Licenciatura de Psicopedagogía:

Métodos, Diseños y Técnicas de Investigación Psicológica

Salvador Chacón Moscoso

Susana Sanduvete Chaves

Dpto. de Psicología Experimental.

Universidad de Sevilla.

TEMA 9
FIABILIDAD DE LAS PUNTUACIONES

Fiabilidad de las puntuaciones

FIABILIDAD

Criterio de calidad relacionado con la precisión de las medidas obtenidas con un test y que proporciona información acerca de:

- **grado en que las puntuaciones empíricas son reflejo de las verdaderas.**
- **grado de ajuste entre puntuaciones empíricas y verdaderas.**

FIABILIDAD

DEFINICIÓN:

En general, podemos afirmar que una medida es fiable si es precisa, estable y consistente.

En este sentido, podemos definir la fiabilidad como el grado de ajuste entre la medición obtenida del comportamiento externo del sujeto (puntuación observada X) y lo que tiene el sujeto del atributo o constructo que en realidad mide el test (rasgo latente ξ).

FIABILIDAD

PROCEDIMIENTOS:

Los procedimientos y técnicas para contrastar la fiabilidad se basan en el modelo lineal de regresión y utilizan como índices el coeficiente de correlación de Pearson, atendiendo a:

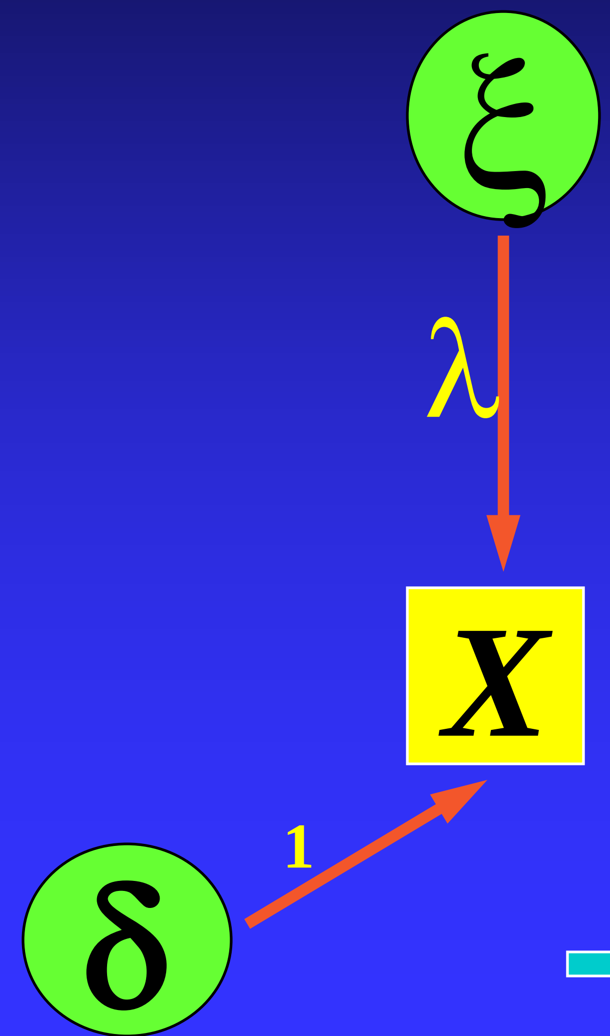
- **la estabilidad temporal de las puntuaciones obtenidas con el test (test-retest y formas paralelas)**
- **la consistencia interna del test (alpha y dos mitades)**

Teoría Clásica de los Tests (Spearman)

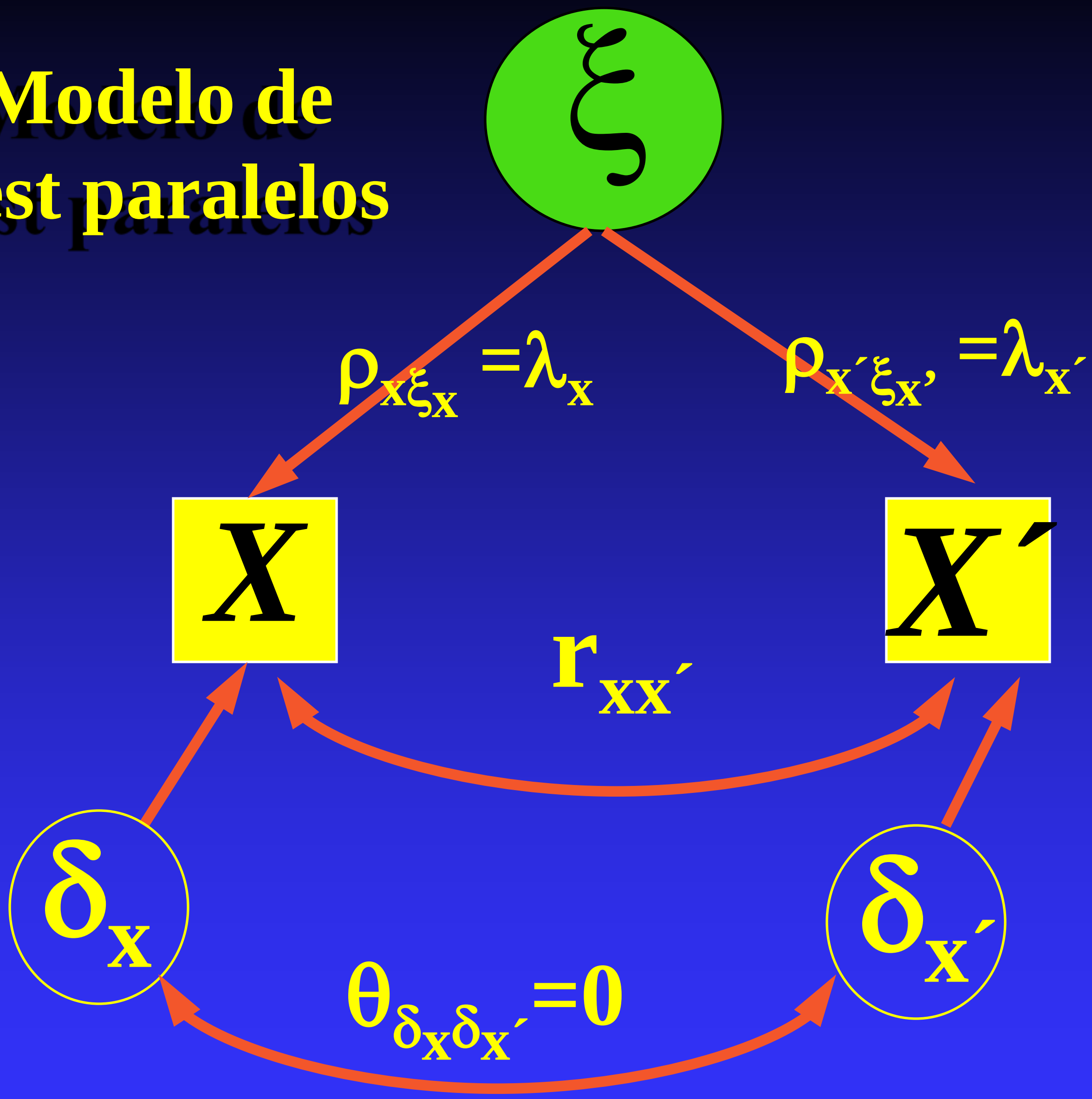
$$X = V + \text{error}$$

Modelo lineal

$$\boldsymbol{X} = \lambda\,\xi + \boldsymbol{\delta}$$



Modelo de test paralelos



$$\hat{\rho}_{xx'} = r_{XX'} = \rho_{X\xi_x} \cdot \rho_{X\xi_{x'}} = \lambda_X \cdot \lambda_{X'} = \lambda_X^2$$

Ello implica:

$$\left.\begin{array}{l} x = \lambda_x \xi + \delta_x \\ x' = \lambda_x \xi + \delta_{x`'} \end{array}\right\} \Rightarrow \begin{array}{l} \lambda_x = \lambda_{x'} \\ \theta_{\delta_x} = \theta_{\delta_{x'}} \end{array}$$

$$\widehat{\rho}_{xx'} = \lambda_x . \lambda_{x'}$$

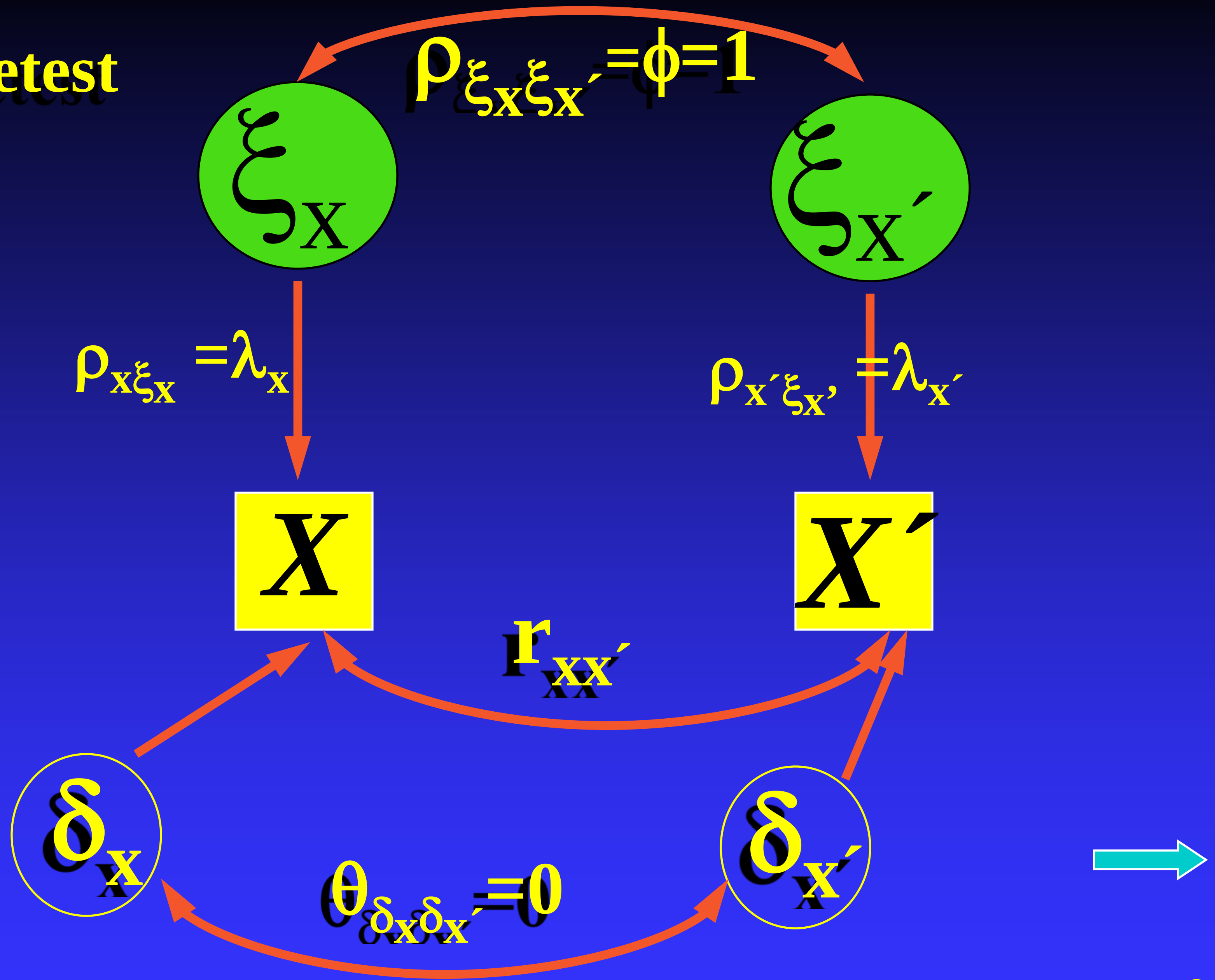
Test-retest

- Consiste en aplicar el mismo test al mismo grupo de individuos en dos ocasiones distintas.

- Posteriormente se calcula el coeficiente de correlación de Pearson entre las dos series de puntuaciones obtenidas.

- En la medida que dicha correlación sea alta entenderemos que el test mide de forma estable a los sujetos estudiados y que, por tanto, presenta una alta fiabilidad.

test-retest



$$\hat{\rho}_{xx'} = r_{XX'} = \rho_{X\xi_x}\cdot\rho_{\xi_x\xi_{x'}}\cdot\rho_{X\xi_{x'}} = \lambda_X\cdot\phi\cdot\lambda_{X'} = \lambda_X\cdot\lambda_{X'} = \lambda_X^2$$

Amenazas a la validez del cálculo Test-retest Caso A:

- En general el coeficiente empírico de fiabilidad tendrá un valor más alto (sobrevalorado) que el que corresponde a la verdadera precisión de la medida, si el tiempo transcurrido entre dos pasadas del mismo test es corto.

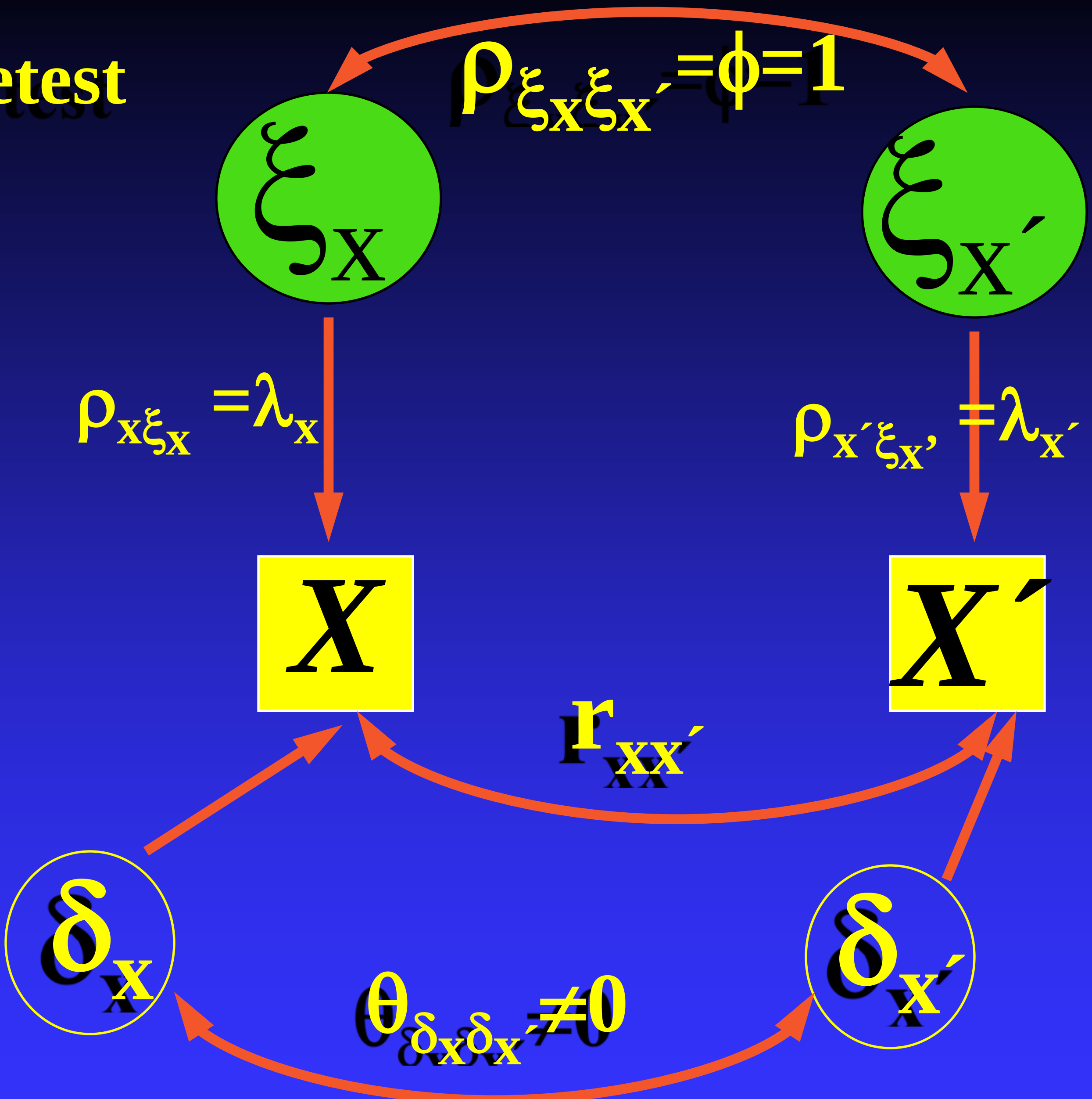
- Pudiera ocurrir que los sujetos den respuestas apoyadas en el recuerdo o aprendizaje previo del propio instrumento de medida

Test-retest Caso A:

- El efecto estable de estas variables extrañas (no contempladas) daría lugar a **errores estables** y **no aleatorios** lo cual produciría una **sobre-estimación** de la fiabilidad (caso A figura siguiente). En este caso el cálculo del coeficiente de fiabilidad viene dado por:

$$\hat{\rho}_{xx'} < r_{XX'} = \lambda_X \cdot \phi \cdot \lambda_{X'} + \theta_{\delta_X \delta_{X'}} > \lambda_X^2$$

test-retest



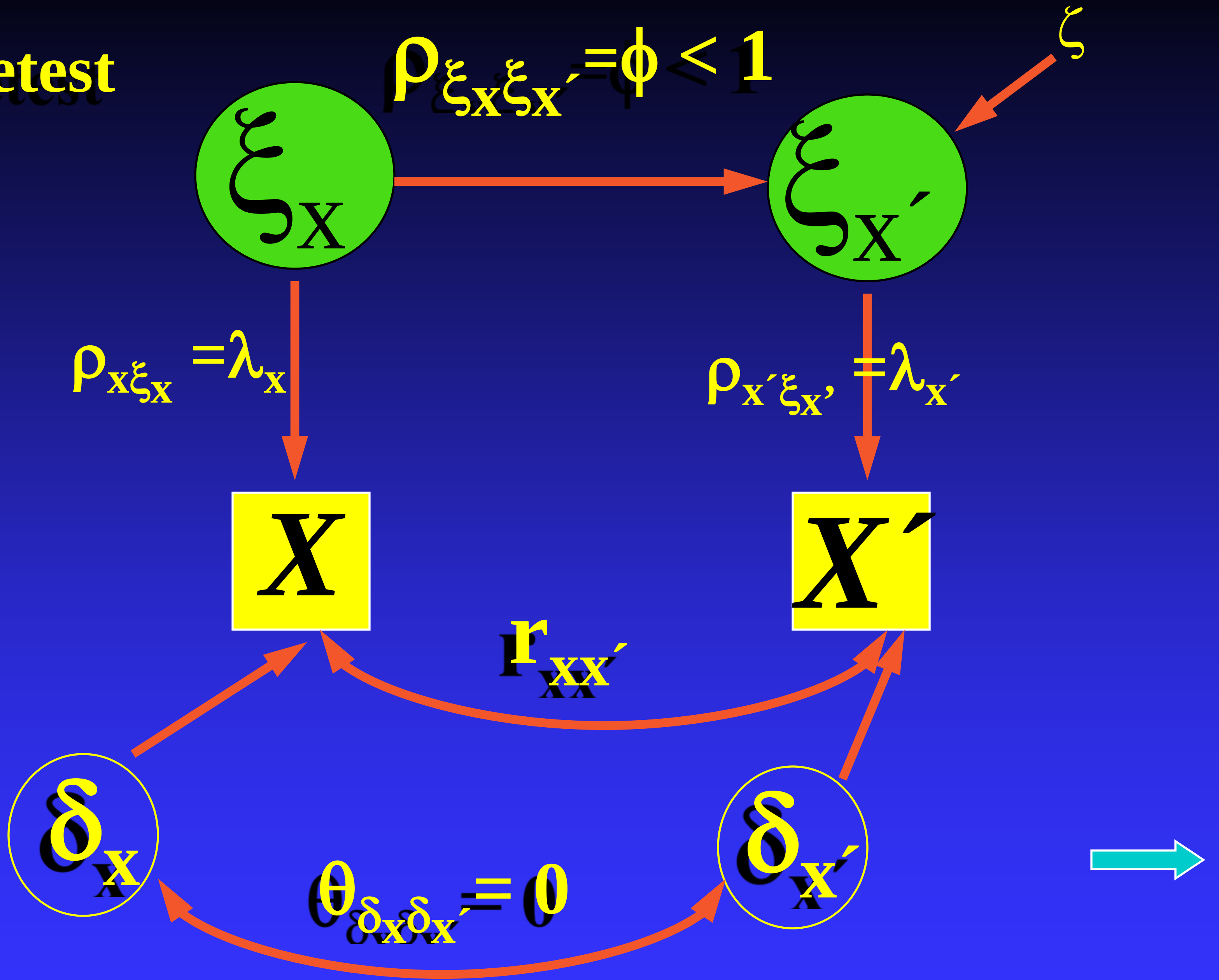
$$\hat{\rho}_{xx'} < r_{XX'} = \lambda_X \cdot \phi \cdot \lambda_{X'} + \theta_{\delta_X \delta_{X'}} > \lambda_X^2$$

Test-retest Caso B:

- **En general el coeficiente empírico de fiabilidad tendrá un valor más bajo (infravalorado) que el que corresponde a la verdadera precisión de la medida, si las dos pasadas están distanciadas en el tiempo de manera que exista una variación en las puntuaciones verdaderas de los sujetos (maduración) y, por tanto, estén relacionadas en un cierto grado (caso B figura siguiente). En este caso el cálculo del coeficiente de fiabilidad viene dado por:**

$$\hat{\rho}_{xx'} > r_{XX'} = \lambda_X \cdot \phi \cdot \lambda_{X'} < \lambda_X^2$$

test-retest



$$\hat{\rho}_{xx'} > r_{XX'} = \lambda_X \cdot \phi \cdot \lambda_{X'} < \lambda_X^2$$

Test-retest. Resumen:

- Por estas razones, podemos considerar que el método test-retest es un procedimiento adecuado cuando se miden rasgos que varían poco en el tiempo, en donde la práctica no ejerce especial influencia y procurando que el intervalo temporal no sea ni excesivamente corto ni muy largo (no existe una respuesta única respecto a cuánto es el tiempo adecuado, depende del tipo de prueba realizada).

- Pruebas tales como atención, rapidez perceptiva, cálculo numérico, etc.. pueden ser adecuadas para ser sometidas al método del test-retest.

Formas paralelas

- Este método consiste en aplicar a un mismo grupo de sujetos dos formas paralelas (dos versiones) del mismo test.

- Dos formas se dicen que son *paralelas* cuando miden los mismos aspectos con el *mismo tipo* de cuestiones (que no iguales).

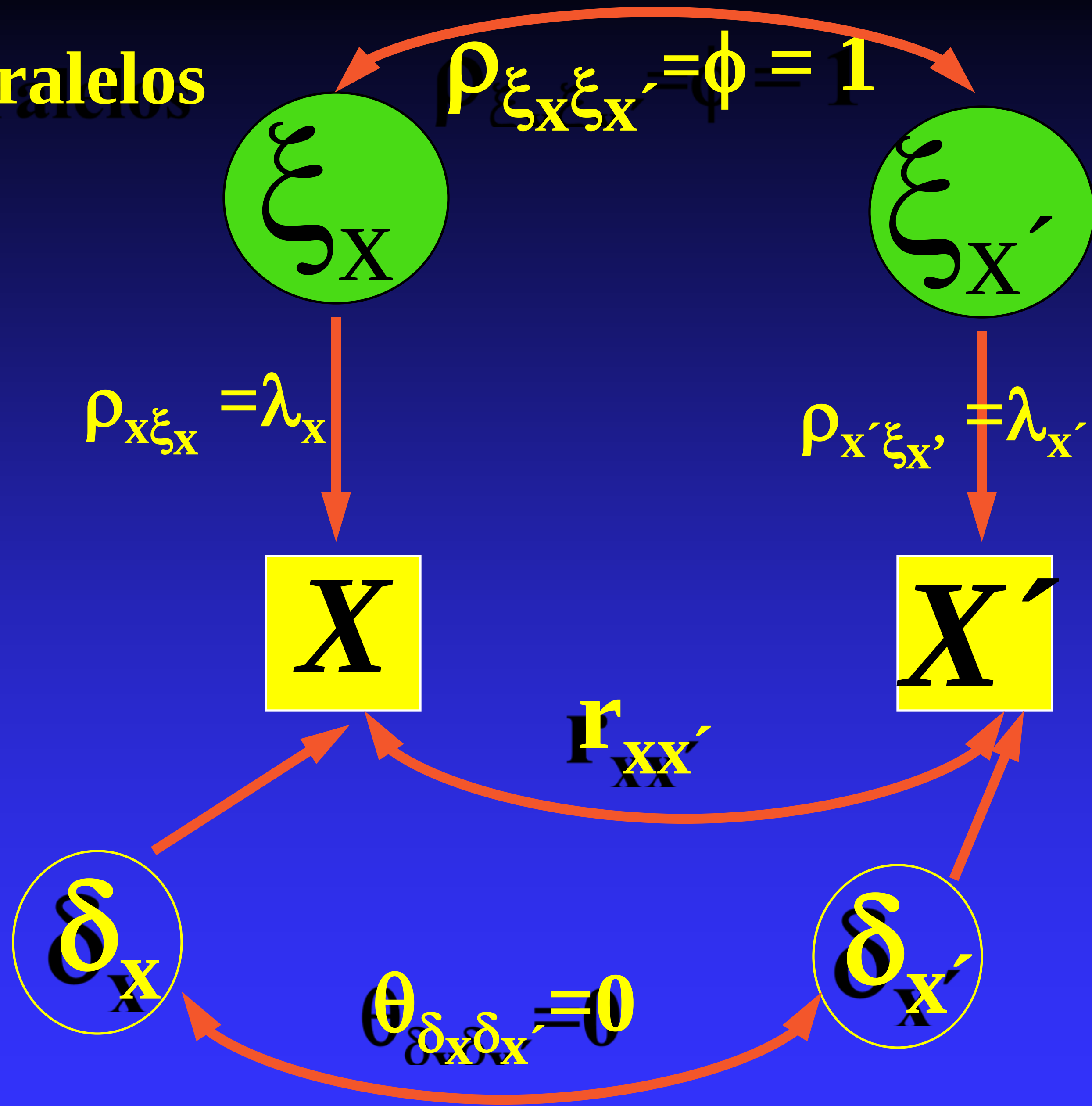
- Posteriormente se calcula el coeficiente de correlación de Pearson entre las dos series de puntuaciones obtenidas con las dos formas.

Formas paralelas

- En la medida que dicha correlación sea alta entenderemos que los dos test miden de mismo modo a los sujetos estudiados y que, por tanto, presentan una alta fiabilidad.

- La característica de las pruebas paralelas es que aún estando constituidas por ítems diferentes, éstos, uno a uno, han de medir el mismo rasgo y de la misma manera, lo que implica igualdad de medias, varianzas y covarianzas entre ambos tests. Lo cual se representa en el diagrama de la figura siguiente.

tests-paralelos



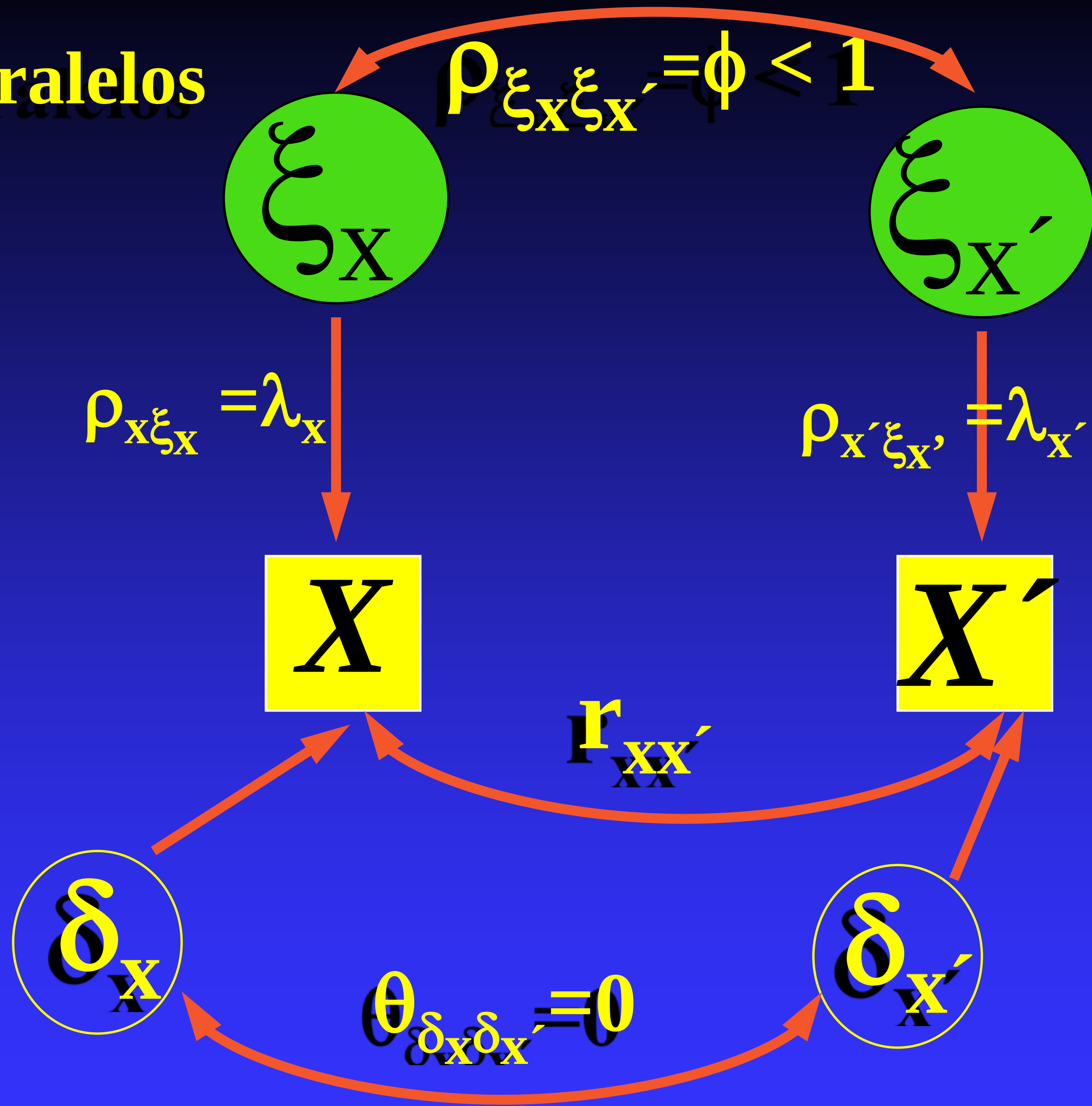
$$\widehat{\rho}_{xx'} = r_{XX'} = \rho_{X\xi_x} \cdot \phi \cdot \rho_{X\xi_{x'}} = \lambda_X \cdot \phi \cdot \lambda_{X'} < \lambda_X^2$$

Formas paralelas

- Es evidente la dificultad que entraña lograr formas *exactamente* paralelas, por lo que parece más razonable hablar de *formas alternativas* mas bien que de formas paralelas, donde se entiende que ambos tests no son por completo equivalentes, sino dos *intentos* de que lo sean.

- La figura siguiente refleja esta situación Se observa que la fiabilidad del test, en este caso, depende no sólo de la ausencia de errores en la medida, sino también del grado de similitud de ambas formas. Así:

tests-paralelos



$$\hat{\rho}_{xx'} > r_{XX'} = \rho_{X\xi_x} \cdot \rho_{\xi_x\xi_{x'}} \cdot \rho_{X\xi_{x'}} = \lambda_X \cdot \phi \cdot \lambda_{X'} < \lambda_X^2$$

Test paralelos. Resumen:

- El método de las formas paralelas sería el procedimiento idóneo si no fuera por la dificultad que entraña la elaboración de pruebas realmente equivalentes.

- Cuando las formas no son totalmente paralelas es difícil distinguir lo que es cambio en la puntuación verdadera de la falta de fiabilidad (Carmines y Zeller,1988 pág. 40).

Test paralelos. Resumen:

- No obstante, este procedimiento para el cálculo del coeficiente de fiabilidad (en la versión que hemos apuntado como de formas alternativa) presenta algunas ventajas respecto al método del test-retest.

- Al tratarse de formas diferentes, no existe el efecto de memoria anteriormente señalado. Por esta misma razón los errores de medida entre ambas aplicaciones tendrán menos probabilidad de estar correlacionados.

Test paralelos. Resumen:

- Algunas dificultades permanecen aún vigentes con este procedimiento.
- Puede presentarse reactividad al test de forma tal que en la segunda pasada haya cambiado la actitud del sujeto, y por otro lado, se mantienen algunos efectos del intervalo temporal entre ambas aplicaciones del test.
- Si el período de tiempo es corto, cierto influjo puede traslucirse (más por efecto de práctica que de memoria) y si el intervalo es largo puede, como el caso del test-retest, cambiar el rasgo a medir del sujeto. Todo ello puede contribuir, al igual que la falta de paralelismo entre las formas del test, a reducir la correlación entre ellos.

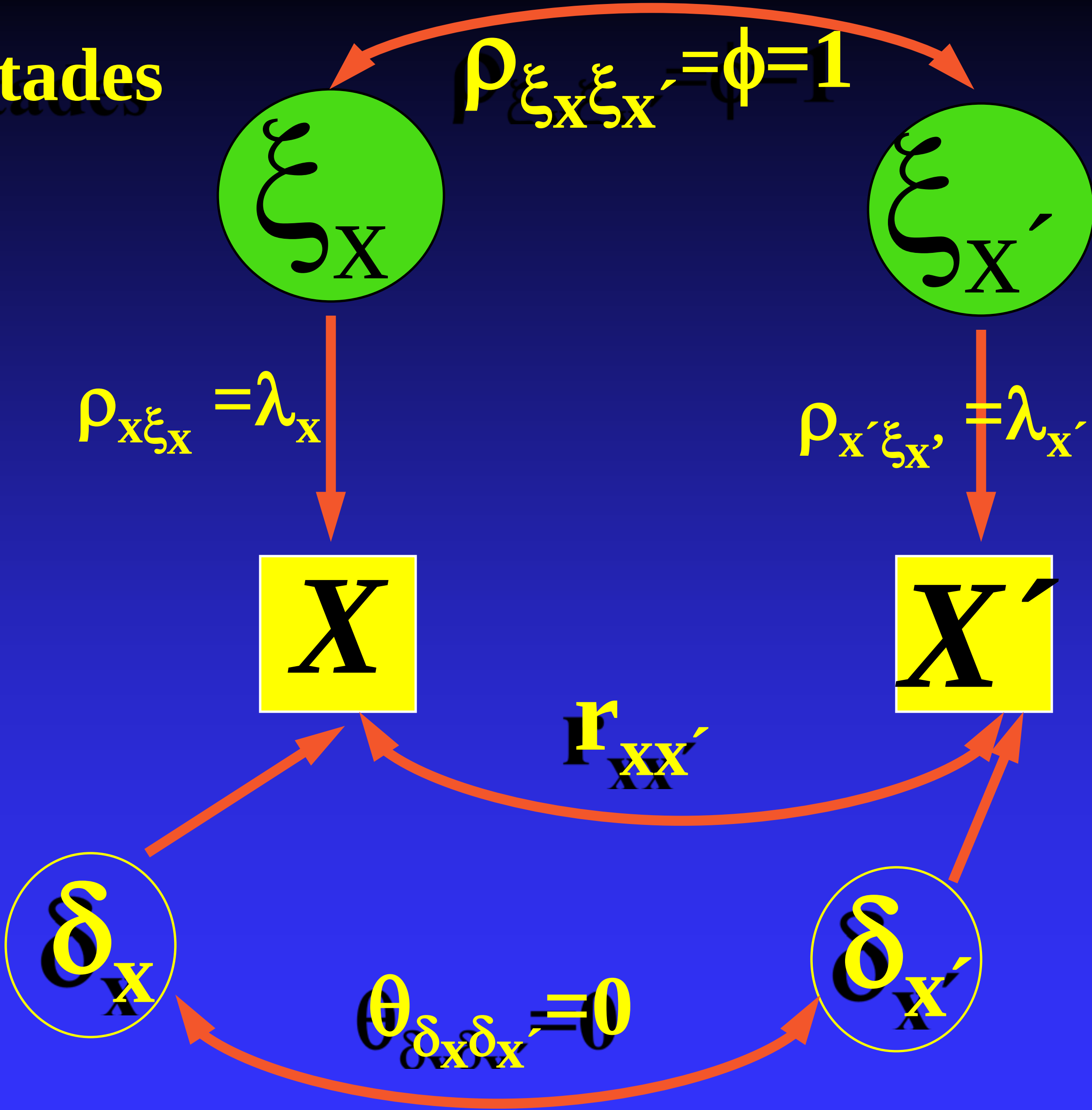
Dos mitades

- Este método consiste en aplicar un único test a un mismo grupo de individuos representativos de una cierta población.

- A continuación se divide el test en dos mitades; es decir, la mitad de los ítems configura uno de los tests, y la otra mitad de los ítems el otro test.

- La correlación de Pearson aplicada con las puntuaciones totales de ambas mitades constituirá el coeficiente de fiabilidad.

Dos mitades



$$\hat{\rho}_{xx'} = r_{XX'} = \rho_{X\xi_x} \cdot \rho_{\xi_x\xi_{x'}} \cdot \rho_{X\xi_{x'}} = \lambda_X \cdot \phi \cdot \lambda_{X'} = \lambda_X \cdot \lambda_{X'} = \lambda_X^2$$

Calculo del coeficiente de fiabilidad:
(Formula de atenuación de Spearman-Brown)

$$\hat{\rho}_{xx'} = R_{XX'} = \frac{2r_{XX'}}{1 + r_{XX'}}$$

Donde:

$r_{XX'}$= Coeficiente de fiabilidad del test mitad (longitud mitad)

$R_{XX'}$ = Coeficiente de fiabilidad del test completo (longitud doble que el test mitad)

Variantes para el cálculo del coeficiente de fiabilidad mediante el procedimiento de las dos mitades

Fórmula de Rulón:

$$\hat{\rho}_{xx'} = 1 - \frac{\sigma_d^2}{\sigma_X^2}$$

Donde:

σ_d^2 = varianza de las diferencias entre las puntuaciones pares e impares

σ_X^2 = varianza del test

Variantes para el cálculo del coeficiente de fiabilidad mediante el procedimiento de las dos mitades

Fórmula de Flanagan y Guttman:

$$\hat{\rho}_{xx'} = 2\left(1 - \frac{\sigma_p^2 + \sigma_i^2}{\sigma_X^2}\right)$$

Donde:

σ_p^2 = varianza de las puntuaciones de los ítems pares

σ_i^2 = varianza de las puntuaciones de los ítems impares

σ_X^2 = varianza del test

Dos mitades. Resumen

- Este procedimiento recoge las ventajas principales de los dos métodos anteriores, al mismo tiempo, que elimina sus mayores inconvenientes.

- El simple hecho de ser una única aplicación supone un cierto ahorro de esfuerzo, tiempo y costo económico.

- Por otro lado, al aplicarse simultáneamente formas distintas de un mismo test, suprimimos el efecto del intervalo temporal a la par que eliminamos el posible efecto de memoria.

Dos mitades. Resumen

- La principal desventaja del método de las dos mitades consiste en que el valor del coeficiente de fiabilidad obtenido depende de cómo hayan quedado repartidos los ítems en cada una de las mitades.

- Posiblemente, tendríamos tantos coeficientes de fiabilidad para un mismo test como formas posibles de agrupaciones (pares-impares, primera-segunda mitad, dos mitades aleatorias), lo que es lo mismo que decir que dicho test carece de un verdadero coeficiente de fiabilidad.

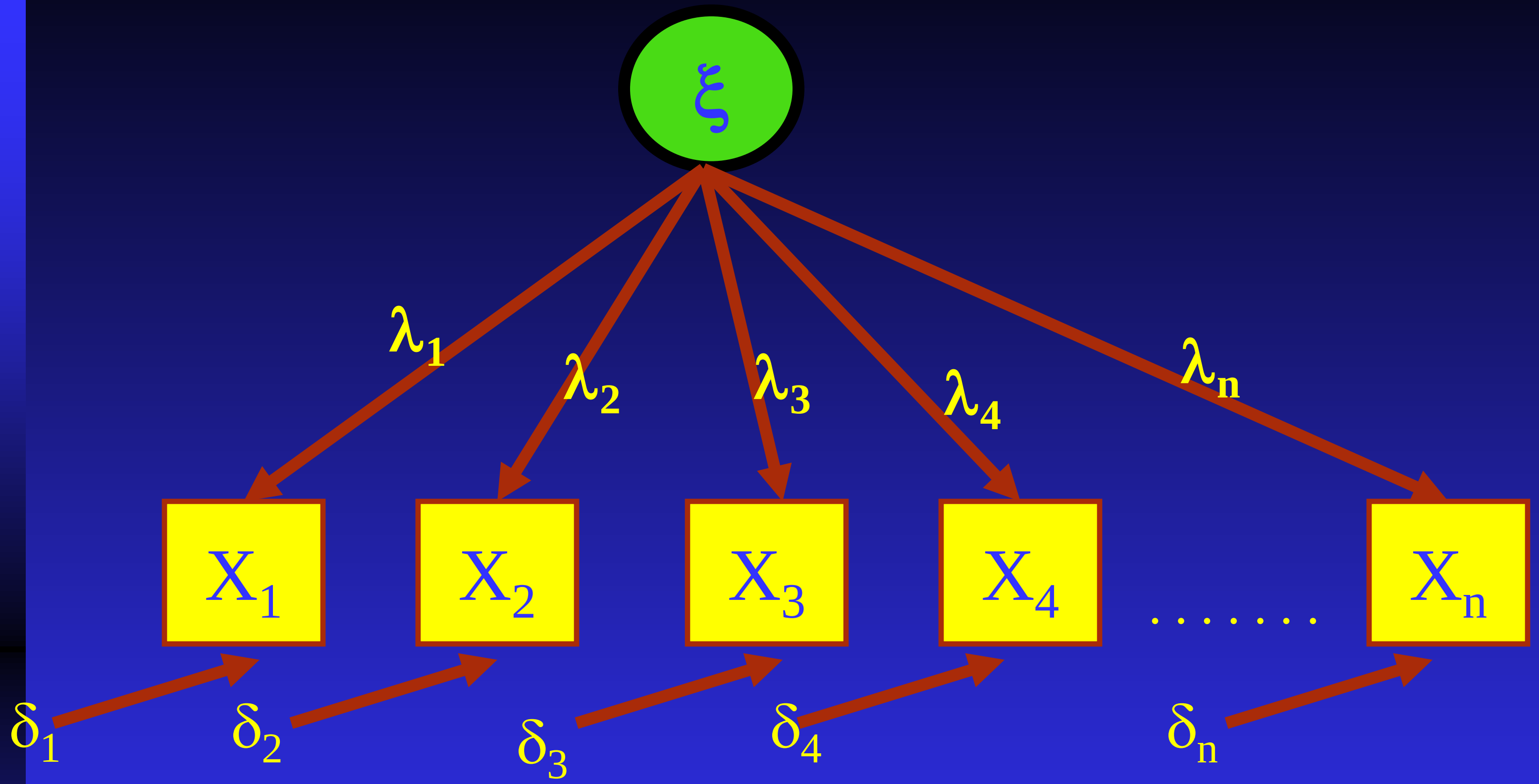
Coeficiente alpha de Cronbach

- Este método determina la fiabilidad de un test, en cuanto consistencia interna.

- Está basado en la correlación media entre todos los ítems de un test.

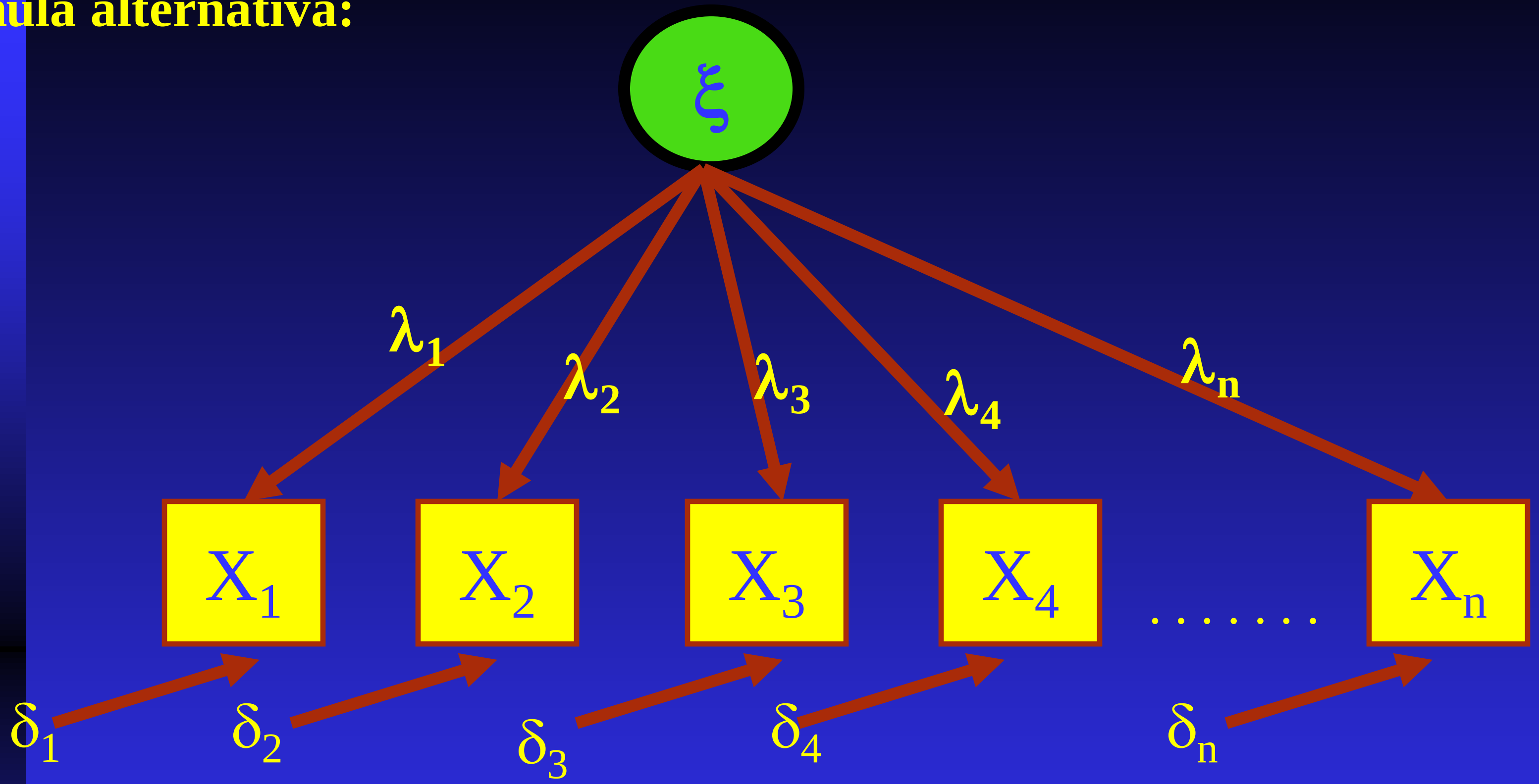
- Para su cálculo, se procede a considerar cada ítem del test como si fuera un test de longitud unidad y, acto seguido, calcular la correlación media de dichos ítems entre sí.

- Por último se aplica la fórmula de Spearman-Brown, que permite determinar la fiabilidad de un test de longitud n veces superior a la de los tests de longitud unidad (el test total tiene *n* ítems).



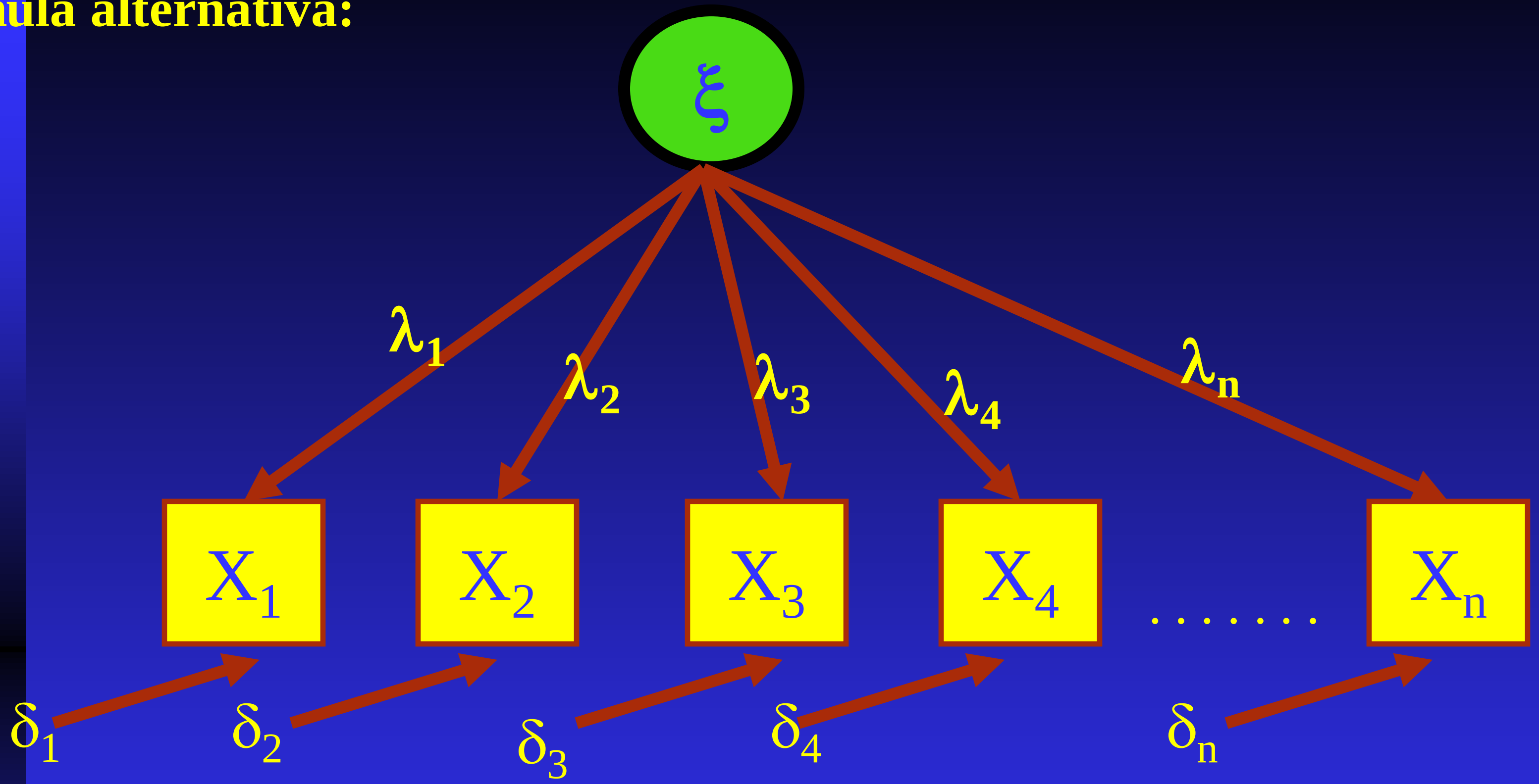
$$\hat{\rho}_{xx'} = \alpha = \frac{n\bar{\rho}_{xx'}}{1+(n-1)\bar{\rho}_{xx'}}$$

Formula alternativa:



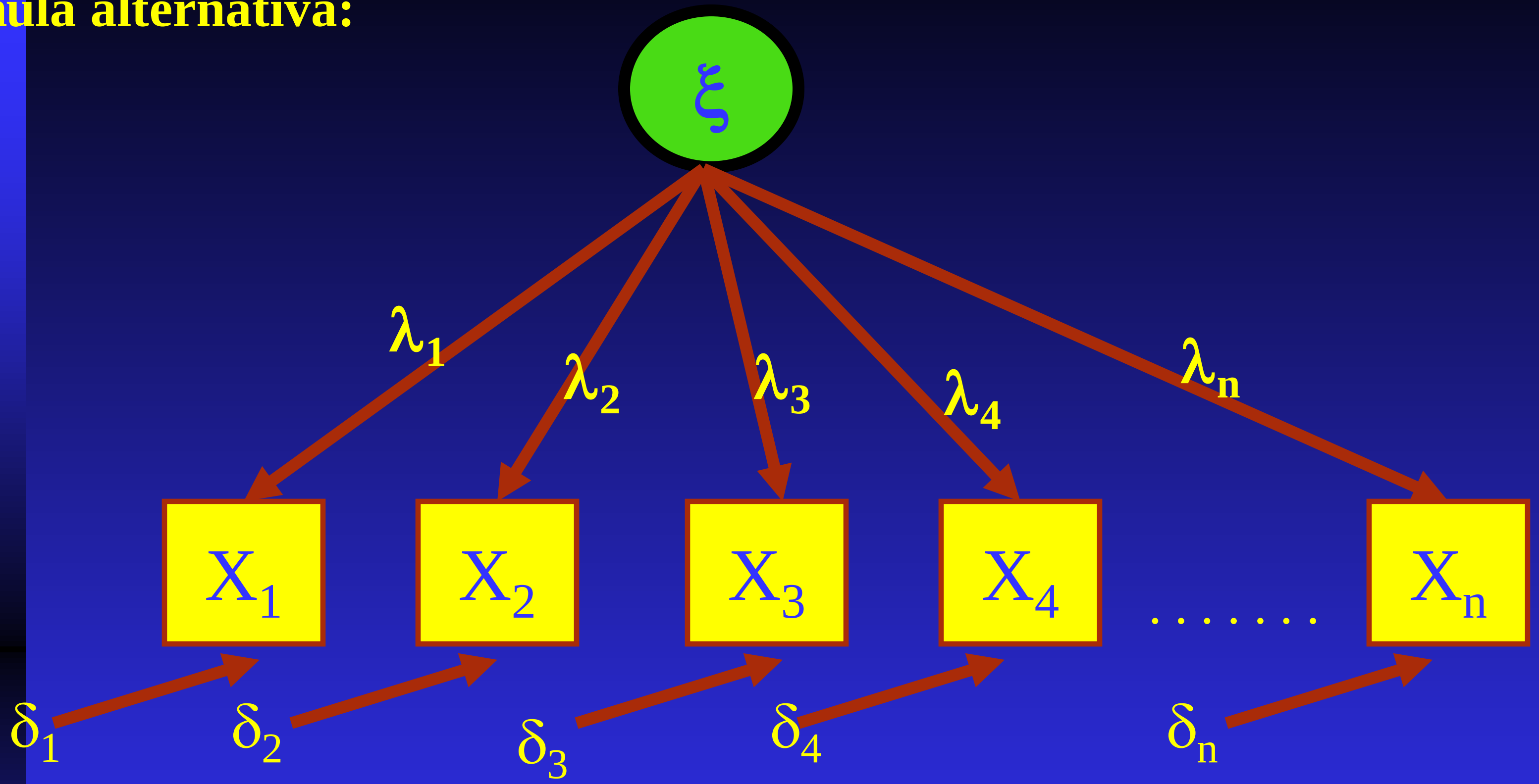
$$\widehat{\rho}_{xx'} = \alpha = \frac{n\lambda^2}{n\lambda^2 + \theta_\delta} \qquad ó \qquad \widehat{\rho}_{xx'} = \alpha = \frac{(n\lambda)^2}{(n\lambda)^2 + n\theta_\delta}$$

Formula alternativa:



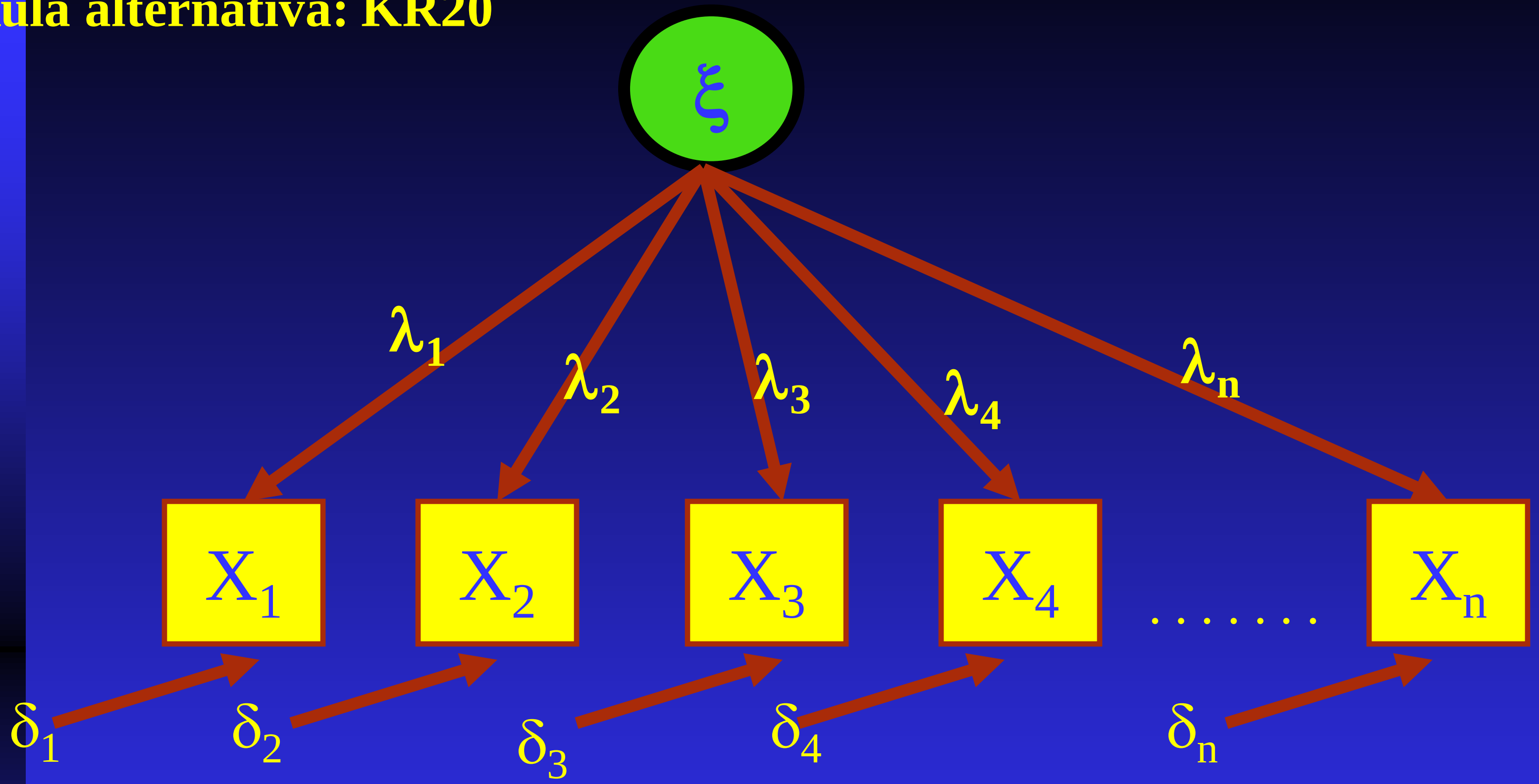
$$\hat{\rho}_{xx'} = \alpha = \frac{\left(\sum_{j=1}^{n} \lambda_j\right)^2}{\left(\sum_{j=1}^{n} \lambda_j\right)^2 + \sum_{j=1}^{n} \theta_{\delta_{jj}}}$$

Formula alternativa:



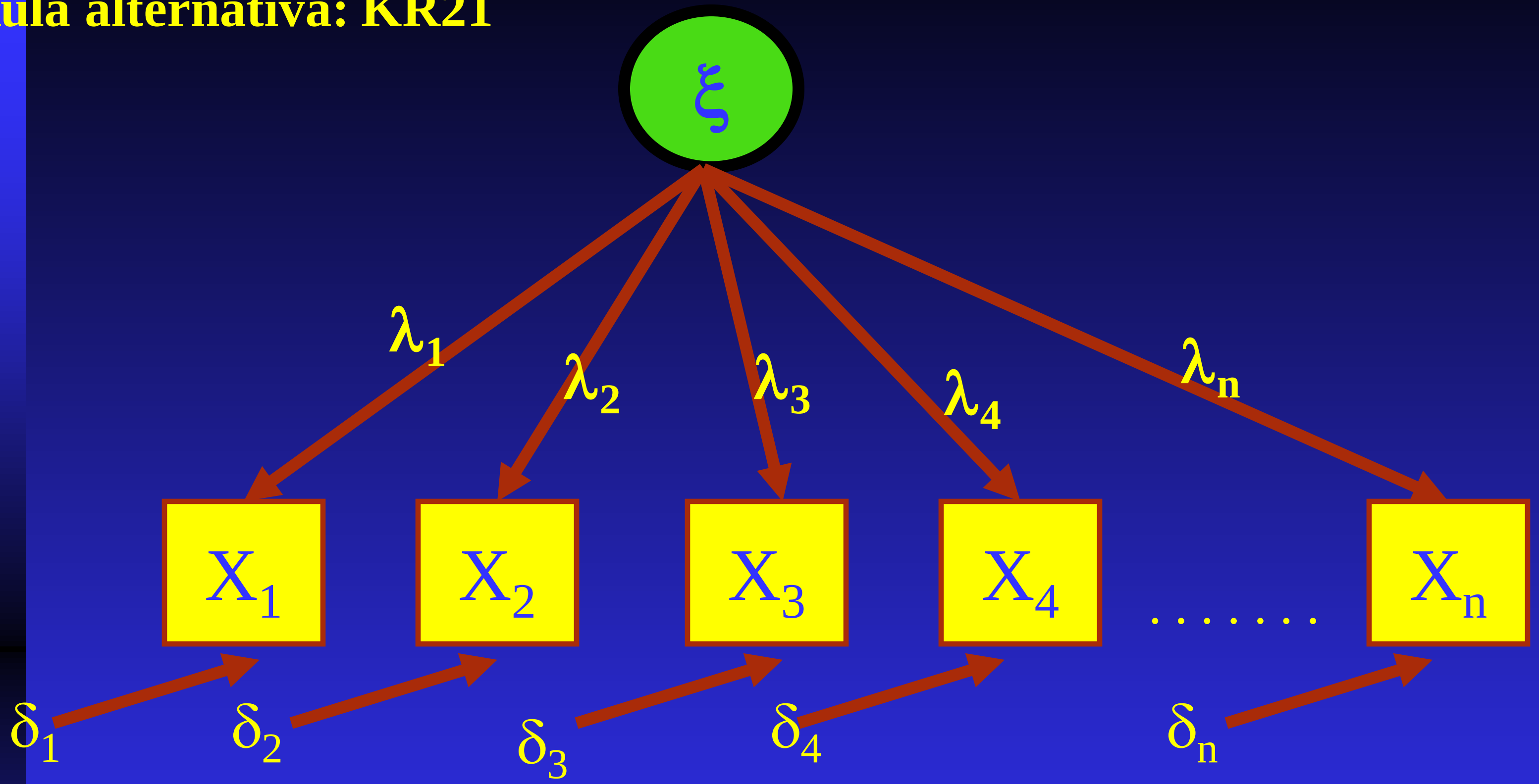
$$\widehat{\rho}_{xx'} = \alpha = \frac{n}{n-1}\left(1 - \frac{\sum_{j=1}^{n}\sigma_j^2}{\sigma_x^2}\right)$$

Formula alternativa: KR20



$$\hat{\rho}_{xx'} = \alpha = KR_{20} = \frac{n}{n-1}\left(1 - \frac{\sum_{j=1}^{n} p_j q_j}{\sigma_x^2}\right)$$

Formula alternativa: KR21



Para ítems con igual dificultad:

$$\hat{\rho}_{xx'} = \alpha = KR_{21} = \frac{n}{n-1}\left(1 - \frac{\bar{X} - \frac{\bar{X}^2}{n}}{\sigma_x^2}\right)$$

Factores que afectan al coeficiente de fiabilidad

- Longitud del test,
- Variabilidad de la muestra.
- Limitación del tiempo y
- Características de los ítems.

- **Longitud del test:**

Formula de atenuación de Spearman-Brown

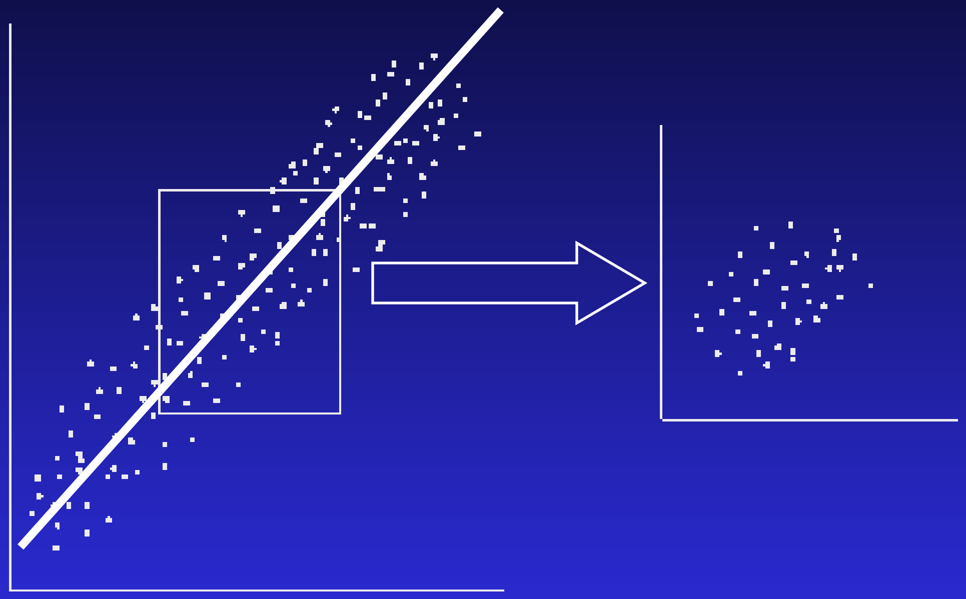
$$R_{XX'} = \frac{k\, r_{xx'}}{1 + (k - 1)\, r_{xx'}}$$

Donde,
$R_{XX'}$ = fiabilidad final lograda a partir de la fiabilidad inicial $r_{xx'}$
k = razón o número de veces que el test resultante contiene la longitud del test original.

k= L/l

Donde,
L= longitud final del test
l=longitud inicial del test.

Efecto de la variabilidad sobre la correlación



$$\sigma^2_{e_1} = \sigma^2_{x_1}\left(1 - \rho_{11'}\right)$$

$$\sigma^2_{e_2} = \sigma^2_{x_2}\left(1 - \rho_{22'}\right)$$

$$\rho_{22} = 1 - \frac{\sigma_1^2}{\sigma_2^2}\left(1 - \rho_{11'}\right)$$

ρ_{11}=coeficiente de fiabilidad en la población 1
ρ_{22}= coeficiente de fiabilidad en la población 2
$\sigma^2{}_1$=varianza empírica en la población 1
$\sigma^2{}_2$=varianza empírica en la población 2

Estimación de la puntuación verdadera de los sujetos en el atributo de interés

En el marco de la Teoría Clásica de los Tests, se han desarrollado varios métodos para estimar la Puntuación Verdadera a partir de la Observada. Según Charter y Feltd (2001), existen por lo menos cinco aproximaciones posibles, aunque sólo dos de ellas tienen un fundamento teórico lo suficientemente riguroso como para recomendar su uso: la Aproximación tradicional, que utiliza el EEM para construir un intervalo de confianza en torno a la Puntuación Observada asumiendo normalidad en la distribución de los errores, y la Aproximación Basada en Regresión (*Fórmula de Kelley*), que permite una estimación puntual o intervalar de la Puntuación Verdadera, a partir de la Puntuación Observada y la media del grupo de referencia.

Aproximación Tradicional para estimar la Puntuación Verdadera

Esta es la técnica más conocida y probablemente muchos usuarios crean que es la única disponible para estimar la Puntuación Verdadera. Se trata de un procedimiento ampliamente recomendado por muchos autores (e.g., Allen y Yen, 1979; Anastasi y Urbina, 1998; Felt y Brennan, 1989; Gulliksen, 1950) que consiste en utilizar el EEM para construir un intervalo de confianza en torno a la Puntuación Observada.

Fundamento

Teóricamente, el procedimiento se fundamenta en asumir que los errores de medida se distribuyen normalmente. Bajo este supuesto es posible utilizar el EEM para construir un intervalo de confianza para la Puntuación Verdadera en torno a la Puntuación Observada.

Sin profundizar en la derivación de la fórmula (para tal efecto ver, por ejemplo, Muñiz, 2001), las ecuaciones para el límite inferior y superior del intervalo de confianza corresponden a:

En ambas ecuaciones *X* es la Puntuación Observada de un sujeto y *Z* representa el valor de la distribución normal asociado a la magnitud del intervalo de confianza que se desea construir. La práctica convencional es construir un intervalo no direccional de un 95% de confianza, al que corresponde un valor *Z*=1.96.

Estimación de la puntuación verdadera mediante regresión lineal

Este procedimiento, originalmente desarrollado por Kelley (1947, p.409) permite obtener una estimación puntual de la Puntuación Verdadera y además construir un intervalo de confianza en torno a ella. Para estimar la Puntuación Verdadera es necesario conocer la fiabilidad y el promedio alcanzado en el test por el grupo de referencia en el cual se calculó la fiabilidad de las puntuaciones observadas. Para construir el intervalo de confianza se requiere contar, además, con la desviación estándar de las puntuaciones observadas.

Estimación de la puntuación verdadera mediante regresión lineal

Fundamento

Esta aproximación se basa en un principio simple: la Puntuación Observada guarda una relación lineal con la Puntuación Verdadera, la que puede modelarse por una recta de regresión lineal, cuya pendiente depende de la fiabilidad de la prueba. La fundamentación del procedimiento puede consultarse en Muñiz (2001) y una exposición simplificada en Gempp (2006). En la práctica, la Puntuación Verdadera *V* es predicha a partir de una Puntuación Observada *X*, conociendo la fiabilidad de la prueba y la media de las puntuaciones de la muestra normativa, mediante la ecuación:

$$v = fiabilidad(x - \bar{X}) + \bar{X}$$

Estimación de la puntuación verdadera mediante regresión lineal

La estimación la Puntuación Verdadera *V* de un sujeto cualquiera es sencilla. Primero es necesario restar a la Puntuación Observada para el sujeto (*X*) el promedio obtenido por el grupo. El resultado debe multiplicarse por la fiabilidad y al producto resultante sumarle el valor del promedio grupal.
Analizando la ecuación anterior se deducen varias características de la Puntuación Verdadera estimada que concuerdan con los principios básicos de la TCT. Por ejemplo, que las puntuaciones verdaderas y observadas tienen la misma media. Además, que si la fiabilidad es igual a 1 (es decir, fiabilidad perfecta y por tanto ausencia de error de medida) las puntuaciones observadas equivalen a las verdaderas para cualquier valor de la distribución. Por último, también es evidente que si la fiabilidad disminuye las puntuaciones observada y verdadera serán más discrepantes para los valores más alejados de la media.
En la práctica, las puntuaciones verdaderas estimadas con este método muestran un típico efecto de regresión a la media: cuando la Puntuación Observada es *mayor que* el promedio de la distribución, la correspondiente Puntuación Verdadera estimada será *menor que* la Puntuación Observada. Lo contrario sucederá cuando la Puntuación Observada sea *menor que* la media.