ESTADÍSTICA: MEDICIÓN, DESCRIPCIÓN E INFERENCIA [1]

Dr. Enerio Rodríguez Arias
Universidad Autónoma de Santo Domingo
erodriguez27@uasd.edu.do

RESUMEN

Partiendo de la teoría representacional de la medición, y de las diferentes escalas de medida derivadas de dicha teoría, se ofrece una exposición sencilla de los conceptos fundamentales de la Estadística, tanto descriptiva como inferencia1. En ese sentido, se describen las medidas de tendencia central y de variabilidad apropiadas para cada escala de medida. Luego, se distingue entre la estimación de parámetros y la verificación de hipótesis como los instrumentos de la Estadística inferencia1. Se reflexiona sobre el establecimiento de un nivel de significación estadística, y se mencionan las pruebas estadísticas para las diferentes escalas de medida. Finalmente, se orienta al lector sobre la interpretación de los resultados de cualquier análisis estadístico.

Palabras clave: Estadística, medición, descripción, inferencia, medición, pruebas, significación.

La estadística es una ciencia de gran utilidad para la investigación empírica, sea en psicología o en cualquier otra ciencia. Generalmente es conocida como la ciencia de los grandes números, porque sus leyes y principios alcanzan su máxima validez en los grandes conjuntos de casos o sucesos. La Estadística se divide en dos partes: La estadística descriptiva y la estadística inferencial. La primera se utiliza para describir con el uso de números los principales rasgos o características de grupos de personas, cosas, o fenómenos. La segunda (la estadística inferencial) se utiliza para sacar conclusiones sobre una población o universo a partir del estudio de una muestra representativa de dicha población o universo. Antes de ver en detalle los conceptos fundamentales de cada una de las dos ramas de la Estadística, definiremos brevemente la teoría representacional de la medición y sus implicaciones estadísticas.

La teoría representacional de la medición sostiene que la medición, en el sentido más amplio, consiste en la asignación de números a objetos o fenómenos de acuerdo con reglas. El hecho de que los números puedan ser asignados bajo diferentes reglas conduce a diferentes clases de escalas y a diferentes clases de medición. El problema se reduce entonces a explicitar las diferentes reglas para la asignación de números a objetos o fenómenos, las propiedades matemáticas de las escalas resultantes, y finalmente las operaciones estadísticas aplicables a las mediciones hechas con cada tipo de escala (Stevens, 1946). En otros términos, para esta teoría, la medición es la correlación de números con entidades que no son números (Michell, 1993), de donde se sigue que no es lícito hacer con los números ninguna operación que no pueda realizarse con las entidades a las que los números representan.

La Estadística descriptiva dispone de un conjunto de formas de medir alguna característica o rasgo de un grupo de sujetos o fenómenos bajo estudio (Aron & Aron, 2001; Freeman, 1965; Urdan, 2005). Para empezar, están las llamadas medidas de tendencia central, que con un solo valor numérico nos permiten hacer la mejor

1- Trabajo inédito, 2009.

descripción numérica de una realidad determinada. Las medidas de tendencia central son: la moda, la mediana, la media aritmética o promedio y la media geométrica. La moda es la medida de tendencia central propia de las escalas nominales. Una escala nominal es la que se compone de un conjunto de categorías o clases, donde ninguna categoría es mayor en jerarquía a las demás; de manera que en una escala nominal todas las clases o categorías son equivalentes en su jerarquía, por ejemplo, la variable sexo es el mejor ejemplo de una escala nominal, pues es la escala nominal más sencilla que hay, ya que sólo tiene dos categorías o clases (masculino y femenino); otros ejemplos de escalas nominales son las variables e*stado civil, nacionalidad, preferencia religiosa, preferencia política*, etc., etc. La moda es la categoría o clase que tiene la mayor cantidad de casos, o como se suele decir, la mayor frecuencia. Por ejemplo, si clasificamos una población de 500 personas por la variable sexo y encontramos que 300 son mujeres y 200 son hombres, entonces podemos decir que en ese grupo la moda para la variable sexo está en la clase o categoría *femenina***.** En síntesis, en cualquier sistema de clasificación, la clase que contiene la mayor frecuencia es la moda; puede haber distribuciones bimodales o plurimodales, si no hay una sola clase con la mayor frecuencia, sino dos (bimodal) o más de dos (plurimodal).

La *mediana* es la medida de tendencia central propia de las escalas ordinales. Las escalas ordinales son escalas de rangos, las cuales incluyen una relación de *mayor que*, por ejemplo, el orden en que llegan los autos o los caballos de una carrera, o el orden en que finalizan los participantes en cualquier competencia. La mediana es el punto que divide en dos partes iguales un conjunto de personas o cosas ordenadas por rangos. Por ejemplo, si ordenamos a los miembros (hijos e hijas) de una familia por el orden de su nacimiento, el hijo o el punto por debajo del cual y por encima del cual queda la misma cantidad de hijos, ese representa la mediana; tal es el caso de una familia de once hijos en la que X es el sexto, de manera que X es la mediana en su familia, pues por encima de él en edad hay 5 hermanos y por debajo de él en edad hay también 5 hermanos. Si en vez de once hijos hubieran sido diez, entonces la mediana no hubiera estado representada por ningún hijo, sino por el punto que divide la serie en dos, es decir, el punto que separa al quinto hijo del sexto hijo, pues por encima y por debajo de ese punto hay cinco hijos. Es importante señalar que en una escala ordinal lo único que cuenta es el orden o rango en que se encuentra cada caso, de manera que la distancia a que se encuentra un caso de otro no importa para nada; por ejemplo, no interesa para nada en una escala ordinal el hecho de que el segundo caso esté muy cerca del primero mientras que el que ocupa el tercer puesto esté muy lejos del segundo; lo único que se tiene en cuenta en una escala ordinal es el orden o rango.

La *media aritmética o promedio* es la medida de tendencia central propia de las escalas de intervalo. Una escala de intervalo se caracteriza porque a partir de un punto cero arbitrario, la distancia que separa a los puntos vecinos es exactamente la misma en cualquier parte de la escala; por ejemplo, la distancia que hay del siete al ocho es exactamente la misma que la distancia que hay del dos al tres; es precisamente porque los intervalos entre los puntos vecinos de la escala son iguales por lo que la escala se llama *escala de intervalo*. Esa realidad es la que permite que podamos sumar los valores de una escala de intervalo y dividir el resultado de la suma entre el número de casos; por ejemplo, si sumamos las edades de los miembros de un grupo compuesto por 25 sujetos y luego dividimos el resultado de esa suma entre 25, esa operación nos dará la edad promedio de dicho grupo o, lo que es lo mismo, la media aritmética de las edades de los miembros de dicho grupo; lo mismo sucede cuando sumamos las calificaciones de todos los alumnos de un curso y dividimos el resultado de esa suma entre el número total de alumnos de dicho curso; obtenemos entonces la calificación promedio del grupo.

La media aritmética o promedio es muy sensible a los valores que se alejan mucho de la mayoría de los valores de una distribución; es decir, unos pocos casos que se desvían de la mayoría, sea hacia abajo o hacia arriba, arrastrarán la media aritmética o promedio en la dirección de la desviación señalada. Cuando esto último ocurre, la media aritmética o promedio no constituye una adecuada representación de una distribución de valores. En las distribuciones siguientes, se puede observar que la media aritmética del grupo A es más representativa de la distribución de valores que las medias aritméticas de los grupos B y C.

Grupo A: 20+19+18+17+16 = 90; M = 90÷5 = 18

Grupo B: 72+06+05+04+03 = 90; M = 90÷5 = 18

Grupo C: 40+25+20+04+01 = 90; M = 90÷5 = 18

Cuando los valores o datos de una variable dependiente (calificaciones en un examen, edades, estaturas, pesos, o ingresos de los miembros de un grupo) se distribuyen en la forma de una curva normal (la conocida campana de Gauss), la moda, la mediana y la media coinciden en el mismo punto. Cuando no coinciden, se dice que la distribución es asimétrica; en este último caso, si la media es más grande que la mediana, entonces la distribución es asimétricamente positiva; si en cambio, la media es más pequeña que la mediana, la distribución es asimétricamente negativa.

La media geométrica es la medida de tendencia central propia de las escalas de razón. Una escala de razón se caracteriza porque, además de suponer la existencia de un punto cero verdadero, consta de intervalos iguales a través de toda la escala. Todos los tipos de medidas estadísticas son aplicables a las escalas de razón, y únicamente con estas escalas se pueden hacer transformaciones logarítmicas. Las escalas de razón son raras en psicología, aunque no totalmente desconocidas; se usan en estudios experimentales de algunos fenómenos sensoriales. Muy frecuente es el uso de la media geométrica para calcular el promedio de varios porcentajes, la ganancia promedio de un consorcio integrado por varias empresas, o el interés promedio generado por una cuenta bancaria durante un determinado período de tiempo.

Además de las medidas de tendencia central, en la estadística descriptiva se utilizan medidas de variabilidad o de dispersión. Estas medidas sirven para determinar qué tan homogéneo o heterogéneo es un grupo que está siendo estudiado. Para las escalas nominales, la principal medida de variabilidad es la *Razón de Variación* (V), la cual se calcula dividiendo la cantidad de casos que caen fuera de la categoría modal entre el número total de casos y multiplicando el resultado por 100 para convertirlo en porcentaje. Por ejemplo, si el número total de casos en una investigación es de 250 y la frecuencia modal es de 150, hay 100 casos que están fuera de la categoría modal; si dividimos esos 100 casos entre los 250 casos nos dará 0.40, y si multiplicamos 0.40 por 100 nos da el 40%; ese 40% es el valor de V en el ejemplo analizado, al cual vamos a llamar el ejemplo (1). Supongamos que tenemos un grupo de 250 casos, y que la categoría modal tiene 200 casos, lo que quiere decir que hay 50 casos fuera de la categoría modal. Dividiendo esos 50 casos entre el total de 250 casos me dará 0.20 que multiplicado por 100 nos da el 20%, que es el valor de V en este segundo ejemplo. Como en el ejemplo (1) V es el 40% mientras que en el ejemplo (2) V es sólo el 20%, el grupo del ejemplo 1 es más heterogéneo que el grupo del ejemplo 2, o lo que es lo mismo, el grupo 2 es más homogéneo que el grupo 1.

Ej. 1: V=100/250 = 0.40 x 100 = 40%

Ej. 2: V= 50/250 = 0.20 x 100 = 20%

Hay varias formas de medir la variabilidad interna de un grupo cuando la variable dependiente en estudio es medida con una escala ordinal. El grupo completo puede dividirse en deciles (diez subgrupos de un diez por ciento cada subgrupo, desde los rangos más bajos hasta los más altos), en quintiles (cinco subgrupos de un 20 por ciento cada uno, desde los rangos más bajos hasta los más altos), en cuartiles (cuatro subgrupos de un 25 por ciento cada uno, desde los rangos más bajos hasta los más altos), y finalmente en percentiles (cien subgrupos de un 1 por ciento cada subgrupo, desde los rangos más bajos hasta los más altos). Los rangos pueden ser de edad, peso, ingreso, resultado de un test o de un examen, etc. Las más usadas medidas de variabilidad para escalas ordinales son la *Desviación Decil* y la *Desviación Semi-Intercuartil*. La desviación decil es la cantidad de rangos que caen entre el noveno decil y el primer decil; eso es lo mismo que decir, la cantidad de rangos que caen entre el percentil 90 y el percentil 10, porque el decil 9 y el percentil noventa coinciden y lo mismo pasa con el decil 1 y el percentil 10. La desviación semi-intercuartil es la diferencia entre el tercer cuartil y el primer cuartil dividida entre 2, que es lo mismo que la diferencia entre el percentil 75 y el percentil 25 dividida entre 2. Mientras mayor es la cantidad de rangos entre los cuartiles o percentiles previamente señalados, mayor es la variabilidad interna del grupo.

Pasemos ahora a las medidas de variabilidad cuando la variable dependiente es medida con una escala de intervalo. La más rústica y sencilla medida de variabilidad para las escalas de intervalo es la *amplitud total (the range)* que es la diferencia entre el valor más grande y el más pequeño de una distribución. La más importante y también la más usada medida de variabilidad es la desviación standard o desviación típica, que generalmente se simboliza con S. Se calcula a partir del grado en que cada puntuación individual se desvía de la media aritmética del grupo; cada desviación individual es elevada al cuadrado; luego se suman las desviaciones al cuadrado; el resultado de esa suma es dividido entre el número de sujetos del grupo; este resultado se conoce como varianza o desviación cuadrática media y se simboliza con una S elevada al cuadrado. Finalmente, a la varianza o desviación cuadrática media se le saca la raíz cuadrada: el resultado es la llamada desviación standard o desviación típica. La utilidad de la desviación típica es que en una distribución normal (campana de Gauss) hay tres desviaciones típicas por debajo de la media y tres desviaciones típicas por encima de la media. Entre -1S y +1S alrededor de la media cae el 68.26% de los casos de una distribución; entre -2S y +2S alrededor de la media cae el 95.44% de los casos; y entre -3S y +3S alrededor de la media cae el 99.74% de los casos de una distribución. El estadístico ruso Pafnuti L. Tchebycheff (1821-1894) demostró que para cualuier forma de distribución, entre -2S y +2S alrededor de la media cae por lo menos el 75% de los casos, y entre -3S y +3S cae por lo menos el 89% de los casos. (Vogt,1999).

Para determinar cuál de dos grupos es más variable o heterogéneo, no podemos hacerlo comparando simplemente sus desviaciones típicas; para ello es necesario calcular el *Coeficiente de Variación*, el cual se obtiene dividiendo la desviación típica de un grupo entre la media aritmética de ese mismo grupo y luego multiplicando el resultado por cien. Por ejemplo, el grupo A tiene una desviación típica de 5 y una media aritmética de 40, mientras que el grupo B tiene una desviación típica de 15 y una media aritmética de 160. El cálculo del Coeficiente de Variación (CV) que aparece a continuación revela que el grupo A es más variable o heterogéneo que el grupo B, a pesar de que A tiene una desviación típica más pequeña que B.

M: Símbolo de la media aritmética. Cuando se trata de la media aritmética de una población, no de una muestra, se usa como símbolo la letra m minúscula en griego, que se llama mu.

S: Símbolo de la desviación standard o típica. Cuando se trata de la desviación típica de una población, no de una muestra, se usa como símbolo la letra s minúscula en griego, que se llama sigma.

Grupo A: M=40; S=5

CV=S/M=5/40 =0.125x100=12.50%

Grupo B: M=160; S=15

CV=S/M=15/160 =0.0937x100=9.37%

Pasemos ahora a la otra rama de la Estadística, la llamada Estadística Inferencial. En la primera página de este trabajo dijimos que la Estadística inferencial se ocupa de sacar conclusiones sobre una población o universo a partir del estudio de una muestra representativa de dicha población o universo; es decir, se ocupa de hacer inferencias estadísticas. La necesidad de realizar inferencias estadísticas se debe al hecho de que por múltiples razones (poblaciones demasiado grandes y tiempo y recursos muy limitados) puede resultar impráctico o imposible estudiar a una población en su totalidad.

Se llama *estadística (statistic*) a cualquier medida estadística obtenida en una muestra, y *parámetro (parameter)* al valor de esa misma medida estadística en la población. Como señalamos antes, los símbolos para representar las estadísticas (medidas de muestra) son letras latinas, mientras que los símbolos para representar los parámetros (valores de la población, generalmente desconocidos) son letras griegas.

Existen dos tipos de inferencia estadística: La estimación de un parámetro a partir de una estadística, y la verificación o prueba de hipótesis sobre poblaciones a partir de los resultados obtenidos en muestras (Hopkins, Hopkins y Glass, 1997; Urdan, 2005). En el caso de la estimación de un parámetro, por ejemplo, podemos estimar (calculando lo que se llama un intervalo de confianza) la estatura promedio de la población masculina dominicana de más de veinte años de edad a partir de la estatura promedio de una muestra de 1,500 hombres dominicanos de más de veinte años de edad. Este es el tipo de inferencia que se utiliza en las encuestas de opinión, en las cuales se determina con un determinado nivel de confianza (generalmente de un 95%) y con un determinado margen de error (generalmente entre 2.5 y 3%) el grado en que los resultados obtenidos en una muestra pueden interpretarse como representativos de la población o universo bajo estudio. El segundo tipo de inferencia estadística es la verificación o prueba de hipótesis sobre parámetros poblacionales. Esta es la llamada estrategia fisheriana (porque fue creada por Ronald Fisher) de la prueba de la hipótesis nula. Para una descripción detallada de los diferentes pasos de la prueba de la hipótesis nula, véase Rodríguez (2005). Aquí nos referiremos únicamente al establecimiento de un nivel de significación estadística.

¿Cuán baja debe ser la probabilidad de un suceso antes de que estemos dispuestos a rechazar la posibilidad de que haya ocurrido? A fin de contestar esta pregunta, debemos considerar algo más que la simple probabilidad de que el suceso ocurra. Debemos considerar también las consecuencias de la decisión de actuar como si el suceso hubiera o no hubiera ocurrido. Un ejemplo algo imaginario podría ayudarnos a aclarar este punto. Supongamos que usted se encuentra frente a una caja de 100 pistolas, cinco de las cuales usted sabe que están cargadas y que se le permite tomar una de las cien. La probabilidad de que esta pistola esté cargada es 5/100, ó .05. ¿Actuaría usted como si esta pistola estuviera cargada o como si no lo estuviera? Al decidir si actúa como si el arma estuviera cargada, usted tomaría en cuenta no sólo la probabilidad de que esté cargada, sino también las consecuencias de actuar como si estuviera cargada. Si se le preguntara si está usted dispuesto a apuntar con el arma hacia su cabeza y apretar el gatillo, usted actuaría como si el arma estuviera cargada. Por el contrario, si se le preguntara si estaría usted dispuesto a usarla para defenderse en un duelo, usted actuaría como si el arma no estuviera cargada. Ambas decisiones son perfectamente racionales; aunque la probabilidad es la misma en ambos casos, las consecuencias no son las mismas. Siguiendo el mismo razonamiento, debemos considerar las consecuencias de actuar como si el resultado de una investigación fuera producido completamente por errores de azar y las consecuencias de actuar como si la variable independiente estuviera asociada con el resultado.

Una decisión errónea al usar la inferencia estadística para analizar los resultados de una investigación puede producir dos clases de error. Una clase de error es “ver demasiado en los datos”. Este es el error de concluir que los cambios en la variable dependiente están relacionados con la variable independiente, cuando, de hecho, el cambio en la variable dependiente fue producido completamente por variables de azar; ese es el error del investigador que se apresura a concluir que existe una relación entre las variables estudiadas, y es con mucha frecuencia uno de los errores más graves que puede cometer un investigador.

La otra clase de error es “no ver lo suficiente en los datos”. Este es el error de concluir que los cambios en la variable dependiente no guardan ninguna relación con la variable independiente, cuando, de hecho, hay una genuina relación entre la variable independiente y la variable dependiente. Ese es el error del investigador excesivamente cauteloso.

El punto que divide aquellas probabilidades que nos conducen a aceptar la posibilidad de que el cambio en la variable dependiente es debido completamente a error de azar, de aquellas que nos conducen a rechazar esta posibilidad es llamado nivel de significación. El nivel de significación determina, y equivale a la proporción de veces (cuando, de hecho, no hay ninguna relación entre la variable independiente y la variable dependiente) que el investigador puede esperar cometer el error de “ver demasiado en los datos”. Si decidimos trabajar con un nivel de significación de .05, estamos, en

Perspectivas Psicológicas/Santo Domingo (Rep. Dom.)
/Vols. 6-7/pp. 172-178/2010/ISSN 1992-5131

realidad, decidiendo considerar como efectos reales aquellos que pudieron haber sido producidos por azar cinco veces en cien. (Esto no quiere decir que un investigador, trabajando con un nivel de significación de .05, cometerá realmente el error de “ver demasiado en los datos” con una probabilidad de.05. La interpretación del nivel de significación como la probabilidad de cometer esta clase de error vale únicamente en aquellas condiciones donde realmente no haya ninguna relación entre las variables independiente y dependiente. Pero, en teoría, nunca realizaremos una investigación bajo tales condiciones, y nunca, por tanto, cometeremos este error con esa frecuencia.

No podemos ser tan precisos acerca de la probabilidad de cometer la otra clase de error (no ver una relación que realmente existe) porque este error se comete cuando un efecto constante está operando en unión al azar y la probabilidad de pasar por alto este efecto depende de la magnitud del mismo.

Podemos, adoptando un nivel de significación pequeño, reducir las posibilidades de cometer el error de “ver demasiado en los datos”. Por el contrario, adoptando un nivel de significación grande, podemos reducir las posibilidades de cometer el error de “no ver lo suficiente en los datos”.

La situación puede ser igualada a la de un joven cazador africano, el cual está pasando su primera noche, solo, en la selva y no puede dormir porque, al oír los ruidos de la selva, piensa que tigres se están moviendo a su alrededor. Eventualmente, él resuelve su problema tapándose las orejas con una manta para no oír los ruidos de la selva, al tiempo que se expone a ser atacado y devorado por un tigre real.

Los ruidos son la variable dependiente y los tigres la variable independiente. Los ruidos casuales en la selva son cambios producidos en la variable dependiente por la operación del azar y el sonido de un tigre real es un cambio en la variable dependiente producido por la variable independiente. Oír tigres cuando no existen es la primera clase de error; no oír el tigre cuando está ahí es la segunda clase de error. Taparse las orejas con la manta es hacer más pequeño el nivel de significación: es decir, hacer menos probable que oiga ruidos casuales de la selva y los atribuya a tigres. Sin embargo, mientras el uso de la manta disminuye la probabilidad de la primera clase de error, también aumenta la probabilidad de la segunda clase. Y por el contrario, la remoción de la manta disminuye la probabilidad de la segunda clase de error (no oír los ruidos de un tigre real), pero aumenta la de la primera clase (atribuir los ruidos casuales a un tigre real). Curiosamente, en este caso la segunda clase de error tiene consecuencias más perjudiciales para el joven cazador, a diferencia de lo que generalmente ocurre en la investigación científica, donde es la primera clase de error (ver más de la cuenta en los datos) la que arroja las consecuencias más perjudiciales. Estas probabilidades son interdependientes, y la mejor solución es mantenerlas a ambas en un mínimo aceptable -la cantidad justa de manta alrededor de las orejas para oír unos ruidos y no otros-.

¿Cuál debe ser esta cantidad justa de manta? No hay una sola respuesta a esta pregunta, pues la gravedad relativa de las dos clases de error difiere de situación a situación. En consecuencia, el nivel de significación que es aceptado, varía algo de científico a científico y de investigación a investigación. Muy pocos científicos, sin embargo, aceptarían una probabilidad mayor de .05 o insistirían en una menor de .001. El hecho de que estas probabilidades son marcadamente pequeñas refleja la precaución del científico que no quiere ver en los datos más de lo que realmente hay en ellos. Nosotros adoptaremos un nivel de significación de .05, porque es el más frecuentemente usado. En esta forma, si la probabilidad de que los resultados de una investigación sean producidos completamente por error de azar es de .05 o menor, nosotros rechazaremos esta posibilidad; y, por el contrario, si esta posibilidad es mayor de .05, lo que realmente queremos decir es que consideramos el error de “ver demasiado en los datos” lo suficientemente grave como para no querer cometer este error más de cinco veces en cien, cuando no hay ninguna relación entre la variable independiente y la variable dependiente. Estamos dispuestos a adoptar un nivel de significación tan grande como éste, porque también queremos reducir a un mínimo aceptable la probabilidad de cometer el error de "no ver lo suficiente en los datos".

Las pruebas o análisis estadísticos más frecuentemente usados en la verificación o prueba de hipótesis sobre parámetros poblacionales son la ji cuadrada de Pearson cuando comparamos los datos expresados en escalas nominales de dos o más grupos independientes; la prueba U de Mann-Whitney cuando los datos de dos grupos independientes están expresados en escalas ordinales y la pruebas t de Student cuando la variable dependiente para dos grupos, sean éstos independientes o relacionados, es medida con una escala de intervalo; en este último caso, si se trata de más de dos grupos independientes, se utiliza la prueba F, llamada así en honor a Ronald Fisher.

Cuando queremos conocer la relación entre dos variables medidas en una escala nominal se usa el coeficiente de contingencia, de Pearson; para escalas ordinales, el coeficiente de asociación ordinal de Goodman y Kruskal; y para escalas de intervalo, la correlación de Pearson (r). Cuando la investigación incluye más de una variable independiente, entonces se utilizan análisis multivariados, tales como Análisis de Varianza y Análisis de Regresión Múltiple.

Cuando después del resultado de un análisis estadístico aparece $p<0.05$, eso significa que ese resultado **es** estadísticamente *significativo*, es decir, que sólo pudo haber ocurrido por azar (por casualidad) menos de cinco veces en cien; si aparece $p<0.01$, significa que ese resultado es *altamente significativo*, pues sólo puede ocurrir por azar menos de una vez en cien; y si aparece $p<0.001$, entonces significa que ese resultado es *altísimamente significativo,* pues sólo puede ocurrir por azar menos de una vez en mil. Si, por el contrario, al lado derecho o debajo del resultado aparece $p>0.05$, entonces eso significa que ese resultado *no es* estadísticamente *significativo*, pues puede ocurrir por azar más de cinco veces en cien.

REFERENCIAS

Aron, A. & Aron, E.N. (2001). *Estadística Para Psicología***.** Buenos Aires: Pearson Educación, S.A.

Freeman, Linton C. (1965). *Elementary Applied Statistics: For Students in Behavioral Science*. New York: John Wiley & Sons, Inc.

Hopkins, K.D., Hopkins, B.R., Glass, G.V. (1997). *Estadística Básica*. México, D.F.: Prentice-Hall.

Michell, Joel (1993). The Origins of the Representational Theory of Measurement: Helmholtz, Hölder, and Russell. *Studies in History and Philosophy of Science, 24*, 185-206.

Rodríguez, Enerio (2005). Estadística y Psicología: Análisis Histórico de la Inferencia Estadística. *Perspectivas Psicológicas, Vol. 5*, Año VI, 96-102.

Salkind, Neil J. (2004). *Statistics for People Who (Think They) Hate Statistics*. Thousand Oaks, C.A.: Sage Publications.

Stevens, S.S. (1946). On the Theory of Scales of Measurement. *Science, 103*, No.2684, 677-680.

Urdan, Timothy (2005). *Statistics in Plain English* (Second Edition), Mahwah, N.J.: Lawrence Erlbaum Associates, Publishers

Vogt, W.P. (1999). *Dictionary of Statistics & Methodology.* Thousands Oaks, C.A.: Sage Publications.