

15A. FIABILIDAD O ACUERDO DE PRUEBAS DIAGNÓSTICAS

ESTUDIOS SOBRE PRUEBAS DIAGNÓSTICAS

La investigación clínica implica en ocasiones la evaluación de pruebas diagnósticas. Aunque habitualmente se utilizan diseños transversales y menos frecuentemente de cohortes, los objetivos y análisis de los estudios sobre pruebas diagnósticas son distintos a los habituales en estos diseños. El estudio de la variabilidad de cualquier medición, y una prueba diagnóstica no es más que una medición, consta de dos aspectos: la Validez o Exactitud que se verá en la Unidad 16 y la Fiabilidad o Acuerdo (Tabla 15A.1).

Tabla 15A.1. Estudios sobre pruebas diagnósticas.

	Fiabilidad (<i>reliability</i>)	Validez (<i>validity</i>)
Concepto	Grado de coincidencia de dos o más mediciones, realizadas sobre la misma muestra por uno o más observadores	Grado de coincidencia con la verdad, representada por una prueba de referencia. Capacidad de distinguir entre estados de salud
Otras denominaciones en la literatura	Acuerdo o concordancia (<i>agreement</i>), reproducibilidad (<i>reproducibility</i>), consistencia (<i>consistency</i>).	Exactitud o precisión (<i>accuracy</i>), eficacia (<i>efficacy</i>), rendimiento diagnóstico (<i>diagnostic accuracy</i>)
Pruebas utilizadas	Índice kappa Índice kappa ponderado Coeficiente de correlación intraclase (CCI)	Sensibilidad y Especificidad Valor Predictivo Positivo y Negativo Curva ROC

FIABILIDAD O ACUERDO DE PRUEBAS DIAGNÓSTICAS

La **Fiabilidad** de una prueba diagnóstica es el grado de coincidencia de dos o más mediciones, realizadas sobre la misma muestra por uno o más observadores. Es el paso inicial y previo en la valoración de una prueba diagnóstica. Es lo primero que se debe hacer cuando no existe un grado de coincidencia adecuado, y es lo único que se puede valorar cuando no existe una prueba de referencia que permita estudiar la validez.

Las pruebas utilizadas para analizar la Fiabilidad dependen de la escala de medida de la prueba diagnóstica. El *índice Kappa* se utiliza para variables categóricas nominales, el *índice kappa ponderado* para variables categóricas ordinales y *Coeficiente de Correlación Intraclase* (CCI) para variables cuantitativas (Tabla 15A.2).

Tabla 15A.2. Pruebas utilizadas para la Fiabilidad de pruebas diagnósticas.

VARIABLES	Nominal	Ordinal	Cuantitativa
Nominal	Índice Kappa		
Ordinal		Índice Kappa ponderado	
Cuantitativa			Coeficiente de Correlación Intraclase (CCI)

Fiabilidad de variables nominales: Índice Kappa

Se pretende evaluar el grado acuerdo o concordancia entre dos clínicos (*A* y *B*) en la identificación de un síntoma que podría ser, por ejemplo la DPN (disnea paroxística nocturna). En el archivo *MedidasAyB.sav* se encuentran los datos de las variables nominales binarias *MedidaA* y *MedidaB* resultado de la identificación del síntoma por los clínicos A y B en 100 sujetos. En la Tabla 15A.3 se ofrece el resultado con *MedidaA* en filas y *MedidaB* en columnas. En cada casilla se muestran los efectivos observados, los esperados y el porcentaje sobre el total.

- El **Acuerdo observado** (A_o) también llamado *Acuerdo específico global* o *Acuerdo porcentual* es la medida de acuerdo más sencilla. Se valora la coincidencia o acuerdo en los resultados positivos (*acuerdo específico positivo*) y negativos (*acuerdo específico negativo*). A partir de la Tabla 15A.3 se obtiene sumando los porcentajes totales de las casillas concordantes (en rojo): $(50,0\%+35,0\%)=85\%$. También se puede obtener sumando los efectivos observados en las casillas concordantes y dividirlo por el número total de sujetos (en negrita): $(50+35)/100=0,85$. Es decir, el cociente entre la suma de pares concordantes y el total de sujetos. Su valor oscila entre 0 (discordancia total) y 1 (**Acuerdo máximo** o total). Tiene el inconveniente de no tener en cuenta los acuerdos que se producen simplemente por azar.
- El **Acuerdo esperado** (A_e) por azar, en el caso de independencia perfecta entre las mediciones, se vería cuando la tabla de contingencia mostrara los valores esperados bajo el supuesto de que no existe relación entre las variables A y B. Se obtiene sumando las frecuencias esperadas de las casillas concordantes y dividiéndolo por el total de casos (en azul): $(33,0+18,0)/100,0=0,51$.

Tabla 15A.3. Tabla de contingencia *MedidaA - MedidaB*.

			MedidaB		Total
			Negativo	Positivo	
MedidaA	Positivo	Observado	5	50	55
		Esperado	22,0	33,0	55,0%
		% Total	5,0%	50,0%	55,0%
	Negativo	Observado	35	10	45
		Esperado	18,0	27,0	45,0
		% Total	35,0%	10,0%	45,0%
Total	Observado	40	60	100	
	Esperado	40,0	60,0	100,0	
	% Total	40,0%	60,0%	100,0%	

En la Figura 15A.1 se representan el acuerdo máximo que sería 1,0, el acuerdo observado, que en este caso es 0,85 y el acuerdo esperado por azar que es 0,51.

- El **acuerdo real** es el observado menos el esperado por azar. En este caso sería: $A_o - A_e = 0,85 - 0,51 = 0,34$.
- El **acuerdo potencial** sería el acuerdo máximo menos el debido al azar. Sería: $1 - A_e = 1 - 0,51 = 0,49$. Es la máxima concordancia real que se puede obtener cuando hay acuerdo perfecto. Es el acuerdo que queda por explicar después de quitar la influencia del azar.
- El **índice kappa** de Cohen es el cociente entre el acuerdo real y el acuerdo potencial. Es la proporción de acuerdo real respecto al potencial. Soluciona el inconveniente del acuerdo esperado teniendo en cuenta la coincidencia esperable por azar. Incluye sólo los acuerdos que supuestamente no se deben al azar.

$$Kappa = (A_o - A_e) / (1 - A_e).$$

En este ejemplo sería $kappa = 0,34 / 0,49 = 0,694$ ($p < 0,001$). El índice kappa oscila entre -1 (existe desacuerdo sistemático) y $+1$ (existe acuerdo sistemático) pasando por 0 (acuerdo debido al azar). Sólo está disponible para las tablas “cuadradas” (tablas en las que ambas variables tienen el mismo número de categorías).

Figura 15A.1. Acuerdos (máximo, observado, esperado, real y potencial) e índice kappa.

Acuerdo máximo = 1,0	
Acuerdo observado = 0,85	
Acuerdo esperado por azar = 0,51	Acuerdo real = 0,34
	Acuerdo potencial = 0,49
Índice kappa = Acuerdo real / Acuerdo potencial = 0,34 / 0,49 = 0,694	

Existen 2 propuestas para valoración del índice kappa que se muestran en la Tabla 15A4. Son una orientación y no deben seguirse estrictamente. El índice kappa más adecuado suele variar en cada situación.

Tabla 15A.4. Valoración del índice kappa

Landis y Koch		Fleiss	
0.81 a 1.00	Excelente		
0.61 a 0.80	Bueno	0.76 a 1.00	Fuerte
0.41 a 0.60	Moderado	0.40 a 0.75	Moderado
0.21 a 0.40	Bajo	< 0.40	Pobre
< 0.20	Muy bajo		

Limitaciones del índice kappa:

- **Influencia de las asimetrías en las concordancias.** Está influido por el desequilibrio entre los resultados positivos y negativos, por lo que varía con la prevalencia de la enfermedad. Una manera que disminuir esta limitación es comunicarlo junto con los acuerdos específicos positivos y negativos. Esta influencia de la asimetría de los resultados positivos y negativos es tanto mayor cuanto mayor es el acuerdo global.
- **Influencia de las asimetrías en las discordancias.** También está influido por la asimetría de partes discordantes o sesgo y además esta influencia es tanto mayor cuanto menor sea el índice kappa.

Videotutorial 15A1Kappa1.avi

Con el archivo *MedidasAyB.sav* se muestra como se evalúa la Fiabilidad de dos mediciones nominales binarias (*MedidasA* y *MedidasB*) con el **índice kappa** a partir del cuadro *Tabla de contingencia*. Ofrece los datos de los acuerdos observados y esperados, a partir de los que se pueden calcular los acuerdos real y potencial.

Fiabilidad de variables ordinales: Índice Kappa ponderado.

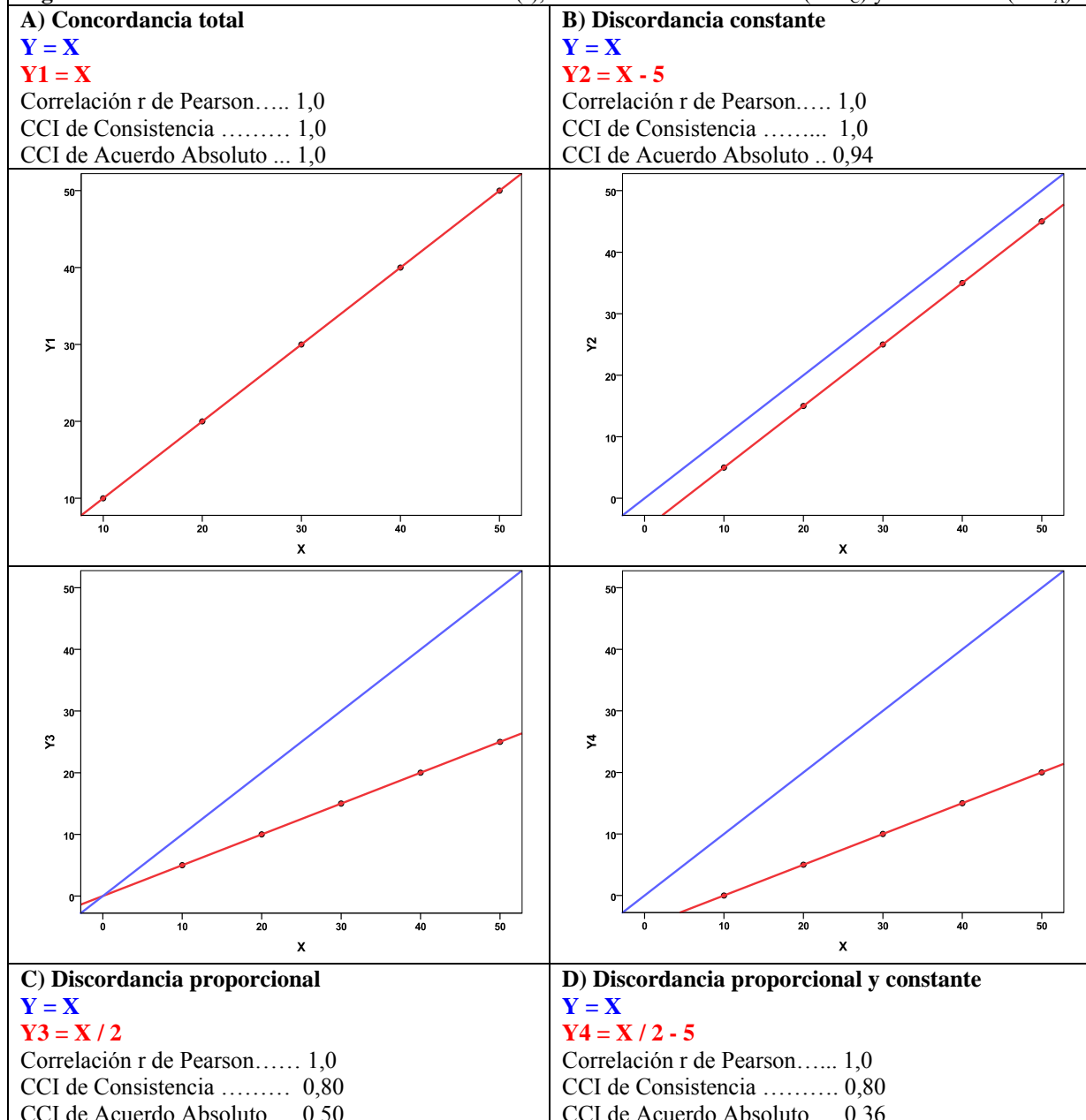
Cuando el índice kappa se aplica a variables ordinales, se asigna el mismo nivel de desacuerdo a la diferencia en un nivel que en dos. Se debe utilizar el índice **kappa ponderado** que diferencia si el nivel de desacuerdo es en un nivel o en más de un nivel. En general el índice kappa ponderado es mayor que el no ponderado.

Fiabilidad de variables cuantitativas: Coeficiente de Correlación Intraclase (CCI).

La fiabilidad o concordancia entre dos variables cuantitativas resultado de aplicar distintos instrumentos de medida o diferentes evaluadores a un mismo proceso, no debe evaluarse mediante el coeficiente de correlación r de Pearson, sino mediante el *Coeficiente de Correlación Intraclase*, CCI (*Intraclass Correlation Coefficient*, ICC). El coeficiente de correlación valora la intensidad de la asociación lineal entre dos mediciones cuantitativas, que es un concepto distinto del grado de acuerdo entre ambas mediciones. La existencia de correlación no implica la presencia de concordancia o acuerdo.

El cálculo del CCI se basa en el modelo de análisis de la varianza con medidas repetidas o intrasujeto o de dos vías: la correspondiente a los sujetos y la correspondiente a los observadores. Si se somete a varios sujetos a varios evaluadores, la variación total obtenida puede descomponerse en tres componentes: la variación debida a los sujetos, la debida a los evaluadores y la variación residual.

Figura 15A.2. Coeficientes de correlación: Pearson (r), Intraclase de Consistencia (CCI_C) y de Acuerdo (CCI_A).



Los valores del CCI oscilan entre 0 (ausencia total de concordancia) y 1 (concordancia absoluta). Su interpretación es similar a la del kappa, que aunque algo arbitraria, se tiende a aceptar las categorías propuestas por Fleiss: <0.40 es bajo o malo, entre 0.41 y 0.75 es regular/bueno o aceptable y >0.75 es muy buena o excelente.

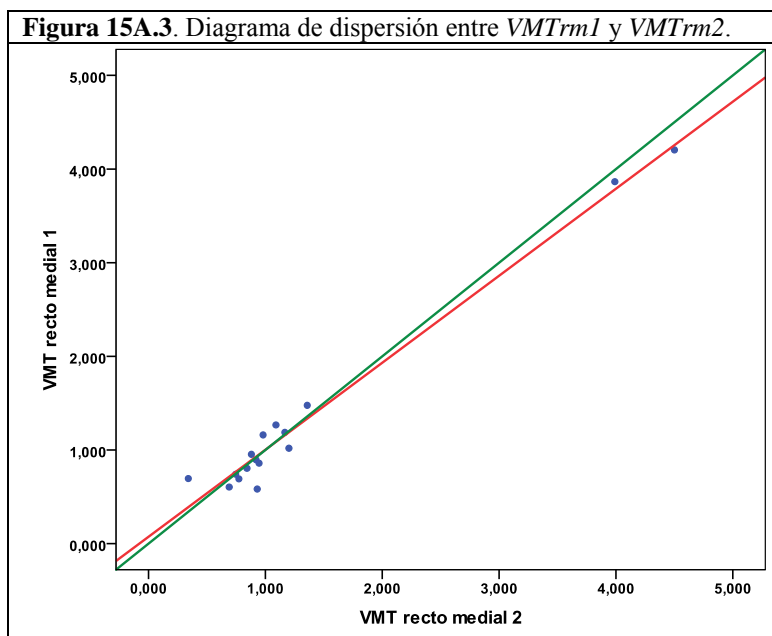
Hay dos tipos de CCI, de Consistencia y de Acuerdo. La diferencia entre ambos y con el coeficiente de correlación de Pearson se observa analizando la relación entre las variables X e Y en los 4 ejemplos de la Figura 15A.2. Cuando las valoraciones de X e Y son idénticas, los tres coeficientes valen 1.

- **CCI de Consistencia, CCI_C .** Valora consistencia o coherencia (*consistency*). No considera discrepancias las diferencias de tipo aditivo. Los ejemplos A y B de la Figura 15A.2 presentan el mismo CCI_C . En este caso el ejemplo B da un valor máximo de 1, teniendo en cuenta que los valores de Y difieren en 5 unidades de los de X. Lo mismo sucede en los ejemplos C y D. Se comporta como un índice de aditividad que valora si es aplicable la transformación $Y = \text{Constante} + X$. Al igual que el r de Pearson no detecta una diferencia constante entre dos series de observaciones. Pero a diferencia del r de Pearson si es sensible a diferencias de tipo proporcional. El r de Pearson se comporta como un índice de linealidad que valora si es aplicable la transformación $Y = \text{Constante} + \text{Pendiente } X$
- **CCI de Acuerdo, CCI_A .** Valora acuerdo absoluto (*absolute agreement*). Contempla como discrepante cualquier tipo de diferencia entre las valoraciones, ya sea constante, proporcional o de otro tipo. Cuantas más fuentes de discrepancia haya y mayor sea, mas bajo será el valor del CCI_A . Valora si es aplicable la transformación $Y = X$.

Ejemplo Oftalmopatía de Graves

Se quiere medir la variabilidad entre dos observadores (1 y 2) para la medición mediante Resonancia Magnética del Volumen Muscular Total de los músculos Recto Medial (VMTrm) de 8 pacientes con oftalmopatía de Graves. Los datos de los dos observadores (variables $VMTrm1$ y $VMTrm2$) de los 16 ojos se encuentran en el archivo *Graves.sav*.

En la Figura 15A.3 se muestra el diagrama de dispersión de ambas variables. La línea oblicua marcada en verde representa la ecuación $VMTrm1 = VMTrm2$, con fiabilidad o acuerdo máximo (el r de Pearson, el CCI_C y el CCI_A valen 1,0). La recta marcada en rojo representa la relación entre $VMTrm1$ y $VMTrm2$. Ambas retas están muy próximas lo que indica que todos los coeficientes de correlación serán muy altos, muy cercanos a 1.



En la Tabla 15A1 se muestran los tres coeficientes de correlación con su nivel de significación.

Tabla 15A.5. Coeficientes de correlación en la relación $VMTrm1 - VMTrm2$.		
Prueba	CC	p
Coefficiente de Correlación r de Pearson	$r = 0,989$	<0,001
Coefficiente de Correlación Intraclase de Consistencia	$CCI_C = 0,987$	<0,001
Coefficiente de Correlación Intraclase de Acuerdo	$CCI_A = 0,988$	<0,001

Videotutorial 15A2CCI1.avi

Con el archivo *Graves.sav* se muestra como se evalúa la Fiabilidad de dos mediciones cuantitativas (*VMTrm1* y *VMTrm2*) con el **Coefficiente de correlación intraclass (CCI) de consistencia y de acuerdo** a partir del cuadro *Análisis de fiabilidad*. También se muestra como se realiza el **Diagrama de dispersión** con el cuadro *Diagrama de dispersión simple* y como se obtienen el **coeficiente de correlación *r* de Pearson** con el cuadro *Correlaciones bivariadas*.