Psicothema 2008. Vol. 20, nº 4, pp. 857-862
www.psicothema.com

ISSN 0214 - 9915 CODEN PSOTEG

Técnicas basadas en la mediana como alternativa a las pruebas clásicas de decisión

Maribel Peró Cebollero, Joan Guàrdia Olmos, Montserrat Freixa Blanxart y Jaume Turbany
Universidad de Barcelona

Habitualmente, en investigación, el constraste de hipótesis se realiza a partir de la distancia entre la medida de todo sujeto muestreado con respecto a la media de la distribución. Esta concepción de la distancia es usada y reutilizada constantemente como fuente de evidencia para la búsqueda de contrastes significativos con relación a modelos de probabilidad teóricos. Esta trabajo pretende aportar algunas concepciones a propósito del uso de nuevos indicadores de distancia con relación a estadísticos menos vinculados a supuestos paramétricos, en concreto el uso de los intervalos de confianza en torno a la mediana. Para ello se han generado vía simulación a partir del EXCEL matrices de datos para un diseño de dos grupos independientes bajo 15 condiciones diferentes en dos situaciones distintas: igualdad de medias y diferencia de medias. Los resultados muestran la elevada especificidad de la comparación de los intervalos de confianza, especialmente si se utiliza un criterio de decisión estricto, pero una sensiblidad no tan adecuada, de hecho en algunos casos funciona mejor la comparación de los intervalos de confianza de medias que los de medianas.

Median decision techniques as an alternative to the classical statistical inference test. In the research area, the criterion to test a hypothesis is frequently the distance of each subject from the mean of the distribution. This concept of distance is constantly used and re-used as a source of evidence in the search for significant contrasts with respect to models of theoretical probability. This paper proposes a number of ideas regarding the use of new distance indicators related to statistics that rely less heavily on parametric assumptions, in particular, the use of median confidence intervals. We have simulated via EXCEL different data samples for a design of two independent groups under 15 different conditions in two different situations: mean equality and mean differences. The results showed the high sensitivity of the comparison of the confidence intervals, especially if a strict decision criterion is used. However, we did not obtain good results for sensitivity. In fact, in some cases, the comparison of the confidence intervals of the mean worked better than those of median.

La tradición del contraste estadístico y sus características han conducido durante mucho tiempo a su consideración como prueba irrefutable de hallazgo científico, de rigor y solvencia y de relevancia en la aportación asociada al rechazo de la hipótesis nula. Esta especie de servidumbre al criterio paramétrico ha generado una cierta «dependencia de la hipótesis alternativa» que lleva a muchos investigadores a la creencia de que si no hay significación paramétrica no hay aportación científica relevante. La realidad señala que esta creencia es incorrecta y que la información contenida en los datos no siempre es suficientemente contundente como para permitir el rechazo de la hipótesis nula, que en la investigación aplicada las alteraciones, anomalías y dificultades que los datos comportan hacen que, en la mayoría de los casos, las premisas paramétricas no sean aplicables. Sin embargo, nadie pone en duda la bondad de utilizar los intervalos de confianza, en cualquiera de sus formatos y generados bajo esos supuestos, para el estudio del contraste estadístico y conjuntamente con la estimación del tamaño del efecto, establecer un criterio más aplicado sin dar la última palabra a la mera aplicación de los criterios estadísticos de las pruebas de decisión (Bailar y Mosteller, 1988; Cohen, 1994; Wilkinson y the Task Force on Statistics Inference, 1999; APA, 2001; Cumming y Finch, 2001; Wolfe y Hanley, 2002; Belia, Fidler, Williams y Cumming, 2005; Cumming y Finch, 2005; o Monterde, Pascual y Frías, 2006). De hecho, Hagen, en 1997, ya apuntaba la idea, por otra parte recurrente, que los intervalos de confianza aportan la misma información que el contraste de hipótesis y se ha hecho patente en el manejo de la media, proporciones y varianza observadas.

No existe la misma tradición en el uso de indicadores resistentes para ese tipo de proceder estadístico, excepto los trabajos que se centran en la formulación de intervalos de confianza para las medianas (Woodruff, 1952; Sheather y McKean, 1987; LeStrelen, 2001; Bonett y Price, 2002; Strelen, 2004; o Dubnicka, 2007). Dado que el uso de la media ha monopolizado el procedimiento estadístico, pocos han sido los trabajos basados en el estudio de la posibilidad verosímil del uso de intervalos de confianza en torno a la

Fecha recepción: 19-11-07 • Fecha aceptación: 20-2-08
Correspondencia: Maribel Peró Cebollero
Facultad de Psicología
Universidad de Barcelona
08035 Barcelona
E-mail: mpero@ub.edu

mediana como criterio de decisión y, menos aún, explorar convenientemente la capacidad de discriminar entre poblaciones sobre las que se muestrea.

Así pues, en el presente trabajo se presentan soluciones en torno a procedimientos estadísticos en relación con la mediana de la distribución observada bajo la perspectiva de un diseño de grupos independientes de dos grupos, y se presenta como posible solución la superposición de los intervalos obtenidos como una alternativa a la significación de la diferencia entre los dos grupos. Este procedimiento obliga a reordenar inmediatamente algunos puntos de vista que son, a nuestro entender, muy sugerentes para el análisis de datos. Utilizar intervalos de confianza para la significación de contrastes puede suponer alguna concepción distinta de lo que implica la simple descripción o exploración e incluso la propia teoría de la decisión estadística.

Intervalos de confianza

La distribución muestral de medias sigue el modelo de la ley normal, por lo que la fómula para la obtención del intervalo de confianza en caso de trabajar con muestra pequeña y varianza poblacional desconocida es:

$$\bar{x} \pm t_{(\alpha,\nu)} \cdot \frac{S}{\sqrt{n}}$$

En el caso de la mediana se han utilizado dos métodos diferentes para la obtención del intervalo de confianza, el basado en el error estándar y el basado en la distribución binomial. Para el primer caso, según Kendall (1945) o Mothes y Torrens-Ibern (1970), si la población es normal con media μ y desviación típica σ y la muestra suficientemente grande, la ley de probabilidad de la mediana tiende a ser una ley normal con las siguientes características:

$$E[mediana] = \mu \qquad VAR[mediana] = \frac{\pi}{2}\frac{\sigma^2}{n} \qquad EE(mediana) = 1{,}253 \cdot \sigma/\sqrt{n}$$

En consecuencia, el intervalo para muestra pequeña y varianza poblacional desconocida se obtiene calculando la siguiente fórmula:

$$Md \pm t_{(\alpha,\nu)} \cdot 1{,}253 \cdot \frac{\sigma}{\sqrt{n}}$$

El segundo método se basa en la obtención de las posiciones correspondientes al límite inferior y superior del intervalo a partir de la aplicación de la distribución binomial, dado que el número de observaciones por debajo del centil k sigue esta ley con parámetros n y k (DeCoster y Burchill, 2000; y Bland, 2003) y la mediana es el punto central de la distribución, k= 0,5, siendo los parámetros que la definen:

Posición que ocupa el valor de la mediana: $\frac{n+1}{2}$

Varianza: n · p · (1 - p)

Error estándar: $\sqrt{n \cdot p \cdot (1-p)}$

En consecuencia, el intervalo de las posiciones se obtiene calculando la siguiente fórmula:

$$\frac{n+1}{2} \pm t_{(\alpha,\nu)} \cdot \sqrt{n \cdot p \cdot (1-p)}$$

Una vez obtenidas las posiciones se redondean al entero más próximo y finalmente se obtienen los valores de la distribución observada que ocupan dichas posiciones.

Freixa, Turbany y Peró (2005) a partir de los datos de dos ejemplos simulados llegan a la conclusión de que el intervalo de confianza de la mediana a partir de la ley binomial es semejante al intervalo que se obtiene según el método de remuestreo *Bootstrap*. Guàrdia, Peró, Freixa y Turbany (2007) llegan a la conclusión de una alta especificidad (valores superiores a .95) en la decisión si el criterio de decisión es estricto.

Estos estudios muestran la adecuada especificidad del procedimiento, pero no se dispone información sobre su senbilidad o potencia; es por este motivo que en el presente trabajo se pretende ampliar las situaciones de simulación y, en consecuencia, no sólo trabajar con muestras que provengan de poblaciones con igualdad de medias como se realizó en Guàrdia et al. (2007), sino que además se pretende trabajar con muestras que provengan de poblaciones con medias diferentes.

Método

Procedimiento

Los datos se han generado vía simulación con el programa EXCEL, bajo el supuesto de trabajar con poblaciones que se distribuyen según el modelo de la ley normal en la población, en concreto se dispone de 4 situaciones de comparación, una con igualdad poblacional de medias (100 - 100) y tres con diferencia poblacional de medias en diferente grado (100 - 110, 100 - 120 y 100 - 130). Además se ha trabajado con diferentes valores de desviación típica (10, 20 y 30) y tamaños de muestra (10, 15, 20, 25 y 30). De todos modos, en todas las comparaciones realizadas se ha trabajado bajo igualdad de desviaciones típicas y diseño de muestras equilibradas. Así pues, cada una de las situaciones simuladas implica 15 condiciones (3 niveles de desviación típica × 5 niveles de tamaño de muestra). Para cada una de las 15 condiciones se han generado 5.000 parejas de muestras.

Análisis de datos

Con el fin de estudiar la igualdad de parámetros bajo las que se han generado las 5.000 parejas de muestras en las 15 condiciones para cada una de las cuatro situaciones generadas, se ha procedido a la obtención del estadístico t de *Student* de grupos independientes, la prueba no paramétrica U de Mann-Whitney y la obtención de los intervalos de confianza, en cada grupo a comparar, para la media y para la mediana a partir del error estándar y a partir de la distribución binomial. El análisis de datos final ha consistido en determinar el porcentaje de decisiones incorrectas para las 4 situaciones estudiadas. En el caso de la t de *Student* de grupos independientes (en este caso también se ha analizado la igualdad de varianzas de las dos muestras) y de la U de Mann-Whitney se ha fijado un α en el 5%, en tanto que para los intervalos de confianza de medias o de medianas se han utilizado dos criterios de decisión. Un criterio más laxo en la decisión ha llevado a rechazar la hipótesis de nulidad si la media o en su caso la mediana del primer grupo no estaba en el intervalo de confianza generado a partir de la media o en su caso la mediana del segundo grupo y viceversa. El segundo criterio, más estricto, ha consistido en rechazar la hipótesis nula en caso de total ausencia de solapamiento de los intervalos de confianza de los dos grupos a comparar.

Resultados

En las tablas 1 a 4 se muestran los porcentajes de errores en la decisión en los 5.000 pares de muestras para las 15 condiciones en cada una de las cuatro situaciones estudiadas. Como se puede observar en las cuatro tablas, el porcentaje de errores al rechazar la hipótesis nula en la condición de aplicación del estadístico t de *Student* de grupos independientes se mantiene alrededor del valor del α nominal fijado (5%).

En la situación de igualdad de medias (tabla 1) se puede observar que si la decisión se toma a partir de pruebas clásicas (t de *Student* o U de Mann-Whitney) el porcentaje de errores está alrederor del α nominal fijado. Pero en el caso de la decisión a partir de los intervalos de confianza el comportamiento es diferencial en función del método y criterio utilizado, así pues, en el caso del criterio de decisión menos estricto, el porcentaje de errores es superior al α nominal fijado, alejándose más de este valor al aumentar la variabilidad, aunque para el intervalo de confianza de la mediana obtenido a partir de la ley binomial hasta un tamaño de 15 sujetos por grupo se mantiene inferior al α nominal fijado y en tamaños de muestra de 30 sujetos por grupo no llega a estar por encima de un 2% del α nominal fijado. Si la decisión se toma a partir del criterio estricto, el porcentaje de error es muy bajo, inferior al 1%, independientemente del estadístico o método utilizado para construir el intervalo.

Las tres situaciones en las que existe diferencia de medias en la población se muestran en las tablas 2 a 4. En las tres tablas se observa un patrón de errores muy similar, pero, obviamente, a medida que aumenta la distancia entre las medias poblacionales el porcentaje de errores en la decisión disminuye, en consecuencia, aumenta la potencia.

En la tabla 2 se muestran estos porcentajes de error cuando la diferencia poblacional de medias es de 10 puntos. El porcentaje de error es muy elevado independientemente de la prueba de decisión utilizada a excepción de la condición de tamaño de muestra de los grupos a comparar de 30 y desviación típica de 10, condición en la que incluso este porcentaje es inferior al α nominal fijado para el estadístico t de *Student*, la U de Mann-Whitney y la decisión a partir de la comparación del intervalo de confianza en el caso de la media para el criterio no estricto de decisión.

En la tabla 3 (diferencia de medias poblacional de 20 puntos) se puede observar que a partir de tamaño de muestra de 20 sujetos en cada grupo, si la desviación típica es pequeña (s= 10), independientemente de la prueba de decisión utilizada, el porcentaje de error en la decisión es sustancialmente inferior al α nominal fijado, siendo en algunos casos nulo. De hecho, cuando la variabilidad es baja (s= 10) para tamaños de muestra de 10 y 15 también se observan estos valores, si la prueba de decisión es la t de *Student*, la U de Mann-Whitney, la comparación de los intervalos de

Tabla 1
Porcentaje de errores en que se rechaza la hipótesis nula cuando en realidad ésta es cierta (medias de las dos poblaciones 100) según los diferentes criterios de decisión estudiados

Medias 100 y 100									
Condición	**C.A. t**	**t Student**	**U M-W**	**$\bar{x}$1**	**Md EE 1**	**Md bin 1**	**$\bar{x}$2**	**Md EE 2**	**Md bin 2**
n= 10_s= 10	6.84%	5.26%	5.22%	8.96%	7.24%	2.18%	0.76%	0.28%	0.12%
n= 10_s= 20	6.18%	5.32%	5.32%	9.30%	7.40%	2.00%	0.54%	0.24%	0.12%
n= 10_s= 30	5.90%	5.34%	5.46%	9.84%	7.80%	2.30%	0.72%	0.38%	0.14%
n= 15_s= 10	5.72%	5.26%	5.4%	11.08%	9.70%	4.36%	0.62%	0.38%	0.30%
n= 15_s= 20	6.12%	5.12%	5.40%	11.08%	9.72%	4.38%	0.52%	0.34%	0.28%
n= 15_s= 30	5.08%	4.74%	5.00%	10.76%	9.42%	4.40%	0.38%	0.46%	0.26%
n= 20_s= 10	5.54%	4.94%	5.08%	11.88%	10.88%	6.14%	0.66%	0.34%	0.44%
n= 20_s= 20	6.14%	5.825	5.48%	12.88%	11.92%	5.76%	0.62%	0.44%	0.40%
n= 20_s= 30	5.76%	5.58%	5.40%	12.38%	10.68%	5.68%	0.72%	0.54%	0.44%
n= 25_s= 10	5.46%	4.90%	4.74%	12.26%	11.62%	6.60%	0.62%	0.40%	0.58%
n= 25_s= 20	4.90%	4.94%	4.96%	12.46%	11.84%	6.76%	0.52%	0.42%	0.58%
n= 25_s= 30	5.62%	5.48%	5.46%	13.12%	11.72%	6.30%	0.40%	0.36%	0.40%
n= 30_s= 10	5.50%	5.26%	5.30%	12.78%	12.3%	6.90%	0.84%	0.56%	0.44%
n= 30_s= 20	5.80%	5.18%	5.52%	13.24%	10.82%	6.42%	0.56%	0.44%	0.30%
n= 30_s= 30	5.38%	5.14%	5.22%	13.54%	12.56%	6.64%	0.62%	0.34%	0.54%

n: tamaño muestra; s: desviación típica; C.A. t: condición aplicación estadístico t de *Student*; t *Student*: criterio decisión estadístico t de *Student*; U M-W: criterio decisión estadístico U de Mann-Whitney; $\bar{x}$1: criterio decisión comparación intervalos de confianza de la media (medias no incluidas en los intervalos); Md EE 1: criterio decisión comparación de intervalos de confianza de la mediana según error estándar (medianas no incluidas en los intervalos); Md bin 1: criterio decisión comparación de intervalos de confianza de la mediana según ley binomial (medianas no incluidas en los intervalos); $\bar{x}$2: criterio decisión comparación de intervalos de confianza de la media (no solapamiento de los intervalos); Md EE 2: criterio decisión comparación de intervalos de confianza de la mediana según error estándar (no solapamiento de los intervalos); y Md bin 2: criterio de decisión comparación de intervalos de confianza de la mediana según ley binomial (no solapamiento de los intervalos)

Tabla 2

Porcentaje de errores en que no se rechaza la hipótesis nula cuando en realidad ésta es falsa (diferencia poblacional de medias: 10) según los criterios de decisión estudiados

				Medias 100 y 110					
Condición	**C.A. t**	**t Student**	**U M-W**	**$\bar{x}$1**	**Md EE 1**	**Md bin 1**	**$\bar{x}$2**	**Md EE 2**	**Md bin 2**
n= 10_s= 10	6.88%	45.22%	46.42%	33.96%	49.38%	67.34%	78.64%	89.52%	94.30%
n= 10_s= 20	6.64%	81.36%	81.28%	72.42%	80.48%	91.20%	96.40%	98.42%	99.32%
n= 10_s= 30	6.30%	88.96%	88.92%	82.66%	86.68%	95.06%	98.18%	99.10%	99.70%
n= 15_s= 10	5.10%	25.58%	27.88%	15.48%	30.88%	40.52%	60.08%	77.68%	82.10%
n= 15_s= 20	6.16%	74.08%	74.82%	60.98%	71.50%	81.28%	93.18%	96.34%	97.60%
n= 15_s= 30	5.84%	86.56%	86.84%	76.50%	81.24%	88.64%	96.88%	98.14%	98.72%
n= 20_s= 10	5.16%	12.56%	14.78%	5.92%	16.94%	22.76%	40.14%	64.12%	65.56%
n= 20_s= 20	5.80%	66.34%	67.80%	50.46%	63.00%	72.40%	89.64%	94.34%	94.80%
n= 20_s= 30	5.88%	81.22%	82.20%	68.26%	75.96%	83.60%	95.86%	97.66%	97.68%
n= 25_s= 10	5.50%	6.56%	8.14%	2.46%	10.68%	14.78%	26.52%	52.40%	53.84%
n= 25_s= 20	5.28%	59.06%	60.54%	42.60%	56.24%	64.88%	85.88%	92.24%	92.00%
n= 25_s= 30	5.40%	79.40%	80.52%	64.30%	71.84%	80.84%	95.02%	97.48%	97.40%
n= 30_s= 10	5.18%	3.26%	3.76%	1.24%	5.64%	8.48%	16.38%	41.74%	43.50%
n= 30_s= 20	5.66%	51.18%	53.56%	34.40%	49.46%	57.52%	81.30%	89.78%	89.86%
n= 30_s= 30	5.50%	74.44%	75.78%	59.80%	68.36%	76.86%	93.18%	96.20%	96.42%

n: tamaño muestra; s: desviación típica; C.A. t: condición aplicación estadístico t de *Student*; t *Student*: criterio decisión estadístico t de *Student*; U M-W: criterio decisión estadístico U de Mann-Whitney; $\bar{x}$1: criterio decisión comparación intervalos de confianza de la media (medias no incluidas en los intervalos); Md EE 1: criterio decisión comparación de intervalos de confianza de la mediana según error estándar (medianas no incluidas en los intervalos); Md bin 1: criterio decisión comparación de intervalos de confianza de la mediana según ley binomial (medianas no incluidas en los intervalos); $\bar{x}$2: criterio decisión comparación de intervalos de confianza de la media (no solapamiento de los intervalos); Md EE 2: criterio decisión comparación de intervalos de confianza de la mediana según error estándar (no solapamiento de los intervalos); y Md bin 2: criterio de decisión comparación de intervalos de confianza de la mediana según ley binomial (no solapamiento de los intervalos)

Tabla 3

Porcentaje de errores en que no se rechaza la hipótesis nula cuando en realidad ésta es falsa (diferencia poblacional de medias: 20) según los criterios de decisión estudiados

				Medias 100 y 120					
Condición	**C.A. t**	**t Student**	**U M-W**	**$\bar{x}$1**	**Md EE 1**	**Md bin 1**	**$\bar{x}$2**	**Md EE 2**	**Md bin 2**
n= 10_s= 10	6.88%	1.28%	1.58%	0.48%	3.80%	9.50%	12.46%	34.86%	50.32%
n= 10_s= 20	6.38%	43.94%	45.38%	33.02%	49.00%	66.28%	78.30%	89.22%	94.20%
n= 10_s= 30	6.62%	71.16%	71.62%	61.00%	71.74%	84.54%	92.00%	96.50%	98.30%
n= 15_s= 10	5.16%	0.04%	0.06%	0.04%	0.42%	0.80%	1.10%	9.86%	12.36%
n= 15_s= 20	5.30%	25.14%	28.18%	15.36%	30.20%	39.82%	59.94%	77.70%	80.32%
n= 15_s= 30	5.42%	58.56%	60.32%	44.98%	57.48%	69.70%	85.98%	92.76%	94.08%
n= 20_s= 10	5.58%	0.00%	0.14%	0.00%	0.06%	0.04%	0.04%	2.12%	2.14%
n= 20_s= 20	5.20%	12.62%	14.42%	6.06%	17.52%	22.32%	39.76%	63.54%	65.32%
n= 20_s= 30	5.26%	46.82%	48.88%	30.54%	46.68%	55.68%	78.08%	89.18%	89.22%
n= 25_s= 10	5.24%	0.00%	0.00%	0.00%	0.00%	0.00%	0.02%	0.48%	0.60%
n= 25_s= 20	5.20%	7.04%	8.12%	2.82%	10.44%	15.225	26.94%	52.56%	53.90%
n= 25_s= 30	5.72%	38.04%	39.28%	22.66%	38.48%	46.66%	69.98%	83.02%	83.58%
n= 30_s= 10	5.46%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.10%	0.04%
n= 30_s= 20	5.26%	3.54%	4.32%	1.50%	5.62%	8.02%	16.74%	40.22%	42.70%
n= 30_s= 30	5.86%	27.3%	29.54%	15.86%	30.92%	37.70%	59.94%	77.92%	79.18%

n: tamaño muestra; s: desviación típica; C.A. t: condición aplicación estadístico t de *Student*; t *Student*: criterio decisión estadístico t de *Student*; U M-W: criterio decisión estadístico U de Mann-Whitney; $\bar{x}$1: criterio decisión comparación intervalos de confianza de la media (medias no incluidas en los intervalos); Md EE 1: criterio decisión comparación de intervalos de confianza de la mediana según error estándar (medianas no incluidas en los intervalos); Md bin 1: criterio decisión comparación de intervalos de confianza de la mediana según ley binomial (medianas no incluidas en los intervalos); $\bar{x}$2: criterio decisión comparación de intervalos de confianza de la media (no solapamiento de los intervalos); Md EE 2: criterio decisión comparación de intervalos de confianza de la mediana según error estándar (no solapamiento de los intervalos); y Md bin 2: criterio de decisión comparación de intervalos de confianza de la mediana según ley binomial (no solapamiento de los intervalos)

confianza para la media y de la mediana obtenidos a partir del error estándar según criterio no estricto.

Cuando la diferencia poblacional de las medias es de 30 puntos (tabla 4), el porcentaje de error es nulo cuando la desviación típica es de 10 y los tamaños de muestra de 20, 25 o 30 en cada grupo comparado, independientemente de la prueba de decisión utilizada, estando alrededor del 0% en las condiciones de tamaños de muestra más pequeños. Cuando la desviación típica es de 20 y los tamaños de muestra de 25 o de 30 el porcentaje de error en la decisión se mantiene también muy bajo, a excepción de la comparación de los intervalos de confianza de medianas. Además, en la comparación de los intervalos de confianza de medias criterio no estricto, el porcentaje de error es inferior al 1% cuando la desviación típica es de 20 a partir de tamaños de muestra de 15 sujetos.

Discusión y conclusiones

A partir de los resultados obtenidos es importante comentar la elevada especificidad del método utilizado, idea ya detectada en Guàrdia et al. (2007). Sin embargo, en el caso de la sensibilidad o potencia no se puede decir lo mismo, pues, cuando en realidad existen diferencias de medias entre las dos poblaciones a comparar, el error tipo II es elevado, siendo en algunos casos muy baja la potencia de la técnica, especialmente si el intervalo de confianza se ha obtenido a partir de la ley binomial. De todos modos, cuando se trabaja con tamaños de muestra de 30 sujetos por grupo la potencia se mantiene en unos valores adecuados, y de hecho en la situación de máxima distancia estudiada la potencia es adecuada en muestras de 20 o más sujetos.

Así pues, según las situaciones y condiciones estudiadas en el presente trabajo, se puede concluir que la técnica de la comparación de los intervalos de confianza de medianas tiene una alta especificidad pero no una alta sensibilidad en todas las condiciones estudiadas, y que en algunos casos funciona mejor la comparación de los intervalos de confianza de medias que los de medianas. De hecho, esta idea no debe extrañar dado que los datos simulados pueden favorecer a la media en contraposición a la mediana, puesto que no debe olvidarse que los datos se han generado bajo población normal. Es por ello que no consideramos que el tema quede agotado en este trabajo, ya que tal como señalan Bonett y Price (2002) la decisión en torno a la mediana puede ser útil en aquellas situaciones en que la distribución de la variable sea sesgada, la distribución sea leptocúrtica, o bien cuando la distribución de la variable se aparte mucho del ajuste a la ley normal. Estos elementos no se han trabajado en el presente estudio y es por este motivo que se considera que la bondad de la comparación de los intervalos de confianza de medianas en la toma de decisión se debería estudiar bajo otras condiciones como, por ejemplo, en muestras con distribuciones sesgadas, tal como se apuntaba en Freixa et al. (2005).

Finalmente, es importante recordar la necesidad de que en futuros trabajos se estudie el efecto del grado de solapamiento de los intervalos de confianza a la hora de tomar la decisión estadística. De hecho, ésta es una cuestión que aún queda por resolver de forma explícita, puesto que los criterios de interpretación no siempre son estadísticamente sólidos y se basan más en la expectativa de los investigadores que en un conocimiento preciso del concepto de nivel de confianza o de valoración de la estimación por intervalo. En esta línea, Cumming y Finch (2005) ya han realizado este es-

Tabla 4

Porcentaje de errores en que no se rechaza la hipótesis nula cuando en realidad ésta es falsa (diferencia poblacional de medias: 30) según los criterios de decisión estudiados

Medias 100 y 130									
Condición	**C.A. t**	**t Student**	**U M-W**	**$\bar{x}$1**	**Md EE 1**	**Md bin 1**	**$\bar{x}$2**	**Md EE 2**	**Md bin 2**
n= 10_s= 10	6.80%	0.02%	0.00%	0.02%	0.06%	0.06%	0.12%	2.26%	7.28%
n= 10_s= 20	6.12%	11.74%	13.36%	7.00%	18.68%	33.22%	42.54%	65.74%	78.22%
n= 10_s= 30	6.04%	43.1%	44.48%	32.10%	48.02%	66.36%	78.04%	89.52%	94.34%
n= 15_s= 10	5.76%	0.00%	0.00%	0.00%	0.00%	0.00%	0.02%	0.10%	0.10%
n= 15_s= 20	5.94%	2.26%	3.12%	0.94%	5.88%	8.88%	15.02%	38.52%	44.40%
n= 15_s= 30	5.66%	25.32%	27.66%	15.06%	29.66%	40.20%	58.92%	77.42%	80.64%
n= 20_s= 10	5.56%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
n= 20_s= 20	5.22%	0.32%	0.50%	0.14%	1.44%	2.14%	4.32%	20.06%	21.54%
n= 20_s= 30	5.54%	12.68%	14.38%	6.44%	17.20%	23.76%	40.60%	64.12%	65.90%
n= 25_s= 10	5.04%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
n= 25_s= 20	5.70%	0.06%	0.06%	0.02%	0.40%	0.68%	1.08%	9.56%	10.48%
n= 25_s= 30	5.48%	6.58%	8.06%	2.88%	10.42%	14.14%	26.90%	52.32%	53.30%
n= 30_s= 10	5.42%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
n= 30_s= 20	5.42%	0.06%	0.06%	0.02%	0.08%	0.12%	0.26%	4.48%	5.225
n= 30_s= 30	5.56%	3.70%	4.06%	1.58%	6.78%	8.86%	16.52%	40.76%	42.06%

n: tamaño muestra; s: desviación típica; C.A. t: condición aplicación estadístico t de *Student*; t *Student*: criterio decisión estadístico t de *Student*; U M-W: criterio decisión estadístico U de Mann-Whitney; $\bar{x}$1: criterio decisión comparación intervalos de confianza de la media (medias no incluidas en los intervalos); Md EE 1: criterio decisión comparación de intervalos de confianza de la mediana según error estándar (medianas no incluidas en los intervalos); Md bin 1: criterio decisión comparación de intervalos de confianza de la mediana según ley binomial (medianas no incluidas en los intervalos); $\bar{x}$2: criterio decisión comparación de intervalos de confianza de la media (no solapamiento de los intervalos); Md EE 2: criterio decisión comparación de intervalos de confianza de la mediana según error estándar (no solapamiento de los intervalos); y Md bin 2: criterio de decisión comparación de intervalos de confianza de la mediana según ley binomial (no solapamiento de los intervalos)

tudio en el caso de la comparación de medias para dos muestras independientes. Idea recogida en Wolfe y Hanley (2002), o Belia, Fidler, Williams y Cumming (2005). Pero esta idea no se ha comprobado si es aplicable al caso de la mediana, a pesar de que de entrada puede ser inadecuado en el caso de los intervalos de confianza obtenidos a partir de la ley binomial ya que son asimétricos respecto a la mediana.

Como conclusión, se podría decir que en los últimos años se ha conseguido incorporar los intervalos de confianza y la estimación del tamaño del efecto, pero a cambio se ha sacrificado una parte de los criterios interpretativos que le dan a las técnicas estadísticas la confiabilidad necesaria. Este y otros trabajos similares deberían contribuir a la mejora en el tratamiento de los datos y, especialmente, en las conclusiones que se deriven de ellos.

Referencias

APA (2001). *Publication manual of the American Psychological Association. Fitth edition*. Washington. American Psychological Association.

Bailar, J., y Mosteller, F. (1988). Guidelines for statistical reporting in articles for medical journals. Amplifications and explanations. *Annals of Internal Medicine*, *108*, 266-273.

Belia, S., Fidler, F., Williams, J., y Cumming, G. (2005). Researchers misunderstand confidence intervals and standard error bars. *Psychological Methods*, *10*(4), 389-396.

Bland, M. (2003). *Confidence interval for a median and other quantiles*. Documento recuperado el 5 de mayo de 2005 del Word Wide Web: http://www-users.york.ac.uk/~mb55/intro/cicent.htm.

Bonett, D.G., y Price, R.M. (2002). Statistical inference for a linear function of medians: Confidence intervals, hypothesis testing and sample size requirements. *Psychological Methods*, *7*(3), 370-383.

Cohen, J. (1994). The earth is round (p<.05). *American Psychologist*, *49*(12), 997-1003.

Cumming, G., y Finch, S. (2001). A primer on the understanding, use and calculation of confidence intervals that are based on central and noncentral distributions. *Educational and Psychological Measurement*, *61*(4), 532-574.

Cumming, G., y Finch, S. (2005). Inference by eye. Confidence intervals and how to read pictures of data. *American Psychologist*, *60*(2), 170-180.

DeCoster, C., y Burchill, C. (2000). *Confidence interval of the median*. Documento recuperado el 5 de mayo de 2005 del Word Wide Web: http://www.umanitoba.ca/centres/MCHP/concept/dict/ci_median.

Dubnicka, S.R. (2007). A confidence interval for the median of a finite population under unequal probability sampling: A model-assisted approach. *Journal of Statistical Planning and Inference*, *137*(7), 2429-2438.

Freixa, M., Turbany, J., y Peró, M. (2005, mayo). *El uso y el abuso de la distancia con respecto a la media aritmética como la mejor medida de sensibilidad en la investigación aplicada en Psicología*. Comunicación presentada en la X Conferencia Española de Biometría, Oviedo (Spain).

Guàrdia, J., Peró, M., Freixa, M., y Turbany, J. (2007, febrero). *Generación mediante simulación de intervalos de confianza en torno a estadísticos resistentes: el caso de la mediana.* Comunicación presentada en el X Congreso de Metodología de las Ciencias Sociales y de la Salud, Barcelona (Spain).

Hagen, R.L. (1997). In praise of the null hypothesis statistical test. *American Psychologist*, *52*(1), 15-24.

Kendall, M.G. (1945). *The advanced theory of statistics. Volume I*. London: Charles Griffin & Company Limited.

Monterde, H., Pascual, J., y Frías, M.D. (2006). Errores de interpretación de los métodos estadísticos: importancia y recomendaciones. *Psicothema*, *18*(4), 848-856.

Mothes, J., y Torrens-Ibern, J. (1970) *Estadística aplicada a la ingeniería*. Barcelona. Ariel.

Sheater, S.J., y McKean, J.W. (1987). A comparison fo testing and confidence interval methods for the median. *Statistics & Probabilitiy Letters*, *6*(1), 31-36.

Strelen, J.C. (2001). *Median confidence intervals*. Documento recuperado el 11 de abril de 2007 del World Wide Web: http://web.informatik.uni-bonn.de/IV/strelen/Forschung/Publikationen/ESM2001.pdf.

Strelen, J.C. (2004). *The accuracy of a new confidence interval method*. Documento recuperado el 11 de abril de 2007 del World Wide Web: http://ieeexplore.ieee.org/iel5/9441/29988/01371373.pdf.

Wilkinson, L., y the Task Force on Statistical Inference (1999). Statistical methods in psychology journals. Guidelines and explanations. *American Psychologist*, *54*(8), 594-604.

Wolfe, R., y Hanley, J. (2002). If ve're so different, why do we keep overlapping? When 1 plus 1 doesn't make 2. *Canadian Medical Association Journal*, *166*(1), 65-66.

Woodruff, R.S. (1952). Confidence intervals for medians and other position measures. *Journal of the American Statistical Association*, *47*, 635-646.