



UNAM
POSGRADO



Programa
Universitario
de Estudios
del Desarrollo
UNAM

Confiabilidad y la Teoría Clásica del Test

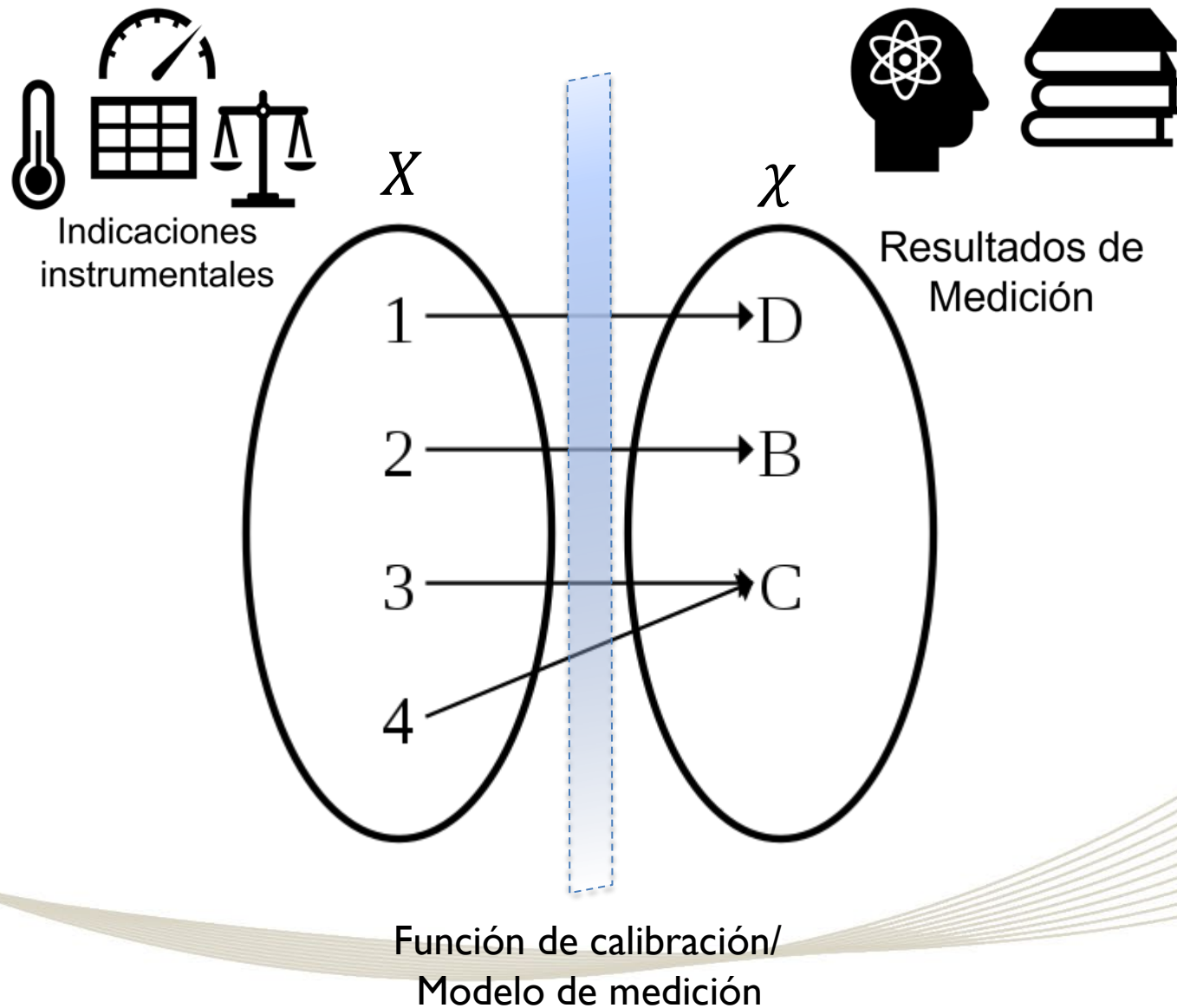
Héctor Nájera
Curtis Huffman

05/09/2024



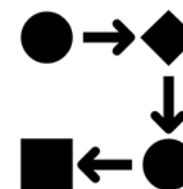
La sesión anterior

- Queremos medir la cantidad χ (concepto abstracto y universal)
- La cantidad χ no es directamente observable, así que inferimos de otra cantidad X , que es “directamente observable” (indicaciones instrumentales).
- Para llevar a cabo esta inferencia necesitamos una regla de correspondencia que exprese χ como función de X como sigue $\chi = f(X)$
- La forma de esta función f no puede ser descubierta ni estimada, debido a que involucra saber los valores tanto de X como de χ , y χ es la variable desconocida que estamos tratando de medir.



La sesión anterior

- Medir es hacer inferencias (causales)
 - Generalizar de las indicaciones instrumentales (observables/conocidos) a los resultados de medición (inobservables/objetos científicos abstractos)
 - Puede ser enmarcado como un problema de predicción (oficio de la estadística)
 - [Center for Philosophy of Science]. (2022, February 25). *Measurement Outcomes as Best Predictors* by Eran Tal [Video]. Youtube.com. <https://www.youtube.com/live/WcTvcMvKIkM?si=nEwuSIHqZU3cXMjr>
 - Los resultados de medición sólo pueden ser inferidos una vez que el instrumento ha sido subsumido bajo un modelo (causal) idealizado que les relaciona (son modelo-dependientes: mind over data).
 - Medir es representar conocimiento causal plausible (en un lenguaje matemático) y combinarlo con datos empíricos para dar respuesta a preguntas causales de utilidad práctica.



Modelo de
medición



Esta sesión

Calidad de nuestras estimaciones: estrategias de identificación

CONFIABILIDAD

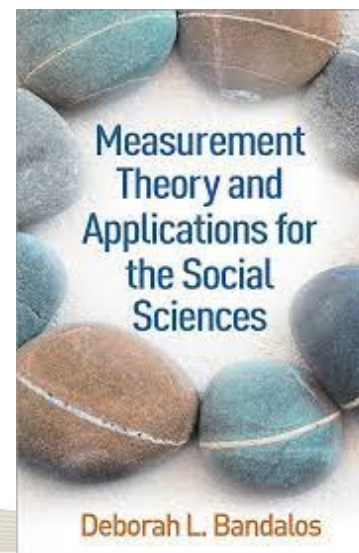
- “An essential condition that all measuring apparatuses must satisfy.” Regnault (1847, 164)
 - Si un instrumento ha de darnos la lectura correcta, debe al menos darnos siempre la misma lectura bajo las mismas circunstancias; de la misma manera, si un cierto tipo de instrumento ha de ser tenido por bueno, todos los instrumentos del tipo deberían concordar entre ellos en sus lecturas.
- “The fundamental question in reliability is to what extent do scores measured at one time and one place with one instrument predict scores at another time and/or another place and perhaps measured with a different instrument?” Ravelle y Condon (2018b)



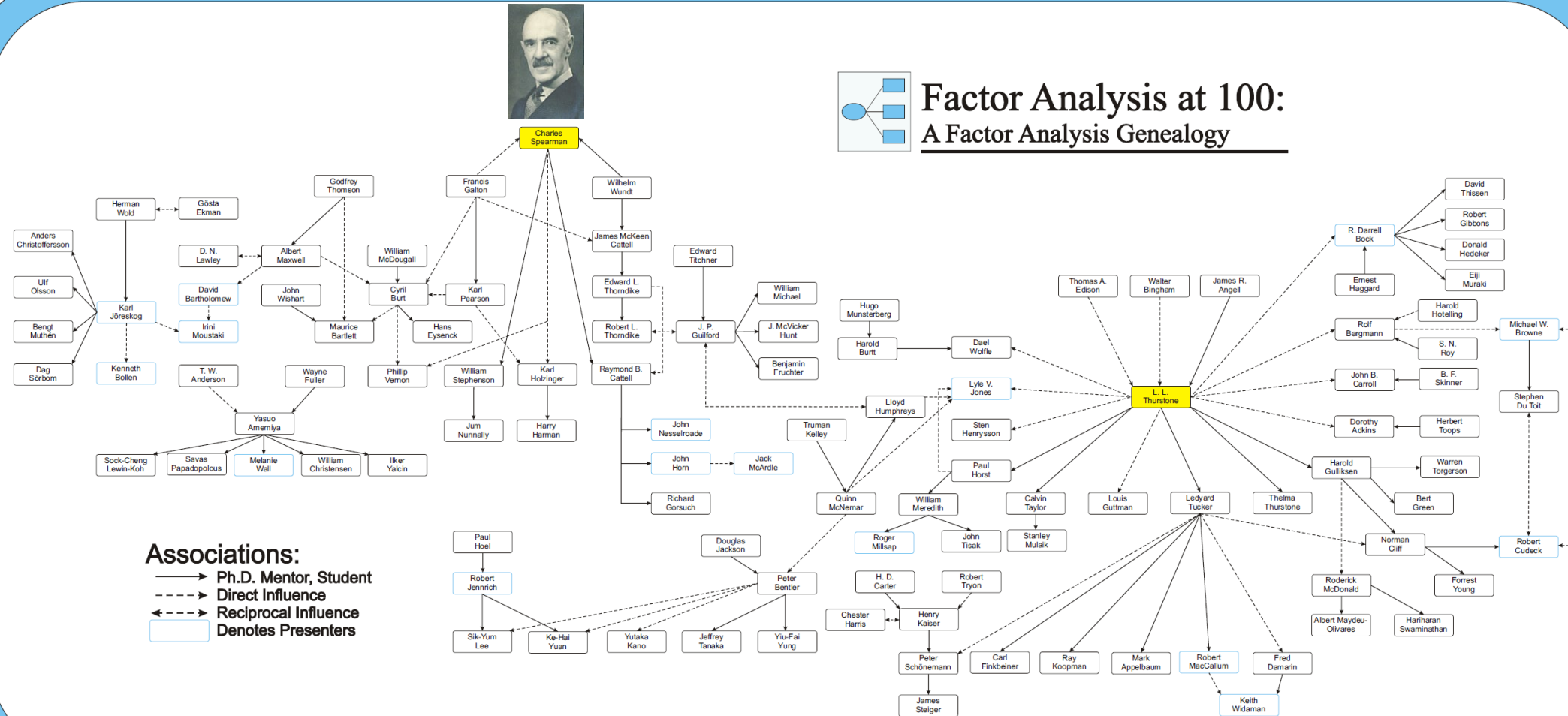
¿Estrategia de identificación?

- ¿Cómo evaluar nuestras presunciones de conocimiento (la calidad de nuestras inferencias)?
- ¿Qué confianza dar a nuestros resultados?
- ¿Cómo cuantificar el error involucrado en nuestras medidas?
- ¿Cómo combinar datos con conocimiento causal (modelo de medición) para producir respuestas a las preguntas anteriores?

Capítulo 7, Introduction to Reliability and the Classical Test Theory Model.



SEM (econometría de variables latentes)



Antecedentes del análisis de datos

- Teoría de los errores para extraer señal de ruido
 - Jacob Bernoulli (1654-1705)
 - Pierre-Simon Laplace (1749-1827)
 - Carl Friedrich Gauss (1777-1855)

Reliability and the Classical Test Theory Model 159

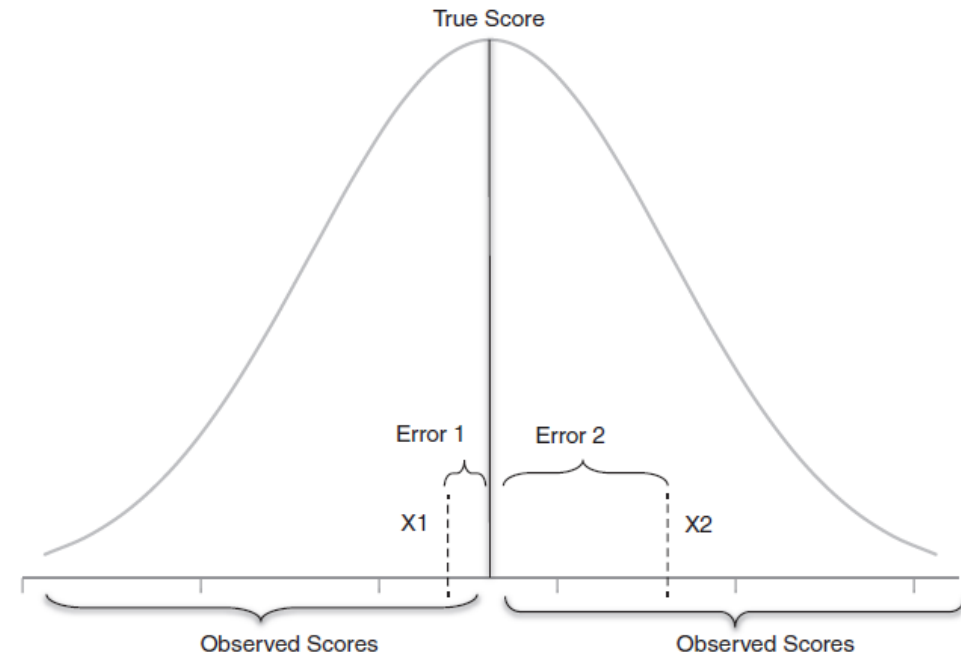
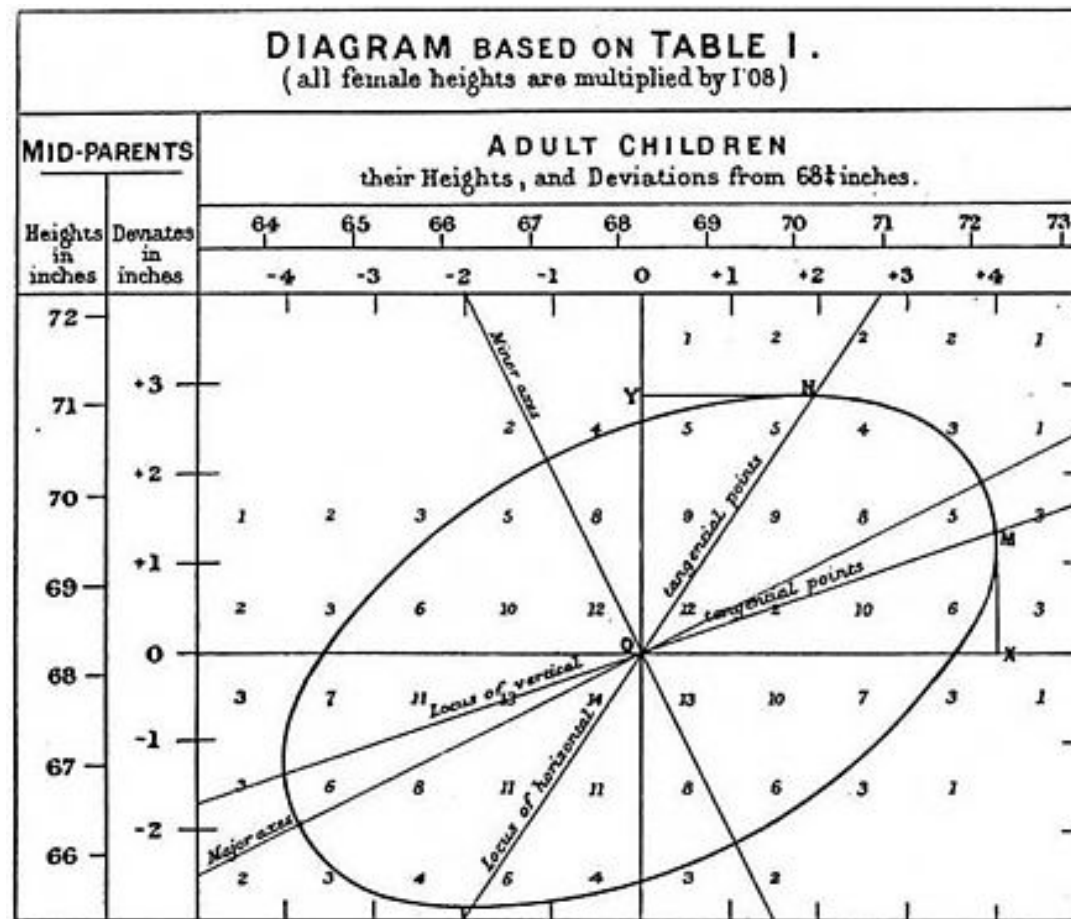


FIGURE 7.1. The distribution of observed scores around the true score.

Antecedentes del análisis de datos

- Francis Galton (1822-1911; regresión a la media)
- Carl Pearson (1857-1936; correlación no es causalidad, pero ¿qué es la causalidad?)

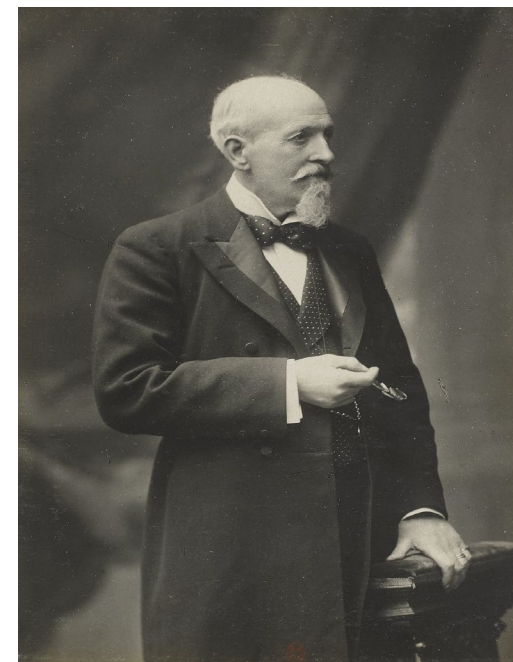


Antecedentes del análisis de datos



Teoría Clásica de los Tests (un poco de historia)

La historia de CTT comienza a la vuelta del siglo XX con la propuesta de Spearman para corregir el coeficiente de correlación por atenuación debida a errores de medición, y cómo obtener el índice de confiabilidad (calidad de las medidas) necesario para llevar a cabo esa corrección



Charles Edward Spearman
1863-1945

La intuición de Spearman (1904)

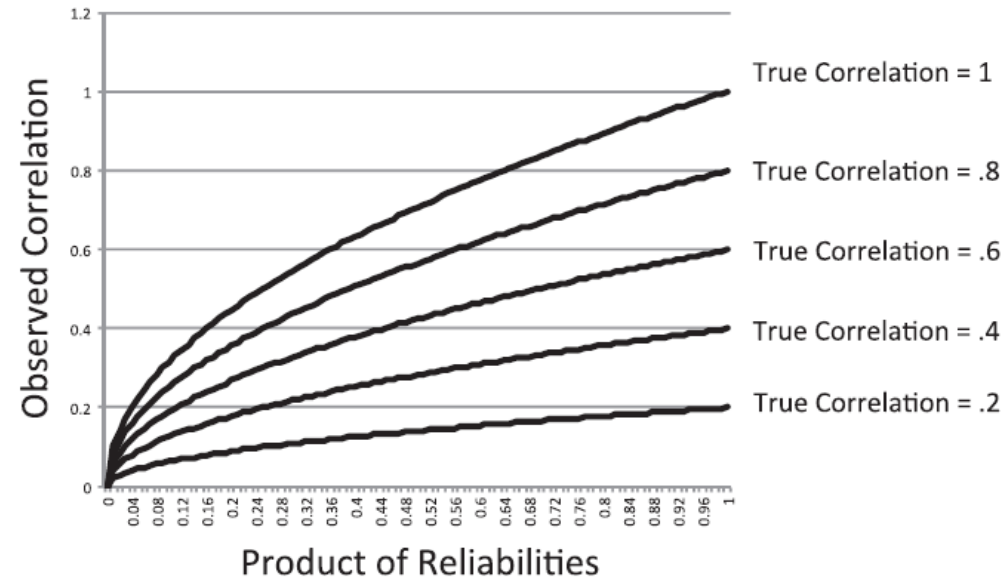


Fig. 1. The observed correlation coefficient is expressed as a function of the product of the reliabilities of the two measures and the true correlation coefficient

© 2015 The Authors
Teaching Statistics © 2015 Teaching Statistics Trust, **38**, 1, pp 25–28

Trafimow, D. (2016). The attenuation of correlation coefficients: A statistical literacy issue. *Teaching Statistics*, 38(1), 25-28.



Teoría Clásica de los Tests (un poco de historia)

"GENERAL INTELLIGENCE," OBJECTIVELY DETERMINED AND MEASURED.

By C. SPEARMAN.

TABLE OF CONTENTS.

Chap. I.	Introductory.	PAGE
	1. Signs of Weakness in Experimental Psychology	202
	2. The Cause of this Weakness	203
	3. The Identities of Science	204
	4. Scope of the Present Experiments	205
Chap. II.	Historical and Critical.	

Spearman, C. (1904). The proof and measurement of association between two things. *American journal of Psychology*, 15(1), 72-101.

Spearman, C. (1904). "General Intelligence," objectively determined and measured. *The American Journal of Psychology*, 15(2), 201-292.

¡15,000 citas en GS!

THE PROOF AND MEASUREMENT OF ASSOCIATION BETWEEN TWO THINGS.

By C. SPEARMAN.

CONTENTS.

	PAGE
Introductory.	72
Part I. Elementary correlation, and "accidental deviations."	
1. Requirements of a good method.	73
2. Standard methods explained.	77
3. Comparison by rank.	79
4. Auxiliary methods.	82
Part II. Correlation of "generalized deviations."	88

Teoría Clásica de los Tests (un poco de historia)

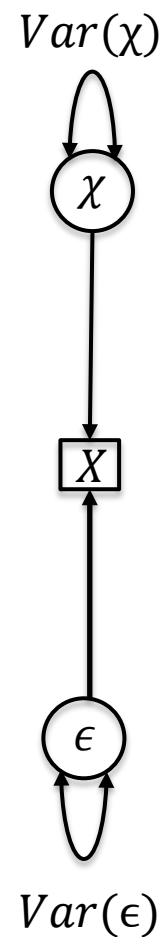
- La intuición fundamental de Spearman fue que es posible descomponer el puntaje observado en dos constructos no-observables: un puntaje latente de interés y otro de error residual, también latente.

$$X = \chi + \epsilon$$

$$Var(X) = Var(\chi) + Var(\epsilon) + 2Cov(\chi, \epsilon)$$

- ¡Aun si la $Cov(\chi, \epsilon) = 0$, el modelo no está identificado! (más incógnitas que ecuaciones).

But another—theoretically far more valuable—property may conceivably attach to one among the possible systems of values expressing the correlation; this is, that a measure might be afforded of the *hidden underlying cause of the variations*. Sup-



La gran contribución de Spearman consiste en haber reconocido que es posible introducir **variables latentes** en el discurso científico a través de sus relaciones causales con variables manifiestas.





¿La gran contribución?

- Lenguaje Causal (Path Analysis)
 - The limits of my language are the limits of mi mind. All I know is what I have words for.” –Ludwig Wittgenstein
 - Sewall Wright (1889-1988) (no se pueden sacar conclusiones causales sin hipótesis causales)
 - Frisch (1895-1973; “Causality is in the mind”)

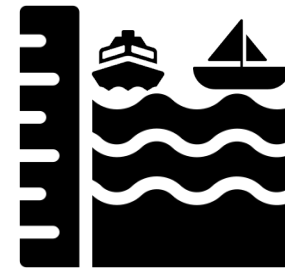
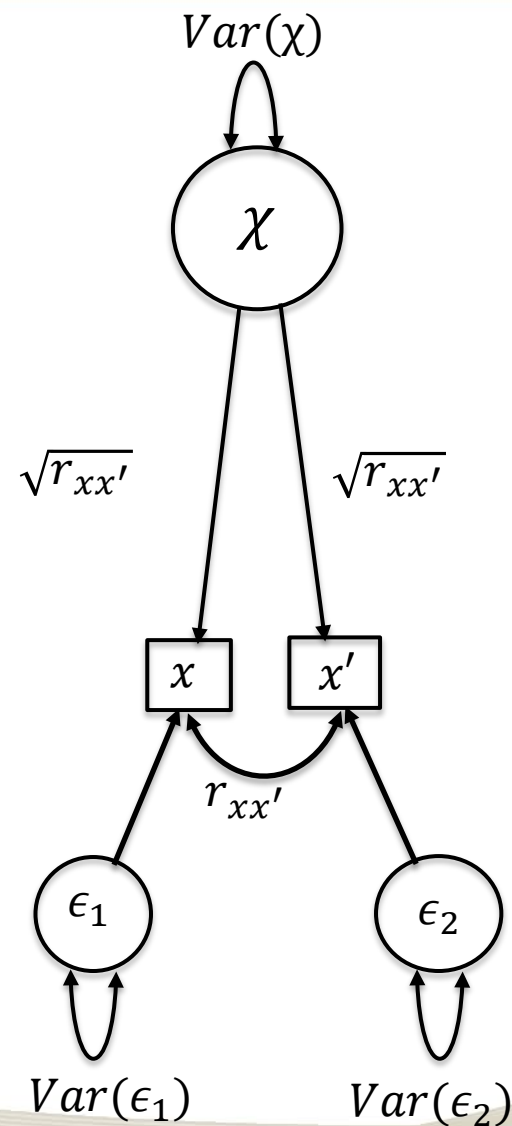
“... we think of a cause as something imperative which exists in the exterior world. In my opinion this is fundamentally wrong. If we strip the word cause of its animistic mystery, and leave only the part that science can accept, nothing is left except a certain way of thinking. [T]he scientific ... problem of causality is essentially a problem regarding our way of thinking, not a problem regarding the nature of the exterior world.” — Frisch (1930), p. 36

Teoría Clásica de los Tests (un poco de historia)

- Si hay dos (o más) indicaciones instrumentales (x y x' , *paralelas*) que comparten la misma fuente de variación (χ), de igual varianza ($Var(X_1) = Var(X_2)$), y errores independientes ($\epsilon_1 \perp \epsilon_2$),

$$\begin{aligned}X_1 &= \chi + \epsilon_1 \\X_2 &= \chi + \epsilon_2\end{aligned}$$

- Entonces la correlación de estas dos medidas, $r_{xx'}$, es un estimado directo de la varianza de χ .



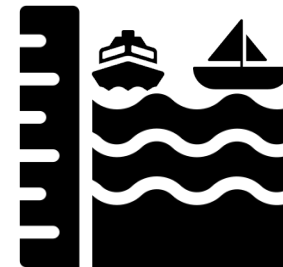
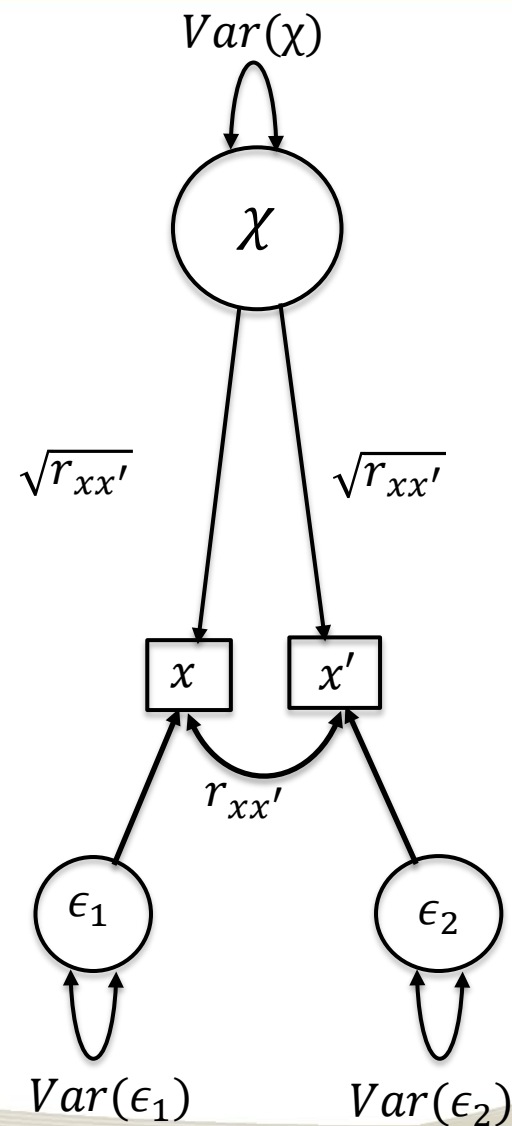
Teoría Clásica de los Tests (un poco de historia)

$$X_1 = \chi + \epsilon_1$$

$$X_2 = \chi + \epsilon_2$$

$$\begin{aligned} \text{Corr}(X_1, X_2) &= \frac{\text{Cov}(X_1, X_2)}{\sqrt{\text{Var}(X_1)\text{Var}(X_2)}} \\ &= \frac{\text{Cov}(\chi + \epsilon_1, \chi + \epsilon_2)}{\sqrt{\text{Var}(X_1)\text{Var}(X_2)}} \\ &= \frac{\text{Cov}(\chi, \chi) + \text{Cov}(\chi, \epsilon_2) + \text{Cov}(\chi, \epsilon_1) + \text{Cov}(\epsilon_1, \epsilon_2)}{\sqrt{\text{Var}(X_1)\text{Var}(X_2)}} \\ &= \frac{\text{Cov}(\chi, \chi)}{\sqrt{\text{Var}(X_1)\text{Var}(X_2)}} = \frac{\text{Var}(\chi)}{\sqrt{\text{Var}(X_1)\text{Var}(X_2)}} \\ &= \frac{\text{Var}(\chi)}{\sqrt{\text{Var}(X)\text{Var}(X)}} = \frac{\text{Var}(\chi)}{\text{Var}(X)} \end{aligned}$$

Aumenta
la calidad



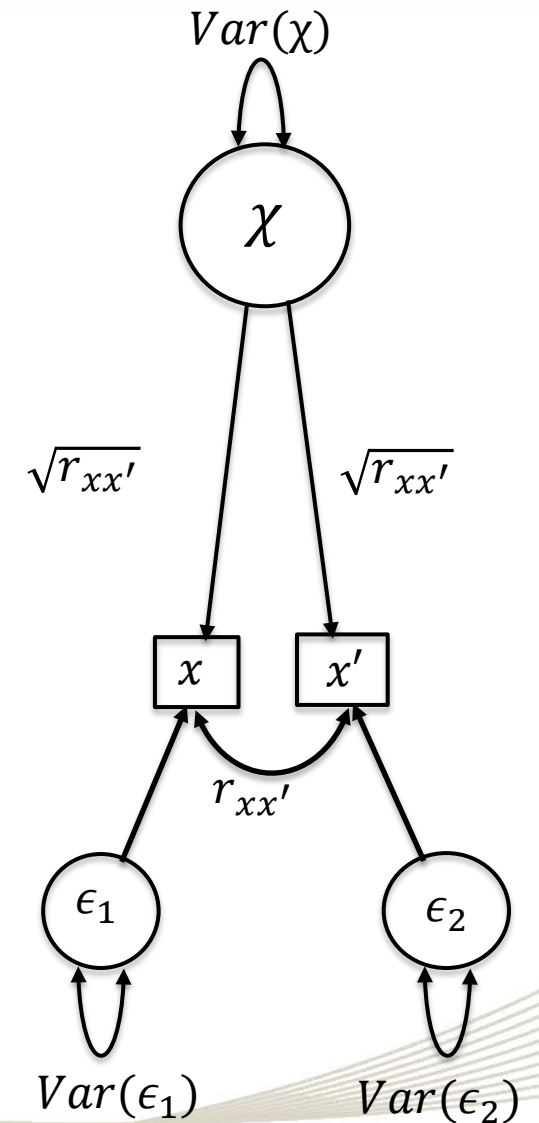
¿Dijo usted calidad?

- ¿De qué hablamos cuando hablamos de confiabilidad?
 - “ [reliability may be considered as dealing with] separating signal from noise”. Ravelle y Condon (2018a)
 - La proporción de la variabilidad en nuestras medidas (scale score) que es atribuible a la propiedad que queremos medir, en relación con la variabilidad total contenida en respuesta de los participantes (la lectura tomada de los objetos medidos). (Dunn, T. J., Baguley, T., & Brunsden, V., 2014)
 - Una función del cociente entre la señal y el ruido.
 - La proporción entre lo relevante y lo irrelevante de nuestras observaciones empíricas (puntajes observados o medidas).
 - La fracción de nuestras mediciones que **no** es irrelevante.
 - La razón entre la varianza confiable y la varianza total de nuestras mediciones.
 - El porcentaje de la variación de nuestras mediciones que no es error.
- **Adecuado para los fines perseguidos (fit-for-purpose)**

¿Es creíble la estrategia bajo TCT?



- Toda la varianza individual es error de medición (las lecturas son completamente homogéneas/paralelas)
- Noten que basta con que $\chi_2 = \chi_1 + c = \chi$ (pueden tener medias diferentes) toda vez que $Cov(\chi, \chi) = Cov(\chi, \chi + c)$, $Corr(\chi_1, \chi_2) = 1$ y $Var(X) = Var(X + c)$. Pero $Var(\epsilon_1) = Var(\epsilon_2)$ sigue siendo necesario
- Las dos medidas son exactamente igual de buenas (reactivas) y sus errores varían igual: **con la misma mezcla de señal y ruido**
- Si lo “desparalelizamos” el modelo vuelve a no estar identificado: tres ecuaciones y 5 incógnitas



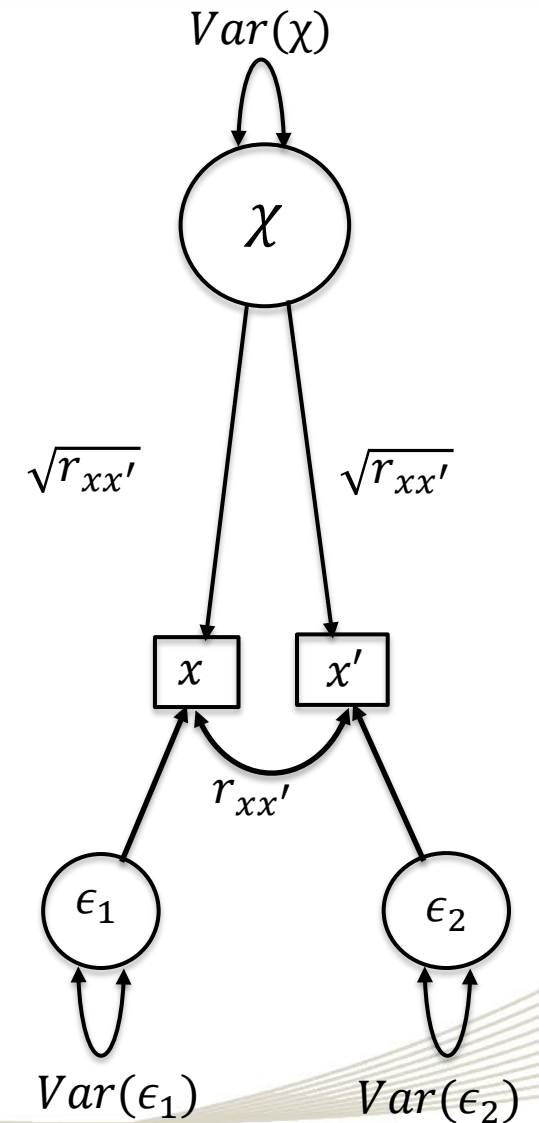
¿Es creíble la estrategia bajo TCT?



- ¿Si x y x' fuesen la misma prueba aplicada en diferentes momentos?

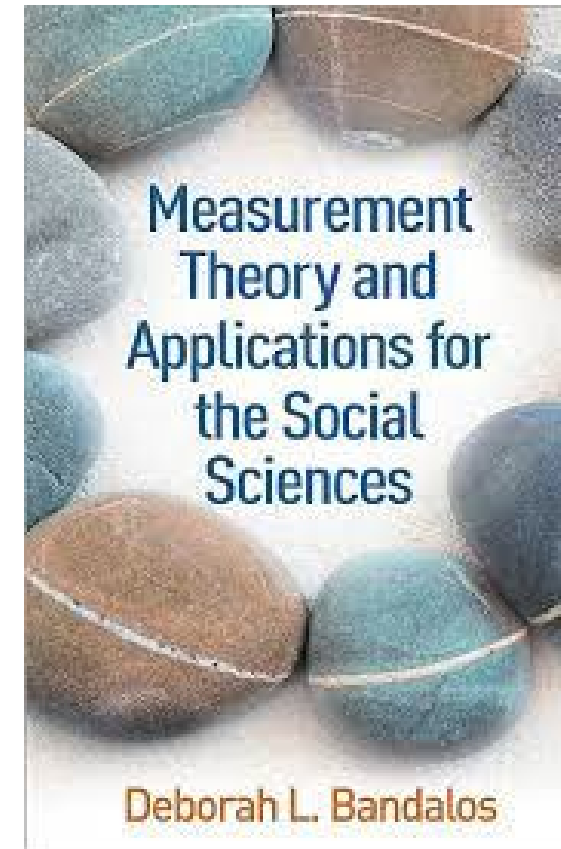
$$Cov(\epsilon_1, \epsilon_2) = 0?$$

- ¿Si x y x' formaran parte de la misma prueba? (reliability of a composite, $X = X'_1 + X'_2$) ?
 - Noten que la confiabilidad de medio test no es, en general, la confiabilidad del test completo. Un test largo suele ser más confiable que su versión corta.



Próxima clase

- Capítulo 8, Methods of Assessing Reliability





Lecturas recomendadas

- Bartholomew, D. J. (1995). Spearman and the origin and development of factor analysis. *British Journal of Mathematical and Statistical Psychology*, 48(2), 211-220.
- Cho, E., & Kim, S. (2015). Cronbach's coefficient alpha: Well known but poorly understood. *Organizational Research Methods*, 18(2), 207-230.
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of applied psychology*, 78(1), 98.
- Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British journal of psychology*, 105(3), 399-412.
- Green, S. B., & Yang, Y. (2009). Commentary on coefficient alpha: A cautionary tale. *Psychometrika*, 74(1), 121-135.
- Levy, P. (1995). Charles Spearman's contributions to test theory. *British Journal of Mathematical and Statistical Psychology*, 48(2), 221-235.
- Miller, M. B. (1995). Coefficient alpha: A basic introduction from the perspectives of classical test theory and structural equation modeling.
- Rigdon, E. E. (1994). Demonstrating the effects of unmodeled random measurement error. *Structural Equation Modeling: A Multidisciplinary Journal*, 1(4), 375-380.
- Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological assessment*, 8(4), 350.
- Streiner, D. L. (2003). Starting at the beginning: an introduction to coefficient alpha and internal consistency. *Journal of personality assessment*, 80(1), 99-103.
- Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International journal of medical education*, 2, 53.
- Trafimow, D. (2016). The attenuation of correlation coefficients: A statistical literacy issue. *Teaching Statistics*, 38(1), 25-28.
- Winne, P. H., & Belfry, M. J. (1982). Interpretive problems when correcting for attenuation. *Journal of Educational Measurement*, 125-134.
- Zimmerman, D. W., & Williams, R. H. (1997). Properties of the Spearman correction for attenuation for normal and realistic non-normal distributions. *Applied Psychological Measurement*, 21(3), 253-270.



Referencias

- Bandalos, D. L. (2018). *Measurement theory and applications for the social sciences*. Guilford Publications.
- Frisch, R. (1930). A dynamic approach to economic theory: Lectures by Ragnar Frisch at Yale University. Lectures at Yale University beginning September, 1930. Mimeographed, 246 pp. Frisch Archives, Department of Economics, University of Oslo.
- Regnault, V. (1847). *Relation des expériences pour déterminer les principales lois et les données numériques qui entrent dans le calcul des machines à vapeur*. Didot.
- Revelle, W., & Condon, D. M. (2018a). Reliability. *The Wiley Handbook of Psychometric Testing: A Multidisciplinary Reference on Survey, Scale and Test Development*, First Edition. Edited by Paul Irwing, Tom Booth, and David J. Hughes., 709-749. <https://personality-project.org/revelle/publications/reliability-final.pdf>
- Revelle, W., & Condon, D. M. (2018, June 10). *Reliability from alpha to omega: a tutorial*. <https://doi.org/10.31234/osf.io/2y3w9>
- Spearman, C. (1904). "General Intelligence," objectively determined and measured. *The American Journal of Psychology*, 15(2), 201-292.
- Spearman, C. (1904). The proof and measurement of association between two things. *American journal of Psychology*, 15(1), 72-101



CONTACTO

Dres. Héctor Nájera y Curtis Huffman
Investigadores (SIN-II)

Programa Universitario de Estudios del Desarrollo (PUED)
Universidad Nacional Autónoma de México (UNAM)

Antigua Unidad de Posgrado (costado sur de la Torre II de Humanidades), planta baja.
Campus Central, Ciudad Universitaria, Ciudad de México, México.

Tel. (+52) 55 5623 0222, Ext. 82613 y 82616

Tel. (+52) 55 5622 0889

Email: hecatalan@hotmail.com

chuffman@unam.mx



*¡Bienvenidos
estudiantes!*

