

# Precisar qué va a observarse

King, Keohane y Verba

Supongamos que...

- En ciencias sociales se busca, a partir de la investigación, extraer inferencias causales válidas

# Sin MIEDO!

- Los autores están preocupados por los sesgos que pueden presentarse en la selección de las observaciones tanto en estudios cuantitativos y CUALITATIVOS

¿Saben qué?

- Se centran en los cualitativos!!! Esto es por que es más difícil escapar a los sesgos por no contar con las bondades de la selección aleatoria

# Retomando capítulos anteriores

- 1) necesidad de extraer la mayor cantidad posible de consecuencias observables de nuestras teorías y en la medida de lo posible, comprobar cuantas más mejor.
- 2) “independencia condicional”
- 3) supuesto de la homogeneidad de las unidades

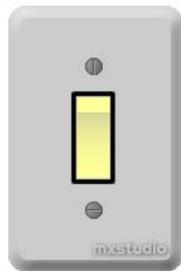
# Independencia Condicional

- Supuesto que señala que se eligen las observaciones y se asignan los valores de las variables explicativas al margen de los de las dependientes.
- Solucionar esto para  $n$  grandes = aleatoriedad
- Solucionar esto para  $n$  pequeños = *problema*

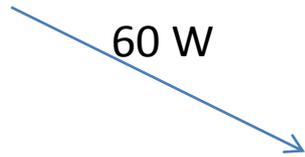
*Se recomienda apelar al supuesto de homogeneidad de las unidades como última línea defensiva*

# Supuesto de homogeneidad de las unidades

- Si dos unidades tienen el mismo valor de la variable explicativa principal, el valor esperado de la dependiente será el mismo



60 W



Mismo efecto esperado, siempre y cuando los interruptores estén en la misma **posición (o apagados o encendidos)**



60 W



**VERSIÓN ESTRICTA**



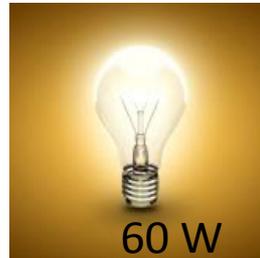
# Supuesto de homogeneidad de las unidades

- Efecto constante: variación similar de los valores de la variable explicativa en dos observaciones, conduce al mismo efecto causal en unidades diferentes, aunque puede que los niveles de dichas variables no sean los mismos.



Apagado  
Bajo

60 W →



*Efecto constante*



Bajo  
Alto

60 W →



**VERSIÓN MENOS  
ESTRICTA**

# Problema de sesgo de selección

- Deviene del no cumplimiento de los supuestos de independencia condicional y de homogeneidad de las unidades.

# ¿Cómo elegir las observaciones para el análisis?

- Caso: no es “una sola observación”, puede referirse a una única unidad en la que se midan muchas variables, o incluso a un amplio ámbito de análisis
- Observación: se define como la medida que se toma de una variable dependiente en una unidad e incluye información sobre los valores de las variables explicativas.

# Elegir qué se va a observar

- Preocupan las observaciones utilizadas para extraer inferencias en cualquiera sea el nivel de análisis en que estemos interesados.
- Selección de observaciones (cuanti) Selección de casos (cuali)

# Diseños de Investigación Imprecisos

- “un diseño de investigación es un plan que muestra mediante el análisis de un modelo y de unos datos, de qué manera esperamos utilizar nuestro material para extraer inferencias”
- Imprecisión: no podemos saber nada sobre sus hipótesis causales.
- Cualitativistas: con frecuencia disponen de suficiente información para hacer que sus diseños sean precisos.

Validez, requiere control, la posibilidad de control está sujeta al nivel de precisión del diseño

- DOS PROBLEMAS GRAVES A TENER EN CUENTA EN LA ELECCIÓN DE LO QUE OBSERVAREMOS
- A) más inferencias que consecuencias observadas
- B) multicolinealidad de las variables

# Más inferencias que observaciones

- Inferencia: proceso mediante el cual utilizamos HECHOS CONOCIDOS para saber algo sobre HECHOS DESCONOCIDOS
- 7 variables causales, 1 dependiente y 3 observaciones: no permitirán conocer cuál de las hipótesis es plausible.
- La solución no se encuentra en profundizar en el caso, incluir más puede ser una buena opción, pero si es inviable, se puede apelar a aumentar el número de observaciones (al hacerlo podemos aportar más pruebas de las consecuencias, en diferentes niveles de análisis)

# Multicolinealidad

- Refiere a cualquier situación en la que podemos predecir perfectamente una variable explicativa en función de una o más de las restantes.
- Ejemplos: tamaño y proximidad geográfica/género y embarazo
- Se recomienda que se maximice el control limitando el número de variables explicativas con las que se quiere hacer inferencias causales

# Límites de la selección aleatoria

- La selección aleatoria es fantástica para los estudios en los que el  $n$  es grande= evita el sesgo de selección mediante una selección aleatoria de las observaciones, por que un criterio aleatorio no se correlaciona con todas las variables explicativas o dependientes posibles. Las posibilidades de que un criterio de selección se correlacione con cualquier variable observada son extremadamente pequeñas.

# Experimentos

- En ciencias sociales los experimentos son escasos pero hay dos puntos de los que aprender en cuanto a su diseño: la selección de las observaciones y la adjudicación de valores de las variables explicativas a las unidades.

# Selección aleatoria: no siempre es útil

- Cuando el  $n$  es muy pequeño (como en los estudios cualitativos) pretender usar la selección aleatoria como criterio de selección de “casos” (u observaciones) puede incurrir en un sesgo de selección. Si es equivalente ola selección de aleatoria de observaciones y la elección aleatoria de uno de los tres criterios de selección posibles, !Existirá una probabilidad de dos contra tres de que la selección se observaciones produzca un sesgo de selección!

# Conocer el sesgo nos permite intentar controlarlo.

- Cuando el criterio de selección correlaciona con la variable dependiente, habrá un sesgo de selección.
- Ejemplo: elegir “casos” por que la investigadora conoce el idioma de cada uno de los países. ( se puede evitar obviando que sabe checo y poniéndose a estudiar húngaro)

# Sesgo de selección: ¿qué casos tenemos que incluir en nuestro estudio?

- Investigación cualitativa: elección de las observaciones resulta crucial para el resultado del estudio y condiciona en qué medida éste puede generar resultados precisos y fiables.

# Seleccionar en función de la variable dependiente (?)

- La selección tiene que permitir que se produzca , al menos, alguna variación en la variable dependiente.
- Cuando las observaciones surgen a partir de un determinado valor de la variable dependiente, no se pueden conocer ninguna de las causas de estas sin que adopten valores diferentes.
- Cualquier criterio de selección que se correlacione con la variable dependiente atenuará la media de estimaciones de los efectos causales.

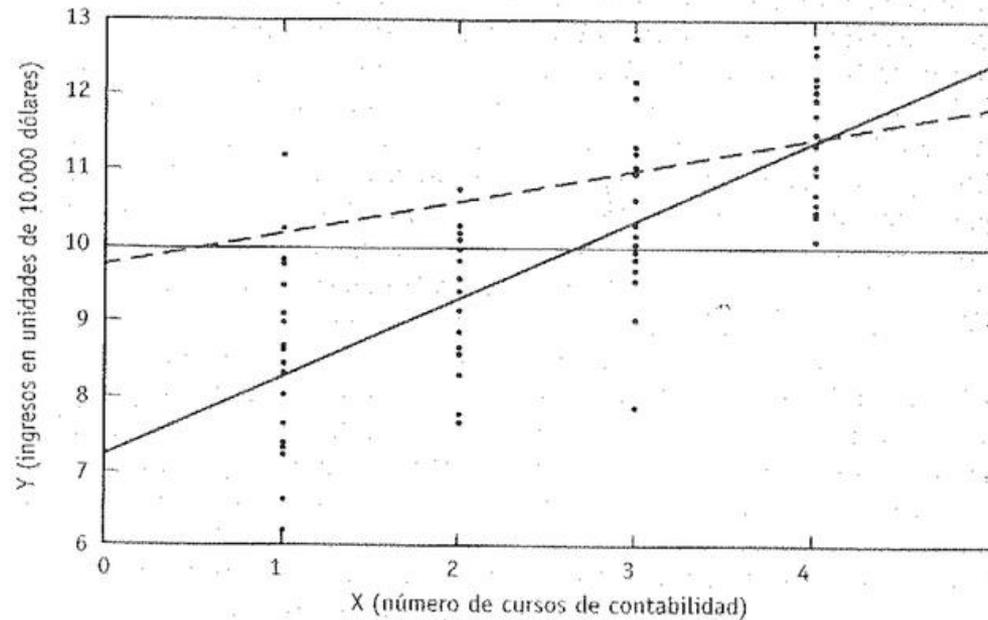
# Sesgo de selección: sobreestimación del efecto causal

- Si sabemos que existe un sesgo de selección, y no tenemos forma de esquivarlo mediante una muestra mejor, al menos sabemos que el VERDADERO EFECTO CAUSAL, TIENE UN LÍMITE MÁS BAJO.
- El grado de subestimación del efecto causal depende de la gravedad del sesgo de selección y de él, habría que tener al menos, cierta idea.

# Ejemplo de subestimación de efecto causal cuando seleccionamos en función de la variable dependiente

4. Precisar qué va a observarse

Gráfico 4.1 Sesgo de selección



# Ejemplos de sesgo de selección producido por el investigador

- Seleccionar casos a partir de datos disponibles
  - Si la disponibilidad de estos datos está relacionada con la variable dependiente
  - EJ.: Participación del presidente de EEUU en decisiones de política internacional importantes (reuniones secretas)

Evitar causas de sesgo: lista de declaraciones públicas del presidente respecto a política exterior

- Ej.: Porter “ventaja competitiva” en los actuales sectores industriales y empresas
  - Eligió 10 países en función de la variable dependiente (con ventajas competitivas) volviéndola prácticamente constante. Sesgo
  - Le permitió desarrollar hip. Sobre causas de la ventajas competitivas viendo lo que los países tenían en común
  - SIN EMBARGO: diseño de investigación le impidió evaluar efectos causales concretos

# Ejemplos de sesgo de selección producidos por el mundo

- Registros de observaciones externos no necesariamente evitan el sesgo en mi investigación
  - Ej.: Y: porcentaje de votos recibe candidato Partido Liberal en Nueva York para asamblea
  - El partido opta por no avalar a nadie en los casos en que va a perder (Nueva York x ej.)
  - El sesgo en este caso es externo pero es sesgo para nuestra investigación (como si lo produjéramos nosotros)

- Ej.: Esculturas de madera y de piedra
  - Sesgo a los historiadores del arte que concluyen que en Europa había más desarrollo artístico que en África

El investigador cuidadoso debe siempre evaluar los sesgos de selección que pudiera haber en las pruebas que dispone  
¿Qué clase de acontecimientos se han podido registrar? ¿Cuáles es posible que se hayan dejado de lado?

# Seleccionar en función de una variable explicativa

- No produce problemas inferenciales porque no restringe el grado de variación posible de la variable dependiente. No introduce SESGO
- Por tanto puede evitarse el sesgo seleccionando casos en función de la variable causal clave (experimentos)
  - También puede lograrse ese efecto introduciendo una Variable de Control
  - O por medio de una Variable irrelevante

# SELECCIÓN INTENCIONADA DE OBSERVACIONES

- Investigación social casi nunca tenemos control de variables explicativas
  - Selección intencionada de las observaciones
  - Para ello: conocer valores variables relevantes
  - Riesgo de sesgo de selección a favor de la hipótesis previa

- 1) Seleccionar observaciones en función de la variable explicatoria
  - Elegir una variable explicatoria sustancial y controlar las otras explicatorias no primordiales
- 2) Seleccionar un abanico de valores de la variable dependiente
  - Tomar aquellos valores especialmente bajos o altos de la variable dependiente
  - **IMPORTANTE:** para que este diseño posibilite inferencias causales deben seleccionarse observaciones representativas. No que se ajusten a la teoría inicial, para poder generalizar e inferir

3) Seleccionar observaciones utilizando tanto las variables explicativas como las dependientes

- Peligro de introducir un sesgo en el resultado.

- Mayor error: variables explicativas y causales variando al unísono

4) Seleccionar observaciones de manera que la variable causal clave sea constante

- considerar variable explicativa como constante es claramente deficiente para conocer el efecto causal

- 5) Seleccionar observaciones de manera que la variable dependiente sea constante
- Tampoco puede conocerse el efecto causal manteniendo constante la variable dependiente

# Comentarios Finales

- Este capítulo analizó la selección de las observaciones para conseguir un diseño que minimice el sesgo relativo al proceso de selección
- NO hay diseños perfectos! Hay estrategias imperfectas pero ÚTILES para dar cierto control sobre el problema del sesgo

- SUMA CAUTELA! A la hora de pensar el diseño de investigación
- Objetivo primordial:
  - obtener más información relevante para contrastar nuestra teoría
  - Sin por ello introducir un sesgo que ponga en peligro la calidad de las inferencias

**MUCHAS GRACIAS!!**