

Tema 3

Calidad de los datos observacionales

3.1 Fiabilidad, precisión, concordancia y validez

3.2 Concordancia global

3.3 Concordancia local

3.4 Concordancia secuencial

3.5 Sesgos de la observación

3.1 Fiabilidad, precisión, concordancia y validez

- Los resultados que se obtienen en una investigación son **creíbles** o **fiables** si se logra garantizar que los datos que se han registrado son **objetivos**.
- Por muy complejo o sofisticado que sea el análisis de los datos, no producirá más que resultados equivocados si los datos de partida son erróneos. Por ello es necesario demostrar que los datos poseen calidad, carecen de error.

“El gobierno es muy propenso a amasar estadísticas: las recogen, las elevan a la enésima potencia, les extraen la raíz cúbica y hacen con ellas bonitos diagramas. Pero lo que nadie debe olvidar es que todas y cada una de esas cifras provienen en primera instancia del vigilante del pueblo, el cual toma nota de los datos que le da la gana.”

(comentarios de un juez británico, citados por Suen y Ary, 1989).

Es decir, **los datos incorrectos producen resultados incorrectos sea cual sea el análisis estadístico.**

3.1 Fiabilidad, precisión, concordancia y validez

- **Fiabilidad de un instrumento de medida:** El instrumento es fiable si proporciona la misma medida al aplicarlo dos veces o más al mismo objeto (se supone que la propiedad que se está midiendo de dicho objeto no cambia).

Ejemplo: un test de inteligencia es fiable si al administrarlo dos o más veces a la misma persona proporciona el mismo CI para esa persona.

- **Precisión de un instrumento de medida:** El instrumento es preciso si el valor de la medida que proporciona es igual al valor “verdadero” de la misma.

Ejemplo: un test de inteligencia es preciso si el CI que proporciona para una persona es igual a su auténtico CI.

- **Validez de un instrumento de medida:** El instrumento es válido si mide aquella propiedad que se supone que ha de medir.

Ejemplo: un test de inteligencia es válido si las puntuaciones de CI que proporciona son realmente una medida de la inteligencia y no de otra propiedad (p.e., de ansiedad ante el test).

3.1 Fiabilidad, precisión, concordancia y validez

- **Fiabilidad de un observador:** El observador es fiable si obtiene el mismo registro al observar dos veces o más la misma conducta, aplicando los mismos sistemas de categorías.
- **Precisión de un observador:** El observador es preciso si el registro que obtiene es idéntico al registro “verdadero” que obtendría un observador infalible o perfecto, aplicando los mismos sistemas de categorías.
- **Validez de un sistema de categorías:** Un sistema de categorías es válido para un problema de investigación concreto si permite obtener soluciones al problema.

<i>Psicometría</i>	<i>Observación sistemática</i>
El instrumento de medida es un test	El instrumento de medida es un observador entrenado a registrar la conducta utilizando sistemas de categorías
La fiabilidad y la precisión son propiedades del test	La fiabilidad y la precisión son propiedades del observador
La validez es una propiedad del test	La validez es una propiedad de los sistema de categorías

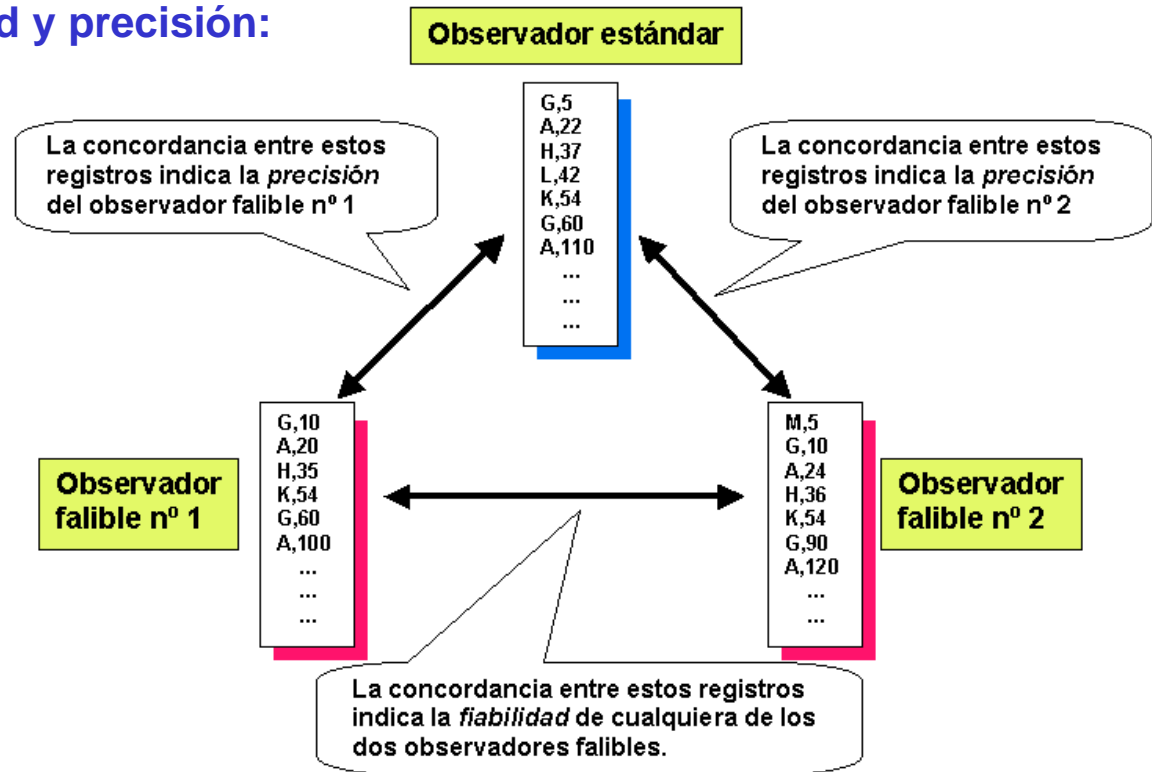
3.1 Fiabilidad, precisión, concordancia y validez

Concordancia entre observadores:

- Los registros o datos obtenidos por observadores entrenados sólo son creíbles si es posible demostrar que los observadores son **intercambiables**.
- Cuanto mayor sea la **concordancia** entre los registros de los observadores, más probable es que éstos sean intercambiables.
- La concordancia entre observadores es el **sine qua non** de la observación sistemática. Es decir, si no existe concordancia, la observación no puede considerarse científica.
- **Fiabilidad observacional:** Se estima como el grado de concordancia entre los registros obtenidos por **dos observadores falibles** al observar de forma independiente la misma conducta empleando los mismos sistemas de categorías.
- **Precisión observacional:** Se estima como el grado de concordancia entre los registros obtenidos por **un observador falible y uno infalible, perfecto o estándar** al observar de forma independiente la misma conducta empleando los mismos sistemas de categorías.

3.1 Fiabilidad, precisión, concordancia y validez

Relación entre fiabilidad y precisión:



- Una fiabilidad alta no indica necesariamente que los datos de los dos observadores falibles estén exentos de error: pueden estar cometiendo los mismos errores.
- Por lo tanto, la **precisión** es un mejor indicador de la calidad de los datos observacionales que la fiabilidad.
- Sin embargo, puesto que es difícil disponer de observadores estándar, es más habitual evaluar simplemente la fiabilidad.

3.1 Fiabilidad, precisión, concordancia y validez

Tipos de fiabilidad observacional:

- **Fiabilidad de equivalencia:** Se comparan los datos obtenidos por dos observadores falibles que:
 - Han sido igualmente entrenados y utilizan los mismos sistemas de categorías.
 - Observan de forma independiente.
 - Observan a los mismos individuos durante las mismas sesiones.

Si los datos concuerdan, se concluye que los observadores son *equivalentes* o *intercambiables*.

- **Fiabilidad de consistencia:** Se comparan los datos obtenidos por un mismo observador falible:
 - Al observar en dos ocasiones una grabación en vídeo de una sesión de observación.
 - Empleando los mismos sistemas de categorías en ambas ocasiones.

Si los datos concuerdan, se concluye que el observador es *consistente*.

3.1 Fiabilidad, precisión, concordancia y validez

Tipos de concordancia entre observadores

(como fiabilidad o como precisión):

- **Concordancia global:** Las medidas conductuales básicas y derivadas (frecuencias, duraciones, tasas, etc.) obtenidas por ambos observadores son muy similares.
- **Concordancia local (o “punto por punto”):** Ambos observadores tienden a registrar las mismas categorías de conducta de forma idéntica en cada unidad o intervalo de tiempo (por ejemplo, segundo a segundo).
- **Concordancia secuencial:** Las secuencias de categorías de conducta obtenidas por los dos observadores son muy similares.
 - *Local* es más estricta y más difícil de lograr que *secuencial*.
 - *Secuencial* es más estricta y más difícil de lograr que *global*.

3.1 Fiabilidad, precisión, concordancia y validez

Tipos de concordancia observacional:

		Unidades de tiempo									
		1	2	3	4	5	6	7	8	9	10
Obs1	B	C	B	C	A	B	B	C	A	A	
Obs2	A	A	B	C	B	C	A	B	B	C	

Concordancia local: el 20% (2/10) de las unidades de tiempo contienen categorías idénticas.

Categ.	Frecuencia		Duración	
	Ob1	Ob2	Ob1	Ob2
A	2	2	3	3
B	3	3	4	4
C	3	3	3	3

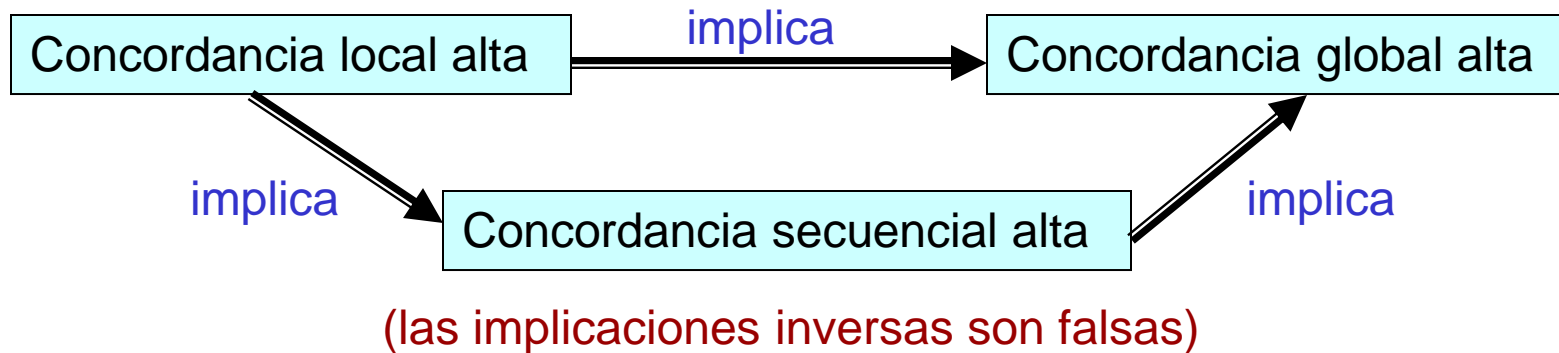
Concordancia global: el 100% de las medidas conductuales globales son idénticas.

Obs1:		B	C	B	C	A	B	C	A
Obs2:	A	B	C	B	C	A	B	C	

Concordancia secuencial: el 87.5% (7/8) de las categorías de las secuencias encajan.

3.1 Fiabilidad, precisión, concordancia y validez

Tipos de concordancia observacional:



- Si se calcula la concordancia **local** y resulta ser alta:
No hay necesidad de calcular las concordancias **secuencial** y **global**, porque ya se sabe que también son altas.
- Si se calcula la concordancia **global** y resulta ser alta:
No es posible concluir que las concordancias **local** y **secuencial** también son altas. Para saberlo, hay que calcularlas.

3.1 Fiabilidad, precisión, concordancia y validez

Tipos de concordancia observacional: El tipo de concordancia a calcular depende de las preguntas que se formula el investigador y de los objetivos a alcanzar en la investigación:

<i>Objetivo</i>	<i>Ejemplo</i>	<i>Tipo de concordancia a calcular</i>
Comparar frecuencias, duraciones o tasas entre grupos	“¿Los bebés prematuros tienden a llorar más por unidad de tiempo que los nacidos a término?”	Global. No hay necesidad de calcular las concordancias local y secuencial.
Buscar patrones en las secuencias de conducta	“¿Tiende el hombre a quejarse después de que la mujer se queja?”	Secuencial. No hay necesidad de calcular la concordancia local. Calcular la concordancia global no es suficiente.
Analizar la sincronía entre categorías	“¿Tiende A a mirar a B mientras B está mirando a A?”	Local. Calcular las concordancias global y secuencial no es suficiente.

3.1 Fiabilidad, precisión, concordancia y validez

Error observacional:

Definición general: **Error = Medida obtenida (X) – Medida verdadera (V)**

- Las medidas verdaderas son desconocidas. Por tanto, el error exacto que se comete al obtener X también lo es. Cuanto mayor es el *error* cometido por un observador falible, menor es su *precisión*.

Dos tipos de error en la observación:

- **Error cuantitativo:** cuando X y V son medidas conductuales cuantitativas, como frecuencias, duraciones, tasas, etc. A mayor error cuantitativo, menor *concordancia global* con un observador estándar.
- **Error cualitativo:** cuando X y V son medidas binarias, como la *presencia / ausencia* de una categoría en una unidad o intervalo de tiempo. A mayor error cualitativo, menor *concordancia local* con un observador estándar.

3.1 Fiabilidad, precisión, concordancia y validez

- **Error cuantitativo:** Puede ser *por exceso* o *por defecto*.

Por ejemplo, si la frecuencia verdadera de una conducta es 15, el observador comete error por exceso si obtiene $F > 15$ y error por defecto si obtiene $F < 15$.

- **Error cualitativo:** Puede ser de *omisión* o de *comisión*.

El observador registra la conducta en la unidad de tiempo (X)	La categoría ocurre realmente en la unidad de tiempo (V)	Error = X - V
Sí: X = 1	Sí: V = 1	No: E = 0
Sí: X = 1	No: V = 0	Sí: E = +1 Error de comisión, "falso positivo"
No: X = 0	Sí: V = 1	Sí: E = -1 Error de omisión, "falso negativo"
No: X = 0	No: V = 0	No: E = 0

3.1 Fiabilidad, precisión, concordancia y validez

Validez observacional:

- **Validez de contenido:** Grado en que las categorías de conducta reflejan adecuadamente el dominio conductual objeto de estudio.
 - Por ejemplo, grado en que un sistema de categorías del juego infantil refleja adecuadamente el juego de los niños de una determinada edad.
- **Validez de constructo:** Grado en que las categorías de conducta son buenos indicadores de constructos inobservables.
 - Por ejemplo, grado en que conductas observables como “pegar”, “dar patadas”, etc. son buenos indicadores del constructo inobservable “agresividad”.
 - Las categorías cuyas definiciones poseen términos *inferenciales* (por ejemplo, causales) deben tener validez de constructo.
 - La validez de constructo *no es un problema* si las categorías se definen de forma estructural y no pretenden utilizarse como indicadores de constructos inobservables.
- La existencia de validez es un problema teórico, que debe abordarse a través del juicio de expertos en el problema objeto de estudio.

3.2 Concordancia global

Datos para el cálculo de la concordancia global:

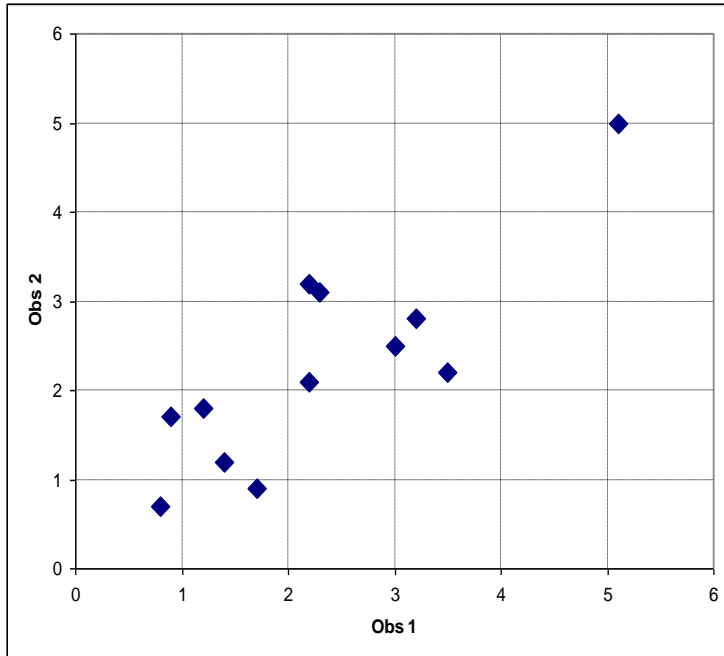
- Se comparan las medidas globales (F , D , t , etc.) de dos o más observadores.
- Las medidas han sido obtenidas:
 - A lo largo de varias sesiones de observación, *o bien*
 - Para varios individuos, *o bien*
 - Para varias categorías de conducta, *o bien*
 - Para varios individuos a lo largo de varias sesiones, etc.
- Lo esencial es que existe una *variable relevante* (sesiones, individuos, categorías) respecto a la cual se desea saber si concuerdan las medidas globales obtenidas por los observadores (*variable irrelevante*).

Ejemplo: Dos observadores han obtenido estas tasas por minuto de la categoría “llorar” en 12 niños (observados en distintas sesiones):

	Niños											
	1	2	3	4	5	6	7	8	9	10	11	12
Obs 1	3.2	5.1	1.2	0.9	1.4	2.2	2.3	2.2	1.7	0.8	3.0	2.5
Obs 2	2.8	5.0	1.8	1.7	1.2	3.2	3.1	2.1	0.9	0.7	2.5	2.2

3.2 Concordancia global

Diagrama de dispersión: Es útil como indicio de la existencia de relación entre las medidas obtenidas por los dos observadores.



- En este ejemplo, las tasas según uno y otro observador parecen mantener una relación positiva (a mayor tasa según *Obs1*, mayor tasa según *Obs2*).
- La correlación lineal de Pearson indica en qué grado están relacionadas las medidas: **$r = 0.87$**

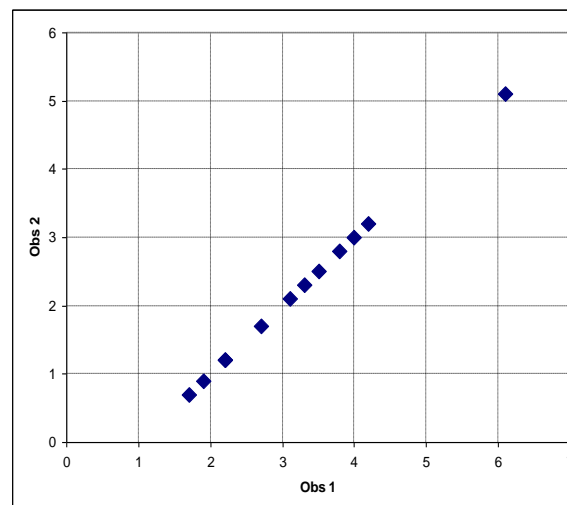
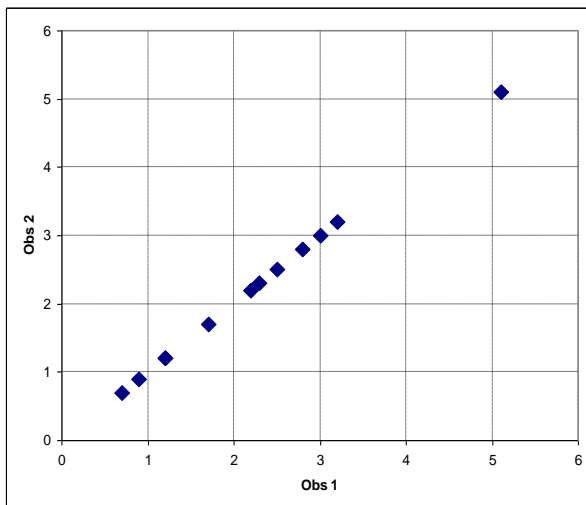
$$[-1 \leq r \leq +1]$$

	<i>Niños</i>											
	1	2	3	4	5	6	7	8	9	10	11	12
<i>Obs 1</i>	3.2	5.1	1.2	0.9	1.4	2.2	2.3	2.2	1.7	0.8	3.0	2.5
<i>Obs 2</i>	2.8	5.0	1.8	1.7	1.2	3.2	3.1	2.1	0.9	0.7	2.5	2.2

3.2 Concordancia global

Concordancia global relativa

- Dos observadores concuerdan de manera relativa si sus datos están *correlacionados*.
- La **correlación lineal r de Pearson** indica la concordancia relativa.
- Útil cuando solo interesa conocer la concordancia respecto al orden de los valores (p.e., acerca de qué niños lloran más) y no respecto a sus magnitudes (p.e., acerca de cuántas veces han llorado).
- Concordancia relativa = concordancia referida a una norma.
- Dos ejemplos de concordancia relativa perfecta, $r = 1$:

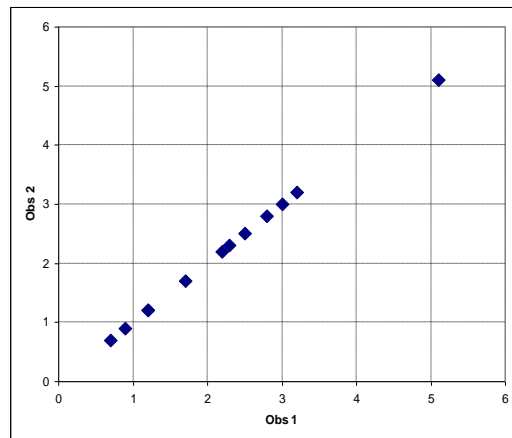


En este caso, las diferencias sistemáticas entre los observadores no son detectadas por r .

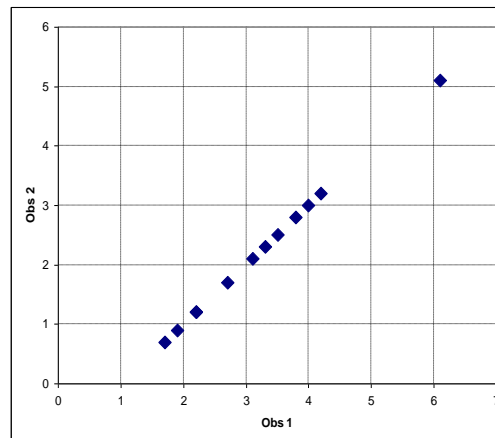
3.2 Concordancia global

Concordancia global absoluta

- Dos observadores concuerdan de manera absoluta si sus datos *coinciden*.
- El **índice de Berk** ρ_c^2 (“coeficiente de correlación intraclase”) indica la concordancia absoluta.
- Este índice detecta diferencias sistemáticas entre los datos de los dos observadores (el coeficiente de Pearson, no).
- Útil cuando interesa conocer la concordancia tanto acerca del orden como de las magnitudes.
- Concordancia absoluta = concordancia referida a un criterio.



$$r = 1, \rho_c^2 = 1$$

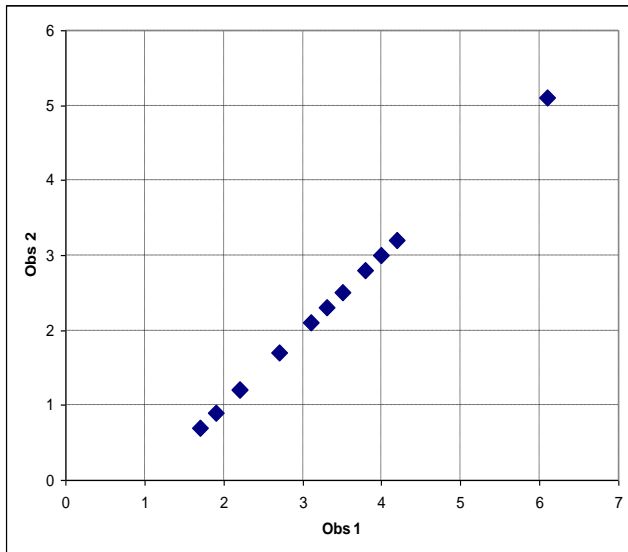


$$r = 1, \rho_c^2 < 1$$

En este caso, las diferencias sistemáticas entre los observadores no son detectadas por r pero sí por el índice de Berk.

3.2 Concordancia global

Concordancia global absoluta y relativa



- En este ejemplo existe concordancia relativa, pero NO concordancia absoluta:

$$r = 1, \rho_c^2 = 0.75$$

- Los puntos del diagrama se sitúan en una línea recta que NO pasa por (0,0).
- La recta no pasa por (0,0) porque los datos de *Obs 1* son sistemáticamente mayores que los de *Obs 2*.

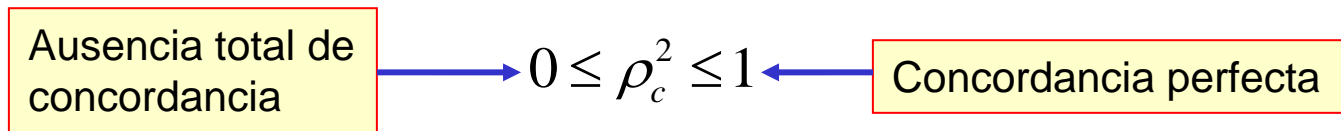
		<i>Niños</i>											
		1	2	3	4	5	6	7	8	9	10	11	12
<i>Obs 1</i>		4.2	6.1	2.2	1.9	2.2	3.8	3.3	3.1	2.7	1.7	4.0	3.5
<i>Obs 2</i>		3.2	5.1	1.2	0.9	1.2	2.8	2.3	2.1	1.7	0.7	3.0	2.5

- Por lo tanto, r no detecta diferencias sistemáticas entre los datos de los dos observadores, pero ρ_c^2 sí las detecta.
- La correlación intraclase de Berk es un indicador de concordancia global más exigente que la correlación lineal de Pearson.

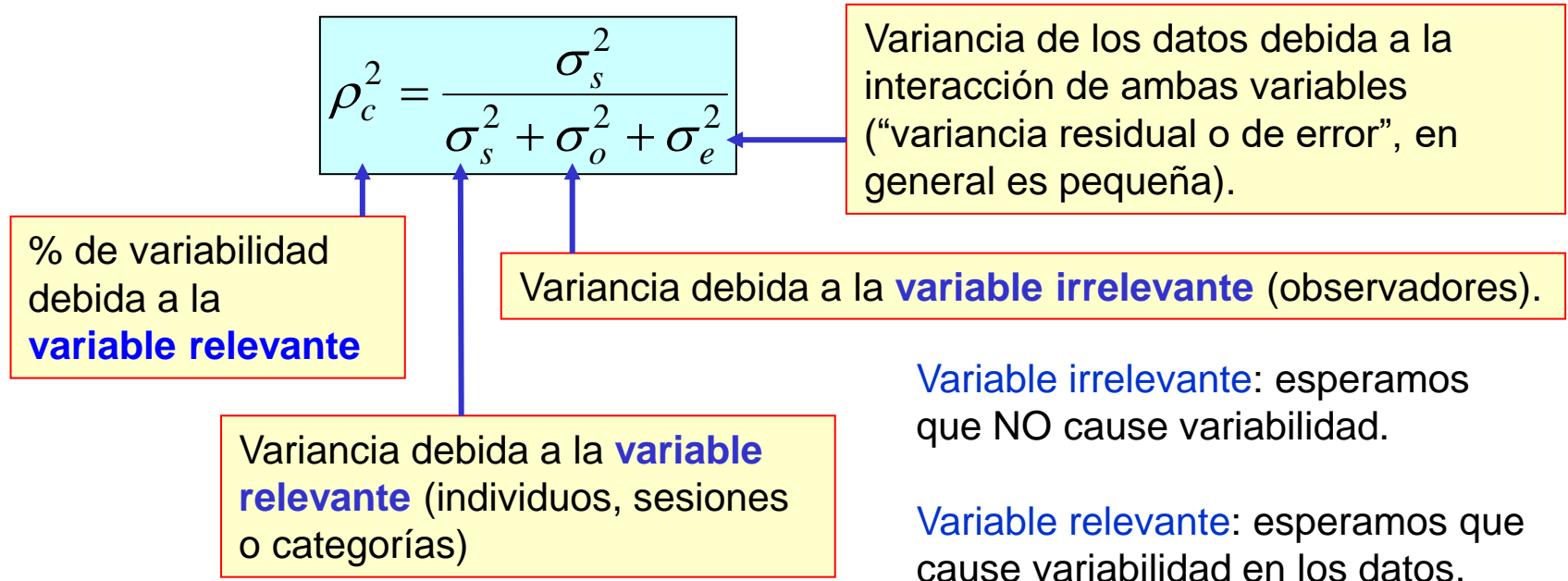
3.2 Concordancia global

Índice de Berk

- Indica la concordancia global absoluta entre dos o más observadores:



- Si un observador obtiene sistemáticamente medidas globales mayores o menores que los otros, entonces $\rho_c^2 < 1$.



3.2 Concordancia global

Índice de Berk

La variancia de estos datos es nula.
Concordancia entre observadores = 1:

		Observadores			
		Obs 1	Obs 2	Obs 3	Obs 4
Individuos	1	8	8	8	8
	2	8	8	8	8
	3	8	8	8	8
	4	8	8	8	8
	5	8	8	8	8

En estos datos sólo existe variancia debida a la variable **relevante** (individuos).
Concordancia entre observadores = 1:

		Observadores			
		Obs 1	Obs 2	Obs 3	Obs 4
Individuos	1	10	10	10	10
	2	8	8	8	8
	3	13	13	13	13
	4	21	21	21	21
	5	5	5	5	5

En estos datos sólo existe variancia debida a la variable **irrelevante** (observadores).
Concordancia entre observadores < 1:

		Observadores			
		Obs 1	Obs 2	Obs 3	Obs 4
Individuos	1	8	10	6	17
	2	8	10	6	17
	3	8	10	6	17
	4	8	10	6	17
	5	8	10	8	17

En estos datos existe variancia debida tanto a la variable **relevante** como a la **irrelevante**.
Concordancia entre observadores < 1:

		Observadores			
		Obs 1	Obs 2	Obs 3	Obs 4
Individuos	1	10	13	9	10
	2	8	6	10	8
	3	13	12	10	12
	4	21	20	21	20
	5	5	6	6	7

3.2 Concordancia global

Índice de Berk

Ejemplo: Cuatro observadores falibles simultáneos e independientes obtienen la frecuencia de una categoría en 5 individuos durante una sesión de observación:

Observadores (k = 4)

	1	2	3	4	S_i
1	5	6	3	5	4.75
2	9	10	7	7	8.25
3	6	10	3	7	6.50
4	1	2	0	1	1.00
5	4	5	2	5	4.00
O_j	5.00	6.60	3.00	5.00	4.90

Individuos (n = 5)

Frecuencia obtenida por el observador 2 para el individuo 5 (X_{52})

Frecuencias medias para cada observador (O_j)

Frecuencias medias para cada individuo (S_i)

Frecuencia media general de la tabla (M)

3.2 Concordancia global

Índice de Berk

Paso 1: Calcular las “sumas de cuadrados” :

- Total:
- Debida a los individuos:
- Debida a los observadores:
- De error o residual:

$$SC_t = \sum X_{ij}^2 - nkM^2$$

$$SC_s = k \sum S_i^2 - nkM^2$$

$$SC_o = n \sum O_j^2 - nkM^2$$

$$SC_e = SC_t - SC_s - SC_o$$

Paso 2: Calcular los “cuadrados medios”:

- Debido a los individuos:
- Debido a los observadores:
- De error o residual:

$$CM_s = SC_s / (n - 1)$$

$$CM_o = SC_o / (k - 1)$$

$$CM_e = SC_e / [(k - 1)(n - 1)]$$

Paso 3: Estimar las variancias:

- Debida a los individuos:
- Debida a los observadores:
- De error o residual:

$$\hat{\sigma}_s^2 = (CM_s - CM_e) / k$$

$$\hat{\sigma}_o^2 = (CM_o - CM_e) / n$$

$$\hat{\sigma}_e^2 = CM_e$$

3.2 Concordancia global

Índice de Berk

Paso 1: Calcular las “sumas de cuadrados” :

- Total: $SC_t = 5^2 + 6^2 + \dots + 5^2 - (5 \times 4 \times 4.90^2) = 163.8$
- Debida a los individuos: $SC_s = 4 \times (4.75^2 + 8.25^2 + \dots + 4.00^2) - (5 \times 4 \times 4.90^2) = 119.3$
- Debida a los observadores: $SC_o = 5 \times (5.00^2 + 6.60^2 + 3.00^2 + 5.00^2) - (5 \times 4 \times 4.90^2) = 32.6$
- De error o residual: $SC_e = 163.8 - 119.3 - 32.6 = 11.9$

Paso 2: Calcular los “cuadrados medios”:

- Debido a los individuos: $CM_s = 119.3 / (5 - 1) = 29.825$
- Debido a los observadores: $CM_o = 32.6 / (4 - 1) = 10.867$
- De error o residual: $CM_e = 11.9 / [(5 - 1) \times (4 - 1)] = 0.992$

Paso 3: Estimar las variancias:

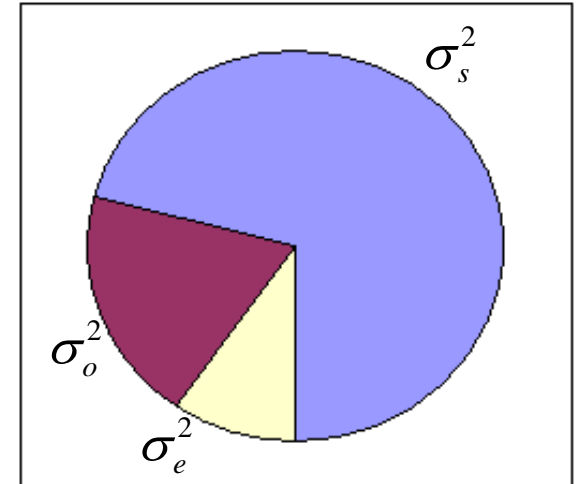
- Debida a los individuos: $\hat{\sigma}_s^2 = (29.825 - 0.992) / 4 = 7.208$
- Debida a los observadores: $\hat{\sigma}_o^2 = (10.867 - 0.992) / 5 = 1.975$
- De error o residual: $\hat{\sigma}_e^2 = 0.992$

3.2 Concordancia global

Índice de Berk

- **Tabla resumen:**

	Sumas de cuadrados	Cuadrados medios	Estimaciones de variancia
<i>Individuos</i>	119.3	29.825	7.208
<i>Observadores</i>	32.6	10.867	1.975
<i>Error o residual</i>	11.9	0.992	0.992
<i>Total</i>	163.8		10.175



- **Cálculo del índice:**

$$\rho_c^2 = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_o^2 + \sigma_e^2} = \frac{7.208}{7.208 + 1.975 + 0.992} = \frac{7.208}{10.175} = 0.708$$

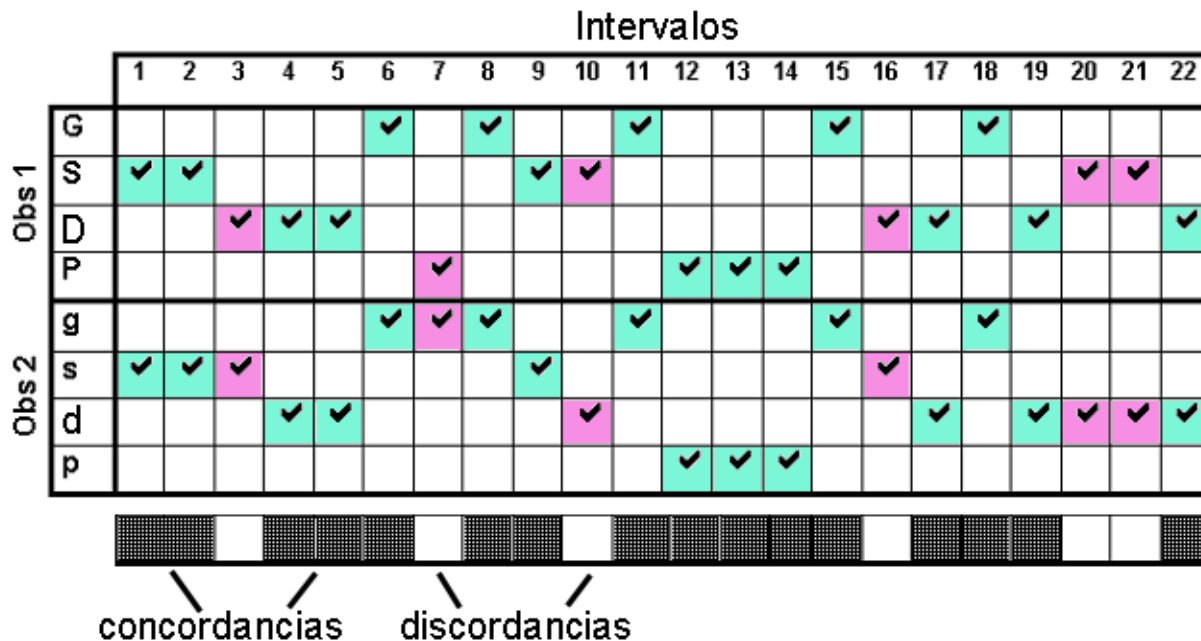
$\rho_c^2 = 0.708$ es la **concordancia global absoluta** promedio de estos 4 observadores respecto a la frecuencia de una cierta categoría en estos 5 individuos.

- Variable relevante (individuos): explica el 70.8% de la variabilidad.
- Variable irrelevante (observadores): explica el 19.4% (= 1.975/10.175).

3.3 Concordancia local

En el registro activado por unidades de tiempo (RAUT): Los observadores concuerdan si en cada intervalo anotan categorías idénticas.

Por ejemplo, un muestreo instantáneo con 22 intervalos:



Ejemplo con un sistema de categorías simplificado de juego infantil:
{Grupal, Solitario, Desocupado, Paralelo}

3.3 Concordancia local

En el registro activado por transiciones (RAT):

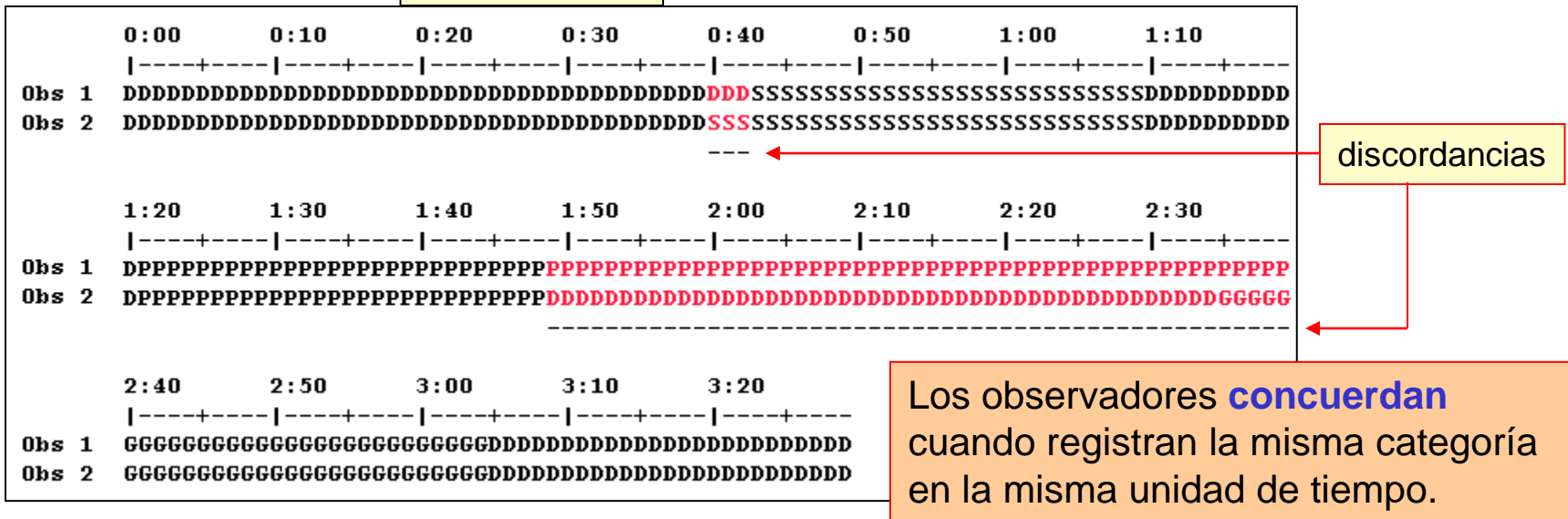
- Los observadores han anotado los tiempos de inicio y fin de las categorías:

Obs 1	Obs 2
D, 0:00-0:43	D, 0:00-0:40
S, 0:43-1:10	S, 0:40-1:10
D, 1:10-1:21	D, 1:10-1:21
P, 1:21-2:40	P, 1:21-1:49
G, 2:40-3:05	D, 1:49-2:35
D, 3:05-3:30	G, 2:35-3:05
	D, 3:05-3:30

Dos tipos de **discordancia** posibles:

Anotar distintos tiempos de inicio y/o fin para una misma ocurrencia de una categoría.

Anotar una categoría diferente durante unas unidades de tiempo.



3.3 Concordancia local

Tabla de concordancias y discordancias ($m \times m$): Se contabilizan las cantidades de concordancias y de discordancias y se disponen en una tabla de contingencia de m filas por m columnas ($m = \text{núm. de categorías}$).

		Obs 2						
		A	B	C	D	E	...	Total
Obs1	A	n_{11}	n_{12}	n_{13}	n_{14}	n_{15}	...	n_{1+}
	B	n_{21}	n_{22}	n_{23}	n_{24}	n_{25}	...	n_{2+}
	C	n_{31}	n_{32}	n_{33}	n_{34}	n_{35}	...	n_{3+}
	D	n_{41}	n_{42}	n_{43}	n_{44}	n_{45}	...	n_{4+}
	E	n_{51}	n_{52}	n_{53}	n_{54}	n_{55}	...	n_{5+}

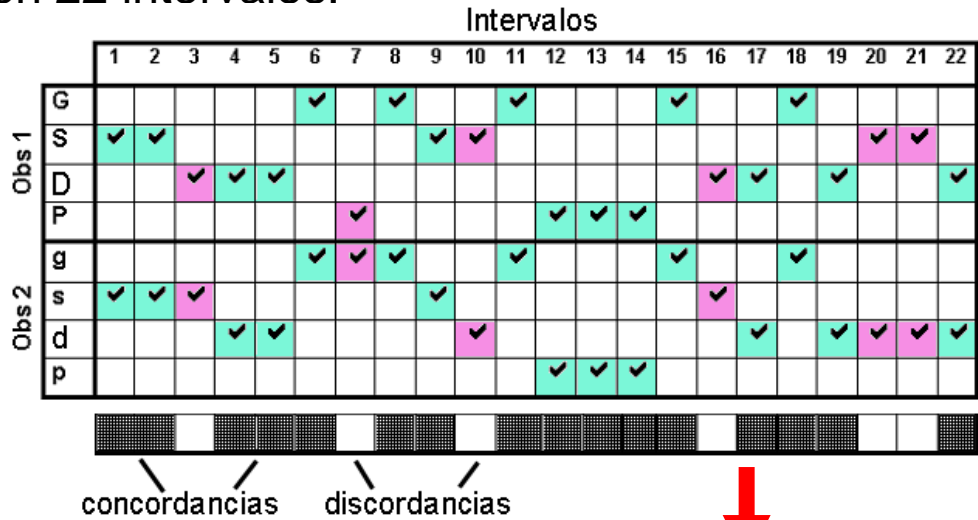
Total	n_{+1}	n_{+2}	n_{+3}	n_{+4}	n_{+5}	...	N	

Categorías según han sido anotadas por Obs 1 (pointing to Obs1 labels)
Categorías según han sido anotadas por Obs 2 (pointing to Obs2 labels)
Cantidad de intervalos, unidades de tiempo o segmentos conductuales en los que Obs 1 anotó A mientras Obs 2 anotó E (**discordancia**) (pointing to n_{15})
Cantidad de intervalos, unidades de tiempo o segmentos conductuales en los que Obs 1 y Obs2 **concordaron** en anotar E (pointing to n_{55})
Diagonal: concordancias
Fuera de la diagonal: discordancias

En esta tabla de contingencia solo pueden representarse categorías EME.

3.3 Concordancia local

Tabla de concordancias y discordancias: Ejemplo para un muestreo instantáneo con 22 intervalos:

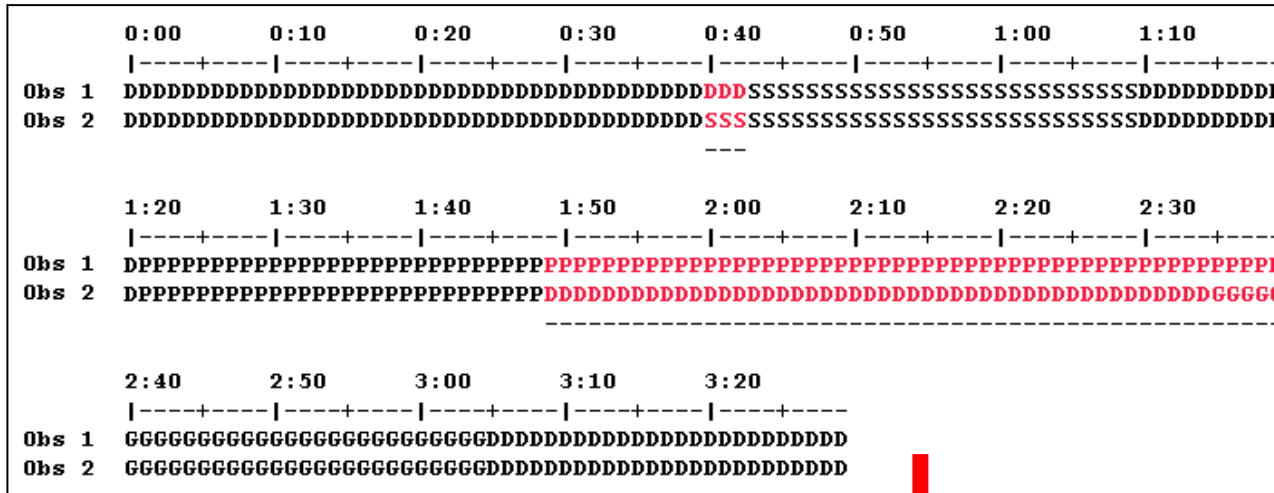


		Obs 2				Total
		g	s	d	p	
Obs 1	G	5	0	0	0	5
	S	0	3	3	0	6
	D	0	2	5	0	7
	P	1	0	0	3	4
Total		6	5	8	3	22

Cantidad total de intervalos

3.3 Concordancia local

Tabla de concordancias y discordancias: Ejemplo para un RAT con tiempos de inicio y fin:



		Obs 2				Total
		g	s	d	p	
Obs 1	G	25	0	0	0	25
	S	0	27	0	0	27
	D	0	3	76	0	79
	P	5	0	46	28	79
Total		30	30	122	28	210

Cantidad total de segundos (duración de la sesión)

3.3 Concordancia local

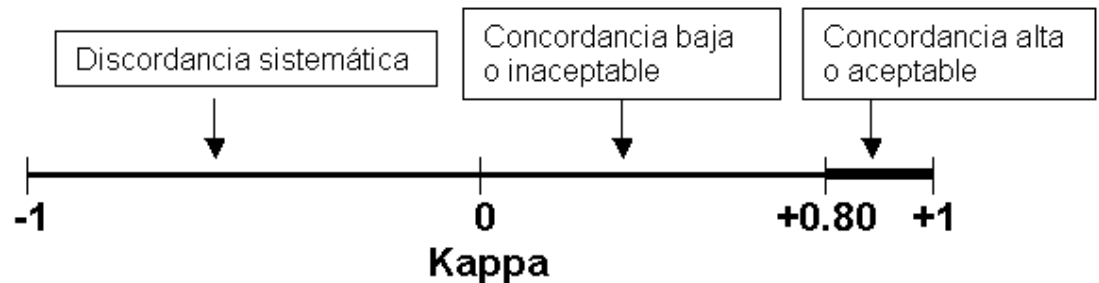
Coeficiente *kappa* (κ) de Cohen:

- Indica el grado de concordancia local entre dos observadores respecto a un sistema de categorías (EME).
- Al calcularlo se tiene en cuenta la posibilidad de que los observadores puedan concordar por azar.
- Pasos:

- Calcular P_o , la **proporción de concordancia observada**.
- Calcular P_e , la **proporción de concordancia esperada** que se estima que tendrían los observadores si uno o ambos registrasen la conducta al azar.

- Calcular ***kappa***:

$$\kappa = \frac{P_o - P_e}{1 - P_e}$$



3.3 Concordancia local

Coeficiente *kappa*:

Proporción de concordancia observada:

$$P_o = \frac{\sum_{i=1}^m n_{ii}}{n_{++}}$$

Proporción de concordancia esperada debida al azar:

$$P_e = \frac{\sum_{i=1}^m n_{i+} n_{+i}}{n_{++}^2}$$

$$P_o = \frac{5+3+5+3}{22} = 0.7273$$

$$P_e = \frac{5 \cdot 6 + 6 \cdot 5 + 7 \cdot 8 + 4 \cdot 3}{22^2} = 0.2645$$

		Obs2				Total
		g	s	d	p	
Obs1	G	5	0	0	0	5
	S	0	3	3	0	6
	D	0	2	5	0	7
	P	1	0	0	3	4
Total		6	5	8	3	22

$$\begin{aligned} K &= \frac{0.7273 - 0.2645}{1 - 0.2645} \\ &= 0.6292 \end{aligned}$$

3.3 Concordancia local

Valor máximo posible de *kappa*:

- Si la concordancia *global* de los observadores no es perfecta, su concordancia *local*, medida con *kappa*, no puede alcanzar el valor +1.
- Al no haber concordancia global perfecta, como mínimo para alguna categoría ocurre que $n_{i+} \neq n_{+i}$. Entonces $\kappa_{max} < +1$.

$$\kappa_{max} = \frac{P_{o max} - P_e}{1 - P_e}$$

$$P_{o max} = \frac{\sum_{i=1}^m \min(n_{i+}, n_{+i})}{n_{++}}$$

La diferencia $1 - \kappa_{max}$ es una medida indirecta de la *discordancia global*.

		Obs 2				Total
		g	s	d	p	
Obs 1	G	5	0	0	0	5
	S	0	3	3	0	6
	D	0	2	5	0	7
	P	1	0	0	3	4
Total		6	5	8	3	22

$$P_{o max} = \frac{5+5+7+3}{22} = 0.9091$$

$$\kappa_{max} = \frac{0.9091 - 0.2645}{1 - 0.2645} = 0.8764$$

Por lo tanto, los observadores como mucho podrían haber alcanzado *kappa* = 0.8764 y obtuvieron *kappa* = 0.6292.

3.3 Concordancia local

Tabla parcial de concordancias y discordancias:

Cuando solo interesa conocer la concordancia respecto al registro de una única categoría de conducta (X), la tabla es 2 x 2.

Obs 1 registra o no (X / no X) la categoría en cada unidad tiempo o intervalo.

Obs 2 registra o no (X / no X) la categoría en cada unidad tiempo o intervalo.

		Obs 2		Total
		X	No X	
Obs 1	X	a	b	a+b
	No X	c	d	c+d
Total		a+c	b+d	a+b+c+d = N

Cantidad de unidades de tiempo o intervalos en los que **Obs 1 registra X.**

Cantidad de unidades de tiempo o intervalos en los que **Obs 1 no registra X.**

Cantidad total de unidades de tiempo o intervalos de la sesión de observación.

Cantidad de unidades de tiempo o intervalos en los que **Obs 2 registra X.**

Cantidad de unidades de tiempo o intervalos en los que **Obs 2 no registra X.**

Cantidad de concordancias: a + d
Cantidad de discordancias: b + c

3.3 Concordancia local

Ejemplo de tabla parcial de concordancias y discordancias:

Un muestreo de intervalo parcial en el que un intervalo puede contener más de una categoría.

		Intervalos																					
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
Obs 1	G						✓		✓			✓	✓	✓		✓			✓				
	S	✓	✓		✓				✓	✓	✓								✓		✓	✓	
	D		✓	✓	✓	✓												✓	✓		✓		✓
	P							✓					✓	✓	✓								
Obs 2	g						✓	✓	✓			✓	✓			✓			✓				
	s	✓	✓	✓					✓	✓								✓		✓	✓		
	d		✓		✓	✓					✓								✓		✓	✓	✓
	p												✓	✓	✓								

		Obs 2		
		s	No s	Total
Obs 1	S	5	4	9
	No S	3	10	13
	Total	8	14	22

3.3 Concordancia local

Cálculo de *kappa* para una tabla parcial de concordancias y discordancias:

		Obs 2		Total
		X	No X	
Obs 1	X	a	b	a+b
	No X	c	d	c+d
Total		a+c	b+d	a+b+c+d = N

Proporción de concordancia
observada:

$$P_o = \frac{a+d}{N}$$

Proporción de concordancia
esperada debida al azar:

$$P_e = \frac{(a+b) \cdot (a+c) + (c+d) \cdot (b+d)}{N^2}$$

		Obs 2		Total
		s	No s	
Obs 1	S	5	4	9
	No S	3	10	13
Total		8	14	22

$$P_o = \frac{5+10}{22} = \frac{15}{22} = 0.6818$$

$$P_e = \frac{9 \cdot 8 + 13 \cdot 14}{22^2} = \frac{254}{484} = 0.5248$$

$$\kappa = \frac{0.6818 - 0.5248}{1 - 0.5248} = 0.3304$$

3.3 Concordancia local

Obtención de tablas 2 x 2 a partir de una tabla $m \times m$

Cuando se han contabilizado las frecuencias de una tabla $m \times m$ (categorías EME) puede obtenerse una tabla 2 x 2 para cada categoría *agrupando filas y columnas* de la tabla original:

Tabla 4 x 4:

		Obs 2				Total
		g	s	d	p	
Obs 1	G	5	0	0	0	5
	S	0	3	3	0	6
	D	0	2	5	0	7
	P	1	0	0	3	4
Total		6	5	8	3	22

Tabla 2 x 2 para la categoría D:

		Obs 2		Total
		d	No d	
Obs 1	D	5	2	7
	No D	3	12	15
	Total	8	14	22

3.3 Concordancia local

Kappa general y *kappas* parciales

- La *kappa* general del sistema de categorías es igual a la **media ponderada** de las *kappas* parciales de las categorías:

The diagram illustrates the formula for the general Kappa coefficient, $\kappa = \frac{\sum w_i \cdot \kappa_i}{\sum w_i}$. A central light blue box contains the formula. Three yellow boxes with red borders are connected to it by blue arrows: one on the left labeled 'Kappa general' points to the formula; one on the top right labeled 'Kappa parcial de la categoría i' points to the κ_i term in the numerator; and one on the bottom right labeled 'Peso de la categoría i' points to the w_i term in the numerator.

$$\kappa = \frac{\sum w_i \cdot \kappa_i}{\sum w_i}$$

- Por lo tanto, la concordancia general **no puede ser mayor** que la concordancia parcial máxima **ni menor** que la concordancia parcial mínima:

$$\min(\kappa_i) \leq \kappa \leq \max(\kappa_i)$$

- Los pesos son función de los totales de filas y de columnas en la tabla $m \times m$:

The diagram illustrates the formula for the weight $w_i = \frac{n_{i+} + n_{+i}}{2} - \frac{n_{i+} \cdot n_{+i}}{n_{++}}$. A central light blue box contains the formula. Two yellow boxes with red borders are connected to it by blue arrows: one on the left labeled 'Peso de la categoría i' points to the formula; and one on the right labeled 'Totales de la fila i y de la columna i' points to the n_{i+} and n_{+i} terms in the numerator.

$$w_i = \frac{n_{i+} + n_{+i}}{2} - \frac{n_{i+} \cdot n_{+i}}{n_{++}}$$

3.3 Concordancia local

Kappa general y kappas parciales

	g	s	d	p	Total
G	5	0	0	0	5
S	0	3	3	0	6
D	0	2	5	0	7
P	1	0	0	3	4
Total	6	5	8	3	22

$$\kappa = 0.6292$$

	g	No g		s	No s	
G	5	0	5	3	3	6
No G	1	16	17	2	14	16
	6	16	22	5	17	22
	d	No d		p	No p	
D	5	2	7	3	1	4
No D	3	12	15	0	18	18
	8	14	22	3	19	22

Se calculan las kappas parciales

	κ_i	w_i	$w_i \cdot \kappa_i$
G	0.8791	$(5+6)/2 - (5 \cdot 6)/22 = 4.1364$	3.6362
S	0.3956	$(6+5)/2 - (6 \cdot 5)/22 = 4.1364$	1.6363
D	0.4954	$(7+8)/2 - (7 \cdot 8)/22 = 4.9545$	2.4544
P	0.8308	$(4+3)/2 - (4 \cdot 3)/22 = 2.9545$	2.4546
Total		16.1818	10.1817

Se comprueba que la media ponderada de las kappas parciales es igual a la kappa general del sistema

$$\kappa = \frac{10.1817}{16.1818} = 0.6292$$

3.3 Concordancia local

Precisión local:

- Si en la tabla de concordancias y discordancias las filas corresponden a un *observador falible* y las columnas a un *observador estándar o perfecto*, entonces el coeficiente *kappa* indica el grado de *precisión del observador falible*.
- Otros indicadores de precisión que complementan la información proporcionada por *kappa*:

- **Sensibilidad del observador** respecto a una categoría de conducta:

Proporción de unidades de tiempo o intervalos en las que el observador falible *registra* la categoría de conducta cuando ésta verdaderamente *ocurre*.

Es una estimación de la probabilidad de que el observador falible detecte *cuándo ocurre* la categoría.

- **Especificidad del observador** respecto a una categoría de conducta:

Proporción de unidades de tiempo o intervalos en las que el observador falible *no registra* la categoría de conducta cuando ésta verdaderamente *no ocurre*.

Es una estimación de la probabilidad de que el observador falible detecte *cuándo no ocurre* la categoría.

3.3 Concordancia local

Coeficientes de precisión local:

		Obs estándar		Total
		X	No X	
Obs falible	X	a	b	a+b
	No X	c	d	c+d
	Total	a+c	b+d	

Valores verdaderos:

a + c = cantidad de unidades de tiempo en las que la categoría X **verdaderamente ha ocurrido.**

b + d = cantidad de unidades de tiempo en las que la categoría X **verdaderamente no ha ocurrido.**

Sensibilidad:

$$\xi = \frac{a}{a+c}$$

Especificidad:

$$\eta = \frac{d}{b+d}$$

Concordancias:

a = cantidad de aciertos en registrar la categoría

d = cantidad de aciertos en no registrarla

Discordancias:

b = cantidad de errores de comisión, o falsos positivos

c = cantidad de errores de omisión, o falsos negativos

3.3 Concordancia local

Coeficientes de precisión local. *Ejemplo:*

		Obs estándar		
		g	No g	
G	G	5	0	5
	No G	1	16	17
		6	16	22
		s		
		s	No s	
S	S	3	3	6
	No S	2	14	16
		5	17	22
		d		
		d	No d	
D	D	5	2	7
	No D	3	12	15
		8	14	22
		p		
		p	No p	
P	P	3	1	4
	No P	0	18	18
		3	19	22

	Sensibilidad ξ	Especificidad η	κ
G	$5/6 = 0.833$	$16/16 = 1$	0.879
S	$3/5 = 0.6$	$14/17 = 0.823$	0.396
D	$5/8 = 0.625$	$12/14 = 0.857$	0.495
P	$3/3 = 1$	$18/19 = 0.947$	0.831

- Si la sensibilidad **o** la especificidad son *bajas*, entonces kappa es *baja*.

Ejemplos: Categorías S y D.

- Si la sensibilidad **y** la especificidad son *altas*, entonces kappa es *alta*.

Ejemplos: Categorías G y P.

3.4 Concordancia secuencial

- Es el grado de semejanza de las secuencias que registran dos observadores.
- *Problema*: Si *no existe un consenso previo* acerca de cuáles son los segmentos de conducta y los observadores *sólo registran el orden en que ocurren las categorías* pero no sus tiempos de inicio y fin, entonces **no** es posible obtener una tabla de concordancias y discordancias partir de las secuencias.
- *Ejemplo* (**D**: Desocupado, **S**: Juego solitario, **J**: Juego conjunto, **P**: Juego paralelo, **G**: Juego grupal):

Obs 1: D S G J G P S P D J D P G
Obs 2: D S J P G S J D P G

¿Qué categorías registró Obs 2 mientras Obs 1 registraba cada una de las categorías?

- Una conjetura posibles (entre muchas otras) acerca de cómo *alinearse* o hacer corresponder las dos secuencias:

D-SGJ-GP-SP-DJDPG
| | | | ||
DS--J--PGS-JD--PG

3.4 Concordancia secuencial

- **Algoritmo de Needleman y Wunsch:** Es un procedimiento que alinea las secuencias de modo que la cantidad de concordancias sea la máxima posible.
- Se utiliza en la biología molecular para alinear secuencias de biomoléculas. Por ejemplo, para comparar dos secuencias de ADN de especies diferentes.
- En general se requiere un *software* específico para aplicar el algoritmo.
- El resultado es el *alineamiento óptimo* de las secuencias. Para el ejemplo anterior es:

DSGJGSPDJDPG
 || | | :: ||||
DS-J-P-GSJDPG

Una vez alineadas las secuencias, puede obtenerse una tabla de concordancias y discordancias y calcular kappa: $\kappa = 0.55$

		Obs 2						Totales
		-	D	S	J	P	G	
Obs 1	-	0	0	0	0	0	0	0
	D	0	2	1	0	0	0	3
	S	1	0	1	0	0	0	2
	J	0	0	0	2	0	0	2
	P	0	0	0	0	2	1	3
	G	2	0	0	0	0	1	3
	Totales	3	2	2	2	2	2	13

3.5 Sesgos de la observación

- En la observación sistemática los errores de medida son debidos a:
 - El **observador**: fallos en la atención, entrenamiento defectuoso, etc.
 - El **sistema de categorías**: definiciones no objetivas o confusas de las categorías, elevado número de categorías, etc.
 - La **tecnología de adquisición de datos**: imprecisión del cronómetro, fallos en el *software*, etc.
- Estos fallos pueden causar sesgos en la observación sistemática, los cuales hacen disminuir la *fiabilidad* y la *precisión*.
- **Sesgo [o error sistemático]**: existe sesgo cuando las medidas conductuales obtenidas son *siempre mayores* [o *siempre menores*] que las medidas verdaderas.
 - *Ejemplo*: un observador registra la categoría “Juego solitario” cuando el niño está realmente jugando solo pero también cuando está desocupado (en lugar de registrar entonces “Desocupado”). La frecuencia de “Juego solitario” tenderá a ser sistemáticamente mayor que la verdadera.

3.5 Sesgos de la observación

Factores que pueden afectar a la fiabilidad o a la precisión:

- Fallos mecánicos de registro
- Deriva del observador
- Complejidad de los sistemas de categorías
- Fallos perceptivos
- Expectativas del observador

Factor que puede afectar a la validez:

- Reactividad de los individuos observados

3.5 Sesgos de la observación

Fallos mecánicos de registro:

- Puesto que los observadores generalmente anotan las ocurrencias de las categorías mientras prestan atención a la conducta, pueden anotar una categoría equivocada en la hoja de registro o pulsar una tecla errónea en el ordenador. Si el número de categorías es grande y su tasa es rápida, habrá mayor probabilidad de fallos.
- Si el observador tiende a anotar erróneamente ciertas categorías, las medidas conductuales que obtendrá tenderán a ser sistemáticamente mayores o menores que las verdaderas.
- En tal caso:
 - La *precisión* del observador será baja.
 - Su *fiabilidad* será baja si se comparan sus datos con otro observador que no tiende cometer esa clase de errores sistemáticos.
- *Solución*: Los observadores deben ser *entrenados* previamente en el uso de la tecnología de adquisición de datos.

3.5 Sesgos de la observación

Deriva del observador:

- Tendencia del observador a desarrollar sus propias versiones de las definiciones de las categorías de conducta, deformando o tergiversando las definiciones originales.
- La deriva puede ocurrir o bien sin propósito específico o bien de modo intencionado (por ejemplo, porque el observador considere que las definiciones no son las adecuadas).
- La presencia de deriva hace disminuir la *precisión* del observador. También hace disminuir la *fiabilidad* si los otros observadores falibles con los que se compara no derivan o si derivan de forma diferente.
- ***Deriva consensuada:*** Dos observadores falibles desarrollan de común acuerdo sus propias versiones de las definiciones de las categorías, lo que da como resultado una *fiabilidad falsamente alta* cuando son comparados.
- *Solución:* Los observadores deben ser *re-entrenados* de forma periódica.

3.5 Sesgos de la observación

Complejidad de los sistemas de categorías:

- La complejidad es mayor:
 - Cuanto mayor es el número de categorías que se han definido.
 - Cuando se han definido sistemas multidimensionales de categorías o cuando varias categorías pueden ocurrir simultáneamente.
 - Cuando las definiciones de las categorías contienen términos interpretativos o inferenciales.
- Cuanta mayor complejidad, más difícil puede ser para el observador *discriminar* entre las categorías y, por lo tanto, más probable que cometa fallos al registrarlas.
- Por lo tanto, al elaborar los sistemas de categorías debe tenerse en cuenta que definir sistemas complejos con la finalidad de obtener información muy detallada sobre la conducta puede repercutir negativamente en la fiabilidad y la precisión.
- *Solución:* Cuando la complejidad es grande es necesario observar *grabaciones audiovisuales* y evitar la observación en vivo.

3.5 Sesgos de la observación

Fallos perceptivos:

- *Causados por la localización espacial del observador:* Cuanto más lejos se encuentre el observador de los individuos observados más probable es que no pueda detectar todas sus conductas.
 - *Soluciones:* Usar vídeo, micrófonos direccionales, binoculares.
- *Causados por la duración y tasa de los estímulos:* Cuanto más corta es la duración de las conductas y/o mayor es su tasa, más difícil es para el observador detectar sus ocurrencias y registrarlas correctamente.
 - *Solución:* Usar vídeo.
- *Causados por déficit de atención:* Cuanto más compleja sea la situación a observar (cantidad de categorías, tasas altas, muchos individuos a observar simultáneamente, etc.), más difícil es para el observador mantener la atención en la conducta durante largos periodos de tiempo y más probable es que cometa errores.
 - *Soluciones:* Usar vídeo, sesiones de observación cortas, muestreo intrasiesional de individuos.

3.5 Sesgos de la observación

Expectativas del observador:

- Existe un “*efecto de la expectativa*” cuando el observador obtiene datos falsos (de forma intencionada o no intencionada) que tienden a favorecer o a desfavorecer las hipótesis o conjeturas de la investigación.
- *Ejemplo:*
 - En una investigación sobre la agresión en las escuelas se informa al observador que se cree que en las escuelas de “clase alta” y “clase media-alta” existe una tasa de agresión menor entre los niños que en las escuelas de “clase media-baja” y “clase baja”.
 - Se han definido categorías de agresión “grave”, “intermedia” y “leve”.
 - Conociendo la conjetura, el observador puede tender a registrar determinadas interacciones que se consideran agresión “intermedia” (por ejemplo, empujones) como agresión “grave” en las escuelas de “clase baja” y como agresión “intermedia” en las de “clase alta”.
- *Solución:* Los investigadores *no deben informar* a los observadores acerca de las hipótesis o conjeturas de la investigación.

3.5 Sesgos de la observación

Control de los sesgos:

- Los observadores deben recibir *entrenamiento previo* en el uso de los sistemas de categorías y de las tecnologías de adquisición de datos.
- Deben ser *re-entrenados periódicamente* en el curso de la investigación.
- Debe calcularse su *precisión* (o su fiabilidad) de forma periódica y deben ser *informados* de la misma. Conocer que la precisión no es buena es un incentivo para mejorarla.
- Los sistemas de categorías deben ser manejables y no demasiado complejos. El vídeo facilita registrar la conducta empleando sistemas complejos.
- Debe elaborarse un *manual de uso* de las categorías.
- Deben diferenciarse claramente las funciones:
 - El *equipo de investigación* planifica la investigación, establece conjeturas, define variables y categorías, entrena a los observadores, analiza los datos, etc.
 - El *observadores* únicamente registran la conducta pero no deben conocer las hipótesis y los objetivos de la investigación.

3.5 Sesgos de la observación

Reactividad de los individuos observados:

- La reactividad es el cambio que se produce en la conducta de los individuos observados por el hecho de que saben o sospechan que son observados.
- Cuando existe reactividad el observador no obtiene un registro de la *conducta espontánea* de los individuos [que es lo que se pretendía] sino de su *conducta como reacción a la situación de observación*.
- Por lo tanto, los datos no reflejarán aquello que se pretendía que reflejasen: **El estudio carecerá de validez.**
- **Indicadores** de posible reactividad:
 - Los individuos observados se enfrentan al observador, se resisten a ser observados, se esconden, sobreactúan, miran a la cámara de vídeo o al observador.
 - Aparecen ciertas conductas inesperadas o conductas que eran esperables no se observan, en comparación con otros estudios similares.

3.5 Sesgos de la observación

Factores que generan reactividad:

- **Los observadores:** Sus rasgos físicos (talla, peso, sexo, procedencia, etnia, etc.) y su conducta pueden afectar la conducta espontánea de los individuos observados.
- **Los observados:** Hay individuos más reactivos que otros, desde los más sensibles a los más ingenuos, en función de su experiencia previa como sujetos de observación. También existen diferencias en la actitud ante el hecho de ser observado.
- **Diferencias entre los observadores y los observados:** Cuanto más diferentes sean observadores y observados, más reactivos pueden ser estos. *Por ejemplo*, se espera mayor reactividad en hombres que son observados por una mujer que en hombres que son observados por un hombre.
- **Las conductas:** Las conductas voluntarias, la conducta verbal y la conducta social en general son más proclives a cambiar por efecto de la reactividad que las conductas automáticas, motrices, posturales y causadas por mediadores hormonales (p.e., respuesta a una agresión).

3.5 Sesgos de la observación

Control de la reactividad:

- **Observador oculto:** La mejor solución pero no siempre factible. El observador se encuentra en el mismo entorno que los observados pero éstos no pueden verle. Típico de observación de especies animales en su entorno natural. Puede no ser ético en observación de humanos.
- **Uso de espejos unidireccionales:** Los individuos observados se hallan en una habitación en cuya pared hay un espejo unidireccional, y el observador se encuentra fuera de la habitación viéndolos a través del espejo. Los observados *pueden sospechar* que están siendo observados, pero no lo saben con certeza [esto puede atenuar su reactividad]. Típico de observación de la interacción infantil.
- **Habitación a la observación:** La presencia del observador es evidente y se espera que los observados se habitúen a ella tras un periodo de tiempo, a partir del cual su reactividad se reduzca o desaparezca.
- En general, hay que *evitar informar* a los individuos observados que van a ser observados. En ciertas situaciones es posible emplear *múltiples observadores*, sin informar a los observados cuáles están realmente observando ni a quién.