

Principios básicos en estadística para neuropsicólogos clínicos e investigadores: utilidad práctica e interpretación de análisis de variables continuas

Javier Oltra-Cucarella¹, Ph.D., Diego Rivera², Ph.D.,
Juan Carlos Arango-Lasprilla^{3,4,5}, Ph.D.

Resumen

Objetivo. La calidad de la evaluación neuropsicológica y el seguimiento de programas de rehabilitación se apoyan en gran medida en tests, escalas e inventarios neuropsicológicos. Sin embargo, la interpretación de las pruebas estadísticas utilizadas en investigación neuropsicológica puede ser difícil para muchos neuropsicólogos. El objetivo de este trabajo es proporcionar un texto científico de consulta para los neuropsicólogos interesados en la realización de estudios de investigación y en la interpretación de resultados científicos.

Método. Se realiza una revisión y explicación detallada de diferentes análisis estadísticos relacionados con variables continuas, utilizando ejemplos de estudios científicos relacionados con la neuropsicología.

Resultados. Con el presente trabajo los clínicos y los investigadores conocerán de manera pormenorizada los fundamentos y la interpretación de los análisis estadísticos relacionados con diferencias de medias, análisis de varianza, correlaciones y regresión lineal.

Conclusión. El presente trabajo pretende ser una herramienta de consulta rápida para clínicos e investigadores interesados en interpretar y realizar investigaciones científicas con personas con daño neurológico. Los profesionales interesados en profundizar en el estudio de cada una de las técnicas deberán acceder a textos especializados donde se explique cada técnica de manera más pormenorizada.

Palabras clave:

Análisis de datos; estadística; investigación; modelo lineal general; neuropsicología

Autor de correspondencia:

Dr. Juan Carlos Arango-Lasprilla
BioCruces Bizkaia Health Research Institute
Cruces University Hospital
48903 Barakaldo, Bizkaia, Spain.

¹ Departamento de Psicología de la Salud, Universidad Miguel Hernández de Elche, Avda de la Universidad s/n Edificio Altamira, 03202 Elche (Alicante, España).

² Departamento de Ciencias de la Salud, Universidad Pública de Navarra. Campus Arrosadia, s/n. 31006. Pamplona, España.

³ BioCruces Bizkaia Health Research Institute, Cruces University Hospital Barakaldo, Bizkaia, Spain. 48903.

⁴ IKERBASQUE. Basque Foundation for Science, Bilbao, Spain. 48013.

⁵ Department of Cell Biology and Histology, University of the Basque Country UPV/EHU, Leioa, Spain. 48940

INTRODUCCIÓN

La neuropsicología es un área de la psicología encargada de estudiar la relación entre el funcionamiento cerebral y el comportamiento humano, tanto en personas sanas como en personas que han sufrido algún tipo de daño cerebral¹⁻³. En la actualidad la labor del neuropsicólogo cobra cada vez más importancia no sólo por su capacidad para evaluar y diagnosticar a pacientes con daño cerebral, sino también por su papel activo y relevante en el proceso rehabilitador⁴.

La calidad de la evaluación neuropsicológica y el seguimiento de programas de rehabilitación se apoyan en gran medida en tests neuropsicológicos, escalas e inventarios. Estos deben cumplir con ciertas características tales como adecuada validez y fiabilidad, adaptación cultural y disposición de datos normativos⁵. Estas características son esenciales puesto que de ello dependerá la adecuada interpretación de la puntuación de una persona^{3,5}. Los datos normativos ajustados por características demográficas han llegado a ser de gran uso y de administración rutinaria en neuropsicología. En general se pueden encontrar normas corregidas por edad y nivel educativo, aunque en algunos casos también se corrige según el sexo. Actualmente se reconoce la importancia de generar, adicionalmente, normas según grupos definidos por raza/origen étnico⁵⁻⁷.

Sin embargo, la adecuada interpretación final de los datos recopilados para conseguir un diagnóstico clínico siempre depende del juicio clínico. Esto es, los tests neuropsicológicos, las escalas e inventarios, y las técnicas de observación son herramientas que sirven de ayuda para que el clínico pueda realizar su labor. Esto supone que, además de los conocimientos clínicos, los neuropsicólogos deben adquirir a su vez conocimientos en estadística para poder realizar de manera adecuada su labor. Por ejemplo, el conocimiento de la estadística permite una adecuada interpretación y equivalencia de las diferentes distribuciones y las puntuaciones derivadas, tales como las puntuaciones z , puntuaciones t , puntuaciones CI , entre otras, usadas para la calificación e interpretación de los tests

neuropsicológicos. De igual modo, es necesario saber cómo calcular las desviaciones estándar, puesto que determinados criterios diagnósticos de algunas patologías (p. ej. deterioro cognitivo leve, demencias) exigen que el rendimiento del paciente se sitúe determinadas desviaciones estándar por debajo de la media⁸. A su vez, el clínico debe tener conocimiento sobre la mayor probabilidad que se tiene de obtener puntuaciones bajas al aplicar una batería de pruebas más o menos larga. Esto es, cuantas más pruebas se administren, mayor probabilidad de obtener puntuaciones bajas, al margen de ser signo o no de deterioro de alguna función cognitiva⁹⁻¹¹.

Otro de los puntos clave en la labor como neuropsicólogo es saber si la intervención está dando resultados positivos. Esto es, el clínico necesita saber si su intervención está favoreciendo la recuperación del paciente, su mayor integración en los diferentes ámbitos de la vida diaria: familia, escuela, trabajo, etc. Actualmente existen herramientas estadísticas que permiten conocer el cambio en el rendimiento neuropsicológico del paciente¹², pero se requiere de conocimiento estadístico para su correcto uso.

Por otra parte, desde los años 30 del siglo pasado (1929), la psicología se ha caracterizado por ser una ciencia que realiza sus aportes científicos a partir de análisis estadísticos basados en grupos muestrales ($n > 1$) gracias a los aportes de Fischer¹³. Los conocimientos estadísticos permiten a los clínicos entender los avances que se están realizando bien en evaluación y diagnóstico, bien en rehabilitación, cuando acuden a revistas científicas. Esto es, les permitirá leer un artículo de manera crítica y determinar si es útil para su práctica clínica. Por ejemplo, conocer si la muestra con la que se ha realizado el estudio puede ser equivalente a la población a la que atiende, valorar si una prueba cumple con los requisitos mínimos de fiabilidad o validez, interpretar los valores de las pruebas estadísticas (valores t , F , r , etc.), dominar conceptos como tamaño del efecto, cambios estadísticamente significativos, cambios clínicamente significativos, etc.

Sin embargo, adquirir una alfabetización en estadística no siempre es una tarea fácil¹⁴, y a pesar de

que desde las universidades se admite su importancia para la formación del psicólogo y se incluye en el currículum formativo de los grados de psicología, no suele contar con la valoración positiva de los estudiantes^{15,16}. Esto puede deberse a que esta asignatura es impartida por estadísticos sin experiencia en el ámbito de la psicología o neuropsicología, por lo que el alumnado no ve una conexión entre el conocimiento estadístico y su uso para la práctica clínica. Además, por lo general se acude a textos complejos con terminología altamente técnica, lo que en muchas ocasiones desmotiva y/o no permiten una adecuada adquisición de las herramientas estadísticas por parte del psicólogo.

El objetivo del presente trabajo es presentar una guía sencilla de algunas de las pruebas estadísticas más usadas en investigación neuropsicológica. La finalidad del presente texto es servir de consulta rápida por parte de los clínicos y los investigadores interesados en analizar e interpretar datos derivados de estudios científicos. Para un conocimiento más exhaustivo de las técnicas estadísticas se remite al lector a los manuales de referencia citados a lo largo del texto.

ANÁLISIS ESTADÍSTICOS Y NEUROPSICOLOGÍA

Comparación de medias: dos grupos

En ocasiones los neuropsicólogos están interesados en conocer si las puntuaciones medias que obtiene un grupo de participantes en una variable (M_1) son diferentes de las puntuaciones medias que obtiene otro grupo de participantes en esa misma variable (M_2). La variable bajo estudio se denomina variable dependiente (VD), mientras que los grupos de comparación son la variable independiente (VI). Mediante una prueba de comparación de medias se resta la media de un grupo de la media del otro grupo, y se divide el resultado entre el error estándar de esa diferencia (esto es, se **estandariza**). Posteriormente se compara la diferencia estandarizada con una tabla de frecuencias de la que se obtiene el valor de probabilidad (o valor de p). Estadísticamente, el valor de p asociado a la diferencia estandarizada indica la probabilidad de encontrar una diferencia entre medias igual o más extrema que la

encontrada si la hipótesis nula (de $\mu_2 - \mu_1 = 0$ en la población) es cierta¹⁷.

Sin embargo, el estadístico que está tomando mayor importancia frente al valor de p es el tamaño del efecto, el cual indica la magnitud de las diferencias con independencia de la significación estadística. El estadístico del tamaño del efecto más utilizado para la diferencia de medias es la d de Cohen¹⁸ el cual indica cuántas desviaciones típicas se desvía la media de un grupo respecto de la media del otro grupo. No obstante, se han calculado estimadores insesgados para muestras pequeñas¹⁹. Siguiendo las guías de Cohen, las diferencias de medias se interpretan como pequeñas, moderadas o grandes cuando el valor de d es igual o superior a 0.20, 0.50 y 0.80 respectivamente. Como indicador adicional, se puede conocer el porcentaje de solapamiento entre las distribuciones de puntuaciones en las dos muestras a partir del valor de la d de Cohen²⁰.

La importancia del tamaño del efecto es que indica si es interesante tomar en consideración la asociación entre variables. Por ejemplo, un tamaño del efecto demasiado grande podría indicar que esa variable diferencia suficientemente entre grupos y no es necesario incluirla como variable de estudio (p.ej., la diferencia de peso entre niños de 4 años y adultos de 30). Por su parte, un tamaño del efecto demasiado pequeño, o incluso insignificante, indica que la distribución de frecuencias entre los grupos se solapa tanto que no es útil para diferenciar entre los grupos (p.ej., la diferencia de peso entre personas de 20 años y personas de 22 años).

Como ejemplo, Garcia-Diaz y sus colaboradores²¹ administraron una tarea de velocidad de procesamiento de la información a una muestra de 20 personas mayores cognitivamente sanas [Media (M) = 46.90, Desviación Típica (DT) = 10.26] y a una muestra de 36 personas con Enfermedad de Parkinson (M = 32.60, DT = 22.04) y deterioro cognitivo leve. La VD es el número de ítems correctos tras 90 segundos. La prueba t para muestras independientes indica que la diferencia en las medias entre los dos grupos es estadísticamente significativa ($t = 2.73$, $p = 0.008$), lo que sugiere que las

personas con Enfermedad de Parkinson y deterioro cognitivo leve identifican menos ítems en la misma prueba durante el mismo periodo de tiempo. La conclusión principal de este análisis es que las personas con Enfermedad de Parkinson y deterioro cognitivo leve presentan un enlentecimiento en la velocidad de procesamiento.

El estadístico d de Cohen para estos estadísticos es igual a 0.83, lo que indica un tamaño del efecto de moderado a grande. Asimismo, una $d = 0.83$ se asocia con un porcentaje de solapamiento de alrededor del 50%, lo que indica que aproximadamente el 50% de las puntuaciones en el grupo con Enfermedad de Parkinson está por debajo de la puntuación más baja observada en el grupo cognitivamente sano.

Es importante saber que aumentando los tamaños muestrales se reduce el error de medida y, por tanto, la desviación típica, lo cual tiene un impacto en el tamaño del efecto. Aunque la diferencia puede ir en cualquier dirección, el aumento del tamaño muestral suele estar asociado a una reducción del tamaño del efecto, lo que indica que los estadísticos del tamaño del efecto asociados a un valor de p menor de 0.05 y obtenidos en muestras pequeñas no son fiables, por lo que deben interpretarse con mucha cautela²².

Comparación de medias: varios grupos

Cuando lo que se pretende es comparar las puntuaciones en una misma VD entre más de dos grupos, la prueba utilizada es el análisis de varianza (ANOVA) de un factor. En el ANOVA se cambia el término VI por el término *factor*, siendo el factor la variable relacionada con los grupos. En el ANOVA, la hipótesis nula es que todas las medias son iguales entre ellas (i.e., $\mu = \mu_2 = \mu_k$, donde k es el número de grupos).

El ANOVA se basa en dos estimaciones de varianza²³. Por un lado, se analiza la varianza dentro de cada grupo, esto es, si las puntuaciones dentro de cada grupo son similares o presentan una marcada dispersión. Esta varianza intra-grupo se considera varianza aleatoria o varianza de error. Por otro lado,

se analiza la diferencia entre medias de los grupos, lo que se considera un reflejo de la diferencia entre los grupos más el error. Si las dos estimaciones de varianza no difieren considerablemente, es decir, si la varianza entre-grupos es similar a la varianza intra-grupos, se concluye que todas las medias provienen de la misma distribución de medias y que cualquier pequeña diferencia se debe al error de muestreo. Por el contrario, si las medias de los grupos difieren más de lo esperable, se puede concluir que esas medias provienen de diferentes distribuciones y se rechaza la hipótesis nula de igualdad de medias²³. Sin embargo, el ANOVA que indica que las medias no son iguales entre los grupos no indica entre qué grupos existen diferencias. Para ello es necesario realizar pruebas adicionales.

En cuanto al tamaño del efecto relacionado con el ANOVA, el más utilizado es η^2 parcial (eta cuadrado parcial), que indica la varianza en la VD atribuible al efecto de interés más el error²³. El tamaño del efecto relacionado con η^2 parcial para cada variable independiente es de 0.010, 0.058 y 0.137 para un tamaño del efecto pequeño, moderado y grande respectivamente²⁴.

Como ejemplo, Smith, Dawes, Smith y Pal²⁵ compararon el procesamiento auditivo entre un grupo de 40 niños y adolescentes con epilepsia rolándica (G1), un grupo de 32 de los hermanos de los miembros del G1 (G2) y un grupo de 99 de los padres de los participantes del G1 (G3). Mediante escucha dicótica se presentaba una palabra en cada oído de manera simultánea, y se les solicitaba que repitieran la palabra que se les presentaba en uno u otro oído. Las puntuaciones (y su DT) de los grupos G1, G2 y G3 fue de 7.3 (3.8), 7.6 (3.7) y 8.9 (2.8) respectivamente. Los datos indican que la variabilidad dentro de los grupos es ligeramente inferior en el G3, y que las medias son similares entre los grupos G1 y G2. El ANOVA indica que la variabilidad entre las medias es superior a la esperable según la variabilidad debida al error de muestreo ($F = 4.3$, $p = 0.015$), con lo que se concluye que existen diferencias entre los grupos en la tarea de memoria. El ANOVA no permite conocer entre qué grupos ocurren las diferencias, pero pruebas post-hoc mostraron que los grupos G1 y G2 obtuvieron puntuaciones significativamente

más bajas que el G3, sin diferencias entre los grupos G1 y G2. El estadístico η^2 parcial es de 0.38, lo que indica que el 38% de la varianza en la puntuación del test de escucha dicótica se asocia con la pertenencia a uno de los grupos.

Extensión del Análisis de Varianza: ANOVA de dos factores

Una de las particularidades del ANOVA es que es una técnica versátil que permite ampliar los análisis considerablemente. La primera extensión del ANOVA es la introducción de uno o más factores en el análisis. Así, el ANOVA permite analizar si las diferencias en la variable dependiente se deben a un factor, a otro factor o la interacción entre ambos. En el caso más simple se incluyen dos factores, lo que permite conocer dos efectos simples y un efecto de interacción. Los efectos simples indican si las diferencias en la VD se deben a los efectos de cada uno de los factores considerados de manera aislada. La interacción, por el contrario, indica si las diferencias entre los diferentes niveles de un factor se repiten en los diferentes niveles del otro factor²³.

Como ejemplo, Testa, Malec, Moessner y Brown²⁶ analizaron la relación entre la edad y el tipo de daño en la independencia en tareas relacionadas con el trabajo. Para ello, reclutaron a 195 personas con traumatismo craneoencefálico (TCE) y 82 personas con daño físico de tipo ortopédico. Los participantes de ambos grupos fueron, a su vez, separados en dos grupos de edad: de 18 a 49 años, y de 50 a 89 años. La VD en este estudio fue la puntuación en el cuestionario *Vocational Independence Scale*. Este diseño incluye dos factores: el factor **Grupo** (F1) y el factor **Edad** (F2). El análisis del factor simple F1 permite conocer si el rendimiento difiere, de manera general, entre personas con TCE y personas con daño ortopédico. Por otro lado, el análisis del factor simple F2 permite conocer si el rendimiento difiere, de manera general, en pacientes más mayores en comparación con pacientes más jóvenes.

Lo más interesante, sin embargo, es la interacción entre factores, que en este caso mostraría si

uno de los grupos (F1) obtiene una puntuación diferente en función de la edad (F2). Los análisis mostraron un efecto principal del grupo ($F = 20.30, p < 0.001$) y un efecto principal de la edad ($F = 8.02, p = 0.005$). También se encontró una interacción significativa ($F = 3.95, p = 0.048$), indicando que las personas mayores que habían sufrido un TCE presentaban la menor independencia vocacional. Esta independencia vocacional era menor que la encontrada en las personas mayores con daño ortopédico y menor que las personas más jóvenes con independencia del daño que hubieran sufrido.

Análisis de covarianza

La segunda ampliación del ANOVA es que permite incluir el efecto de una tercera variable, en este caso continua, en el análisis de la relación entre factores y la VD. Cuando se incluye una variable continua (denominada **covariable**) en un análisis de varianza, la nomenclatura cambia a Análisis de Covarianza (ANCOVA). En el ANCOVA, se analiza la relación entre factores y VD después de ajustar los valores de la VD en función de las diferencias en la covariable introducida (y que correlaciona con la VD)²³. Mediante el ANCOVA, se ajustan las medias de los grupos en la VD al valor que tomaría esta VD si todos los sujetos puntuaran igual en la covariable, por lo que las diferencias existentes se deben únicamente a los factores introducidos en el análisis (más otros desconocidos que no se han podido controlar). Esto indica que los efectos de la covariable en el ANCOVA se basan en una regresión lineal (ver más abajo).

Como ejemplo, Wobrock y sus colaboradores²⁷ compararon el rendimiento en tareas de funciones ejecutivas entre un grupo de 11 personas con primer episodio de esquizofrenia, 13 personas con esquizofrenia crónica, 18 personas con trastorno bipolar y 23 controles sanos. La VD es el número de intentos administrados en el *Wisconsin Card Sorting Test* (WCST). Para realizar el ANCOVA compararon la puntuación en el WCST entre los grupos controlando el sexo, la edad y el nivel educativo. En este ANCOVA introdujeron dos factores categóricos (grupo y sexo) y dos covariables (edad y nivel educativo). Los resultados mostraron que,

controlando por edad, sexo y nivel educativo, existían diferencias significativas en el WCST, con el grupo de controles sanos realizando la tarea en un menor número de intentos ($M = 88.45$, $DT = 16.77$) en comparación con las personas con primer episodio de esquizofrenia ($M = 110.91$, $DT = 18.56$), las personas con esquizofrenia crónica ($M = 104.09$, $DT = 18.42$) y las personas con trastorno bipolar ($M = 103.25$, $DT = 20.92$). Los autores no indicaron el efecto de la edad o el nivel educativo sobre la VD, por lo que no es posible conocer si las puntuaciones en el WCST disminuyeron en función de la edad o los años de escolaridad.

ANOVA de medidas repetidas

La tercera ampliación del ANOVA, y quizá de las más importantes para interpretar estudios sobre intervenciones cognitivas en personas con daño neurológico, es el análisis de una VD en varios momentos temporales (intra-sujetos), manteniendo la posibilidad de incluir tanto factores entre-sujetos como covariables. En este caso, denominado ANOVA de medidas repetidas (o ANOVA mixto si incluye factores entre-sujetos y factores intra-sujetos), se analizan tanto los efectos de los factores simples como la interacción entre factores, como se describe en la sección dedicada al ANOVA de dos factores. La diferencia fundamental entre el ANOVA de dos factores y el ANOVA de medidas repetidas es la independencia de los datos. En el ANOVA de dos factores entre-sujetos, cada factor incluye datos de dos grupos diferentes y, por tanto, independientes²³. En el caso del ANOVA de medidas repetidas, el factor intra-sujetos incluye valores obtenidos en los mismos sujetos en varios momentos diferentes (p.ej., antes y después de la intervención), por lo que los datos están relacionados entre sí y no son independientes.

Como ejemplo, De Simone, Perri, Fadda, Calta-girone y Carlesimo²⁸ analizaron el rendimiento de 80 personas con Deterioro Cognitivo Leve (DCL) y 62 controles sanos en una tarea de memoria verbal. El grupo DCL fue, a su vez, dividido en dos grupos: un grupo de 39 personas con DCL que progresaron a Enfermedad de Alzheimer y un grupo de 41 personas con DCL que se mantuvieron

estables en un periodo de tres años. Este diseño incluyó una variable entre-sujetos con dos niveles (factor *Grupo*, F_1) y una variable intra-sujetos (factor *Tiempo*, F_2) con dos niveles: puntuaciones en la última serie de recuerdo inmediato vs. puntuaciones en el recuerdo diferido.

El ANOVA de medidas repetidas mostró un efecto principal del factor grupo ($F = 79.03$, $p < 0.001$), un efecto principal del factor tiempo ($F = 330.8$, $p < 0.001$), y una interacción significativa ($F = 9.33$, $p < .001$). En cuanto al factor grupo, los controles sanos obtuvieron en global puntuaciones superiores a los dos grupos con DCL, mientras que las personas con DCL que progresaron a Enfermedad de Alzheimer obtuvieron puntuaciones en global significativamente más bajas que las personas con DCL que se mantuvieron estables. En cuanto al factor tiempo, las puntuaciones en la última serie del recuerdo inmediato fueron, en general y contando con todos los grupos, mayores que las puntuaciones en el recuerdo diferido.

Lo más importante en este diseño, sin embargo, es la interacción entre factores. La interacción entre grupo y tiempo mostró que las personas en los dos grupos con DCL mostraron una pérdida entre la última serie del recuerdo inmediato y la serie de recuerdo diferido similar ($p = 0.07$), significativamente más marcada que la encontrada en el grupo de controles sanos ($ps < 0.001$). Por lo tanto, la interacción significativa entre factores muestra que lo que ocurre en alguno de los grupos de un factor no ocurre en alguno de los grupos del otro factor. En un estudio de intervención, se escogen dos grupos de personas con patología neurológica. A uno de los grupos se le realiza una intervención cognitiva, mientras que el otro grupo no la recibe, y ambos grupos son evaluados con las mismas pruebas en los mismos momentos temporales (antes y después de la intervención en el grupo experimental). El investigador está interesado en encontrar una interacción significativa entre los factores grupo y tiempo, de manera que las puntuaciones en un grupo antes y después sean diferentes a las del otro grupo. Si el grupo sin intervención se mantiene estable, se espera que el grupo con intervención mejore; si el grupo sin intervención empeora, se

espera que el grupo con intervención mejore o se mantenga estable.

Análisis de asociación: Correlaciones

En ocasiones los investigadores están interesados en conocer si los cambios en una variable continua se relacionan con cambios en otra variable continua, dentro de un grupo. En este caso, los análisis de correlaciones bivariadas (dos variables) permiten conocer si el aumento o disminución en los valores de una variable se asocian con un aumento o una disminución en los valores de otra variable²³. Las correlaciones bivariadas más utilizadas en investigación son la *r* de Pearson, para variables que se distribuyen normalmente, y la *Rho* de Spearman para variables que no se distribuyen normalmente.

El estadístico de correlación puede tomar valores entre ± 1 . Los valores positivos indican que el aumento en los valores de una variable se asocia con el aumento en los valores de la otra variable. Los valores negativos de *r* indican que los valores de una variable disminuyen conforme aumentan los valores de la otra variable. Una correlación no significativa o con valores cercanos a 0 indica que no existe asociación entre variables y, por tanto, los cambios en una variable no tienen relación con los cambios en la otra variable.

El tamaño del efecto indica una asociación pequeña, moderada o grande para valores $r = 0.10$, $r = 0.30$ y $r = 0.50$ respectivamente¹⁸. Por otra parte, el cuadrado de la correlación (r^2) indica la proporción de varianza en una variable explicada por la otra variable, y puede tomar valores entre 0 y 1. Cuanto mayor es la varianza explicada más cercano es r^2 a 1, mayor es la asociación entre variables y mejor pueden predecirse los valores de una variable en función de las puntuaciones en la otra variable²³. Es muy importante tener presente que, como en el caso de las diferencias de medias, los análisis de correlación indican asociación, pero en ningún caso causalidad.

Como ejemplo, Borghesani y sus colaboradores²⁹ analizaron la asociación entre la edad y la velocidad de procesamiento con otras variables cognitivas

relacionadas con la memoria, la fluencia verbal, el razonamiento y la flexibilidad cognitiva en 165 personas mayores cognitivamente sanas. Sus resultados mostraron que la edad correlacionó de manera significativa con la velocidad de procesamiento ($r = -0.36$), la memoria ($r = -0.24$), la fluencia verbal ($r = -0.15$), el razonamiento ($r = -0.43$) y la flexibilidad cognitiva ($r = -0.40$). El signo negativo de las correlaciones indica que las puntuaciones en los diferentes tests cognitivos disminuía conforme aumentaba la edad. En cuanto a la velocidad de procesamiento, los resultados mostraron una asociación significativa con la memoria ($r = 0.47$), la fluencia verbal ($r = 0.62$), el razonamiento ($r = 0.64$) y la flexibilidad cognitiva ($r = 0.34$). El signo positivo de las correlaciones indica que las puntuaciones en los diferentes tests aumentaban conforme aumentaban las puntuaciones en tests de velocidad de procesamiento. Sin embargo, pese a la significación estadística, las medidas de tamaño del efecto indican que la relación entre variables es generalmente baja, explicando porcentajes de varianza entre el 2.25% y el 22%. Las asociaciones más importantes se encontraron entre la velocidad de procesamiento y la fluencia verbal ($r^2 = 38.4\%$), por un lado, y la velocidad de procesamiento y el razonamiento por otro ($r^2 = 41\%$).

Análisis de asociación: regresión lineal

Los análisis de correlación indican la asociación entre dos variables continuas, esto es, cómo los cambios en una variable afectan a otra variable. Si dos variables están relacionadas se puede predecir, con cierto grado de confianza, los valores de una variable en función de los valores de la otra variable. Este análisis de predicción se realiza a través de análisis de regresión³⁰. Cuando la variable dependiente que se quiere predecir, denominada **criterio**, es una variable continua el análisis adecuado es la regresión lineal. Las variables independientes utilizadas para predecir la variable criterio, denominadas **predictores**, pueden ser variables categóricas y variables continuas.

Los análisis de regresión se utilizan para responder a dos cuestiones generales de investigación³⁰. En primer lugar, aportan información acerca de la asociación de variables. Si se utiliza una variable

dicotómica como predictor de una variable continua, el coeficiente asociado al predictor es igual al que se obtendría con un análisis de diferencia de medias. En segundo lugar, los coeficientes obtenidos en los análisis de regresión son útiles para predecir los valores de la variable criterio a partir de los valores del conjunto de variables predictoras.

Dos son los estadísticos más relevantes para los investigadores en un análisis de regresión lineal. Los estadísticos relacionados con cada predictor se denominan **coeficientes**, e indican el cambio medio en la variable criterio por cada aumento de una unidad en la variable predictora. Cuando hay dos o más predictores, el coeficiente de cada predictor indica el cambio medio en la variable criterio por cada aumento de una unidad en el predictor manteniendo el resto de predictores constantes. El segundo estadístico relevante es R^2 , que indica el porcentaje de varianza de la variable criterio que se explica por la combinación de predictores. R^2 puede tomar valores entre 0 y 1. Los valores más cercanos a 1 indican una relación absoluta entre predictores y criterio, mientras valores cercanos a 0 indican que no existe asociación entre la variable criterio y los predictores. R^2 es útil para calcular el tamaño del efecto del modelo, con valores de 0.02, 0.13 y 0.26 indicando un tamaño del efecto pequeño, moderado y grande respectivamente¹⁸.

Como ejemplo, Chang y sus colaboradores³¹ analizaron el tiempo de estancia hospitalaria en 330 pacientes con primer ictus isquémico. Con el fin de identificar variables relacionadas con el tiempo de estancia recogieron datos sobre el sexo del paciente, la edad, la gravedad del ictus, independencia funcional, tiempo pasado desde el ictus (<24h o >24h), comorbilidad (hipertensión, diabetes mellitus, hipercolesterolemia), tabaquismo, problemas cardíacos, subtipo de ictus (oclusión de pequeño vaso o no), y niveles de colesterol y triglicéridos. La gravedad del ictus se analizó mediante la escala *National Institutes of Health Stroke Scale* (NIHSS), con un rango de 0 a 38 y mayores puntuaciones indicando mayor gravedad. La independencia funcional se analizó mediante el Índice de Barthel Modificado (rango 0-20), en el que mayores puntuaciones indican mayor independencia

funcional. La variable dependiente en el análisis de regresión fue el tiempo de estancia (en valores logarítmicos), con el que los autores pretendían conseguir la normalidad de la distribución de los datos. Aunque no es el objetivo del presente trabajo, esta transformación indica que los autores malinterpretaron los supuestos del análisis de regresión. Según el Modelo Lineal General son los residuos, y no las variables, quienes deben distribuirse normalmente³².

Los resultados indicaron que el tiempo medio de estancia fue de 11 días (DT = 14). La mayoría de los pacientes (62%) sufrió un ictus por oclusión de pequeño vaso, seguido de aterosclerosis (25%) y cardioembolismo (12%). La mayoría de los pacientes (81%) ingresó dentro de las primeras 24 horas tras el ictus. De los 330 pacientes, 17 (5%) murieron durante la hospitalización aguda, 86 (26%) fueron trasladados al servicio de rehabilitación intensiva, y 227 (69%) fueron dados de alta y regresaron a su residencia habitual u otros servicios de atención médica.

Los datos del análisis de regresión se presentan en la tabla 3-1. El tiempo de estancia aumentó en hombres respecto a mujeres, en no-fumadores respecto a fumadores, en los casos con mayor gravedad, en casos con mayor dependencia funcional y en los casos con tipo de ictus diferente a la oclusión de pequeño vaso.

Sin embargo, lo relevante de los análisis de regresión es que no solo indican si existe una relación estadísticamente significativa, sino que aportan la posibilidad de conocer cuál es la asociación de cada predictor con la variable criterio controlando el resto de predictores. En el ejemplo de la tabla 1 se observa que los hombres presentan una estancia media 1.24 días superior que las mujeres cuando se controlan los demás predictores. Respecto a la historia de tabaquismo, la estancia media se reduce de media en 0.84 días en los fumadores respecto a los no-fumadores.

Además de la asociación entre predictores y variable criterio, los análisis de regresión permiten predecir los valores de la variable criterio a partir

Tabla 3-1. Regresión lineal sobre tiempo de estancia hospitalaria

Predictores	Coefficiente β	Error estándar	p
Sexo (0=mujer)	0.212	0.073	.004
Tabaquismo	-0.178	0.088	.043
NIHSS	0.060	0.018	.001
NIHSS cuadrático	-0.002	0.000	.001
MBI	-0.020	0.010	.042
Oclusión de pequeño vaso	-0.436	0.080	<.001
Constante (<i>Intercept</i>)	2.019	0.339	<.001

Error de estimación: 3.186. $R^2 = 36.9\%$. Significación del modelo <.001. La constante (o *intercept* en inglés), se interpreta como la puntuación media en la variable criterio cuando la puntuación en todas las variables predictoras es 0. Tomado de Chang et al., 2002.

de los diferentes valores de los predictores. En el caso del ejemplo de la tabla 3-1, se podría predecir la estancia hospitalaria con la fórmula:

$$\hat{Y}_i = \beta_1 * X_{i1} + \beta_2 * X_{i2} + \beta_j * X_{ij} + \varepsilon_{ij}$$

donde \hat{Y}_i es la puntuación predicha para el sujeto i , β_j es el coeficiente de la variable j , X_{ij} es la puntuación del sujeto i en la variable j , y ε_{ij} es el error. Utilizando los valores de la tabla 1, la predicción del tiempo de estancia hospitalaria para un varón fumador con una puntuación NIHSS=10, una puntuación en el Índice de Barthel=15, y diagnóstico de ictus por oclusión de pequeño vaso sería:

$$\hat{Y}_i = 2.019 + (0.212 * 1) - (0.178 * 1) + (0.060 * 10) - (0.002 * 10^2) - (0.020 * 15) - (0.436 * 1) = 1.717$$

Puesto que la variable criterio era el logaritmo del tiempo de estancia, hay que invertir el valor del logaritmo ($e^{1.717}$), lo cual indica que la estancia hospitalaria en este paciente ascendería a 5.57 días.

Las diferencias entre los valores reales y los valores predichos por el modelo de regresión en cada individuo se conocen como **residuos**, e indican si las predicciones del modelo son adecuadas. Cuanto mayores son los residuos, menor es la capacidad predictora del modelo y mayor es el error²³. Es importante que los investigadores tengan en cuenta que, en los análisis de regresión, el

análisis de residuos es esencial para comprobar la adecuación del modelo. Aunque el tamaño de la muestra no modifica los coeficientes asociados a cada predictor, las muestras pequeñas se asocian con mayores residuos y, por tanto, un mayor error de la predicción, lo cual puede llegar a invalidar la utilidad para predecir la variable criterio con esos predictores^{32,33}.

DISCUSIÓN

La estadística es una herramienta fundamental para el correcto desarrollo de la neuropsicología como especialidad clínica para la evaluación, el diagnóstico y la intervención en patologías neurológicas. Sin embargo, dominar conceptos estadísticos puede llegar a ser complejo para muchos neuropsicólogos¹⁴⁻¹⁶. Esta dificultad provoca que muchos profesionales no cuenten con las herramientas necesarias para interpretar correctamente los resultados de los estudios científicos a partir de los cuales fundamentar la elección de pruebas para la evaluación neuropsicológica, la interpretación de los resultados obtenidos en las pruebas seleccionadas, la elección de las técnicas de intervención más eficaces, o la interpretación de los resultados tras la aplicación de diferentes intervenciones cognitivas. El presente trabajo pretende proporcionar a los neuropsicólogos que desarrollan labores clínica e investigadora una herramienta de consulta rápida para la elección de las herramientas estadísticas más adecuadas en

función de la pregunta de investigación que pretenden responder.

Con ayuda del presente trabajo, los neuropsicólogos pueden consultar de manera rápida y sencilla los análisis estadísticos utilizados en trabajos de investigación de su interés, y conocer si las técnicas utilizadas son adecuadas. Asimismo, pueden identificar de manera rápida si los estudios consultados incluyen la información relevante asociada a los valores de significación estadística y tamaños del efecto. De esta manera, les facilita la interpretación de la importancia de los resultados desde un punto de vista estadístico (significación) y clínico (tamaño del efecto).

Es importante remarcar, no obstante, que el presente trabajo se centra en las pruebas estadísticas utilizadas más frecuentemente para el análisis de datos cuando la variable de estudio es continua. Futuros trabajos proporcionarán información para interpretar puntuaciones derivadas de los tests neuropsicológicos⁵, así como herramientas para entender e interpretar datos cuando la variable de estudio es dicotómica. Así, por ejemplo, en neuropsicología se estudia si las variables cognitivas se asocian a la probabilidad de desarrollar deterioro cognitivo (Sí/No)⁹, o también la utilidad de diferentes puntos de corte en los tests neuropsicológicos para identificar

un verdadero deterioro cognitivo a través de valores como la sensibilidad y la especificidad³⁴.

Como resumen, el presente trabajo pretende ser una herramienta de consulta rápida para clínicos e investigadores interesados en interpretar y realizar investigaciones científicas con personas con daño neurológico, de una manera sencilla y asociada a ejemplos prácticos fácilmente comprensibles. De esta manera, se puede convertir en un texto de consulta para estudiantes de psicología como para profesionales que desarrollen su labor en unidades de neuropsicología. No obstante, la consulta de este texto debe ser solo el primer paso para un conocimiento más profundo de las herramientas estadísticas. Una vez identificada la técnica estadística más adecuada para responder a una pregunta de investigación, los investigadores deben acudir a los manuales especializados para conocer más profundamente cada técnica^{23,30}. Uno de los principales mensajes de este trabajo es que, como se ha comentado en la sección dedicada a la regresión lineal, la correcta utilización de las pruebas estadísticas depende de que se cumplan unos supuestos estadísticos^{32,35}. Los clínicos e investigadores en neuropsicología deben conocer los supuestos de cada análisis estadístico antes de reclutar su muestra, realizar sus estudios o cuando interprete los datos derivados de su investigación.

Declaración de intereses:

Los autores declaran no tener conflicto de intereses

REFERENCIAS

1. Barth JT, Pliskin N, Axelrod B, et al. Introduction to the NAN 2001 Definition of a Clinical Neuropsychologist. NAN Policy and Planning Committee. *Arch Clin Neuropsychol Off J Natl Acad Neuropsychol*. 2003; 18(5):551-555.
2. Boller F, Grafman J, eds. *Handbook of Neuropsychology*. Amsterdam: Elsevier; 1997.
3. Lezak MD, Howieson DB, Loring DW. *Neuropsychological Assessment*. 4th ed. New York: Oxford University Press; 2004.
4. Goldstein LH, McNeil JE, eds. *Clinical Neuropsychology: A Practical Guide to Assessment and Management for Clinicians*. 2nd ed. Chichester, West Sussex ; Malden, MA: Wiley-Blackwell; 2013.

5. Strauss EH, Sherman EHS, Spreen O. *A Compendium of Neuropsychological Tests. Administration, Norms and Comments*. New York: Oxford University Press; 2006.
6. Fletcher-Janzen E, Strickland TL, Reynolds CR. *Handbook of Cross-Cultural Neuropsychology*. New York: Kluwer Academic; 2000.
7. Neil V. *Cross-Cultural Neuropsychological Assessment. Theory and Practice*. Mahwah, NY: Lawrence Erlbaum; 2000.
8. Petersen R, Smith G, Waring S, Ivnik R, Tangalos T, Kokmen E. Mild cognitive impairment: Clinical characterization and outcome. *Arch Neurol*. 1999; 56(3):303–308. doi:10.1001/archneur.56.3.303
9. Oltra-Cucarella J, Sánchez-SanSegundo M, Lipnicki DM, et al. Using the base rate of low scores helps to identify progression from amnesic MCI to AD. *J Am Geriatr Soc*. 2018; 66(7):1360-1366. doi:10.1111/jgs.15412
10. Oltra-Cucarella J, Sánchez-SanSegundo M, Rubio-Aparicio M, Arango-Lasprilla JC, Ferrer-Cascales R. The Association Between the Number of Neuropsychological Measures and the Base Rate of Low Scores. *Assessment*. July 2019;107319111986464. doi:10.1177/1073191119864646
11. Rivera D, Olabarrieta-Landa L, Brooks BL, et al. Multivariate Base Rates of Low Scores on Tests of Learning and Memory Among Latino Adult Populations. *J Int Neuropsychol Soc*. 2019; 25(08):834-844. doi:10.1017/S135561771900050X
12. Crawford JR, Garthwaite PH. Statistical Methods for Single-Case Studies in Neuropsychology: Comparing the Slope of a Patient's Regression Line with those of a Control Sample. *Cortex*. 2004; 40(3):533-548. doi:10.1016/S0010-9452(08)70145-X
13. Arnau Grass J. *Diseños experimentales en psicología y educación*. México: Trillas; 1990.
14. Cardona LZ. ¿Cómo contribuir a la alfabetización estadística? *Rev Virtual Univ Católica Norte*. 2011; 1(33):234-247.
15. Arango-Lasprilla JC, Olabarrieta-Landa L, Rivera D, Olivera Plaza SL, De Los Reyes Aragón CJ, Quijano MC. Situación actual de la neuropsicología en Colombia. In: Arango-Lasprilla JC, Rivera D, eds. *Neuropsicología En Colombia: Datos Normativos, Estado Actual y Retos a Futuro*. Manizales, Colombia: Editorial Universidad Autónoma de Manizales; 2015:21-46.
16. Fonseca-Aguilar P, Olabarrieta-Landa L, Rivera D, et al. Situación actual de la práctica profesional de la neuropsicología en México. *Psicol Desde El Caribe*. 2015; 32(3):343-364.
17. Cassidy SA, Dimova R, Giguère B, Spence JR, Stanley DJ. Failing Grade: 89% of Introduction-to-Psychology Textbooks That Define or Explain Statistical Significance Do So Incorrectly. *Adv Methods Pract Psychol Sci*. 2019;2(3):233-239. doi:10.1177/2515245919858072
18. Cohen J. A power primer. *Psychol Bull*. 1992; 112(1):155–159. doi:10.1037/0033-2909.112.1.155
19. Fritz CO, Morris PE, Richler JJ. Effect size estimates: Current use, calculations, and interpretation. *J Exp Psychol Gen*. 2012;141(1):2–18. doi:10.1037/a0024338
20. Zakzanis KK. Statistics to tell the truth, the whole truth, and nothing but the truth: formulae, illustrative numerical examples, and heuristic interpretation of effect size analyses for neuropsychological researchers. *Arch Clin Neuropsychol*. 2001; 16(7):653–67. doi:10.1016/S0887-6177(00)00076-7
21. Garcia-Diaz AI, Segura B, Baggio HC, et al. Structural Brain Correlations of Visuospatial and Visuosperceptual Tests in Parkinson's Disease. *J Int Neuropsychol Soc*. 2018; 24(01):33-44. doi:10.1017/S1355617717000583
22. Button KS, Ioannidis JP, Mokrysz C, et al. Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci*. 2013; 14(5):365–76. doi:10.1038/nrn3475
23. Tabachnick BG, Fidell LS. *Using Multivariate Statistics*. Sixth Ed. New Jersey: Pearson Education Inc; 2013.
24. Richardson JTE. Eta squared and partial eta squared as measures of effect size in educational research. *Educ Res Rev*. 2011; 6(2):135-147. doi:10.1016/j.edurev.2010.12.001
25. Smith AB, Dawes P, Smith S, Pal DK. A specific deficit of auditory processing in children with Rolandic Epilepsy and their relatives. *Epilepsy Behav*. 2017; 72:135-139. doi:10.1016/j.yebeh.2017.04.043

26. Testa JA, Malec JF, Moessner AM, Brown AW. Outcome After Traumatic Brain Injury: Effects of Aging on Recovery. *Arch Phys Med Rehabil.* 2005; 86(9):1815-1823. doi:10.1016/j.apmr.2005.03.010
27. Wobrock T, Ecker UKH, Scherk H, Schneider-Axmann T, Falkai P, Gruber O. Cognitive impairment of executive function as a core symptom of schizophrenia. *World J Biol Psychiatry.* 2009; 10(4-2):442-451. doi:10.1080/15622970701849986
28. De Simone MS, Perri R, Fadda L, Caltagirone C, Carlesimo GA. Predicting progression to Alzheimer's disease in subjects with amnesic mild cognitive impairment using performance on recall and recognition tests. *J Neurol.* 2019; 266(1):102-111. doi:10.1007/s00415-018-9108-0
29. Borghesani PR, Madhyastha TM, Aylward EH, et al. The association between higher order abilities, processing speed, and age are variably mediated by white matter integrity during typical aging. *Neuropsychologia.* 2013; 51(8):1435-1444. doi:10.1016/j.neuropsychologia.2013.03.005
30. Vittinghoff E, Glidden DV, Shiboski SC, McCulloch CE. *Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models.* New York: Springer; 2005.
31. Chang K-C, Tseng M-C, Weng H-H, Lin Y-H, Liou C-W, Tan T-Y. Prediction of Length of Stay of First-Ever Ischemic Stroke. *Stroke.* 2002; 33(11):2670-2674. doi:10.1161/01.STR.0000034396.68980.39
32. Williams MN, Gómez Grajales CA, Kurkiewicz D. Assumptions of Multiple Regression: Correcting Two Misconceptions. *Pract Assess Res Eval.* 2013; 18(11):1-14.
33. Osborne JW. Normality of residuals is a continuous variable, and does seem to influence the trustworthiness of confidence intervals: A response to, and appreciation of, Williams, Grajales, and Kurkiewicz (2013). *Pract Assess Res Eval.* 2013; 18(12):1-9.
34. De Jager CA, Hogervorst E, Combrinck M, Budge MM. Sensitivity and specificity of neuropsychological tests for mild cognitive impairment, vascular cognitive impairment and Alzheimer's disease. *Psychol Med.* 2003; 33(6):1039-1050. doi:10.1017/s0033291703008031
35. Cohen J, Cohen P, West SG, Aiken LS. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences.* 3rd ed. Mahwah, N.J: L. Erlbaum Associates; 2003.