SOBRE EL CONTRASTE DE HIPÓTESIS

Conceptos claves de un curso de Estadística para psicólogos

En este texto se atienden los temas fundamentales que conforman el temario del programa de la asignatura Estadística III, de la Cátedra de Estadística del Departamento Metodológico de la escuela de Psicología de la UCV

Dimas Sulbarán

22/09/2012

ÍNDICE

1. Contraste de hipótesis 3
   1.1. La lógica del contraste estadístico de hipótesis. 6
      1.1.1. Ronald A. Fisher y la Prueba de Significación (P-Values) 7
      1.1.2. La teoría de Neyman-Pearson y la Prueba de Hipótesis (Hypotesis Testing) 7
   1.2. Pasos del contraste de hipótesis. 9
   1.3. Tipos de error. Error tipo I y error tipo II 10
   1.4. Potencia del contraste 11
      1.4.1. Función de potencia del contraste. 12
      1.4.2. Cálculo de la potencia del contraste 12
   1.5. Tipos de hipótesis 13
   1.6. Factores que influyen en el rechazo de la hipótesis nula 15
   1.7. Tipos de contrastes 16
      1.7.1. Contraste uni y bilaterales. 17
      1.7.2. Contraste de hipótesis paramétricos 18
         1.7.2.1. Contraste de hipótesis sobre la media. 20
            Contraste de hipótesis sobre la media para una muestra. *20*
            Contraste de hipótesis sobre la media para dos muestras independientes (con datos no pareados). *21*
            Contraste de hipótesis sobre la media para dos muestras, con datos pareados. *21*
         1.7.2.2. Contraste de hipótesis sobre la porción 22
         1.7.2.3. Contraste de hipótesis sobre la varianza, de una y dos vías. 22
            Cálculo de ANOVA de una vía. *23*
            Cálculo de ANOVA de dos vías. *25*
            Representación gráfica de los diseños factoriales. *28*
      1.7.3. Contraste de hipótesis No paramétricos (para poblaciones no normales) 29
         1.7.3.1. Pruebas para una muestra 30
            Prueba de chí cuadrado. *30*
            Prueba binomial. *31*
            Prueba de Kolmogorov-Smirnov. *32*
         1.7.3.2. Pruebas para dos muestras independientes 33
            Prueba de Chí cuadrado para dos muestras independientes. *33*

Prueba de la mediana para dos grupos independientes. ... *34*

Prueba U de Mann-Whitney ... *34*

1.7.3.3. Pruebas para dos muestras relacionadas ... 35

Prueba de cambios de McNemar ... *36*

Prueba de Wilcoxon ... *36*

1.7.3.4. Pruebas para varias muestras independientes ... 37

Prueba de la mediana para más de dos grupos independientes. ... *37*

Prueba H de Kruskal-Wallis ... *38*

1.7.3.5. Pruebas para varias muestras relacionadas ... 38

Prueba Q de Cochran ... *39*

Prueba de rangos de Friedman ... *39*

Bibliografía ... 41

Contraste de Hipótesis

El contraste de hipótesis es una estrategia metodológica fundada en la inferencia estadística que permite evaluar si una propiedad que el investigador supone (hipotetiza) existe en una población, es compatible con lo observado en una muestra experimental. En esta situación se encontraría la lógica inductiva de Carnap, una teoría según la cual la esencia del razonamiento inductivo consiste en la determinación de valores de probabilidad de las hipótesis científicas en razón de los datos de experiencia. El contraste de hipótesis se caracteriza por "constituir uno de los instrumentos más apropiados para la toma de decisiones en condiciones de incertidumbre". (López Casuso, 1996, pág. 297). Las propuestas teóricas fundamentales para el contraste de hipótesis surgen a principios del S.XX con base en los trabajos pioneros de autores como Charles Sanders Peirce, Student (William Gosset), Ronald Fisher y los desarrollos de Jerzy Neyman, Egon Pearson y Richard Von Misses. (Reichenbach, 1949).

En términos generales, se puede definir el contraste de hipótesis como: "una función de decisión o estrategia que lleva a aceptar o rechazar la hipótesis nula H_0" (López Casuso, 1996, pág. 304). Mediante esta teoría, se aborda el problema científico considerando su traducción en una hipótesis estadística determinada H_o, y una hipótesis alternativa H_1, a partir de las cuales se intenta decidir cuál de las dos es la hipótesis verdadera, tras aplicar el problema estadístico a un cierto número de experimentos.

La lógica del contraste de hipótesis se resume en la tesis de que una muestra aleatoria del conjunto total tiende a representar el comportamiento real de la población en la medida en que aumenta el número de observaciones. En un contexto de incertidumbre con relación a la validez de nuestras hipótesis de investigación, elementos teóricos matemáticos tan robustos como los de la probabilidad pueden generar los argumentos para verificar, según Reichenbach; confirmar, según Carnap o falsear, según Popper, una vez que nuestros constructos se traducen a términos matemáticos. En cualquier caso, el salto inductivo de los datos a la teoría es un tema que sigue generando polémica entre los filósofos de la ciencia.

En la práctica dos grandes corrientes caracterizan el contraste estadístico de hipótesis, nos referimos a las tradiciones derivadas de los trabajos de Ronald Fisher y aquellas de los trabajos de Jerzey Neyman y Egon Pearson. Del primero heredamos las pruebas de significación. Por su parte, los trabajos de Neyman-Pearson se traducen en el contraste de hipótesis por intervalos. Aunque, formalmente, ambas propuestas representan enfoques encontrados en la práctica se han unido en un matrimonio forzado que se evidencia en la mayoría de la producción investigativa contemporánea.

La práctica del contraste de hipótesis está fuertemente orientada al control de los errores de tipo I y II en estadística, que definen respectivamente, la posibilidad de tomar un suceso falso como verdadero, o uno verdadero como falso. En otras palabras, el contraste de hipótesis implica el interés por validar determinados modelos, en los cuales se supone los efectos de determinadas variables independientes sobre alguna variable dependiente. En el caso del error tipo I, al cual se le conoce como alfa de la prueba, el autor corre el riesgo de atribuir efectos a una variable independiente sobre una variable dependiente que no existen. Por su parte, el error tipo II, también conocido como beta, implica la posibilidad de que el investigador descarte los efectos de la variable independiente sobre la variable dependiente por fallas en el diseño.

La mayoría de la práctica estadística para el contraste de hipótesis se basa en el control del error tipo I (Kerlinger & Lee, 2002), favoreciendo los argumentos con relación al nivel de confianza. Sin embargo, la propuesta alternativa de Neyman y Pearson, desarrollada en obras como *On the use and interpretation of certain test criteria for purposes of statistical inference* (1928), sugiere la necesidad de considerar la posibilidad de controlar el error tipo II o beta, favoreciendo la potencia del contraste. No obstante, pese a la amplia trayectoria de esta discusión, autores como (Cohen, 1992), sostienen haber observado en sendos metanálisis de numerosas investigaciones la existencia de altas debilidades con relación a la potencia del contraste.

El contraste de hipótesis a la luz de los desarrollos de las teorías estadísticas pertinentes implica un trabajo racional por la búsqueda de un equilibrio entre el control del error tipo I y la potencia del contraste. En función de ello, los autores sugieren trabajar con algunos elementos críticos asociados con el rechazo de la hipótesis nula como son: el tamaño del efecto, el nivel de significación o alfa y el tamaño de la muestra o número de observaciones. Básicamente, la potencia del contraste se mueve como una función de estos elementos que se resume en la necesidad de incrementar el índice de discrepancia o efecto relativo de la variable independiente sobre la variable dependiente. Se entiende el índice de discrepancia como una razón del tamaño de las diferencias entre los parámetros contrastados dividido por la varianza relativa de la población.

En términos metodológicos el concepto de contraste de hipótesis abriga una familia de pruebas estadísticas con el fin de procurar argumentos cuantitativos para sostener la aseveración con relación a la validez de los hallazgos de una investigación. Sin embargo, al estudiar a nivel operativo esta familia de pruebas, el investigador encontrará una diversidad de modelos que han sido desarrollados con el fin de garantizar la mayor eficiencia en el contraste. Básicamente, la investigación ha tomado en cuenta dos criterios fundamentales para clasificar las distintas pruebas de contraste de hipótesis: a) el diseño de la investigación y b) el cumplimiento de los supuestos asociados con el uso de pruebas paramétricas. Según el diseño, hoy en día se cuenta con un complejo conjunto de pruebas que se derivan de la combinación entre: a) el número de muestras, b) el número de observaciones y c) los niveles de independencia de las observaciones. En cuanto al cumplimiento de los supuestos, la literatura agrupa las pruebas estadísticas de contraste en las categorías: paramétricas y no paramétricas, tomando en cuenta los criterios: a) nivel de medición de las variables y b) supuestos con relación a la forma de la distribución de la población. (Siegel & Castellan, 1995).

De esta manera, hoy en día se cuenta con un número importante de pruebas estadísticas para el contraste de hipótesis, en cuya base se encuentran las teorías matemáticas de la probabilidad frecuentista. (Mayo & Cox, 2006; Neyman, 1977). En cualquier caso, las pruebas estadísticas de contraste permiten determinar un estimado de la varianza sistemática y diferenciarla de la varianza de error, en las relaciones funcionales en estudio. De modo que, el objetivo de la prueba será en todo caso, determinar el grado de probabilidad de que un evento cualquiera, entendido como valor estimado, ocurra en el universo de eventos posibles determinado por la distribución del error muestral.

1.1. La lógica del contraste estadístico de hipótesis.

Aunque es difícil hablar de una lógica del contraste estadístico de hipótesis, en términos generales, la idea se resume al planteamiento Charles Sanders Peirce (1839-1914). Se reconoce a Peirce como un defensor de la teoría inductiva de la probabilidad. La mayor contribución conceptual de Peirce a la estadística fue su entendimiento del rol de la aleatorización en los procesos de inferencia; concepto que marca un hito en la investigación psicológica experimental. La propuesta de este autor se enumera de la siguiente manera: hay un tipo de pensamiento conjetural, las abducciones, que generan hipótesis. Los explicativos asociados a la deducción que permiten predecir las consecuencias experimentales que se deberían observar si la hipótesis fuera correcta. Los ampliativos o extensivos que definen la inducción, cuyo objeto permite verificar las hipótesis propuestas mediante la experimentación (generando grados estadísticos de corroboración). (Wiener, 1978). De esta manera heredamos la tradición que sostiene que: Si las observaciones que tales experimentos implican no corresponden a las consecuencias derivadas de las hipótesis, se generan nuevas hipótesis. Si corresponden, se repiten las observaciones, con un grado creciente de confianza.

1.1.1. Ronald A. Fisher y la Prueba de Significación (P-Values)

No es posible introducir el punto del contraste de hipótesis estadísticas sin hacer referencia los trabajos acerca de la inferencia inductiva de Ronald Fisher. (Fisher, 1935).

En el enfoque de Fisher el científico debe especificar la hipótesis nula de que la muestra proviene de una población hipotética infinita, con distribución muestral conocida (la distribución por excelencia de las variables en la naturaleza es la normal). Se rechaza la hipótesis nula si la estimación muestral difiere de la media de la distribución con una probabilidad menor al nivel de significación acordado, que el mismo Fisher aconseja fijar en un 5%. (Cowles & Davis, 1982).

En síntesis, propuso una estrategia para testear la hipótesis nula H_0 y usar el valor probabilístico o valor p (*p-value*) para decidir acerca de la fuerza de la evidencia, de modo que: "si el valor p es pequeño esto implica que la realización observada del estadístico de prueba constituye un evento raro, o bien que la hipótesis nula postulada es inválida". (Fisher, 1935).

En la perspectiva de Fisher, el principal objetivo es utilizar los datos como evidencia inductiva en contra de la hipótesis nula. El resultado de la prueba es un valor de p_{obs}, o *p-value*, y al procedimiento le denomina *prueba de significación*.

1.1.2. La teoría de Neyman-Pearson y la Prueba de Hipótesis (Hypotesis Testing)

De acuerdo con (Rivadulla, 1991), tras rechazar la viabilidad del método Bayesiano y el Fisheriano de máxima verosimilitud, como procedimientos adecuados de estimación matemática, Neyman sugiere el método de estimación por intervalos de confianza, en otras palabras, por estimación de los intervalos inferior y superior del parámetro poblacional desconocido. Son nociones claves de esta propuesta los conceptos de límites de confianza, coeficiente de confianza e intervalo de confianza.

En la teoría de N-P, el problema de probar hipótesis se presenta cuando las circunstancias obligan a elegir entre dos cursos de acción: A ó B, elección que depende de la distribución de probabilidad desconocida de una variable X; la decisión quedará determinada por los valores muestrales de esta variable X. La acción A es preferida si la distribución pertenece a un conjunto posible de distribuciones de X, y B es preferida si pertenece a otro conjunto. Al decir de Neyman:

"la elección entre dos acciones A y B se interpreta como la aceptación de una hipótesis H y el rechazo de otra H*. Si la regla adoptada lleva a la acción A diremos que la hipótesis H es aceptada (y por lo tanto H* es rechazada); por otra parte si la aplicación de esta regla lleva a la acción B diremos que la hipótesis H es rechazada (y H* es aceptada).

Es importante aclarar que los términos de "*aceptación*" o "*rechazo*" de hipótesis se han popularizado en el discurso estadístico, al punto que son aceptados y de uso común entre los investigadores. No obstante, se debe tener presente su correcta interpretación y desechar implicaciones adicionales que puedan ser sugeridas por la intuición. Así, aceptar la hipótesis H_0 significa sólo decidirse a favor de la acción A y no hacia B. Por tanto, no significa que calificamos la hipótesis H_0 como verdadera; también rechazar H_0 solamente significa elegir la acción B, lo cual no implica que pensemos que H_0 es falsa.

La teoría de N-P fija reglas de comportamiento inductivo para la toma de decisiones: se decide en favor de una u otra hipótesis, sin considerar la creencia personal del investigador en alguna de ellas. De este modo, los autores introducen el concepto de comportamiento inductivo, en contraposición al de inferencia inductiva de Fisher. Es claro que para llegar a estas reglas el investigador elabora su razonamiento sobre la base de las teorías de Probabilidad y Estadística; y aclaran los autores: este razonamiento es totalmente deductivo. (Birnbaum, 1977).

El principio básico en la teoría de N-P se vincula con la forma de juzgar los procedimientos estadísticos: éstos deberán ser evaluados en términos de sus propiedades probabilísticas, (*performance characteristics* en palabras de Neyman), es decir, según su comportamiento promedio o de largo plazo: un procedimiento con buenas propiedades probabilísticas mostrará, si se lo usa repetidamente, una buena performance, en promedio.

De ahí que, Neyman y Pearson interpretan la prueba de hipótesis como regla de decisión en el contexto del muestreo repetido; Fisher estaría de acuerdo en aceptar esta interpretación en problemas comerciales o tecnológicos, pero no para corroborar hipótesis científicas: en estos casos el muestreo repetido genera confusión al no existir un problema de decisión bien definido. Creía firmemente que las pruebas de hipótesis no tenían como objetivo la toma de decisiones finales e irrevocables, como se dejaba entrever en los ejemplos presentados por N y P.

La postura de los autores acerca de sus diferentes teorías no era conciliadora. Fisher fué en todo momento un crítico del trabajo de N-P; Las diferencias, tanto filosóficas como prácticas fueron ferozmente defendidas por Fisher y Neyman como "una batalla" que tuvo efecto destructivo en la profesión de estadística y que perduró mientras Fisher estuvo vivo.

La forma en que actualmente se presenta la teoría de pruebas de hipótesis, en la gran mayoría de los textos, es una amalgama de ambas aproximaciones.

Autores como Aron y Aron (2001), describe la situación de la siguiente forma: aunque el debate Fisher vs. Neyman-Pearson sigue presente entre los estadísticos, fué resuelto silenciosamente en los libros de texto escritos entre 1954 y 1960, principalmente por no estadísticos para enseñar a estudiantes de ciencias sociales las "reglas de la estadística".

La teoría de las pruebas de significación de Fisher que históricamente apareció primero, fue mezclada con los conceptos de Neyman-Pearson y enseñada como la estadística per-se. Le llamamos a esto "teoría híbrida" de la inferencia estadística, y esto continúa aún sin mencionar siquiera que ni Fisher ni Neyman-Pearson hubieran visto con buenos ojos el florecimiento de este Matrimonio Forzado."

1.2. Pasos del contraste de hipótesis.

Vistos los elementos fundamentales de la inferencia estadística debemos abordar el tema con relación a los pasos fundamentales para el contraste de hipótesis (Kerlinger & Lee, 2002, pág. 265). Recordemos que el contraste de hipótesis hace referencia a la traducción de los constructos en modelos matemáticos, en el contexto de la aleatoriedad. Para ello debemos:

1. Establecer dicha hipótesis en términos de una relación a nivel paramétrico.
2. Enunciar la hipótesis nula o de ausencia de relación.
3. Calcular, con base en datos empíricos, los estadísticos de la prueba.

4. Definir la regla de decisión, referencia fundamental para establecer las conclusiones con relación a las pruebas.
5. Establecer la inferencia, implica extrapolar las observaciones de los datos a la teoría.

1.3. Tipos de error. Error tipo I y error tipo II

Estrictamente hablando, el contraste estadístico de hipótesis se reduce a la toma de una decisión por parte del investigador en un contexto de probabilidad, más específicamente de error o acierto. Dado que los argumentos estadísticos de la prueba siempre representarán sólo una ganancia a los niveles de incertidumbre general con relación a la naturaleza del fenómeno estudiado, existirá, por tanto, una posibilidad de acertar las características "verdaderas" o también de equivocarnos en la decisión que tomemos con respecto a las interpretaciones de los resultados de nuestras observaciones. Los autores han reducido a dos clases los tipos de error que se pueden cometer en el contexto del contraste de hipótesis estadística, a saber: error tipo I y II.

La teoría del contraste de hipótesis sostiene que pueden existir uno de dos estados, mutuamente excluyentes, en la naturaleza: los casos en los cuales la H_o es cierta o aquellos en los cuales la H_1 es cierta. Razón por la cual, tras estudiar el comportamiento de una muestra, se llega a una de las dos decisiones mencionadas. Lo cual implicará en cualquier caso el riesgo de cometer alguna de las formas de error comentados en el párrafo anterior. La tabla a continuación ilustra las combinaciones de dichos eventos y la condición de la decisión tomada según el caso.

		Decisión	
		No acepta H_0	Acepta H_0
Evento	H_0 es cierta	Error tipo *I* $p=\alpha$	Decisión correcta $p=1-\beta$
	H_0 es falsa	Decisión correcta $p=1-\alpha$	Error tipo *II* $p=\beta$

De esta manera, se entiende por error tipo I al caso en el cual se rechaza la hipótesis nula cuando esta es verdadera. En otras palabras, no se acepta la hipótesis nula y se asume que las observaciones arrojan diferencias "estadísticamente significativas" cuando en "verdad" no es así.

De esta manera, el investigador asumirá típicamente que los efectos de una determinada variable independiente sobre una determinada variable dependiente son “verdaderos” cuando en “realidad” no es así. Por ejemplo: xxxxxx

Por su parte, el error tipo II es el caso contrario al error tipo I. se dice que un investigador ha cometido un error tipo II cuando acepta la hipótesis nula en lugar de la alternativa siendo ésta en “realidad” falsa. En otras palabras, se acepta la hipótesis nula y se asume que las observaciones **NO** arrojan diferencias “estadísticamente significativas” cuando en “verdad” no es así. De esta manera, el investigador asumirá típicamente que los efectos de una determinada variable independiente sobre una determinada variable dependiente **NO** “existen” cuando en “realidad” no es así. Por ejemplo: xxxxxx

1.4. Potencia del contraste

Se llama *potencia del contraste* o *potencia de la prueba,* a la probabilidad de rechazar la hipótesis nula cuando esta es falsa, se conoce también como $1 - \beta$. (Kerlinger & Lee, 2002).

Por lo general, en las pruebas de hipótesis se especifica el valor de α y se diseña la prueba de tal forma que el valor de β sea pequeño. Esto es, la probabilidad de error tipo I se controla de forma directa, mientras que la probabilidad de error tipo II se controla de manera indirecta con el tamaño de la muestra, ya que a más datos, β será menor. En otras palabras, se suele asumir que con muestras grandes la potencia de la prueba será mayor. Sin embargo, parte de las críticas al esquema tradicional fundado en el contraste de significación de alfa se basan en la afirmación de que cuando el tamaño de la muestra se incrementa en exceso se llega a tener una potencia excesiva, que lleva al extremo de rechazar la hipótesis nula para efectos muy pequeños de la variable independiente sobre la variable dependiente. (Cohen, 1992).

Según (Glass & Stanley, 1974), en general, puede plantearse lo siguiente con respecto al poder para el contraste de hipótesis:

a. El poder para el contraste de H_0 aumenta proporcionalmente al tamaño n de la muestra, para cualquier valor del parámetro en cuestión, por ejemplo: p=0.40
b. Para determinado parámetro que hay que contrastar, por ejemplo, p=0.40, el poder del contraste de H_0 aumenta proporcionalmente a α, como cuando α pasa de 0.01 a 0.05.

c. Para valores fijos de α y n, el poder del contraste de h_0 aumenta a medida que el parámetro se aparta del valor supuesto para él en h_0. Por ejemplo, dado n=100 y α=0.01, el poder del contraste de h_0: p=0 es mayor para un p= 0.60 que para un p=0.40

1.4.1. Función de potencia del contraste.

Al contrastar una hipótesis sobre el parámetro de una población, la probabilidad de rechazar la hipótesis nula siendo falsa es una función del parámetro. Esta función p(θ) se conoce como "función de potencia del contraste" (López Casuso, 1996, pág. 317). Un contraste que maximiza la función de potencia para todo valor del parámetro se conoce como "uniformemente más potente". (López Casuso, 1996, pág. 324). Dada por la ecuación:

$$p(\theta) = 1 - \beta = P\left[\hat{\theta} \in S_1/\theta \in \Omega_1\right]$$

1.4.2. Cálculo de la potencia del contraste

Siguiendo a (Vargas Sabadías, 1995, pág. 344), el grado de falsedad de la hipótesis nula se puede determinar si se contrasta ésta con una hipótesis alternativa específica, en que se fija un valor concreto del parámetro, como puede ser: $H_1 = \mu_1 = \mu_0 + 2$

Como medida del grado de falsedad de la hipótesis nula se utiliza el *índice de discrepancia,* Γ, que proporciona una medida de la diferencia entre la hipótesis nula y la alternativa, en términos tipificados, a partir de la ecuación: $\Gamma = \frac{d}{\sigma}$, donde $d = \mu_1 - \mu_0$

Este factor es una función lineal de los demás elementos que intervienen en la determinación de la potencia, de forma que, si se aumenta el índice de discrepancia, crece la potencia, ya que resulta más probable rechazar la hipótesis nula cuando aumenta su grado de falsedad.

En cuanto al tamaño de la muestra, un aumento en la magnitud de la discrepancia favorecerá el uso de muestras menos grandes para alcanzar resultados potentes.

La práctica estadística está colmada de conflictos a la hora de determinar el índice de discrepancia más adecuado; en estas situaciones, se recomienda tomar algún valor convencional de forma similar al procedimiento para establecer el nivel de significación. Los valores habituales que se toman en estas situaciones son específicos de cada prueba. Por ejemplo, para el contraste de medias, se suelen tomar valores como: $\Gamma = 0.2, \Gamma = 0.5\ o\ \Gamma = 0.8$, según se aprecie una potencia más pequeña, mediana o alta.

La potencia de la prueba, de la que conocemos sus elementos constitutivos: α, n y Γ, se define como: $1 - \beta = P(rechazar\ H_0|H_1\ es\ verdadera) = P(rechazar\ H_0|\mu = \mu_1)$, bajo estas condiciones, se establece la siguiente función de potencia del contraste:

$$1 - \beta = P\big(\hat{\theta} \in S_1 \mid \theta \in \ \Omega_1\big)$$

Supongamos el caso emblemático del contraste de medias típico, se pueden dar tres condiciones con relación al valor de a, donde $a = \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}$, siendo z_α el valor crítico en la función de probabilidad correspondiente al nivel de significación α para una prueba unilateral. El primero: contraste de una cola por la derecha $1 - \beta = 1 - P(\,\bar{x} < a \mid \mu = \mu_1\,)$, una cola por la izquierda $1 - \beta = 1 - P(\,\bar{x} > a \mid \mu = \mu_1\,)$ y dos colas $1 - \beta = 1 - P(\,a < \bar{x} < a \mid \mu = \mu_1\,)$

Para llevar a cabo el contraste de probabilidad para la hipótesis alternativa, se calcula el valor de Z_β como: $z_\beta = \frac{\mu_0 - (\mu_0 + d)}{\sigma}\sqrt{n}$, cuya distribución se comporta como una normal $Z \sim N(0,1)$ y de lo cual se deduce que $z_\beta = z_\alpha - \frac{d}{s_{\bar{x}}}$. Por ejemplo, se presenta el caso en que se desea calcular la potencia del contraste para una distribución con las siguientes características n= 25, $\bar{x} = 205$; s=9, $s_{\bar{x}} = 1.8$, $\alpha = 0.05$, $z_{0.05} = 1.64$

$$H_0: \mu = 205$$
$$H_1: \mu < 205$$

De este modo procedemos a sustituir cada uno de los términos de la ecuación:

$z_\beta = \frac{205-(201)}{9} 5 = \frac{4}{9} 5 = 2.22$, de modo que p = 0.0132, es decir, la probabilidad de aceptar una hipótesis nula falsa es menor a 1.32 por ciento. Por tanto, la potencia del contraste es igual a $1 - \beta = 1 - P = 0.9868$ o casi un 99% de probabilidad de rechazar una hipótesis nula falsa.

1.5. Tipos de hipótesis

El contraste de hipótesis es la práctica científica que, como tal, busca articular los elementos teóricos, metodológicos y técnicos en la construcción del conocimiento científico. De este modo, el contraste de hipótesis involucra dos tipos de hipótesis fundamentales, según el campo discursivo en el cual se les ubica: el primero es el campo propiamente teórico y que definiremos como hipótesis sustantivas y el otro es el campo metodológico donde las hipótesis se definen operacional y estadísticamente. A nivel sustantivo, las hipótesis expresan una afirmación conjetural acerca de la relación entre dos o más variables. Mientras que las hipótesis sustantivas nos hablan del plano formal del conocimiento, es decir, nos describe las relaciones teóricas entre las variables, estrictamente hablando estas no pueden ser probadas a menos que se traduzcan en definiciones operacionales (Kerlinger & Lee, 2002). De modo que, el investigador deberá traducir sus enunciados teóricos a elementos operacionales y reducir estos a formulaciones estadísticas que le permitan generar los argumentos para la "verificación" o "falsación" de sus hipótesis sustantivas. Una hipótesis estadística se define como "una predicción sobre cómo resultarán los estadísticos utilizados al analizar los datos cuantitativos de un problema de investigación" (Kerlinger & Lee, 2002, pág. 251). Con fines operativos, han sido clasificadas en hipótesis nula e hipótesis alternativa.

Kerlinger y Lee (2002) nos ilustran de forma interesante el contraste de hipótesis a partir de una analogía con el sistema judicial. En términos concretos los autores refieren:

> La condición de la H_o resulta similar a la de un acusado en un juicio, en el cual es considerado "inocente" hasta que se pruebe que es culpable. Si el juicio resulta en un veredicto de "no culpable", ello no quiere decir que el acusado sea "inocente", tan sólo significa que no pudo demostrarse la culpa más allá de la duda razonable. Cuando el investigador no logra rechazar la H_o, eso no significa que la H_o sea verdadera, sino que no pudo demostrarse su falsedad más allá de la duda razonable. (Kerlinger & Lee, 2002, pág. 253).

Vale la pena aclarar que, en lo referente a las hipótesis estadísticas, debemos tener presente la siguiente distinción, en términos epistemológicos, de las mismas: a) el primero, con relación al supuesto con respecto a la forma de la distribución de la variable en la naturaleza, b) el segundo, con relación a los valores que pueden asumir los parámetros desconocidos de la población de interés. (López Casuso, 1996, pág. 298). Este segundo tipo de hipótesis con base en conjeturas con relación a los parámetros es la forma más popular de la práctica estadística inferencial, correspondiente al contraste de hipótesis a partir de pruebas paramétricas y que ocupará el apartado siguiente. El tratamiento de hipótesis estadísticas No paramétricas será objeto de estudio en un apartado posterior.

1.6. Factores que influyen en el rechazo de la hipótesis nula

Ciertos factores constitutivos de las pruebas de contraste moderan la posibilidad de rechazar la hipótesis nula, rayando en la posibilidad de cometer un error tipo I. por esta razón, los autores coinciden en la necesidad de diseñar de forma razonada las pruebas de contraste con el fin de contrarrestar las debilidades probadas en el control del error tipo I y II o β. Dado que todas las pruebas de contraste parten de la estimación de los parámetros con base en la función de distribución muestral; la cual, se encuentra definida por los parámetros de la variable aleatoria x, la varianza de la variable y el primer momento de la distribución, los elementos que influyen en el rechazo de la hipótesis nula incluyen:

- El alfa establecido por el investigador.
- El tamaño de la muestra.
- La varianza de error.
- El tipo de contraste.
- El tamaño del efecto.

Por ejemplo, ilustremos el caso con la ecuación fundamental para el contraste de medias para poblaciones infinitas, de las cuales se conoce la varianza poblacional, definida como:

$$\frac{\mu - \bar{x}}{\sigma}\sqrt{n} = \pm z_{\alpha}$$

α= probabilidad de error tipo I. si el valor de alfa se incrementa, el valor crítico para el contraste lo hace proporcionalmente y lo que en otra circunstancia bastaría para rechazar la Ho podría no ser suficiente y terminar aceptándola, aunque sea falsa.

n= tamaño de la muestra. Se puede apreciar como le imparte un respaldo geométrico a la diferencia relativa entre las medias contrastadas. Sin embargo, favorece el riesgo de que pequeñas diferencias alcancen un resultado estadísticamente significativo por lo que debe evaluarse críticamente el valor de n en el contraste.

$\sigma =$ la varianza de error. Algebraicamente se evidencia como un incremento en la varianza de error atenta contra el rechazo de la hipótesis nula, pues reduce el valor relativo de la diferencia entre las medias sometidas a contraste.

$\mu - \bar{x}$= al tamaño del efecto. En la medida en que aumenta la diferencia entre la media estimada y el valor de la media poblacional, la magnitud del numerador crece proporcionalmente y esta distancia es favorable para la confianza y la potencia del contraste.

Finalmente, la posibilidad de aceptar o "rechazar" la hipótesis nula de forma pertinente estará condicionada al tipo de contraste que se lleve a cabo. Este punto merece un tratamiento particular tal como se expondrá a continuación.

1.7. Tipos de contrastes

Los tipos de contraste de hipótesis se clasifican, fundamentalmente, por el diseño de la investigación a la cual obedecen. Por su parte, los diseños atenderán a los siguientes criterios para su clasificación: a) número de grupos a contrastar (uno, dos o k grupos), b) el nivel de independencia de los grupos (relacionados o no relacionados), c) los supuestos acerca de la distribución (nivel de medición y forma de la distribución poblacional) y d) la direccionalidad de la hipótesis (unilateral o bilateral). De acuerdo con esto se ha generado la siguiente tabla de referencia:

Tabla x. *pruebas estadísticas para el contraste de hipótesis.*

Número de grupos.	**Nivel de independencia.**	**Nivel de medición.**		**Prueba recomendada**	**Tipo de prueba.**
Una muestra.	No aplica.	Nominal	Binario	Binomial.	No Paramétrica
			Politómico	Ji cuadrada.	
		Ordinal.		Z de Kolmogorov-Smirnov	
		Escala.		t de Student para una muestra.	Paramétrica.

Dos muestras	No relacionados	Nominal	Binario	Prueba exacta de Fisher	No Paramétrica.
			Politómico	Ji cuadrada.	
		Ordinal		U de Mann-Whitney	
		Escala.		t de Student para grupos independientes.	Paramétrica.
	Relacionados	Nominal.	Binario	Mc Nemar	No Paramétrica.
			Politómico	Ji cuadrada.	
		Ordinal		Wilcoxon.	
		Escala.		t de Student para grupos relacionados.	Paramétrica.
K muestras	No relacionados	Nominal		Ji cuadrada.	No Paramétrica.
		Ordinal		Kruskal-Wallis	
		Escala.		ANOVA de una o dos vías	Paramétrica.
	Relacionados	Nominal.		Q de Cochran	No Paramétrica.
		Ordinal		Prueba de rangos de Friedman	
		Escala.		ANOVA intragrupo	Paramétrica.

Además de lo anterior, las pruebas de hipótesis obedecen al tipo de hipótesis planteada con relación al rango de probabilidad de error para el rechazo de la hipótesis nula o error tipo I. este punto hace referencia a la direccionalidad de la hipótesis (contrastes uni y bilaterales).

1.7.1. Contraste uni y bilaterales.

Al fijar el valor de α, por consecuencia se está determinando el tamaño de la región crítica para la función de distribución dada, quedando por especificar la situación de esa región crítica. En general pueden darse tres casos: a) que toda la región crítica quede en la parte derecha del espacio muestral; es decir, la región crítica corresponde a valores mayores que valor dado, b) que toda la región crítica esté situada en la parte izquierda del espacio muestral; es decir, que la región crítica corresponda a valores menores que un valor dado, y c) que la región crítica esté dividida en los extremos superior e inferior de la distribución muestral; es decir, que la región crítica esté formada por la unión de los valores menores y mayores que un par de valores dados, respectivamente. Estos tres casos corresponden a tres tipos de contraste de hipótesis conocidos como: a) contraste de una cola por la derecha, b) contraste de una cola por la izquierda, y c) contraste de dos colas. En los dos primeros casos las regiones crítica y de aceptación están determinadas por un solo valor crítico; mientras que en el tercero, existen dos valores críticos para acotar las zonas de aceptación y rechazo de H_0. (López Casuso, 1996, pág. 320).

La actuación del investigador deberá guiarse por el siguiente principio: minimizar las probabilidades de cometer los dos tipos de error del contraste de hipótesis, a saber: los errores tipo I y II; en otras palabras, aquel que reduce la probabilidad de α y β. (López Casuso, 1996, pág. 319). De los tres tipos de contraste estudiados, será mejor, según el principio antes mencionado, aquel que minimice β para todos los valores alternativos de μ, en otras palabras, aquel que maximice la potencia del contraste.

Como lo define muy bien López Casuso (1996), el contraste más apropiado para cada caso, depende de cuál sea la hipótesis alternativa. Como regla general se puede enunciar la siguiente: dada H_0: $\theta = \theta_0$, entonces se puede dar alguna de las siguientes situaciones:

- si la hipótesis alternativa es h_1: $\theta > \theta_0$, el mejor contraste es el de una cola por la derecha.
- si la hipótesis alternativa es h_1: $\theta < \theta_0$, el mejor contraste es el de una cola por la izquierda.
- si la hipótesis alternativa es h_1: $\theta \neq \theta_0$, el mejor contraste es el de dos colas.

1.7.2. Contraste de hipótesis paramétricos

Como hemos mencionado hasta ahora, el contraste o prueba de hipótesis es un desarrollo de la moderna estadística inferencial. Hemos resaltado el aporte de las primeras técnicas estadísticas inferenciales para el contraste de hipótesis, en la tradición Fisheriana, fundadas en el manejo de una serie de supuestos con relación a la forma de la distribución de las poblaciones de las cuales se generan las observaciones y se extraen los datos. El supuesto fundamental es que las puntuaciones de la población se distribuyen de manera normal. (Kerlinger & Lee, 2002). Éstas técnicas estadísticas se les conoce por el nombre genérico de *pruebas paramétricas*. Tales técnicas proporcionan información con relación a nuestras hipótesis del tipo: “si las suposiciones acerca de la forma de la distribución de la población son válidas, entonces podemos concluir que…” (Siegel & Castellan, 1995, pág. 24).

Como hemos observado en los párrafos anteriores, la discusión con relación al uso de pruebas paramétricas o no paramétricas se reduce al problema de los supuestos. De acuerdo con autores como Kerlinger y Lee (2002) y Siegel (1995), el problema de los supuestos ha devenido en que estadísticos e investigadores en general consideren que la violación de los mismos es un asunto crítico que podría traducirse en la invalidación de los resultados de la investigación. Para este análisis presentaremos una revisión de los tres supuestos más importantes, de acuerdo con la literatura, y la evidencia para considerar robustos a los métodos paramétricos.

Supuesto de normalidad

Desde los trabajos pioneros de Fisher (1935), uno de los fundamentos del contraste de hipótesis ha sido el supuesto de que la distribución de las puntuaciones para la población de origen ocurre en el contexto de la normalidad estadística según la tradición de Gauss. La clave de este supuesto es el control estadístico de la varianza de error, la cual se supone se comporta de manera normal y sostiene las declaraciones con respecto a las diferencias estadísticamente significativas.

Pruebas como la Kolmogorov-Smirnov de una muestra nos permiten evaluar el supuesto de normalidad en las observaciones realizadas.

Homogeneidad de la varianza

En el análisis de varianza se supone que las varianzas dentro de los grupos son estadísticamente las mismas, es decir, se supone que las varianzas son homogéneas de un grupo a otro, dentro de los límites de la variación aleatoria. (Kerlinger & Lee, 2002, pág. 372).

Pruebas como el contraste de varianzas de Levene se suelen emplear para evaluar el supuesto de homogeneidad de las varianzas.

Continuidad e intervalos iguales de medida

De acuerdo con las consideraciones estudiadas con relación a los procesos de medición y, particularmente, a la cuantificación en ciencias sociales, el supuesto con relación a la continuidad e intervalos iguales de medida se reduce a una afirmación muy simple: si los datos recolectados se derivan de un proceso de medición que no favorecen la construcción de una media que resulte en un parámetro pertinente y válido, entonces todo lo demás carecerá de sentido. (Hand, 1996).

Con base en los requerimientos expuestos, se ha generado una familia de pruebas que reciben el calificativo común de pruebas paramétricas y se resumen en la siguiente tabla:

Diseño			
Número de grupos.	**Nivel de independencia.**	**Estimador**	**Prueba recomendada**
Una muestra		Media	t de Student para una muestra.
Dos muestras	No relacionados		t de Student para grupos independientes.
	Relacionados		t de Student para grupos relacionados.
K muestras	No relacionados	Varianza	ANOVA de una o dos vías
	Relacionados		ANOVA intragrupo

Los puntos a continuación se pasearán por cada una de estas propuestas y darán cuenta de los principales aportes de cada uno de estos paradigmas para el análisis de los datos.

1.7.2.1. Contraste de hipótesis sobre la media.

Contraste de hipótesis sobre la media para una muestra.

En cualquier caso, el contraste de hipótesis para las medias obedecerá al supuesto de que nuestras variables se distribuyen de manera normal y son mediciones con un nivel de medida de intervalo o superior.

$t_{obt} = \frac{\bar{x}-\mu}{\frac{s_x}{\sqrt{n}}}$, donde $S_x = \sqrt{\frac{\sum(x_i-\bar{x})^2}{n-1}}$

Es particularmente útil, antes de adentrarnos en el análisis de las pruebas para el contraste de medias para grupos independientes o relacionados, la revisión de la ecuación generalizada para la prueba *t* (McGuigan, 1977) y que se define como:

$$t_{obt} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} - 2(r_{12})\left(\frac{S_1}{\sqrt{n_1}}\right)\left(\frac{S_2}{\sqrt{n_2}}\right)}}$$

Las dos ecuaciones para el cálculo de *t* que se expondrán a continuación se derivan de la ecuación anterior. Para entender el concepto general que implica la prueba, basta con sustituir el coeficiente de correlación r_{12} en la fórmula por los valores 0 y 1. Donde 0 significa independencia o ausencia de relación y 1 significa total dependencia o relación.

Contraste de hipótesis sobre la media para dos muestras independientes (con datos no pareados).

$$t_{obt} = \frac{(\bar{X}_1 - \bar{X}_2) - \left(\mu_{\bar{x}_1 - \bar{x}_2}\right)}{\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Donde, S^2 hace referencia a la varianza de la distribución de cada grupo respectivamente. Si asumimos que nuestra hipótesis nula corresponde a una ausencia total de efectos de la variable independiente, lo que se traduciría en una diferencia de medias poblacional $\mu_{\bar{x}_1 - \bar{x}_2} = 0$, entonces la ecuación se traduce en:

$$t_{obt} = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Recordemos que se pierde un grado de libertad por cada desviación típica que calculamos. Dado que se debe calcular S_1 y S_2 para el caso de dos muestras, perdemos dos grados de libertad, uno por cada muestra, `por tanto: $gl = n_1 + n_2 - 2$

Contraste de hipótesis sobre la media para dos muestras, con datos pareados.

$$t = \frac{\bar{d}}{\frac{S_d}{\sqrt{n}}} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sum\left(d_i - \bar{d}\right)^2}{n(n-1)}}}$$

Donde, *d* equivale a la diferencia entre las calificaciones de la variable dependiente para cada par de sujetos.

1.7.2.2. Contraste de hipótesis sobre la porción

Para contrastar hipótesis sobre la proporción “p” de una población nos fundamos en el teorema de Laplace-DeMoivre y suponemos que la distribución dela población para una variable binaria se ajusta a una distribución normal, de modo que:

$$\hat{p} \sim N\left[p, \frac{pq}{n}\right]$$

De esta manera, nos basaremos en la distribución z para llevar a cabo la inferencia con relación al contraste de proporciones, siguiendo el esquema utilizado para el contraste de medias. La fórmula ajustada al cálculo de porciones se define como:

$$z_{obt} = \frac{\hat{p} - p_o}{\sqrt{\frac{pq}{n}}}$$

Donde $\hat{p}$ es el valor estimado de p y p_o el valor de la proporción para la hipótesis nula.

1.7.2.3. Contraste de hipótesis sobre la varianza, de una y dos vías.

En los puntos anteriores se ha utilizado la media como el estadístico básico para evaluar la hipótesis nula, en lo sucesivo se presentarán las pruebas de contraste de hipótesis que se basan en la varianza como estimadores. Llamaremos prueba F a una de las pruebas más importantes en esta línea.

La prueba F se fundamenta en el análisis de la varianza total de la variable dependiente, permitiendo el contraste entre las varianzas intergrupales o atribuible a los cambios en la variable independiente con respecto a las varianzas intragrupales o de error. El resultado se contrastará contra la tabla de valores críticos para la distribución muestral de F.

La prueba F es apropiada para los análisis de experimentos en los cuales los puntajes pueden ser utilizados para formar dos estimados independientes de la varianza poblacional. Típicamente, se asocian los análisis de varianza o ANOVA (acrónimo en inglés de *analysis of variance*) con los resultados de experimentos en los cuales se dispone de más de dos grupos de comparación, en otras palabras, los llamados diseños multigrupos; tanto para grupos independientes como para grupos relacionados. En diseños de un solo factor (una variable independiente) o factoriales, con factores múltiples (más de dos variables independientes).

De esta manera se distinguen dos acepciones fundamentales para el análisis de varianza, a saber: el análisis de varianza de una vía y el análisis de varianza de dos vías. En cualquiera de los casos, el empleo de la prueba F nos permite alcanzar una comparación general de las distribuciones de los distintos grupos de observaciones y nos indica si existe una diferencia significativa entre al menos dos de los grupos definidos.

Cálculo de ANOVA de una vía.

En definitiva, la prueba F se reduce al cálculo de F_{obt} por medio de la siguiente fórmula:

$$F_{obt} = \frac{S_B^2}{S_W^2}$$

Para el cálculo de F_{obt}, se procede a elaborar una tabla que reúna la información necesaria, que se suele denominar "Tabla de resumen para el Análisis de varianza o ANOVA", con los resultados de las sumas de cuadrados, las medias cuadráticas, los grados de libertad y el valor de F, que adopta la siguiente forma:

Fuente de variación	Suma de cuadrados (SC)	Grados de libertad	Cuadrados medios (CM)	Valor de la función F
Entre grupos	SS_B	$k-1$	$S_B^2 = \frac{SS_B}{k-1}$	$F_{obt} = \frac{S_B^2}{S_W^2}$
Dentro de los grupos	SS_W	$n-k$	$S_W^2 = \frac{SS_w}{n-k}$	
Total	SS_{total}	$N-1$		

Donde:

$SS_{total} = suma\ de\ cuadrados\ total$

$SS_w = suma\ de\ cuadrados\ dentro\ de\ los\ grupos\ (W = within)$

$SS_B = suma\ de\ cuadrados\ entre\ los\ grupos\ (B = between)$

$k - 1 = grados\ de\ libertad\ entre\ grupos$

$n - k = grados\ de\ libertad\ dentro\ de\ los\ grupos$

$N - 1 = grados\ de\ libertad\ total$

$S_B^2 = cuadrados\ medios\ entre\ los\ grupos$

$S_w^2 = cuadrados\ medios\ dentro\ de\ los\ grupos$

$F_{obt} = valor\ estimado\ de\ F$

Cálculo de las sumas de cuadrados totales

$$SS_{total} = SS_B + SS_w = \sum_{i=1}^{n} (X_i - \bar{X}_G)^2$$

Cálculo de las sumas de cuadrados dentro de los grupos:

$$SS_w = \sum_{i=1}^{n} (X_i - \bar{X}_1)^2 + \sum_{i=1}^{n} (X_i - \bar{X}_2)^2 + \cdots + \sum_{i=1}^{n} (X_i - \bar{X}_k)^2$$

Cálculo de las sumas de cuadrados entre los grupos:

$$SS_B = n(\bar{X}_1 - \bar{X}_G)^2 + n(\bar{X}_2 - \bar{X}_G)^2 + \cdots + n(\bar{X}_k - \bar{X}_G)^2$$

Con estos resultados parciales sólo falta completar el resto de las operaciones, tal como se presentan en cuadro de análisis de varianza y tendremos el valor de F observado.

La regla de decisión se define como sigue:

Si $F_{obt} \geq F_{crit}$, rechazamos la H_0.

Si $F_{obt} < F_{crit}$, conservamos la H_0.

Tamaño del efecto

Una de las bondades de la prueba de ANOVA es que permite calcular un estimado del tamaño del efecto con los mismos elementos empleados para la construcción de la Fobt. El resultado viene dado por el calculo del coeficiente eta cuadrada o η^2. La fórmula se define como:

$$\eta^2 = \frac{SS_B}{SS_{total}}$$

El resultado es un valor entre 0 y 1 que se interpreta de la misma forma que el coeficiente de regresión estimado para los modelos de regresión lineal.

Cálculo de ANOVA de dos vías.

En términos básicos, el análisis de varianza de dos vías nos permite, en un experimento, evaluar los efectos de dos variables independientes y la interacción entre las mismas sobre las variaciones de una variable dependiente.

En el análisis de varianza de un factor, la suma total de los cuadrados se descompone en la suma de cuadrados dentro de los grupos y la suma de cuadrados entre los grupos o lo que es lo mismo, en suma de cuadrados por efectos atribuidos a la variable independiente y suma de cuadrados atribuidos a las variables no controladas y el error aleatorio.

La situación es muy similar en el análisis de varianza de dos factores. No obstante, en este caso, descomponemos la varianza total en cuatro componentes: a) la suma de cuadrados para la primera VI, b) la suma de cuadrados para la segunda VI, c) la suma de cuadrados para la interacción entre las dos VV II y d) la suma de cuadrados para el error.

En definitiva, la prueba F para el ANOVA de dos vías se reduce al cálculo de tres valores estimados de F_{obt} por medio de las siguientes fórmulas:

$$F_{obt} = \frac{S_R^2}{S_W^2}, \text{para los efectos principales del Factor A.}$$

$$F_{obt} = \frac{S_C^2}{S_W^2}, \text{para los efectos principales del Factor B.}$$

$$F_{obt} = \frac{S_{RC}^2}{S_W^2}, \text{para los efectos de interacción del Factor A con el Factor B.}$$

Para el cálculo respectivo de F_{obt}, se procede a elaborar una tabla que reúna la información necesaria, que se suele denominar "Tabla de resumen para el Análisis de varianza o ANOVA", con los resultados de las sumas de cuadrados, las medias cuadráticas, los grados de libertad y el valor de F, que adopta la siguiente forma:

Fuente de variación	Suma de cuadrados (SC)	Grados de libertad	Cuadrados medios (CM)	Valor de la función F
Factor A	SS_R	$r-1$	$S_r^2=\dfrac{SS_R}{r-1}$	$F_{obt}=\dfrac{S_r^2}{S_W^2}$
Factor B	SS_C	$c-1$	$S_C^2=\dfrac{SS_C}{c-1}$	$F_{obt}=\dfrac{S_C^2}{S_W^2}$
Interacción	SS_{rC}	$(r-1)(c-1)$	$S_{rC}^2=\dfrac{SS_{rC}}{(r-1)(c-1)}$	$F_{obt}=\dfrac{S_{RC}^2}{S_W^2}$
Dentro de los grupos	SS_W	$rc(n-1)$	$S_W^2=\dfrac{SS_w}{rc(n-1)}$	
Total	SS_{total}	$N-1$		

Donde:

$SS_{total} = suma\ de\ cuadrados\ total$

$SS_w = suma\ de\ cuadrados\ dentro\ de\ los\ grupos\ (W = within)$

$SS_r = suma\ de\ cuadrados\ entre\ las\ filas\ (r = rows)$

$SS_c = suma\ de\ cuadrados\ entre\ las\ columnas\ (c = columns)$

$SS_{rc} = suma\ de\ cuadrados\ de\ la\ interacción$

$r-1 = grados\ de\ libertad\ para\ las\ filas$

$c-1 = grados\ de\ libertad\ para\ las\ columnas$

$(r-1)(c-1) = grados\ de\ libertad\ para\ la\ interacción$

$rc(n-1) = grados\ de\ libertad\ dentro\ de\ los\ grupos$

$N-1 = grados\ de\ libertad\ total$

$S_r^2 = cuadrados\ medios\ entre\ las\ filas$

$S_c^2 = cuadrados\ medios\ entre\ las\ columnas$

$S_{rc}^2 = cuadrados\ medios\ de\ la\ interacción$

$S_w^2 = cuadrados\ medios\ dentro\ de\ los\ grupos$

$F_{obt} = valor\ estimado\ de\ F, según\ el\ caso$

Cálculo de las sumas de cuadrados totales

$$SS_{total} = \sum_{i=1}^{n} (X_i - \bar{X}_G)^2$$

Donde: $\bar{X}_G$ corresponde a la media general de la distribución de los valores de la VD.

Cálculo de las sumas de cuadrados entre las filas:

$$SS_r = \sum_{i=1}^{r}(X_{i.} - \bar{X}_r)^2 = \sum_{i=1}^{r=1}(X_{1.} - \bar{X}_1)^2 + \sum_{i=1}^{r=2}(X_{2.} - \bar{X}_2)^2 + \cdots + \sum_{i=1}^{r=k}(X_{k.} - \bar{X}_k)^2$$

Cálculo de las sumas de cuadrados entre las columnas:

$$SS_c = \sum_{j=1}^{c}\left(X_{.j} - \bar{X}_c\right)^2 = \sum_{j=1}^{c=1}(X_{.1} - \bar{X}_1)^2 + \sum_{j=1}^{c=2}(X_{.2} - \bar{X}_2)^2 + \cdots + \sum_{j=1}^{c=k}(X_{.k} - \bar{X}_k)^2$$

Cálculo de las sumas de cuadrados para la interacción

$$SS_{rc} = \left[\sum_{i=1}^{r}\sum_{j=1}^{c}\left(\bar{X}_{ij} - \bar{X}_G\right)^2\right] - SS_r - SS_c$$

$$= \left[\sum_{i=1}^{r=1}\sum_{j=1}^{c=1}(\bar{X}_{11} - \bar{X}_G)^2 + \sum_{i=1}^{r=1}\sum_{j=1}^{c=2}(\bar{X}_{12} - \bar{X}_G)^2 + \cdots + \sum_{i=1}^{r=n}\sum_{j=1}^{c=k}(\bar{X}_{nk} - \bar{X}_G)^2\right] - SS_r - SS_c$$

Cálculo de las sumas de cuadrados dentro de los grupos:

$$SS_w = \sum_{i=1}^{n}(X_i - \bar{X}_1)^2 + \sum_{i=1}^{n}(X_i - \bar{X}_2)^2 + \cdots + \sum_{i=1}^{n}(X_i - \bar{X}_k)^2$$

Con estos resultados parciales sólo falta completar el resto de las operaciones, tal como se presentan en cuadro de análisis de varianza y tendremos el valor de F observado para cada caso.

La regla de decisión general se define como sigue:

Si $F_{obt} \geq F_{crit}$, rechazamos la H_0.

Si $F_{obt} < F_{crit}$, conservamos la H_0.

Representación gráfica de los diseños factoriales.

De acuerdo con (León & Montero, 2003), dada la complejidad de las hipótesis puestas a prueba en los diseños factoriales, se requiere de dos clases de gráficos que permitan ilustrar los dos tipos de efectos que se pueden producir en este tipo de situaciones experimentales, a saber: los efectos principales y los de interacción. En este sentido, se recomienda el uso de gráficos de perfiles con las medias de las categorías generadas. El protocolo consiste en generar un plano bidimensional definido por el eje de las abscisas, en el cual se representarán los valores para una de las variables independientes y el eje de las ordenadas para representar la escala asociada a la medición de la variable dependiente. En este plano se presentarán las variaciones producto de la segunda variable independiente con el uso de líneas distintas para unir las medias de los grupos en función de los cambios en la primera variable independiente. Veamos el siguiente ejemplo para el caso de un diseño 2X2.

Dada la siguiente matriz de datos:

		Factor A		
		A1	*A2*	*total*
Factor B	*B1*	$\bar{X}_{A1B1}$	$\bar{X}_{A2B1}$	$\bar{X}_{B1}$
	B2	$\bar{X}_{A1B2}$	$\bar{X}_{A2B2}$	$\bar{X}_{B2}$
	total	$\bar{X}_{A1}$	$\bar{X}_{A2}$	$\bar{X}_{G}$

Debemos decidir:

1. Cuál es el factor que definirá los valores para el eje de las abscisas. Tomaremos como ejemplo al factor A para el eje de las abscisas.
2. Cómo se identificarán las líneas para cada valor de la segunda variable. Para el ejemplo, definimos líneas distintas para los cambios en el factor B como:

   ----- B1

   ___ B2

De este modo, generamos el plano bidimensional según el protocolo:

Variable Dependiente	x_n			Factor B
	…			B1 = ------
	x_3			B2= ———
	x_2			
	x_1			
		A1	A2	
		Factor A		

Los valores concretos de las medias por cada categoría A,B define un punto en el plano.

1.7.3. Contraste de hipótesis No paramétricos (para poblaciones no normales)

Más recientemente, una rama de la estadística inferencial ha desarrollado una serie de técnicas de análisis que no se basan en rigurosas suposiciones con respecto a la forma de las distribuciones de las poblaciones de las cuales se derivan las observaciones y el nivel de medición de los datos de investigación. Por ejemplo, las pruebas no paramétricas no dependen del supuesto de normalidad de las puntuaciones de la población. (Kerlinger & Lee, 2002). Estas técnicas se han denominado por analogía a las pruebas paramétricas como pruebas *No paramétricas* o *pruebas de distribución libre.* En estos casos, seremos capaces de generar conclusiones del tipo: “sin considerar la forma de la población, podemos concluir que…” (Siegel & Castellan, 1995, pág. 25).

Con base en las restricciones expuestas para el uso de las pruebas paramétricas, se ha generado una familia de pruebas que responden análogamente a los diseños experimentales fundamentales y reciben el calificativo común de pruebas no paramétricas. Las mismas se resumen en la siguiente tabla:

<table>
<tr><th>Número de grupos.</th><th>Nivel de independencia.</th><th colspan="2">Nivel de medición.</th><th>Prueba recomendada</th></tr>
<tr><td rowspan="3">Una muestra.</td><td rowspan="3">No aplica.</td><td rowspan="2">Nominal</td><td>Binario</td><td>Binomial.</td></tr>
<tr><td>Politómico</td><td>Ji cuadrada.</td></tr>
<tr><td colspan="2">Ordinal.</td><td>Z de Kolmogorov-Smirnov</td></tr>
<tr><td rowspan="5">Dos muestras</td><td rowspan="3">No relacionados</td><td rowspan="2">Nominal</td><td>Binario</td><td>Prueba exacta de Fisher</td></tr>
<tr><td>Politómico</td><td>Ji cuadrada.</td></tr>
<tr><td colspan="2">Ordinal</td><td>U de Mann-Whitney</td></tr>
<tr><td rowspan="2">Relacionados</td><td>Nominal.</td><td>Binario</td><td>Mc Nemar</td></tr>
<tr><td colspan="2">Ordinal</td><td>Signos y Wilcoxon.</td></tr>
<tr><td rowspan="4">K muestras</td><td rowspan="2">No relacionados</td><td colspan="2">Nominal</td><td>Ji cuadrada.</td></tr>
<tr><td colspan="2">Ordinal</td><td>Kruskal-Wallis y la Mediana</td></tr>
<tr><td rowspan="2">Relacionados</td><td colspan="2">Nominal.</td><td>Q de Cochran</td></tr>
<tr><td colspan="2">Ordinal</td><td>Prueba de rangos de Friedman</td></tr>
</table>

Por motivos de tiempo y espacio se omitirán los detalles con relación al cálculo de los estadísticos en cuestión. Sin embargo, los puntos a continuación se pasearán por cada una de estas propuestas y darán cuenta de los principales aportes de cada uno de estos paradigmas para el análisis de los datos.

1.7.3.1. Pruebas para una muestra

Prueba de chí cuadrado.

Cuando trabajamos con bases de datos cualitativas, la información que es capaz de proporcionarnos el estadístico estudiado, en los términos expuestos, hace referencia únicamente a la frecuencia o cantidad de individuos que cumplen con una determinada característica (simple o conjunta). En ningún caso se nos informa sobre la magnitud con la cual se presenta el carácter en cuestión. De modo que, un análisis de los datos daría lugar al tipo de pregunta siguiente: difieren de forma significativas las frecuencias observadas de las que se esperarían si se cumpliera la hipótesis de nulidad?

En cualquier caso, la prueba de ji cuadrado permite contrastar el comportamiento de las frecuencias observadas en una distribución con la distribución de las frecuencias esperadas teóricamente cuando se cumple la hipótesis nula. (Siegel & Castellan, 1995).

Bondad de ajuste: estructura de los datos y estimación de frecuencias esperadas.

Para comparar la distribución de unas frecuencias observadas contra aquellas esperadas teóricamente, debemos ser capaces de establecer qué frecuencias deben ser esperadas, cuando se cumple la hipótesis de nulidad. La hipótesis nula, en su expresión más ortodoxa, se comporta como una distribución uniforme y establece la equivalencia estadística de las frecuencias para cada una de las categorías de la variable. No obstante, otras aplicaciones del método pueden incluir valores teóricos específicos para las frecuencias esperadas en cada una de las categorías de la variable de estudio. En cualquier caso, la prueba de bondad de ajuste determinará la diferencia estadística entre la distribución observada y la teórica.

Su cálculo se resume a la siguiente ecuación:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

Donde:

χ^2= valor estimado de chi cuadrado para los datos observados.

O_i= el número de casos (frecuencia) observados en la iésima casilla.

E_i= el número de casos (frecuencia) observados en la iésima casilla, cuando la hipótesis nula es verdadera.

k= el número de categorías.

La significación estadística de este valor estimado de chi cuadrado puede ser determinada con el uso de tablas que resumen los valores críticos para distintas distribuciones muestrales que se presentan como anexos en la mayoría de los textos de estadística inferencial. La distribución de probabilidad para los valores estimados de chi cuadrado se comporta como una distribución chi cuadrada con k-1 grados de libertad.

Prueba binomial.

En ocasiones nos topamos con mediciones que obedecen a un sistema de categorización que se reduce a sólo dos valores, entonces nos referimos a estas observaciones como datos dicotómicos.

Observado el comportamiento de la distribución de datos dicotómicos, el investigador suele hacerse una pregunta: se están distribuyendo los datos de la misma forma en que lo harían según un determinado esquema teórico? En otras palabras, se busca contrastar si la proporción de aciertos observados se ajusta a la proporción teórica de una distribución binomial.

La prueba binomial permite esclarecer si una variable dicotómica sigue o no un determinado modelo de probabilidad. En concreto, permite contrastar la hipótesis de que la proporción observada de aciertos se ajusta a la proporción teórica de una distribución binomial. (Pardo & Ruíz, 2005).

El contraste de hipótesis binomial obedece al comportamiento teórico para la distribución muestral de porciones y para su cálculo se suele tomar dos rutas: a) para el caso de muestras pequeñas se define por la siguiente función de densidad: $p(x) = f(x) = \binom{n}{m} p^n q^{n-m}$, donde: $\binom{n}{m} = \frac{n!}{n!\,(n-m)!}$

Para el caso de muestras grandes, el contraste atiende al teorema de Laplace-DeMoivre y supone que

$$Bin(x) \sim N\left[np, \sqrt{npq}\right]$$

De este modo, cuando disponemos de n>30, la fórmula ajustada para el cálculo de porciones se define como:

$$z_{obt} = \frac{X_i - np}{\sqrt{npq}}$$

Donde X_i es el valor estimado de observaciones de aciertos, n es el número total de observaciones y p el valor de la proporción para la hipótesis nula (p= 0.5).

Prueba de Kolmogorov-Smirnov.

Al igual que las pruebas chí-cuadrado para una muestra y binomial, la prueba de Kolmogorov-Smirnov (K-S) para una muestra también es una prueba de bondad de ajuste que sirve: para contrastar la hipótesis nula de que la distribución de las variables se ajusta a una determinada distribución teórica de probabilidad. Pero a diferencia de las primeras, diseñadas para variables cualitativas, la prueba K-S responde al tratamiento de variables cuantitativas.

Para contrastar la hipótesis nula de bondad de ajuste, la prueba K-S se basa en la comparación de dos funciones de distribución (funciones de probabilidad acumulada): una empírica u observada y una teórica.

Para obtener la función de distribución empírica $F_{(xi)}$ se ordenan los valores de forma ascendente. Tras esto, la función de distribución empírica para cada valor de X_i se obtiene de la siguiente manera: $F_{(xi)} = i/n$ (donde i se refiere al rango correspondiente para cada observación).

Por su parte, la forma de obtener la función de distribución teórica $F_0(X_i)$ depende de la distribución concreta propuesta para la hipótesis, tales como: uniforme, normal, poison, etc.

Una vez obtenidas las distribuciones empírica y teórica, el estadístico de K-S se calcula a partir de la diferencia D, entendida como: $D = F(X_i) - F_0(X_i)$, más grande existente entre ambas: $Z_{K-S} = max|D_i|\sqrt{n}$

El estadístico Z se distribuye según el modelo de probabilidad normal $Z{\sim}N(0,1)$.

1.7.3.2. Pruebas para dos muestras independientes

Prueba de Chí cuadrado para dos muestras independientes.

En ocasiones, el investigador puede verse en la necesidad de responder a la pregunta: cuán diferentes son los comportamientos de las frecuencias observadas para una determinada variables entre una muestra de individuos agrupados en dos o más grupos? Dado que las medidas realizadas obedecen a un nivel de medición tan bajo como nominal.

La hipótesis que está siendo probada generalmente es aquella que postula que los grupos difieren respecto de alguna característica y, por tanto, respecto a la frecuencia relativa con que los miembros de los grupos caen dentro de alguna categoría. (Siegel & Castellan, 1995). Al igual que en el caso de la bondad de ajuste para una muestra, la esencia del contraste consiste en evaluar estadísticamente las diferencias entre las frecuencias observadas y las esperadas teóricamente. De hecho, su cálculo es una derivación de la fórmula para la bondad de ajuste cuando la distribución está definida por una tabla de doble entrada, tal como se presenta a continuación:

Su cálculo se resume a la siguiente ecuación:

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\left(n_{ij}-E_{ij}\right)^2}{E_{ij}}$$

Donde:

χ^2= valor estimado de chi cuadrado para los datos observados.

n_{ij}= el número de casos (frecuencia) que cumple con la doble condición observados en la i-ésima fila y j-ésima columna.

E_{ij}= el número de casos (frecuencia) que cumple con la doble condición observados en la i-ésima fila y j-ésima columna, cuando la hipótesis nula es verdadera.

r= el número de categorías por filas.

c= el número de categorías por columnas.

Dado que se trata de una tabla de doble entrada y bajo el supuesto de independencia para la hipótesis de nulidad, la frecuencia esperada por celda debería ser proporcional a la distribución total de filas y columnas. A diferencia del caso para la bondad de ajuste con una muestra, se calcula la frecuencia esperada en cada celda con la fórmula:

$$E_{ij} = \frac{n_{i}.n._{j}}{N}$$

Donde:

$n_{i\cdot} = \sum_{i=1}^{I} n_{ij}$ = el número total de casos (frecuencia) observados en la i-ésima fila, para todas las columnas.

$n_{\cdot j} = \sum_{j=1}^{I} n_{ij}$= el número total de casos (frecuencia) observados en la j-ésima columna, para todas las filas.

$N = n_{\cdot\cdot} = \sum_{i=1}^{I} \sum_{i=1}^{j} n_{ij}$= número total de observaciones realizadas.

La distribución de probabilidad para los valores estimados de chi cuadrado se comporta como una distribución chi cuadrada con (r-1)x(c-1) grados de libertad.

Prueba de la mediana para dos grupos independientes.

La prueba de la mediana es una prueba no paramétrica que podemos considerar un caso especial de la prueba de chi-cuadrado, pues se basa en esta última. Su objetivo es comparar las medianas de dos muestras y determinar si pertenecen a la misma población o no. Esta prueba permite contrastar la hipótesis acerca de si la diferencia observada en las medianas de los grupos es suficiente para asumir que los grupos provienen de poblaciones distintas. Esta prueba requiere que los datos se midan en al menos una escala ordinal.

Es una prueba bastante sencilla en su procedimiento. Se basa en el cálculo de la mediana de todos los datos conjuntamente. Después, se divide cada muestra en dos subgrupos: uno para aquellos datos que se sitúen por encima de la mediana y otro para los que se sitúen por debajo. La prueba de chi-cuadrado determinará si las frecuencias observadas en cada grupo difieren de las esperadas con respecto a una distribución de frecuencias que combine ambas muestras.

Prueba U de Mann-Whitney

La prueba U de Mann-Whitney funciona como una apreciable alternativa a la prueba t para diferencias de medias cuando no se cumplen los supuestos en los que se basa la misma, tales como nivel de medición de intervalo o superior, normalidad y homoscedasticidad. De acuerdo con (Pardo & Ruíz, 2005), la prueba U de Mann-Whitney fue propuesta originalmente por Wilcoxon (1945) para el caso de grupos de igual tamaño. Pero fueron Mann y Whitney (1947) los primeros en extender el procedimiento al caso de grupos de tamaños desiguales y los primeros también en proporcionar tablas para poder utilizarlo con muestras pequeñas. Fueron precisamente las aportaciones de Mann y Whitney las que más contribuyeron a la divulgación del procedimiento; de ahí, que generalmente, sea conocida como prueba de Mann-Whitney. Sin embargo, autores como (Siegel & Castellan, 1995) la siguen presentando como prueba de Wilcoxon-Mann-Whitney.

Para calcular el estadístico U se asigna a cada uno de los valores de las dos muestras su rango para construir los elementos de las siguientes ecuaciones:

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$$

Donde: n_1 y n_2 son los tamaños respectivos de cada muestra; R_1 y R_2 es la suma de los rangos de las observaciones de las muestras 1 y 2, respectivamente. El estadístico U se define como el mínimo de U_1 y U_2.

Dado que se asume que las dos muestras se han extraído de poblaciones idénticas, cabe esperar que U_1 y U_2 sean aproximadamente iguales, más allá del error aleatorio. Si los valores de U_1 y U_2 son muy disímiles, se asume cierta evidencia de que las muestras proceden de poblaciones distintas. Por consecuencia, la H0 de que ambos promedios son iguales podrá rechazarse si U_1 (o U_2) es demasiado grande o pequeño.

1.7.3.3. Pruebas para dos muestras relacionadas

Las pruebas de este apartado permiten analizar datos provenientes de diseños con medidas repetidas. Todas ellas se ajustan a los diseños del tipo pre-post. Sin embargo, debemos discriminar de este grupo de pruebas aquella que se adecúe al nivel de medición de nuestras variables, a saber: dicotómicas, politómicas u ordinales.

Prueba de cambios de McNemar

Una variable de los diseños de medidas repetidas consiste en medir una misma variable dicotómica, en dos momentos diferentes. La situación sería similar si en lugar de tomar dos medidas a los mismos sujetos se tomara una medida a pares de sujetos igualados en algún criterio. Esta situación es particularmente útil para evaluar el cambio entre dos momentos, cuando las opciones de respuesta se reducen a sólo dos opciones.

Este diseño permite contrastar la hipótesis de nulidad para la igualdad de porciones, entre dos medidas antes-después. La hipótesis de igualdad de porciones también puede plantearse en términos del número de cambios que se producen en una dirección determinada, la cual se complementa con los cambios en la otra opción. Para contrastar esta hipótesis, el estadístico de McNemar (1947) compara los cambios que se producen entre el anes y el después en ambas direcciones y determina la probabilidad asociada a ese número de cambios, asumiendo como hipótesis nula la igualdad de proporciones.

Para su cálculo definimos A como el número casos observados cuyas respuestas cambiaron de + a – y D como el número observado de casos que cambiaron de – a +, y (A + D)/2 como el número de casos esperados en las celdillas A y D. de esta manera, al desarrollar y reducir términos de la ecuación de chí-cuadrado para grupos independientes, se obtiene:

$$\chi^2 = \frac{(A-D)^2}{A+D}$$

Para la decisión del contraste tomamos como referencia que la distribución muestral de χ^2 cuando la hipótesis nula es verdadera, se distribuye asintóticamente como chí-cuadrada con un grado de libertad, dado que gl=(r-1)(c-1)= (2-1)(2-1)= 1.

Prueba de Wilcoxon

La prueba de los rangos con signo de Wilcoxon es una prueba no paramétrica para comparar la media de dos muestras relacionadas y determinar si existen diferencias entre ellas. Se utiliza como alternativa a la prueba t de Student para grupos relacionados cuando no se puede suponer la normalidad de dichas muestras o las medidas no alcanzan un nivel de medición de intervalo. Debe su nombre a Frank Wilcoxon, que la publicó en 1945. Se utiliza cuando la variable subyacente es continua pero no se presupone ningún tipo de distribución particular.

El estadístico de (Wilcoxon, 1945), se basa en ordenar de forma ascendente las diferencias absolutas D_i, donde $D_i = |X_i - Y_i|$ (i= 1, 2, …, n) entre cada par de observaciones (x,y). además, deberá señalar el signo correspondiente a cada diferencia observada.

La hipótesis de nulidad sostiene que los tratamientos X e Y son equivalentes, en otras palabras, muestras de la misma población, con la misma mediana y la misma distribución. Si H0 es verdadera, deberíamos encontrar un equilibrio en los rangos de las distribuciones de X e Y, por tanto, la suma de éstos sería equivalente, salvando por supuesto las fluctuaciones atribuibles al azar. De modo que: $P(X_i < Y_i) = P(X_i > Y_i)$. En el caso contrario, aceptaríamos que los grupos son diferentes y con estas evidencias nos aventuraríamos a rechazar la hipótesis nula.

1.7.3.4. Pruebas para varias muestras independientes

Prueba de la mediana para más de dos grupos independientes.

La extensión de la prueba de la mediana permite determinar si k grupos independientes (no necesariamente del mismo tamaño) se han extraído de la misma población o de poblaciones distintas, pero con medianas iguales. Esta prueba es particularmente útil para los casos en los cuales se viola el supuesto de normalidad por la presencia de valores atípicos o aberrantes en la distribución y cuando los datos se encuentran en una escala inferior a intervalo.

Para llevar a cabo la prueba de la mediana, se procede a construir una tabla de contingencia de 2 x k, con los números en el cuerpo de la tabla representando las frecuencias de signos más (+) cuando las puntuaciones están por encima de la mediana y menos (-) si las puntuaciones se ubican por debajo de la mediana, en cada uno de los k grupos.

Para probar la hipótesis nula de que las k muestras provienen de la misma población con respecto a las medianas, debemos calcular el valor estadístico de chí-cuadrado en los términos convencionales. La distribución de probabilidad de chí-cuadrado atenderá a la distribución con k-1 grados de libertad. Aunque se trata de una distribución de doble entrada, recordemos que: gl= (r-1)(k-1) = (2-1)(k-1)= k-1.

Prueba H de Kruskal-Wallis

La prueba de Mann-Whitney fue extendida al caso de más de dos muestras por William Kruskal y W. Allen Wallis (Kruskal y Wallis, 1952) es un método no paramétrico para probar si un grupo de datos proviene de la misma población. Intuitivamente, es idéntico al ANOVA con los datos reemplazados por categorías. Sólo en el caso en que se violen, significativamente, los supuestos de las pruebas paramétricas la H de Kruskal-Wallis resulta más potente que la prueba F.

Ya que es una prueba no paramétrica, la prueba de Kruskal-Wallis no asume normalidad en los datos, en oposición al tradicional ANOVA y el uso del estadístico F. además, permite trabajar con variables medidas en un nivel de medición ordinal. Sí asume, bajo la hipótesis nula, que los datos vienen de la misma distribución. Una forma común en que se viola este supuesto es con datos heterocedásticos.

La fórmula que resume el concepto y cálculo de la prueba se define como:

$$H = \frac{12}{N(N+1)} \sum_{j=1}^{k} n_j \left(\bar{R}_j - \bar{R}\right)^2$$

Bajo la hipótesis nula de que los *j* rangos promedio poblacionales son iguales, el estadístico H se distribuye según el modelo de probabilidad *chí*-cuadrado, con *j*-1 grados de libertad.

1.7.3.5. Pruebas para varias muestras relacionadas

Las pruebas que se revisarán en este apartado permiten analizar datos provenientes de diseños con múltiples medidas repetidas. Todas ellas se ajustan a diseños del tipo longitudinal. Sin embargo, debemos discriminar de este grupo de pruebas aquella que se adecúe a nuestros datos según el nivel de medición de nuestras variables, a saber: dicotómicas u ordinales. La prueba de Cochran (Cochran, 1950) permite contrastar la hipótesis de igualdad de proporciones para un número j de observaciones correspondientes a variables dicotómicas. Por su parte, la prueba de Friedman (Friedman, 1937) nos será de utilidad para el contraste de variables con nivel de medición ordinal.

Prueba Q de Cochran

El estudio de más de dos proporciones relacionadas es una generalización del procedimiento para el caso de dos proporciones relacionadas. Dado que la variable en estudio mantiene el estatus de medición nominal dicotómica, con la particularidad de que en este caso necesitamos comparar más de dos grupos de medidas relacionadas.

De acuerdo con los planteamiento de (Cochran, 1950) a *n* sujetos se les toma un número *j* de medidas de una variable dicotómica. Se trata en esencia de un diseño de ANOVA de una vía para medidas repetidas, para medidas dicotómicas.

En el trabajo clásico de (Cochran, 1950) el autor demuestra que, si la hipótesis nula es verdadera, en otras palabras, si no hay diferencias en el tipo de respuesta arrojada por el sujeto para cada condición, más allá de aquellas que se definen por el error aleatorio, las respuestas se distribuirán de forma aleatoria entre las filas y columnas de la tabla generada.

De acuerdo con (Cochran, 1950), típicamente, los datos están organizados en una tabla de contingencia o de doble entrada con r número de filas y c número de columnas, en la cual cada columna representa una muestra y cada fila (r) un grupo de observaciones. Se define el test criterio como:

$$Q = \frac{c(c-1)\sum_{j=1}^{c}\left(\bar{T}_j - \bar{T}\right)^2}{c(\sum_{i=1}^{n} u_i) - (\sum_{i=1}^{n} u_i^2)}$$

Donde T es el número total de sucesos en la j-ésima muestra (columna) y u_i el número total de sucesos en las filas. Si la probabilidad de sucesos es la misma en todas las muestras, el límite de la distribución de Q, cuando el número de filas es grande, es la distribución chí-cuadrado con c-1 grados de libertad. La relación entre esta prueba y la prueba chí-cuadrado ordinaria, está dada por los niveles de independencia de los grupos. (Cochran, 1950, pág. 266).

Prueba de rangos de Friedman

La prueba de rangos de (Friedman, 1937) sirve para comparar un número p de promedios poblacionales cuando se trabaja con muestras relacionadas. El diseño basa del cual se derivan las observaciones para la aplicación de este tipo de pruebas es el de múltiples medidas repetidas, una forma análoga a la prueba ANOVA de un factor con medidas repetidas, pero para variables con nivel de medida ordinal.

Las ventajas de este estadístico frente a la prueba de ANOVA son las mismas que se han señalado para el caso de la aplicación de la prueba de Kruskal-Wallis, dependerá de la violación de los supuestos requeridos para la aplicación de la prueba de ANOVA.

Se tiene de esta manera una base de datos conformada por un número *p* de muestras o tratamientos agrupados en columnas y por una muestra de *n* sujetos o bloques independientes entre sí e independientes de los tratamientos. Para el cálculo del estadístico de Friedman, se procede a transformar las puntuaciones originales a rangos R_{ij}. Con esto bastará para cumplir con los elementos requeridos en la fórmula fundamental para el cálculo del estadístico chí-cuadrado de Friedman, tal como lo presenta en su obra (Friedman, 1937, pág. 679) y que define como:

$$\chi_r^2 = \left[\frac{12}{np(p+1)}\sum_{j=1}^{p}\sum_{i=1}^{n} R_{ij}^2\right] - 3n(p+1)$$

El estadístico χ_r^2 de Friedman se distribuye en concordancia con la distribución de probabilidad de chí-cuadrado para p-1 grados de libertad.

Referencias

Abelson, R. (1998). *La estadística razonada: reglas y principios.* Barcelona: Paidós.

American Psychological Association. (01 de Junio de 2010). *Ethical Principles of Psychologists and Code of Conduct.* Recuperado el 13 de Marzo de 2013, de http://www.apa.org/ethics/code/index.aspx#

American Psychological Association. (2010). *Manual de Publicaciones de la Amercan Psychological Assciation* (Tercera ed.). (M. Guerra Frias, Trad.) D.F., México: Manual Moderno.

Anastasi, A., & Urbina, S. (1998). *Test Psicológicos.* D.F., México: Prentice Hall.

Aparicio M., G. (1985). *Teoría subjetiva de la probabilidad: fundamentos, evolución y determinación de probabildiades.* Recuperado el 01 de Marzo de 2013, de http://eprints.ucm.es/7818/1/01.pdf

Asamblea Nacional Constituyente. (30 de Diciembre de 1999). Constitución de la República Bolivariana de Venezuela. *Gaceta Oficial Nº 36.860*. Caracas, Venezuela: Asamblea Nacional.

Asamblea Nacional de la República Bolivariana de Venezuela. (12 de Diciembre de 2010). Reforma de Ley Orgánica de Ciencia y Tecnología. *Gaceta Oficial Nº 39.575*. Caracas, Venezuela: Asamblea Nacional.

Babbie, E. (1988). *Métodos de investigació por encuesta.* (J. Utrilla, Trad.) D.F., México: Fondo de Cultura Económica.

Balluerka, N., & Vergara, A. (2002). *Diseños de Investigación Experimental en Psicología.* Madrid, España: Prentice Hall.

Birnbaum, A. (1977). The Neyman-Pearson theory as decision theory, and as inference theory; with a criticism of the Lindley-Savage argument for Bayesian theory. *Synthese, 36*(1), 19-49.

Brown, C., & Ghiselli, E. (1969). *El método científico en Psicología.* (E. Prieto, Trad.) Buenos Aires, Argentina: Paidós.

Campbell, D., & Stanley, J. (1970). *Diseños experimentales y cuasi-experimentales en la investigación social.* Buenos Aires, Argentina: Amorrortu.

Campbell, N. (1921). *What is Science?* New York: Dover Publications.

Carnap, R. (1934). On the character of philosophic problems. *Philosophy of Science, 51*(1), 5-19.

Carnap, R. (1950). *Logical Foundations of Probability.* Chicago, U.S.A.: The University of Chicago Press.

Chávez, H. (01 de Noviembre de 2001). Decreto con fuerza de Ley de la función púiblica de la Estadística. Caracas, Venezuela: Gaceta Oficial Nº 37.321.

Cochran, W. (Diciembre de 1950). The comparison of percentages in matched samples. *Biometrika, 37*(3/4), 256-266.

Cohen, J. (1992). Cosas que he aprendido (hasta ahora). *Anales de Psicología, 8*(1-2), 3-17.

Comisión de Ética, Bioética y Biodiversidad. (Diciembre de 2010). Código de ética para la vida. (T. e. Ministerio del Poder Popular para la Ciencia, Ed.) Caracas, Venezuela: Ministerio del Poder Popular para la Ciencia, Tecnología e Industrias Intermedias.

Congreso de la Republica de Venezuela. (11 de Septiembre de 1978). Ley de Ejercicio de la Psicología. *Gaceta Oficial nº 2306*. Caracas, Venezuela.

Cowles, M., & Davis, C. (1982). On The Origins of .05 Level of Significance. *American Psychologist, 37*(5), 553-558.

Diez Calzada, J. A. (1992). En torno a la lógica de la inferencia. *Enrahonar: Quaderns de filosofía*, 91-97.

Dingle, H. (1950). A theory of measurement. *The British Journal for the Philosophy of Science*, 5 - 26.

Federación de Psicólogos de Venezuela. (1981). Código de Ética Profesional del Psicólogo de Venezuela. *II Asamblea Nacional Ordinaria de la Federación de Psicólogos de Venezuela.* Barquisimeto: Autor.

Fisher, R. (Julio de 1925). Theory of Statistical Estimation. *Mathematical Proceedings of the Cambridge Philosophical Society, 22*(5), 700-725.

Fisher, R. A. (1935). The Logic of Inductive Inference. *Journal of the Royal Statistical Society*, 39 - 82.

Friedman, M. (1937). The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance. *Journal of the American Statistical Association, 32*(200), 675-701.

Gaito, J. (1980). Measurement scales and statistics: resurgence of an old misconception. *Psychological Bulletin*, 564 - 567.

Glass, G., & Stanley, J. (1974). *Métodos Estadísticos Aplicados a las Ciencias Sociales.* (E. Galvis, & E. Guzman, Trads.) Madrid, España: Prentice Hall.

Grande, I., & Abascal, E. (1989). *Métodos de análisis multivariante para la investigación comercial.* Barcelona: Ariel.

Gujarati, D., & Porter, D. (2010). *Econometría* (Quinta ed.). (P. Carril Villareal, Trad.) D.F., México: Mc Graw Hill.

Gutiérrez, H., & de la Vara, R. (2012). *Análisis y Diseño de Experimentos* (Tercera ed.). D.F., México: Mc Graw Hill.

Hand, D. J. (1996). Statistics and the Theory of Measurement. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 445 - 492.

Hernández, R., Fernández, C., & Baptista, P. (2010). *Metodología de la Investigación* (Quinta ed.). D.F., México: Mc Graw Hill.

Kerlinger, F., & Lee, H. (2002). *Investigación del comportamiento. Métodos de investigación en ciencias sociales. (4a Ed).* México: Mc Graw Hill.

Knapp, T. (1990). Treating ordinal scales as interval scales: an attempt to resolve the controversy. *Nurse Research*, 121 - 123.

Kruskal, W., & Wallis, W. (Diciembre de 1952). Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association, 47*(260), 583-621.

Landau, D., & Lazarsfeld, P. (1978). Quetelet Adolphe. En W. Kruskal, & J. Tanur, *International Encyclopedia of Statistics* (Vol. 2, págs. 824-834). New York, USA: The Fee Press.

León, O., & Montero, I. (2003). *Métodos de Investigación en Psicología y Educación* (Tercera ed.). Madrid, España: Mc Graw Hill.

Lezama, L. (2011). Puntuaciones relacionadas con las normas. *Psicología*, 107-143.

López Casuso, R. (1996). *Cálculo de probabilidades e inferencia estadística* (Tercera ed.). Caracas, Venezuela: Publicaciones UCAB.

Lord, F. (1953). On the statistical treatment of football numbers. *The American Psychologist*, 750 - 751.

Luce, R. D. (1997). Quantification and symmetry: Commentary on Michell Quanttative Science and the definition of measurement in Psychology. *British Joumat of Psychology*, 395 - 398.

Magnusson, D. (1969). *Teoría de los test.* Trillas.

Magnusson, D. (1975). *Teoría de los test.* México: Biblioteca Técnica de Psicología.

Mann, H., & Whitney, D. (Marzo de 1947). On a test of wether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics, 18*(1), 50-60.

Marx, M., & Hillix, W. (1983). *Sistemas y teorías psicológicas contemporáneas* (3a ed.). Buenos Aires, Argentina: Paidós.

Mayo, D., & Cox, D. (2006). Frequentist statistics as a theory of inductive inference. *Lecture Notes-Monograph Series, 49*, 77-97.

Mayo, D., & Cox, D. R. (2006). Frequentist statistics as a theory of inductive inference. *Lecture Notes-Monograph Series*, 77-97.

McGuigan, F. (1977). *Psicología Experimental: enfoque metodológico* (Segunda ed.). (A. Fabre, Trad.) D.F., México: Trillas.

McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika, 12*(2), 153-157.

Michell, J. (1986). Measurement scales and statistics: a clash of paradigms. *Psychological Bulletin, 100*(3), 398 - 407.

Montgomery, D. C. (1984). *Design adn Análysis of Experiments* (Second ed.). New York, United States of América: John Wiley & Sons, Inc.

Morris, C., & Maisto, A. (2009). *Psicología* (13a ed.). D.F., México: Pearson Educación.

Muñíz, J. (1998). La medición de lo Psicológico. *Psicothema*, 1 - 21.

Narens, L., & Luce, R. D. (1986). Measurement: The Theory of Numerical Assignments. *Psychological Bulletin*, 166 - 180.

Navarro, A. (1989). *La Psicología y sus múltiples objetos de estudio.* Caracas, Venezuela: Consejo de Desarrollo Científico y Humanístico.

Neyman, J. (1977). Frequentist probability and frequentist statistics. *Synthese, 36*, 97-131.

Neyman, J., & Pearson, E. (1928). On the use and interpretation of certain test criteria for purposes of statistical inference: part I. *Biometrika, 20*(1/2), 175-240.

Pagano, R. (2011). *Estadística para las ciencias del comportamiento.* (M. Torres, Trad.) D.F., México: CENGAGE Learning.

Pardo, A., & Ruíz, M. (2005). *Análisis de datos con SPSS 13 Base.* Madrid, España: Mc Graw Hill.

Popper, K. (1962). *La Lógica de la Investigación Científica.* Madrid, España: Tecnos.

Recalde, L. C. (2009). Los axiomas de la cantidad de Hölder y la fundamentación del continuo lineal. *Matemáticas: Enseñanza Universitaria*, 101 - 121.

Reichenbach, H. (1949). *The Theiry of Probability.* Los Angeles, U.S.A.: University of California Press.

Restrepo, L., & Gonzalez, J. (2003). La Historia de la Probabilidad. *Revista Colombiana de Ciencias Pecuarias*, 83 - 87.

Rivadulla, A. (1991). *Probabilidad e inferencia científica.* Barcelona: Anthropos.

Saavedra, N. (2000). La axiomática de Kolmogorov: fundamentos de la teoría de la probabilidad. *Números, 43*, 185-190.

Sáchez Carrion, J. (2001). Estadística, orden natural y orden social. *Papers*, 33 - 46.

Sáiz Roca, M., de la Casa Rivas, G., Dolores Saíz, L., Ruiz, G., & Sánchez, N. (2009). Fundación y establecimiento de la Psicología Científica. En M. Sáiz Roca, *Historia de la Psicología* (págs. 55 - 150). Barcelona: UOC.

Sáiz, M. (2009). Los tiempos de reacción. La ecuación personal y el impulso nervioso. En M. Sáiz Roca, *Historia de la Psicología* (págs. 43 - 46). Barcelona: UOC.

Sáiz, M., & Sáiz, D. (2009). La Psicología Científica Británica. En M. Sáiz Roca, *Historia de la Psicología* (págs. 97 - 113). Barcelona: UOC.

Siegel, S., & Castellan, N. (1995). *Estadística No Paramétrica* (Cuarta ed.). (L. Aragón, & L. Fierros, Trads.) D.F., México: Trillas.

Sojo, V. (2004). *Ética en Investigación Psicológica con Humanos.* Manuscrito No publicado, Universidad Central de Venezuela, Escuela de Psicología, Caracas.

Stahl, S. (2006). The evolution of the normal distribution. *Mathematics Magazine*, 96 - 113.

Stevens, S. (Abril de 1935). The Operational Basis of Psychology. *The American Journal of Psychology, 47*(2), 323-330.

Stevens, S. (1935). The operational definition of psychological concepts. *Psychological Review, 42*(6), 517-527.

Stevens, S. (1946). On the theory of scales of measurement. *Science*, 677 - 680.

Student. (Marzo de 1908). The Probable Error of a Mean. *Biometrika, 6*(1), 1-25.

Thomas, H. (1982). IQ interval scales, and normal distribution. *Psychological Bulletin*, 198 - 202.

Vargas Sabadías, A. (1995). *Estadística Descriptiva e Inferencial.* Publicaciones de Universidad de Castilla-La Mancha.

Velleman, P., & Wilkinson, L. (1993). Nominal, ordinal, interval and ratio typologies are misleading. *The American Statistician*, 65 - 72.

Walker, H. (1978). Pearson, Karl. En W. Kruskal, & J. Tanur, *International Encyclopedia of Statistics* (Vol. 2, págs. 691-698). New York, U.S.A.: The Free Press.

Wiener, P. (1978). Peirce, Charles Sanders. En W. Kruskal, & J. Tanur, *International Encyclopedia of Statistics* (Vol. 2, págs. 698-702). New York, U.S.A.: The Free Press.

Wilcoxon, F. (Diciembre de 1945). Individual comparisons by ranking methods. *Biometrics Bulletin, 1*(6), 80-83.