

coeficiente de concordancia – coefficient of concordance

Authored by
memjavad

November 17, 2025

RECOMMENDED CITATION

memjavad (2025). *coeficiente de concordancia – coefficient of concordance*. Spanish Psychological Databases. Retrieved from <https://spanish.arabpsychology.com/?p=4933>

Coeficiente de Concordancia W de Kendall

Primary Disciplinary Field(s): [Estadística no paramétrica](#), Psicometría, Investigación Social, [Teoría de la Decisión](#)

1. Definición Central y Propósito

El **Coeficiente de Concordancia W de Kendall** es una medida estadística no paramétrica diseñada para evaluar el grado de acuerdo o asociación entre k conjuntos de clasificaciones u ordenaciones (rangos) de N objetos o sujetos. A diferencia de las medidas de correlación bivariada, como el coeficiente de [Spearman](#), que solo comparan dos conjuntos de clasificaciones, el Coeficiente W es particularmente útil cuando se tienen tres o más jueces, evaluadores o criterios que rankean el mismo conjunto de elementos. Su propósito fundamental radica en determinar si las clasificaciones proporcionadas por los diferentes evaluadores son significativamente similares, es decir, si existe un **consenso** general en la forma en que se ordenan los elementos.

Formalmente, el valor de W varía entre 0 y 1. Un valor de 1 indica una **concordancia perfecta**, lo que significa que todos los evaluadores han asignado exactamente el mismo rango a cada uno de los N objetos. Por el contrario, un valor de 0 sugiere una ausencia total de acuerdo, indicando que la clasificación de los objetos es esencialmente aleatoria entre los evaluadores. Es importante destacar que W no es una medida de la magnitud de la correlación en el mismo sentido que el coeficiente de Pearson, sino más bien una medida de la proporción de la varianza máxima posible que es explicada por la suma de los rangos asignados. De esta manera, proporciona una herramienta robusta para la validación de juicios expertos o la consistencia interna en estudios de evaluación multicriterio.

El Coeficiente W opera bajo la premisa de que los datos son de naturaleza ordinal, lo que lo sitúa firmemente dentro del ámbito de la **estadística no paramétrica**. Esta característica lo hace invaluable en campos como la psicología, la investigación de mercados y la investigación social, donde a menudo las escalas de medición utilizadas (como las escalas de Likert o las clasificaciones de preferencia) solo garantizan propiedades de orden, pero no de intervalo o razón. Al no requerir supuestos sobre la distribución subyacente de los datos (como la normalidad), Kendall's W ofrece una solución flexible y aplicable a una amplia gama de escenarios evaluativos.

2. Origen Histórico y Desarrollo

El desarrollo del Coeficiente de Concordancia se atribuye al estadístico británico [Maurice Kendall](#), quien lo introdujo a mediados del siglo XX. Su creación surgió de la necesidad de extender y generalizar los métodos de correlación de rangos existentes, que hasta entonces se habían

limitado principalmente a la comparación de solo dos conjuntos de clasificaciones (como el coeficiente tau de Kendall o el rho de Spearman). Kendall reconoció que en muchas aplicaciones prácticas--particularmente en estudios donde múltiples jueces subjetivos intervenían--era crucial medir la coherencia global de estos juicios de manera simultánea.

Históricamente, la estadística de rangos floreció como una respuesta a las limitaciones de los métodos paramétricos en el manejo de datos no normales o cuando los datos eran inherentemente ordinales. La contribución de Kendall fue crucial al proporcionar una métrica única y fácilmente interpretable que condensaba la información de k^2 comparaciones en un solo valor. Antes de W , abordar la concordancia entre múltiples jueces requería calcular $\binom{k}{2}$ coeficientes de correlación bivariada y luego promediarlos, un proceso tedioso y que no ofrecía una prueba estadística clara sobre la significancia del acuerdo colectivo.

La fórmula y la prueba de significancia asociadas con W se basan en la distribución de chi-cuadrado (χ^2), lo que permitió a los investigadores no solo cuantificar el grado de acuerdo, sino también determinar la probabilidad de que dicho acuerdo observado pudiera haber ocurrido simplemente por azar. Esta dualidad--cuantificación y prueba de hipótesis--consolidó al Coeficiente W como una herramienta estándar en el análisis de datos multivariados ordinales, especialmente en el contexto de la **evaluación inter-jueces** (inter-rater reliability) cuando el número de jueces es superior a dos.

3. Fundamentos Matemáticos: La Fórmula W

La esencia matemática del Coeficiente W se centra en la medición de la variación de las sumas de los rangos asignados a cada objeto. Si N es el número de objetos clasificados y k es el número de evaluadores, cada evaluador asigna rangos de 1 a N . La estadística clave es la suma de los rangos para cada objeto j , denotada como R_j . Si hay un acuerdo perfecto, todos los objetos tendrán sumas de rangos muy distintas (el objeto clasificado primero por todos tendrá la suma mínima, y el último, la suma máxima). Si no hay acuerdo, las sumas de rangos serán muy similares entre sí, acercándose al valor medio esperado.

El cálculo de W se basa en la variabilidad de estas sumas de rangos. Se calcula la cantidad S , que es la **suma de los cuadrados de las desviaciones** de cada R_j respecto al valor medio de la suma de rangos esperada, \bar{R} . El valor medio esperado de la suma de rangos es $\bar{R} = \frac{k(N+1)}{2}$. El coeficiente W se obtiene entonces normalizando S dividiéndolo por el valor máximo posible de S , denotado como S_{\max} , que ocurriría si la concordancia fuera perfecta. La fórmula general es $W = \frac{S}{S_{\max}}$.

El valor de S_{\max} se calcula como $S_{\max} = \frac{k^2 (N^3 - N)}{12}$. Por lo tanto, el Coeficiente de Concordancia W de Kendall se define matemáticamente como:

$$W = \frac{12S}{k^2 (N^3 - N)}$$

Donde $S = \sum_{j=1}^N (R_j - \bar{R})^2$. Esta normalización asegura que W siempre se encuentre en el intervalo $[0, 1]$. Un aspecto crucial de la aplicación práctica de W es la necesidad de ajustar la fórmula cuando existen **empates** (*ties*) en las clasificaciones, aunque el ajuste es complejo y a menudo se maneja mediante software estadístico, que reduce el denominador (S_{\max}) para reflejar la menor variabilidad posible debido a los empates.

4. Interpretación y Prueba de Significancia

La interpretación del Coeficiente W es directa: cuanto más se acerca W a 1, mayor es el grado de **concordancia** entre los evaluadores. Un W cercano a 0 indica que las clasificaciones son esencialmente independientes y aleatorias. Sin embargo, la simple magnitud de W no es suficiente; es crucial determinar si el acuerdo observado es estadísticamente significativo, es decir, si es poco probable que haya ocurrido por casualidad.

Para probar la significancia, W se relaciona con la distribución de chi-cuadrado (χ^2). Bajo la hipótesis nula (H_0) de que no hay acuerdo real ($W=0$), la estadística de prueba se calcula como: $\chi^2 = k(N-1)W$. Esta estadística de prueba se distribuye aproximadamente como una chi-cuadrado con $N-1$ grados de libertad. Si el valor p asociado a esta χ^2 es menor que el nivel de significancia preestablecido (típicamente 0.05), se rechaza la hipótesis nula, concluyendo que existe un acuerdo significativo entre los k evaluadores.

En términos de directrices prácticas, si bien no existen umbrales universales y estrictos, un W superior a 0.70 generalmente se considera indicativo de una **concordancia fuerte** y aceptable para propósitos de investigación o validación de criterios. Valores entre 0.50 y 0.70 sugieren una concordancia moderada. Es vital recordar que la interpretación de W debe ir acompañada del análisis de significancia estadística, ya que un valor moderado de W puede ser significativo si el número de objetos (N) es grande. Además, el investigador debe considerar el contexto del estudio; en áreas subjetivas (como la evaluación artística), un W más bajo podría ser aceptable que en áreas objetivas (como la calibración instrumental).

5. Relación con la Correlación de Rangos de Spearman

Una característica fundamental del Coeficiente de Concordancia W es su estrecha relación matemática con el **Coeficiente de Correlación de Rangos de Spearman** (ρ). De hecho, W puede interpretarse como una medida promedio de la correlación de rangos entre todos los pares posibles de evaluadores. Si se calcula el coeficiente de Spearman para cada par de evaluadores (hay $\binom{k}{2}$ pares) y luego se promedian estos valores, existe una relación lineal directa entre este promedio y W .

La relación matemática que vincula W con el promedio de los coeficientes de Spearman ($\bar{\rho}$) es la siguiente:

$$W = \frac{\bar{\rho} + 1}{k - 1}$$

Esta fórmula demuestra que W es esencialmente una transformación del promedio de las correlaciones de Spearman. Si el promedio de las correlaciones de Spearman es 1 (acuerdo perfecto en todos los pares), entonces $W = \frac{1 + 1}{k - 1} = 1$. Si el promedio de las correlaciones es -1 (desacuerdo perfecto en todos los pares, aunque esto es raro en la práctica de múltiples jueces), W tendería a 0 o incluso a ser negativo en la fórmula, pero en la práctica W está limitado a ser no negativo ya que mide la concordancia y no la discordancia.

Esta equivalencia es crucial porque ayuda a conceptualizar W . Mientras que Spearman ρ cuantifica la fuerza y dirección de la relación monótona entre dos variables, W sintetiza la consistencia global. Si W es alto, implica que la correlación promedio entre cualquier par de jueces es también alta y positiva. Esta conexión proporciona una base teórica sólida y permite a los investigadores que están familiarizados con la correlación bivariada extender su comprensión al análisis multivariado de concordancia.

6. Aplicaciones Prácticas y Contextos de Uso

El Coeficiente W de Kendall es una herramienta excepcionalmente versátil y se utiliza en una amplia variedad de disciplinas que requieren la agregación o validación de juicios ordinales. En la **Psicometría** y la **Psicología Clínica**, se emplea frecuentemente para evaluar la fiabilidad inter-juez (inter-rater reliability) cuando varios terapeutas, observadores o diagnosticadores clasifican síntomas o comportamientos. Por ejemplo, se puede usar para medir la concordancia entre tres psiquiatras que clasifican la gravedad de una condición en una escala ordinal.

En la **Investigación de Mercados** y la **Investigación Social**, W es indispensable para analizar la consistencia de las preferencias. Si un grupo focal de consumidores clasifica varios productos o atributos según su atractivo, W determina si existe una preferencia colectiva significativa. De manera similar, en el campo de la **Gestión de Calidad** y la **Ingeniería de Sistemas**, se utiliza para evaluar el consenso entre expertos que priorizan fallas, riesgos o requisitos de diseño.

Finalmente, su utilidad se extiende a los **concursos** y las **evaluaciones educativas**. Cuando múltiples jueces evalúan el desempeño de competidores (por ejemplo, en gimnasia, patinaje artístico o concursos de belleza) mediante rangos, W proporciona una medida objetiva de si los jueces están utilizando criterios consistentes o si sus clasificaciones son idiosincrásicas. En todos estos contextos, el valor principal de W reside en su capacidad para transformar una serie compleja de clasificaciones individuales en una métrica única y estadísticamente interpretable de acuerdo colectivo.

7. Limitaciones y Consideraciones Críticas

A pesar de su utilidad, el Coeficiente de Concordancia W de Kendall presenta ciertas limitaciones

que deben ser consideradas por el investigador. Una crítica fundamental es que W solo mide la **magnitud del acuerdo general**, pero no ofrece información sobre la naturaleza específica del desacuerdo. Si W es bajo, el investigador sabe que hay poca concordancia, pero no sabrá qué jueces están en desacuerdo ni en qué elementos específicos se concentran las discrepancias. Para obtener esa información, se requeriría un análisis posterior de las correlaciones bivariadas o el uso de análisis de conglomerados.

Otra limitación importante se refiere a la **prueba de significancia** basada en la aproximación de chi-cuadrado. Esta aproximación es menos precisa cuando el número de objetos (N) es pequeño (generalmente, menor a 5 o 7) o cuando hay un número muy elevado de empates en las clasificaciones. En tales casos, el valor p derivado puede ser inexacto, llevando potencialmente a conclusiones erróneas sobre la significancia del acuerdo. Para muestras pequeñas, se recomienda consultar tablas exactas o utilizar métodos de remuestreo (bootstrapping) si el software lo permite.

Además, W asume que la variable subyacente que se clasifica es continua, aunque solo se observan los rangos ordinales. Si los datos son verdaderamente categóricos nominales, W no es apropiado. Si los datos son de intervalo o razón, el Coeficiente de Correlación Intraclase (ICC) es a menudo preferido sobre W , ya que el ICC aprovecha la información de la magnitud de las diferencias, no solo el orden. Por lo tanto, la elección de W debe basarse en la confirmación de que la escala de medición empleada es, en efecto, **ordinal** y que el número de evaluadores es mayor a dos.

Further Reading

[Kendall's W \(Wikipedia, English\)](#)

[Estadística no paramétrica \(Wikipedia, Spanish\)](#)

[Coeficiente de correlación de rangos de Spearman \(Wikipedia, Spanish\)](#)

[Maurice Kendall \(Wikipedia, English\)](#)