población es muy probable encontrar resultados muestrales como el que hemos obtenido, podemos
- a) Mantener la hipótesis nula
 - b) Considerar suficiente la diferencia encontrada
 - c) Rechazar la hipótesis nula
 - d) Considerar que la hipótesis nula es poco probable
- p8: En una distribución muestral normal los valores más probables son:
- a) Los más altos
 - b) Los más cercanos al estadístico
 - c) Los más cercanos al parámetro
 - d) Los más extremos
- p9: Según la hipótesis nula el índice en la población es 10 y se considera poco probable todo resultado muestral alejado en más de 5 puntos)Qué resultados muestrales justifican el rechazo de la hipótesis nula?
- a) Entre 5 y 15
 - b) Sólo menor que 5
 - c) Sólo mayor que 15
 - d) Menor que 5 y mayor que 15
- p10: Si en una investigación se concluye que no ha habido cambio en la población, ello implica que:
- a) Hay suficiente diferencia entre el resultado en la muestra y el resultado en la población
 - b) Se mantiene la hipótesis nula
 - c) Es poco creíble que la muestra pertenezca a esa población
 - d) El resultado obtenido en la muestra se encuentra en la zona de rechazo de la hipótesis nula