



2018

ESTADÍSTICA NO PARAMÉTRICA

Francisco Sánchez Villarreal
FACULTAD DE CIENCIAS, UNAM

INTRODUCCION

Tradicionalmente se suele clasificar a la estadística en descriptiva e inferencial. La primera se socia a la presentación de tablas gráficas y mediciones que describen en forma general el comportamiento de una muestra o una población. La inferencia estadística por otra parte es asociada con la labor de investigación científica en la búsqueda de la relaciones causales u asociaciones entre fenómenos diversos. En sus inicios, la investigación y desarrollo en torno a la estadística inferencial se orientaron a los métodos llamados “paramétricos”, tales métodos de estimación y procedimientos de pruebas de hipótesis se apoyan a una serie de supuestos para su aplicación y correctas conclusiones. Los científicos especialistas en ciencias que requieren del uso de inferencia estadística, pero cuyas observaciones no siempre cumplen con los supuestos necesarios para la aplicación de las pruebas clásicas, tales como la normalidad la homogeneidad de varianzas y la exactitud de las mediciones, motivaron a la investigación de estadísticos matemáticos en el terreno de procedimientos inferenciales que no plantearan tantos requisitos y que sin embargo, coadyuvarán a conclusiones científicamente válidas.

La estadística no paramétrica se desarrolló rápidamente debido a que sus procedimientos cumplen en general con las siguientes ventajas:

- Los métodos no paramétricos hacen pocos supuestos acerca de las poblaciones a partir de las cuales se obtienen las observaciones.
- En general, son fáciles de aplicar y los paquetes de análisis estadístico más populares incluyen opciones no paramétricas en sus menús.
- Se pueden aplicar en situaciones donde los métodos tradicionales no son factibles de aplicación, tales como aquellas en las que las observaciones no son mediciones en el sentido preciso de la palabra, sino solamente rangos.

El desarrollo de la inferencia estadística no paramétrica, cuyo nombre más adecuado sería inferencia estadística de distribución libre, se inició vigorosamente por la tercera década del siglo pasada. R. Savage señala el año de 1936 como el punto de arranque debido a la publicación de un artículo sobre correlación por rangos debido a Hotelling y Pabst. Otros autores como Scheffé identifican sus antecedentes en fechas anteriores. En el período de 1900 a 1911 se publican algunos artículos debidos a Karl Pearson y años después se presenta la prueba de los signos en la primera edición de Statistical Methods for Research Workers (1925) escrita por Ronald A. Fisher. Por el año de 1943 Scheffé escribía: “Solamente una pequeña fracción de la extensa literatura dedicada a la estadística matemática está orientada al caso no paramétrico y la mayor parte ha sido escrita en esta década. Esperamos, sin embargo, un rápido crecimiento para esta rama. Las perspectivas de una teoría libre de supuestos específicos acerca de la forma de las distribuciones de las poblaciones debe excitar la mente tanto de teóricos como prácticos, ya que tal teoría combina la elegancia estructural con amplias posibilidades de aplicación.”

En 1947 F. Wilcoxon promueve los métodos estadísticos no paramétricos con la publicación de un libro titulado Some Rapid Approximate Statistical Procedures,. Para el año siguiente M.G. Kendall publica el primer libro orientado a métodos apoyados en rangos: Rank Correlation Methods. Kendall cita las posibilidades de aplicación de tales métodos en campo tan diversos como psicología, educación, economía y experimentación industrial. Por 1956 Sidney Siegel publica el que posiblemente sea el texto más popular de métodos estadísticos no paramétricos de la actualidad: Nonparametric Statistical for the Behavioral Sciences.

Para 1962 el impulso de la estadística no paramétrica es tal que Savage publica sus Bibliography of Nonparametric Statistical con referencia a más de 3000 artículos relacionadas con el tema.

J. Hájek uno de los más notables investigadores de esta rama de la estadística escribe en 1969 “Los métodos no paramétricos constituyen una de las ramas más exitosas de la estadística moderna. Ellos son ampliamente aplicables, de rápida ejecución y fáciles de entender. La teoría no paramétrica es elegante y dos décadas de vigoroso desarrollo no han agotado sus posibilidades de estimular a los investigadores”. En 2002 esta afirmación tiene plena validez

Medición

Se le atribuyen al científico escocés del siglo XIX, Lord Kelvin las siguientes palabras: “Cuando uno puede medir aquello de lo que está hablando y expresarlo en números, sabe algo acerca de ello; pero cuando no puede medirlo, cuando no puede expresarlo en números sus conocimientos es escaso e insatisfactorio; podrá ser un principio de conocimiento a la etapa de una ciencia”. Esta puede ser una afirmación exagerada, pero ilustra claramente la importancia que la medición ha tenido para los científicos.

En términos muy simplistas decimos que medimos cuando somos capaces de asociar números a objetos y eventos de acuerdo a ciertas reglas. El problema reside entonces en determinar claramente las reglas de asignación de números a los objetos y fenómenos, las propiedades matemáticas de los grupos a estructuras de las escalas resultantes y el tipo de operaciones estadística aplicables a las mediciones efectuadas con cada tipo de escala.

Las escalas son posibles en la medida en que guardan un isomorfismo entre lo que podemos hacer con las características de los objetos y las propiedades de los números.

S.S. Stevens publicó en 1946 en la revista Science un artículo titulado On the theory of scales of measurement en donde propone una clasificación que contempla cuatro tipos básicos de mediciones:

- Nominal
- Ordinal
- Intervalar
- Razón

La escala nominal representa la menos restringida asignación de números. Los números asignados a los objetos y fenómenos solamente representan etiquetas y cualquier otro tipo de símbolo podría servir para el mismo fin, como por ejemplo, las letras. La identificación de las entidades federativas de México mediante los números 1 al 32 ilustran la aplicación de una escala de este tipo, la posterior identificación de las personas de acuerdo al lugar de nacimiento por algún número del 1 al 30 no tiene una intención cuantitativa. Lo mismo podrá decirse de los números que identifican a los 11 jugadores de un equipo de fútbol. Si en la clasificación de un grupo de empleados por el turno en que trabajan, se designa a los pertenecientes al turno matutino con 1 al turno vespertino con 2, el valor del número no refleja magnitud de la característica, sino tan solo un medio de diferenciación.

En la escala nominal la operación de escalamiento consiste en partir de una clase dada y formar un conjunto de subclases que se excluyen mutuamente. La única relación válida es la equivalencia. Los miembros de cualquier subclase deben ser equivalentes en la propiedad medida. La relación de equivalencia es reflexiva, simétrica y transitiva.

Es claro que no tendría sentido obtener el promedio de los números que identifican a un grupo de personas de acuerdo al lugar de nacimiento, hacerlo y obtener un valor como 16.3 no significa que en promedio la gente nació cerca de Michoacán. La única operación válida con esos números es el conteo de las frecuencias absolutas y posteriormente la obtención de frecuencias relativas. Con la disposición de frecuencias es válida la identificación del valor que aparece con mayor frecuencia y la realización de pruebas cuyos datos de ingreso son frecuencias como aquellas basadas en la Binomial, Multinomial o Ji cuadrada.

La escala ordinal está asociada a la operación de ordenamiento por rango. Entre los ejemplos clásicos de este tipo se pueden referir: la escala de dureza de los minerales; la escala de medición de la magnitud de temblores de tierra propuesta por Mecalli, la cual solamente descubre en forma gruesa los efectos percibidos por la gente; las escalas de preferencia de productos que utilizan los investigadores de mercados en el público.

La diferencia entre la escala nominal y la ordinal, es que ésta incorpora además de una relación de equivalencia un a relación de orden. Entre tres elementos arbitrarios A, B, y C; es posible decir $A > B$ y a su vez $B > C$. La escala ordinal tiene una estructura llamada isotónica porque preserva el orden de un grupo.

Las estadísticas válidas en el nivel ordinal son las cuantiles (cuartiles, deciles, percentiles) y la medida de tendencia central que la caracteriza es la mediana. No es estrictamente válido el cálculo de las medias y desviación estándares, pues estas estadísticas suponen el conocimiento de algo más que el orden relativo de los datos. Entre las pruebas estadísticas que se basan en este tipo de escala se pueden mencionar las asociadas a coeficientes de correlación por rangos, como el de Spearman y el de Kendall; pruebas como la de la Mediana y la U de ;Mann y Whitney.

La escala intervalar corresponde a la cuantificación en el sentido común de la palabra. Casi todas las estadísticas son aplicables en este caso, a menos que se requiera el conocimiento de un punto cero verdadero. El cero es una escala intercalar es motivo de una convención arbitraria, como en el caso de la medición de la temperatura en grados centígrados, donde el cero se asocia al punto de fusión del hielo en condiciones normales, pero no a la ausencia total de calor en el concepto físico del término.

La diferencia entre las escala ordinal y la escala intervalar es que en la primera, dados tres elementos que ocupan puntos consecutivos de la escala $A > B > C$ no necesariamente presentan proporcionalidad en las magnitudes de la característica. La diferencia entre A y B no necesariamente es la misma que entre B y C, sin embargo, en la escala intervalar esta proporcionalidad si se preserva. La diferencia entre 23 grados y 24 grados es de un grado, lo mismo que entre 5 y 6 grados centígrados.

Un valor x medido en una escala intervalar se puede transformar en otro valor x' de otra escala por medio de una relación de la forma $x' = a + bx$. Nuestras escalas para medir el tiempo bajo diferentes calendarios son ejemplos clásicos de este tipo

Valores medidos en escala intervalar son susceptibles de cálculo para todo tipo de estadísticas, como media aritmética, media geométrica, varianza, desviación estándar, coeficiente de correlación producto momento de Perraño, etc. Y aquellas pruebas basadas en estas estadísticas, siempre y cuando se cumplan los supuestos adicionales de comportamiento estadístico de las mediciones.

Las escalas de razón se relacionan de manera más frecuentes con las magnitudes físicas tales como distancia, peso, resistencia eléctrica; también llamadas fundamentales u derivadas como densidad, fuerza y elasticidad.

En la escala de razón el cero significa ausencia de la característica, por ejemplo la temperatura medida en grados Kelvin o de escala absoluta asociada el cero a lo que en grados centígrados corresponde a -273 y que de acuerdo con la física, a tal temperatura las moléculas de cualquier elemento dejan de moverse, esto es hay una ausencia absoluta de calor.

La transformación de una unidad x en otra unidad x' medidas ambas en escala de razón se efectúa mediante una relación de proporcionalidad de la forma $x' = ax$. Para convertir yardas a metros multiplicamos por la constante aproximada 0.914 .

Todas las operaciones aritméticas tienen cabida en la escala de razón y en consecuencia se calculan medias, varianzas, potencias, etc. Las pruebas estadísticas son aplicables en la medida en que las observaciones cumplen supuestos probabilísticos.

S.S. Stevens presenta la siguiente tabla resumen de las propiedades de las escalas de medición:

Escala	Observaciones Válidas	Estructura Matemática	Estadísticas Permisibles
Nominal	Determinación de equivalencia	Grupo de permutación $x' = f(x)$ $f(x)$ es uno a uno	Número de casos Moda
Ordinal	Determinación de mayor a menor	Grupo isotónico $x' = f(x)$ Donde $f(x)$ es una función monótona creciente	Mediana Percentiles
Intervalar	Determinación de igualdad de intervalos	Grupo lineal general $x' = a + bx'$	Media Desviación estándar Correlación producto momento
Razón	Determinación de igualdad de razones	Grupo similar $x' = ax$	Coeficiente de variación.

Prueba de los Signos

I. Introducción

Esta prueba, junto con la de Wilcoxon se utiliza para probar hipótesis que involucran a dos grupos correlacionados, como es el caso de los experimentos en los cuales cada sujeto es su propio control. Para este tipo de experimentos existen alternativas no paramétricas más poderosas, como la prueba de Wilcoxon, pero esta prueba requiere un nivel de medición en la variable observada que permita establecer una magnitud relativa entre las diferencias de las diferentes parejas.

La prueba de los signos en cambio, no hace ningún supuesto a cerca de la forma de la distribución de las diferencias ni pide que todos los sujetos se tomen de la misma población. Las parejas pueden incluso, provenir de distintas poblaciones con respecto a edad, sexo, etc... El único requisito es que el investigador haya logrado en cada pareja un nivel de medición ordinal.

Se supone que la muestra se compone de n parejas relacionadas $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$. A cada pareja se le asocia un signo (+) si el valor de X_i sobrepasa al correspondiente Y_i , un signo (-) si sucede lo contrario y no se asocia ningún signo (de hecho se elimina la pareja) cuando $X_i = Y_i$. Los supuestos básicos de la prueba son los siguientes:

- a) Las parejas (X_i, Y_i) , $1, 2, \dots, n$ son mutuamente independientes
- b) La escala de medición es ordinal en cada pareja
- c) Las parejas son internamente consistentes, en el sentido de que si $P(+)>P(-)$ para una pareja, entonces $P(+)>P(-)$ para todas las parejas. La misma afirmación sería válida en los casos $P(+)<P(-)$ y $P(+)=P(-)$.

II. La Hipótesis

La hipótesis nula que se plantea en la prueba de los signos tiene la siguiente expresión:

$$H_0: P(X_i > Y_i) = P(X_i < Y_i) = \frac{1}{2}$$

donde X_i es el valor observado antes del tratamiento (o después del tratamiento), Y_i es el valor observado después del tratamiento (o antes del tratamiento). En otro sentido la hipótesis nula expresa que la diferencia de las medianas de las dos condiciones es cero.

La hipótesis alternativa para pruebas de una cola pueden ser:

$$H_1: P(X_i > Y_i) > P(X_i < Y_i)$$

ó bien

$$H_1: P(X_i > Y_i) < P(X_i < Y_i)$$

En el caso de proponer una hipótesis alternativa de dos colas, ésta tiene la siguiente forma:

$$H_1: P(X_i > Y_i) \neq P(X_i < Y_i)$$

III. Estadística de Prueba

Se ha mencionado que en la aplicación de la prueba de los signos se asocia un signo (+) si en la pareja observada el valor X_i sobrepasa a su correspondiente Y_i ; se asocia un signo (-) si sucede lo contrario; las parejas con valores iguales y que por lo tanto carecen de signo son eliminadas de la muestra.

De acuerdo con la hipótesis nula se espera que el número de signos (+) sea muy parecido al número de signos (-). Es decir, si H_0 fuera verdadera aproximadamente la mitad de las parejas tendría signo (+) y la otra mitad signo (-). La hipótesis nula H_0 se rechaza cuando ocurre muy pocas parejas de determinado signo.

La probabilidad asociada con la ocurrencia de un número particular de signos (+) y (-) puede calcularse haciendo uso de la distribución binomial. Se establecen los valores de los parámetros en la siguiente forma $p = \frac{1}{2}$ y n = número de parejas en la muestra. Como es sabido existen varias versiones de tablas de valores calculados para la distribución Binomial. El empleo de estas tablas es conveniente sobre todo con valores $n \leq 25$.

Cuando se usan tablas acumulativas de la Binomial, el argumento x se toma como el número menor de signos en el cálculo de la probabilidad asociada a pruebas de una sola cola, en caso de tener alternativas de dos colas, esa probabilidad debe duplicarse. El cálculo de la probabilidad mencionada se realiza mediante la siguiente expresión:

$$\Pr = \sum_{t=0}^x \binom{n}{t} \left(\frac{1}{2}\right)^n$$

donde

$$\binom{n}{t} = \frac{n!}{(n-t)!t!}$$

Cuando $n > 25$ en el cálculo de la expresión anterior se puede emplear la aproximación Normal a la distribución Binomial.

La estadística z , correspondiente a una variable aleatoria que se distribuye Normal estándar, se presenta a continuación

$$Z = \frac{x - np}{\sqrt{np(1-p)}} \quad \text{o bien al incluir corrección por continuidad} \quad Z = \frac{x \pm 0.5 - np}{\sqrt{np(1-p)}}$$

El valor 0.5 se suma si $x < np$ o se resta en caso contrario.

Como $p = \frac{1}{2}$ se tendrá

$$Z = \frac{x - \frac{n}{2}}{\sqrt{\frac{n}{4}}} \quad \text{o con la corrección} \quad Z = \frac{x \pm 0.5 - \frac{n}{2}}{\sqrt{\frac{n}{4}}}$$

Este valor se compara con tablas de Normal estándar para obtener la probabilidad asociada.

IV. Ejemplo

En un estudio de los efectos de la ausencia del padre sobre el desarrollo de los hijos, 17 matrimonios que habían sido separados por la guerra y cuyo primer hijo nació durante la ausencia del padre, fueron entrevistados individualmente. A cada uno de los cónyuges se le pidió que examinara diversos temas concernientes al hijo que había pasado su primer año sin padre, referentes a las relaciones disciplinarias de padre a hijo durante los años que siguieron a su regreso. Estas declaraciones fueron extraídas de grabaciones de las entrevistas, y un psicólogo, que conocía a cada familia, fué llamado para que estimara las declaraciones en cuanto al grado de conocimiento exhibido por cada padre acerca de la esencia de la disciplina paterna. La predicción suponía que la madre, debido a una unión más prolongada y cerca con el niño y a diversas circunstancias, típicamente asociadas con la separación del padre a causa de la guerra, tendría mayor conocimiento de las relaciones disciplinarias de su esposo con su hijo que él mismo.

Ho: La mediana de las diferencias es cero.

El número de esposas cuyo conocimiento de sus propias relaciones disciplinarias con sus hijos es mayor que el de sus esposos es igual al de esposas cuyo conocimiento de las relaciones disciplinarias es menor que el de sus esposos, son similares.

H1: La mediana de las diferencias es positiva

El conocimiento de las relaciones disciplinarias de la esposa con sus hijos es mayor en la mayoría de los casos.

Se selecciona un nivel de significancia del 5%

Los psicólogos midieron el conocimiento de la relación disciplinaria con una escala ordinal de uno a cinco en la que uno significaba gran conocimiento y cinco muy escaso o nulo. Los resultados se concentran en la siguiente tabla

Pareja	Madre	Padre	Signo	Nota
1	4	2	+	
2	4	3	+	
3	5	3	+	
4	5	3	+	
5	3	3	0	Se anula
6	2	3	-	
7	5	3	+	
8	3	3	0	
9	1	2	-	
10	5	3	+	
11	5	2	+	
12	5	2	+	
13	4	5	-	
14	5	2	+	
15	5	5	0	
16	5	3	+	
17	5	1	+	

Note que la muestra original de 17 parejas se eliminan 3 por tener el mismo valor, por ello el cálculo se hará con n=14 y para x=3, pues el signo (-) fue el menos frecuente.

Procediendo al cálculo por el método exacto:

$$\begin{aligned}
 \text{Pr} &= \sum_{t=0}^3 \binom{14}{t} \left(\frac{1}{2}\right)^{14} \\
 &= (1 + 14 + 91 + 364) \frac{1}{16,384} \\
 &= 0.0286865
 \end{aligned}$$

Como Pr es menor a 0.05 la hipótesis nula se rechaza. Entoces hay suficiente evidencia para decir que las esposas conocen mejor las relaciones disciplinarias de sus esposos con sus hijos.

Prueba de Wilcoxon

I. Introducción.

La prueba de Wilcoxon, al igual que la prueba de los signos, constituyen las alternativas no paramétricas para probar hipótesis referentes a dos grupos correlacionados, como sucede en los experimentos que plantean condiciones “antes” y “después” para el mismo grupo de sujetos. La prueba de Wilcoxon resulta de gran utilidad cuando el investigador tiene dudas acerca de la distribución de la variable observada y considera que la normalidad no se cumple razonablemente, requisito establecido para el empleo de la prueba de t para dos grupos relacionados.

La prueba de los signos y la prueba de Wilcoxon utilizan información referente a las diferencias de las parejas de observaciones relacionadas. Sin embargo, la prueba de los signos utiliza solamente información acerca de la dirección de las diferencias. Si el nivel de medición de la variable permite agregar la magnitud relativa, de tal suerte que se pueda dar mayor peso a una diferencia grande que a una pequeña, entonces, la prueba de Wilcoxon constituye una prueba mas poderosa.

Se considera que la muestra se compone de n parejas relacionadas $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$. A cada pareja se le asocia su diferencia $D_i = X_i - Y_i$. Para las variables aleatorias D_i se tienen los siguientes supuestos:

- a) Las D_i son variables aleatorias continuas
- b) Son mutuamente independientes
- c) Tienen la misma mediana
- d) La distribución de las D_i es simétrica

II. La Hipótesis

Si se denota por M_d la mediana común de las D_i , es posible establecer la hipótesis nula de forma simple o compuesta, de acuerdo al deseo de investigador de hacer una prueba de una cola o una prueba de dos colas.

Para el caso de pruebas de una sola cola se tiene

$$\begin{array}{lll} \text{Ho: } M_d \leq 0 & \text{vs} & \text{Hi: } M_d > 0 \\ \text{o bien} & & \\ \text{Ho: } M_d \geq 0 & \text{vs} & \text{Hi: } M_d < 0 \end{array}$$

Cuando se desea una prueba de dos colas las hipótesis nula y alternativa se definen como sigue:

$$\text{Ho: } M_d = 0 \quad \text{vs} \quad \text{Hi: } M_d \neq 0$$

III. La Estadística de Prueba

La estadística de prueba se denotará por T y se define como la suma de los rangos R_i asignadas aquellas parejas (X_i, Y_i) con el signo menos frecuente

$$T = \sum_{i=1} R_i$$

Los valores de T con diferentes tamaños de muestra y niveles de significancia para pruebas de una o dos colas fueron tabulados por Wilcoxon.

Cuando el tamaño de la muestra n es mayor de 25, la tabla anexa no se puede usar. Sin embargo, la distribución de la suma de los rangos de T para tales casos se aproxima a la normal. Para estandarizar se consideran las siguientes expresiones

$$U_T = \frac{n(n+1)}{4}$$

$$\sigma_T = \frac{\sqrt{n(n+1)(2n+1)}}{\sqrt{24}}$$

de donde

$$Z = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$$

IV. Ejemplo (1)

Un psicólogo desea comprobar si la asistencia al jardín de niños tiene algún efecto en la capacidad de percepción social cuenta con un instrumento cuyo nivel de medición se sitúa en el nivel ordinal.

En el experimento se contó además con 8 pares de gemelos idénticos. Un gemelo de cada par seleccionado al azar fue asignado al jardín por un tiempo, el otro permaneció fuera de la escuela. Al final del período los 16 niños se sometieron al instrumento de medición de la percepción social.

La hipótesis nula es Ho: la percepción social en los niños asignados al jardín es igual a la de los niños que se quedaron en casa.

La hipótesis alternativa Hi: la percepción social de los niños asignados al jardín es diferente de la de los niños que se quedaron en casa.

Los datos y cálculos se resumen en el siguiente cuadro

Puntaje de Niños asignados al Jardín	Puntaje de Niños asignados a su Casa	Diferencias	Rango de Diferencias	Rango de Signos Menos Frecuente
82	63	19	7	1
69	42	27	8	
73	74	-1	-1	
43	37	6	4	
58	51	7	5	
56	43	13	6	3
76	80	-4	-3	
65	62	3	2	
				T=4

Nota: Al tomar el rango de la Diferencia se toman los valores en valor absoluto, es por ello que el número 2 se le asigna a la diferencia 3.

Al consultar la tabla se observa que para N=8 una T=4 permite rechazar Ho con un nivel de $\alpha=0.05$.

Ejemplo (2)

A 28 parejas de gemelos se les sometió a una prueba psicológica para determinar si el gemelo que nació primero tendía a ser más agresivo que el otro.

Ho: la agresividad no esta relacionada con el orden en el nacimiento.

H1: la agresividad de los gemelos que nacieron primero es mayor.

Se toma un nivel $\alpha=0.05$.

Gemelo 1 86, 71, 77, 68, 91, 72, 77, 91, 70, 71, 88, 87, 68, 92
 72, 77, 64, 65, 85, 70, 87, 65, 71, 60, 75, 92, 70, 66

Gemelo 2 88, 77, 76, 64, 96, 72, 65, 90, 65, 80, 81, 72, 65, 90
 76, 75, 64, 70, 87, 76, 86, 64, 76, 60, 63, 91, 76, 72

Al relacionar el cálculo de T de forma semejante al primer ejemplo (promediando los rangos de las observaciones de igual valor) se obtiene T=166.5 para el signo (-), después de eliminar 3 observaciones ligadas queda n = 25 y al estandarizar se obtiene el valor de Z:

$$Z = \frac{166.5 - \frac{25(26)}{4}}{\sqrt{\frac{25(26)(51)}{24}}}$$

$$Z = 0.10762764$$

Como el valor de Z es menor a 1.64 Ho no se rechaza la hipótesis nula.

Prueba de U de Mann-Whitney

I. Introducción

Con frecuencia se pretende comparar dos muestras independientes y las condiciones de los datos no permiten hacer los supuestos de normalidad e igualdad de varianzas que requiere para su aplicación, la prueba t en muestras pequeñas ($n < 30$). Sin embargo, si se ha logrado al menos un nivel ordinal en los datos, la prueba U de Mann-Whitney resulta una excelente alternativa.

Hodges y Lehmon mencionan que al menos teóricamente, la potencia de la prueba U con respecto a la potencia de la t, nunca es menor al 83% y en el caso de que los supuestos de la t se cumplen la potencia de la U alcanza el 95%.

Se considera que los datos provienen de dos muestras aleatorias X_1, X_2, \dots, X_{n_1} , de tamaño n_1 , y Y_1, Y_2, \dots, Y_{n_2} , de tamaño n_2 .

Los supuestos básicos de la prueba son:

- a) Las muestras se han tomado aleatoriamente y en forma independiente, no solamente entre los dos grupos considerados, sino además dentro de cada grupo.
- b) Ambas muestras provienen de variables continuas (sin embargo, se puede tolerar un moderado número de “ligas” u observaciones con el mismo valor)
- c) La escala de medida es al menos ordinal.

II. La hipótesis.

Sean $F_X(x)$ y $F_Y(y)$ las funciones de distribución correspondiente a las dos variables X, Y. La variable aleatoria X, se afirma que es estocásticamente más pequeña que Y si $F_X(a) < F_Y(a)$ para toda a.

La hipótesis que se plantea es:

$H_0: F_X(a) = F_Y(a)$ para toda a

contra cualquiera de las alternativas

$$H_1: \begin{cases} F_X(a) > F_Y(a) \\ F_X(a) < F_Y(a) \\ F_X(a) \neq F_Y(a) \end{cases} \text{ Para alguna } a$$

Lo anterior se puede considerar equivalente a

$$H_0: P(X > Y) = P(X < Y) = \frac{1}{2}$$

contra cualquiera de las alternativas

$$H_1: \begin{cases} P(X > Y) > P(X < Y) \\ P(X > Y) < P(X < Y) \\ P(X > Y) \neq P(X < Y) \end{cases}$$

III. La Estadística de Prueba

Considérese la sucesión de observaciones de ambas poblaciones mezcladas, pero identificables, como aparecerían si se ordenasen de menor a mayor. Para cada par de observaciones X_i, Y_j , con $i=1,2,\dots,n_1; j=1,2,\dots,n_2$ se define

$$W_{ij} = \begin{cases} 1 & \text{si } X_i < Y_j \\ 0 & \text{si } X_i > Y_j \end{cases}$$

La suma de las w 's es el estadístico U_x

$$U_x = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} W_{ij}$$

Un resultado semejante se define para y , lo que da U_y . Ambas U_x y U_y cumplen con la relación $U_x + U_y = n_1 n_2$ de donde

$$U_x = n_1 n_2 - U_y$$

a) Media y Varianza de la U_x

Apoyándose en la hipótesis nula, la media y la varianza de U_x se pueden calcular a partir de las siguientes expresiones

$$E(U_x) = \frac{n_1 n_2}{2}$$

$$V(U_x) = \frac{1}{12} n_1 n_2 (n_1 + n_2 + 1)$$

Estos resultados son de particular utilidad en el caso de muestras grandes, pues por el teorema del límite central se tiene que la variable Z expresada por

$$Z = \frac{U_x - n_1 n_2 / 2}{\sqrt{\frac{1}{12} n_1 n_2 (n_1 + n_2 + 1)}}$$

se distribuye aproximadamente como una normal estándar o normal $(0,1)$

IV. Ejemplo

En un estudio sobre el problema de la generalización de la imitación que un grupo de ratas había aprendido, al ser colocadas en una situación nueva. Cinco ratas fueron adiestradas para imitar a ratas "líderes" en un laberinto en T. Se les enseñó a seguir a sus líderes cuando tenían hambre para lograr un incentivo de comida. Después, las cinco ratas fueron transferidas a una situación consistente en

evitar descargas eléctricas, donde la imitación de las ratas líderes les permitieran liberarse de dichas descargas. Su conducta al evitar descargas fue comparada a la de cuatro controles sin adiestramiento previo para seguir a sus líderes.

La hipótesis fue que las cinco ratas adiestradas en la imitación transferían este adiestramiento a la nueva situación y alcanzarían el criterio de aprendizaje correspondiente a la nueva situación antes que las cuatro ratas del grupo control. La comparación se hace en términos de los ensayos de cada rata previos a un criterio de 10 respuestas en 10 ensayos.

Sean denotadas por T las ratas sometidas al entrenamiento previo y sean denotadas por C las ratas del grupo control.

La hipótesis nula H_0 : ambos grupos de ratas responden de forma semejante, se plantea por

$$H_0: P(T > C) = P(T < C) = \frac{1}{2}$$

La hipótesis alternativa H_1 : Las ratas sometidas al entrenamiento T aprenden en menos sesiones que las ratas control C

$$H_1: P(T < C) > \frac{1}{2}$$

se selecciona un nivel de significancia para la prueba de 5%

Rata	1	2	3	4	5
Tratamiento (ensayos)	78	64	45	75	82
Control (ensayos)	110	51	70	53	-

se cumple la siguiente relaciones

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

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

donde

R_1 es la suma de los rangos asignados al grupo cuyo tamaño muestral es n_1

R_2 es la suma de los rangos asignados al grupo cuyo tamaño muestral es n_2

Para aplicarse el ejemplo mencionado, se requiere la sucesión mezclada y etiquetada, a la que además, se le asignan rangos

Rango	1	2	3	4	5	6	7	8	9
	45	51	53	64	70	75	78	82	110
	T	C	C	T	C	T	T	T	C

IV. Procedimiento de Cálculo

Ordenar la sucesión mezclada e identificada

45	51	53	64	70	75	78	82	110
T	C	C	T	C	T	T	T	C

Calcular el número de puntajes T que preceden a cada puntaje C

T	C	C	T	C	T	T	T	C
	1	1		2				5

La suma $1+1+2+5=9$ es el valor de U. En las tablas para $n=4$ y $n=5$ se puede observar que

$$P(U \leq 9) = 0.452$$

Como el valor $0.452 > 0.05$ correspondiente al nivel de significancia seleccionada se puede concluir que H_0 no se rechaza.

b) Procedimiento de Cálculo

$$R_T = 1+4+6+7+8=26$$

$$U_T = 4(5) + \frac{5(6)}{2} - 26 = 9$$

$$U_T = 9$$

valor calculado que coincide con el del procedimiento 1

c) Procedimiento Propuesto

Las dos muestras se colocan, no importa que no estén ordenadas, al margen de una matriz en la siguiente forma

C\T	82	78	64	75	45
51					
110					
53					
70					

En cada celda de la matriz, se asigna el valor 1 si la observación del renglón supera a la observación de la columna y un cero si sucede lo contrario. En caso de empate se asigna $1/2$

C\T	82	78	64	75	45
51	0	0	0	0	1
110	1	1	1	1	1
53	0	0	0	0	1
70	0	0	1	0	1

La estadística U_T está dada por la suma de las celdas en la matriz, en este caso $U_T = 9$.

d) El Caso de Ligas o Empates

Teóricamente este caso no debe darse, pero debido a la limitación de los dispositivos de medición de los fenómenos en estudio, es frecuente tener observaciones con empates o ligas. En general la perturbación causada por rangos ligados no es importante.

Cuando son rangos los resultados de las mediciones se acostumbra asignar el promedio de los rangos correspondientes a las observaciones ligadas y hacer algunos ajustes al cálculo manual.

En el caso particular del procedimiento de cálculo de la estadística U mediante la matriz, se ha mencionado que a la celda correspondiente a observaciones ligadas se le asigne 1/2.

e) Tamaños de Muestra Grandes

Si los tamaños de muestra son superiores a 20 observaciones es posible utilizar la variable Z definida en 3; la cual se distribuye aproximadamente como una normal (0,1).

Suponga que se tienen dos muestras de tamaño $n_1=16$ y $n_2=23$ correspondientes al grupo A y al grupo B. Se plantea la hipótesis

$$H_0: P(B > A) = P(B < A) = \frac{1}{2}$$

$$H_1: P(B > A) = \frac{1}{2}$$

Nivel de significancia $\alpha=0.05$.

Grupo A 13, 12, 12, 10, 10, 10, 10, 7, 7, 7, 7, 7, 6, 9, 8, 8

Grupo B 16, 17, 15, 15, 15, 14, 14, 14, 13, 13, 13, 12, 12,
12, 12, 11, 11, 10, 10, 10, 6, 8, 8

El cálculo de U mediante el procedimiento de la matriz da el valor siguiente

$$U=64$$

de donde

$$W = \frac{U - n_1 n_2 / 2}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}$$

$$W = \frac{64 - (16)(23)/2}{\sqrt{\frac{16(23)(16 + 23 + 1)}{12}}}$$

$$W=-3.43$$

Como se cumple que $P(Z \leq -3.43) < 0.05$ se rechaza H_0 . Por lo tanto se puede concluir que el grupo B es superior al grupo A.

Análisis de Varianza por Rangos de Kruskal-Wallis

I. Introducción.

En el año de 1952 Kruskal y Wallis propusieron un equivalente no paramétrico a la prueba F empleada en el análisis de un criterio de clasificación. La prueba que desarrollaron se presentó como una extensión a k muestras independientes de la prueba U de Mann y Whitney para dos muestras independientes y fue identificada a partir de entonces con los nombres de sus autores.

La prueba de Kruskal-Wallis es recomendable para aquellos problemas en los que los supuestos de normalidad u homogeneidad de varianzas no se cumplen adecuadamente y por lo tanto, emergen como una fuente limitante en la aplicación de la prueba F.

Desde el punto de vista de la potencia se le asigna 95.5% relativa a la F, por lo que se le puede considerar bastante confiable.

Otra prueba no paramétrica con la cual se puede resolver el mismo problema es la extensión de la prueba de la mediana para k grupos independientes, sin embargo, la prueba de Kruskal-Wallis incorpora más información ya que toma en consideración los rangos de las observaciones y no solamente la posibilidad de que éstos sean inferiores o superiores a la mediana, como es el caso de la prueba antes mencionada.

Los datos consisten de k muestras independientes de tamaños (n_i , $i=1,2,\dots,k$) no necesariamente iguales, los cuales se pueden esquematizar en la siguiente forma:

Muestra 1	Muestra 2	...	Muestra k
X11	X21	...	Xk1
X12	X22	...	Xk2
.
.
.
X1n1	X2n2	X2n2	Xknk

Los supuestos básicos que sustentan la prueba son:

- a) Las k muestras son aleatorias
- b) Las k muestras son independientes
- c) Las variables observadas son continuas
- d) La escala de medición es al menos ordinal

II. La Hipótesis

Las hipótesis nula y alternativa se plantean en dos formas opcionales, primero con referencia a la distribución.

Ho: Las distribuciones de las k poblaciones son idénticas

H1: Al menos una de las k distribuciones es diferente

o con referencia a las medias ya que la prueba es sensible a sus diferencias

Ho: Las medianas de las k muestras son iguales

H1: De las k poblaciones al menos una tiene la mediana diferente

III. La Estadística de Prueba

Designada por H, la estadística de prueba sigue una distribución aproximada a la Ji-cuadrada con k-1 grados de libertad. La aproximación es mejor a medida que se dispone de tamaños de muestra más grandes.

En el procedimiento de cálculo de la estadística de prueba se sustituyen los valores X_{ij} por sus rangos correspondientes (R_{ij}) al tomar en consideración todas las observaciones en forma simultánea.

$$H = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1)$$

donde

$$\begin{aligned} k &= \text{Número de muestras} \\ n_i &= \text{Tamaño de la } i\text{-ésima muestra} \\ N &= \sum_{i=1}^k n_i \text{ Total de observaciones} \\ R_i &= \text{Suma de rangos de la } i\text{-ésima muestra} \end{aligned}$$

La significancia de un valor calculado de H se verifica mediante tablas de Ji-cuadrada, pero para los casos $k \leq 3$ y $n_i \leq 5$ se aconseja el empleo de valores tabulados.

Cuando ocurre que dos o más observaciones tienen el mismo valor se asigna a cada una de ellas el promedio de los rangos que les corresponderían. Estos empates o ligas pueden dar lugar a ciertas perturbaciones, razón por la cual se recomienda introducir un factor de corrección, el cual consiste en dividir H entre la siguiente expresión.

$$1 - \frac{\sum T}{N^3 - N}$$

donde $T = t^3 - t$ (t es el número de observaciones con los mismos puntajes)

IV. Ejemplo

Un especialista en limnología pretende establecer si hay diferencia en el PH del agua de 4 lagunas, para ello tomó series de 8 muestras en cada lago, excepto en la laguna número 3, en la cual tomó 7 muestras. Se utilizará la prueba de Kruskal-Wallis para probar la hipótesis.

H_0 : El PH es el mismo en los 4 lagos

H_1 : El PH es diferente en los 4 lagos

El nivel de significancia de la prueba es de $\alpha=0.05$

PH LAGO 1	PH LAGO 2	PH LAGO 3	PH LAGO 4
7.68	7.69	7.74	7.71
7.70	7.70	7.75	7.71
7.72	7.71	7.77	7.74
7.73	7.73	7.78	7.79
7.73	7.74	7.80	7.84
7.76	7.74	7.81	7.85
7.78	7.78	7.81	7.87
7.80	7.81		7.91
n1=8	n2=8	n3=7	n4=8

Una vez que se mezclan y ordenan las observaciones de los 4 grupos, los rangos que se les asigna quedan como sigue

1	2	13.5 **	6 **
3.5 **	3.5 **	16	6 **
8	6 **	18	13.5 **
10 **	10 **	20 **	22
10 **	13.5 **	23.5 **	28
17	13.5 **	26 **	29
20 **	20 **	26 **	30
23.5 **	26 **		31
93	94.5	143	165.5

(*) Rangos ligados

$$N=8+8+7+8=31$$

$$\begin{aligned}
 H &= \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1) \\
 &= \frac{12}{31(31+1)} \left[\frac{93^2}{8} + \frac{94.5^2}{8} + \frac{143^2}{7} + \frac{165.5^2}{8} \right] - 3(31+1) \\
 &= 7.7336
 \end{aligned}$$

Número de grupos de rangos ligados m=7

$$\begin{aligned}
 \sum T &= \sum (t_i^3 - t_i) \\
 &= (2^3 - 2) + (3^3 - 3) + (3^3 - 3) + (4^3 - 4) + (3^3 - 3) + (2^3 - 2) + (3^3 - 3) \\
 &= 168
 \end{aligned}$$

La H corregida tiene el valor

$$H_c = \frac{H}{1 - \frac{\sum T}{N^3 - N}} = \frac{7.336}{1 - \frac{168}{31^3 - 31}}$$

$$H_c = 7.378$$

El valor de tablas de Ji-cuadrada con 3 grados de libertad y 0.95 es de 7.815, por lo tanto Ho no se puede rechazar.

Prueba de Friedman para Análisis de Varianza con Dos Criterios de Clasificación

I. Introducción

Con frecuencia la distribución de la variable en estudio o limitaciones en los dispositivos de medición dan lugar a severos desvíos en los datos con relación a las hipótesis de trabajo que sustentan el análisis clásico de la varianza, como es el caso de la normalidad y homocedasticidad. Estos desvíos pueden inducir al investigador a llegar a conclusiones equivocadas en las pruebas de hipótesis mediante la F de Snedecor, a pesar de la conocida robustez de esta prueba.

Para el caso del análisis de la varianza con dos criterios de clasificación sin interacciones, Friedman propuso una prueba no paramétrica que ha sido probada con excelentes resultados.

Los datos se organizan en una tabla de doble entrada con n renglones y k columnas. Los renglones representan bloques o sujetos y las columnas tratamientos. La X_{ij} observación de la tabla (i -ésimo bloque, j -ésimo tratamiento) es sustituida por el rango que corresponde dentro del renglón. Así R_{ij} es el rango de la observación correspondiente al i -ésimo sujetos o bloque ($i=1,2,\dots,n$) del j -ésimo tratamiento ($j=1,2,\dots,k$). Se denotará por R_j la suma de los rangos de la j -ésima columna. La suma de los rangos de cualquier renglón será $\frac{k(k+1)}{2}$ por ser la suma de los primeros k naturales.

Los datos pueden esquematizarse en la siguiente forma tabular

		Tratamientos				
		1	2	...	k	
Bloques	1	R_{11}	R_{12}	...	R_{1k}	$\frac{k(k+1)}{2}$
	2	R_{21}	R_{22}	...	R_{2k}	$\frac{k(k+1)}{2}$

	n	R_{n1}	R_{n2}	...	R_{nk}	$\frac{k(k+1)}{2}$
		R_1	R_2	...	R	$\frac{nk(k+1)}{2}$

Los supuestos básicos que permiten el empleo de la estadística de Friedman son los siguientes:

- a) Las observaciones de los diferentes bloques (renglones) son independientes.
- b) Las observaciones de diferentes columnas están relacionadas por algún criterio de asociación o apareamiento.

II. La Hipótesis

Ho: Los rangos asignables a las observaciones de un bloque cualquiera son igualmente probables para todos los tratamientos.

$$H_0: R_1 = R_2 = \dots = R_k$$

H1: Los rangos asignables a las observaciones de un bloque cualquiera tienen mayor probabilidad de ser distintos para cierto tratamiento.

$$H_1: R_j \neq R_{j'} \quad \forall j \neq j'$$

III. La Estadística de Prueba

La estadística T propuesta por Friedman, permite determinar si existen diferencia significativa entre las sumas de los rangos de los k tratamientos, se distribuye aproximadamente como una Ji-Cuadrada con k-1 grados de libertad y tiene la siguiente expresión:

$$T = \frac{12}{nk(k+1)} \sum_{j=1}^K R_j^2 - 3n(k+1)$$

IV. Relación de la Estadística T de Friedman con el Coeficiente de concordancia de Kendall

Kendall y Babington introdujeron en 1939 en forma independiente una estadística llamada coeficiente de concordancia, la cual se denotará por W. La estadística W mide el grado de concordancia en términos de rangos entre los n bloques.

La expresión de la W de Kendall en términos de la T de Friedman es como sigue

$$W = \frac{T}{n(k-1)}$$

Cualquier hipótesis a probar mediante la W puede ser probada mediante la R, ya que si la R excede en $1-\alpha$ un cuantil, entonces W excede también en $1-\alpha$ su propio cuantil.

V. Ejemplo

Se realizó un experimento con cerdos de guinea para probar 4 tipos de dieta. Se establecieron 5 bloques de 4 animales cada uno, los criterios de bloqueo fueron dados por las condiciones de laboratorio. A cada uno de los cuatro cerdos le fue asignado una dieta. La información registrada fue la ganancia de peso medida en Kg. en un periodo de una semana.

Ho: La ganancia de peso es la misma para las cuatro dietas.

H1: La ganancia de peso no es la misma para las cuatro dietas.

$$\alpha=0.05$$

Bloque	1	2	3	4
1	1.5	2.7	2.1	1.3
2	1.4	2.9	2.2	1.0
3	1.4	2.1	2.4	1.1
4	1.2	3.0	2.0	1.3
5	1.4	3.3	2.5	1.5

La estadística T calculada tiene un valor de 7.815 lo que permite rechazar la hipótesis H_0 .

El valor del coeficiente de concordancia de Kendall es de $W=0.521$

Prueba de Bondad de Ajuste de Kolmogorov

I Introducción

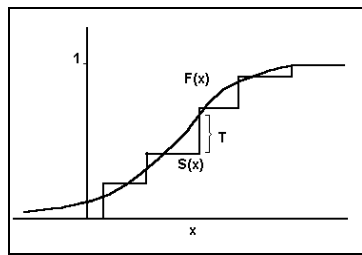
Esta prueba fue introducida por Kolmogorov en 1933 como una alternativa a la prueba basada en la Ji-cuadrada.

Una prueba de bondad de ajuste usualmente envuelve el análisis de una muestra aleatoria de alguna distribución desconocida con objeto de probar la hipótesis nula de que la distribución desconocida $F(x)$ es de hecho conocida y corresponde a una distribución específica. Esto es la hipótesis nula define una distribución $F^*(x)$.

La forma lógica de comparación de la muestra aleatoria con la distribución propuesta $F(x)$ es su distribución empírica $S(x)$.

El problema reside entonces en determinar el tipo de prueba que mida la discrepancia entre $S(x)$ y $F^*(x)$.

Kolmogorov sugirió tomar como base la mayor distancia entre ambas funciones como se ilustra en la gráfica siguiente:



La prueba de Kolmogorov puede ser preferida sobre la Ji-cuadrada para bondad de ajuste si el tamaño de muestra es pequeño, porque la prueba de Kolmogorov es exacta, aun para muestras pequeñas, mientras que la Ji-cuadrada asume que el número de observaciones es suficientemente grande.

Datos:

Una muestra aleatoria X_1, X_2, \dots, X_n que se supone asociada con una distribución $F^*(x)$.

II. Hipótesis:

a) Dos colas

$H_0: F(x) = F^*(x)$ para toda $x \in (-\infty, \infty)$

$H_a: F(x) \neq F^*(x)$ para al menos un valor de x

b) Una cola

$H_0: F(x) = F^*(x)$ para toda $x \in (-\infty, \infty)$

$H_a: F(x) < F^*(x)$ para al menos una X

c) Una cola

Ho: $F(x) = F^*(x)$ para toda $x \in (-\infty, \infty)$

Ha: $F(x) > F^*(x)$ para al menos un valor de x

III. Estadística de Prueba

Sea $S(x)$ la distribución empírica basada en la muestra X_1, X_2, \dots, X_n . La estadística de prueba para cada tipo de hipótesis es T_1 , la cual se define:

a) Sea T_1 la mayor distancia vertical entre $S(x)$ y $F^*(x)$

$$T_1 = \sup_x |F^*(x) - S(x)|$$

b) Sea $T_1(+)$ la mayor distancia vertical positiva con $F^*(x)$ por arriba de $S(x)$.

$$T_1(+) = \sup_x (F^*(x) - S(x))$$

c) Sea $T_1(-)$ la mayor distancia vertical negativa con $F^*(x)$ por abajo del $S(x)$.

$$T_1(-) = \sup_x (F^*(x) - S(x))$$

IV. Regla de decisión.

Se rechaza Ho al nivel de significancia $\alpha\%$ si la estadística T_1 , $T_1(+)$ ó $T_1(-)$ excede el valor del cuartil $(1-\alpha)\%$.

Ejemplo

(Sidney Siegel). Supongamos que un investigador está interesado en confirmar, por medios experimentales, la observación sociológica de que los negros americanos parecen tener una jerarquía de preferencias entre distintos matices de piel. Para probar cuán sistemáticas son las preferencias de los negros por el tono de piel, nuestro ficticio investigador decide tomar fotografías de cada uno de diez sujetos negros. El fotógrafo obtiene cinco copias de cada fotografía, cada copia ligeramente diferente de las otras en cuanto a la oscuridad, así, las cinco copias pueden obtenerse según el color de la piel, de la más oscura a la más clara. la fotografía que presenta el color de la piel más oscuro recibe el rango 1, la siguiente oscuridad el 2, y así sucesivamente hasta la más clara, que recibe el rango 5. Después se le pide a cada sujeto que escoja, de entre las cinco impresiones, su propia fotografía. Si el matiz de la piel no es importante para los sujetos, las fotografías de cada rango a menudo serán escogidas igualmente, exceptuando las diferencias aleatorias. Si el matiz de la piel es importante, como se ha supuesto, entonces los sujetos consistentemente mostrarán preferencias a uno de los rangos extremos.

I Hipótesis nula: Ho: no hay diferencias entre el número esperado de elecciones para cada uno de los cinco negros y las diferencias observadas son meramente variaciones casuales esperadas en una muestra aleatoria proveniente de una población rectangular donde $f_1=f_2=\dots=f_5$. H1: las frecuencias f_1, f_2, \dots, f_5 , no son todas iguales.

- II. Prueba estadística. la prueba de una muestra de Kolmogorov-Smirnov se escoge porque el investigador dese comparar una distribución de puntajes observados en una escala ordinal con distribución teórica.
- III. Nivel de significancia. Sean $\alpha=0.01$ y N (el número de sujetos negros que sirvieron como tales en el estudio)=10.
- IV. Distribución muestral. diferentes valores críticos de D de la distribución muestral se presenta en la tabla de valores críticos de D para la prueba de una muestra de Kolmogorov-Smirnov, junto con sus probabilidades asociadas de ocurrencia conforme a H_0 .
- V. Región de rechazo. Consiste en todos los valores de D son tan grandes que la probabilidad asociada con su ocurrencia conforme a H_0 es igual o menor que $\alpha=0.01$.
- VI. Decisión. En este estudio hipotético, cada sujeto escoge una de las cinco impresiones de la misma fotografía. Supongamos que un sujeto escoge la impresión 2 (la siguiente a la más oscura), cinco sujetos escogen la impresión 4 (la siguiente a la más clara) y cuatro sujetos escogen la impresión 5 (la impresión más clara). La tabla siguiente muestra los datos y los muestra en la forma apropiada para aplicarlos a la prueba de una muestra de Kolmogorov-Smirnov.

Rango de la foto escogida (1 es el color de piel más oscuro)					
	1	2	3	4	5
f=número de sujetos que escogen a cada rango	0	1	0	5	4
Fo(x)=distribución acumulativa de elecciones conforme a Ho.	1/5	2/5	3/5	4/5	5/5
S₁₀(x)=distribución acumulativa de elecciones observadas	0/10	1/10	1/10	6/10	10/10
$F_o(x) - S_{10}(x)$	2/10	3/10	5/10	2/10	0

Nótese que $F_o(x)$ es la distribución acumulativa teórica conforme a H_0 , donde H_0 supone que cada una de las 5 impresiones recibiría una quinta parte de las elecciones. $S_{10}(x)$ es la distribución acumulativa de las elecciones observadas de los 10 sujetos. En el último renglón de la tabla encontramos la desviación absoluta del valor real de la muestra con respecto a su valor esperado tomado comparativamente. Así la primera desviación absoluta del valor real de la muestra con respecto a su valor esperado tomado comparativamente. Así la primera desviación absoluta es de 2/10 obtenida sustituyendo 0 de 1/5.

La inspección del último renglón de la tabla señala que la D para todos los datos es de 5/10, o sea 0.500. Se muestra que para $N=10$, $D \geq 0.500$ tiene una probabilidad asociada conforme a H_0 de $p < 0.01$. En vista de que p asociada con el valor observado de D es menor que $\alpha=0.01$, nuestra decisión en este estudio es rechazar H_0 a favor de H_1 . Concluimos que nuestros sujetos muestran preferencias significativas entre los colores de piel.

Resumen del procedimiento. En el cálculo de la prueba de Kolmogorov-Smirnov se siguen estos pasos:

1. Se especifica la función acumulativa teórica, es decir, la distribución acumulativa esperada conforme a H_0 .
2. Se disponen los puntajes observados en una distribución acumulativa, comparando cada intervalo de $S_N(x)$ con el intervalo correspondiente de $F_0(x)$.
3. En cada peldaño de las distribuciones acumulativas, se sustrae $S_N(x)$ de $F_0(x)$
4. En la tabla de Kolmogorov se encuentra la probabilidad (de dos colas) asociada con la ocurrencia conforme a H_0 de valores tan grandes como el valor observado de D . Si esa p tiene un valor igual o menor que α , se rechaza H_0 .

V. Potencia

La prueba de una muestra de Kolmogorov-Smirnov trata las observaciones individuales separadamente y, así, a diferencia de la prueba Ji-cuadrada para una muestra, no pierde necesariamente información al combinar categorías. Cuando las muestras son pequeñas y, por consiguiente, las categorías adyacentes deben combinarse antes de que Ji-cuadrada pueda calcularse apropiadamente, la prueba X^2 es definitivamente menos poderosa que la de Kolmogorov-Smirnov. Además, para muestras muy pequeñas, la prueba X^2 no es aplicable en modo alguno; sin embargo, la de Kolmogorov-Smirnov sí lo es. Estos hechos sugieren que la prueba de Kolmogorov-Smirnov puede en todos los casos ser más poderosa que su alternativa, la prueba X^2 .

Un nuevo análisis mediante la prueba X^2 de los datos del ejemplo anterior destaca la potencia superior de la prueba de Kolmogorov-Smirnov. En la norma en que los datos están presentados en la tabla de preferencias hipotéticas, X^2 no puede calcularse, porque las frecuencias esperadas son solamente 2 cuando $N=10$ y $k=5$. Tenemos que combinar las categorías adyacentes para incrementar la frecuencia esperada por celdilla, lo que nos lleva al análisis de dos categorías que aparecen en la tabla siguiente. Cualquier elección de un sujeto es simplemente clasificada como inclinación al tono de piel claro u oscuro; las gradaciones más finas serán ignoradas.

	Oscuro (rangos 1,2)	Claro (rangos 3,4)	Total
Frecuencia de elección	1	9	10

Para estos datos, X^2 (no corregida por continuidad)=3.75. Se muestra que la probabilidad asociada con la ocurrencia conforme a H_0 de ese valor, cuando $gl=k-1=1$, está entre 0.10 y 0.05. Esto es, $0.10 > p > 0.05$. Este valor de p no nos permite rechazar H_0 al nivel 0.01 de significación.

Nótese que la p encontrada con la prueba Kolmogorov-Smirnov es menor que 0.01, mientras la de la prueba X^2 es mayor que 0.05. Esta diferencia nos da alguna indicación de la potencia superior de la prueba Kolmogorov-Smirnov.

La Prueba de Ji Cuadrada para las Tablas de Contingencia

I. Introducción

Son frecuentes los estudios en los cuales, los elementos de una muestra deben ser categorizados de acuerdo a dos o más criterios de clasificación y es de interés para el investigador, determinar si los criterios de clasificación son independientes. En este tipo de problemas el análisis de los resultados mediante el uso de la distribución Ji-cuadrada aplicada a tablas de contingencia es sin duda, uno de los procedimientos estadísticos más populares.

En un sentido más estricto, una tabla de contingencia es un arreglo matricial de c renglones y k columnas en cuyas $C \times K$ celdas se registran conteos o frecuencias referentes a la muestra y cuyos elementos se han medido en un nivel nominal.

En este punto resulta conveniente aclarar que las hipótesis a probar mediante tablas de contingencia, aún cuando los procedimientos de cálculo son los mismos, tienen básicamente dos sentidos diferentes:

- a) Como hipótesis de igualdad de proporciones en los diferentes niveles de cierta clasificación, cuando las observaciones provienen de dos o más poblaciones.
- b) Como hipótesis de independencia entre dos criterios de clasificación aplicables a los elementos de una misma población, cuando el objetivo es valorar la relación que existe entre los dos criterios o variables.

Como se mencionó, ambos casos son tratados idénticamente desde el punto de vista de los cálculos estadísticos, pero las diferencias básicas entre las dos aplicaciones justifican discusiones separadas.

II. La Hipótesis

1) Hipótesis de Igualdad de Proporciones

Suponga que se tiene k poblaciones y que se extrae una muestra aleatoria de tamaño n_j de cada población ($j=1,2,\dots,k$). Cada observación de las k muestras puede ser clasificada en una de c diferente categorías. Se denotará por O_{ij} el número de observaciones de la i -ésima categoría perteneciente a la j -ésima muestra. Denótese además por m_i el total de observaciones pertenecientes a todas las muestras que quedan contenidas en la i -ésima categoría.

La información se dispone en forma tabular de la manera siguiente:

Muestras (Tratamientos)					
Clases	1	2	...	k	Total
1	O11	O12	...	O1k	m1
2	O21	O22	...	O2k	m2
.
.
.
c	Oc1	Oc2	...	Ock	mc
Total	n1	n2	...	nk	N

En la tabla se debe verificar lo siguiente:

$$\sum_{i=1}^c m_i = \sum_{j=1}^k n_j = N$$

$$\sum_{i=1}^c O_{ij} = n_j \quad \sum_{j=1}^k O_{ij} = m_i$$

Se consideran los siguientes supuestos básicos en el planteamiento de la hipótesis

- a) Las k muestras son aleatorias
- b) Los resultados de las diferentes muestras son mutuamente independientes
- c) Cada observación puede ser categorizada en una y sólo una de las c diferentes clases

Sea P_{ij} la probabilidad de que un elemento de la j-ésima población seleccionado al azar, quede clasificada en la i-ésima categoría.

La hipótesis nula y alternativa se plantea como sigue:

$$H_0: P_{i1} = P_{i2} = \dots = P_{ic}$$

La probabilidad (proporción) de pertenecer a cualquiera de las c clases es la misma para cualquier elemento de la i-ésima muestra.

$$H_1: P_{ij} \neq P_{ij'} \text{ para al menos una pareja } j \neq j'$$

La probabilidad de pertenecer a cualquier de las c clases es diferente para al menos una clase.

2) Hipótesis de Independencia

Suponga que se dispone de una muestra aleatoria de tamaño N y que las observaciones de la muestra pueden clasificarse de acuerdo a dos criterios. Al usar el primer criterio cada observación puede asociarse con uno de los c renglones y al usar el segundo criterio se asocia con una de las k columnas.

La disposición de las observaciones es igual que en 1) con la excepción de que en este caso, las n no se establecen previamente , sino que son aleatorias.

Los supuestos para este caso son los siguientes:

- a) Cada observación tiene la misma probabilidad de ser clasificada en el i-ésimo renglón y en la j-ésima columna, independientemente de cualquier otra observación.
- b) Las observaciones pueden ser clasificadas en una de c diferente categorías de acuerdo al segundo criterio.

La hipótesis nula correspondiente se puede plantear en los siguientes términos:

Ho: El evento “la observación pertenece al i-ésimo renglón” es independiente del evento “la misma observación pertenece a la j-ésima columna” para toda i,j.

La proposición anterior puede traducirse en términos probabilísticos de la siguiente forma

Sea P_i la probabilidad de pertenecer al i-ésimo renglón

P_j la probabilidad de pertenecer a la j-ésima columna

Ho: $P_{ij} = P_i P_j$ para toda i,j

H1: $P_{ij} \neq P_i P_j$ para alguna i,j

III. La Estadística de Prueba

1) La Estadística de Prueba con Distribución Ji-Cuadrada Aproximada

Para ambos casos 1 y 2 de la sección anterior se define la frecuencia esperada E_{ij} correspondiente a la celda con coordenadas (i,j) como

$$E_{ij} = \frac{min_j}{N}$$

La estadística de prueba se denota por T^2 y su fórmula de cálculo es:

$$T^2 = \sum_{i=1}^c \sum_{j=1}^K \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

La distribución de T se aproxima a una Ji-cuadrada con (c-1)(k-1) grados de libertad.

Atención especial en las tablas de contingencia merecen las tablas 2x2. Para este caso la fórmula para el cálculo de T se puede simplificar con la siguiente expresión.

$$T^2 = \frac{N(O_{11}O_{22} - O_{12}O_{21})^2}{m_1m_2n_1n_2}$$

Sin embargo, es posible mejorar el cálculo de T con la introducción de Yates para la continuidad, ya que se aproxima la distribución de una continua, como lo es la Ji-cuadrada.

La introducción de la corrección para la continuidad cambia la expresión de T en la forma siguiente:

$$T^2 = \frac{N\left(\left|O_{11}O_{22} - O_{12}O_{21}\right| - \frac{N}{2}\right)^2}{m_1m_2n_1n_2}$$

2) El Coeficiente de Contingencia

Para medir el grado de relación o dependencia de las clasificaciones de una tabla de contingencia se utiliza el coeficiente de contingencia CC.

$$CC = \sqrt{\frac{T^2}{T^2 + N}}$$

Un valor alto de CC significa un alto grado de asociación, cuyo valor máximo está determinado por el número de filas y columnas de la tabla, sin embargo en ningún caso es superior a 1.

3) La V de Cramér.

Cramér introdujo una variable del coeficiente de contingencia conocida como la V de Cramér. Tiene la ventaja de alcanzar un valor máximo de 1 sin importar la estructura de la tabla.

$$V = \sqrt{\frac{T^2}{N(m-1)}}$$

Donde m es el mínimo entre el número de columnas y renglones.

4) El Coeficiente Phi de Pearson

Se calcula mediante el cociente de la estadística T (Ji-Cuadrada) dividida entre el tamaño de muestra y después se extrae raíz cuadrada.

$$\phi = \sqrt{\frac{T^2}{N}}$$

5) Ji Cuadrada de Mantel-Haenszel

Para medir la asociación lineal entre variables medidas en nivel ordinal se frecuenta el uso del coeficiente de correlación de Spearman.

Este coeficiente se calcula a partir de los rangos de dos grupos de observaciones que corresponden a las variables, cuya asociación se pretende medir. Una alternativa de la Ji Cuadrada de Mantel-Haenszel, medida de asociación lineal, que se calcula multiplicando el cuadrado del coeficiente de correlación de Pearson por el número de casos menos 1.

$$MH = (N - 1)R^2$$

La estadística MH tiene un grado de libertad y no es aplicable a mediciones efectuadas en nivel nominal.

6) La Prueba exacta de Fisher

Una popular prueba no paramétrica aplicable a tablas de contingencia 2x2 es la prueba exacta de Fisher. la prueba se apoya en el cálculo de la probabilidad exacta de observar un conjunto particular de frecuencias en una tabla 2x2, cuando los totales marginales se consideran fijos.

Si se considera la tabla:

O_{11}	O_{12}	$O_{1\cdot}$
O_{21}	O_{22}	$O_{2\cdot}$
$O_{\cdot 1}$	$O_{\cdot 2}$	N

La probabilidad exacta se obtiene mediante una distribución hipergeométrica

$$P = \frac{\binom{O_{\cdot 1}}{O_{11}} \binom{O_{\cdot 2}}{O_{12}}}{\binom{N}{O_{1\cdot}}}$$