¿Qué es un problema de identificación?

.....

Ernesto San Martín

Millennium Nucleus on Intergenerational Mobility: From Modelling to Policy MOVI
Laboratorio Interdisciplinario de Estadística Social, LIES UC
Facultad de Matemáticas, Pontificia Universidad Católica de Chile
The Economics School of Louvain, Uniersité catholique de Louvain, Belgium

December, 2021

Laboratorio
Interdisciplinario de
Estadística Social

Perspective

- **¿Cuál es la tarea de un metodólogo?**

Perspective

- **¿Cuál es la tarea de un metodólogo?**

  The sociologist studies man in society: the methodologist studies the sociologist at work (Lazarsfeld, 1959).

Perspective

- **¿Cuál es la tarea de un metodólogo?**

  The sociologist studies man in society: the methodologist studies the sociologist at work (Lazarsfeld, 1959).

- **¿Hay alguna herramienta para ello?**

Perspective

- **¿Cuál es la tarea de un metodólogo?**

  The sociologist studies man in society: the methodologist studies the sociologist at work (Lazarsfeld, 1959).

- **¿Hay alguna herramienta para ello?**

  Conditional independence is the key concept of structural modelling (Novick, 1979).

Perspective

- **¿Cuál es la tarea de un metodólogo?**

  The sociologist studies man in society: the methodologist studies the sociologist at work (Lazarsfeld, 1959).

- **¿Hay alguna herramienta para ello?**

  Conditional independence is the key concept of structural modelling (Novick, 1979).

- **¿Cuál es la finalidad epistemológica?**

Perspective

- **¿Cuál es la tarea de un metodólogo?**

  The sociologist studies man in society: the methodologist studies the sociologist at work (Lazarsfeld, 1959).

- **¿Hay alguna herramienta para ello?**

  Conditional independence is the key concept of structural modelling (Novick, 1979).

- **¿Cuál es la finalidad epistemológica?**

  In many fields the objective of the investigator's inquisitiveness is not just a "population" in the sense of a distribution of observable variables, but a physical structure projected behind this distribution, by which the latter is thought to be generated. The word "physical" is used merely to convey that the structure concept is based on the investigator's ideas as to the "explanation" or "formation" of the phenomena studied, briefly, on his theory of these phenomena, whether they are classified as physical in the literal sense, biological, psychological, sociological, economic or otherwise (Koopmans and Reiersol, 1950).

Desafío epistemológico

- **Lógica de la investigación empírica:**

$$\text{Datos/Evidencia} \quad + \quad \text{Supuestos/Teoría} \Longrightarrow \text{Conclusiones/Recomendaciones}$$

Desafío epistemológico

- **Lógica de la investigación empírica:**

$$\text{Datos/Evidencia} \quad + \quad \text{Supuestos/Teoría} \Longrightarrow \text{Conclusiones/Recomendaciones}$$

- Esta lógica no es coherente con afirmaciones del estilo

  Hay evidencia que muestra ...
  La evidencia sugiere que ...

Desafío epistemológico

- Lo que determina los cursos de acción son las conclusiones o recomendaciones.
- Pero estas dependen de los supuestos:

$$\text{Datos/Evidencia} \quad + \quad \text{Supuestos/Teoría}_1 \quad \Longrightarrow \quad \text{Conclusiones/Recomendaciones}_1$$

$$\text{Datos/Evidencia} \quad + \quad \text{Supuestos/Teoría}_2 \quad \Longrightarrow \quad \text{Conclusiones/Recomendaciones}_2$$

- Si mantenemos los datos/evidencia fijos y si cambiamos los supuestos, puede que las conclusiones cambien, y que el cambio sea dramático.

Desafío epistemológico

- **¿A qué nos referimos cuando usamos la expresión supuestos?**

Desafío epistemológico

- **¿A qué nos referimos cuando usamos la expresión supuestos?**
- No nos referimos a las hipótesis estadísticas que pueden ser falseadas usando la teoría de tests de hipótesis. Estas dependen de los observables.

Desafío epistemológico

- **¿A qué nos referimos cuando usamos la expresión supuestos?**
- No nos referimos a las hipótesis estadísticas que pueden ser falseadas usando la teoría de tests de hipótesis. Estas dependen de los observables.
- Nos referimos a supuestos que no pueden falsearse empíricamente: son las ideas que los investigadores tienen acerca de la formación o explicación del fenómeno. Estas dependen de no-observables.

Desafío epistemológico

- Existe un problema de identificación cuando las respuestas a la siguientes dos preguntas son diferentes:

  ¿Qué se puede aprender de un conjunto de datos? y ¿Qué se quiere aprender de esos datos?

- Aprender de un conjunto de datos significa especificar la distribución de probabilidades que los genera.

Ejemplo: modelo Rasch

- Sea $\{Y_{ij} : i = 1, \ldots, I; j = 1, \ldots, J\}$ una secuencia de variables aleatorias utuamente independientes tales que

$$Y_{ij} = \begin{cases} 1, & \text{si la persona } i \text{ responde correctamente el item } j; \\ 0, & \text{en cualquier otro caso.} \end{cases},$$

donde $Y_{ij} \sim \text{Bern}\,(p_{ij})$.

- El modelo Rasch se focaliza sobre *comparaciones relativas*. Por lo tanto, introducimos la siguiente reparametrización:

$$p_{ij} = \frac{\lambda_{ij}}{1 + \lambda_{ij}}, \quad \lambda_{ij} \geqslant 0,$$

que es equivalente a

$$\lambda_{ij} = \frac{P(Y_{ij} = 1)}{P(Y_{ij} = 0)}.$$

Ejemplo: modelo Rasch

- Por lo tanto, si $\lambda_{ij} > 1$, entonces la persona i tiene una probabilidad de responder correctamente el item j mayor que de responderlo incorrectamente.

- Una interpretación similar puede ser dada para $\lambda_{ij} < 1$ o $\lambda_{ij} = 1$.

Ejemplo: modelo Rasch

- Por lo tanto, si $\lambda_{ij} > 1$, entonces la persona i tiene una probabilidad de responder correctamente el item j mayor que de responderlo incorrectamente.
- Una interpretación similar puede ser dada para $\lambda_{ij} < 1$ o $\lambda_{ij} = 1$.
- ¿Por qué este modelo se usa en medición educacional? Ver, por ejemplo, Claro et al. (2009).
- Puede ser moda, conveniencia teórica, conveniencia práctica ...

Ejemplo: modelo Rasch

- Por lo tanto, si $\lambda_{ij} > 1$, entonces la persona i tiene una probabilidad de responder correctamente el item j mayor que de responderlo incorrectamente.
- Una interpretación similar puede ser dada para $\lambda_{ij} < 1$ o $\lambda_{ij} = 1$.
- ¿Por qué este modelo se usa en medición educacional? Ver, por ejemplo, Claro et al. (2009).
- Puede ser moda, conveniencia teórica, conveniencia práctica ...
- La pregunta debería ser: ¿Cómo debe ser usado este modelo?

Ejemplo: modelo Rasch

- Rasch quiere comparar individuos, para lo cual busca *una regla, un estándar*.
- Pero nada de esto está presente en los parámetros p_{ij}'s o λ_{ij}'s. De los datos solo se puede aprender la tasa o proproción de 1's, que puede ser vista como un estimador de p_{ij}.

Ejemplo: modelo Rasch

- Rasch quiere comparar individuos, para lo cual busca *una regla, un estándar*.

- Pero nada de esto está presente en los parámetros p_{ij}'s o λ_{ij}'s. De los datos solo se puede aprender la tasa o proproción de 1's, que puede ser vista como un estimador de p_{ij}.

- El interés teorico conlleva a querer aprender algo de los datos que no está en los datos.

- Es por ello que se introducen dos parámetros, uno que se relaciona con cada individuo (ϵ_i), el otro con cada item (η_j):

$$\lambda_{ij} = \frac{\epsilon_i}{\eta_j} \quad i = 1, \ldots, I, \ j = 1, \ldots, J.$$

Ejemplo: modelo Rasch

- ¿Pueden escribirse los parámetros $\{(\epsilon_i, \eta_j)\}$ como una función de $\{\lambda_{ij}\}$?
- Respuesta: no.
- Es decir, los parámetros $\{(\epsilon_i, \eta_j)\}$ no son propiedades de los datos bajo estudio.
- Tenemos entonces un problema de identificación.

Ejemplo: modelo Rasch

- De hecho, para $\{\epsilon_i, \eta_j\}$ y $\lambda_{ij} = \{\epsilon_i', \eta_j'\}$ se puede tener que

$$\frac{\epsilon_i}{\eta_j} = \frac{\epsilon_i'}{\eta_j'} \qquad \forall\,(i,j) \in \{1,\ldots,I\} \times \{1,\ldots,J\},$$

  lo que es equivalente a

$$\frac{\epsilon_i}{\epsilon_i'} = \frac{\eta_j}{\eta_j'} \qquad \forall\,(i,j) \in \{1,\ldots,I\} \times \{1,\ldots,J\},$$

  lo que necesariamente implica que

$$\frac{\epsilon_i}{\epsilon_i'} = \frac{\eta_j}{\eta_j'} = \alpha \qquad \forall\,(i,j) \in \{1,\ldots,I\} \times \{1,\ldots,J\},$$

  donde α es una constant.

- Por lo tanto, $\{\epsilon_i, \eta_j\}$ no son identificados y en consecuencia no pueden ser estadísticamente interpretados.

Ejemplo: modelo Rasch

- Para remover la indeterminación inherente, escogemos un item, por ejemplo $j = 1$, como un *item estándar* en el sentido que fijamos una "unidad". Así, $\eta_1 = 1$ y por lo tanto

$$\epsilon_i = \lambda_{i1} = \frac{P(Y_{i1} = 1}{P(Y_{i1} = 0)}.$$

- El parámetro que caracteriza a la persona i corresponde al *betting odd* de responder correctamente el item estándar 1. En particular,

$$\epsilon_i > \epsilon_{i'} \iff P(Y_{i1} = 1) > P(Y_{i'1} = 1).$$

Ejemplo: modelo Rasch

- Para remover la indeterminación inherente, escogemos un item, por ejemplo $j = 1$, como un *item estándar* en el sentido que fijamos una "unidad". Así, $\eta_1 = 1$ y por lo tanto
$$\epsilon_i = \lambda_{i1} = \frac{P(Y_{i1} = 1}{P(Y_{i1} = 0)}.$$

- El parámetro que caracteriza a la persona i corresponde al *betting odd* de responder correctamente el item estándar 1. En particular,
$$\epsilon_i > \epsilon_{i'} \Longleftrightarrow P(Y_{i1} = 1) > P(Y_{i'1} = 1).$$

- Pero aquí no hay nada psicológico!

Ejemplo: modelo Rasch

- Similarmente,

$$\eta_j = \frac{P(Y_{i1} = 1)}{P(Y_{i1} = 0)} \frac{P(Y_{ij} = 0)}{P(Y_{ij} = 1)}.$$

- Entonces
  - $\eta_j > 1 = \eta_1 \iff P(Y_{i1} = 1) > P(Y_{ij} = 1)$ para toda persona i.
  - $\eta_j > \eta_k \iff P(Y_{ik} = 1) > P(Y_{ij} = 1)$ para toda persona i.

Ejemplo: modelo Rasch

- Similarmente,

$$\eta_j = \frac{P(Y_{i1}=1)}{P(Y_{i1}=0)}\,\frac{P(Y_{ij}=0)}{P(Y_{ij}=1)}.$$

- Entonces
  - $\eta_j > 1 = \eta_1 \Longleftrightarrow P(Y_{i1}=1) > P(Y_{ij}=1)$ para toda persona i.
  - $\eta_j > \eta_k \Longleftrightarrow P(Y_{ik}=1) > P(Y_{ij}=1)$ para toda persona i.

- De nuevo, nada psicológico o educacional hay aquí!

Ejemplo: modelo Rasch

- El modelo Rasch permite comparar características individuales ϵ_i y características de los items η_j:

$$\epsilon_i > \eta_j \iff P(Y_{ij} = 1) > P(Y_{ij} = 0) \iff \lambda_{ij} > 1.$$

Ejemplo: modelo Rasch

- El modelo Rasch permite comparar características individuales ϵ_i y características de los items η_j:

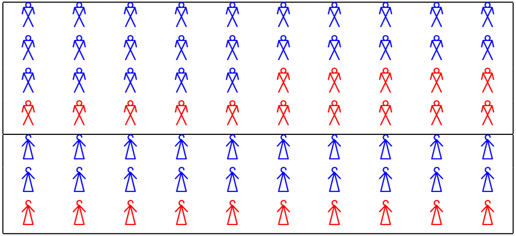
$$\epsilon_i > \eta_j \iff P(Y_{ij} = 1) > P(Y_{ij} = 0) \iff \lambda_{ij} > 1.$$

- λ_{ij} es un parámetro que explica una determinada característica de los datos: ahora dicho parámetro tiene una interpretación teórica.

San Martín, González & Tuerlinck (2009, 2015)

Ejemplo de una encuesta

- Vamos a suponer que de 70 personas encuestadas, 45 declaran que votarán **apruebo** en el plebiscito de salida, mientras que 25 declaran que votarán **rechazo**.
- El **espacio muestral** corresponde al conjunto de etiquetas o identificadores de cada uno de los encuestados.
- Los 70 encuestados son clasificados según su sexo biológico: hombre y mujer .
- Esta información puede representarse de la siguiente manera:



Probabilidades marginales

- La proproción (o probabilidad) de **apruebo** es igual a

$$P(\textbf{apruebo}) = \frac{45}{70}$$

–esto es, 64.29 %.

- La proporción (o probabilidad) de **rechazo** es igual a

$$P(\textbf{rechazo}) = \frac{25}{70}$$

–esto es, 35.71 %.

- Las probabilidades anteriores se calculan teniendo en cuenta **toda la población de encuestados**: por eso se llaman **probabilidades marginales**.

Probabilidades condicionales

- Las proporciones de **apruebo** y **rechazo** pueden reportarse también bajo una determinada **condición**, por ejemplo el sexo biológico.

- Así, la proporción de **apruebo** **entre los hombres** o **bajo la condición que el encuestado es hombre**, lo que escribimos como $P(\textbf{apruebo} \mid 🚹)$, está dada por

$$P(\textbf{apruebo} \mid 🚹) = \frac{25}{40}.$$

- La proporción de **apruebo** **entre las mujeres** o **bajo la condición que la encuestada es mujer**, lo que escribimos como $P(\textbf{apruebo} \mid 🚺)$, está dada por

$$P(\textbf{apruebo} \mid 🚺) = \frac{20}{30}.$$

- Cada una de estas probabilidades se calcula con respecto a la población definida por la **condición**, por lo que se llaman **probabilidades condicionales**.

- Es posible relacionar la probabilidad marginal $P(\textbf{apruebo})$ con las probabilidades condicionales $P(\textbf{apruebo} \mid \Gentsroom)$ y $P(\textbf{apruebo} \mid \Ladiesroom)$.

- Para ello solo es necesario considerar la proporción de hombres y de mujeres entre los encuestados:

$$P(\Gentsroom) = \frac{40}{70}, \qquad P(\Ladiesroom) = \frac{30}{70}.$$

Ley de Probabilidades Totales

- Teniendo en cuenta entonces que de los 45 apruebo, 25 son hombres y 20 son mujeres, podemos hacer la siguiente descomposición:

$$
\begin{aligned}
P(\textbf{apruebo}) &= \frac{45}{70} \\
&= \frac{25+20}{70} \\
&= \frac{25}{40} \times \frac{40}{70} + \frac{20}{30} \times \frac{30}{70} \\
&= P(\textbf{apruebo} \mid \male)\, P(\male) \; + \; P(\textbf{apruebo} \mid \female)\, P(\female)
\end{aligned}
$$

- La descomposición anterior siempre se puede hacer: basta descomponer el espacio muestral en subpoblaciones o grupos, y luego caracterizar las probabilidades condicionales de una variable de interés para cada grupo. Esto se llama **Ley de Probabilidades Totales**.