

## Métodos No-Paramétricos de Uso Común

### (*Non Parametric Methods of Common Usage*)

Badii, M.H., A. Guillen, L.A. <sup>1</sup>Araiza, E. <sup>2</sup>Cerna, J. <sup>2</sup>Valenzuela & J. <sup>2</sup>Landeros. UANL, San Nicolás, N.L., [mhbadiiz@gamil.com](mailto:mhbadiiz@gamil.com), <sup>1</sup>ITESM, Mty., <sup>2</sup>UAAAN, Saltillo Coah., México

**Abstract.** Non parametric or distribution free methods include the techniques of choice to be used in situations where assumptions on the type of the distribution are lacking. In this paper several non parametric models ( $X^2$  squared, Tests of Median, McNemar, Kolmogorov-Smirnov, Cochran, Cox, Wilcoxon, Mann-Whitney, Kruskal-Wallis, Friedman, Spearman, etc.) of common usage are discussed and an example for each model is provided.

**Keywords:** Hypothesis testing, non-parametric statistics, statistical models

**Resumen.** Los métodos no-paramétricos incluyen las técnicas de selección a utilizarse en condiciones en donde carecen supuestos sobre el de distribución de los datos. En este artículo se discuten varios modelos comunes no-paramétricos o de tipo de distribución libre ( $X^2$  cuadrada, Tabla de contingencia, Wald-Wolfwitz, pruebas de: mediana, McNemar, Kolmogorov-Smirnov, Cochran, Cox, Wilcoxon, Mann-Whitney, Kruskal-Wallis, Friedman, Spearman, etc.) y para cada modelo se presenta un ejemplo práctico.

**Palabras clave:** Estadística no-paramétrica, modelos estadísticos, pruebas de hipótesis

#### Introducción

El campo de la estadística de diseños de distribución libre o no paramétrica ha tenido un desarrollo tan grande, que ahora tenemos, para casi cualquier diseño de investigación, pruebas estadísticas alternativas válidas que podemos utilizar para producir resultados acerca de una hipótesis. Teniendo pruebas alternativas válidas, necesitamos algunas bases racionales para elegir entre ellas (Badii & Castillo, 2009a, 2009b; Badii et al., 2010; Badii et al., 2011<sup>a</sup>, 2011<sup>b</sup>). En este trabajo los procedimientos se centran en la estadística no paramétrica donde va presentar un análisis de las bases para elegir entre varias pruebas aplicables a un diseño de investigación determinado.

**Concepto, ventajas y limitaciones.** Las pruebas no paramétricas son aquellas en las que no existen supuestos sobre la distribución de los parámetros de la población. Se aplican con mayor frecuencia a los datos nominales y ordinales, si bien pueden emplearse también para analizar datos continuos transformados a una escala ordinal. La estadística no paramétrica presenta las siguientes *ventajas*.

1. Cuando es posible hacer solamente supuestos débiles acerca de la naturaleza de las distribuciones que fundamentan los datos.
2. Cuando solo será posible poco más que categorizar los datos por falta de una escala de medición adecuada.
3. Cuando es posible asignar rangos a los datos. Por ejemplo, puede asignarse rangos por nivel socioeconómico, textura o sabor; o bien por el género.
4. Como la estadística de distribución libre usa recuentos, rangos o los signos de diferencias de observaciones pareadas, suelen ser, aunque no siempre, rápida y fácil de aplicar y aprender.

**Desventajas.** Si se sabe que la forma de la población original es razonablemente cercana a una distribución normal, o si los datos pueden transformarse de modo que éste sea el caso, entonces los procedimientos de distribución libre extraen menos información que la que hay

disponible en los datos. En particular, la eficiencia de los procedimientos no paramétricos es bastante alta para muestras pequeñas. Por otra parte, la eficiencia puede no ser importante para muestras muy grandes.

**Principales métodos no-paramétricos**

- A) Prueba de X<sup>2</sup> de Bondad de Ajuste
- B) Prueba de Kolmogorov-Smirnov con una Muestra
- C) Prueba de Kolmogorov-Smirnov con dos Muestras
- D) Prueba de Rangos de Wilcoxon
- E) Prueba de Wilcoxon-Mann-Whitney con Dos Muestras
- F) Prueba de Kruskal-Wallis con k Muestras
- G) Prueba de los Signos
- H) Prueba de la Mediana
- I) Prueba de la Mediana para k Muestras
- J) Prueba de la Mediana para la Clasificación de Dos Vías
- K) Prueba de Friedman para la Clasificación de Dos Vías
- L) Prueba de Desigualdad de Chebyshev
- M) Coeficiente de Correlación de Rangos de Spearman
- N) Prueba de Asociación del Cuadrante de Olmstead-Tukey

Daniel (1978) provee otra clasificación de los diseños no-paramétricos como se indica en el siguiente cuadro (Brown & Mood, 1951; Theil, 1950; Hollander, 1970) (Tabla H).

Tabla H. Diferentes diseños no-paramétricos.

Tipo de problema	Tipo de data				
	1 muestra	2 muestras independientes	2 muestras relacionadas	Mas de 2 muestras independientes	Mas de 2 muestras relacionadas
<b>Localidad</b>	prueba de rangos de Wilcoxn	1 Prueba rápida de Tukey 2. Prueba de mediana 3. Mann-Whitney	prueba de rangos pareados de Wilcoxn	1. Prueba de mediana 2. Kruskal-Wallis 3. Prueba de ordenados alternos 4. Comparación multiple	1. Prueba de Friedman 2. Prueba de Durbin de bloques incompletos 3. Prueba de ordenados alternos 4.Comparacion

					multiple
<b>Dispersión</b>		1. Prueba de Mood 2. Prueba de Moses			
<b>Bondad de ajuste</b>	1. Ji-cuadrada 2. K-S = Kolmogorov-Smirnov	Kolmogorov-Smirnov			
<b>Asociación</b>	1. Spearman 2. Kendall 3. Olmstead-Tukey	Ji-cuadrada		1. Ji-cuadrada 2. Kendall	
<b>Regresión</b>	1. Brwon-Mood 2. Theil	Prueba de paralelismo			
<b>Micelánea</b>	1. Binomial 2. Cox-Stuart	1. Ward-Wolfowitz 2. Hollander 3. Rueba exacta de Fisher 4. Ji-cuadrada homogénea	McNemar		Cochran

A continuación se describen y ejemplifican los modelos de uso mas común en la práctica y en la literatura.

**Distribución de ji-cuadrada**

Vamos a suponer que en una muestra particular un conjunto de sucesos posibles  $E_1, E_2, E_3, \dots, E_k$  se observa que ocurren con frecuencias  $0_1, 0_2, 0_3, \dots, 0_k$ , llamadas *frecuencias observadas*, y que según las leyes de las probabilidades, se espera que sucedan con frecuencias  $e_1, e_2, e_3, \dots, e_k$  llamadas *frecuencias esperadas o teóricas*.

A menudo deseamos saber si las frecuencias observadas difieren significativamente de las esperadas. La medida que determina discrepancia entre las frecuencias observadas y esperadas en un experimento que involucra K celdas se define por el siguiente estadístico que es:

$$\chi^2 = \frac{(o_1 - e_1)^2}{e_1} + \frac{(o_2 - e_2)^2}{e_2} + \dots + \frac{(o_k - e_k)^2}{e_k} = \sum_{j=1}^k \frac{(o_j - e_j)^2}{e_j} \quad (1)$$

donde  $\chi^2$  es el valor de una variable aleatoria cuya distribución muestral se aproxima muy cercanamente a una distribución Ji-cuadrada con K-1 grados de libertad (número de clases -1);  $o_i$  = frecuencias observadas para la i-ésima celda;  $e_i$  = frecuencias esperadas para la misma celda.

Para la  $H_0: o_i = e_i$  ; Hay concordancia, mientras que para  $H_A: o_i \neq e_i$  ; no hay concordancia, de otra palbra, se rechaza la  $H_0$  si  $\chi^2_c \geq \chi^2_{\text{tabulado}}$ ; en el caso contrario se apoya la  $H_a$ .

Para los datos sobre la escala nominal, dicha medida de la discrepancia existente entre las frecuencias observadas y esperadas también se puede calcular como:

$$\chi^2 = \sum \frac{o_j^2}{e_j} - N \quad (2)$$

Donde,  $N$  = la frecuencia total, y  $\sum o_j = \sum e_j = N$

Si  $\chi^2 = 0$ , las  $o_i$  y las  $e_i$  coinciden completamente; mientras que si  $\chi^2 > 0$ , las  $o_i$  y las  $e_i$  no coinciden exactamente. A valores más grandes de  $\chi^2$ , mayor discrepancia entre las frecuencias observadas y esperadas. Si consideramos muestras de tamaño  $N$  tomadas de una población normal con desviación estandar  $\sigma$ , y si para cada muestra calculamos, se obtiene para  $\chi^2$  una distribución de muestreo, llamada *distribución ji - cuadrada*, que viene dada por:

$$Y = Y_0 (\chi^2)^{\frac{1}{2}(v-2)} e^{-\frac{1}{2}\chi^2} \tag{3}$$

donde  $v = n - 1$  es el número de grados de libertad que se define como el número  $n$  de observaciones independientes en la muestra (o sea, el tamaño de la muestra) menos el número  $k$  de parámetros de la población, y  $Y_0$  es una constante que depende del  $v$  de tal manera que el área total bajo la curva es 1. La distribución ji-cuadrada correspondiente a varios valores de  $v$  se muestran en la Fig. 1. El máximo valor de  $Y$  ocurre en  $\chi^2 = v - 2$ .

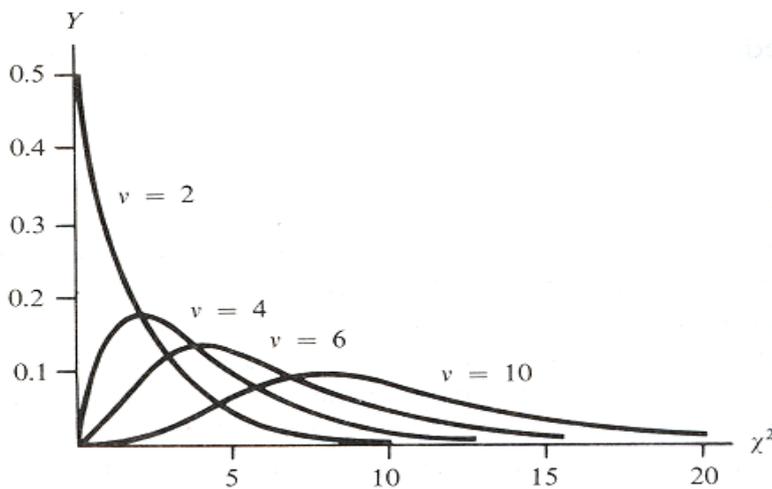


Figura 1. Distribución de Ji-cuadrada para varios valores de grado de libertad ( $v = 2 - 10$ ).

Para el cálculo de un estadístico tal como (ecuación 3), es necesario emplear tanto observaciones de muestras como propiedades de ciertos parámetros de la población. Si estos parámetros son desconocidos, hay que estimarlos a partir de la muestra.

En la práctica, las frecuencias esperadas se calculan sobre la base de una hipótesis  $H_0$ . Si bajo tal hipótesis el valor calculado para  $\chi^2$  es mayor que algún valor crítico, debemos concluir que las frecuencias observadas difieren *significativamente* de las frecuencias esperadas y rechazaremos  $H_0$  al correspondiente nivel de significación; en caso contrario, la apoyamos. Este procedimiento se llama el test o contraste ji-cuadrada de hipótesis o significación.

Hay que hacer constar que debe mirarse con suspicacia en circunstancias en las que  $\chi^2$  sea *demasiado próximo a cero*, pues es raro que las frecuencias observadas coincidan *demasiado bien* con las frecuencias esperadas. Para examinar tales situaciones, podemos determinar si el valor calculado de  $\chi^2$  es menor que  $\chi^2$  a nivel de  $\alpha = 0.05$  o  $\chi^2$  a nivel de  $\alpha = 0.01$ , en cuyo caso se decide que el acuerdo es demasiado bueno al nivel de significación.

**Ejemplo 1.** En un estudio sobre el color de las flores de rosal se seleccionó un muestra aleatoria de 100 flores compuestas de 84 flores amarillas y 16 verdes. Prueba de la hipótesis nula en la cual la muestra proviene de una población con la proporción 3:1 para las flores amarillas y verdes

respectivamente. Las frecuencias esperadas para las flores amarillas y verdes son 75 y 25 respectivamente. Con base a las frecuencias observadas y teóricas se calcula como sigue y con el valor de  $\chi^2$  tabulada para  $K-1 = 2-1 = 1$  grado de libertad.

$$\chi^2 = \frac{(84 - 75)^2}{75} + \frac{(16 - 25)^2}{25} = 4.320$$

Tabla 1. Relación entre el valor chi-cuadrada y el grado de probabilidad del error.

$\alpha$	0.10	0.05	0.025	0.01
$\chi^2$	2.760	3.841	5.024	6.635

A medida que disminuye el  $\alpha$ , aumenta el valor de Ji-cuadrada.

$\chi^2_{0.05, 1} = 3.841$  por lo tanto  $\chi^2_c$  (4.320) es superior al  $\chi^2_{\alpha, 1}$ , y se rechaza la hipótesis nula con una probabilidad  $0.025 < P(\chi^2 \geq 4.320) < 0.05$ .

**Ejemplo 2.** En 1865, Mendel presentó las relaciones de la población  $F_2$  de la cruce de dos Chicharos monohíbridos, como indica la Tabla 2.

Tabla 2. Relación de la población  $F_2$  de la cruce según Mendel, 1865.

Semillas	Observado experimental(o)	Relación teórica	Relación teórica en 8023 (e)	Desviación (o - e)	$(o-e)^2 / e$
Amarillas	6022	3	6017.25	+4.75	0.00375
Verdes	2001	1	2005.75	-4.75	0.01125
Suma (N)	8023	4	8023.00	0.00	$\chi^2 = 0.015$

El número teórico esperado se calcula  $E = p(\text{probabilidad}) \times n$  (tamaño de la muestra).

Semillas amarillas (E) = $(\frac{3}{4}) 8023 = 6017.25$
Semillas verdes (E) = $(\frac{1}{4}) 8023 = 2005.75$
$\chi^2 = 0.015 < \chi^2_{0.05, 1} = 3.841$

Conclusión: Se apoya la  $H_0$  ya que los valores experimentales se ajustan a los esperados por lo que las diferencias o desviaciones no son significativas.

**Ejemplo 3.** También Mendel presentó las relaciones proporcionales de 9, 3, 3, 1 de la población  $F_2$  de un híbrido en el chícharo de acuerdo a la Tabla 3.

Tabla 3. Relación de la población  $F_2$  de la cruce con hipótesis de proporciones de 9, 3, 3, 1.

Semillas	(o)	(e)	(o-e)	$(o-e)^2 / e$
Redondas y amarillas	315	312.75	+2.25	0.0162
Redondas y verdes	108	104.25	+3.75	0.1349
Arrugadas y amarillas	101	104.25	-3.25	0.1349
Arrugadas y verdes	32	34.75	-2.75	0.2176
Sumas	556	556		$\chi^2 = 0.4700$

Conclusión:  $\chi^2 = 0.47 < \chi^2_{0.05, 3} = 7.815$ , por lo tanto los valores observados experimentalmente se ajustan a los teóricos esperados; es decir, hay concordancia.

La prueba de ji-cuadrada puede utilizarse para determinar la calidad del ajuste mediante distribuciones teóricas (normal o binomial) de distribuciones empíricas. Las tablas cuyas las frecuencias observadas ocupan una sola fila, se llaman una tabla de *clasificación de entrada única*. Extendiendo estas ideas, podemos llegar a tablas de doble entrada ( $h \times k$ ) en las que las frecuencias observadas ocupan  $h$  filas y  $k$  columnas. Tales tablas se suelen llamar *tablas de contingencia*.

Correspondiendo a cada frecuencia observada, hay una frecuencia esperada que se calcula sujeta a ciertas hipótesis de acuerdo con las leyes de las probabilidades. Estas frecuencias, que ocupan las celdas de una tabla de contingencia, se llaman *frecuencias de celda*. La frecuencia total en cada fila o en cada columna se llama la *frecuencia marginal*. Para investigar el acuerdo entre las frecuencias observadas y las frecuencias esperadas, calculamos el estadístico

$$\chi^2 = \sum_j^k (o_j - e_j) / e_j \quad (4)$$

donde  $o_j$  y  $e_j$  representan, respectivamente, las frecuencias observadas y frecuencias esperadas de la  $j$ -ésima celda. El número de grado de libertad viene dado por  $h > 1$  y  $k > 1$  por  $(h-1)(k-1)$ .

Los contrastes de significación para las tablas  $h \times k$  son similares a los de las tablas  $1 \times k$ . Las frecuencias esperadas se hallan sujetas a una hipótesis particular  $H_0$ . Una hipótesis común es suponer que las dos clasificaciones son mutuamente independientes. Las tablas de contingencia se pueden generalizar a más dimensiones. Así, por ejemplo, puede tener tablas  $h \times k \times 1$ , donde están presentes tres clasificaciones.

**Ejemplo 4.** Las frecuencias observadas obtenidas en un estudio de tendencias políticas según la orientación política se presentan en la Tabla 4.

Tabla 4. Orientación política y sus regidores.

Afiliación	Liberales	Conservadores	Totales
Regidores	5 (7.5)	10 (7.5)	15
No regidores	15 (12.5)	10 (12.5)	25
Totales/orientación	20	20	40

Prueba de la hipótesis nula en la cual no hay diferencias entre las frecuencias observadas y teóricas de dicha orientación. Cálculo de frecuencias teóricas según sigue.

$F_e = (\text{Totales/hilera}) (\text{totales/ columna}) / N$ , o sea,  $F_{11}$  y  $F_{12} = 15 \times 20 / 40 = 7.5$ , asimismo  $F_{21}$  y  $F_{22} = 25 \times 20 / 40 = 12.5$

$$\chi^2 = \frac{(5-7.5)^2}{7.5} + \frac{(10-7.5)^2}{7.5} + \frac{(15-12.5)^2}{12.5} + \frac{(10-12.5)^2}{12.5} = 2.66, \quad \text{g.l.} = (2-1)(2-1) = 1$$

$\chi^2_{0.05, 1} = 3.841 > \chi^2_c (2.66)$ , por lo tanto se acepta la hipótesis nula con una probabilidad  $0.10 < p (\chi^2 \geq 2.66) < 0.25$  y se concluye que no existe diferencias significativas entre las frecuencias observadas y esperadas en lo que se refiere a la orientación política.

**Corrección de Yates.** Cuando se aplican resultados de distribuciones continuas de datos discretos, pueden hacerse ciertas correcciones a la continuidad. Una corrección similar existe cuando se usa la distribución ji-cuadrada. La corrección consiste en reformular la ecuación de  $\chi^2$  como

$$\chi^2(\text{corregido}) = \sum \frac{(\hat{o}_j - e_j - 0.5)^2}{e_j} \quad (5)$$

En general, la corrección se hace sólo cuando el número de grados de libertad es igual a 1. Para grandes muestras, esto da prácticamente los mismos resultados que el  $\chi^2$  sin corregir, pero pueden surgir dificultades cerca de los valores críticos. Para pequeñas muestras donde cada frecuencia esperada está entre 5 y 10, es quizás mejor comparar ambos valores de  $\chi^2$ , corregido y sin corregir. Si ambos llevan a la misma conclusión acerca de la hipótesis, tal como el rechazo al nivel de significación 0.05, rara vez surgen dificultades. Si conducen a diferente conclusión, uno debe pensar en aumentar el tamaño de la muestra.

**Cálculos alternativos de Ji-cuadrada**

Existen fórmulas sencillas para calcular  $\chi^2$  que implican tan sólo las frecuencias observadas. Lo que sigue da los resultados para las tablas de contingencia 2 x 2 y 2 x 3.

A	B	A+B
C	D	C+D
A+C	B+D	N

$$\chi^2 = \frac{N(AD - BC)^2}{(A + B)(C + D)(A + C)(B + D)} \tag{6}$$

o

$$\chi^2 = \frac{N(\frac{1}{2}AD - BC \frac{1}{2} - \frac{1}{2}N)^2}{(A + B)(C + D) + (A + C)(B + D)} \tag{7}$$

	I	II	III	Total
A	$a_1$	$a_2$	$A_3$	$N_A$
B	$b_1$	$b_2$	$B_3$	$N_B$
Total	$N_1$	$N_2$	$N_3$	$N$

$$\chi^2 = \frac{N}{N_A} \left[ \frac{a_1^2}{N_1} + \frac{a_2^2}{N_2} + \frac{a_3^2}{N_3} \right] + \frac{N}{N_B} \left[ \frac{b_1^2}{N_1} + \frac{b_2^2}{N_2} + \frac{b_3^2}{N_3} \right] - N \tag{8}$$

donde hemos usado el resultado general válido para todas las tablas de contingencia

**Coefficiente de contingencia**

Una medida del grado de interrelación, asociación o dependencia de las clasificaciones en una tabla de contingencia viene dada por:

$$C = \sqrt{\frac{\chi^2}{\chi^2 + N}} \tag{9}$$

Dicha relación se llama el coeficiente de contingencia. Cuanto mayor es  $C$ , mayor es el grado de asociación. El número de filas y de columnas en la tabla de contingencia determina el máximo valor de  $C$ , que nunca es mayor que 1. Si el número de filas y columnas de una tabla de contingencia es igual a  $k$ , el valor máximo de  $C$  está dado por  $\sqrt{(k-1)/k}$

**Bondad de ajuste por el método Kolmogorov-Smirnov**

La prueba recibe su nombre de A. Kolmogorov y N.V. Smirnov, dos matemáticos rusos quienes la introdujeron en la década de 1930. Esta prueba proporciona una alternativa de bondad de ajuste cuando se desea determinar qué tan buena se conforma la distribución de los datos de la muestra a alguna distribución teórica sobre escala ordinal. Esta prueba se basa sobre el estadístico “ $D_i$ ” que se calcula como:

$$|D_i| = |F_{oACUM} - F_{TACUM}| \tag{10}$$

Donde:  $D_i$  = Distancia que existe entre valores absolutos de la diferencia entre  $f_o$ , done el valor más grande de  $|D_i|$  se llama  $D_{max}$ . Al compararse con el valor tabulado de  $D_i$  la hipótesis nula ( $H_o$ ) se puede aceptar o rechazar.

**Ejemplo 5.** La preferencia a los alimentos con diferentes niveles de vitaminas (Tabla 5a).  $H_o$ : No hay preferencia alimenticia.  $H_A$ : Hay preferencia alimenticia.

Tabla 5a. Analisis de datos según Kolmogorov-Smirnov.

	Niveles o clases de vitaminas ( $X_i$ )					
	1	2	3	4	5	N
Frecuencia observada ( $F_o$ )	2	18	10	4	1	35
Frecuencia teórica ( $F_t$ )	7	7	7	7	7	35
$F_{Oacum}$	2	20	30	34	35	
$F_{Tacum}$	7	14	21	28	35	
$ D_i $	5	6	9	6	0	

$D_{I_{max}} = 30 - 21 = 9 > D_i_{MAX} 5,35 = 7$  (valor tabulado de la tabla de contingencia), entonces se rechaza el  $H_o$  con una probabilidad de  $0.002 < 0.005$ , y se concluye que las personas no tienen la misma preferencia en seleccionar los alimentos con diferentes grados de vitaminas.

### Bondad de ajuste Kolmogorov-Smirnov para datos continuos

La bondad de ajuste Kolmogorov-Smirnov fue originalmente creada para datos continuos (proporcional, intervalos, etc. [Tabla 5b],) sin embargo, fue aplicada posteriormente a los datos discretos. En este caso se calcula el estadístico  $D_i$  o  $D'_i$  para apoyar o rechazar la  $H_o$ :  $D_i = |F_{o_{real}} - F_{t_{real}}|$  (11),  $D'_i = |F_{o_{real-1}} - F_{t_{real}}|$  (12), donde,  $F_{o_{real}} = f_{Acumuladas}/N_i$  (la proporción de una muestra medida) y  $F_{t_{real}} = X_i/\text{valor máxima}$ . Si  $D_i < D_{\alpha, n}$  se apoya la hipótesis nula, en el caso contrario se la rechaza.  $D_{i_{Max}} = D_{12} = 0.3707$  o  $D'_{i_{Max}} = D'_{12} = 0.3040$  son superiores al valor  $0.3376$  ( $D_{0.05, 15} = 0.3376$  proviene de una tabla para este modelo) ó  $D'_{\alpha, n}$  y, por lo tanto, se rechaza la uniformidad de la distribución vertical (Tabla 5b) en función de tiempo de observación (se rechaza la hipótesis nula), donde  $P > 0.05$ .

Tabla 5b. La distribución vertical del incremento en altura de una especie vegetal en un ecosistema con altura máxima igual a 25 metros.

Orden	$x_i$	$f_{obs}$	$f_{acum}$	$F_{obs-real}$	$f'_{real}$	$D_i$	$D'_i$
1	1.4	1	1	0.0667	0.0560	0.0107	0.0560
2	2.6	1	2	0.1337	0.1040	0.1230	0.0373
3	3.3	1	3	0.2000	0.1320	0.0680	0.0017
4	4.2	1	4	0.2667	0.1680	0.0987	0.0320
5	4.7	1	5	0.3333	0.1880	0.1453	0.0787
6	5.6	2	7	0.4667	0.2240	0.2427	0.1093
7	6.4	1	8	0.5300	0.3560	0.2773	0.2107
8	7.7	1	9	0.6000	0.3080	0.2920	0.2253
9	9.3	1	10	0.6667	0.3720	0.2947	0.2280
10	10.6	1	11	0.7333	0.4240	0.3093	0.2427
11	11.5	1	12	0.8000	0.4600	0.3400	0.2733
12	12.4	1	13	0.8667	0.4960	<b>0.3707</b>	<b>0.3040</b>
13	18.6	1	14	0.9333	0.744	0.1893	0.1227
14	22.3	1	15	1.0000	0.8920	0.1080	0.0413

**K muestras independientes: Prueba de la mediana**

La prueba de la mediana determina si K grupos independientes (no necesariamente de igual tamaño) han sido tomados de la misma población o de poblaciones con medianas iguales. Es útil cuando la variable en estudio ha sido medida por lo menos en una escala *ordinal*.

*Suposiciones:*

1. Las dos muestras deben ser extraídas aleatoria o independientemente de sus respectivas poblaciones.
2. La escala de medición de la variable de respuesta debe ser al menos ordinal.

*Hipótesis:*

H<sub>0</sub>: Las medianas de las poblaciones son iguales, por lo que los tratamientos experimentales tienen el mismo efecto. Lo que normalmente se expresa como:

$$H_0: M_1=M_2=M_3=M_4.....M_k$$

H<sub>a</sub>: Las medianas de las poblaciones difieren significativamente, por lo que los tratamientos tienen efectos diferentes.

$$H_a: M_1 \neq M_2 \neq M_3 \neq ..... M_k$$

Para probar la hipótesis de nulidad de que las K muestras proceden de la misma población con respecto a las medianas, calculamos el valor de  $\chi^2$  con la fórmula:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^k \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \tag{13}$$

donde O<sub>ij</sub> = número observado de casos clasificados en la hilera i de la columna j. E<sub>ij</sub> = número de casos esperados conforme a H<sub>0</sub> para ser clasificados en la hilera i de la columna j;  $\sum_{i=1}^r \sum_{j=1}^k =$

indica sumar todas las celdas.

Regla de decisión. La hipótesis nula (H<sub>0</sub>) se rechazará si  $\chi^2_c > \chi^2_{t, r-1, \alpha}$ ; donde r es el número de poblaciones o renglones de una tabla de contingencia de r x 2.

**Ejemplo 6.** Se realizó un experimento con cuatro variedades de maíz, con el fin de determinar si existen diferencias significativas en cuanto a su rendimiento. Las cuatro variedades fueron asignadas aleatoriamente a 34 parcelas experimentales ubicadas en la misma localidad. El rendimiento en Kg fue registrado para cada una de las 34 parcelas, obteniéndose los siguientes resultados (Tabla 6).

Tabla 6. Rendimientos en las 34 parcelas.

Variedades	Rendimiento										
V1	83	91	94	89	89	96	92	90	91		n <sub>1</sub> =9
V2	91	90	81	83	83	84	88	91	89	84	n <sub>2</sub> =10
V3	101	100	91	93	96	95	94				n <sub>3</sub> =7
V4	78	82	81	77	79	81	80	81			n <sub>4</sub> =8
											N=34

*Hipótesis:* H<sub>0</sub>: M<sub>1</sub>=M<sub>2</sub>=M<sub>3</sub>=M<sub>4</sub>; *los medianas de rendimiento son iguales entre las cuatro variedades.*

H<sub>a</sub>= M<sub>1</sub>≠M<sub>2</sub>≠M<sub>3</sub>≠M<sub>4</sub> al menos un par de medianas es diferente.

*Procedimiento:*

1. Se ordenan las 34 observaciones en orden creciente de magnitud, sin perder de vista que la población pertenece cada observación.
2. Se calcula la mediana (Me) del conjunto de observaciones ya ordenadas en el inciso 1. De esta manera la mediana resulta ser = 89
3. Se clasifican las 34 observaciones de acuerdo a la siguiente tabla de contingencia de 4x2 (r X c).

Variedades	Observaciones < 89	Observaciones > 89	Totales Hileras
V1	3(4.7)	6(4.2)	9
V2	7(5.2)	3(4.7)	10
V3	0(3.7)	7(3.2)	7
V4	8(4.2)	0(3.7)	8
Totales/Columna	18	16	34

El cálculo del estadístico de prueba se presenta a continuación:

$$X^2 = \frac{(3 - 4.7)^2}{4.7} + \frac{(6 - 4.2)^2}{4.2} + \dots + \frac{(0 - 3.7)^2}{3.7} = 17.97$$

De la tabla se obtiene con g.l. = 4-1=3 y  $\alpha = 0.05$ ,  $\chi^2 = 7.82$

Conclusión: Se rechaza la  $H_a$ , la mediana de rendimiento es diferente, al menos en dos variedades.

### Dos muestras independientes: prueba Wald-Wolfwits

Esta prueba es útil cuando se desean hacer inferencias sobre dos poblaciones independientes. Mediante esta prueba se puede probar si las dos poblaciones difieren, ya sea en alguna medida de tendencia central o de dispersión, o bien en su distribución de probabilidad.

*Supuestos:*

- 1) Las dos muestras deben ser tomadas aleatorias e independientemente de sus respectivas poblaciones.
- 2) La escala de medición de la variable de respuesta debe ser al menos ordinal.

*Hipótesis:*

$H_0$ : La distribución de probabilidades de las dos poblaciones es la misma, es decir  $F_a(x) = F_b(x)$ .

$H_a$ : La distribución de probabilidades de las dos poblaciones es diferente es decir  $F_a(x) \neq F_b(x)$ .

*Estadístico de prueba  $T_o$  : Regla de decisión:* rechaza la hipótesis nula ( $H_0$ ) si  $T_o < W$ , donde, W es el valor crítico del lado izquierdo de la distribución y se busca en una tabla correspondiente.

**Ejemplo 7.** Se desea estudiar el efecto de dos dietas diferentes (lirio acuático y harina de pescado), sobre el incremento del peso en la carpa herbívora *Ctenopharyngodon idella*. Con este fin se dispuso de 22 lotes de 20 peces cada uno, a 12 de los lotes se les alimento diariamente (durante dos meses) con lirio acuático. A los 10 restantes se les alimentó de igual manera, pero con harina de pescado. Al término de los dos meses se registro la ganancia promedio del peso (g) por lote, obteniéndose los siguientes resultados (Tabla 7).

Tabla 7. Datos de ganancias promedias por diferentes alimentos.

Dieta	Ganancia promedio del peso (gr. por lote)											
	Lirio acuático (x)	134	146	104	119	124	161	107	83	113	129	97
Harina de pescado (y)	70	118	101	85	107	132	94	86	90	95		
$N_x=12$												
$N_y=10$												

*Hipótesis:*

$H_0$ :  $F_x(z) = F_y(z)$ ; es decir las dietas tienen el mismo efecto sobre la ganancia del peso.

$H_a$ :  $F_x(z) \neq F_y(z)$ ; es decir las dietas no tienen el mismo efecto sobre la ganancia del peso.

*Procedimiento:*

- 1) Ordenar las 22 observaciones en forma creciente de magnitud, sin perder de vista a que lapoblación pertenece cada observación.

70(y)	83(x)	85(y)	86(y)	90(y)	94(y)	95(y)	97(y)	101(y)	104(x)	107(x)	107(y)
113(x)	118(y)	119(x)	123(x)	124(x)	129(x)	132(y)	134(x)	146(x)	161(x)		

- 2) Calcular  $T_0$  (número de corridas) a partir del arreglo ordenado obtenido en el paso 1.

70(y) 83(x) 85(y)86(y)90(y)94(y)95(y) 97(x) 101(y) 104(x)107(x) 107(y)  
 1 2 3 4 5 6 7

113(x) 118(y) 119(x)123(x)124(x)129(x) 132(y) 134(x)146(x)161(x)  
 8 9 10 11 12

De esta manera el valor del estadístico de prueba para este primer arreglo es  $T_1=12$

Como en este problema existe un empate en el valor 107 entre las observaciones de “x” y “y”, se hace un segundo arreglo donde se permuten de posición las observaciones donde existe el empate (la observación de y se cambia a la primera posición).

70(y) 83(x) 85(y)86(y)90(y)94(y)95(y) 97(x) 101(y) 104(x)107(x) 107(y)  
 1 2 3 4 5 6 7

113(x) 118(y) 119(x)123(x)124(x)129(x) 132(y) 134(x)146(x)161(x)  
 8 9 10 11 12

Cambiando quedaría así

70(y) 83(x) 85(y)86(y)90(y)94(y)95(y) 97(x) 101(y) 104(x)107(y) 107(x)  
 1 2 3 4 5 6 7

113(x) 118(y) 119(x)123(x)124(x)129(x) 132(y) 134(x)146(x)161(x)  
 8 9 10 11 12

El valor del Estadístico para este arreglo es  $T_2 =12$

- 3) Se obtiene el valor crítico W en una tabla correspondiente para este modelo. Este valor se obtiene utilizando los tamaños de muestra de cada grupo ( $N_1= 10$ ,  $N_2 =12$ ) y un nivel  $\alpha = 0.5$ , el cual = 8.

- 4) Decisión: se rechaza  $H_0$ : si  $T_0 < W$ , dado que el valor de  $T_0 = 12$  y  $W = 8$ , se aoya la  $H_0$ , es decir,

las dos dietas tuvieron el mismo efecto sobre la ganancia promedio del peso.

**Dos muestras pareadas: prueba de McNemar para la significancia de los cambios**

Esta prueba es particularmente apropiada para los diseños de antes y después en los que cada persona es utilizada como su propio control. Los datos que se utilizan deben de encontrarse en una escala nominal y ordinal.

La potencia y eficiencia de la técnica (cuando es posible por la escala y otros aspectos de los datos la utilización de la prueba de t), es de cerca de 95% para  $A + D = 6$  y declina a medida que aumenta el tamaño de  $A + D$ , hasta una eficiencia final asintótica de 63%. Si las frecuencias esperadas son muy pequeñas deberá utilizarse la prueba binomial antes que la de McNemar.

ANTES	DESPUES	
	-	+
+	A	B
-	C	D

Puesto que  $A + D$  representan el número total de personas que cambiaron se espera conforme a la hipótesis de nulidad que  $(A + D)$  casos cambiaran en una dirección y el resto en otra. Por lo que desarrollando y reuniendo términos, tenemos:

Primero se recordara que

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} \tag{14}$$

Sin embargo, en nuestro caso los que nos interesa son las diferencias en las celdillas A y D, luego:

$$\chi^2 = A - \frac{(A + D)^2}{2(A + D)} + D - \frac{(A + D)^2}{2(A + D)} \tag{15}$$

$$\text{Por lo tanto: } \chi^2 = \frac{(A - D)^2}{A + D} \text{ con gl} = 1 \tag{16}$$

Incluyendo una corrección por continuidad (ya que se utiliza una distribución continua  $\chi^2$ , para aproximar una discreta), tenemos que:

$$\chi^2 = \frac{(|A - D| - 1)^2}{A + D} , \text{ con gl} = 1 \tag{17}$$

Si el valor de  $\chi^2$  calculado es mayor que el de tablas, se rechaza  $H_0$ , es decir que si hay un efecto significativo de antes y después del tratamiento.

**Ejemplo 8.** Se desea saber si los niños que llegan a una guardería tienen la misma inclinación por relacionarse con otros niños o con un adulto el primer día de clases que un mes después (Tabla 8).

Tabla 8. Inclínación de los niños entre sí y con los adultos en diferentes fechas.

Primer día de clases			
		Con otro niño	Con un adulto
Un mes después	Con un adulto	14	4
	Con otro niño	3	4

$$\chi^2 = \frac{(|14 - 4| - 1)^2}{14 + 4} = \frac{9^2}{18} = 4.5$$

$\chi^2$  de tablas es menor que 4.5 con una  $P < 0.025$  por lo que se rechaza  $H_0$  por lo tanto los niños muestran una tendencia significativa a cambiar su preferencia por relacionarse con otros niños conforme pasa el tiempo.

### Prueba de rangos con signo de Wilcoxon

Esta prueba es considerada como una alternativa a la prueba de t para dos muestras pareadas. El procedimiento de ambas pruebas se basa en el cálculo de diferencias ( $D_i = x_i - y_i$ ) entre pares de observaciones, pero en la prueba de Wilcoxon se asignan rangos a las diferencias. Cabe mencionar que, en esta prueba la hipótesis se plantea en torno a la mediana de las diferencias ( $M_d$ ), mientras que en la prueba de t se plantea sobre la media de diferencias ( $\mu_D$ ).

La eficiencia de esta prueba es de 95% para muestras pequeñas. Cuando las suposiciones para la prueba paramétrica de t se satisfacen la eficiencia asintótica cercana a  $H_0$  de esta prueba comparada con la de t es de 95.5%.

#### Suposiciones:

1. Cada  $D_i$  es una variable aleatoria continua.
2. La distribución de cada  $D_i$  es simétrica.
3. Las  $D_i$  son independientes.
4. La escala de medición es al menos de intervalo en las  $D_i$ .

*Hipótesis:* Sea  $M_d$  la mediana de las  $D_i$ .

- A. Hipótesis bilateral:  $H_0 : M_d = 0$  vs  $H_A : M_d \neq 0$
- B. Hipótesis unilateral:  $H_0 : M_d \leq 0$  vs  $H_A : M_d > 0$
- C. Hipótesis unilateral :  $H_0 : M_d \geq 0$  vs  $H_A : M_d < 0$

La hipótesis alternativa en B puede interpretarse como “los valores de  $x$  son mayores que los de  $y$ ”, y en el caso de C como “los valores de  $y$  son mayores que los de  $x$ ”.

#### Estadístico de prueba:

$W_0^+ = \sum R_i^+$  donde  $\sum R_i^+ \Rightarrow$  es la suma de rangos asociados a las  $D_i$  con signo [+].

Alternativamente se puede utilizar como estadístico a  $D W_0^- = \sum R_i^-$

#### Regla de decisión:

Se utiliza la tabla de valores críticos de  $W$  para la prueba de Wilcoxon.

- 1.- Hipótesis bilateral: Rechazar  $H_0$  si  $W_0 < W_{\frac{\alpha}{2}}$  ó  $W_0 > W_{1-\frac{\alpha}{2}}$ . Donde,  $W_{1-\frac{\alpha}{2}}$  y  $W_{\frac{\alpha}{2}}$ , son los valores críticos del lado izquierdo y derecho respectivamente. El valor de  $W_{\frac{\alpha}{2}}$  se obtiene

directamente en la tabla, y  $W_{1-\frac{\alpha}{2}}$  mediante la relación:  $W_{1-\frac{\alpha}{2}} = \frac{n(n+1)}{2} - W_{\frac{\alpha}{2}}$ . El término  $\frac{n(n+1)}{2}$ , se obtiene de la última columna de la misma tabla.

2.- Hipótesis unilateral: Rechazar  $H_0$  si  $W_o < W_{1-\alpha}$  donde  $W_{1-\alpha} = \frac{n(n+1)}{2} - W_{\alpha}$

3.- Hipótesis unilateral: Rechazar  $H_0$  si  $W_o < W_{\alpha}$

En 2 y 3 el nivel  $\alpha$  de significancia no se divide entre dos por tratarse de hipótesis unilaterales.

**Ejemplo 9.** Se desea determinar si existe algún efecto sobre el número de mazorcas producidas por planta de maíz, al eliminar un fertilizante de una compañía que dejó de producir dicho producto. En el caso de que la falta de dicho fertilizante tenga efectos negativos, se buscaron alternativas necesarias para suplir este producto.

Se utilizaron para el experimento, ocho parcelas divididas en dos porciones cada una; aplicándose en una de las mitades el fertilizante común y dejando a la otra como control. Los resultados se reportan en # de mazorcas por porción de parcela (Tabla 9)

Tabla 9. Rendimiento por proporción de parcela.

No. de parcela	1	2	3	4	5	6	7	8
Fertilizante (x)	31	20	18	17	9	8	10	7
Control (y)	18	17	14	11	10	7	5	6

Hipótesis nula ( $H_0$ ):  $M_d = 0$ . No existen diferencias significativas en el número de mazorcas, entre la porción control y la tratada con fertilizante. Hipótesis ( $H_A$ ):  $M_d \neq 0$

*Procedimiento:*

1) Se determina la diferencia ( $D_i = x_i - y_i$ ) para cada par de observaciones.

No. de parcela	1	2	3	4	5	6	7	8
$D_i = x_i - y_i$	13	3	4	6	-1	1	5	1

Cuando  $D_i = 0$  se elimina esta pareja de datos.

2) Se calculan los valores absolutos de las  $D_i$ , y se ordenan de menor a mayor, con el fin de asociarles rangos.

No. de parcela	5	6	8	2	3	7	4	1
$ D_i $	1	1	1	3	4	5	6	13
Rango ( $R_i$ )	2	2	2	4	5	6	7	8

Como existen tres  $|D_i|$  con valores de 1, no existe razón alguna para asignarle un rango menor a uno de ellos y uno mayor a la otra, por lo tanto, en éste y solamente en éste caso, se les asigne a estas  $|D_i|$  el promedio aritmético de sus rangos, esto es  $(1+2+3)/3=2$ .

3) Se calcula el estadístico de prueba  $W_o^+ = \sum R_i^+$ .

Dado que existen 7  $|D_i|$  con signo (+), la suma de sus rangos es igual a 34, por lo que  $W_o^+ = 34$ . Alternativamente, podemos utilizar a  $W_o^-$  como estadístico, la hipótesis nula ( $H_0$ ) se rechazaría si el valor de  $W_o^+$  es muy grande (lo que significaría que  $W_o^-$  es muy pequeño y por lo tanto difieren mucho) ó bien si el valor de  $W_o^+$  es muy pequeño lo que implica que  $W_o^-$  es muy grande y de nuevo difieren.

4) Puesto que estamos tratando de probar una hipótesis bilateral, entonces se debe obtener los valores de  $W_{\alpha/2}$  y  $W_{1-\alpha/2}$ .

$W_{\alpha/2}$  se busca en una tabla correspondiente para  $n=8$  (número de observaciones pares) y un  $\alpha=0.05$

$$W_{0.05/2} = W_{0.025} = 4$$

$$W_{1-\alpha/2} = W_{0.975} = \frac{n(n+1)}{2} - W_{0.025}$$

$$= 36 - 4 = 32$$

El término  $n(n+1)/2 = 36$ , se busca en la última columna de la tabla.

5) Decidimos de Rechazar  $H_0$  si  $W^+_o < W_{0.025}$  ó  $W^+_o > W_{0.975}$ ; es decir, si  $W^+_o < 4$  ó  $W^+_o > 32$ . Como  $W^+_o = 34$  es mayor que  $W_{0.975} = 32$  rechazamos  $H_0$ , y concluimos que existen diferencias significativas en el número de mazorcas entre la porción con fertilizante y la control. Revisando los valores originales del número de mazorcas, nos podemos percatar que estos son más grandes en la porción con fertilizante.

### K muestras pareadas: prueba de Q de Cochran

La prueba de McNemar puede utilizarse para 2 muestras relacionadas, una extensión de N prueba se puede utilizar para 3 o más muestras. Esta extensión se conoce como Prueba de Q de Cochran, y nos sirve para determinar si 3 o más conjuntos igualados de frecuencias o proporciones difieren entre si.

#### Suposiciones:

- 1) Se tiene K muestras pareadas o relacionadas.
- 2) La escala de medición de la variable de respuesta es nominal o se ha dicotomizado la información ordinal, en éxito y fracaso, (+) o (-) etc.

La potencia y exactitud de esta prueba no se conoce exactamente, sin embargo, cuando se cuenta con datos en cualquiera de las escalas mencionadas anteriormente no se puede utilizar una prueba paramétrica por lo que esta es nuestra única opción; pero si los datos no están en estas escalas la prueba puede ser disipadora de información.

Los datos se colocan en una tabla formada de N hileras y K columnas, donde N es el número de datos para cada K y K = a la técnica o ítem a probar (tratamiento). Estos datos se prueban contra una hipótesis.

$H_0$  : Las K poblaciones son iguales ( sin diferencias en la probabilidad de “éxito” en cada condición).

$H_A$  : Al menos un par de poblaciones son diferentes.

Cochran ha demostrado que si la hipótesis de nulidad es verdad, al no ser demasiado pequeño el número de hileras, tenemos que el estadístico de prueba es:

$$Q = \frac{k(k-1) \sum_j^k T_{.j}^2 - (k-T)(T_{..})^2}{k(T_{..}) - \sum_i^r B_i^2} \quad (18)$$

y que Q se distribuye aproximadamente como chi-cuadrada con  $gl = k-1$ ,

$T_j$  = total por tratamiento o columna

$B_i$  = total por bloque o hilera o renglón

$T_{..}$  = total general

$J = 1 \dots k$  (número de tratamientos o columnas)

$I = 1 \dots r$  (número de bloques o hileras)

*Decisión:* Si  $Q > X^2_{k-1}$ , se rechaza  $H_0$ .

**Resumen del procedimiento**

1. Para los datos dicótonos, se asigna un puntaje de 1 a cada “éxito” y un puntaje de 0 a cada “fracaso”.
2. Se arreglan los puntajes en una tabla  $k \times N$  usando  $k$  columnas y  $N$  hileras = al número de casos en cada grupo  $h$ .
3. Se determina el valor de  $Q$  sustituyendo los valores observados en la fórmula.
4. La significación del valor observado de  $Q$  puede determinarse consultando la tabla  $C$ , porque  $Q$  está distribuida aproximadamente como chi-cuadrada con  $gl = k-1$ . Si la probabilidad asociada con la ocurrencia conforme a  $H_0$  de un valor tan grande como el valor observado de  $Q$  es igual o menor que  $\alpha$ , se Rechaza  $H_0$ .

**Ejemplo 10.** Se desea conocer si tres técnicas indirectas para detectar yodo radioactivo en la glándula tiroides, son igualmente efectivas. Con este fin se seleccionaron aleatoriamente 10 ratas (bloques), en cada una de las cuales se utilizaron las tres técnicas para detectar yodo (Tabla 10).

*Hipótesis:*

*Hipótesis nula ( $H_0$ ): Las tres técnicas son igualmente efectivas para detectar yodo.*

*Hipótesis alternativa ( $H_A$ ): Al menos dos técnicas son diferentes.*

*Procedimiento:*

Tabla 10. Datos de yodos radioactivos (1 si se detectaba yodo y 0 si no se detectaba).

(Bloque)	I	II	III	Total por bloque ( $B_i$ )
1	1	1	0	2
2	1	1	1	3
3	0	1	1	2
4	0	1	0	1
5	0	1	0	1
6	1	0	0	1
7	1	0	0	1
8	0	0	0	0
9	1	1	1	3
10	1	0	0	2
Totales/tratamiento( $T_j$ )	6	7	3	Total general( $T_{..}$ ) = 16

$$\sum_{j=1}^3 T_j^2 = 6^2 + 7^2 + 3^2 = 94$$

$$\sum_{i=1}^3 B_i^2 = 2^2 + 3^2 + 2^2 + \dots + 2^2 = 34$$

$$K = 3$$

$$r = 10$$

$$T_{..} = 16$$

Sustituyendo estos valores en el estadístico de prueba  $Q_0$ , tenemos:

$$Q_o = \frac{k(k-1)\sum_{j=1}^k T_{.j}^2 - (k-1)(T_{..})^2}{k(T_{..}) - \sum_{i=1}^r B_i^2}$$

$$Q_o = \frac{3(2)(94) - (2)(16)^2}{3(16) - 34} = \frac{564 - 512}{14}$$

$$Q_o = 3.75$$

Utilizando la tabla, se busca el valor de  $\chi^2$  con  $gl = 2$ , y  $\alpha = 0.05$  ( $\chi^2_{2,0.05}=5.9$ ), como  $Q_o = 3.71$  es menor que  $\chi^2_{2,0.05} = 5.9$  no se rechaza  $H_o$ , y se concluye que las técnicas para detectar yodo son igualmente efectivas.

### Prueba de Friedman

La diferencia más notable de esta prueba con la de Cochran, es que esta se basa en los rangos y por lo tanto requiere que las muestras igualadas estén, por lo menos en escala ordinal. No se conoce mucho sobre la potencia de esta prueba, sin embargo, Friedman (1937) ha realizado un estudio empírico comparando esta con la prueba paramétrica de F, obteniendo muy buenos resultados, siendo imposible determinar cual de las dos pruebas es más eficiente. Esto sugiere que esta prueba se debe utilizar preferentemente a la de Cochran.

Es por esta razón que a esta prueba se le considera como un análisis de varianza no paramétrico para un diseño experimental en bloques. Por lo tanto hay que cumplir con dos suposiciones:

- 1) Se tiene k muestras relacionadas.
- 2) La escala de medición de la variable a probar está al menos en escala ordinal.

Los datos se colocan en una tabla de 2 clasificaciones con N hileras y K columnas; donde las hileras representan cada observación y las columnas los tratamientos, luego a los datos de cada hilera se les asigna un rango que va desde 1-k.

$H_o$ : Las k poblaciones (Tratamientos) son iguales, es decir, no difieren significativamente.

$H_A$ : Al menos un par de poblaciones es diferente.

La prueba de Friedman determina la probabilidad de que las diferentes columnas de rangos (muestras) proceden de la misma población. El estadístico a probar es T.

$$T_o = \frac{12}{rk(k+1)} \sum_{j=1}^k (R_j)^2 - 3r(k+1) \tag{19}$$

Rj = Suma de rangos por tratamiento o columna

k = número de tratamientos o columnas

N = r = número de hileras

La  $T_o$  se puede distribuir de acuerdo a 3 casos:

- 1) Cuando  $k = 3$  y  $r \leq 9$ : su distribución se busca en una tabla de T para  $k = 3$
- 2) Cuando  $k = 4$  y  $r \leq 4$ : su distribución se busca en una tabla de T para  $k = 4$
- 3) Cuando k o r son grandes: su distribución se busca en una tabla de  $\chi^2$  con  $gl = k-1$

**Ejemplo 11.** Se desea conocer si existe diferencia significativa en el tiempo de coagulación de la sangre, entre cuatro venenos diferentes de serpiente. Para llevar a cabo el experimento se eligieron cuatro camadas ( $k = 4 = 4$  bloques) de cuatro ratones ( $r = 4$ ) cada una, obteniéndose los siguientes resultados en minutos (Tabla 11).

$H_o$ : No hay diferencias significativas en tiempo de coagulación entre 4 venenos.

$H_A$ : Sí existen diferencias al menos en dos venenos.

Tabla 11. Resultados de la coagulación de sangre bajo 4 tratamientos.

Bloque	Tratamiento			
Camada	Veneno A	Veneno B	Veneno C	Veneno D
1	4.2	3.8	4.9	4.0
2	4.3	4.3	5.1	3.6
3	4.2	4.1	5.0	3.9
4	4.9	3.9	4.9	4.0

Procedimiento:

1) Dentro de cada bloque se ordenan las observaciones de menor a mayor y se asignan rangos. Lo que entonces se obtiene, es una tabla donde se sustituyen las observaciones por sus rangos (Tabla 11a).

Tabla 11a. Arreglo de las observaciones por el rango.

Camada	Veneno A	Veneno B	Veneno C	Veneno D
1	3.0	1.0	4.0	2
2	2.5	2.5	4.0	1
3	3.0	2.0	4.0	1
4	3.5	1.0	3.5	2
Suma de rangos por Tratamiento $R_i$	$R_1$	$R_2$	$R_3$	$R_4$

Tenemos que el estadístico de prueba  $T_o$  es:

$$T_o = \frac{12}{(4)(4)(5)} (12^2 + 6.5^2 + 15.5^2 + 6^2) - 3(4)(5)$$

$$T_o = (0.15)(462.5) - 60 = 9.375$$

El valor crítico de T se busca en tablas con  $k = 4$ ,  $r = 4$  y  $\alpha = 0.05$ , (se utiliza el valor más cercano a éste) donde  $T_{4, 4, 0.05} = 7.5$ . Como  $T_o = 9.37$  fue mayor que  $T_{4, 4, 0.05} = 7.5$  rechazamos  $H_o$  y concluimos que existen diferencias significativas en el tiempo de coagulación de la sangre entre los cuatro venenos, es decir tienen efectos diferentes.

### Correlación de rango de Spearman

Se puede usar un método no paramétrico para medir la correlación de dos variables X e Y. Este procedimiento es utilizado frecuentemente y resulta interesante por su sencillez de los cálculos que implica. En esta correlación de rango de Spearman en el lugar de usar valores precisos de las variables, o cuando su precisión no es alcanzable, a los datos se les puede asignar un rango de 1 a N ordenándolos por su tamaño, importancia, etc. Cuando se ha asignado un rango a X e Y, éste coeficiente de correlación de rango de Spearman viene dado por  $r_s$  y se calcula mediante la fórmula siguiente:

$$r_s = 1 - \frac{6 \sum d^2 i}{n(n^2 - 1)} \tag{20}$$

Las hipótesis que se prueban con mayor frecuencia son:

- a)  $H_o$ : X y Y son mutuamente independientes.

$H_A$ : X y Y no son mutuamente independientes.

b)  $H_0$ : X y Y son mutuamente independientes.  
 $H_A$ : existe una tendencia a formar parejas entre los valores grandes de X y Y.

c)  $H_0$ : X y Y son mutuamente independientes.  
 $H_A$ : existe una tendencia a formar parejas de los valores grandes de X y valores pequeños de Y.

Las hipótesis especificadas en el inciso (a) Conducen a una prueba bilateral y se utiliza cuando se desea descubrir cualquier desviación de la independencia, mientras que los caso de (b) y (c); Se utiliza respectivamente cuando se desea saber si es posible concluir que las variaciones están directamente o indirectamente relacionadas.

Si  $n$  esta entre 4 y 30, se compara con los valores críticos de  $r_s$  de tablas; pero si los valores de  $n$  son mayores de 30 se calcula mediante la fórmula:

$$Z = r_s \sqrt{n-1} \tag{21}$$

Y se compara con los valores de Z de tablas.

En la prueba bilateral si:

$$H_0: r_s < r_s^*$$

$$H_A: r_s > r_s^*$$

$$H_0: r_s > -r_s^*$$

$$H_A: r_s < -r_s^*$$

En la prueba unilateral si:

$$H_0: r_s < r_s^*$$

$$H: r_s > r_s^*$$

$$H_0: r_s > -r_s^*$$

$$H_A: r_s < -r_s^*$$

Si el número de valores numéricamente iguales, es grande, se puede utilizar la corrección por valores iguales dado por Siegel y Castellan. El factor de corrección, para los valores iguales es el siguiente:  $T = \frac{t^3 - t}{12}$ , Donde, t es el número de observaciones para un rango particular. Cuando se utiliza el factor de corrección  $r_s$ , se calcula mediante la fórmula siguiente:

$$r_s = \frac{\sum X^2 + \sum Y^2 - \sum d^2 i}{2\sqrt{\sum X^2 \sum y^2}} \tag{22}$$

$\sum X^2 = \frac{n^2 - n}{12} - \sum T_x$	$T_x$ =La suma de los valores de T para los diversos rangos de igual valor numérico igual en X.
$\sum Y^2 = \frac{n^3 - n}{12} - \sum T_y$	$T_y$ =La suma de los valores de T para los diversos rangos de igual valor numérico igual en Y.

Siegel y Castellan señalan que a menos que sea excesivo el número de cantidades con igual valor numérico, la corrección produce una diferencia muy pequeña en el valor de  $r_s$ . Cuando en # de valores iguales es pequeño, puede seguirse el procedimiento habitual.

**Ejemplo 12.** En un estudio de la relación entre la edad y los resultados del electroencefalograma (EEG), se obtuvieron datos de 20 personas con edades entre 20 y 60 años. Los resultados obtenidos se ilustran en la Tabla 12.

Tabla 12. La relación entre la edad y EEG de 20 individuos.

Edad y resultado del eeg de 20 individuos		
Individuo	Edad (X)	EEG (Y)
1	20	98
2	21	75
3	22	95
4	24	100
5	27	99
6	30	65
7	31	64
8	33	70
9	35	85
10	38	74
11	40	68
12	42	66
13	44	71
14	46	62
15	48	69
16	51	54
17	53	63
18	55	52
19	58	67
20	60	55

Hipótesis:

$H_o$  : El rendimiento del EEG y la edad son mutuamente independientes.

$H_A$  : Existe una tendencia del EEG de disminuir con la edad.

Regla de decisión: Se rechaza  $H_o$  si el valor calculado de  $r_s$  es menor que  $-0.3789$

La desviación (di) se calcula por:  $d_i = \text{Rango}(Y) - \text{Rango}(X)$ .

$$r_s = 1 - \frac{6 \sum d^2 i}{n (n^2 - 1)}$$

$$r_s = 1 - \frac{6 (2340)}{20 ((20)^2 - 1)}$$

$$r_s = -0.76$$

Utilizando la fórmula del Coeficiente de Correlación por Rangos de Spearman y clasificando por rango, tenemos el siguiente arreglo de datos (Tabla 12a).

Tabla 12a. Arreglo de datos por rango según Spearman.

Individuo	Rango (X)	Rango (Y)	D <sub>i</sub>	D <sub>i</sub> <sup>2</sup>
1	1	18	-17	289
2	2	15	-13	169
3	3	17	-14	196
4	4	20	-16	256
5	5	19	-14	196
6	6	7	-1	1
7	7	6	1	1
8	8	12	-4	16
9	9	16	-7	49
10	10	14	-4	16
11	11	10	1	1
12	12	8	4	16
13	13	13	0	0
14	14	4	10	100
15	15	11	4	16
16	16	2	14	196
17	17	5	12	144
18	18	1	17	289
19	19	9	10	100
20	20	3	17	289

$$\sum d^2_i = 2340$$

$r^*_{s\ 0.05\ (1),\ 20} = -0.3789$ ,  $r_s > r^*_s$  se apoya la  $H_a$  con  $P > 0.05$ . Conclusión: Se concluye que las dos variables se encuentran inversamente relacionadas.

**Prueba de rangos para 2 muestras (Mann-Whitney)**

El procedimiento no paramétrico ha propuesto para la prueba entre la “dispersión” o “variabilidad” de 2 poblaciones. La prueba *no paramétrica* es *análoga* a la prueba de “t” de 2 muestras. Originalmente el método fue propuesto por Wilcoxon (1945) y después aplicado por Mann y Whitney (1947). La prueba Mann-Whitney como muchos otros procedimientos no paramétricos se usa para los *datos arreglados en rangos*. En este caso los datos se deben organizar en rangos de mayor a menor marcando 1 para datos mayor hasta n para datos menores o viceversa, donde N (tamaño total) =  $n_1 + n_2$ . La prueba de hipótesis esta basada sobre la significancia del estadístico ( $U_c$ ) Mann-Whitney:

$$U_c = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1 \quad \text{si } R_1 < R_2 \quad (23)$$

Donde,  $U_{c\ \alpha(2),n_1,n_2}$  = valor tabulado ( $n_1 < n_2$ );  $n_1, n_2$  = número de observaciones en las muestras 1 y 2,  $R_1$  = la suma de rangos de observaciones en la muestra 1.

$$U' = n_2 n_1 + \frac{n_2(n_2 + 1)}{2} - R_2 \quad \text{si } R_1 > R_2$$

$R_2$  =  $\sum$  de rangos de observaciones en la muestra 2, por lo tanto, donde,  $U = n_1 n_2 - U$ , si  $U$  ó  $U' > U_{\alpha(2),n_1,n_2}$  en tabla correspondiente al modelo, se rechaza la hipótesis nula.

En resumen, se puede notar que después de ordenar los datos en rango de 2 muestras se calcula  $U$  ó  $U'$ . Cuando se necesita calcular ambos rangos, se usa:  $R_1 + R_2 = \frac{N(N + 1)}{2}$ .

**Ejemplo 13.** Verificar la hipótesis nula en la cual no hay diferencia entre las estaturas de estudiantes de ambos sexos (Tabla 13).

$H_0$ : Las estaturas de los alumnos masculinos y femininos son iguales.  
 $H_A$ : Las estaturas de los alumnos masculinos y femininos no son iguales.

Tabla 13. Datos correspondiente a las estaturas en centímetros.

Estaura masculina	Rango masculino	Estatura femenina	Rango femenino
193	1	175	7
188	2	173	8
185	3	168	10
183	4	165	11
180	5	163	12
178	6		
170	9		
$n_1 = 7$	$R_1 = 30$	$n_2 = 5$	$R_2 = 48$

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

$$U = (7)(8) + \frac{(7)(8)}{2} - 30$$

$$U = 35 + 28 - 30, U = 33, U_{0.05(2), 7,5} = U_{0.05(2), 5,7} = 30$$

$U_c (33) > U_{\text{tabulado}}(30)$  se rechaza la  $H_0$

$$0.01 < P(U \geq 33) < 0.02$$

**Ejemplo 14.** Para acelerar la tasa de crecimiento mensual de plantas, a un grupo de ellas se aplicó el fertilizante a base de nitrógeno. Los resultados se demuestran en la Tabla 14.

Tabla 14. Datos obtenido sobre el crecimiento (en centímetros) de plantas

Con nitrógeno	Rango 1	Sin nitrógeno	Rango 2
44	9	32	3.5
48	12	40	7
36	6	44	9
32	3.5	44	9
51	13	34	5
45	11	30	2
54	14	26	1
56	15		
$n_1 = 8$	$R_1 = 83.5$	$n_2 = 7$	$R_2 = 36.5$

$H_0$ : El crecimiento de plantas con el tratamiento no es mayor que las plantas sin tratamiento.

$H_A$ : El crecimiento de plantas con aplicación del nitrógeno es mayor que del sin nitrógeno.

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

$$U' = (7)(8) + \frac{(7)(8)}{2} - 36.5$$

$$U' = 56 + 28 - 36.5$$

$$U' = 47.5$$

$$U_{0.05(1), 8, 7} = U_{0.05(1), 7, 8} = 43$$

$$U_c (47.5) > U_{tab.} (43), \text{ se rechaza } H_0$$

$$0.01 < P < 0.025$$

**Prueba de Kruskal-Wallis**

Si se colectan datos an base a un diseño completamente aleatorio con más de 2 muestras (k>2), sería posible probar la diferencia entre los grupos, de forma no-paramétrica. Para esto, se usa el modelo de Kruskal-Wallis (Kruskal & Wallis, 1952), frecuentemente, denominado como análisis de varianza por rangos. Esta prueba tiene el poder estadístico de 95% de ANOVA de una via (Andrews, 1954). La prueba de Kruskal-Wallis se aplica cuando las k muestras no proceden de una población normal y/o cuando las varianzas de k poblaciones son heterogéneas. Hay que notar que en los modelos no-paramétricos no se usan los parámetros poblacionales en el enunciado de las hipótesis, ni tampoco, los parámetros ni las estimaciones muestrales se emplean en los cálculos de las pruebas. La estadística de este modelo se calcula de manera siguiente.

$$H = \{12 / [N(N+1)]\} (\sum R_i^2 / n_i) - 3(N+1) \tag{24}$$

Donde,  $n_i$  es el número de observaciones por grupo  $i$ ,  $N = \sum n_i$  (número total de observaciones en  $k$  grupos),  $R_i$  es la suma de rangos de  $n_i$  observaciones en grupo  $i$ . El procedimiento de conducir esta prueba se demuestra en el ejemplo 15 y la Tabla 15.

**Ejemplo 15.** Se desea saber la abundancia de las personas en tres localidades y para este fin, un investigador realiza cinco colecciones en cada uno de tres localidades.

$H_0$ : La abundancia de las personas es similar en tres localidades.

$H_a$ : La abundancia de las personas no es similar en tres localidades.

Los datos se demuestran en la siguiente tabla con valores de los rangos en los paréntesis.  
 $\alpha = 0.05$

Tabla 15. Promedio de personas por km<sup>2</sup> en 5 muestras de cada uno de tres localidades.

Localidad A	Localidad B	Localidad C
14.0 (15)	8.4 (11)	6.9 (8)
12.1 (14)	5.1 (2)	7.3 (9)
9.6 (12)	5.5 (4)	5.8 (5)
8.2 (10)	6.6 (7)	4.1 (1)
10.2 (13)	6.3 (6)	5.4 (3)
$n_1 = 5$	$n_2 = 5$	$n_3 = 5$
$R_1 = 64$	$R_2 = 30$	$R_3 = 26$
$N = n_1 + n_2 + n_3 = 5 + 5 + 5 = 15$		
$H = \{12 / [N(N+1)]\} (\sum R_i^2 / n_i) - 3(N+1)$ $\{12 / [15(16)]\} [(64^2/5) + (30^2/5) + (26^2/5)] - 3(15-1)$ $= 8.72$ $H_{0.05, 5, 5, 5} = 5.78$ <b>Por tanto, se rechaza la <math>H_0</math></b> $0.005 < P < 0.01$		

## Referencias

- Badii, M.H. & J. Castillo. 2009<sub>a</sub>. Muestreo Estadístico: Conceptos y Aplicaciones. UANL, Monterrey, 225 pp.
- Badii, M.H. & J. Castillo. 2009<sub>b</sub>. Distribuciones probabilísticas de uso común. Daena, 4(1): 149-178.
- Badii, M.H., A. Guillen & L.A. Araiza. 2010. Estimaciones estadísticas: Un acercamiento analítico. Daena. 5(1): 237-255.
- Badii, M.H., A. Guillen, E. Cerna & J. Landeros. 2011<sub>a</sub>. Dispersión espacial: El requisito esencial para el muestreo. Daena International J. of Good Conscience. 6(1): 40-71.
- Badii, M.H., A. Guillen, E. Cerna & J. Valenzuela. 2011<sub>b</sub>. Nociones introductorias de muestreo estadístico. Daena International J. of Good Conscience. 6(1): 89-105.
- Andrews, F. C. 1954. Asymptotic behavior of some rank tests for analysis of variance. Ann. Math. Statist. 25: 724-735.
- Brown, G. W., & A. M. Mood. 1951. Median tests for linear hypotheses. In: Proceedings of the Second Berkely Symposium on Mathematica Statistics and Probability. Berkeley and Los Angeles, 1951, J. Neyman (ed.). pp. 159-166.
- Buckle, N., C. Kraft, & C. van Eeden. 1969. An approximation to the Wilcoxon-Mann-Whitney Distribution. J. Amer. Statist. Assoc. 64: 591-599.
- Conover, W. J. 1980. Practical Nonparametric Statistics. 2<sup>nd</sup> ed. John Wiley, New York.
- Daniel, W. W. 1978. Applied Nonparametric Statistics. Houghton Mifflin Co., Boston Massachusetts.
- Gibbons, J. D. 1978. Nonparametric Methods for Quantitative Analysis. Holt. Rinehart, and Winston, New York.
- Harter, H. L. 1970. Order Statistics and their use in Testing and Estimation. US Government Printing Office. Washington, D.C.
- Harter, H. L., H. J. Khamis, & R. E Lamb. 1984. Modified Kolmogorov-Smirnov tests for goodness of fit. Communic. Statist.-Simula. Computa. 13: 293-323.
- Hodges, J. L. 1990. Improved significance probabilities of the Wilcoxon test. J. Educ.Statist. 15: 249-265.
- Hollander, M. 1970. A distribution-free test for parallelism. J. Amer. Statist. Assoc. 65: 287-294.
- Iman, R. L., & J. M. Davenport. 1976. New approximations to the exact distribution of the Kruskal-Wallis test statistic. Communic. Statist.-Theor Meth. A5: 1335-1348.
- Iman, R. L., & J. M. Davenport. 1980. Approximation of the critical region of the Friedman statistic. Communic. Statist.-Theor Meth. A9: 571-595.
- Kruskal, W. H., & W. A. Wallis. 1952. Use of ranks in one criterion analysis of variance. J. Amer. Statist. Assoc. 47: 583-621.
- Martín Andrés, A. 1991. Una revisión de metodos clásicos no paramétricos para comparar dos proporciones de medias muestrales. Communic. Statist.-Simuta. Computa. 20: 551-583.
- Milton, R. C., 1964. An extended table of critical values for the Mann-Whitney two sample statistic. J. Amer. Statist. Assoc. 59: 925-934.
- Skillings, J. H. & G. A. Mack. 1981. On the use of a Friedman-type statistic in balanced and unbalanced block designs. Technometrics 23: 171-177.
- Theil, H. 1950. A rank-Invariant method of linear and polynomial regression analysis. Inederl. Akad. Wetensch. Proc. Ser. A. 53: 386-392.
- Thomas, G. E. 1989. A note on correcting for ties with Spearman's  $\rho$ . J. Statist. Computa. Simula. 31: 37-40.