Psicothema
2000. Vol. 12, Supl. nº 2, pp. 396-398

ISSN 0214 - 9915 CODEN PSOTEG

Diseños de respuesta múltiple: aplicación de la teoría de la detección de señales al análisis de la inferencia visual de series de tiempo

Manuel Morales Ortiz, M. L. Domínguez y T. Jurado
Universidad de Sevilla

El objetivo de la investigación fue estudiar si la evaluación de la efectividad de un tratamiento mediante inspección visual dependía del tipo de representación gráfica. Se estudiaron 3 sujetos a los que se les presentó 3.000 gráficos (1.000 de líneas, 1.000 de barras y 1.000 de caja), representando cada uno de ellos los resultados de un diseño A-B. Los resultados indicaron que los sujetos cometieron mayor número de errores cuando los datos se presentaban mediante líneas o barras.

Visual inspection and type of graphs used. The objective of this research was to see if the evaluation of the effectiveness of a given treatment through visual inspection depends on the type of graphs used. Three subjects were studied with 3000 graphs (1.000 line, 1.000 bar and 1.000 box-plots), each one representing the results of an A-B design. The results indicate that the subjects committed more errors when the data was presented by lines or bars instead of box-plots.

Durante mucho tiempo las técnicas gráficas han sido la principal forma de evaluar la efectividad del tratamiento en el área de la modificación de conducta (Parsonson y Baer, 1986). Sin embargo, varios autores (Matyas y Greenwood, 1990), han puesto de manifiesto la inconsistencia existente entre las inferencias estadísticas y las derivadas de la inspección visual. Se ha encontrado que esta incosistencia depende de las características estadísticas existentes en los datos (Jones, Weinrott y Vaught, 1978; Knapp, 1983, etc.), y del entrenamiento de los sujetos (Wampold y Furlong, 1981). Sin embargo, pocos trabajos han evaluado si la inferencia visual depende del tipo de técnica gráfica.

Por tanto, el objetivo del presente trabajo es estudiar si la evaluación de la efectividad de un tratamiento mediante la inspección visual depende de la técnica gráfica utilizada. En concreto, nuestra hipótesis considera que los errores serán menores con el gráfico de caja que con el gráfico de líneas o el de barras.

Método

Sujetos

Realizaron el experimento 3 profesores que impartían la asignatura de Análisis de datos en el departamento de Psicología Experimental de la Universidad de Sevilla. Estos sujetos no habían participado en ningún estudio previo sobre percepción gráfica.

Correspondencia: Manuel Morales Ortiz
Facultad de Psicología
Universidad de Sevilla
41005 Sevilla (Spain)
E-mail: morales@psicoexp.us.es

Material

Se construyeron 3.000 gráficos de (1.000 de caja, 1.000 de líneas y 1.000 de barras), representando puntuaciones procedentes de un diseño de replicación intrasujeto A-B. La estructura de los datos fue siempre la misma:

$$Y = B + D + S + E$$

donde Y es el valor de la conducta en un determinado momento temporal; B es el nivel previo a la intervención (B= 40); D es el efecto de la intervención definido como la diferencia entre las medias (D = 5); S es la varianza existente en los datos (S_{lbo}= S_t = 64, y E es una variable aleatoria normalmente dstribuida, N(0,1). Los datos de cada gráfico se obtuvieron utilizando las rutinas para generación de variables aleatorias N(0,1) del paquete estadístico SPSS, v. 6. Estas variables fueron transformadas en otras con los valores de tendencia central y de dispersión que previamente se han indicado. Aproximadamente, la mitad de los gráficos reflejaban un efecto de la intervención pequeño, pero estadísticamente significativo (D = 5), y la otra mitad no (D = 0). Los gráficos de líneas y de barras se construyeron mediante el programa Harvard Graphics v. 3, y los de caja mediante el programa SPSS, v. 5. Un ordenador PC compatible presentó en blanco sobre fondo negro cada uno de los gráficos y registró las respuestas de los sujetos.

Procedimiento

Se utilizó un diseño factorial 3x2 intrasujeto (gráfico x tratamiento). Se presentaron tres tipos de gráficos (box-plot, barras y líneas) y dos tipos de tratamientos (efectivos y no efectivos). Cada sujeto se sentó frente al ordenador, indicándosele que el objetivo del estudio era evaluar la efectividad de cada tratamiento cuyos resultados aparecían representados gráficamente con distinto formato. Mientras veían los 6 gráficos de prueba de cada tipo se les ins-

truía sobre las teclas que tenían que utilizar para responder y se les explicaba que siempre aparecerían dos conjuntos de datos. Uno de ellos correspondiente a la fase de no intervención (línea base), y otro correspondiente a la intervención (tratamiento). A todos ellos se les indicó que trataran de responder lo más rápido que pudieran, pero sin perder seguridad en sus respuestas.

Los sujetos respondieron a todos los gráficos en varias sesiones. En cada sesión respondían a un total de 100 gráficos, descansando 5 minutos cuando habían evaluado 50 gráficos. En total, la prueba nunca duró más de una hora. En cada ensayo se presentaban durante 5 segundos un gráfico y a continuación aparecía una pantalla en la que se le pedía al sujeto que indicara su respuesta entre cuatro alternativas: 1= seguro no efectivo; 2 = probablemente no efectivo; 3 = probablemente efectivo, y 4 = seguro efectivo. Esta pantalla no desaparecía hasta que el sujeto no pulsaba la tecla correspondiente a una de las 4 opciones. En cada sesión sólo se presentó el mismo tipo de gráficos (barras, líneas o box-plots). El orden de presentación de los estímulos dentro de cada sesión fue aleatorio y el orden entre las sesiones fue contrabalanceado mediante cuadrado latino.

Resultados

Los resultados correspondientes a cada uno de los sujetos fueron analizados independientemente mediante modelos logit utilizando el programa GLIM, v.4 (Francis, B; Green, M. Payne, C., 1994). Para cada uno de ellos, se construyó una tabla de contigencia A x B x C donde A = tipo de gráfico (A1= box-plot, A2= barras y A3= líneas), B= tipo de intervención (B1= efectiva , B2= no efectiva), y C= tipo de respuesta (C1= seguro no efectivo, C2= probablemente no efectivo, C3= probablemente efectivo, y C4= seguro efectivo). Para la interpretación de los resultados se calculó la odds-ratio correspondiente al parámetro significativo. Así, en el caso de que fuera significativo el parámetro A(2)B(2)C(4), se calculó la razón entre la frecuencia de la respuestas tipo C4 y la frecuencia de respuestas C1 (C4/C1) tanto en el caso de los estímulos efectivos como en el de los no efectivos. A continuación, se compararon dichas razones siendo siempre el denominador la razón entre las respuestas comparadas correspondiente a los estímulos no efectivos. Siempre se compararon las respuestas C4, C3 y C2 con la C1.

Podemos considerar que la odds-ratio es un indicador de la ejecución global del sujeto en un determinado tipo de gráfico, ya que interesa que se detecten tanto los estímulos efectivos como los no efectivos. Mientras mayor sea la odds-ratio, mayor diferencia habrá entre la razón de las respuestas correspondientes a los estímulos efectivos y la de los no efectivos. Es decir, esto indica un mayor número de respuestas (C4, C3 o C2) frente a respuestas C1 en el caso de los estímulos efectivos y un menor número de respuestas (C4, C3 o C2) frente a respuestas C1 en el caso de los estímulos no efectivos. Por tanto, podemos considerar que mientras mayor sea el valor de la odds-ratio, menor será el número de errores cometido[1].

Sujeto 1

El análisis logit indicó que el modelo saturado era el único que se ajustaba. El modelo que sólamente incluyó asociaciones entre dos variables [AB, AC, BC] no presentó un ajuste adecuado (scaled deviance= 62.128; residual df= 6; $P(\chi^2, 6) \geq 62.168) < .05$). Se encontró que la odds-ratio entre las respuestas «seguro efectivo» y «seguro no efectivo» (C4/C1) fue mayor en el box-plot (69.8) que en el gráfico de líneas (5.9) y que en el de barras (6.5), (parámetros A(3)B(2)C(4), z= 6.47 y A(2)B(2)C(4), z= 6.21). En relación con los estímulos efectivos las razones C4/C1 fueron 21.64 para el gráfico de caja, 1.18 para el gráfico de líneas y 1.17 para el gráfico de barras. En los estímulos no efectivos no hubo grandes diferencias entre las razones (0.31 para el box-plot, 0.18 para el gráfico de barras y 0.2 para el gráfico de líneas). También resultaron ser significativos los parámetros A(3)B(2)C(3), z= 4.29 y A(2)B(2)C(3), z= 3.61). La odds-ratio entre las respuestas «probablemente efectivo» y «seguro no efectivo» (C3/C1) fue mayor en el gráfico de caja (14.64) que en los otros gráficos (3.2 para el gráfico de líneas y 4.09 para el gráfico de barras). Las razones C3/C1 en los estímulos efectivos fueron muy diferentes entre el gráfico de caja (8.93) y los otros tipos de gráficos presentados (1.47 para el el gráfico de líneas y 1.76 para el gráfico de barras). En relación con los estímulos no efectivos las razones fueron muy parecidas entre los tres tipos de gráficos (0.61 para el box-plot, 0.43 para el gráfico de barras y 0.46 para el gráfico de líneas) . Finalmente, la odds-ratio entre las categorías de «probablemente no efectivo» y «seguro no efectivo» (C2/C1) también resultó ser significativa (parámetro A(3)B(2)C(2), z= 4.56 y parámetro A(2)B(2)C(2), z= 4.01). En el box-plot fue 8.85, en el gráfico de líneas fue 1.67 y la del gráfico de barras fue 2.02. En relación con los estímulos efectivos, la razón C2/C1 fue 4.07 para el box-plot, 1.58 para el gráfico de barras y 1.42 para el gráfico de líneas. En relación con los estímulos no efectivos no hubo grandes diferencias, aunque algo mayores para el gráfico de barras y el de líneas (0.46 para el box-plot, 0.78 para el gráfico de barras y 0.85 para el gráfico de líneas).

Sujeto 2

El análisis logit indicó que el modelo saturado era el único que se ajustaba, ya que el modelo que sólamente incluyó asociaciones entre dos variables [AB, AC, BC] no presentó un ajuste adecuado (scaled deviance= 20.08; residual df= 6; $P(\chi^2, 6) \geq 20.08) < .01$). Se encontró que los parámetros A(3)B(2)C(4) y A(2)B(2)C(4) fueron significativos (z= 2.34 y z= 3.22 respectivamente). La odds-ratio fue mucho más grande en el box-plot (200) que en el gráfico de líneas (52.7) o que en el gráfico de barras (38.59). La razón entre «seguro efectivo» y «seguro no efectivo» (C4/C1) fue considerablemente mayor en el box-plot (22) que en los otros gráficos (3.69 para líneas y 11.19 para barras), cuando los estímulos representaron tratamientos efectivos. Sin embargo, en los tratamientos inefectivos las razones fueron 0.11 para los gráficos de caja, 0.29 para los de barras y 0.07 para los de líneas. Asimismo, el parámetro A(3)B(2)C(3) también resultó ser significativo (z= 2.71). La odds-ratio fue 33.5 para el gráfico de caja y 11.2 para el gráfico de líneas. La razón entre «probablemente efectivo» y «seguro no efectivo» (C3/C1) fue también mayor en el gráfico de caja (18.09) que en el gráfico de líneas (7.84) en los estímulos que reflejaron tratamientos efectivos. Finalmente, la razón entre las respuestas anteriores (C3/C1) fue mayor en el gráfico de líneas (0.7) que en el box-plot (0.54), cuando los tratamientos fueron inefectivos.

Sujeto 3

El modelo saturado fue el único que se ajustó, ya que el modelo que sólo incluía todas las asociaciones entre dos variables pre-

sentó grandes discrepancias entre las frecuencias observadas y las esperadas (scaled deviance = 19.694; residual df = 6; $P(\chi^2, 6) \geq 19.694) = .003$). Se encontró que el parámetro A(2)B(2)C(4) fue significativo (z = 3.12). La odds-ratio fue mayor en el box-plot (67.75) que en el gráfico de barras (21.14). Sin embargo, la proporción de respuestas entre «seguro efectivo» y «seguro no efectivo» (C4/C1) fue mayor en el gráfico de barras (10.78) que en el box-plot (2.71) en los estímulos con tratamientos efectivos. Asimismo, cuando se presentaron tratamientos inefectivos, la proporción C4/C1 fue también mayor para el gráfico de barras (0.51) que para el gráfico de caja (0.04). También se encontró significativo el parámetro A(2)B(2)C(2), (z = 2.02). La odds-ratio fue mayor para el gráfico de caja (5.8) que para el gráfico de barras (3). Asimismo, la proporción de respuestas entre «probablemente no efectivo» y «seguro no efectivo» (C2/C1) fue mayor en el gráfico de caja (1.74) que en el gráfico de barras (1.59) en los estímulos que presentaron tratamientos efectivos. Por último, la proporción entre dichas respuestas (C2/C1) fue mayor en el gráfico de barras (0.53) que en el gráfico de caja (0.30), cuando los estímulos presentaron tratamientos inefectivos.

Discusión

Los resultados de este estudio nos permiten establecer varias conclusiones. La primera es que la inferencia visual en diseños n = 1 depende del tipo de técnica gráfica utilizada. Considerando todos los errores (ante estímulos efectivos y no efectivos), los tres sujetos tuvieron mejor ejecución con el box-plot que con el resto de los gráficos.

Una segunda conclusión de nuestro estudio es que se confirma la hipótesis de que representar directamente la información relevante facilita el procesamiento de la información y, por tanto, reduce el número de errores. Esta hipótesis fue planteada originalmente por Cleveland & Mc Gill (1984) y posteriormente contrastada por distintos autores (Spence & Lewandowsky, 1990; Hollands & Spence, 1992; Morales, Jurado y Domínguez, 1999, etc.).

Finalmente, hemos de indicar que los resultados de nuestro estudio sólo indican la conveniencia de utilizar el gráfico de caja frente a otros tipos de gráficos cuando se pretende estimar el tamaño del efecto. Sin embargo, también es importante extraer información acerca de la tendencia existente en los datos cuando se trabaja con diseños n=1. Aunque no hay estudios que comparen el gráfico de caja con el de líneas, Hollands & Spence (1992) encontraron que el gráfico de líneas era mejor que el gráfico de barras para evaluar la tendencia existente en una serie de datos. Asimismo, también es importante conocer cómo afecta el grado de autocorrelación de las puntuaciones a la inferencia visual mediante el gráfico de caja, ya que existen trabajos (Jones, Weinrott & Vaught, 1978; Matyas & Greenwood, 1990), que han encontrado errores en la estimación visual del tamaño del efecto cuando se utilizan gráficos de líneas. Por tanto, es necesario realizar nuevos estudios para abordar estos problemas.

Agradecimientos

Este trabajo fue financiado por una beca PB93-1173 de la Dirección General de Investigación Científica y Técnica del Ministerio de Educación y Ciencia Español y por la Consejería de Educación y Ciencia de la Junta de Andalucía.

Notas

[1] Aquí consideramos como error responder «seguro no efectivo» frente a «probablemente efectivo» o frente a «probablemente no efectivo» cuando los estímulos son efectivos. Obviamente, también es un error responder «probablemente no efectivo» en vez de «seguro no efectivo» cuando los estímulos son no efectivos.

Referencias

Cleveland, W. S. & McGill, R. (1984). Graphical Perception: Theory, experimentation and application to the development of graphical methods. *Journal of the American Statistical Association, 79*, 531-554.

Francis, B.; Green, M. & Payne, C. (1994). *The GLIM system. Release 4 manual*. Oxford: Clarendon press.

Hollands, J. G. & Spence, I. (1998). Judging proportions with graphs: The sumation model. *Applied Cognitive Psychology, 12*, 173-190.

Jones, R.R.; Weinrott, M. & Vaught, R.S. (1978). Effects of serial dependency on the agrreement between visual and statistical inference. *Journal of Applied Behavior Analysis, 11*, 277-283.

Knapp, T.J. (1983). Behavioral analysts' visual appraisal of behavior change in graphic display. *Behavioral Assessment, 5*, 155-164.

Matyas, T. A. & Greenwood, K. M. (1990). Visual analysis of single-case time series: effects of variability, serial dependence, and magnitude of intervention effects. *Journal of Applied Behavior Analysis, 23*, 341-351.

Morales, M., Jurado, T. & Domínguez, M. L. (1999). Evaluación del sesgo en el análisis exploratorio de datos multivariados. *Comunicación presentada en el VI Congreso de Metodología de las Ciencias Sociales y de la Salud. Oviedo.*

Spence, I. & Lewandowsky, S. (1990). Graphical perception. En J. Fox and J. S. Long (Eds.). *Modern methods of data analysis*, (pp. 13-57). Newbury Park, CA: Sage.

Wampold, B. E. & Furlong, M. J. (1981). The heuristics of visual inference. *Behavioral Assessment, 3*, 79-92.