

## El resumen de los resultados en los artículos de investigación de Ciencias Comportamentales

Manuel Miguel Ramos-Álvarez<sup>1</sup> y Andrés Catena Martínez<sup>2\*</sup>

<sup>1</sup>Universidad de Jaén, <sup>2</sup>Universidad de Granada

**Resumen:** El objetivo principal será la aplicación de desarrollos metodológicos recientes de cara a la adecuada elaboración del apartado de resultados propio de los artículos de investigación, atendiendo especialmente al campo de la Psicología y disciplinas comportamentales afines. Para ello, se exponen las generalidades sobre el resumen de información, resaltando especialmente la explicación detallada sobre la elección de la medida de error. En una segunda parte se presentan aspectos más específicos, discutiendo con detenimiento la adaptación de la medida de variabilidad error en función de dos parámetros básicos: el tipo de diseño de investigación y el grado de generalidad de la hipótesis de investigación. El trabajo se cierra con un ejemplo de aplicación de los conceptos expuestos con unos breves apuntes sobre la forma de agregar barras de error a los gráficos a partir de un programa de Hoja de Cálculo, como Microsoft Excel<sup>®</sup>.

**Palabras clave:** Resultados del informe; medida de variabilidad de error; diseños entregrupos *versus* intrasujetos; diseños factoriales *versus* unifactoriales; hipótesis generales *versus* específicas.

**Title:** Summary of results in research papers of behavioural sciences.

**Abstract:** The main aim of this paper is the application of recent methodological developments in order to write adequately the results section of research papers, particularly in the field of Psychology and other behavioural sciences. Thus, we illustrate some general issues about how the information should be summarized, emphasizing the need to explain in detail the choice of the error measure. The second part of the paper introduces more specific issues, as the adaptation of the error variability with reference to two parameters: the research design and the degree of generalization of the research hypothesis. This paper concludes with an applied example of the above issues, including some brief remarks about the way to add error bars to figures and graphics using a spread sheet program (i.e. Microsoft Excel<sup>®</sup>).

**Key words:** Report results; measure of error variability; inter vs. within group designs; factorial vs. unifactorial designs; and general vs. specific hypotheses.

Una de las características más definitorias del proceso científico de investigación es la objetividad, lo que depende de la puesta en común pública de los resultados de la investigación y la principal vía de difusión para lograrlo es la publicación de artículos de investigación en revistas especializadas. Dentro de la elaboración de tales artículos, la redacción del apartado de resultados suele ser una de las más importantes, a la vez que complejas. Nuestro objetivo con este artículo será la aplicación de desarrollos metodológicos recientes de cara a la adecuada elaboración de este tipo de apartado.

Es habitual presentar los resultados primero de forma cualitativa para proceder a continuación a la descripción de los análisis estadísticos del tipo inferencial (recordar el contraste de Hipótesis y la estimación de parámetros). Normalmente se presentan estadísticos descriptivos (promedios, desviaciones típicas, proporciones, etc.) con objeto de resumir los resultados y posteriormente los estadísticos relacionados con el contraste de Hipótesis. Por ejemplo, en relación a F de Snedecor se indican sus grados de libertad, el valor obtenido, el nivel de significación y el término error que se empleó en su cálculo:  $F(3,126)=6.35$ ;  $p<.05$ ;  $MC_{\text{Error}} 425.657$ ; a los que se recomienda añadir la estimación del efecto de tratamiento, i.e.  $\eta^2=0.35$ , y la potencia estadística, i.e.  $1-\beta=0.40$ . La información del segundo tipo está claramente bien establecida (ver las recomendaciones de la APA, 1994) pero la primera, la que tiene que ver con el resumen de los datos, es más controvertida y por ello será nuestro principal objetivo.

En lo que sigue, primero se introducirán las generalidades sobre el resumen de información, lo que nos llevará especialmente a plantear las opciones disponibles para la medida de error, pues es lo más complejo, lo que se abordará en un segundo apartado. Una vez expuestas las generalidades, pasaremos a cubrir los aspectos específicos: por un lado cómo adaptar la medida de variabilidad error en función del tipo de diseño de investigación –apartado tercero- y cómo adaptarla en función del grado de generalidad de la hipótesis de investigación –apartado cuarto-.

El trabajo termina de una manera muy práctica, con un ejemplo de aplicación de los conceptos expuestos –apartado quinto- y unos breves apuntes sobre la forma de agregar barras de error a los gráficos a partir de un programa de Hoja de Cálculo, como Microsoft Excel<sup>®</sup>, con objeto de que cualquier investigador pueda realizar los análisis, a pesar de no disponer de programas costosos de carácter especializado.

### 1. Introducción al resumen de los datos

El resumen de los resultados de la investigación puede realizarse en formato de tablas o bien mediante una representación gráfica. En ocasiones se incluyen los dos tipos de información, a pesar de la redundancia; de manera que los gráficos aparecen en el apartado de los resultados y las tablas en un anexo. La más frecuente a su vez es incluir en los gráficos exclusivamente las medidas de tendencia central (v.gr. los promedios) y en las tablas tanto la tendencia central como la variabilidad (i.e. Desviación Estándar).

Pero a este respecto también se recomienda que los propios gráficos sinteticen los dos tipos de medida. Pensemos que el objetivo es comunicar la información más relevante pero de manera que se facilite su interpretación por parte del

\* Dirección para correspondencia [Correspondence address]: Manuel M. Ramos-Álvarez. Departamento de Psicología. Universidad de Jaén. Paraje Las Lagunillas s/n –Edificio D-2, Buzón 19-. 23071 Jaén (España). E-mail: [mramos@ujaen.es](mailto:mramos@ujaen.es)

lector. Los gráficos proporcionados por la perspectiva de análisis exploratorio de datos (EDA, ver Behrens, 1997), como por ejemplo el de “cajas y barbas” serían bastante recomendables entonces puesto que son muy intuitivos a la vez que completos. El problema fundamental es que su utilización en el propio informe de investigación no está muy extendida. Por ello siguen predominando los gráficos más clásicos, histogramas y diagramas de dispersión.

Sea cual sea el formato, tablas o gráficos, éste incluirá la medida de tendencia central para cada una de las condiciones experimentales así como una medida de variabilidad. Respecto a las del primer tipo, está muy clara la tendencia predominante: se obtiene el promedio –la media aritmética– de cada condición experimental. Por ejemplo, en un diseño unifactorial de tres niveles de manipulación, la media de las observaciones (supongamos que hay  $n=10$  por cada nivel) en cada nivel por separado. Hasta aquí es indiferente que sea un diseño Entregrupos o Intrasujetos. Igualmente, en un diseño factorial, por ejemplo del tipo  $2 \times 3$  (la var. independiente A tiene 2 niveles y B que tiene tres niveles); habría que obtener las medias de cada combinación de niveles, luego seis medias en el ejemplo. Alternativamente, se puede optar por estadísticos de tendencia central robustos como la mediana o la media recortada (ver Smith y Prentice, 1993). Estos últimos son de utilidad cuando los datos a resumir incumplen alguno de los supuestos que impone la estadística clásica. Por ejemplo, si la variable dependiente es el Tiempo de Reacción lo más probable es que optemos por este tipo de medidas robustas pues al menos el supuesto de Normalidad no se verá cubierto. Parece, pues, que efectivamente esta parte no presenta problemas, sólo opciones.

## 2. El resumen de la variabilidad ¿Qué tipo de estimación del error es la más apropiada?

Las dudas están en las medidas de variabilidad, ya que a este respecto hay discrepancias entre los especialistas de metodología (Judd, McClelland y Culhane, 1995) e incluso entre las normas de las editoriales. Pensemos que el objetivo de la comunicación de la información sobre variabilidad es facilitar la comprensión del error contenido en los datos de la investigación. Y esto dependerá de consideraciones inferenciales. Es decir, la medida de variabilidad podría obedecer a fines meramente descriptivos o bien ir más allá y cubrir una función inferencial. A diferencia de la práctica más extendida, la medida de variabilidad que se incluye en un informe de Ciencias Comportamentales tiene que obedecer en última instancia al fin inferencial. Habitualmente, o no se incluye medida de variabilidad o si se hace; ésta tiene la misma lógica que la media: por ejemplo se calcula la desviación típica insesgada (o la cuasivarianza) de cada condición experimental una por una.

Según las recomendaciones que han surgido recientemente (ver Wilkinson and the Task Force on Statistical Inference, 1999), la medida de variabilidad que se incluya en el

informe, con la función inferencial, dependerá del tipo de diseño, es decir del esquema de investigación que se ha seguido. Cuando nos movemos en el esquema experimental o Cuasiexperimental, es decir orientado a la búsqueda de relaciones causales, y se pretende cubrir un fin teórico, es decir poner a prueba teorías; entonces el fin último debería de ser juzgar en qué medida las medias muestrales observadas en el experimento constituyen un buen estimador de las medias en la población. En comparación, si el enfoque es más bien aplicado o no se ha trabajado bajo el esquema experimental, entonces el fin último podría ser más bien juzgar en qué medida las medias muestrales observadas en el estudio son replicables. Esto debería marcar la primera decisión a adoptar. Si el objetivo es el segundo, el de replicabilidad, entonces la medida de variabilidad más adecuada es el error estándar de la media (SEM en adelante, según las siglas del término en inglés, *Standard Error of Means*, también llamado error típico de la media). En cambio, si el objetivo es el primero, de estimación de las medias, entonces la medida de variabilidad vendrá dada por los intervalos confidenciales más bien (ver Estes, 1997). Ambas medidas están además muy relacionadas ya que los intervalos confidenciales en realidad toman el valor del error estándar SEM para computar el error máximo tolerado en función de un nivel de significación, atendiendo a la distribución del estadístico de contraste. Recordemos brevemente estos dos conceptos básicos.

$$\sigma_{\mu} = \frac{\sigma}{\sqrt{n}}; SEM \equiv \hat{\sigma}_{\mu} = \frac{\sqrt{MC_{\epsilon}}}{\sqrt{n}}$$

Que expresa la definición básica a nivel poblacional (la desviación de la Media) y la estimación correspondiente a partir de los resultados observados en la muestra, o SEM propiamente. En pocas palabras, la medida SEM de variabilidad se define como el cociente entre la raíz de la Media de Cuadrados Error y la raíz del tamaño muestral.

Y cuando expresa una medida de variabilidad por ejemplo en una representación gráfica, se extiende a ambos lados de la media según:

$$\mu_j \pm SEM$$

Como se puede apreciar, estrictamente la medida SEM se define sobre la base de la media poblacional. Expresa así el grado de fluctuación que esperaríamos en torno al valor de la verdadera media en la población, en sucesivas repeticiones de la investigación. Pero, usualmente en Ciencias Comportamentales el valor de la media poblacional no se conoce y de hecho interesaría más bien estimar el grado de fluctuación en torno a la media muestral que se ha observado en el experimento. En cuyo caso, la fórmula anterior lleva asociado un sesgo importante. Pero se ha demostrado que un factor corrector de aproximadamente  $\sqrt{2}$  podría corregir di-

cho sesgo (Estes, 1997). En definitiva, si nuestro objetivo es el de medir la replicabilidad en torno a las medias muestrales, la fórmula quedaría definitivamente como:

$$\bar{X}_j \pm SEM \sqrt{2}$$

Por otro lado, el intervalo confidencial se define según:

$$\bar{X}_j \pm (SEM) \left( {}_{\alpha}Val.Crit_{gl} \right)$$

Es decir, el Intervalo Confidencial extiende a ambos lados de la media el valor del error estándar, SEM, pero ponderado en función del valor crítico asociado al valor del estadístico. Como ya sabemos el valor crítico del estadístico vendrá dado por un nivel de significación (alfa) y unos determinados grados de libertad (gl). Por ejemplo, en la prueba más universal el valor crítico corresponde a un modelo F en el que se introducen dos fuentes de información, los grados de libertad del numerador y los del denominador.

Estas fórmulas son de carácter general, lo que nos permitirá adaptarlas a cualquier tipo de diseño como veremos a continuación.

A su vez, en cualquiera de los dos casos, hay que adoptar una nueva decisión que dependerá del tipo de hipótesis que guía la misma. Cuando la Hipótesis de investigación se plantea en términos generales entonces la variabilidad que se representa en torno a las medias se estima a partir de una medida error tal y como se acaba de plantear. En cambio, cuando la hipótesis de investigación es de carácter más específico, del tipo de un contraste o comparación de medias, entonces hay que aplicar un factor corrector sobre las fórmulas precedentes. Es fácil demostrar (ver Loftus y Masson, 1994 y Estes, 1997) que la relación viene dada de la siguiente manera:

$$\left| \bar{X}_1 - \bar{X}_2 \right| > \sqrt{2} \bullet Error$$

Es decir, en el caso de que la hipótesis sea más concreta, sobre las fórmulas básicas que se plantearon más arriba, habría que multiplicar por la raíz cuadrada de dos; lo que permite nuevamente una gran comparabilidad entre todas las aproximaciones. Esto es aplicable, además, tanto a la perspectiva basada en los errores SEM como a la perspectiva de los Intervalos Confidenciales.

Finalmente, la tercera decisión que hay que adoptar tiene que ver con el tipo de diseño concreto. En todos los casos se recomienda que la medida de error sea de carácter global (*pooled* en la literatura especializada) pero dependiendo del tipo de diseño dicha estimación podría variar. En la Figura 1 se facilita al lector el cálculo de los errores para todos los diseños básicos, partiendo del conocimiento de la tabla resumen ANOVA prototípica. A partir de aquí se puede genera-

lizar con facilidad a otros casos más complejos (v.gr. trifactoriales).

### 3. La medida de error en función del tipo de diseño

En el caso unifactorial –una única variable independiente–, la medida se obtiene directamente a partir del error de la tabla ANOVA. Por ejemplo, en el tipo unifactorial Entregrupos se emplearía la Media de Cuadrados Intra (o S/A como también se denomina) con “N-a” grados de libertad asociados. En el caso unifactorial Intrasujetos el error corresponde a la interacción “Sujetos por Tratamientos” y lleva asociados “(a-1)(n-1)” grados de libertad. Recordar que en el diseño Intrasujetos (o también llamado de medidas repetidas), todos los participantes reciben todos los niveles de la variable independiente, como por ejemplo si manipulamos tres dosis de un fármaco y todos los participantes reciben los tres tipos de dosis.

En los diseños factoriales –con más de una variable independiente, como por ejemplo la motivación y la cantidad de horas de estudio–, el cómputo se complica en la medida en la que hay diseños en los que se calcula más de un error. Por ejemplo, en el caso factorial Entregrupos no hay más problemas pues se tomaría como error el único que aparece en la tabla ANOVA, es decir la Media de Cuadrados Intra nuevamente (o S/AB) con “N-ab” grados de libertad, donde “ab” se refiere al número de condiciones experimentales o combinaciones de niveles de sendas variables independientes.

En cambio, en el diseño factorial Intrasujetos, puesto que hay que calcular un término error específico a cada fuente de variación, la estimación global podría venir dada por un promedio de todos ellos. Es decir, se suman las Sumas de Cuadrados de todas las fuentes de error y esta suma se divide entre la suma de los grados de libertad que llevan asociadas todas las fuentes. Por ejemplo, en el caso bifactorial, de partida tendríamos tres términos: AxS, BxS y AxBxS, cada una de las tres con sus grados asociados. Pues bien, sumamos las tres sumas de cuadrados por un lado y los grados de libertad por el otro “(a-1)(n-1) + (b-1)(n-1) + (a-1)(b-1)(n-1)” y dividimos el resultado de la primera suma entre el de la segunda. En realidad es fácil demostrar que la suma de grados de libertad especificada se reduce a “(n-1)(ab-1)”; es decir sería como reducir las condiciones experimentales a un nuevo término formado por “ab” niveles de tratamiento y plantear el error con la lógica general de un diseños Intrasujetos: la interacción de sujetos con estos tratamientos.

Finalmente, en el diseño factorial Mixto (combina la manipulación Entregrupos con la del tipo Intrasujetos), puesto que también hay más de un error posible, tendríamos que proceder como en el diseño precedente. Se suman las sumas de cuadrados asociadas a los errores y se divide el resultado por la suma de los grados de libertad que les corresponden. Por ejemplo, en un diseño bifactorial Mixto en el que A se ha manipulado Entregrupos y B Intrasujetos, tendríamos

dos errores,  $S/A$  y  $BxS/A$ . En consecuencia, sumáramos las dos sumas de cuadrados correspondientes y también los grados de libertad, " $a(n-1) + a(b-1)(n-1)$ ". Entonces dividiríamos el resultado de la SSCC entre el de los ggl. No obstante, también aquí se puede demostrar fácilmente que la suma de los grados de libertad se reduce a " $(n-1)ab$ ".

Para terminar, queda una aclaración respecto al tipo de error más conveniente en función del diseño. En todos los casos, el error básico es el mismo independientemente de las medias que se desee plantear, lo que varía es el factor corrector en todos los casos. Pero hay una excepción a esta regla general, puesto que el diseño factorial Mixto mezcla los dos tipos de manipulación, el error dependerá del tipo de medias comparadas. Cuando la Hipótesis del investigador versa sobre las medias en los niveles de la variable manipulada Entre entonces se utiliza el error global que acabamos de explicar. Esto sería análogo a adoptar una perspectiva de efectos simples del tipo A en  $b_k$ . En cambio, cuando la hipótesis versa sobre los niveles de la variable Intrasujetos (o efectos simples del tipo B en  $a_j$ ), el término error que se recomienda es el que corresponde a la interacción, es decir el error del tipo  $BxS/A$  con " $a(b-1)(n-1)$ " grados de libertad asociados. Es decir, en este último caso no se recomienda la medida global de los errores.

En la Figura 1 se resumen todos los cálculos pertinentes a este apartado. La definición básica de error corresponde al estadístico SEM. A partir del mismo se puede realizar la representación básica tanto a partir de barras SEM como a partir de los Intervalos Confidenciales. A su vez, en las dos tablas de la mencionada figura se proporciona la definición básica de error SEM en función del tipo de diseño y de Hipótesis. Todo lo que hay que hacer, pues, es sustituir dicho error y sus grados de libertad asociados dentro de las fórmulas básicas que aparecen en primer lugar, respecto a la representación basada en SEM y la que se basa en los Intervalos Confidenciales.

Lo que se ha planteado hasta aquí es la recomendación general que tiene más fuerza, pero no deja de ser controvertida.

#### 4.- La medida de error en función del tipo de hipótesis

Sabemos que el tipo de hipótesis en realidad marca diferentes tipos de error, más allá de lo que hemos expuesto: o bien se obtiene una medida de variabilidad concreta para las condiciones experimentales o bien una medida global (*pooled*) a través de las mismas. Esta problemática es compleja y su puede consultar en cualquier fuente especializada (Ramos, Catena y Trujillo, 2004; Rosenthal y Rosnow, 1985). Si el investigador analiza los datos según un enfoque global, realizando el ANOVA, entonces el error que corresponde a la medida de variabilidad será del tipo general; es decir una

misma medida para todas las condiciones experimentales. Por el contrario, el análisis detallado de contrastes o comparaciones llevará usualmente a errores más específicos, es decir que diferentes condiciones experimentales requerirán diferentes errores.

No obstante, esto no deja de ser nuestra propia aproximación y no se puede tomar como una recomendación general. Si de base los especialistas discrepan en torno al error más conveniente que se debe utilizar para realizar el contraste de hipótesis (v.gr. en ANOVA), claramente habrá un paralelismo en cuanto al error más indicado para resumir la información en torno a la variabilidad. Por ejemplo, autores como Estes (1991, 1997) tienden hacia los errores del tipo global, básicamente porque es la mejor forma de facilitar la comparación entre diferentes experimentos. En cambio, otros autores como Rosenthal (Hallam y Rosenthal, 2000) o Loftus (Loftus y Masson, 1994) tienden a defender errores de carácter más específico o concreto, porque esto evita problemas en torno a los supuestos de la técnica, a la vez que incrementa la potencia estadística. Es evidente que las dos posturas tienen motivos razonables. Luego, a pesar de que nuestra postura particular está más cerca de la segunda, se podría recomendar en general que el investigador incluya la información suficiente como para facilitar al lector la decisión última.

De hecho, los autores van más allá, recomendando incluso un tipo de formato u otro en función del tipo de error. Si nos decantamos por la aproximación basada en SEM entonces se recomienda resumir los datos en forma de tablas, mientras que la aproximación según intervalos confidenciales quedará mejor sintetizada mediante representaciones gráficas. El motivo es claramente funcional: se debe de facilitar siempre en la medida de lo posible la comprensión de los datos.

De aquí, surge pues una paradoja. Según se ha comprobado, los investigadores que en realidad dan a su investigación un enfoque eminentemente teórico, tienden a incluir en las representaciones gráficas las barras SEM en torno a las medias muestrales y frecuentemente según errores que son específicos a cada condición experimental. Luego, se estarían infringiendo todas las recomendaciones que hemos ido planteando. Esto por ejemplo constituye una tendencia muy extendida en el ámbito de las investigaciones psicofisiológicas o psicobiológicas, por citar algún campo destacado, en las que las propias editoriales exigen al investigador que incluya una barra de error SEM en los gráficos y posteriormente no se preocupan de las interpretaciones de los mismos. Es decir, el origen de los hábitos inadecuados no necesariamente hay que buscarlo en los investigadores sino en las propias editoriales que marcan los cánones. Esto que acabamos de expresar lo puede constatar cualquier lector echando un vistazo a las publicaciones que aparecen en el campo señalado o en cualquier otro.

**REPRESENTACIÓN BASADA EN SEM:**

$$SEM = \sqrt{\frac{MC_{\varepsilon}}{n}}; \mu_j \pm SEM$$

Para medias muestrales:  $\bar{X}_j \pm SEM \sqrt{2}$

**REPRESENTACIÓN BASADA EN INTERVALOS CONFIDENCIALES:**

$$IC = \bar{X}_j \pm (SEM) (\alpha Val.Crit_{gl})$$

Para la comparación de medias:  $IC \sqrt{2}$

**ESPECIFICACIÓN DEL ERROR EN DISEÑOS UNIFACTORIALES:**

FUENTE	FÓRMULAS ERROR ENTREGUPOS	FÓRMULAS ERROR INTRASUJETOS
Error ANOVA	$MC_{S/A}; gl = a(n-1)$	$MC_{AxS}; gl = (a-1)(n-1)$
Medias individuales	$\sqrt{\frac{MC_{S/A}}{n}}$	$\sqrt{\frac{MC_{AxS}}{n}}$
Contraste de medias en A	$\sqrt{2} \sqrt{\frac{MC_{S/A}}{n}}$	$\sqrt{2} \sqrt{\frac{MC_{AxS}}{n}}$

**ESPECIFICACIÓN DEL ERROR EN DISEÑOS BIFACTORIALES:**

FUENTE	FÓRMULAS ERROR ENTREGUPOS	FÓRMULAS ERROR ERROR INTRASUJETOS	FÓRMULAS ERROR ERROR MIXTO -A:Entre y B:Intra-
Errores ANOVA	$MC_{S/AB}$	$MC_{AxS}; MC_{BxS}; MC_{AxBxS}$	$MC_{S/A}; MC_{BxS/A}$
Error Global	$MC_{S/AB}; gl = ab(n-1)$	$MC_{Intra1} = \frac{SC_{AxS} + SC_{BxS} + SC_{AxBxS}}{gl_{Intra1}};$ $gl_{Intra1} = (a-1)(n-1) + (b-1)(n-1) + (a-1)(b-1)(n-1) = (n-1)(ab-1)$	$MC_{Intra2} = \frac{SC_{S/A} + SC_{BxS/A}}{a(n-1) + a(b-1)(n-1)}$ $gl_{Intra2} = a(n-1) + a(b-1)(n-1) = (n-1)(ab)$
Medias individuales	$\sqrt{\frac{MC_{S/AB}}{n}}$	$\sqrt{\frac{MC_{Intra1}}{n}}$	$\sqrt{\frac{MC_{Intra2}}{n}}$
Contraste de medias en A	$\sqrt{2} \sqrt{\frac{MC_{S/AB}}{n}}$	$\sqrt{2} \sqrt{\frac{MC_{Intra1}}{n}}$	$\sqrt{2} \sqrt{\frac{MC_{Intra2}}{n}}$
Contraste de medias en B	$\sqrt{2} \sqrt{\frac{MC_{S/AB}}{n}}$	$\sqrt{2} \sqrt{\frac{MC_{Intra1}}{n}}$	$\sqrt{2} \sqrt{\frac{MC_{BxS/A}}{n}}$

Figura 1: Resumen de las fórmulas relevantes al cálculo de SEM e Intervalos Confidenciales para los diseños más destacados.

### 5. Ejemplo de aplicación a una investigación

Vamos a ejemplificar los cálculos precedentes mediante una investigación experimental extraída del contexto psicofísico. Esta corresponde al efecto de la probabilidad estimular (abreviada “ProbEs” en adelante) en la exactitud de respuesta. El diseño incluyó la manipulación de tres niveles de variable independiente, probabilidades P.25, P.50 y P.75, aplicados a diferentes muestras de 10 participantes cada uno. Luego, se trata de un diseño unifactorial Entregrupos. A cada persona se le midió la exactitud de respuesta según el índice  $d'$  de la Teoría de Detección de Señales. Para poder realizar las estimaciones necesitamos las medias y la tabla ANOVA de este diseño, que reproducimos a continuación.

	P.25	P.50	P.75
Media	0.43	0.67	0.42
Var.	0.015	0.002	0.048

	Fuente	SC	gl	MC	Fk
Entre (A)->	ProbEs	0.424	2	0.212	9.687
Intra(S/A)>	Error	0.591	27	0.022	
	Total	1.015	29	0.035	

De manera que el error equivale a 0.022, la Media de Cuadrados Intra (o S/A). De hecho, en este diseño este término es equivalente al promedio de los errores (varianzas) individuales. Siguiendo las fórmulas de la Figura 1, se ha confeccionado un libro de trabajo, ver la Figura 2, con todas las estimaciones posibles.

En la parte superior del mismo se detallan los cálculos básicos del Error Global, tal y como intervienen en la fórmula del Error Estándar de la Media, SEM. Es decir, se divide la Media Cuadrados Intra del ANOVA global entre el tamaño muestral (aquí 10 sujetos por condición experimental) y al resultado se le calcula la raíz cuadrada. Así llegaríamos a un valor SEM global de 0.047. A continuación se ejemplifican los cálculos necesarios para la representación gráfica basada en este tipo de error SEM y justo debajo los que implicaría la representación basada en Intervalos Confidenciales. En ambos casos además se ha detallado la estimación para medias individuales, como se haría desde una perspectiva global, y también la estimación en la que interesa más bien comprar medias entre sí, como sucedería cuando el investigador dispone de hipótesis más específicas. Entre ambas, la única diferencia es que hay que aplicar el factor corrector basado en la raíz cuadrada de dos.

Así, pues, para la representación basada en SEM y según las medias individuales, suponiendo que el centro de la representación lo queremos ubicar en las medias muestrales más que poblacionales, tendríamos que multiplicar el valor SEM por la raíz cuadrada de dos (ver Estes, 1997). Lo que nos da el valor 0.066. Si en cambio la representación SEM está guiada según contrastes o comparaciones específicos

entre pares de medias entonces a su vez el valor precedente se multiplicaría por la raíz cuadrada de dos, es decir 0.094.

	A	B	C	D
17				
18	Global	0,022	0,022	0,022
19		10	10	10
20		0,047	0,047	0,047
21				
22		SEM		
23		Cada Media Individual $\bar{X}_j \pm SEM\sqrt{2}$		
24		0,066	0,066	0,066
25				
26		Comparar Medias en A		
27	Factor = $\sqrt{2}$	0,094	0,094	0,094
28				
29		Intervalos Confidenciales		
30		Cada Media Individual		
31	Media	0,43	0,67	0,42
32	SEM	0,047	0,047	0,047
33	Val. Crit	2,052	2,052	2,052
34	Prod	0,096	0,096	0,096
35	IC-	0,330	0,577	0,320
36	IC+	0,522	0,769	0,512
37				
38	Factor = $\sqrt{2}$	Comparar Medias en A		
39	Media	0,43	0,67	0,42
40	SEM	0,047	0,047	0,047
41	Val. Crit	2,052	2,052	2,052
42	Prod	0,136	0,136	0,136
43	IC-	0,290	0,537	0,280
44	IC+	0,562	0,809	0,552
45				

Figura 2: Ejemplificación del cálculo de SEM e Intervalos Confidenciales.

En cambio, si la representación la vamos a basar en los Intervalos Confidenciales entonces necesitamos además del valor del error SEM global, el valor crítico que corresponde al estadístico de contraste. En este caso éste es una t-student con alfa y los grados de libertad del error (ver Ramos *et. al.*, 2004 o cualquier manual de análisis estadístico). Supongamos que trabajamos uniformemente con un nivel de significación de 0,05 para todas las decisiones relativas a este diseño. Además, tal y como nos recuerda la tabla del ANOVA, los grados de libertad del error en este diseño vienen dados por  $N-a$ , ó  $30-3 = 27$ . Luego, determinaríamos el valor crítico de una t-sudent con 0.05 y 27 grados de libertad como parámetros. El propio libro de trabajo nos permite obtener este tipo de valores que en el ejemplo se estima en 2.052 (ver las funciones estadísticas que incluye el programa Microsoft Excel ©). Luego, ya tenemos todos los valores que intervienen en el Intervalo Confidencial, SEM  $-0.047$ - y el calor crítico  $-2.052$ -. Si nuestro objetivo es representar el Intervalo Confidencial para cada medida individual entonces basta con multiplicar los dos valores precedentes, es decir 0.096. El lí-

mite inferior del intervalo se obtiene restando a la media correspondiente el valor del error máximo que acabamos de estimar; y el límite superior sumando dicho valor a la media. Por ejemplo, para el primer nivel de probabilidad estimular, el Intervalo se concretaría de la siguiente manera:

$$IC \left\langle \begin{array}{l} 0,43 - 0,096 \\ 0,43 + 0,096 \end{array} \right\rangle = \left\langle \begin{array}{l} 0,330 \\ 0,522 \end{array} \right\rangle$$

Para efectuar la representación según la aproximación basada en la comparación de medias, todo lo que acabamos de explicar es idéntico, pero cuando estimamos el error máximo, el producto del error SEM y el valor crítico hay que corregirlo según el factor dado por  $\sqrt{2}$ . Es decir, a las medias se les suma y se les resta 0.136 en lugar de 0.096.

Ya hemos visto que la elección entre las cuatro opciones indicadas vendrá dada por el enfoque de la investigación así como por el tipo de hipótesis. En el ejemplo concreto que estamos analizando, puesto que el interés es claramente teórico y no se trata de un estudio de replicación, nos inclinaremos por los intervalos confidenciales. A su vez, puesto que el investigador disponía de hipótesis concretas (Por ejemplo, si la probabilidad estimular es el 50% o superior entonces el efecto sobre la capacidad sensorial es máximo. Además, podría haberse supuesto que la discriminación sensorial es muy baja en comparación cuando la probabilidad estimular es inferior al 50%), a su vez los intervalos confidenciales se establecerían en torno a la comparación de medias. Luego, definitivamente, nos quedaríamos con los intervalos confidenciales: [0.290; 0.562] para el nivel P.25, [0.537; 0.809] para el nivel P.50 y [0.280; 0.552] para el nivel P.75.

## 6. Cómo agregar barras de error a los gráficos del diseño

Una vez estudiadas las consideraciones teóricas precedentes, el investigador habrá realizado los cálculos necesarios en torno a los errores que considere más oportunos. Dichos cálculos se depositan en una fila del libro de trabajo paralela a la

fila de las medias. Es decir, la representación gráfica que buscamos incluirá las medias y los errores asociados a cada una de las mismas. Además, se puede también especificar en otra fila los títulos o etiquetas que corresponden a cada uno de los niveles de la variable independiente. En la Figura 2 aparece el libro de trabajo que nos servirá como ejemplo. Este corresponde a la investigación experimental sobre el efecto de la probabilidad estimular en la exactitud de respuesta. Como se puede apreciar el diseño incluyó la manipulación de tres niveles de variable independiente, probabilidades P.25, P.50 y P.75, aplicados a diferentes muestras de 10 participantes cada uno. Tras introducir los datos obtenidos para cada uno de los 30 participantes, se procede a la estimación de los estadísticos descriptivos, media y varianza. Supongamos para empezar que la medida de variabilidad error que se ha decidido incluir en el gráfico es la variabilidad individual de cada condición experimental. Para obtenerla, basta con especificar los tamaños muestrales de cada nivel de la variable (fila  $n_i$ ); es decir 10 para cada uno de los tres niveles. Entonces obtenemos la raíz cuadrada del cociente entre la varianza –insesgada– y el tamaño muestral. Por ejemplo, para el primer nivel de la variable, la fórmula del libro de trabajo que nos permite obtener dicha estimación es  $=RAIZ(B14/B15)$ ; puesto que en la casilla B14 se encuentra la varianza de este grupo y en B15 su tamaño muestral.

En definitiva, el gráfico se basará en la fila que incluye las medias y en la de los errores SEM. Aunque, tal y como venimos planteando la estimación del error podría ser algo más compleja, dejamos este aspecto hasta el final del apartado, pues es irrelevante para el objetivo actual.

Además, puesto que se han introducido en la primera fila de la tabla de datos, también especificaremos los nombres de los tres niveles. Para iniciar el gráfico, se selecciona con el ratón la fila de los títulos y la de las medias. De momento dejamos los errores. Entonces se pulsa sobre el botón de herramientas que invoca al asistente para gráficos (si el lector no tiene esta herramienta entonces puede hacer lo mismo pulsando sobre el menú [Insertar] y después sobre el submenú [Gráfico...]). Todo esto se ha indicado en la Figura 3.

	A	B	C	D	E	F	G	H
1			d'					
2	Id. Suj	P.25	P.50	P.75				
3	S1	0,43	0,66	0,36				
4	S2	0,42	0,72	0,39				
5	S3	0,44	0,59	0,55				
6	S4	0,45	0,68	0,15				
7	S5	0,38	0,64	0,10				
8	S6	0,59	0,71	0,58				
9	S7	0,46	0,74	0,14				
10	S8	0,57	0,67	0,64				
11	S9	0,14	0,61	0,64				
12	S10	0,38	0,71	0,61				
13	Media	0,43	0,67	0,42	0,51			
14	Var.	0,02	0,00	0,05	0,02			
15	n <sub>j</sub>	10	10	10	Total			
16	SEM <sub>j</sub>	0,039	0,015	0,069				
17								

Figura 3: Agregación de barras Error a un gráfico de resultados- Paso 1.

El programa despliega entonces una ventana como la de la figura 4 que corresponde al primer paso del asistente para gráficos. Aquí hay que señalar el tipo de gráfico. Puesto que nuestra variable independiente es categórica –no es numérica–, lo más indicado sería un histograma. Luego, el que viene por defecto señalado nos sirve –la primera de las categorías de la parte izquierda–. Como subtipo dejamos también el más sencillo pues al no haber terceras variables no es conveniente un gráfico tridimensional.

Una vez seleccionado el tipo y el subtipo, pulsamos sobre el botón [Siguiente >], en la parte inferior-derecha de la ventana. El paso 2 se refiere a los datos de origen (confirmar donde están los valores del eje de abscisas y de ordenadas), el paso 3 a Opciones de gráfico (títulos de las leyendas y del gráfico, formatos de letra, etc.) y el cuarto y último a la ubicación del mismo en el libro de trabajo. De todo esto lo fundamental es la especificación de los datos pero como ya se habían seleccionado previamente, no tenemos que indicar nada en este punto. Todas las definiciones restantes son secundarias al objetivo que buscamos y por lo tanto pulsamos [Finalizar] ya en el primer paso del asistente, saltando todo lo demás, para que nos devuelva la representación gráfica.

El resultado es un gráfico rudimentario que por defecto se ubica en la misma hoja del libro de trabajo en el que nos situamos actualmente. Este gráfico solamente contiene las medias. En consecuencia, el segundo paso será añadirle las barras de error. Para esto, tenemos que situarnos con el ratón sobre el gráfico y pulsar sobre alguno de los tres histogramas con objeto de que se habiliten las opciones de las se-

ries de datos. Entonces pulsamos dos veces seguidas sobre el botón principal del ratón. Se despliega de esta manera la ventana que sirve para especificar los parámetros en torno a la serie de datos, como el que se muestra en la Figura 5.

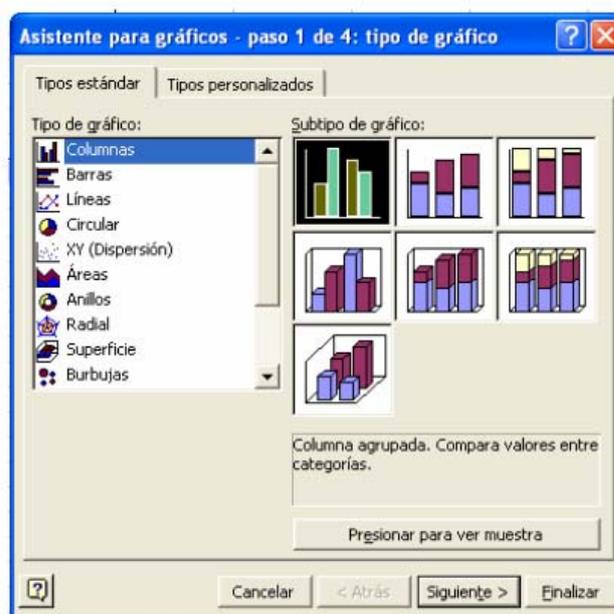


Figura 4: Agregación de barras Error a un gráfico de resultados- Paso 2.

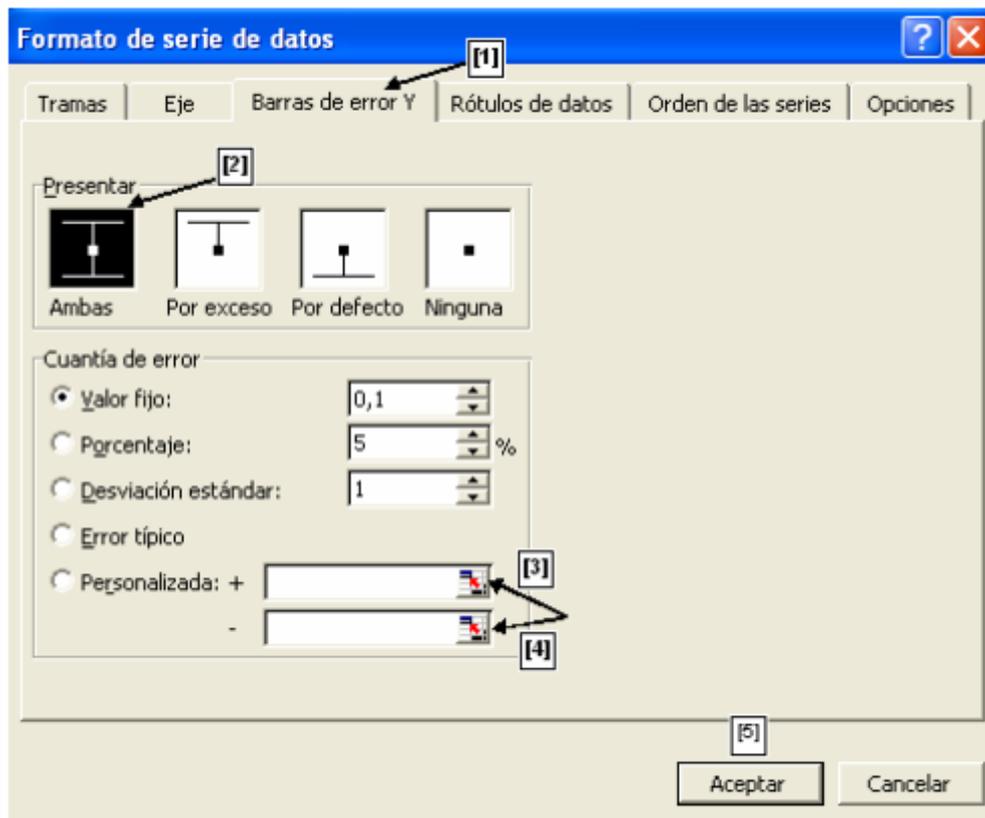


Figura 5: Agregación de barras Error a un gráfico de resultados- Paso 3.

En esta ventana es donde tenemos que efectuar prácticamente todas las operaciones que nos quedan. Por este motivo se ha incluido un número en la figura para guiar la secuencia. Primero [1] de todo hay que seleccionar la pestaña “Barras de Error Y” mediante el ratón. A continuación [2] se elige el formato de las barras. Lo más frecuente es incluir el error hacia ambos lados de la media y por ello seleccionamos el primer formato que aparece por la parte izquierda como en la figura. En tercer lugar queda indicarle al programa el tipo de error que se va a definir. Aquí es donde estriba el paso más crítico desde el punto de vista teórico. Aunque, como se puede apreciar en la figura, el programa incluye la posibilidad de calcular el error, debemos desconfiar de estas opciones ya que en la mayoría de los diseños, los cálculos no son adecuados. El investigador debe de calcular siempre el error por su cuenta de manera expresa (aunque sea el mismo para todas las medias) y emplear el programa simplemente para realizar la representación gráfica. Y esta es precisamente la aproximación que se ha seguido en el ejemplo. Para esto tenemos que utilizar la opción Personalizada, indicando al programa la fila en la que se ubican los errores. Este mismo rango es el que hay que especificar tanto en la pequeña ventana que aparece a la derecha de “+” como en la que aparece a la derecha de “-”. Es decir, el objetivo es que el mismo va-

lor se sume (por encima de) a la media y se le reste (por debajo de). Pulsamos sobre el botoncillo con una flecha roja como se indica en [3] y esto despliega una nueva ventana que permite seleccionar el rango del libro de trabajo, tal y como se indica en la Figura 6.

Seleccionamos entonces con el ratón el rango desde B16 hasta D16 y el programa se encarga de traducirlo en las coordenadas adecuadas. Una vez que el rango ha entrado adecuadamente en la subventana, pulsamos sobre el botón de la flecha roja —está resaltado con una flecha en la figura—; lo que nos devuelve a la ventana de definición básica. Todo lo que nos queda en el último paso [4] es hacer exactamente lo mismo pero en la ventana “-”. Una vez que están definidos los dos rangos se puede dar por finalizado el proceso y pulsar el botón [Aceptar] en la ventana como la de la Figura 5, indicado como paso [5].

Esto nos devuelve al libro de trabajo del que partimos con el gráfico correspondiente ubicado en el mismo. El lector puede moverlo un poco y realizarle algunos cambios de formato. En nuestro caso se ha cambiado el color de los histogramas y se ha añadido el título de los dos ejes de coordenadas. El resultado es el gráfico que se muestra en la parte superior (aparatado A) de la Figura 7.

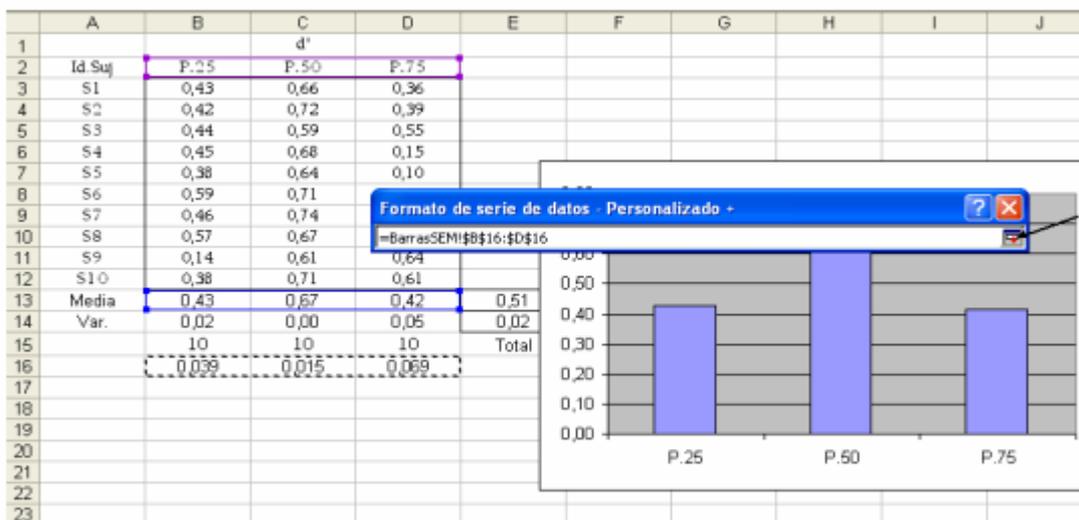
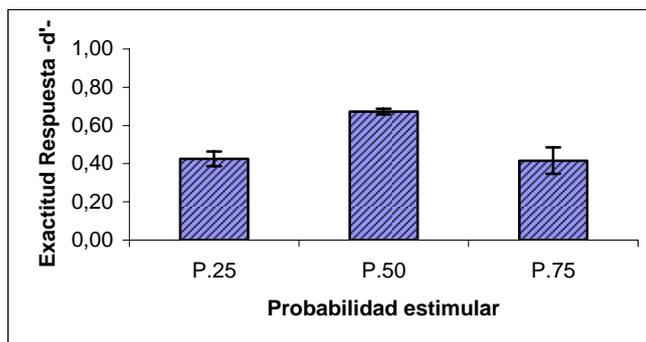


Figura 6: Agregación de barras Error a un gráfico de resultados- Paso 4.

**A) Basado en Errores individuales**



**B) Basado en Intervalos Confidenciales**

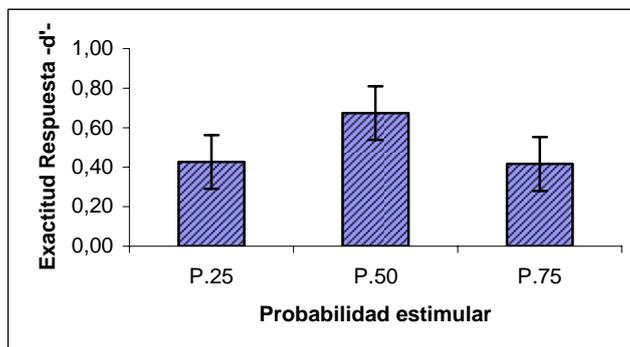


Figura 7: Comparación de dos tipos de barras Error en un gráfico de resultados.

El ejemplo hasta aquí se ha basado en la representación de los Errores SEM individuales para cada uno de los tres niveles de manipulación. Lo que es adecuado únicamente desde el punto de vista descriptivo o bien desde el punto de vista inferencial, cuando se tiene sospechas del incumplimiento

del supuesto de Homocedasticidad –recordar que este implica la homogeneidad de varianzas–.

No obstante, tal y como se argumentó más arriba para este mismo ejemplo, desde el punto de vista inferencial, lo más adecuado sería posiblemente la representación basada

en los Intervalos Confidenciales. En concreto, el error máximo que sumar y restar a cada una de las tres medias se estimó en 0,136. Para actualizar la representación gráfica bastaría con sustituir los valores de error del libro de trabajo empleado para la misma, en las casillas B16, C16 y D16, por

el valor 0,136 en las tres. Así se obtendría el gráfico que aparece en la parte inferior (aparatado B) de la Figura 7, más adecuado que el precedente desde el punto de vista inferencial.

## Referencias

- American Psychological Association (APA) (1994). *Publication manual of the American Psychological Association* (4th ed.). Washington, D. C.: Author.
- Behrens, J.T. (1997). Principles and procedures of exploratory data analysis. *Psychological Methods*, 2 (2), 131-160.
- Estes, W.K. (1991). *Statistical models in behavioral research*. Hillsdale, NJ: Erlbaum.
- Estes, W.K. (1997). On the communication of information by displays of standard errors and confidence intervals. *Psychonomic Bulletin and Review*, 4, 330-341.
- Hallam, M. y Rosenthal, R. (2000). Interpreting and reporting results. En H.E.A. Tinsley, y S.D. Brown (Eds.). *Handbook of applied multivariate statistics and mathematical modelling* (pp. 125-149). San Diego: Academic Press.
- Judd, C.M., McClelland, G.H. y Culhane, S.E. (1995). Data Analysis: Continuing Issues in the Everyday Analysis of Psychological Data. *Annual Review of Psychology*, 46, 433-465.
- Loftus, G.R. y Masson, M.E.J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, 1 (4), 476-490.
- Ramos, M.M.; Catena, A. y Trujillo, H.M. (2004). *Manual de métodos y técnicas de investigación en Ciencias del Comportamiento*. Madrid: Biblioteca Nueva.
- Rosenthal, R. y Rosnow, R.L. (1985). *Contrast analysis: Focused comparisons in the analysis of variance*. Cambridge, England: Cambridge University Press.
- Smith, A.F. y Prentice, D.A. (1993). *Exploratory Data Analysis*. En G. Keren y Ch. Lewis (Eds.). *A Handbook for Data Analysis in the Behavioral Sciences 2. Statistical issues* (pp. 349-390). Hove: Lawrence Erlbaum Associates.
- Wilkinson, L. and the Task Force on Statistical Inference (1999). Statistical methods in psychology journals: Guidelines and Explanations. *American Psychologist*, 54 (8), 594-604. Que se puede obtener de manera gratuita en la dirección: <http://www.apa.org/journals/amp/amp548594.html>

(Artículo recibido: 16-7-04, aceptado: 10-5-05)