

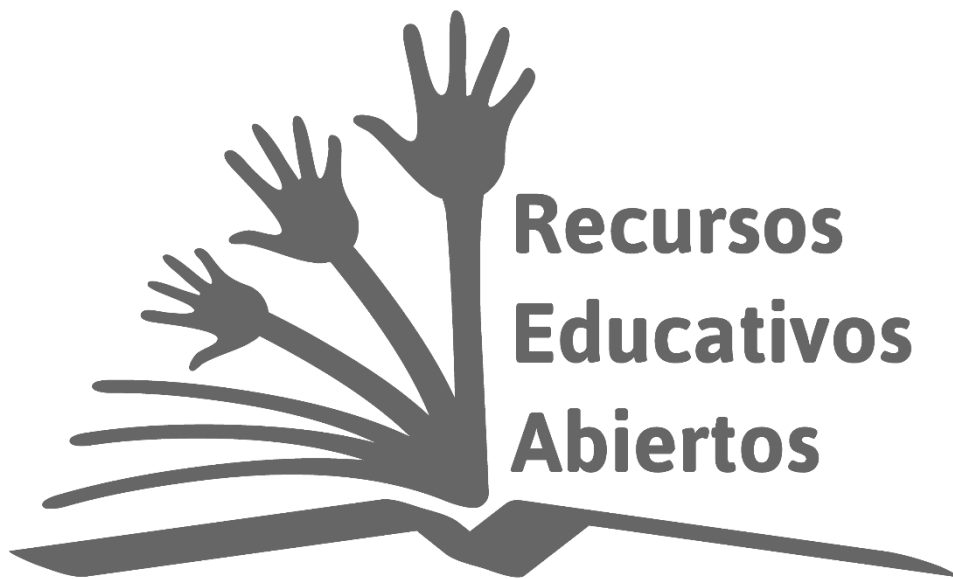
- 2

**Introducción a la Estadística No
Paramétrica para estudiantes de
educación**

Introducción a la Estadística No Paramétrica para estudiantes de educación
Autor: Lic. Jorge Lorenzo
Cátedra de Estadística y Sistemas de Información Educativa
Facultad de Filosofía y Humanidades
Universidad Nacional de Córdoba
Año: 2020
Ilustración de tapa: Gage Walker - Unplash
Photographer, Tree Hugger |
Instagram: [_gagewalker_](#)



**Attribution-NonCommercial-ShareAlike
4.0 International (CC BY-NC-SA 4.0)**



Introducción

El trabajo con datos y los análisis que estos suponen es un tema en que los educadores se involucran cada vez más. Casi todas las carreras de ciencias de la educación incluyen en su oferta académica áreas curriculares dedicadas a la investigación de tipo cuantitativa y, por ende, la labor con datos. Por otra parte, la construcción de indicadores educativos, su lectura y análisis es una faena en la que los profesionales de las ciencias de la educación se implican para el diseño de programas y políticas educativas. Sin embargo, el trabajo cotidiano de las escuelas está escasamente atravesado por la lógica del análisis estadístico. En lo personal, considero que esto es producto de urgencias y problemas que enfrentan las gestiones y los docentes, más preocupados por la formación de alumnas y alumnos, lo cual conlleva un tiempo considerable. A pesar de ello, la escuela es un reservorio de registros de escolares, docentes, familias, etc. Todo ello representa un cúmulo de datos que las gestiones utilizan muy escasamente. En mi opinión, no se trata de falta de interés de parte de docentes y directivos, sino que muchas veces los textos técnicos sobre metodología de la investigación y, especialmente estadística, prestan poca o escasa atención a la necesaria motivación para adentrarse en la labor de recolectar y analizar datos escolares. Por otra parte, la disciplina estadística aparece vinculada estrechamente a la investigación académica y grandes volúmenes de datos, situación que aleja a las docentes de un genuino interés por coleccionar información de su escuela, analizarla y utilizarla para comprender su contexto.

Este texto no pretende ser un manual exhaustivo de técnicas de recolección y análisis de datos. Se trata más bien de una excusa para interesar a aquellos que trabajan en las escuelas y brindarles un conjunto de ejemplos sencillos de problemas que se pueden abordar desde la estadística. Es por ello que escogí elaborar un material accesible centrado en modelos estadísticos no paramétricos. Estos son modelos elegantes y potentes para trabajar con colecciones poco numerosas de datos, como pueden ser las poblaciones de alumnas y alumnos en las aulas, o incluso la misma escuela. Asimismo, los modelos no paramétricos prescinden de muchos supuestos intrínsecos en las técnicas de análisis paramétricas, para las cuales quedan invalidados los datos nominales y ordinales; paradójicamente, tales datos son los más frecuentes en los ámbitos escolares.

El libro se divide en varias etapas de complejidad creciente, comienza con la comparación de dos muestras independientes, seguido de la comparación de más de dos muestras. Asimismo, incorpora el análisis de muestras relacionadas, sean dos o más. En cada sección se esboza una breve introducción a un modelo de análisis estadístico, seguido de un ejemplo. Esta decisión se basó en la necesidad de no perder rigor analítico, pero sin abundar en detalles que terminen por

abrumar al público al que va dirigido. Es así que, si la lectura de la racionalidad técnica de un modelo se torna obtusa, se puede recurrir al ejemplo para aclarar el panorama. Por otra parte, cada ejemplo se resuelve mediante el uso de un software, esto obedece a que actualmente casi nadie emplea el cálculo manual para resolver los problemas. En este caso también espero que sirva de aliciente para que los interesados se embarquen en el estudio de estos programas. Aunque los utilizados en este texto son comerciales, existen numerosas alternativas libres que bien pueden suplantarlos. Al final del texto se dejan algunos enlaces informativos de proyectos de software libre para el análisis estadístico.

Espero que este material sirva como disparador de curiosidades por los análisis estadísticos aplicados a problemas educativos. Si es así, el lector o lectora encontrará mucha bibliografía especializada donde podrá enriquecer su conocimiento. La finalidad última de esta obra es animarlos en ese camino.

Análisis de Muestras Independientes

Prueba exacta de Fisher y contraste Chi cuadrado para tablas de 2x2

Esta es una prueba que sirve para analizar datos discretos tanto nominales como ordinales, cuando las muestras son pequeñas. En este caso, las variables en estudio se distribuyen en dos categorías mutuamente excluyentes y se supone que se han obtenido mediciones de esa variable en dos muestras independientes. En un diseño de muestras de este tipo, ambas deben ser obtenidas por uno de los siguientes métodos: a) pueden obtenerse al azar de dos poblaciones distintas, o b) pueden obtenerse de la misma población, pero sus miembros deberán ser asignados a dos tratamientos independientes y excluyentes. En ambos casos, no es necesario que el tamaño de la muestra sea el mismo. Las mediciones se representan en una tabla de contingencia de 2x2 como la siguiente:

Tabla de distribución de frecuencias: en celdas y marginales de fila y columna

Variable	Grupo I	Grupo II	Combinación
Categoría I	A	B	A+B
Categoría II	C	D	C+D
Totales	A+C	B+D	N

La variable que aquí se consigna como categoría I y II, pueden ser cualesquiera variables que admita una dicotomía, como, por ejemplo: por encima y por debajo de la media, deprimido – no deprimido, ansioso – estable, animo – desánimo, aciertos – errores, motivación alta o baja, etc.

La prueba exacta de Fisher, permite determinar si la proporción de casos en cada una de las celdas difiere entre los grupos analizados. Es decir que A, B, C y D, denotan las frecuencias absolutas en cada una de las categorías de las variables, según el grupo al que pertenecen las unidades de análisis.

El contraste que se realiza mediante la prueba exacta de Fisher, se basa en el supuesto de la hipótesis nula que asume que en la población los dos factores en la tabla no están relacionados. En otras palabras, la H_0 establece que no existe asociación entre las variables.

Dadas las frecuencias marginales en la tabla (A+B, C+D, A+C y B+D), y asumiendo el supuesto de la hipótesis nula, la prueba exacta de Fisher calcula la probabilidad de obtener frecuencias en las celdas, tanto o más desiguales que las que se hubieren observado, siendo cierta la H_0 .

El planteo estadístico de la prueba se representa como sigue:

$$H_0: \chi^2=0$$

$$H_1: \chi^2>0$$

Es decir, el estadístico F se contrasta mediante una distribución χ^2 (chi cuadrada); por lo tanto, al no existir asociación entre las variables, este estadístico será igual a cero o estará muy próximo a este valor. En caso de existir asociación entre las variables, el estadístico χ^2 se apartará de cero. Puesto que la distribución del estadístico χ^2 depende de los grados de libertad de la tabla, y esta prueba se basa en tablas 2x2, la distribución correspondiente es la de χ^2 con 1 grado de libertad.

Cuando el número de casos es pequeño, esta probabilidad puede ser calculada exactamente, contando todas las posibles tablas que pueden ser construidas en base a las frecuencias marginales. Esta es la explicación subyacente para la prueba exacta de Fisher, es decir computar la probabilidad exacta, bajo la H_0 , de obtener la distribución de frecuencias observadas en las celdas, o una que es aún más desigual. La prueba exacta de Fisher, es una de las más poderosas pruebas unidireccionales para analizar datos cuando las variables han sido medidas en escalas dicotómicas y nominales.

Ejemplo de aplicación

En una cátedra de la U.N.C. los docentes estuvieron interesados en determinar si la situación laboral de los alumnos influía en el puntaje promedio obtenido en las tres evaluaciones a lo largo del año en esa materia. Para ello, dividieron la muestra de alumnos en aquellos que trabajaban y los que no lo hacían. Asimismo, dividieron el puntaje promedio entre aquellos que obtuvieron siete o más en los exámenes. La tabla resultante es la que se presenta a continuación:

Tabla de contingencia para las variables rendimiento académico y situación laboral

Variable	Trabaja	No Trabaja	Combinación
menos de siete	35	22	57
Siete o más	20	39	59
Totales	55	61	116

La lectura de la tabla indica que la muestra está compuesta por 116 participantes (total), 59 de ellos han alcanzado promedios de siete o más altos (marginal de fila), mientras que 57 tienen promedios menores a siete (marginal de fila). Asimismo, se observa que, del total de alumnos

participantes, 55 son trabajadores (marginal de columna), mientras que 61 de ellos no lo hacen (marginal de columna).

Según el supuesto de la hipótesis nula, la situación laboral no guarda relación con el rendimiento académico. Con los datos obtenidos, se utilizó la prueba exacta de Fisher para determinar si la distribución de frecuencias observadas en la tabla, se condice con lo esperado por la H_0 . El análisis de datos se realizó con el software SPSS®. Un resumen de los resultados de las comparaciones se muestra en la tabla a continuación:

Resultados obtenidos en la prueba de Fisher

	Estadístico	p - valor
una cola	.076	>0.05
dos colas	.076	>0.05

Según lo que se muestra en los resultados, las variables se encuentran asociadas ya que la posibilidad de encontrar una tabla como la obtenida bajo el supuesto que la hipótesis nula es verdadera es menor a una probabilidad asociada de $p=0.05$. De esta manera, hemos podido descartar la hipótesis nula, aunque nos resta determinar la fuerza de la asociación. Para esto se puede aplicar la aproximación χ^2 en tablas de 2X2, mejorada por la corrección de Yates la cual reduce los valores absolutos de la diferencia entre las frecuencias observadas y esperadas, multiplicando por 0.5 antes de elevar al cuadrado. La corrección de Yates hace a la estimación más conservadora y se la aplica cuando la tabla contiene solo frecuencias observadas pequeñas, de modo que algunas frecuencias esperadas resultan menores que 10. La tabla que se muestra a continuación resulta del cálculo del estadístico χ^2 para los datos de nuestro ejemplo:

Estadístico obtenido a partir de la tabla de contingencia para las variables rendimiento académico y situación laboral

Chi-cuadrado (df=1)	8.800	p= 0.0030
---------------------	-------	-----------

Se constata que el estadístico χ^2 también es significativo, como era de esperar. Entonces se utiliza el coeficiente de asociación V de Cramer para obtener la magnitud de la asociación, tal como se desarrolla en la ecuación que sigue:

$$V = \sqrt{\frac{\chi^2}{n}} = \sqrt{\frac{8.8}{116}} = 0,27$$

El coeficiente de asociación obtenido es de una magnitud moderada a débil. Por lo tanto, es factible concluir que la situación laboral de los estudiantes se relaciona con su rendimiento académico, aunque no podría considerarse esta situación como determinante. Nos resta ahora determinar el sentido de esta asociación, para ello procedemos a reconstruir la tabla a partir de las frecuencias relativas transformadas a porcentajes.

Tabla de contingencia para las variables rendimiento académico y situación laboral en %

Variable	Trabaja	No Trabaja	Combinación
menos de siete	30.17%	18,96%	49,14%
Siete o más	17,24%	33,62%	50,86%
Totales	47,41%	52,58%	100%

Al reducir las celdas al porcentual relativo al total de casos, observamos que la categoría de aquellos alumnos que trabajan y obtienen menos de siete, es mayor a la observada en los alumnos trabajadores (30% contra 19% según redondeo de los porcentuales). La situación inversa se observa cuando consideramos aquellos alumnos que trabajan y obtienen notas de siete o más (17% contra 34%, según redondeo de los porcentuales). Habiendo hecho esto, es posible ampliar la conclusión afirmando que los alumnos trabajadores son los que tienden a obtener mayormente notas menores a siete. En contraste, los alumnos que no trabajan tienden a obtener notas de siete o superiores.

Prueba Binomial

Existen muestras en las cuales los individuos que la componen son clasificados en dos categorías amplias. v.g. Varón – Mujer; Soltero – Casado, Internado – Ambulatorio, etc. En tales casos, la población queda dividida en dos grupos y, por tanto, se denomina como población binaria o dicotómica. El investigador deberá tener presente que no todas las poblaciones pueden dicotomizarse de esta manera sin que se pierda información sustantiva. Pero en el caso en que esto sea posible, el conjunto de las observaciones podrá clasificarse como 0 y 1, o bien, como éxitos y fracasos. Este sistema de clasificación permite trabajar con variables nominales dicotómicas bajo el modelo estadístico binomial. Habiendo hecho esta clasificación en la población, tenemos que la probabilidad de muestrear un objeto en la primera categoría se denomina p , y la probabilidad de muestrear un objeto en la otra categoría se denomina q . Siendo X cada observación, se tiene que:

La probabilidad de $X = 1 = p$,

y

La probabilidad de $X = 0 = 1 - p = q$

Esto es, la probabilidad de que X sea igual a 1 se define como p , entonces la probabilidad de X sea igual a 0, queda definida como $1-p$. Un supuesto de la distribución binomial es que las probabilidades son constantes para todo X , independientemente del número de unidades muestreadas. Otra manera de decirlo es que los ensayos son independientes.

Aunque el valor de p puede variar de una población a otra, es un valor fijo para una determinada población. Aún si se conoce el valor de p para una población, no es posible esperar que una muestra aleatoria de esa población contenga exactamente los valores de p y $1-p$, ya que el muestreo aleatorio impide que la muestra duplique exactamente esos valores. Pero sí cabe esperar que los valores muestrales se aproximen con bastante fidelidad al parámetro. Por lo tanto, si conocemos la distribución del parámetro, es factible suponer que valores muestrales próximos al mismo, implican que la muestra proviene de dicha población; pero si existen diferencias marcadas entre los valores muestrales y los paramétricos, es lícito suponer que la muestra no ha sido extraída de la población a la que refiere el parámetro.

Si contamos con el valor paramétrico, es posible evaluar la probabilidad que los valores observados en una muestra, revelen desviaciones significativas del mismo. Lo cual nos indicaría que la muestra obtenida, no pertenece a la población de referencia. Por lo dicho, la hipótesis nula

en este caso determina la distancia entre la probabilidad estimada (población) y observada (muestra). Esta hipótesis adquiere la siguiente expresión estadística:

$$H_0 = p = p_o$$

$$H_1 = p \neq p_o$$

El planteo de hipótesis se lee de la siguiente forma: si las probabilidades observadas en la muestra (p_o), son iguales a las registradas en la población (p), se asume que la muestra bajo análisis ha sido extraída de dicha población. Si estas difieren sustancialmente, se invalida la hipótesis nula. La racionalidad de la prueba se basa en asumir que la H_0 es verdadera. La hipótesis será rechazada (o aceptada) cuando sea contrastada con los datos obtenidos de un experimento.

Por conveniencia cuando $X=1$ se denominan éxitos a estas observaciones; cuando $X=0$ los resultados se cuentan como fracasos. Nótese que este tipo de clasificación debe traducirse en términos operacionales según el tipo de variables que se estudian en el experimento. Además, en una serie de n observaciones, y es igual a la suma de todos los resultados cuando $X=1$. Entonces, en una muestra de tamaño n , la probabilidad de obtener k objetos en una categoría y $n - k$ objetos en la otra categoría, está dada por la siguiente ecuación:

$$P[y = k] = \binom{n}{k} p^k q^{n-k}$$

p es la proporción de observaciones esperadas cuando $X=1$

q es la proporción de observaciones esperadas cuando $X=0$

$\binom{n}{k}$ es el factorial de $n = n! / k! (n - k!)$

Luego será necesario formular la probabilidad de X bajo la hipótesis nula planteada, que se denota como B_x . Al comparar esta probabilidad con el nivel α adoptado para el experimento, se procederá a rechazar la hipótesis nula, dependiendo de los resultados obtenidos. La probabilidad de B_x , cuando ambos resultados son equiprobables es de $p=(B_x)= 1/2$. Destacamos que no todos los casos B_x , pueden ser tratados como equiprobables. Por ejemplo, la cantidad de nacimientos en un mes en una clínica de neonatología puede asimilarse a esta probabilidad, pero si estudiamos el fenómeno de pertenencia al quintil más alto y más bajo de la población, no podremos utilizar este supuesto ya que sabemos de antemano que la proporción de personas en el quintil más alto es siempre menor a los que quedan comprendidos en el quintil más bajo. Por lo tanto, en el caso en que la probabilidad exacta es conocida y diferente de $1/2$, será necesario calcularla de acuerdo con la hipótesis nula.

El cálculo de la probabilidad depende del tamaño de la muestra. Cuando esta es igual o mayor a 30, se utiliza la aproximación normal a la binomial $N(0,1)$, transformándose los valores de X de frecuencia observada a unidades tipificadas (generalmente puntuaciones Z). La prueba binomial se realiza mediante una aproximación a la normal si se cumplen las siguientes condiciones: el tamaño de la muestra n debe ser mayor o igual que 30 y el mínimo de $n \cdot p$ y $n \cdot (1-p)$ debe ser mayor que 5, siendo p la probabilidad de éxitos de la variable binomial.

Ejemplo con probabilidad de $B_x = 1/2$

En este ejemplo optaremos por una demostración sencilla de la aplicación de la prueba binomial, dado que la misma ofrece múltiples aplicaciones.

En una selección de personal se tomaron diferentes pruebas psicomotoras a personas que deberían realizar trabajos de precisión en el armado de circuitos electrónicos. El encargado de dicha selección dividió por la mediana a las personas seleccionadas de acuerdo a los puntajes combinados de dos de esas pruebas. Al observar los resultados e intuyendo un proceso de distribución binomial, decidió realizar esta prueba con probabilidad $B_x=0.5$. En este ejemplo se utilizó el software Statistica® para el procesamiento de datos. Los resultados se muestran en la tabla a continuación.

Test Binomial

	Categoría	n	Proporción Observada	Test Prop.	Significación (2-colas) ^a
Grupo 1	sobre Mdn	35	0.49	0.50	0.906
Grupo 2	bajo Mdn	37	0.51		
Total		72	1.00		

^a Basada en Aproximación Z.

En este caso, se ha utilizado la prueba binomial para comparar la diferencia de proporciones entre los grupos. Aunque no es el uso habitual de esta prueba, sirve como ejemplo de un test simple para determinar si dos grupos provienen de poblaciones centradas en diferentes parámetros. En este ejemplo en particular, es posible suponer que los individuos muestren diferencias en el rendimiento de acuerdo a los puntajes de los test con que han sido evaluados, y por tanto hubiera proporciones distintas. Los resultados muestran que tal hipótesis no se sostiene.

Prueba chi cuadrada χ^2

A menudo los atributos estudiados solo pueden ser medidos en escalas nominales o bien ordinales, con solo unas pocas categorías amplias. Los estudios en estos casos, se interesan por conocer si existe relación o asociación entre las variables medidas de esta manera. La prueba chi cuadrada (χ^2), permite decidir si existe relación entre dos variables medidas en escalas nominales u ordinales con pocas categorías. Al describir la prueba exacta de Fisher ya hemos introducido el contraste basado en χ^2 , por lo tanto, comenzaremos con un ejemplo que desarrollaremos para comprender mejor esta técnica de análisis.

Ejemplo de aplicación

En una investigación se analizó si el trabajo de las madres se relacionaba con el rendimiento escolar de sus hijos. La muestra en estudio estuvo compuesta por 43 niños de 9 y 10 años. La muestra estuvo distribuida como se muestra en la tabla a continuación:

Rendimiento de los escolares según si la madre sea trabajadora

	Trabaja= si	Trabaja= no	Total
Bueno	13	9	22
Regular	4	17	21
Total	17	26	43

Nótese que el arreglo de la tabla es similar a la que se presentó en el apartado de la prueba exacta de Fisher utilizando el contraste chi cuadrado. A los fines de este ejemplo, solo utilizaremos el contraste χ^2 , dado que la prueba de Fisher solo aplica a tablas 2x2, mientras que el contraste basado en χ^2 puede aplicarse a tablas con mayor número de celdas (filas y columnas). En otras palabras, el ejemplo que se desarrollará puede extenderse a otras tablas de contingencia con mayor número de filas y columnas.

La hipótesis nula para el cálculo del chi cuadrado

La H_0 en este caso, refleja la situación en la que los atributos son independientes entre sí. En este ejemplo, la H_0 estipula que no existe asociación entre el rendimiento escolar del niño y el hecho de que la madre trabaja o no. Esto equivale a decir que la proporción de madres que trabajan y tienen niños con buen rendimiento escolar, será similar a la proporción de madres que no trabajan y tienen niños con buen rendimiento escolar.

Por lo antes dicho, la H_1 para este ejemplo, estipula que los atributos están relacionados, por lo tanto, existirá asociación entre ellos. Si la H_1 es direccional, se estimará que la proporción de madres que trabajan y que tienen niños con rendimiento escolar regular, será mayor que la proporción de madres que no trabajan y que tienen niños con rendimiento escolar regular (la H_1 podría haberse propuesto a la inversa, pero es menos verosímil).

Partiendo del supuesto que la H_0 es correcta, y que las muestras son aleatorias e independientes, es posible calcular desde las frecuencias marginales, el número de niños en cada grupo de madres que se espera que tengan un buen rendimiento escolar. Luego será posible observar si existen diferencias entre esas frecuencias esperadas (si la H_0 es cierta) y las que han sido efectivamente observadas en la población.

La prueba χ^2 se basa precisamente en la comparación de las frecuencias observadas (f_o), con aquellas frecuencias que cabría esperar (f_e), si fuera cierta la H_0 . Si las frecuencias observadas son iguales o semejantes a las frecuencias esperadas (según se espera si fuera cierta la H_0), entonces los atributos son independientes y no hay asociación. Como corolario, se aceptará la H_0 . Al contrario, si las frecuencias observadas son diferentes de las frecuencias esperadas, (siempre bajo el supuesto que es cierta la H_0), entonces los atributos no son independientes y se rechazará la H_0 . En términos estadísticos las hipótesis se plantean de la siguiente manera:

$$H_0: f_o \cong f_e; \chi^2 \cong 0$$

$$H_1: f_o \neq f_e; \chi^2 > 0$$

Nótese que en el caso en que no hay asociación entre las variables las frecuencias observadas y esperadas son prácticamente iguales, por tanto χ^2 tiende a cero. Al contrario, cuando la asociación entre las variables es importante, las frecuencias observadas y esperadas son diferentes y χ^2 tiende a ser mayor que cero.

Cálculo de las frecuencias esperadas

Para calcular las frecuencias esperadas en cada celda cuando las muestras son aleatorias e independientes, basta multiplicar las frecuencias marginales de la fila y la columna a las que pertenece la celda y dividir este producto por el número total de casos. Por ejemplo: la frecuencia esperada para la categoría Madre que trabaja – buen rendimiento escolar del hijo, se obtiene del producto $(17 * 22) / 43$, por ende, la frecuencia esperada en esa categoría es: 8,69. Al repetir este mismo procedimiento en todas las celdas se obtienen las frecuencias esperadas para la tabla completa. A continuación, se muestra el procedimiento resumido en la siguiente tabla:

Relación entre las frecuencias observadas y esperadas: método de cálculo

	Trabaja= si	Trabaja= no	Total
Bueno	fo=13 fe=(17*22)/43=8,69	fo=9 fe=(26*22)/43=13,3	22
Regular	fo= 4 fe=(17*21)/43=8,3	fo=17 fe=(26*21)/43=12,69	21
Total	17	26	43

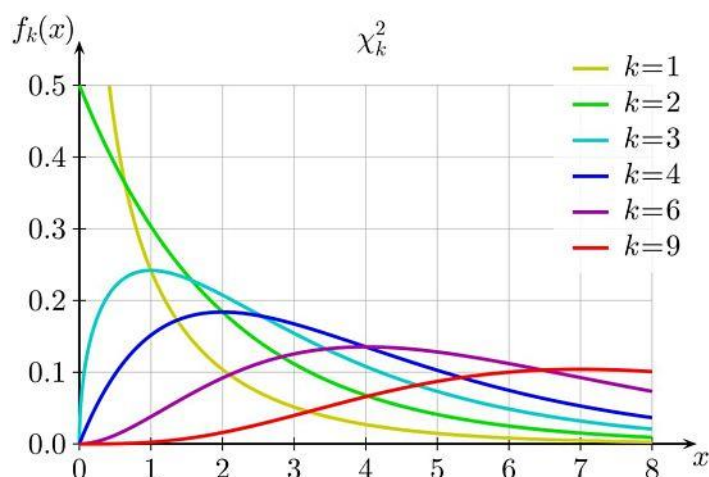
Hemos obtenido de esta manera las frecuencias esperadas bajo la hipótesis nula, a partir de la multiplicación de los marginales de fila y columna divididos por el total de casos. Señalamos entonces que las frecuencias esperadas reciben ese nombre porque son las que hubiéramos observado de ser verdadera la hipótesis nula, entonces, el estadístico χ^2 servirá para estimar las diferencias entre las frecuencias observadas y esperadas, y con ello determinar si existe (o no), asociación entre las variables.

El estadístico χ^2 , es una comparación entre las frecuencias observadas y las esperadas, bajo la hipótesis de la independencia de los eventos (H_0). Su cálculo se representa por la siguiente fórmula:

$$\chi^2 = \frac{\sum (fo - fe)^2}{fe}$$

El estadístico χ^2 , se obtiene tomando el cuadrado de la diferencia entre las frecuencias observadas y las esperadas de ser cierta la H_0 , y dividiendo ese cuadrado por la frecuencia esperada. La sumatoria de ese procedimiento a través de todas las celdas de la tabla, arroja el valor de χ^2 . En el supuesto de la H_0 , la diferencia entre las fo y las fe , sería cero. Así, bajo el supuesto de la H_0 , el valor de χ^2 , será cero. Al contrario, cuanto mayor la diferencia entre fo y fe , el valor de χ^2 , será distinto de cero. En una situación real de experimentación, la diferencia $fo - fe$, serán distintas de cero. El problema estriba entonces en decidir que probabilidad existe que, dada una diferencia entre las frecuencias, ésta se deba al mero azar u otros factores causales. Para ello se utiliza la distribución de muestreo de la χ^2 . Cabe destacar que la distribución del estadístico depende de los grados de libertad de la tabla, esto es lo mismo que decir que χ^2 depende de la cantidad de filas y columnas que contiene la tabla de contingencia.

Grafica de distribución del estadístico χ^2



En esta gráfica se muestra la distribución del estadístico de acuerdo a los grados de libertad que posee la tabla de contingencia. Como referencia se debe tener en cuenta que la tabla más pequeña que se puede construir tiene dos filas y dos columnas, por lo que sus grados de libertad es igual a uno ($k=1$). A medida que se agregan filas y columnas las tablas se hacen más grandes y sus grados de libertad aumentan, dando lugar a una familia de curvas que se distribuyen según este parámetro ($k= 1, 2, 3 \dots n$).

La distribución χ^2 , si fuera cierta la H_0 , es asimétrica. Se aprecia que si la H_0 , es cierta la probabilidad de encontrar valores cercanos a cero es alta, y se hace cada vez menor cuando el valor de χ^2 se aleja de cero.

Los grados de libertad en tablas de contingencia expresan el número de celdas cuyas frecuencias pueden variar libremente, si se conocen las frecuencias marginales. En una tabla de 2×2 , los grados de libertad serán igual a uno. Así, el cálculo de los grados de libertad para una tabla de doble entrada con sendas filas y columnas, se calcula a partir de la siguiente ecuación:

$$gl = (filas - 1) * (columnas - 1)$$

En el caso que estamos usando como ejemplo se verifica que $gl = (2-1) * (2-1) = 1$

Tabla de distribución de χ^2

Los valores críticos de χ^2 se consignan en una tabla donde las columnas indican los grados de libertad y las hileras los niveles de significación. Dado un valor de χ^2 , con un determinado nivel de significación y determinados grados de libertad, se busca el valor crítico en la tabla y se compara con el valor observado bajo la H_0 . Si el valor observado es igual o mayor al registro de la tabla, podemos rechazar la H_0 , de acuerdo a un valor de significación establecido. Esto significa

que las variables se encuentran asociadas. El paso siguiente es encontrar la magnitud de dicha asociación. En los programas estadísticos este paso se resume en la probabilidad de encontrar el valor de χ^2 bajo la H_0 . Cuanto más bajo este valor, menos argumentos existen para sostener la idea de independencia de los atributos. Por convención suele usarse el valor $p=0.05$ como punto de corte a partir del cual se puede rechazar H_0 .

El grado de asociación entre las variables

La prueba χ^2 proporciona información acerca de la existencia o no de asociación entre variables (y si tal asociación es diferente de la esperada por azar), pero no dice nada acerca de la magnitud de la asociación. Cuando interesa conocer el grado de asociación entre las variables, se emplea generalmente el coeficiente de contingencia C de Pearson, el cual se define como: la raíz cuadrada del cociente entre el valor de χ^2 y el valor de χ^2 más el número de casos n .

Cálculo del coeficiente de asociación C de Pearson

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}}$$

El coeficiente de asociación C de Pearson, (que no debe ser confundido con el coeficiente de correlación r de Pearson), estima la asociación entre atributos (variables). Cuando estos son independientes, el coeficiente es igual a cero, y se hace mayor a medida que aumenta la dependencia de los atributos. El valor máximo del coeficiente depende del número de hileras y filas en la tabla de contingencia, pero es siempre menor que la unidad. El hecho que este coeficiente no tenga un valor máximo establecido para cualquier tabla, dificulta su interpretación, pero al ser constante permite la comparación entre resultados obtenidos con tablas de igual tamaño (el valor máximo que toma el coeficiente C , se ha calculado para tablas de distinto tamaño y se pueden encontrar publicadas en algunos libros de estadística).

Otro coeficiente de asociación que puede utilizarse para estimar el grado de asociación entre los atributos es el coeficiente V de Cramer, cuya fórmula es la siguiente:

Formula del coeficiente V de Cramer

$$V = \sqrt{\frac{\chi^2}{\min(f - 1; c - 1) * n}}$$

En la fórmula la expresión $\min. (f - 1; c - 1)$ alude al número de filas o columnas menos uno. En la fórmula entrará la expresión que tenga el menor valor de los dos. Entonces V se define como: la raíz cuadrada del cociente entre el valor χ^2 observado y el número de columnas o hileras menos

uno (dependiendo de cuál sea el menor), multiplicado por el número de casos. Este coeficiente puede aplicarse a tablas de cualquier tamaño y alcanza un máximo de uno.

Advertencias en el uso de la prueba χ^2

La prueba requiere que las frecuencias esperadas en cada celda no sean demasiado pequeñas, ya que, si este es el caso, la prueba pierde potencia. Las recomendaciones más citadas para la correcta aplicación de la prueba son:

Tablas de 2x2; $gl=1$: cuando n es igual o menor que 20, se recomienda el uso de la prueba exacta de Fisher. Cuando n es mayor que 20 y menor que 40, la prueba χ^2 puede utilizarse si las frecuencias esperadas alcanzan un valor no menor que 5. Si existen frecuencias esperadas menores que 5, conviene utilizar la prueba exacta de Fisher.

Tablas de $n \times k$; $gl>1$: puede utilizarse la prueba χ^2 si menos del 20% de las celdas registran valores menores que 5 en las frecuencias esperadas, y si no hay celdas con frecuencias esperadas menor que 1. Si estos requisitos no son cubiertos por los datos en la forma en que originalmente fueron recolectados, se recomienda combinar las categorías adyacentes para incrementar las frecuencias esperadas en las celdas que se traten. Esto ajusta los valores observados de χ^2 a la distribución teórica.

Resolución del ejemplo

El planteo del ejemplo requiere determinar si existe asociación entre dos atributos que son: el rendimiento lector de los estudiantes y la situación laboral de la madre. Mediante el estadístico χ^2 es posible determinar si se cumple que las variables están relacionadas, luego y si efectivamente los atributos guardan relación, calcular el grado o magnitud de la misma. Para resolver el ejemplo recurrimos al software estadístico Infostat®, cuyo resultado se muestra en la siguiente tabla:

Resultado del análisis de datos Infostat

Estadístico	Valor	gl	Valor p
Chi Cuadrado Pearson	7,21	1	0,0073
Coef.Conting.Cramer	0,29		
Coef.Conting.Pearson	0,38		

El primer contraste nos informa que el valor de probabilidad de ocurrencia de χ^2 es menor al valor crítico establecido: $p=0,0073 < p=0,05$, por lo cual es factible desestimar la hipótesis nula. En tal caso, asumimos que existe asociación entre las variables. En otras palabras, se aprecia que la situación laboral de la madre se asocia al rendimiento lector de las hijas e hijos. La siguiente

información que se ofrece es la magnitud de la asociación para las variables. En este caso, el coeficiente $C=0,38$ y $V=0,29$. La diferencia de magnitud entre ambos coeficientes se debe a que V de Cramer es más conservador que C de Pearson. En cualquier caso, los coeficientes determinan que hay asociación y que la misma es de magnitud moderada. Por lo tanto, es de esperar que la situación laboral de la madre nos informe algo sobre el posible rendimiento de los escolares.

Aplicación de χ^2 a la prueba de la mediana

Algunos experimentos se interesan por conocer si existen diferencias en la tendencia central, cuando los datos pertenecen a escalas ordinales. Si las muestras son independientes, se puede someter a prueba la hipótesis de que ambas provienen de una misma población, y tendrán por ello una tendencia central similar. Por tratarse de atributos medidos en escalas ordinales, se utiliza la prueba de la mediana. En este caso la hipótesis nula postula que:

$$H_0: \theta_1 = \theta_2$$

$$H_1: \theta_1 \neq \theta_2$$

La letra griega theta (θ) se utiliza para simbolizar la mediana. La hipótesis de investigación H_1 postulará que las dos medianas son diferentes si es bidireccional, o bien que una es mayor que la otra si es unidireccional.

La mediana combinada

La prueba de la mediana consiste en calcular la mediana de las dos muestras combinada (θ_c). Estos datos pueden resumirse en una tabla 2x2, donde las columnas denoten los valores de la mediana del grupo 1 (θ_1) y la mediana del grupo 2 (θ_2). Las hileras denotarán en este caso los valores que se encuentran por encima de la mediana combinada y por debajo de ella. Esta tabla puede resumirse de la siguiente manera:

	Muestra 1 (θ_1)	Muestra 2 (θ_2)
Valores por encima de la mediana combinada (θ_c)		
Valores por debajo de la mediana combinada (θ_c)		

En una tabla de este tipo tendremos que, si las medianas calculadas para las dos muestras son iguales o están muy próximas entre sí, la proporción de casos que caerán por encima y por debajo de la mediana combinada serán similares. Al contrario, si las dos muestras difieren en cuanto a la tendencia central, la proporción de casos que caen por encima y por debajo de la mediana será diferente. En otras palabras, la muestra que tenga la mayor tendencia central, acumulará más casos en la celda correspondiente a la categoría valores por encima de la mediana combinada, y viceversa. Así se apreciará la concentración de casos opuestos por el vértice, que es propia cuando los atributos están asociados. Bajo tales circunstancias, el estadístico χ^2 será significativo. Por lo tanto, si dos muestras difieren en cuanto a su mediana, el procedimiento descrito permite utilizar

una tabla de 2x2 y calcular un valor de χ^2 . Si este resulta significativo, se podrá rechazar la hipótesis nula de ausencia de asociación entre variables. Aplicado a la prueba de la mediana, permitirá rechazar también la hipótesis nula que postula que las medianas son iguales por haber sido extraídas de la misma población.

Nota: si el número total de casos en ambos grupos es pequeño, puede utilizarse la prueba exacta de Fisher. Si el número total de casos es grande y los $gl = 1$, puede utilizarse la prueba de χ^2 . En estos casos lo que se evalúa siempre es la hipótesis nula de no diferencia en torno a la mediana combinada.

Ejemplo:

En una empresa se desea conocer si el salario inicial de los empleados se distribuye de manera proporcional entre las mujeres y los varones. Puesto que, al momento de entrada en actividad de los empleados, todos cuentan con las mismas antigüedades, la distribución del salario solo se ve afectada por el puesto que ocupa. En este sentido los salarios más bajo corresponden a los empleados administrativos, mientras que los más altos se otorgan a quienes ocupan cargos gerenciales. La hipótesis de investigación refiere a que la proporción de mujeres y varones no es la misma según la categoría del empleo, lo cual se verá reflejado en la disparidad de mujeres y varones en la distribución por grupos de salarios. La hipótesis deriva de la inspección de la tabla que se presenta a continuación.

Tabla de contingencia Genero y Percentil por Grupo en Salario Inicial

Genero	Grupos en Salario Inicial				Total
	1	2	3	4	
Hombre	6	35	120	97	258
Mujer	113	58	24	21	216
Total	119	93	144	118	474

Como puede observarse, del total de empleados el 45,6% son mujeres, y el 54,4% son varones; la diferencia porcentual según género es mínima (diferencia=8,8%). Los rangos de salario se dividieron en cuatro, de menor a mayor, donde los sueldos más bajos corresponden a la categoría 1, y los más altos a la categoría 4. La distribución porcentual es: 1= 25,1%; 2=19,6%; 3=30,4% y 4=24,9%. Las frecuencias en las celdas muestran que las mujeres tienden a agruparse en las categorías salariales más bajas, mientras que los varones lo hacen en las categorías más altas. Corresponde determinar si la hipótesis de investigación se sostiene, empleando para ello el estadístico χ^2 .

$$H_0: \chi^2 \approx 0$$

$$H_1: \chi^2 > 0$$

De no existir asociación entre las variables (los atributos son independientes), el estadístico tiende a cero, al contrario, si existe asociación entre las variables (los atributos no son independientes), el estadístico tiende a ser mayor que cero. La tabla que se muestra a continuación muestra el resultado del cálculo de la prueba χ^2 y su correspondiente valor de probabilidad bajo la hipótesis nula. El cálculo del estadístico se realizó utilizando el software SPSS®.

Prueba chi-cuadrado

	Valor	gl	Significación
Chi-cuadrado de Pearson	212,797	3	p<0,000
N de casos válidos	474		

Según muestra la tabla la probabilidad de haber obtenido el valor de $\chi^2 = 212,797$ bajo el supuesto que la hipótesis nula es verdadera, es menor al 1 por mil. Por lo tanto, corresponde rechazar tal hipótesis. Es posible entonces analizar el grado de asociación entre las variables. En la tabla que se muestra a continuación, se ofrece el valor de los coeficientes de asociación V de Cramer y C de Pearson.

Medidas simétricas de asociación

	Valor
V de Cramer	0,670
N de casos válidos	474

Según lo muestra la tabla, el coeficiente V de Cramer, revela que el grado de asociación entre las variables es fuerte. Por lo tanto, cabe concluir que, tal como lo sospechaban los investigadores, existe un sesgo en la distribución en los puestos de trabajos según el género de los empleados.

Prueba W de Wilcoxon y U de Mann – Whitney

Esta prueba permite la comparación de dos grupos independientes en una variable de tratamiento. La hipótesis nula supone que ambos grupos fueron extraídos de la misma población. La condición para su utilización es que los datos extraídos deben corresponder al menos a una escala ordinal. Por sus características, la prueba es un buen sustituto de la prueba paramétrica t de Student, cuando los supuestos para ésta no se cumplen o cuando los datos no han sido recogidos en escala de intervalo.

La hipótesis nula

Como se mencionó, la prueba permite comparar dos grupos, v.g. X e Y , bajo el supuesto de que ambos fueron extraídos de la misma población y que por ello tendrán la misma distribución. Este supuesto es el de la H_0 . La hipótesis alternativa H_1 parte del supuesto de que uno de los grupos, por caso X , no pertenece a la misma población donde se extrajo Y . La comparación implica el contraste entre las medianas de ambos grupos, considerando que, si ambos fueron extraídos de la misma población, sus medianas no serán estadísticamente diferentes. La H_1 puede aceptarse si la probabilidad de $p(X > Y) > 1/2$. Esto es, la mayor parte de las observaciones extraídas de la población X , serán mayores que las extraídas de la población Y . Al contrario, se deberá aceptar la H_0 cuando $p(X > Y) = 1/2$. Esto es, la probabilidad de encontrar valores observados de X que son mayores que Y , es igual a la probabilidad de encontrar valores de Y mayores que X . Una forma de plantear las hipótesis nula y alternativa es considerando los valores de las medianas de los grupos X e Y . En este caso:

$H_0: \theta_X = \theta_Y$ La mediana de los grupos X e Y serán iguales o estarán muy próximas

H_1 unidireccional $\theta_X > \theta_Y$ La mediana del grupo X será mayor que la del grupo Y (o viceversa)

H_1 bidireccional $\theta_X \neq \theta_Y$ Las medianas de los grupos X e Y serán diferentes.

La mediana combinada y el estadístico W

Para calcular el estadístico W , se combinan las observaciones de los grupos X e Y , luego se ordenan por rangos de manera ascendente. En este ordenamiento se considera el tamaño algebraico, en donde los rangos inferiores tendrán los valores negativos mayores (en caso de existir). Si se considera el grupo X con m observaciones, se puede calcular el valor de W_x , que corresponde a la sumatoria de todos los rangos para ese grupo. De la misma manera es posible calcular el valor de W_y .

Para comprender mejor el concepto de sumatoria de rangos cabe recordar la definición de sumatoria de números enteros: el resultado de la sumatoria de números enteros, $1+2+3+4+\dots+n$, se calcula con la siguiente ecuación:

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

Por lo tanto, para una muestra cuyo n es igual a la suma del tamaño de los grupos X e Y, la sumatoria de rangos para ambos grupos, deberá ser igual a la sumatoria de los rangos del grupo combinado, esto se expresa en la siguiente ecuación:

$$W_x + W_y = n = \frac{n(n+1)}{2}$$

Bajo el supuesto que la H_0 es verdadera se esperaría que el promedio de rangos en los grupos X e Y fueran aproximadamente iguales. Ahora si la sumatoria de rangos es muy grande o muy pequeña en uno de los grupos, entonces es lógico suponer que las muestras no se extrajeron de la misma población. La distribución muestral del estadístico W cuando la H_0 es verdadera es conocido, y con ese dato se puede determinar la probabilidad asociada con la ocurrencia según la H_0 , de cualquier valor de W tan extremo como el valor observado.

Cuando las muestras de los grupos X e Y son pequeñas, el estadístico W tiene una distribución conocida. Sin embargo, conforme se incrementa el número de sujetos en las muestras la distribución muestral de W se aproxima a una distribución normal con media y varianzas conocidas. Por lo tanto, los parámetros media y varianza para esta distribución se calculan mediante las siguientes ecuaciones:

$$\mu_w = \frac{n_1(n+1)}{2}$$

$$\sigma_w^2 = \frac{n_2(n+1)}{12}$$

En este caso n_1 es el tamaño del grupo X y n_2 el tamaño de grupo Y, por lo tanto, el tamaño de la muestra total se obtiene de la suma de $n_1 + n_2 = n$. La significación del valor W se determina por medio de la puntuación estandarizada Z . La ecuación se muestra a continuación.

$$Z = W \pm 0.5 - \frac{\mu_w}{\sigma_w^2}$$

Observaciones Empatadas

Puesto que las variables se arreglan de acuerdo al rango de X e Y, pueden surgir empates en los rangos. Cuando existe dos o más empates, se supone que las observaciones son realmente diferentes, pero dado el nivel en el que han sido medidas imposibilita determinar la magnitud de tales diferencias. Para resolver los empates se da a cada observación empatada el promedio de los rangos que tendrían si no hubieran ocurrido los empates. Por ejemplo, si tres observaciones se hallan empatadas en la primera posición (la inferior), a cada puntuación se le asignará el rango de 2 que resulta del promedio de los rangos asignados, esto es $(1 + 2 + 3) / 3 = 2$, el siguiente rango que se asignaría sería el 4, porque los rangos 1, 2 y 3 ya fueron asignados para resolver los empates mediante el cálculo del rango promedio.

Ejemplo: En una escuela, se ha decidido analizar el rendimiento de los alumnos de sexto grado en la evaluación final integradora de lengua. Sucede que las docentes deciden para este estudio realizar una comparación entre alumnos que asistieron al turno tarde y quienes lo hicieron en el turno mañana. Para descartar que exista un sesgo en la evaluación según el turno al que asisten los alumnos, tomaron dos muestras aleatorias de veinte alumnos; cada muestra representa un conjunto de escolares para cada turno. Puesto que la evaluación final integradora no corresponde estrictamente con una escala métrica, decidieron utilizar para la comparación de grupos una alternativa no paramétrica. La prueba W de Wilcoxon y U de Mann-Whitney. El análisis de datos se realizó utilizando el software Statistica®. El resultado de este estudio piloto se muestra en la tabla que se presenta a continuación:

Estadísticos descriptivos para cada grupo

Rango de Grupos	n	Promedio de Rangos	Suma de Rangos
Turno Mañana	11	10.09	111.00
Turno Tarde	9	11.00	99.00
Total	20		

En la tabla se muestran el tamaño de cada grupo según el turno al que pertenece el escolar, conformando el tamaño total de la muestra. Dado que se trata de un estudio piloto y las docentes no adelantan ninguna diferencia en relación al turno en que cursan los estudiantes, se suponen implícitas las siguientes hipótesis de trabajo:

$H_0: \theta_{\text{Turno Tarde}} = \theta_{\text{Turno Mañana}}$: esto es, las medianas de los grupos de ambos turnos serán iguales (o estarán muy próximas entre sí).

H_1 unidireccional $\theta_{\text{Turno Tarde}} \neq \theta_{\text{Turno Mañana}}$: esto es, las medianas de los grupos será diferente.

Por lo que puede observarse en la tabla anterior, se aprecia que el promedio de los rangos no difiere sustantivamente, por lo que se espera que no existan diferencias entre los grupos. Para completar este contraste de hipótesis, se recurre a la distribución del estadístico W y la probabilidad de ocurrencia bajo la hipótesis nula.

Contraste estadístico U y W

Mann-Whitney U	45.0
Wilcoxon W	111.0
Z	-0.343
Signification Bilateral	0.732

Como muestra la tabla, el contraste estadístico revela una probabilidad de 0,7, que para el caso de la hipótesis nula se considera alta. En otras palabras, el supuesto de no existencia de diferencias en el rendimiento de los alumnos de ambos turnos que tomaron parte en el estudio, se sostiene. Dado que éste es un ensayo piloto llevado adelante por las docentes basado en una muestra aleatoria del total de alumnos que participaron en la evaluación, existen razones fundamentadas para descartar la suposición que el turno al que asiste los alumnos haya tenido una influencia sobre el rendimiento.

Prueba poderosa de rangos ordenados

Como se observa en la tabla anterior, el contraste estadístico incluye la prueba de U de Mann – Whitney. La razón de que ambos estadísticos se presenten juntos, es que tanto la prueba de Wilcoxon como la de Mann – Whitney supone para la H_0 que los grupos a comparar fueron tomados de la misma población, por lo tanto, los datos tendrán distribuciones iguales, lo cual implica además que la variabilidad de los datos (más precisamente las varianzas) serán iguales. Sin embargo, en ocasiones se desea probar la $H_0: \theta_X = \theta_Y$, sin suponer que las distribuciones subyacentes son las mismas, pues éstas no se pueden conocer previamente, o bien puede existir el problema de rangos restringidos. La diferencia en las distribuciones es conocido en estadística como el problema de *Behrens – Fisher*, y cuando el investigador sospecha del mismo, la prueba de Wilcoxon puede contener sesgos. Se recomienda utilizar entonces una alternativa denominada prueba poderosa para rangos ordenados. El procedimiento inicial es el mismo que para la prueba Wilcoxon, salvo que, una vez ordenados los valores de los grupos X e Y , se calcula para cada valor de X , el número de observaciones de Y con un rango inferior. Este procedimiento permite conocer la ubicación de las puntuaciones X y se denomina con U (YX). Luego se calcula la media de U (YX), mediante la siguiente ecuación:

$$U(YX) = \frac{\sum_{i=1}^n U(YX)}{m}$$

donde m es el tamaño del grupo X.

De la misma manera, se debe encontrar las ubicaciones de cada valor de Y, esto es para cada valor de Y, el número de X con un rango inferior. Esto se denomina $U(XY)$. Luego se calcula la media de $U(XY)$, mediante la siguiente ecuación:

$$U(XY) = \frac{\sum_{i=1}^n U(XY)}{n}$$

donde n es el tamaño del grupo Y.

El siguiente paso consiste en encontrar un índice de variabilidad de $U(YX)$ y de $U(XY)$, los cuales se calculan mediante las siguientes ecuaciones:

$$V_x = [U(YX_i) - U(YX)]^2$$

donde la variabilidad se representa por el cuadrado de la desviación de cada valor de $U(YX_i)$, respecto del promedio (puede apreciarse en este caso que es una aproximación al cálculo de varianza de los rangos de X respecto de Y).

De la misma forma, el índice de variabilidad de $U(XY_i)$ se calcula de la siguiente forma:

$$V_y = [U(XY_i) - U(XY)]^2$$

igualmente este cálculo representa la variabilidad de Y

Obtenidos ambos índices de variabilidad, el estadístico U se calcula mediante la siguiente ecuación:

$$U = \frac{m U(YX) - n U(XY)}{\sqrt{V_x + V_y + U(YX) * U(XY)}}$$

Existe una tabulación para el estadístico U, cuando las muestras son pequeñas. Conforme se aumenta el tamaño de las muestras, la distribución de U se aproxima a la distribución normal. Estos valores pueden utilizarse para evaluar la probabilidad de haber obtenido un valor dado de U bajo el supuesto que la H_0 es verdadera. Puesto que esta distribución es la misma que se utiliza para Wilcoxon – Mann – Whitney, el valor del estadístico W y U se presentan juntos.

En el ejemplo que hemos utilizado en este apartado, el estadístico U podría haberse interpretado si las calificaciones se hubieran hecho sobre un rango restringido de valores, por ejemplo, en vez de usar una escala de 1 a 10, se hubiese empleado una escala de 1 a 4, lo que hubiera supuesto un problema para el cálculo de los empates en la asignación de los rangos.

Nota final: en este trabajo hemos presentado de manera separada el cálculo del estadístico W y U , pero dada la similitud entre ambos, muchos manuales de estadística lo presentan como una prueba única bajo el título de prueba Mann-Whitney-Wilcoxon, prueba de suma de rangos Wilcoxon, o prueba de Wilcoxon-Mann-Whitney).

Prueba de Kolmogorov – Smirnov para dos muestras

Esta prueba permite evaluar si dos muestras independientes se han extraído de la misma población, o de dos poblaciones con similar parámetro. Se trata de una prueba unidireccional, y permite decidir si los datos de la población de donde se extrajo una de las muestras, son estocásticamente mayores que los valores de la población de donde se extrajo la otra muestra. Por su definición es una excelente alternativa para evaluar la situación en la que se plantee un experimento con dos grupos. En tal caso, el estadístico se utiliza para probar la predicción de que las puntuaciones del grupo experimental serán mayores que las puntuaciones de grupo control.

La prueba Kolmogorov – Smirnov, (KS por su abreviatura), centra su interés en el acuerdo entre dos distribuciones acumulativas, que en este caso es el acuerdo entre dos conjuntos de valores muestrales. Si las dos muestras han sido extraídas de la misma población (o de dos poblaciones con idéntico parámetro), entonces las distribuciones acumulativas de ambas tendrían que estar muy cercanas. Las diferencias entre ambas, serían las mismas que se observarían por el efecto del error aleatorio. Pero si las distribuciones acumulativas de las dos muestras manifiestan ser dispares, esto sugiere que las mismas provienen de dos poblaciones distintas. De este modo, una desviación suficientemente grande entre las distribuciones acumulativas de las dos muestras, es la evidencia que se requiere para rechazar la H_0 .

La hipótesis nula en el contexto de un experimento

La prueba KS es particularmente apta para evaluar los resultados de experimentos con muestras pequeñas cuya variable dependiente esté expresada en escala ordinal. En este apartado repasaremos brevemente los elementos esenciales de un experimento controlado. En primer lugar, supongamos que tenemos una muestra de individuos y seleccionamos de ella dos grupos, asignando cada individuo aleatoriamente a uno y otro, generando de este modo un grupo A y otro B. Luego medimos una variable x , sea esta cualquier variable que se exprese en una escala ordinal. Finalmente, comparamos ambos grupos. El resultado esperado en este caso es que el efecto de tal comparación expresara que no existen diferencias entre A y B en los niveles de la variable x , por cuanto no hemos realizado ninguna manipulación en uno de los grupos. Dicho de otro modo, las únicas diferencias apreciables entre A y B son debidas al azar. Para que se trate de un verdadero experimento deberíamos realizar una manipulación sobre uno de los grupos. Generalmente, esto se lo denomina tratamiento y consiste en someter a uno de los grupos a una manipulación controlada. En nuestro ejemplo, digamos que el grupo B ha sido sometido a un tratamiento que se espera produzca un aumento en los niveles de la variable x , que ahora pasa a llamarse variable dependiente. La variable independiente será el tratamiento y los grupos se diferencian por haber

sido sometidos (o no) al tratamiento. El grupo A será el grupo control, pues es el que se toma como referencia para evaluar las diferencias con respecto al grupo B, denominado grupo experimental, en los niveles de la variable x . Si el tratamiento es efectivo, se espera que los niveles de la variable x en el grupo B, sean mayores que en el grupo A. En otras palabras, será posible apreciar una diferencia significativa en la variable x , entre el grupo control y el grupo experimental.

La hipótesis nula deriva de esta situación, pues supone que no existen diferencias entre los grupos. Es decir, el efecto del tratamiento no produce diferencias entre ellos en la variable dependiente. La hipótesis de trabajo o experimental expresará que el tratamiento es el factor que diferencia a los grupos, y es factible ser medido por el efecto que produce sobre los niveles de la variable x . Ambas hipótesis quedarán expresadas estadísticamente de la siguiente manera:

$$H_0: \bar{x}_A = \bar{x}_B$$

$$H_1: \bar{x}_B > \bar{x}_A$$

La hipótesis nula expresa que no existen diferencias entre los grupos, por tanto, las puntuaciones medias en la variable dependiente serán iguales o muy cercanas entre sí. Si el grupo sometido a tratamiento es el B, la hipótesis de trabajo expresará que las puntuaciones medias de ese grupo serán mayores que las del grupo control (no tratado), en este caso el grupo A. Nótese que el uso de la prueba KS, necesita que la hipótesis de trabajo se exprese de manera unidireccional.

Distribución de la frecuencia acumulada

Para aplicar la prueba de KS se debe determinar la distribución de la frecuencia acumulada para cada muestra de observaciones, utilizando los mismos intervalos para ambas distribuciones. Entonces, en cada intervalo se resta el valor anterior al valor siguiente. Así, la prueba se centra en las mayores de las desviaciones observadas.

Se define a $Sm(X)$ como la distribución acumulativa observada para una muestra de tamaño m , esto es $Sm(X) = K/m$, donde K es el número de datos iguales o menores que X .

Se define a $Sn(X)$ como la distribución acumulativa observada en la otra muestra de tamaño n , esto es $Sn(X) = K/n$ donde K es el número de datos iguales o menores que X .

La distribución del estadístico Kolmogorov – Smirnov para pruebas unidireccionales es el siguiente:

$D_{m,n} = \max [S_m(X) - S_n(X)]$ (diferencia máxima en la dirección predicha)

para pruebas bidireccionales será

$D_{m,n} = \max |S_m(X) - S_n(X)|$ (diferencia máxima absoluta)

En cada caso, la distribución $D_{m,n}$ es conocida y se cuentan con tabulaciones de las probabilidades asociadas con la ocurrencia de valores tan grandes de una $D_{m,n}$ observada según la H_0 , de que las dos muestras provienen de una misma población. Las tabulaciones existen para las pruebas unidireccional o bidireccional. Como mencionamos anteriormente, la prueba estadística resulta más robusta cuando se sabe la dirección que tomarán las diferencias y por tanto se pueden establecer hipótesis unidireccionales.

El número de intervalos

Al utilizar la prueba de KS en datos cuyo tamaño y número de intervalos son arbitrarios, resulta conveniente emplear tantos intervalos como sea posible. Cuando se usan pocos intervalos, la desviación vertical máxima $D_{m,n}$ de las dos distribuciones acumuladas, puede oscurecerse al presentar los datos con pocos intervalos.

Conforme se incrementa el tamaño de las muestras, la distribución se aproxima a la de χ^2 con $gl=2$. Así, es posible determinar la significación de una $D_{m,n}$ observada, con el cálculo descripto en la siguiente ecuación.

$$\chi^2 = \frac{4D^2 m, n * (m * n)}{m + n}$$

La ecuación expresa la distribución de referencia de la siguiente manera: cuatro por el cuadrado de la diferencia máxima observada, por producto de m y n . Todo ello dividido por la suma de los respectivos tamaños muestrales m y n .

La prueba de Kolmogorov – Smirnov, cuando la hipótesis nula es unidireccional y la muestra es pequeña, resulta más poderosa que la prueba chi cuadrada o la prueba de la mediana. Asimismo, cuando las muestras son pequeñas, esta prueba es tanto o más eficaz que la prueba de Wilcoxon – Mann – Whitney, aunque es preferible ésta última cuando las muestras son grandes.

Ejemplo: En una investigación se utilizaron dos protocolos de evaluación de estudiantes. En el protocolo denominado c1, los alumnos respondían a una serie de preguntas que podían resolver

trabajando de manera colaborativa (Grupo A), o de forma individual con la ayuda de conexión a Internet (Grupo B). En este punto los investigadores no avanzaron la dirección de la diferencia entre los grupos, por lo cual dejaron establecida una hipótesis bidireccional. La variable independiente fue la modalidad de trabajo en la resolución del protocolo de evaluación; la variable dependiente fue el puntaje obtenido en dicha evaluación. Dado que la evaluación se expresa en una escala ordinal, se decidió realizar la comparación de grupos apelando a la prueba KS. Las hipótesis planteadas y los resultados de esta primera etapa de la investigación se muestran a continuación.

$$H_0: \bar{x}_A = \bar{x}_B$$

$$H_1: \bar{x}_A \neq \bar{x}_B$$

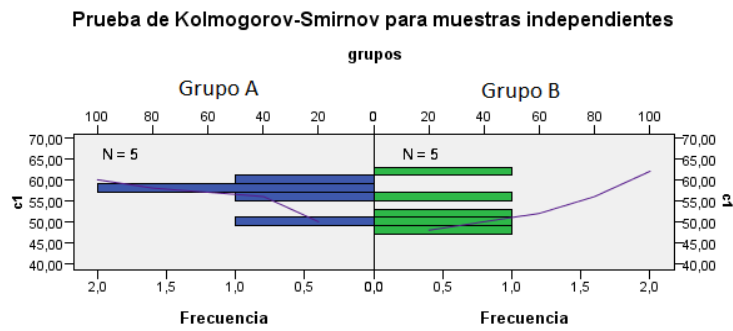
Resultado del experimento

	Hipótesis nula	Prueba	Sig.	Decisión
1	La distribución de c1 es la misma entre las categorías de grupos.	Prueba de Kolmogorov-Smirnov para muestras independientes	,819	Conserve la hipótesis nula.

Se muestran significaciones asintóticas. El nivel de significación es ,05.

Tal como se observa en el resumen de la prueba, los resultados muestran que no existen diferencias significativas entre los grupos, por lo cual resulta conveniente conservar la hipótesis nula. La prueba se realizó utilizando el software estadístico SPSS®, el cual provee más información sobre el contraste de hipótesis. La gráfica que se muestra a continuación detalla la distribución de frecuencias de los casos que contiene el grupo A y B, en relación a las puntuaciones obtenidas en la evaluación realizada con el protocolo c1. Además, se complementa con la tabla donde se detallan las máximas diferencias observadas ($D_{m,n}$), y el nivel de significación de la prueba. Resulta importante destacar que, por defecto, el software toma como valor de referencia para el contraste de hipótesis un valor de $\alpha=0,05$ para el error tipo I.

Tabla complementaria para el estadístico KS



N total	10
Máximas diferencias extremas	
Absoluta	,400
Positivo	,200
Negativo	-,400
Estadístico de prueba	,632
Significación asintótica (prueba bilateral)	,819

Ejemplo continuación: En la misma investigación analizó si existían diferencias en el otro protocolo utilizado para la evaluación de estudiantes, en este caso denominado c2. Las restantes condiciones del experimento no variaron, es decir, los alumnos respondían a una serie de preguntas que podían resolver trabajando de manera colaborativa (Grupo A), o de forma individual con la ayuda de conexión a Internet (Grupo B). En este punto los investigadores avanzaron la dirección de la diferencia entre los grupos, por lo cual dejaron establecida una hipótesis unidireccional. La variable independiente fue la modalidad de trabajo en la resolución del protocolo de evaluación; la variable dependiente fue el puntaje obtenido en dicha evaluación. Dado que la evaluación se expresa en una escala ordinal, se decidió realizar la comparación de grupos apelando a la prueba KS. Las hipótesis planteadas y los resultados de esta primera etapa de la investigación se muestran a continuación.

$$H_0: \bar{x}_A = \bar{x}_B$$

$$H_1: \bar{x}_B > \bar{x}_A$$

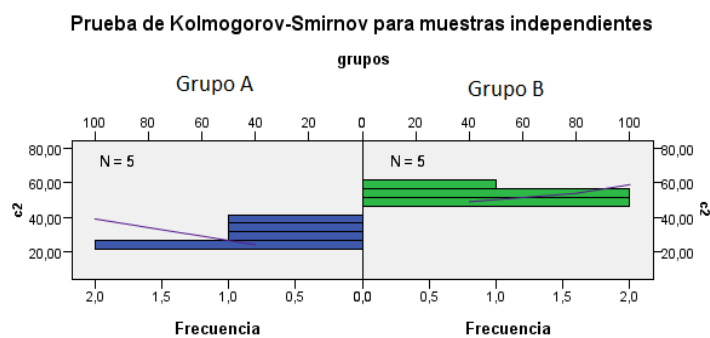
Resultado del experimento

	Hipótesis nula	Prueba	Sig.	Decisión
1	La distribución de c2 es la misma entre las categorías de grupos.	Prueba de Kolmogorov-Smirnov para muestras independientes	,013	Rechace la hipótesis nula.

Se muestran significaciones asintóticas. El nivel de significación es ,05.

Tal como lo muestra el resumen de la prueba, los resultados muestran que existen diferencias significativas entre los grupos, por lo cual resulta conveniente rechazar la hipótesis nula. Nuevamente, la prueba se realizó utilizando el software estadístico SPSS®, el cual provee más información sobre el contraste de hipótesis. La gráfica que se muestra a continuación detalla la distribución de frecuencias de los casos que contiene el grupo A y B, en relación a las puntuaciones obtenidas en la evaluación realizada con el protocolo c2. Además, se complementa con la tabla donde se detallan las máximas diferencias observadas ($D_{m,n}$), y el nivel de significación de la prueba. Resulta importante destacar que, por defecto, el software toma como valor de referencia para el contraste de hipótesis un valor de $\alpha=0,05$ para el error tipo I.

Tabla complementaria para el estadístico KS



N total	10
Máximas diferencias extremas	
Absoluta	1,000
Positivo	1,000
Negativo	,000
Estadístico de prueba	1,581
Significación asintótica (prueba bilateral)	,013

Aunque en este ejemplo no se ha profundizado sobre cuestiones que hace a las diferencias entre los protocolos c1 y c2 de evaluación, resulta claro que al utilizar el primer protocolo no se registran diferencias entre quienes respondían al mismo trabajando de manera colaborativa (Grupo A), y quienes lo hacían de forma individual con la ayuda de conexión a Internet (Grupo B). Sin embargo, si la evaluación se basa en el protocolo c2, los estudiantes que lo responden de manera individual con la apoyatura de conexión a Internet, obtienen un rendimiento superior.

Prueba de rachas de Wald-Wolfowitz

Esta prueba permite contrastar la hipótesis de que dos muestras independientes proceden de poblaciones con distribuciones continuas idénticas. Esta es la base de la hipótesis nula, mientras que la hipótesis alternativa supone que las poblaciones difieren en algún parámetro específico, por ejemplo, en la tendencia central (también pueden considerarse diferencia en otras características).

Para realizar el contraste, primero se combinan las observaciones de ambas muestras, pero ordenándolas de menor a mayor. Luego se cuenta el número de rachas o valores sucesivos de una misma muestra en la secuencia ordenada. Por ejemplo, supongamos que tenemos dos conjuntos de valores:

$$A = \{1, 3, 5\}$$

$$B = \{2, 4, 6\}$$

Los valores ordenados de ambas muestras quedan ordenados como sigue:

$$AB = \{1, 2, 3, 4, 5, 6\}$$

En este caso, hemos diferenciado con colores los valores provenientes de los conjuntos A y B, de modo que se pueda visualizar la sucesión de rachas en el conjunto ordenado AB. Por tratarse de los valores pares e impares de la secuencia de los primeros seis números naturales, resulta que el total de rachas es igual a 6. El estadístico de prueba es el número de rachas, que se denomina R. Dependiendo de los tamaños de las muestras, el valor de R estará comprendido entre 2 y $n_1 + n_2$. En otras palabras, si la hipótesis nula es correcta, las observaciones de ambas muestras aparecerán mezcladas, alcanzando el extremo $n_1 + n_2$, tal como se muestra en nuestro ejemplo con los grupos A y B. Por lo tanto, si las poblaciones de origen son diferentes, las observaciones de cada muestra tenderán a aparecer juntas y el número de rachas será pequeño, a tal punto que, si los conjuntos de observaciones no se superponen solo habrá 2 rachas, una para cada grupo.

Cuando los tamaños muestrales son $n > 10$ la distribución del estadístico R se aproxima a la distribución normal con parámetros:

$$\mu_R = \frac{2n_1n_2}{n_1 + n_2} + 1$$

$$\sigma_R^2 = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}$$

Ejemplo: En una investigación se solicitó a una muestra aleatoria de estudiantes de una carrera universitaria, que valoraran en una escala Lickert de cinco puntos su grado de acuerdo con el requisito del examen de ingreso. La escala se confeccionó de manera tal que el valor 1 sea completo desacuerdo y el valor 5 sea completo acuerdo. Los investigadores procuraron que la muestra obtenida estuviera compuesta de estudiantes del Interior provincial (Grupo 1), y de la Capital provincial (Grupo 2). El total de la muestra $n=20$ (Capital, $n=10$; Interior, $n=10$). Los resultados obtenidos mostraron que el grado de acuerdo con el examen de ingreso fue de media 2,6, con una desviación estándar de 1,14. Sin embargo, los investigadores sospecharon que las valoraciones dadas por los estudiantes del Interior podrían ser diferentes de las que realizaron los estudiantes de la Capital. Por lo cual decidieron evaluar si las valoraciones se diferenciaban tomando como variable independiente la procedencia del estudiante que realizó la encuesta. Las hipótesis derivadas de este planteo se muestran a continuación:

$$H_0 = \bar{x}_{g1} = \bar{x}_{g2}$$

$$H_1 = \bar{x}_{g1} \neq \bar{x}_{g2}$$

El criterio de rechazo de la hipótesis nula se realiza con bajo el contraste de la distribución normal, para un nivel de significación $\alpha=0.05$, donde:

$$Z \leq \pm z_\alpha$$

Este contraste estadístico se realizó utilizando el software estadístico Statistica®. La siguiente tabla ofrece el resultado de la prueba estadística:

Prueba de Wald-Wolfowitz. Variable independiente Procedencia del estudiante.

Wald-Wolfowitz Runs Test. Variable Procedencia								
Significación alfa $p < 0,05$								
n1	n2	Media G1	Media G2	Z	Nivel p	Z ajustado	Nº rachas	Nº empates
10	10	3,0	2,2	-0,46	0,64	0,229	10	7

La tabla muestra que existe una diferencia entre los promedios de ambos grupos, en donde la media de los estudiantes del Interior es ligeramente más alta que la de los estudiantes de Capital. Sin embargo, el contraste indica que el valor Z observado es mayor que el Z crítico. Por tanto, el nivel de significación estadístico resulta mayor al valor de $\alpha=0.05$. El número de rachas observadas es de 10, que se halla en un punto intermedio entre la cantidad extrema mayor y menor de posibles rachas a observar.

Nota: como se observa en la última columna de la tabla, el número de empates es alto (N° empates=7), lo cual requiere tomar con precaución los resultados observados. En casos como estos, resulta conveniente utilizar la prueba de Kolmogorov-Smirnov.

Pruebas de reacciones Extremas de Moses

Muchas veces realizamos un contraste entre medias para determinar diferencias entre grupos, pero solo se inspecciona el grado de variabilidad de las distribuciones mediante la varianza y la desviación estándar. Suele ocurrir que las diferencias significativas observadas no se deben a un efecto conjunto de todos los valores de un grupo, sino a una variación extrema de un conjunto muy acotado de valores que traccionan el promedio hacia un extremo. En tal caso, es necesario analizar si la diferencia encontrada se debe al grado de dispersión de las distribuciones. La prueba de reacciones extremas de Moses, resulta de mucho valor si sospechamos que esta es la situación cuando encontramos diferencias entre medias en distribuciones sesgadas.

Ejemplo: dos grupos toman parte de una evaluación de ciencias naturales. Los resultados se vuelcan en una tabla con los estadísticos descriptivos pertinentes y se observa una diferencia en el rendimiento promedio del grupo 1= 9,3, que es mayor al grupo 2= 6,15. Junto con esta información se ofrece la tabla de distribución de notas en ambos grupos y se observa que las notas máximas, están sobrerrepresentadas en el grupo 1, tal como se muestra a continuación:

Distribución de notas en el grupo 1 y 2: tabla de frecuencias

Recuento notas	grupo		Total
	1,00	2,00	
	3,00	0	
5,00	0	3	3
6,00	0	1	1
7,00	2	4	6
8,00	1	2	3
9,00	1	1	2
10,00	9	0	9
Total	13	13	26

Lo que se observa en la tabla es que la diferencia de promedios entre grupos pudiera deberse a una reacción extrema en la cantidad de calificaciones máximas dadas al grupo 1 (nueve individuos obtuvieron esa calificación contra ninguno del grupo 2). Para verificar si esta diferencia es debida a un sesgo en la distribución se decidió utilizar la prueba de reacciones extremas de Moses, antes de analizar las diferencias encontradas. Para este contraste se utilizó el software estadístico SPSS®.

El resultado del estadístico se muestra a continuación:

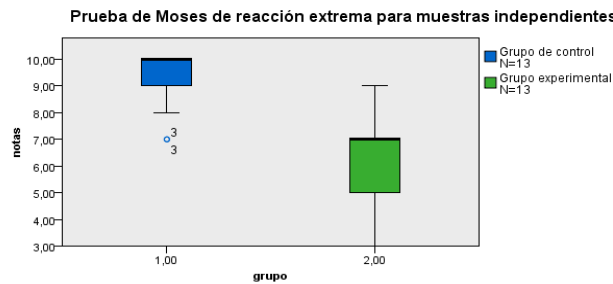
Resumen de contrastes de hipótesis

	Hipótesis nula	Prueba	Sig.	Decisión
1	El rango de notas es el mismo entre las categorías de grupo.	Prueba de Moses de reacción extrema para muestras independientes	,008 ¹	Rechaza la hipótesis nula.

Se muestran significaciones asintóticas. El nivel de significación es .

¹Se muestra la significación exacta para esta prueba.

El resumen del contraste de hipótesis, sugiere rechazar la hipótesis nula que propone que las dos distribuciones de nota tienen el mismo rango de notas. Por lo tanto, las diferencias significativas que pudieran encontrarse en un contraste de medias, está más relacionado a los valores extremos presentes en el grupo 1. La gráfica que se muestra a continuación desglosa la diferencia entre grupos, donde es factible observar la restricción de valores del grupo 1, ampliamente concentrado en la nota más alta.



N total¹	26
Grupo de control observado	
Estadístico de prueba ¹	14,000
Significación exacta (prueba unilateral) ¹	,008
Grupo de control recortado	
Estadístico de prueba ¹	14,000
Significación exacta (prueba unilateral) ¹	,008
Valores atípicos recortados de cada extremo¹	1,000

¹El estadístico de prueba es la amplitud.

En conclusión, la utilización de este estadístico junto con uno que aporte información sobre la diferencia de medias, especialmente si el resultado es significativo, ayuda a determinar mejor el fenómeno y las causas de las diferencias encontradas. En este ejemplo, es sencillo darse cuenta del sesgo en la distribución de notas, dado que se dispone de la tabla original. Pero este no siempre es el caso, ya que muchas veces, solo se reportan las medidas de tendencia central y variabilidad, y no resulta sencillo inferir a partir de ellas si una distribución de valores está sesgada.

Prueba de Irwin-Fisher para diferencias en proporciones

Las pruebas que hemos descripto hasta el momento se basan en la comparación de dos muestras, considerando principalmente su tendencia central. La prueba de Irwin-Fisher, permite comparar si dos muestras aleatorias provienen de dos poblaciones independientes, o de la misma población, utilizando la distribución hipergeométrica. La hipótesis nula en este caso se basa en la igualdad de proporciones de éxitos en ambas poblaciones, esto es p_1 y p_2 respectivamente. En tal caso, la hipótesis nula queda planteada de la siguiente manera:

$$H_0: p_1 = p_2 = p_0$$

donde p_0 es igual a valor propuesto común para ambas distribuciones.

La hipótesis alternativa predice que:

$$H_1: p_1 \neq p_2$$

El razonamiento para interpretar los resultados de esta prueba es el mismo que el ya visto en los anteriores ejemplos. En primer lugar, se calcula la mediana combinada de las dos muestras y luego se contabiliza la proporción de casos que se ubican por encima o por debajo de la misma. La hipótesis nula predice que la proporción de casos por encima y por debajo de la mediana, es la misma para ambos grupos. Estadísticamente la diferencia se expresa cómo:

$$\hat{\Delta} = \hat{p}_1 - \hat{p}_2$$

La fórmula expresa el diferencial de casos que caen por encima de la mediana en cada uno de los grupos, por lo tanto, y bajo la hipótesis nula, se espera que dicha diferencia sea nula o lo más pequeña posible. Cuando la cantidad de casos excede los 50, se utiliza la aproximación a la distribución normal, en vez de la hipergeométrica.

Ejemplo: En una escuela se pone a prueba la eficacia de dos métodos para el aprendizaje de la programación. La diferencia entre ellos radica en el tipo de software utilizado por los estudiantes, llamaremos a cada método programación por bloques o método A y programación estándar o método B. Los alumnos fueron asignados a cada método de manera aleatoria. Luego, todos los estudiantes fueron evaluados en función del tiempo en minutos empleado para resolver un problema.

Para este problema se utilizó el software estadístico InfoStat®. El resultado de la comparación se muestra en la siguiente tabla:

Resultado de la prueba de Irwin-Fisher

Se consideran éxitos a valores: mayores o iguales que la mediana	
Método A	Bloque
Método B	Estandar
n(1)	8
n(2)	11
p1	1,0
p2	0,27
p1-p2	0,73
p(bilat.)	0,0034

Según los resultados obtenidos se aprecia que existen diferencias significativas entre los grupos en el éxito en la resolución de la tarea. Todos los casos del grupo que aprendió programación por el método de bloques, se encuentra por encima de la mediana combinada, mientras que solo el 27% del grupo que aprendió con el método de programación estándar se halla por encima de la mediana. La diferencia de proporciones resulta estadísticamente significativa.

Análisis de Varianza unifactorial por rangos de Kruskal - Wallis

Hasta aquí hemos tratado casos en que analizamos la comparación entre dos grupos, pero en ocasiones se nos presenta la situación en el que participan más de un grupo. La solución cuando tratamos con variables en escala métrica es el Análisis de Varianza. Los modelos no paramétricos atienden a esta posibilidad cuando trabajamos con escalas ordinales, restringiendo las comparaciones a una variable dependiente. Es decir, el modelo de análisis de varianza unifactorial, compara diferencias entre más de dos grupos considerando una sola variable independiente. Por lo tanto, el análisis de varianza unifactorial de Kruska–Wallis es una prueba que permite decidir si k muestras independientes provienen de diferentes poblaciones. Por efecto aleatorio, sabemos que los valores de cada una de las muestras evaluadas invariablemente serán diferentes, pero la pregunta es si tales diferencias son las que efectivamente encontraríamos en la población o si estas representan la clase de variación que es esperable encontrar por azar.

La prueba de Kruskal-Wallis, somete a prueba la Hipótesis Nula (H_0) que las k muestras provienen de la misma población o de poblaciones idénticas y por ende tendrán la misma mediana. Siendo θ_j la mediana para el j -ésimo grupo, la H_0 estipulará que:

$$H_0: \theta_1 = \theta_2 = \theta_3 \dots = \theta_k$$

La H_1 estipulará que al menos una de las medianas de los k grupos es diferente, simbólicamente:

$$H_1: \theta_i \neq \theta_j$$

Donde $\theta_1, \theta_2 \dots \theta_k$ representan las distintas medianas de los k grupos. La H_0 será rechazada si es posible demostrar que al menos una de las medianas es diferente.

Como se dijo, la prueba requiere que la medición de la variable dependiente esté al menos en escala ordinal. Los datos se arreglan en una tabla de doble entrada donde cada columna representa a cada grupo o muestra sucesiva. En el cálculo de la prueba Kruskal-Wallis, cada una de las n observaciones se reemplazan por sus rangos. De este modo, todas las puntuaciones de las k muestras se combinan en una sola serie ordenada por rangos. La puntuación más pequeña se reemplaza por el rango 1, la puntuación que le sigue en tamaño ascendente se reemplaza por el rango 2 y la puntuación mayor se reemplaza por el rango n , donde n es el número total de observaciones independientes de las k muestras. Una vez obtenidos los rangos, se debe encontrar la suma de rangos en cada una de las k muestras. A partir de esta sumatoria de rangos, se puede calcular el rango promedio para cada muestra. Ahora bien, si todas las muestras provienen de la

misma población, el promedio de rangos será similar o estará muy próximo entre las k muestras. Si las muestras (o al menos una de ellas), no proviene de la misma población, habrá diferencia entre esos rangos. La prueba Kruskal-Wallis, evaluará la diferencia entre los rangos promedio para averiguar si son lo suficientemente dispares, de tal manera que no sea probable que las muestras hayan sido extraídas de la misma población. Una forma de calcular el estadístico es la siguiente:

$$kW = \frac{12}{n(n+1)} \sum_{j=1}^k n_j (r_j - \bar{r})^2$$

Donde

k : número de muestras o grupos

n_j : número de casos en la j -ésima muestra

n : número de casos en la muestra combinada (sumatoria de n)

r_j : sumatoria de rangos en la j -ésima muestra o grupo

$\bar{r} = (n + 1) / 2 =$ promedio de los rangos en la muestra combinada

Los términos de la sumatoria se realizan a través de las k muestras.

Cuando hay más de tres grupos, esto es $k > 3$, y cuando el número de observaciones por grupo es mayor que 5, la distribución del estadístico kW se aproxima a la distribución Chi cuadrado, con $gl = k - 1$. Esta aproximación puede mejorarse si se aumenta el número de grupos o se aumenta la cantidad de observaciones en cada grupo. En este caso, si el valor de kW es igual o mayor que el valor tabulado de Chi cuadrado, en el nivel de significación previamente determinado y para los $gl = k - 1$, es posible rechazar la hipótesis nula. Cuando $k=3$ y el número de casos en cada una de las muestras es pequeño, las probabilidades asociada a los valores de kW , se obtienen de una tabla para esta prueba. Si el valor encontrado es igual o mayor que el tabulado, para ese nivel de significación, se rechaza la hipótesis nula. Finalmente, cuando existen observaciones empatadas, éstas deben resolverse de la misma manera en que se explicaron en el apartado de la prueba de Kruskal – Wallis. Debe notarse que, a mayor número de empates, menor la varianza entre las muestras, lo cual hace inviable la aplicación de la misma en tanto se tratan de muestras con valores homogéneos entre sí. Invariablemente esto no permitirá decidir sobre la diferencia entre grupos.

Ejemplo: En un experimento realizado en una escuela, se aplicaron tres métodos diferentes para el uso del tiempo de estudio para la asignatura geografía. En cada uno de ellos la parcelación de los tiempos fue diferentes: grupo 1: tres bloques de 20 minutos; grupo 2: dos bloques de 30 minutos, y grupo 3: un bloque de 60 minutos. Se aplicó el mismo protocolo de evaluación, denominado c1, para todos los alumnos. En este punto los investigadores no avanzaron la dirección de la diferencia entre los grupos, por lo cual dejaron establecida una hipótesis

bidireccional. La variable independiente fue la modalidad de trabajo con las parcelas de tiempo descritas, la variable dependiente fue el puntaje obtenido en la evaluación. Dado que la evaluación se expresa en una escala ordinal, se decidió realizar la comparación de grupos apelando Análisis de Varianza unifactorial por rangos de Kruskal - Wallis. Las hipótesis planteadas y los resultados de esta primera etapa de la investigación se muestran a continuación.

$$H_0: \theta_1 = \theta_2 = \theta_3$$

La H_1 estipulará que al menos una de las medianas de los k grupos es diferente, simbólicamente:

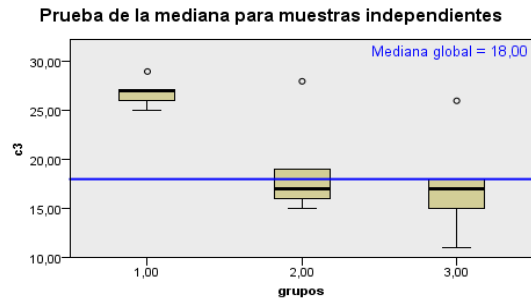
$$H_1: \theta_i \neq \theta_j$$

Los resultados de esta investigación se muestran a continuación. El análisis de datos se realizó con el software estadístico SPSS®. En el primer cuadro donde se resume el contraste de hipótesis se puede observar que se rechaza la hipótesis nula a un valor de significación $p=0.05$. De esto se deduce que al menos una de las muestras evidencia valores diferentes del resto, o los tres grupos son diferentes y por tanto es plausible suponer que no provienen de la misma población. Para determinar cuál o cuáles grupos se apartan significativamente de la mediana combinada, es necesario realizar un contraste entre grupos. Esto se muestra en los cuadros siguientes donde se grafica la prueba de la mediana para muestras independientes.

Resumen de contrastes de hipótesis				
	Hipótesis nula	Prueba	Sig.	Decisión
1	Las medianas de c3 son las mismas entre las categorías de grupos.	Prueba de la mediana para muestras independientes	,013	Rechace la hipótesis nula.

Se muestran significaciones asintóticas. El nivel de significación es ,05.

Como muestra la prueba de la mediana, son los individuos del grupo 1 los que obtienen mayores puntajes en la evaluación. El grupo 2 y 3 muestran diferencias, pero sus valores de mediana están próximos entre sí, por lo cual cabe concluir que la diferencia encontrada se sitúa entre el grupo 1 y los otros grupos.



N total	17
Mediana	18,000
Estadístico de prueba	8,743
Grados de libertad	2
Significación asintótica (prueba bilateral)	,013

Según lo establecido en el planteo del problema, la mejor estrategia de estudio resulta de parcelar una hora de estudio en tres segmentos de 20 minutos con descansos entre esos momentos. El peor rendimiento se observa cuando el tiempo de estudio de una hora no se halla parcelado.

Muestras Relacionadas

Prueba del cambio de McNemar

En la sección anterior hemos repasado la comparación de dos o más muestras cuando estas se suponen que son independientes. En todos los casos, la hipótesis nula establece que existen diferencia entre grupos si se demuestra que las muestras no han sido extraídas de la misma población. En otras situaciones, como las que presentaremos en esta sección, el supuesto implica que existe una sola muestra que se evalúa respecto de una variable independiente, en dos momentos diferentes. Las pruebas estadísticas, de una sola muestra, se denominan de pares replicados, es decir se trata de evaluar un tratamiento o más bien, el efecto de un tratamiento sobre los sujetos que conforman la muestra. En tal caso, cada sujeto pasa a ser su propio control, lo que supone que el tratamiento implica un antes y un después.

Cuando la variable se representa de manera dicotómica y ha sido medida en escala nominal u ordinal, resulta útil la prueba del cambio de McNemar. Para probar la significación del cambio, se utiliza una tabla de contingencia de 2x2, para representar el primero y el segundo conjunto de respuestas de los mismos individuos, en dos momentos distintos, tal como lo representa la tabla que se muestra a continuación.

	Después	
Antes	-	+
+	A	B
-	C	D

En cada caso los valores de las celdas A, B, C y D, son frecuencias observadas. En este arreglo de tabla, los casos que muestran cambios antes \rightarrow después, se agrupan en la celda A cuando la dirección del cambio es de + a -, en la celda D se agrupan los casos que muestran cambios antes \rightarrow después en la dirección de - a +. Las celdas B y C denotan a los individuos que permanecieron sin cambio. Por lo antes dicho, A + D representa el total de casos que cambiaron su respuesta.

La hipótesis nula entonces estipula que el número de cambios en cada dirección es el mismo, por lo cual el total de cambios en una u otra dirección, se representa con una frecuencia esperada igual a: $A + D/2$. Así, se puede calcular un estadístico χ^2 con $g/1= 1$, para las celdas donde se espera encontrar el cambio. El estadístico se representa por la siguiente ecuación:

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

La aproximación por medio de la distribución χ^2 llega a ser más precisa si se hace una corrección por continuidad. Esta corrección es necesaria para eliminar la imprecisión que se produce cuando se utiliza una distribución teórica continua de χ^2 para aproximarse a una distribución observada discreta. Cuando las frecuencias esperadas son pequeñas, la aproximación resulta pobre. El propósito de la corrección por continuidad es eliminar esta fuente de imprecisión. En tal caso, la expresión de la ecuación se corrige de la siguiente manera para tablas con grados de libertad igual a 1.

$$\chi^2 = \frac{(|A - D| - 1)^2}{A + D}$$

Si la frecuencia esperada, esto es $(A+D)/2$ es muy pequeña (menor a 5), se debe utilizar la prueba binomial.

Ejemplo: En un estudio se pidió a los docentes que valoraran sus prácticas de evaluación en relación a los contenidos de la asignatura. En un primer momento, las y los participantes manifestaron su grado de acuerdo con las prácticas actuales, luego realizaron un taller donde tuvieron oportunidad de revisar sus propias prácticas. Al final de la actividad se les pidió manifestar su acuerdo con la posibilidad de cambiar dichas prácticas. El diseño de esta investigación tuvo dos momentos, un Antes del taller y Después del taller. La decisión de cambiar las prácticas se valoró con una escala dicotómica donde: 1= acuerdo, y 0= desacuerdo. El análisis de datos se realizó utilizando la prueba de McNemar.

En la siguiente tabla se muestra la distribución de frecuencias relativas observadas en los dos momentos del estudio. El análisis de datos se realizó con el software estadístico Statistica®.

Distribución de frecuencias en acuerdos y desacuerdos antes y después de realizar el taller de prácticas de evaluación

	Después	
Antes	1	0
0	22	14
1	12	24

Los resultados de la prueba de McNernman se muestran en la siguiente tabla:

	n=72
Chi- Cuadrado	0.038 ^a
Significación p>0.05	0.845 ^b

a Corrección por continuidad

b McNemar Test

Tal como se muestra en la tabla, la tasa de cambio observada en las prácticas evaluativas de los 72 participantes del estudio no resultó significativa según los resultados de la prueba. Lo cual demuestra que la cantidad de docentes que decidieron cambiar sus prácticas en un sentido, resultó similar a aquellos que decidieron cambiar las prácticas en sentido opuesto (celdas A y D). En igual proporción se repartieron aquellos docentes que no decidieron cambiar sus prácticas de evaluación.

Prueba de los signos

Esta prueba estadística adquiere su nombre porque está basada en la dirección de la diferencia entre dos mediciones, más que en su magnitud. La prueba se puede utilizar cuando es posible obtener un par de mediciones, en la que se pueden establecer diferencias expresadas con el cuantificador “*mayor que*”; en tal caso se supone que la variable tiene una distribución continua. El par de mediciones obtenido puede extraerse de una sola muestra medida dos veces, o de una muestra con pares replicados. En uno u otro caso, se tendrán dos mediciones que se pueden denotar como x e y .

En esta prueba, la hipótesis nula se puede expresar como: $p \{X > Y\} = p \{X < Y\} = 1/2$. Es decir, la proporción de veces en que x es mayor que y , es la misma que la proporción de veces en que ocurre que y es mayor que x . Por lo tanto, la distribución de probabilidades es igual a $1/2$. Lo anterior se cumple cuando la muestra de donde se extrajeron las mediciones de x e y tienen la misma mediana (en este sentido es similar a la prueba Binomial). En este caso, x e y son las dos puntuaciones obtenidas por el par replicado o por el mismo sujeto en dos momentos diferentes. Por lo que otra forma de expresar la hipótesis nula es: la mediana de las diferencias entre x e y es cero.

Cuando se aplica la prueba, se anota la dirección de las diferencias entre cada par x e y ; estas diferencias pueden ser positivas, negativas o nulas. Cuando la hipótesis nula es verdadera, se espera que al menos la mitad de las diferencias entre las x e y , sean positivas, y la otra mitad sean negativas. Cuando ocurren pocas diferencias en un sentido u otro, es posible rechazar la hipótesis nula.

Con muestras pequeñas, esto es $n < 35$, se utiliza la prueba Binomial con $p = q = 1/2$. Hay que tener en cuenta que los casos en los que la diferencia entre x e y es cero, quedan excluidos del análisis, por lo cual es frecuente que n se vea reducido. Cuando $n > 35$ se utiliza la aproximación normal a la Binomial y se emplean los valores Z de dicha distribución.

Empates: para la prueba de los signos ocurre un empate cuando no es posible discriminar entre los valores de un par de observaciones. Todos los casos empatados son excluidos del análisis y n se decrementa en $n - e$; donde e es el número de empates.

Ejemplo: En una escuela se analizó la relación entre docentes que conformaban pares pedagógicos. El estudio se propuso determinar en qué medida las decisiones didácticas eran concordantes o discordantes en cuanto a un tema determinado de la planificación curricular. Las

diferencias se estimaban mediante la aplicación de una escala en la que cada docente se puntuaba como “más importante decisor”. La diferencia surgida de tales puntuaciones se analizó mediante la aplicación de la prueba de los signos. La tabla que sigue ejemplifica el tipo de valoraciones de la pareja pedagógica y su codificación: En la pareja A, se verifica que el docente a cargo es el que tiene más responsabilidad en la toma de decisión, mientras que el par pedagógico estima que deberían ser iguales o ser él quien asuma mayor responsabilidad. Por lo tanto, la discrepancia se califica con el signo + (positivo). En la pareja B, se produce el efecto contrario, por lo tanto, tal discrepancia se califica con el signo – (negativo). En este caso, el uso de los signos son indicativos de discrepancias en las puntuaciones de responsabilidad percibida. Finalmente, la pareja C manifiesta acuerdo en quien debe asumir la responsabilidad en la decisión y por tanto se califica con 0 (cero).

Discrepancias y concordancias en la percepción de responsabilidad la toma de decisiones en cada pareja pedagógica, donde x representa el docente a cargo, e y el par pedagógico.

Pareja	Docente	Par Pedagógico	Dirección de la diferencia	Signo	
A	4	3	$x_d > y_{pp}$	+	Discrepancia: el docente considera que es él quien debe tener mayor peso en la decisión
B	3	4	$x_d < y_{pp}$	-	Discrepancia: el par pedagógico considera que es él quien debe tener mayor peso en la decisión
C	3	3	$x_d = y_{pp}$	0	Existe acuerdo en quien tiene mayor peso en la decisión

La tabla representa solo tres de las parejas analizadas de un total de 72 estudiadas.

El análisis de datos se realizó utilizando el software Statistica®. La salida de dicho análisis se muestra en la siguiente tabla:

Resultado de la prueba de los signos aproximación a la normal N(0;1)

Diferencias Negativas	20
Diferencias Positivas	33
Empates	19
Total	72 (53)
Z	-1.648
Significación (2-colas)	$p=0.099$

El resultado muestra que, si bien se observan discrepancias en las valoraciones, en conjunto la actividad de los docentes con sus pares pedagógicos no manifiesta diferencias significativas. La proporción en que discrepan en un sentido o en otro es relativamente proporcional, y se registra

un número importante de empates. Además, el nivel de significación alcanzado con la aproximación a la distribución normal estándar es mayor que el Z crítico $p=0,05$.

Prueba de rangos asignados de Wilcoxon

La prueba de los signos solo analiza la dirección de la diferencia en cada uno de los pares de medidas. La prueba de Wilcoxon, utiliza esta información, pero además permite estimar la magnitud relativa de esa diferencia, por lo cual la variable debe ser medida al menos en escala ordinal. Las diferencias entre las puntuaciones entre los miembros del par igualado, al cual denominaremos d , representa la diferencia en ese par entre los miembros x e y . Para la prueba de Wilcoxon se deben poner en columna todas las diferencias, ordenándolas por rango y numerándolas a partir de 1 comenzando por la más pequeña de esas diferencias. Luego se debe añadir el signo de la diferencia, para identificar el rango de la diferencia cuándo es positiva o negativa. Dado que esta se presenta como una prueba más robusta que la prueba de los signos, utilizaremos el ejemplo anterior para explayarnos sobre la misma. En tal caso, la tabla que presentamos anteriormente tomaría la siguiente forma:

Discrepancias y concordancias en la percepción de responsabilidad la toma de decisiones en cada pareja pedagógica, donde x representa el docente a cargo, e y el par pedagógico.

Pareja	Docente	Par Pedagógico	d	Rango d	
A	4	3	1	2	Discrepancia: el docente considera que es él quien debe tener mayor peso en la decisión
B	2	5	-3	-1	Discrepancia: el par pedagógico considera que es él quien debe tener mayor peso en la decisión
C	3	3	0	1	Existe acuerdo en quien tiene mayor peso en la decisión

La tabla representa solo tres de las parejas analizadas de un total de 72 estudiadas.

Como se aprecia, los rangos se asignan de las menores diferencias a las mayores, respetando el signo de tal diferencia. Así, a la pareja B se le asigna un rango con valor negativo en tanto la diferencia en la valoración tiene ese signo.

La hipótesis nula es que los pares x e y son equivalentes, y por tanto tienen la misma mediana y una misma distribución continua. Si la H_0 es verdadera se debería encontrar algunas diferencias a favor de x y otras tantas a favor de y . Es decir, si se suman los rangos con signos positivos y negativos, se esperaría que ambas sumas fueran iguales. Los casos en que la diferencia entre x e y sea igual a cero, se excluyen del análisis. Por tanto, n es la suma total de rangos, menos los que están empatados en cero. Cuando los pares están empatados en la misma magnitud, se utiliza el rango promedio que les correspondería. Los valores de significación de la prueba pueden

obtenerse con una tabla, pero cuando la muestra es grande, esto es $n > 15$, la suma de los rangos $S(+)$, se distribuye de manera normal $N(0; 1)$.

Ejemplo: ya habíamos verificado en el ejemplo anterior que la cantidad de discrepancias positivas y negativas en la muestra de docentes y pares pedagógicos era relativamente proporcional. Entonces, puesto que la prueba de rangos asignados toma en consideración este dato, los resultados obtenidos con la misma matriz, deberían confirmar los resultados anteriores. También se utilizó el software Statistica®, para este análisis. Los resultados se muestran en la siguiente tabla:

**Resultado de la prueba de rangos asignados de Wilcoxon;
aproximación a la normal $N(0;1)$**

	N	Media Rangos	Suma de Rangos
Rangos Negativos	20	28.45	569.00
Rangos Positivos	33	26.12	862.00
empates	19		
Total	72		
Significación (2-colas) $Z=-1.306$; $p= 0.192$			

Por tratarse del mismo ejemplo dado anteriormente no nos extenderemos en la interpretación de los resultados, más que para confirmar que en ambos casos se rechaza la hipótesis nula.

Prueba Q de Cochran

Las pruebas descritas en los apartados anteriores trataban el caso de comparar dos muestras, o más bien, la misma muestra en dos condiciones diferentes, como puede ser antes y después de un tratamiento. En otras condiciones los sujetos de una muestra pueden pasar por más de un tratamiento, o situaciones de antes, durante y después. En tal caso debe recurrirse a análisis estadísticos que puedan tener esa particularidad en cuenta; es decir, es el equivalente a la comparación de más de dos grupos.

La evaluación de k muestras supone una H_0 que estipula que esas muestras fueron extraídas de la misma población, o de poblaciones idénticas. Cuando se tiene el caso de más de dos muestras a comparar, es necesario desestimar las comparaciones por pares, ya que ello en primer lugar requiere tener en cuenta un número innecesariamente alto de contrastes, pero además aumenta notablemente la probabilidad de rechazar la H_0 en cualquiera de los pares contrastados. La prueba para más de dos muestras permite obtener un valor global de significación para la prueba de hipótesis, que justifica el procedimiento para encontrar diferencias entre cualquier par de las k muestras.

La prueba Q de Cochran es una ampliación de la prueba de McNemar para dos muestras relacionadas, y proporciona un método para evaluar si tres o más conjuntos de frecuencias o proporciones igualadas, difieren significativamente entre ellas. La igualación debe realizarse en las características relevantes de los sujetos, o en el hecho de que los mismos sujetos son evaluados repetidas veces. La prueba Q de Cochran es aplicable a datos obtenidos en escala nominal con más de dos categorías, o en escala ordinal o de intervalo con categorías dicotómicas.

Ejemplo: Una administración escolar decide cotejar cuál de las siguientes modalidades prefieren las y los docentes para recibir información administrativa de su centro educativo. Las novedades podrán comunicarse de tres maneras diferentes: a) personalmente, b) telefónicamente y c) por correo. Se seleccionan tres muestras igualadas de sujetos ($N=10$) con las características mencionadas y se evalúa el éxito en la comunicación de la información (exitosa=1 no-exitosa =0).

La hipótesis nula en este caso es que la probabilidad de que las y los docentes acuerden en la comunicación de la información, es la misma en las tres condiciones. La hipótesis alternativa es que la probabilidad de que acuerden en la comunicación de la información es mayor en alguna de las condiciones.

En este caso, la prueba Q de Cochran evalúa la hipótesis nula acerca de si la proporción (o frecuencia) de respuestas de una clase particular, es la misma en cada columna o, de existir diferencias, estas son debidas al azar. La ecuación propuesta por Cochran para el cálculo del estadístico Q es la siguiente:

$$Q = \frac{k(k-1) \left[k \sum_{j=1}^k (\sum_{i=1}^n G_j)^2 \right]}{k \sum_{i=1}^n L_i - \sum_{i=1}^n L_i^2}$$

Donde:

G_j = número total de éxitos en la j -ésima columna

\bar{G}_j = media de G_j

L_i = número total de éxitos en el i - esimo renglón

El estadístico Q está distribuido como χ^2 con $gl=k-1$. Esto proporciona un valor de significación asociado para ciertos grados de libertad. La implicación es que la proporción (o frecuencias) de éxitos difiere significativamente entre las muestras. Por lo tanto, la H_0 puede ser rechazada a un nivel de significación particular. Para la resolución del problema planteado se utilizó el software Statistica®.

Resultado del estadístico Q

Cochran's Q Test: N: 20			
Q = 8.727, gl = 2, p < 0.012			
	Suma	Porcentual 0's	Porcentual 1's
PERSONAL	13.0	35.0	65.0
TELEFONO	9.0	55.0	45.0
CORREO	5.0	75.0	25.0

El valor de χ^2 asociado permite rechazar la H_0 que estipula que modalidades prefieren las y los docentes para recibir información administrativa de su centro educativo, es equivalente. El encabezado de la tabla muestra que un valor de $Q=8.727$, tiene una probabilidad de ocurrencia asociada de menor a 0.0127, con $gl=2$. Luego, la suma de éxitos en cada una de las formas de recepción de la información se muestra en la columna Suma de la tabla. La proporción de éxitos y fracasos para las distintas modalidades de contacto se muestran en las columnas % de 0's y % de 1's respectivamente. En conclusión, de la interpretación de la tabla se deduce que las y los docentes prefieren el contacto personal por sobre el telefónico, y a este por sobre el correo.

Análisis de Varianza bifactorial por rangos de Friedman

Cuando los datos de k muestras igualadas están al menos en escala ordinal, se puede utilizar el análisis de la varianza de Friedman para evaluar la hipótesis nula de que las muestras fueron extraídas de la misma población. En tal situación, la cantidad de casos n deberá ser la misma en cada una de las muestras, o bien se puede analizar la misma muestra a través de distintas condiciones.

El análisis de varianza de Friedman evalúa la H_0 de que los k grupos igualados, o las medidas repetidas para un mismo grupo, provienen de la misma población o de poblaciones con la misma mediana. La H_1 supone que alguna de las medianas de los k grupos será diferente, y por tanto proviene de una población diferente. Bajo la H_0 la prueba supone que las variables tienen la misma distribución continua subyacente, por ello la prueba requiere que la variable dependiente esté medida al menos en una escala ordinal.

Para la prueba de Friedman, los datos deben arreglarse en una tabla de doble entrada conteniendo n renglones y k columnas. Los renglones representan a los sujetos y las columnas las diferentes condiciones o tratamientos. Los datos que emplea la prueba son los rangos de las puntuaciones obtenidas, en cuyo caso, las puntuaciones en cada renglón se ordenan separadamente por rangos. Esto es, estudiando los sujetos en k condiciones los rangos se ordenan de 1 a k . En la siguiente tabla se muestra un ejemplo de clasificación de un sujeto que obtuvo tres puntuaciones en una prueba de resolución de problemas bajo diferentes condiciones. En la Condición A obtuvo 18 puntos, en la Condición B 14 puntos y en la Condición C 15 puntos. Ordenados por rangos de menor a mayor se tiene 1 para la Condición B, 2 para la Condición C, y 3 para la Condición A. Como solo hay tres condiciones el total de los rangos es igual a k . Conviene señalar que los empates en los rangos, provocan que el estadístico F de Friedman se incremente, por lo cual se emplea una forma alternativa para el cálculo que disminuye esta posibilidad. Por la extensión de este trabajo no desarrollaremos este modo de cálculo, para ello se sugiere consultar el trabajo de Siegel y Castellan (1998).

Transformación de puntuaciones originales a rangos de un sujeto en tres condiciones diferentes

Sujeto 1	Condición A	Condición B	Condición C
Puntuación Original	18.0	14.0	15.0
Puntuación por Rangos	3	1	2

La prueba de Friedman determina la probabilidad de que diferentes columnas de rangos (muestras), provengan de la misma población o de poblaciones con la misma mediana. Es decir que las k variables (condiciones) tengan la misma mediana.

Bajo el supuesto que la H_0 es verdadera, las distribuciones de los rangos en cada columna estarían dispuestos por azar, y se esperaría casi la misma frecuencia de rangos por columna. Es decir, si los datos fueran aleatorios, la suma de los rangos en cada columna sería igual a: $n(k+1)/2$. Esto representa la suma esperada de rangos por columna cuando los datos siguen un patrón aleatorio. Si las puntuaciones de los sujetos fueran independientes de las condiciones, el conjunto de rangos en cada columna representaría una muestra aleatoria de una distribución discreta rectangular de los k rangos. Por el contrario, si las puntuaciones dependen de las condiciones, los rangos variarían en cada columna. Un planteo alternativo para la H_0 sería que el promedio de los rangos para las diferentes columnas sería aproximadamente el mismo.

La prueba de Friedman, permite determinar si los rangos totales R , para cada condición, difieren de los valores esperados por azar. Para el cálculo del estadístico de Friedman, se necesita conocer: n , número de sujetos o tamaño muestral; k , número de columnas, o condiciones; R , suma de rangos por columna y ΣR^2 , sumatoria de los cuadrados de los rangos de todas las columnas.

Las probabilidades asociadas a valores de F , cuando la H_0 es verdadera son conocidas, y en base a ellas se puede aceptar o rechazar esta hipótesis a un valor determinado de significación.

Cuando el número de condiciones es mayor que cinco ($k > 5$) y el tamaño de la muestra es grande, el valor de la F de Friedman se distribuye aproximadamente como χ^2 con $gl = k - 1$. Cuando el valor de F de Friedman resulta significativo, la suma de los valores de los rangos en cada columna es diferente y ello significa que las condiciones determinan la distribución de los valores.

Ejemplo: un grupo de psicólogos diseñó un programa de entrenamiento para la resolución de problemas prácticos. Dicho programa se basa en la aplicación de ciertas herramientas lógicas y probabilísticas a situaciones cotidianas. Se realizó una investigación para determinar el alcance de lo aprendido en el programa a distintas situaciones problemáticas. Para ello, se diseñaron diferentes condiciones o situaciones donde un grupo de participantes debían resolver una serie de problemas. Cada condición se definió de la siguiente manera, Condición A: denominada de transferencia directa dado que los problemas a resolver eran similares a los que se utilizaron como ejemplos en el programa de entrenamiento, Condición B: denominada de transferencia indirecta, en donde los problemas solo tenían una relación conceptual con los utilizados en el programa de

entrenamiento, y Condición C: denominada creativa, ya que los problemas planteados requerían la reestructuración de los conceptos aprendidos en el programa de entrenamiento. Se utilizó como variable dependiente una serie de 20 problemas de papel y lápiz, representativos de las tres condiciones mencionadas. En las siguientes tablas se muestran los puntajes obtenidos por los sujetos participantes ($n=20$).

Dado que este estudio requiere comparar una misma muestra en tres condiciones diferentes, resulta pertinente emplear el análisis de varianza bifactorial de Friedman. El análisis de los datos se realizó con el software estadístico Statistica®. Los resultados se muestran en la siguiente tabla:

Resultados del Anova bifactorial de Friedman

ANOVA Friedman y Coeficiente de Concordancia Kendall				
ANOVA $\chi^2 (n=20, gl=2) = 13.027 p < 0.0014$				
Coeficiente de Concordancia= 0.325 rango promedio $r = 0.29$				
	Rango promedio	Suma de Rangos	Media	DE
Condición A	2.50	50.0	15.6	2.6
Condición B	2.07	41.5	14.9	2.1
Condición C	1.42	28.5	11.6	3.9

En la tabla se muestran las tres condiciones (k renglones), el promedio del rango, la suma de rangos en cada una de las k condiciones, la media y la desviación estándar de los puntajes obtenidos. De haberse registrado una distribución aleatoria de los resultados, esto es, que las condiciones no afectan a los puntajes en las pruebas, la distribución de las sumas de rangos debería haber estado próxima a $n(k+1)/2$, es decir $20*(3+1)/2=40$. Sin embargo, se aprecia que las sumas difieren de tal valor. Es decir, para la condición A y C, la suma de rangos es significativamente diferente de lo esperado por azar. La confirmación de que existen diferencias significativas está dada por la comprobación de la distribución del estadístico χ^2 . En este caso, se observa una probabilidad muy baja de ocurrencia de este resultado ($p < 0.0014$), por lo cual corresponde rechazar la H_0 .

Cuando el valor F de Friedman es significativo, esto implica que al menos dos condiciones son diferentes. Pero, el estadístico no indica cuál de las condiciones es efectivamente distinta. Por lo tanto, las comparaciones posteriores se deben realizar entre todos los pares posibles. Entonces, sean R las diferencias entre los rangos para cada una de las condiciones, las comparaciones se realizan mediante la siguiente ecuación:

$$|Ru - Rv| > z_{\alpha/k(k-1)} \sqrt{\frac{nk(k+1)}{6}}$$

La expresión $|Ru - Rv|$ indica la diferencia entre dos rangos cualquiera, la cual puede ser reemplazada por el promedio de los rangos. Ahora, si la diferencia entre los rangos (o el promedio de los rangos) excede el valor crítico dado por la ecuación, se debe concluir que las dos condiciones evaluadas son diferentes. El valor $Z_{\alpha/k(k-1)}$ es el valor de la abscisa de la distribución normal donde se encuentra el $Z_{\alpha/k(k-1)}$ % de la distribución. Ese valor crítico determina la probabilidad de rechazar la H_0 a un nivel de significación dado. Por convención dicho valor es $\alpha=0,05$.

Tomando los datos del ejemplo planteado, donde $k=3$, el número de comparaciones a realizar está dado por $k(k-1) / 2$, esto es, $3(3-1) / 2=3$. Las comparaciones se estiman como se muestra a continuación (que deben leerse como diferencias en las puntuaciones en el rango a – b, etc.)

- 1) $Ra - Rb= 50 - 41= 9$
- 2) $Ra - Rc= 50 - 28= 22$
- 3) $Rb - Rc= 41 - 28= 13$

Se establece luego el valor crítico de $\alpha= 0,05$, y el número de comparaciones, que en este caso es 3. El valor Z para un ese número de comparaciones es igual a 2.394. Este valor se halla tabulado. Luego al reemplazar estos valores en la fórmula, se obtiene el valor crítico de comparación:

$$|Ru - Rv| > z_{\frac{\alpha}{k(k-1)}} \sqrt{\frac{nk(k+1)}{6}} = 2,394 * \sqrt{\frac{20 * (3 + 1)}{6}} = 15,14$$

El valor crítico de 15,14, es tomado como referencia para determinar cuál de las diferencias resulta significativa. Ahora bien, dado que las diferencias encontradas fueron:

- 1) $Ra - Rb= 50 - 41= 9$
- 2) $Ra - Rc= 50 - 28= 22$
- 3) $Rb - Rc= 41 - 28= 13$

Se desprende que solo la segunda comparación excede el valor de Z crítico, por lo cual es la que resulta significativa. En este caso es la diferencia entre la condición a y c.

Guía de lecturas complementarias

Como esbozamos al comienzo de esta obra, la finalidad de la misma es interesar a los estudiantes en las técnicas de análisis de datos no paramétricos. Los ejemplos desarrollados han sido simplificados y tratados con la menor exigencia de cálculos posibles. Sin embargo, para poder comprender mejor cada una de las técnicas de análisis ejemplificadas, es conveniente recurrir a textos que profundicen estos temas. A continuación, propongo algunos manuales que, a consideración personal, estimo que son los mejores para avanzar en un estudio más pormenorizado de los métodos de análisis descriptos:

Siegel, S. y Castellan, N. J. (1995). *Estadística no paramétrica, aplicada a las ciencias de la conducta*. 4a. edición. México. Editorial Trillas.

Velasquez Leonett V. A. (2005). *Métodos Estadísticos no-paramétricos*. Editorial EAE, Editorial Academia Espanola.

Loureiro Pérez, E. (2011). *Estadística No Paramétrica*. Editorial Ediciones Cooperativas.

Castillo, C. Merino, V. y Caycho, C. (2019) *Manual de estadística no paramétrica aplicada a los negocios*. Editorial(es): Universidad de Lima. Fondo Editorial.

Guía de software no comercial para análisis estadístico

Los ejemplos que se han desarrollado en esta obra han utilizado los programas estadísticos SPSS®, Statistica® e Infostat®. Los dos primeros son de uso comercial, el tercero también lo es, pero permite una descarga gratuita para estudiantes. Más adelante comentaremos sobre el mismo. La finalidad de este apartado es presentar alternativas libres que reemplacen a los de uso comercial. En todos los casos se tratan de software profesionales pero que no por ello resignan las funciones más importantes del análisis estadístico.

R Commander

Es una alternativa al software SPSS basada en el desarrollo del lenguaje de programación R. El enlace de descarga se muestra a continuación:

<https://www.rcommander.com/>

Esta página está en inglés, por lo cual solo puede ser aprovechada por aquellos que dominen el idioma. Es por ello que en el siguiente enlace les dejo un manual en español escrito por el Prof. Dr. Antonio José Sáez Castillo del dpto de Estadística e Investigación Operativa Universidad de Jaén, donde se explica cómo descargar, instalar y trabajar con este programa.

<https://cran.r-project.org/doc/contrib/Saez-Castillo-RRCmdrv21.pdf>

PSPP

Este software libre ha sido diseñado para reemplazar la mayoría de las funciones de estadística básica de su contraparte comercial SPSS. Su interfaz es muy similar y su uso es intuitivo. Es un excelente programa para comenzar a trabajar con análisis estadísticos simples; como en el caso de R Commander, es una interfaz gráfica basada en R. El programa puede descargarse desde el siguiente enlace:

<https://www.gnu.org/software/pspp/get.html>

Esta página también está en inglés, pero podrán ubicar fácilmente los enlaces de descarga para diferentes sistemas operativos. En el siguiente enlace les dejo un manual donde se explica paso a paso, cómo descargar el programa y sus funcionalidades básicas, la obra fue redactada por Javier Torrente: Introducción Guiada a PSPP: uso en la estadística descriptiva e inferencial.

[https://www.e-ucm.es/download/\[Torrente\]%20Manual%20PSPP.pdf](https://www.e-ucm.es/download/[Torrente]%20Manual%20PSPP.pdf)

Infostat versión estudiantil

Infostat es un software desarrollado en la Facultad de Agronomía de la Universidad Nacional de Córdoba (Argentina). Cuenta con una versión profesional de pago y una alternativa gratuita para estudiantes. Esta última se puede descargar e instalar como cualquier programa en versiones de Windows, aunque no está disponible para otros sistemas operativos como Linux. El siguiente enlace los conduce a la página donde se describe el software, su funcionalidad y contiene el enlace de descarga.

<https://www.infostat.com.ar/>

Para descargar la versión estudiantil, deberán navegar en el sitio web hasta la pestaña que se indica como versión estudiantil y seguir el instructivo. El sitio web está en español, de modo que es muy simple navegar por ella.