

Tiempos de reacción y psicología cognitiva: Dos procedimientos para evitar el sesgo debido al tamaño muestral

Manuel Perea*

Universitat de València

Los investigadores en psicología cognitiva suelen emplear cierto número de procedimientos de recorte de datos a la hora de efectuar los análisis del tiempo de reacción. Miller (1988, 1991) ha mostrado que dos de los procedimientos más habituales en los análisis del tiempo de reacción, la mediana y la media restringida (medias obtenidas tras la exclusión de aquellos datos que se hallen más allá de dos desviaciones típicas de la media del participante) son estimadores sesgados de la media poblacional en distribuciones que, como las del tiempo de reacción, muestran una asimetría positiva. Y lo que es más importante, dicho sesgo depende del tamaño muestral, lo que puede inducir sesgos indeseables incluso cuando la hipótesis nula es correcta. Este trabajo analiza dos procedimientos habituales en la literatura psicológica: medias semi-restringidas (medias obtenidas tras la sustitución a los puntos de corte de aquellos datos que se hallen más allá de dos desviaciones típicas de la media del participante) y la eliminación de datos más allá de unos puntos de corte fijos. El empleo de puntos de corte fijo no se encuentra afectado por el sesgo dependiente del tamaño muestral. Por su parte, las medias semi-restringidas sólo muestran un escaso sesgo dependiente del tamaño muestral. Finalmente, se examinan las implicaciones de los resultados para los análisis de los tiempos de reacción.

Palabras clave Simulación, distribución ex-Gaussiana, Tiempos de reacción.

El tiempo de reacción es, indudablemente, la variable dependiente más frecuentemente empleada en la experimentación en psicología cognitiva. Un fenómeno habitual en estos experimentos es la existencia de unas pocas

* Esta investigación fue subvencionada por una beca de la Dirección General de Investigación Científica y Técnica (PB/97-1379). Quisiera agradecer los comentarios de dos revisores anónimos. La correspondencia sobre este trabajo deberá ser enviada a Manuel Perea. Departament de Metodologia. Facultat de Psicologia. Av. Blasco Ibáñez, 21. 46010-València (Spain). (e-mail: mperea@uv.es)

puntuaciones extremas, usualmente causadas por una distracción de los participantes o por un fallo del instrumental. La inclusión de tales puntuaciones en los análisis de datos podría contaminar los resultados, dado que la media muestral se ve fácilmente influida por la existencia de puntuaciones extremas. Por ello, los investigadores suelen emplear técnicas de recorte de los datos antes de realizar el cálculo de los tiempos de reacción medios por condición, una medida que cabría contemplar como un tipo de "póliza de seguros" (Anscombe, 1960).

Naturalmente, las técnicas de recorte de datos no son exclusivas de la psicología y, de hecho, surgieron hace varios siglos en el contexto de la observación astronómica (para una revisión histórica, véase Barnett y Lewis, 1995). Desde entonces se han diseñado una serie de técnicas de recorte de datos cada vez más sofisticadas. Sin embargo, en psicología cognitiva existe el problema añadido de que las distribuciones de los tiempos de reacción no son simétricas, sino que suelen mostrar cierta asimetría positiva (véase Luce, 1986). En este ámbito, Miller (1988, 1991) ha mostrado que dos de las más habituales técnicas de recorte de datos empleadas en psicología cognitiva, las medianas muestrales y las medias restringidas (es decir, las medias obtenidas tras excluir los datos que caigan más allá de cierto número de desviaciones típicas de la media muestral, usualmente 2, 2'5 o 3 desviaciones típicas) dan lugar a estimaciones sesgadas de la media poblacional. Este problema no es, por sí mismo, particularmente importante ya que usualmente los experimentos están diseñados no tanto para saber cuál es el tiempo de reacción "real", sino para comparar los tiempos de reacción de las diferentes condiciones experimentales.

No obstante, Miller (1988, 1991) indicó un problema de mucha mayor importancia y que será el objeto de este trabajo. El sesgo que producen las medianas o las medias restringidas varía de acuerdo con el tamaño de la muestra, de manera que dos muestras extraídas de la misma población pero de diferente tamaño darán lugar a estimaciones distintas de la media poblacional. Dicho problema resulta especialmente importante cuando el tamaño muestral es pequeño (menos de 15 datos por condición), si bien cabe señalar que con tamaños muestrales mayores las diferencias son prácticamente inapreciables. En el caso de la mediana, las estimaciones se acercan a los valores de la media poblacional a medida en que aumenta el tamaño muestral y el sesgo sólo ocurre con los tamaños muestrales pequeños. En cambio, en el caso de las medias restringidas, las estimaciones obtenidas con tamaños muestrales muy pequeños son cercanas a la media poblacional (básicamente porque se elimina una menor proporción de datos) y el sesgo ocurre con tamaños muestrales mayores. La infraestimación de la media poblacional en las medias restringidas ocurre debido a que, por la asimetría positiva de la distribución de los tiempos de reacción, es más

probable que se excluyan datos en la parte derecha de la distribución que en la parte izquierda.

Es importante señalar que el problema del sesgo dependiente del tamaño muestral no afecta únicamente a los experimentos que emplean *a priori* diseños desequilibrados (v.g., estudios atencionales que manipulan la probabilidad de aparición de un estímulo o la probabilidad de aparición de una señal "válida"; por ejemplo, Briand y Klein, 1987; Remington y Pierce, 1984), sino que también puede influir sobre los experimentos que empleen en un diseño equilibrado (con el mismo número de participantes por condición): debido a que las puntuaciones erróneas —o las observaciones extremas al emplear medias restringidas— se eliminan de los análisis de los tiempos de reacción, ello podrá dar lugar a que el número de datos por condiciones experimentales sea diferente. No obstante, como indicó un revisor anónimo, el hecho de que haya un número diferente de valores perdidos por condición sería interesante por sí mismo.

Miller (1991) finaliza su trabajo indicando que se requieren nuevos estudios para determinar los mejores procedimientos de recorte, aunque no sugiere ninguna alternativa en particular —aparte de un procedimiento recursivo computacionalmente tedioso— sino más bien que la mediana y que las medias restringidas no son, en principio, los procedimientos más adecuados de recorte de datos cuando el número de datos por condición es escaso. Cabe señalar que el empleo de diferentes estadísticos "resistentes" a las puntuaciones extremas como la trimedia, las medias recortadas o el bipeso (*biweight*) también muestran cierto sesgo dependiente del tamaño muestral en muestras pequeñas (Perea, 1993a, 1993b). Ponsoda y Alcázar (1996) observaron que el empleo de una variedad de medias restringidas resistentes a las puntuaciones extremas (v.g., empleo de la mediana en lugar de la media, y empleo de las diferencias absolutas de la mediana, en lugar de la desviación típica, a la hora de realizar el procedimiento de exclusión de las puntuaciones extremas) reducía el sesgo dependiente del tamaño muestral, aunque no lo eliminaba. Por su parte, van Selst y Jolicoeur (1994) ofrecen diversos procedimientos recursivos de tratamiento de puntuaciones extremas basados en las medias restringidas, pero que siguen mostrando cierto sesgo debido al tamaño muestral en muestras pequeñas.

En este trabajo se analizan dos técnicas de recorte de datos que son muy habituales en la investigación en psicología cognitiva, una de las cuales —como se mostrará de forma analítica— no se halla influida por el sesgo dependiente del tamaño muestral: 1) Empleo de medias semi-restringidas, en la que los datos que se encuentren más allá de los valores de corte (dos o bien tres desviaciones típicas de la media por participante y condición) no son excluidos, sino que son reemplazados por los valores de corte (v.g., Forster y Veres, 1998); 2) Empleo de puntos de corte fijo, en el que aquellas puntuaciones que superen, por exceso o defecto, dos puntos de corte dados

por el experimentador (v.g., 200 ms y 1.200 ms; Perea y Gotor, 1997) son excluidas de los análisis de datos.

SIMULACIONES

La mayor parte de los modelos que analizan las distribuciones de los tiempos de reacción se basan en la distribución ex-Gaussiana, que es la convolución de la distribución Gaussiana (o normal, con los parámetros μ y σ) y la distribución exponencial (responsable de la asimetría, con parámetro único λ , que hace la función de media), y que se ajusta particularmente bien a las distribuciones de datos observados (v.g., Heathcote, 1996; Heathcote et al., 1991; Hockley, 1984; Ratcliff, 1979). En concreto, fueron consideradas doce situaciones experimentales a partir de variaciones de la distribución ex-Gaussiana, en las que se variaba la asimetría (es decir, el parámetro λ) y dispersión (variando el parámetro σ) de la distribución de datos de los tiempos de reacción. Tales situaciones son paralelas a las empleadas en trabajos anteriores sobre este tema (v.g., Hockley, 1984; Miller, 1988, 1991; Perea, 1993a, 1993b; Ponsoda y Alcázar, 1996; van Selst y Jolicoeur, 1994). En todos los casos la media poblacional ha sido de 600 ms. El programa de ordenador realizado al efecto creaba, a partir del generador de números pseudoaleatorios de Wichmann y Hill (1984), 10.000 muestras aleatorias independientes de tamaños 4, 5, 6, 8, 10, 15, 20, 25, 35, 50 y 100 de cada una de las doce distribuciones ex-Gaussianas consideradas.

Las medias semi-restringidas muestran cierto efecto debido a una interacción del tamaño muestral con el sesgo (ver figuras 1a y 1b). En el peor caso (distribuciones con una elevada asimetría), empleando un criterio de dos desviaciones típicas el sesgo diferencial es menor de 16 ms, que es mucho menor que el obtenido empleando medias restringidas en similares condiciones (51 ms; Miller, 1991). Dicho sesgo es, lógicamente, menor cuando el criterio es el de tres desviaciones típicas (menos de 7 ms, ver figura 1b). En ambos casos, la influencia de la dispersión es bastante pequeña, como ya indicó Miller (1988, 1991).

Por otra parte, cuando se emplean puntos de corte fijo, con los puntos de corte situados en 250 y 1.200 ms, el porcentaje de datos excluidos era de 1'97% (que aumentaba a 4'41% con unos puntos de corte de 250 y 1.000 ms). Pero lo más importante es que no hay ninguna interacción entre el tamaño muestral y el sesgo (véase figuras 2a y 2b). La falta de interacción tiene una fácil explicación analítica: La media de una distribución continua de probabilidad, $f(x)$, que ha sido recortada mediante la imposición de un límite inferior i y de un límite superior s es

$$E[x / (i < x < s)] = \frac{\int_i^s xf(x) dx}{\int_i^s f(x) dx}$$

Fig. 1.A

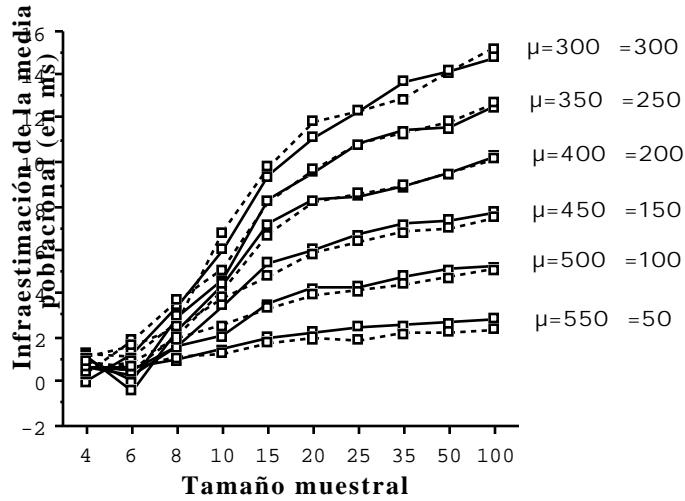


Fig. 1.B

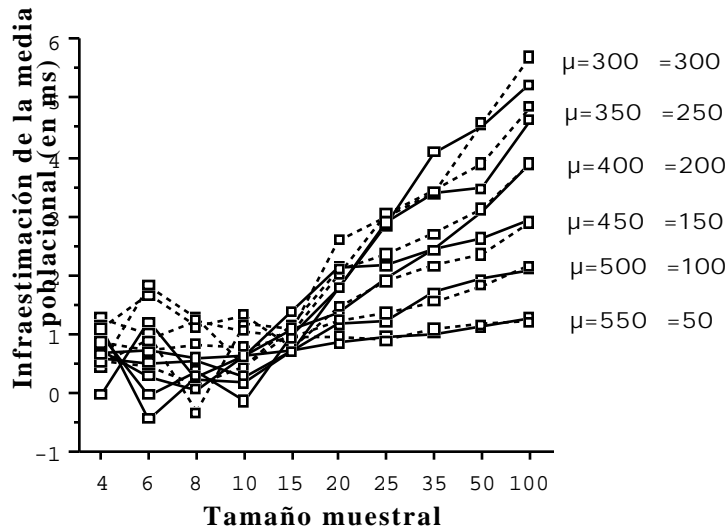


Figura 1. Infraestimación de la media poblacional con el empleo de medias semi-restringidas en la distribución ex-Gaussiana. Las distintas líneas representan los diferentes valores de la distribución Gaussiana (μ , λ) y la exponencial (λ). Las líneas continuas representan una $\lambda = 20$ ms y las líneas discontinuas una $\lambda = 50$ ms. El Panel A emplea un criterio de dos desviaciones típicas. El panel B es análogo al panel A pero empleando un criterio de tres desviaciones típicas.

Fig. 2.A

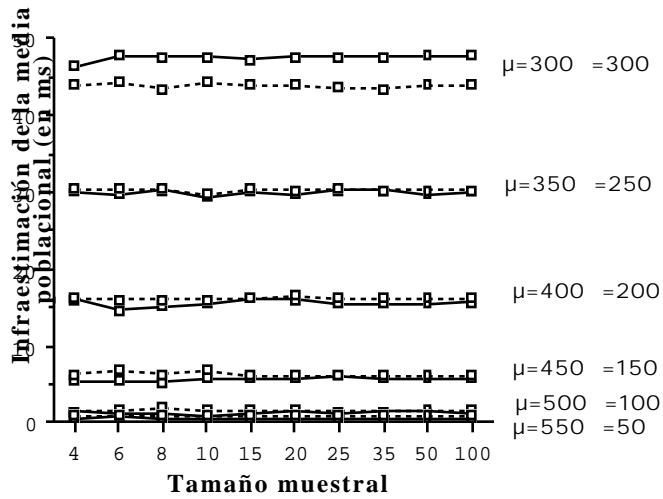


Fig. 2.B

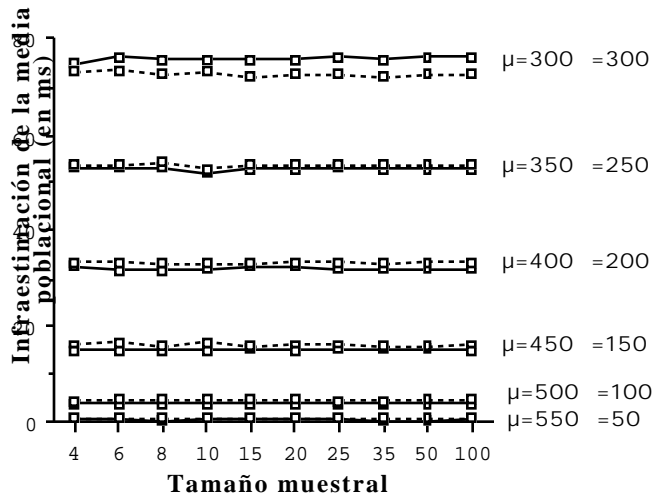


Figura 2. Infraestimación de la media poblacional con la eliminación de puntuaciones a partir de ciertos puntos de corte fijo en la distribución ex-Gaussiana. Las distintas líneas representan los diferentes valores de la distribución Gaussiana (μ , τ) y la exponencial (τ). Las líneas continuas representan una $\tau = 20$ ms y las líneas discontinuas una $\tau = 50$ ms. El Panel A emplea un criterio de 250 y 1.200 ms como puntos de corte. El panel B es análogo al panel A pero empleando un criterio de 250 y 1.000 ms como puntos de corte.

Evidentemente, esta fórmula para la media de la distribución recortada es independiente del tamaño muestral, con lo que el procedimiento de recorte de los datos a partir de dos puntos de corte fijos es independiente del tamaño muestral. Únicamente cabría añadir que, debido a las inherentes características de este procedimiento, hay cierto sesgo dependiente de la asimetría positiva de la distribución poblacional así como de los valores de los puntos de corte: cuanto mayor sea la asimetría, mayor será el sesgo; y cuanto mayor sea el punto de corte superior, menor será el sesgo.

CONCLUSIONES

El principal resultado de este estudio es que un procedimiento simple y habitual en los análisis del tiempo de reacción como es el empleo de puntos de corte fijo no se ve influido por el sesgo dependiente del tamaño muestral, como se ha mostrado analíticamente. Un posible problema del empleo de puntos de corte fijo es que la infraestimación de las medias poblacionales puede ser bastante grande, como se aprecia en los resultados. No obstante, hay que recordar que habitualmente lo que interesa es comparar los tiempos de reacción a través de diferentes condiciones más que en conocer cuál es el tiempo de reacción medio "real" en una determinada condición experimental.

Por su parte, las medias semi-restringidas sólo se ven afectadas por el sesgo dependiente del tamaño muestral en muy pequeña medida. Así, cabe señalar que las medias semi-restringidas se hallan menos influidas que las medias restringidas por el sesgo ocasionado por el tamaño muestral. El sesgo dependiente del tamaño muestral con un criterio de dos desviaciones típicas no supera los 16 ms en el caso de la distribución con un mayor grado de asimetría (recuérdese que tal valor era de 51 ms en el caso de las medias restringidas, véase Miller, 1991).

En definitiva, los procedimientos mostrados en el presente trabajo son, además de habituales y sencillos, una buena elección para evitar el sesgo causado por el tamaño muestral en distribuciones asimétricas —como las del tiempo de reacción— cuando se trabaja con un número relativamente escaso de datos por condición experimental. No obstante, se necesitan más trabajos sobre el tema que permitan una mejor comprensión de los diversos procedimientos de recorte de datos (Ulrich y Miller, 1994), así como su posible influencia sobre la potencia estadística. En este sentido, es importante recalcar que pese a que las técnicas de recorte de datos homogeneizan varianzas y podría facilitar el rechazo de la hipótesis nula, la probabilidad de error de tipo I se mantiene al nivel nominal (véase Ratcliff, 1993). Por otra parte, hay que recordar que ya Ratcliff (1978) señaló que "el tiempo de reacción medio por condición es de uso limitado como estadístico" (p. 99), abogando por un mayor interés en el estudio de procedimientos que permitan usar toda la información de las distribuciones de los tiempos de reacción

(véase Heathcote, 1996; Heathcote, Popiel y Mewhort, 1991). Un buen ejemplo de este tipo de aproximación al análisis de datos aparece en el reciente trabajo de Plourde y Besner (1997) al efectuar un análisis distribucional para determinar el *locus* de los efectos de frecuencia de uso.

ABSTRACT

Response times and cognitive psychology: Two procedures to avoid the bias due to sample size

A number of trimming procedures have been proposed to minimize the influence of outliers on the analyses of reaction times. Previous studies have shown that two of the most popular trimming procedures, medians and restricted means (i.e., means obtained after discarding all the original observations that fall more than a given number of standard deviations from the overall participants' average), are biased estimators of population means in the ex-Gaussian distribution (Miller, 1988, 1991). More important, this bias depends on sample size. This paper analyzes, by simulation and analytical procedures, two common trimming procedures to examine whether their estimations are independent of sample size: 1) semi-restricted means (similar to restricted means, except that outliers are set out to the cutoff values instead of being excluded); 2) truncated means, in which data falling beyond two fixed cutoff values given by the experimenter are excluded. Semi-restricted means show a slight bias depending on the sample size, which is quite smaller than when using restricted means. In addition, estimations provided by truncated means are independent of sample size, as demonstrated analytically.

Key words Simulation, ex-Gaussian distribution, Response times

REFERENCIAS

- Anscombe, F. J. (1960). Rejection of outliers. *Technometrics*, 2, 123-147.
- Barnett, V. y Lewis, T. (1995). *Outliers in statistical data*. New York: Wiley.
- Briand, K. A. y Klein, R. M. (1987). Is Posner's "beam" the same as Treisman's "glue"? On the relation between visual orienting and feature integration theory. *Journal of Experimental Psychology: Human Perception and Performance*, 13, 228-241.
- Forster, K. I. y Veres, C. (1998). The prime lexicality effect: Form-priming as a function of prime awareness, lexical status, and discrimination difficulty. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 24, 498-514.
- Heathcote, A. (1996). RTSYS: A DOS application for the analysis of reaction time data. *Behavior Research Methods, Instruments, and Computers*, 28, 427-445.
- Heathcote, A., Popiel, S. J. y Mewhort, D. J. K. (1991). Analysis of response time distributions: An example using the Stroop task. *Psychological Bulletin*, 109, 340-347.

- Hockley, W. E. (1984). Analysis of response time distributions in the study of cognitive processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 6, 598-615.
- Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. New York: Oxford University Press.
- Miller, J. (1988). A warning about median reaction time. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 539-543.
- Miller, J. (1991). Reaction time analysis with outlier exclusion: Bias varies with sample size. *Quarterly Journal of Experimental Psychology*, 43A, 907-912.
- Perea, M. (1993a). Sobre el sesgo de los estadísticos resistentes en los análisis del tiempo de reacción. *Psicológica*, 14, 33-42.
- Perea, M. (1993b). Sobre el empleo de M-estimadores en los análisis de los tiempos de reacción: Una simulación. *Psicológica*, 14, 209-218.
- Perea, M. y Gotor, A. (1997). Associative and semantic priming effects occur at very short SOAs in lexical decision and naming. *Cognition*, 67, 223-240.
- Ponsoda, V. y Alcázar, M. A. (1996). Reaction time analysis with outlier exclusion: A simple method providing bias nearly independent of sample size. *Psicológica*, 17, 31-40.
- Plourde, C. E. y Besner, D. (1997). On the locus of the word frequency in visual word recognition. *Canadian Journal of Experimental Psychology*, 51, 181-194.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85, 59-108.
- Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin*, 86, 446-461.
- Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, 114, 510-532.
- Remington, R. y Pierce, L. (1984). Moving attention: Evidence for time-invariant shifts of visual selective attention. *Perception and Psychophysics*, 35, 393-399.
- Ulrich, R. y Miller, J. (1994). Effects of outlier exclusion on reaction time analysis. *Journal of Experimental Psychology: General*, 123, 34-80.
- van Selst M. y Jolicoeur, P. (1994). A solution to the effect of sample size on outlier elimination. *Quarterly Journal of Experimental Psychology*, 47A, 631-650.
- Wichmann, B. A. y Hill, J. D. (1984). An efficient and portable pseudo random number generator. *Applied Statistics*, 33, 123.

(Revisión aceptada: 23/3/99)