



UNAM
POSGRADO



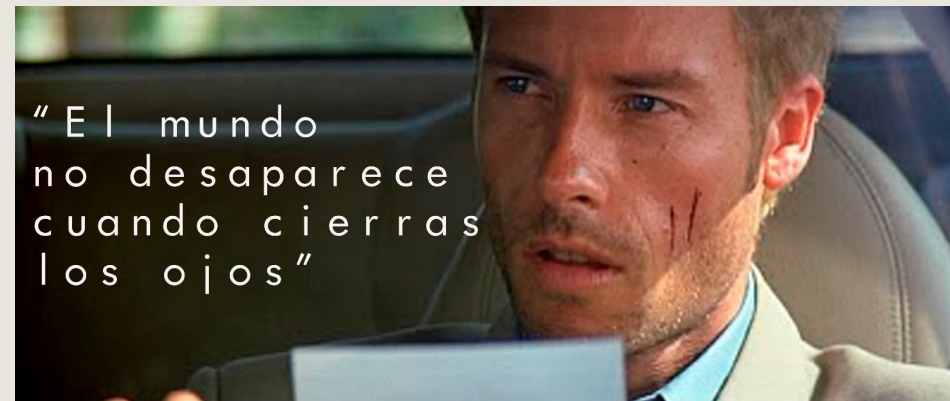
Programa
Universitario
de Estudios
del Desarrollo
UNAM

Modelos multinivel

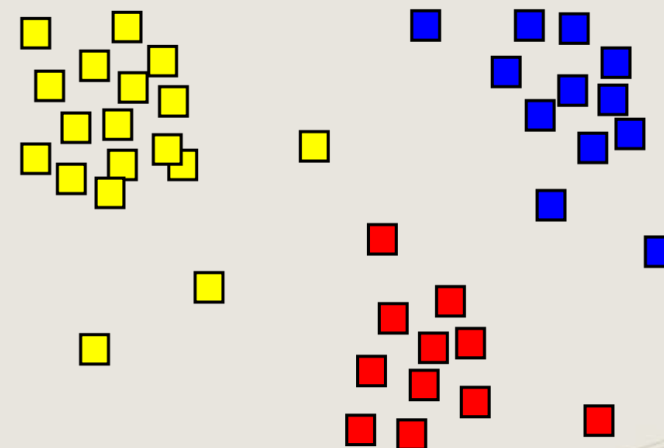
Dr. Héctor Nájera
Dr. Curtis Huffman

¿Modelos con memoria?

- ¿Nada aprendido en una categoría sirve en otra?
- ¿No es mejor suponerles iguales?

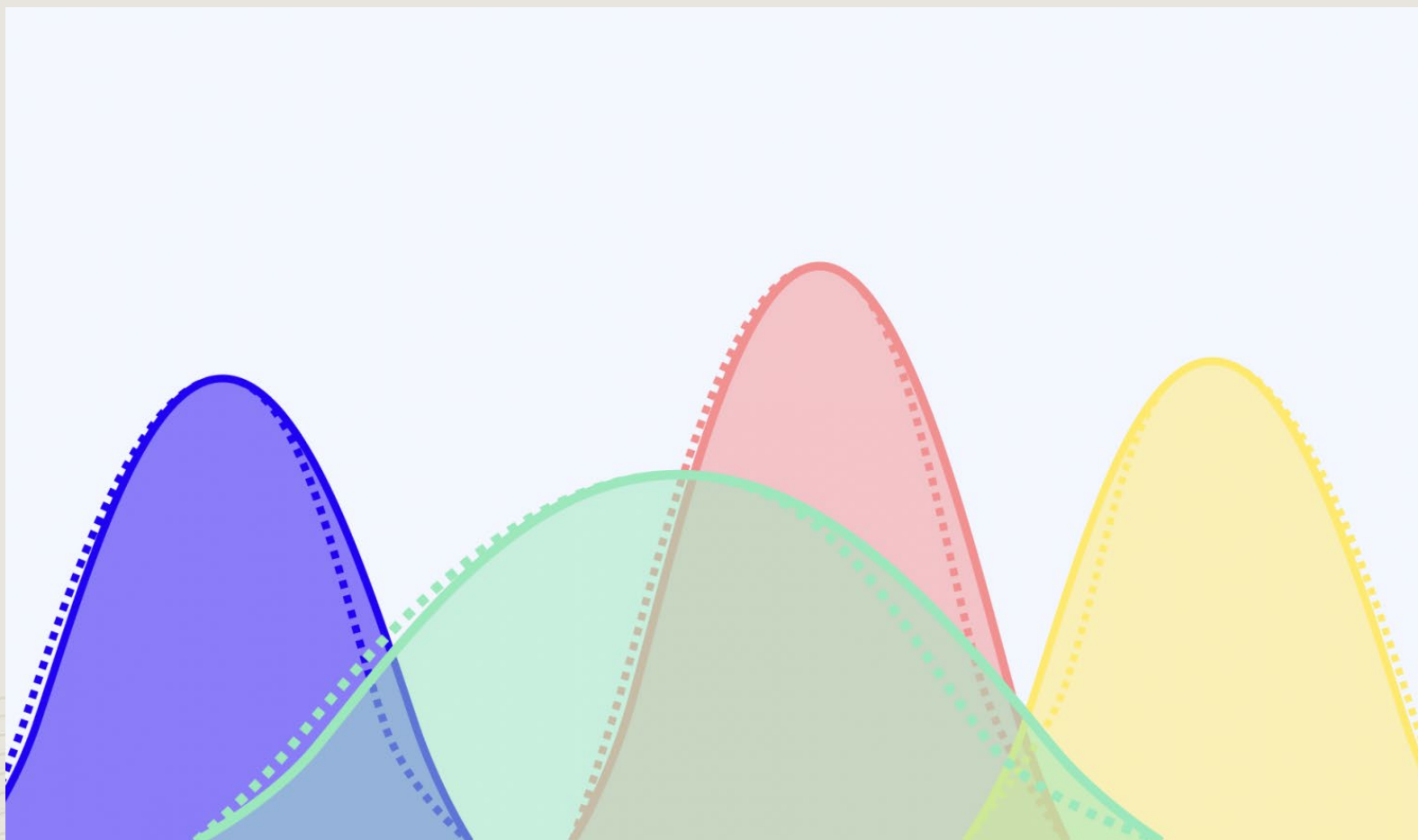


- ¿Nada que aprender de la población de grupos?
- Aprender de cada grupo al tiempo que aprendemos de la población de grupos (efectos variables)



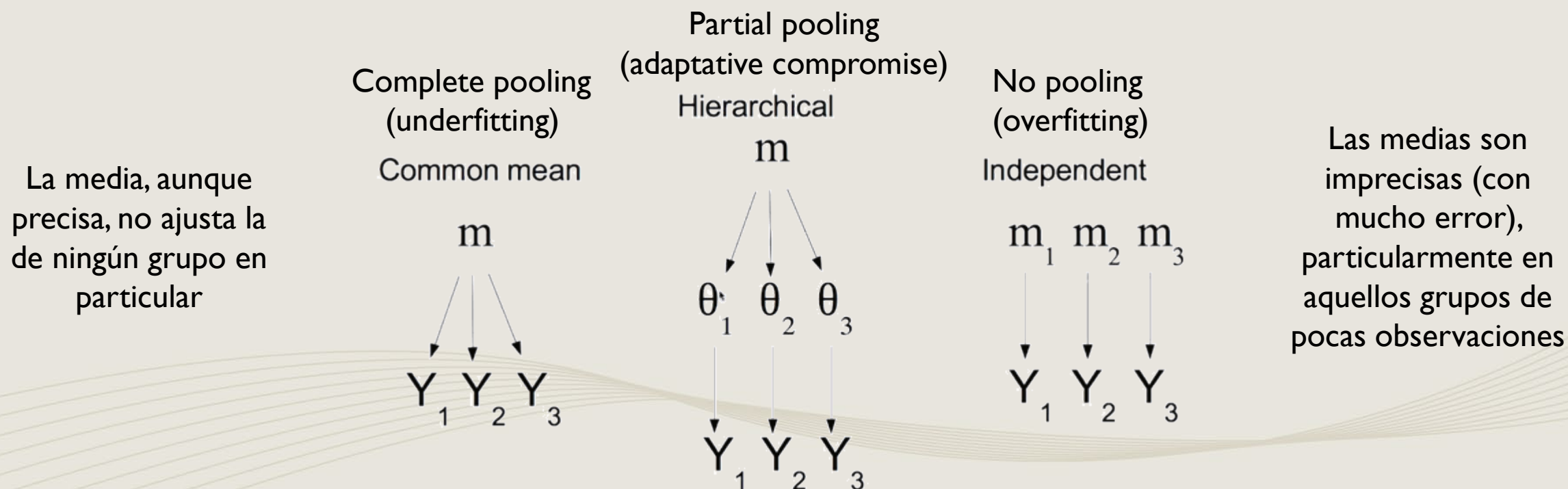
¿Dijo usted categorías?

- El nombre del juego en modelación multinivel es *heterogeneidad*



¿Dijo usted categorías?

- El nombre del juego en modelación multinivel es *heterogeneidad*
- Pueden pensarse como un punto medio entre ajustar modelos diferentes para diferentes subconjuntos de datos y ajustar un mismo modelo para el conjunto completo de datos

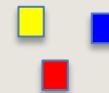


- Los modelos multinivel/jerárquicos/efectos-mixtos/efectos-aleatorios son modelos anidados (modelamos el prior, con un hiperprior común a los grupos, dejándolo aprender adaptativamente de los datos)
 1. Modelo de los grupos observados
 2. Modelo de la población de grupos o individuo
- Transfiere información entre grupos al aprender sobre su población



Ventajas de la modelación multinivel

- Mejora la estimación de cada grupo
 - Mejora las estimaciones de observaciones repetidas (aprenden más rápido y mejor)
 - Mejora las estimaciones de muestras desbalanceadas
 - regulariza (resiste overfitting) tomando prestada (transfiriendo) información de otros grupos más numerosos
 - Estima explícitamente la variación entre individuos o grupos (con suficientes datos/grupos)
 - Preserva la incertidumbre (conserva la variación) evitando pérdida de información al transformar los datos



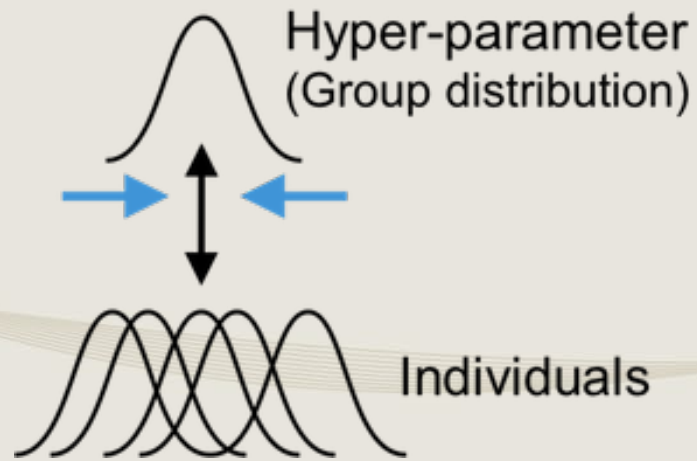
Ventajas de la modelación multinivel

- Incluyen información respecto a cómo fueron recolectados y ordenados los datos
 - Estructuras anidadas conceptual, experimental, temporal, geográficamente que sugieren que tratarles como independientes es dejar (“cosa buena sobre le mostrador”) sin explotar información relevante



Ventajas de la modelación multinivel

- Permite compartimentalizar la incertidumbre
- La estructura de dependencia entre los parámetros permite que la estimación, incluyendo la relativa a subconjuntos de datos, tome “prestada” información (fuerza) de todo el conjunto de datos
 - Cada parámetro individual es informada simultáneamente por datos de las demás unidades (todos los individuos informan los parámetros de nivel superior, que por su parte restringen a todos los parámetros individuales)



Modelos multinivel

- Requieren más estructura
 - Definir las distribuciones de las que surgen las características de los grupos
 - Pueden ser difíciles de
 - estimar (a veces)
 - interpretar (muchos compartimentos)
 - comparar (contrastes más sutiles entre modelos)



¿Dijo usted anidados?

- Los modelos multinivel involucran múltiples parámetros de tal manera que los valores creíbles de algunos parámetros dependen significativamente de los valores de otros parámetros.
- En otras palabras, hay cadenas de dependencia entre los parámetros.



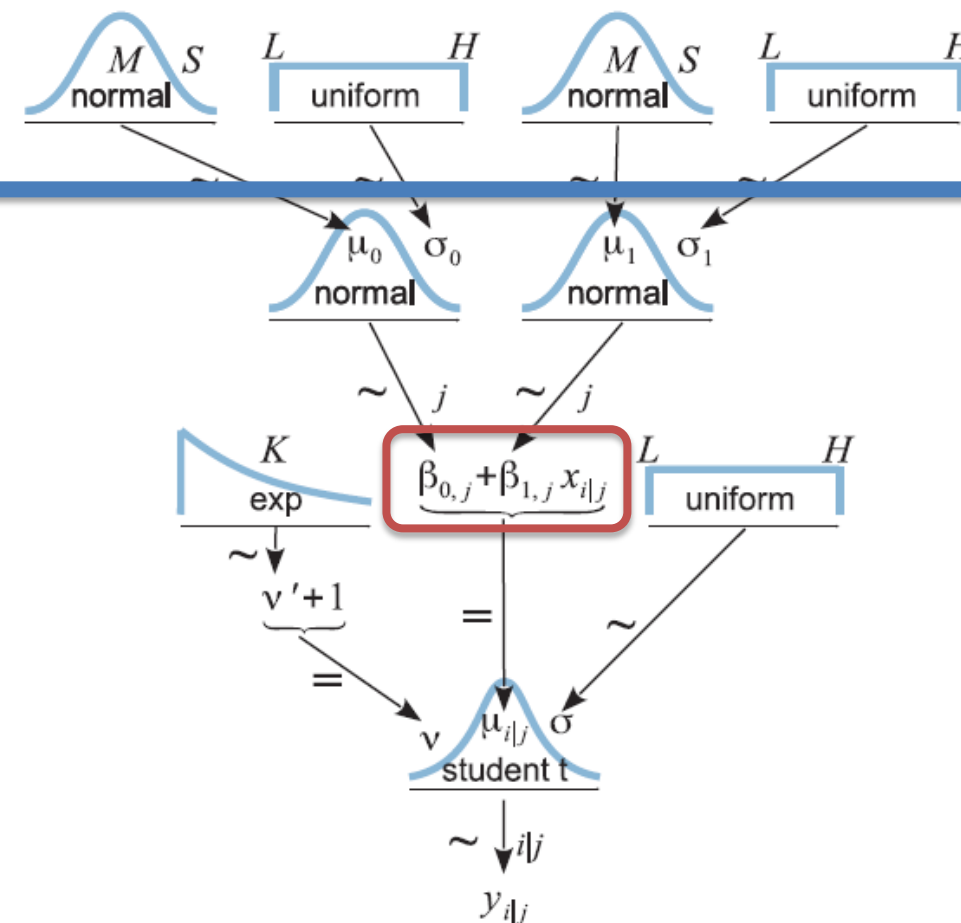
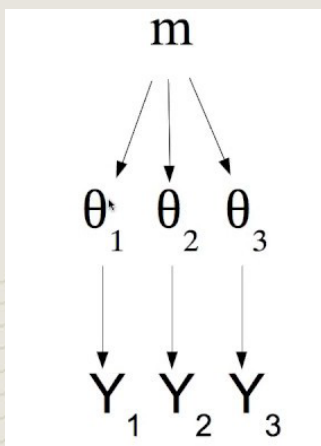
Otra vez, ¿dijo usted jerárquico?

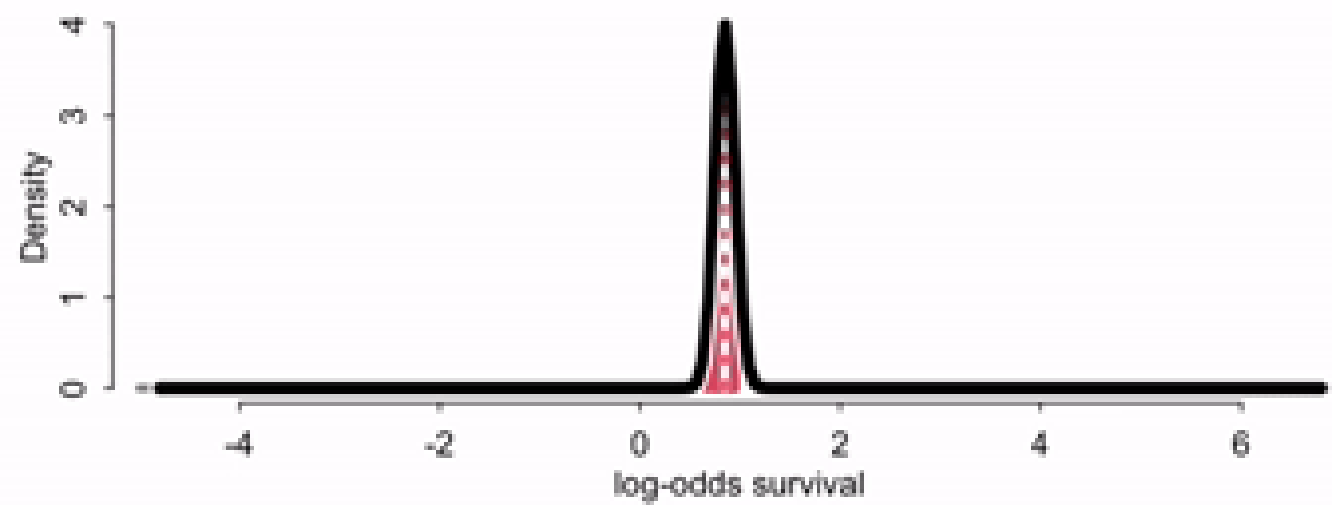
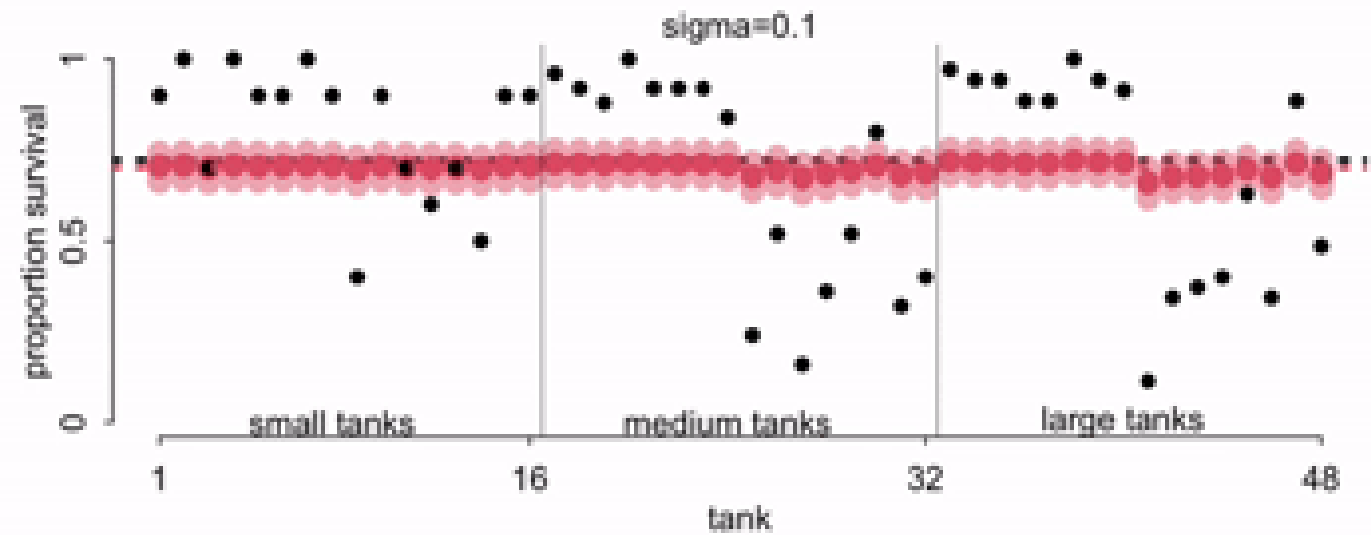
Los parámetros de niveles superiores son informados por **TODOS** los datos

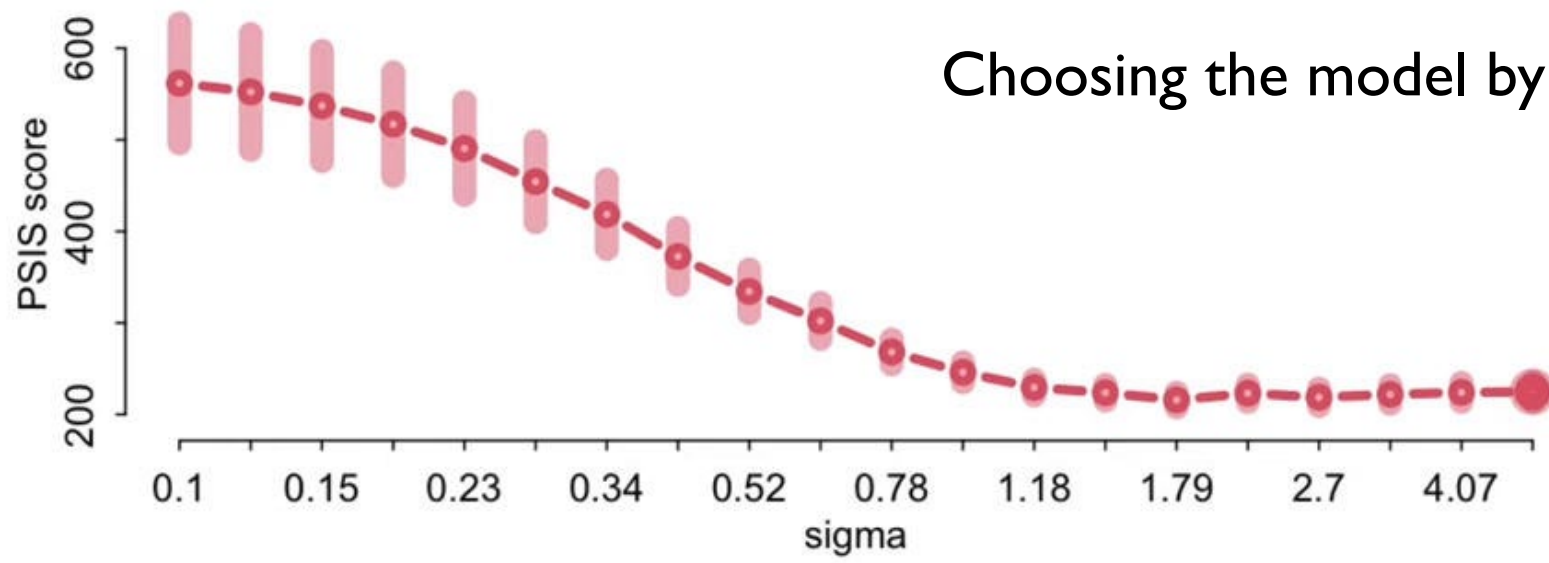
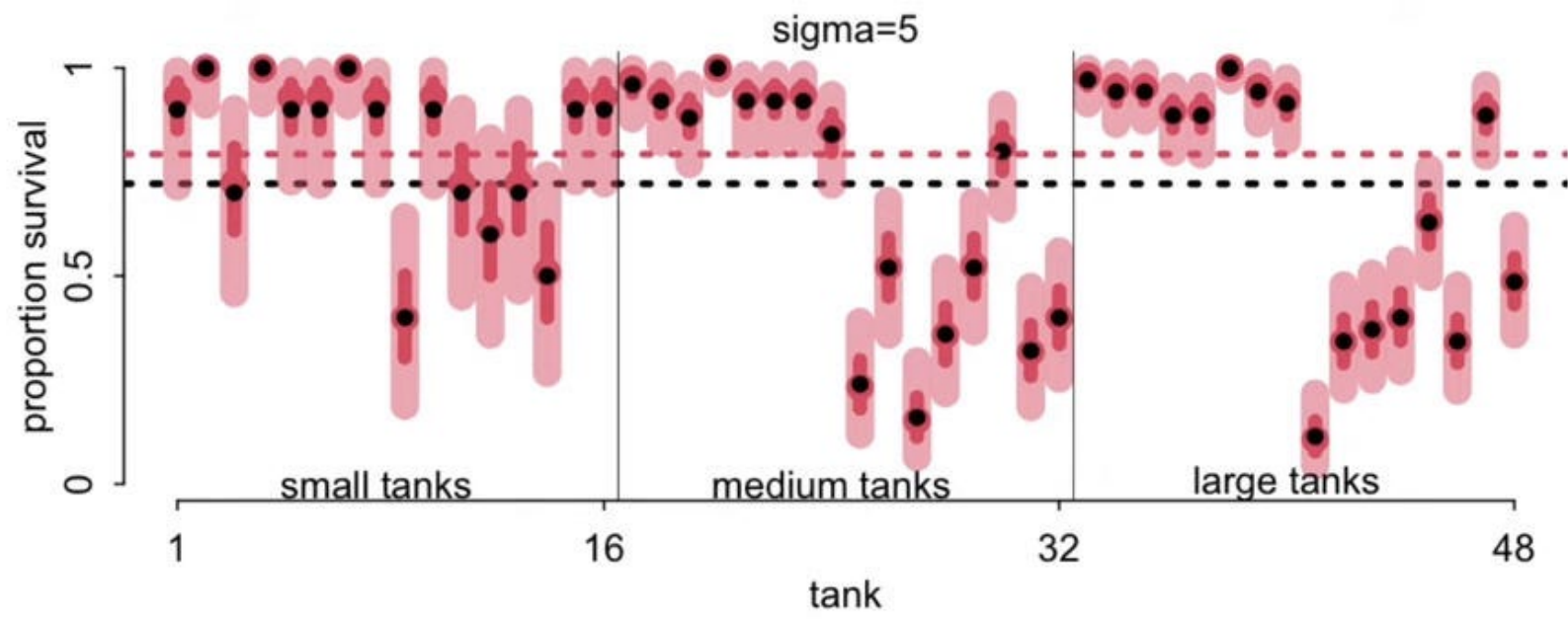
Reexpresan la verosimilitud como una cadena de dependencia entre parámetros

$$p(\theta, \omega | D) \propto p(D | \theta, \omega) p(\theta, \omega) \\ = p(D | \theta) p(\theta | \omega) p(\omega)$$

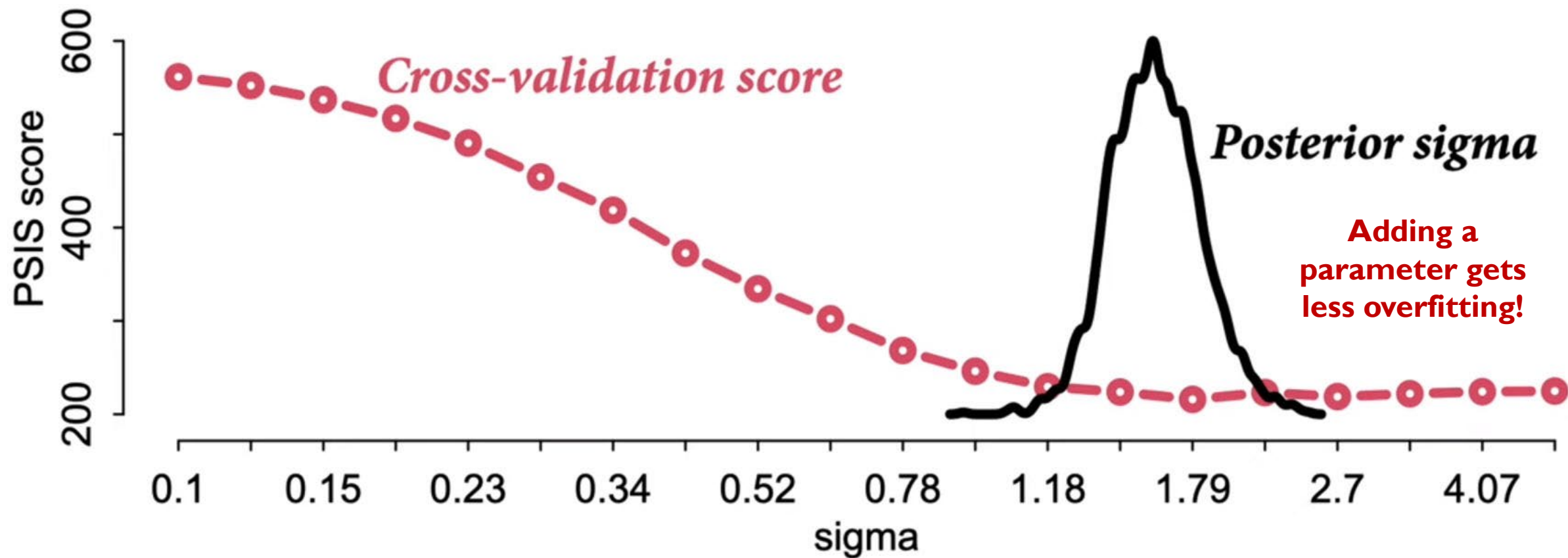
Los datos terminan por depender sólo de θ

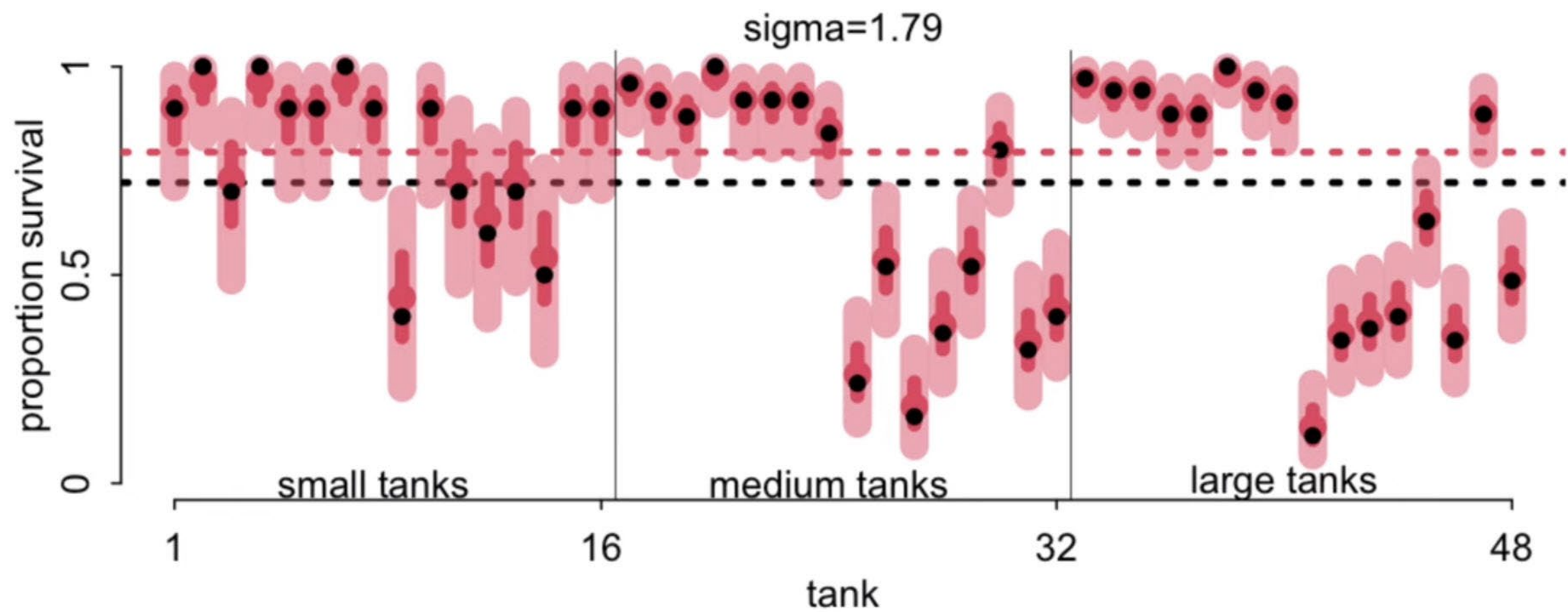






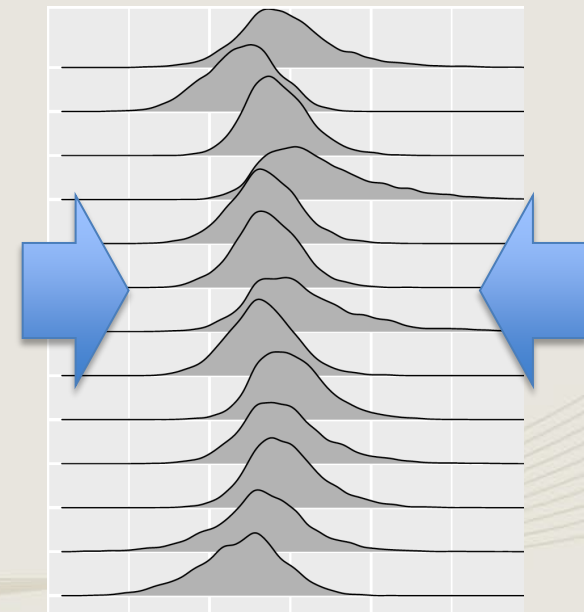
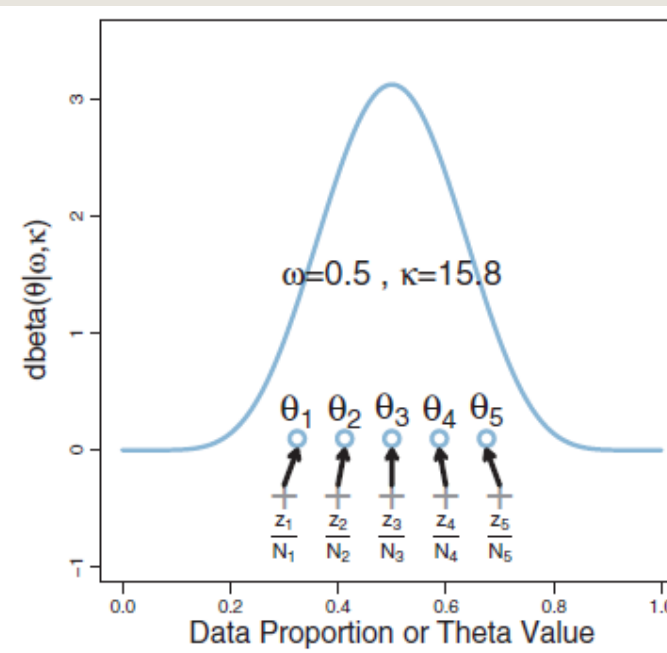
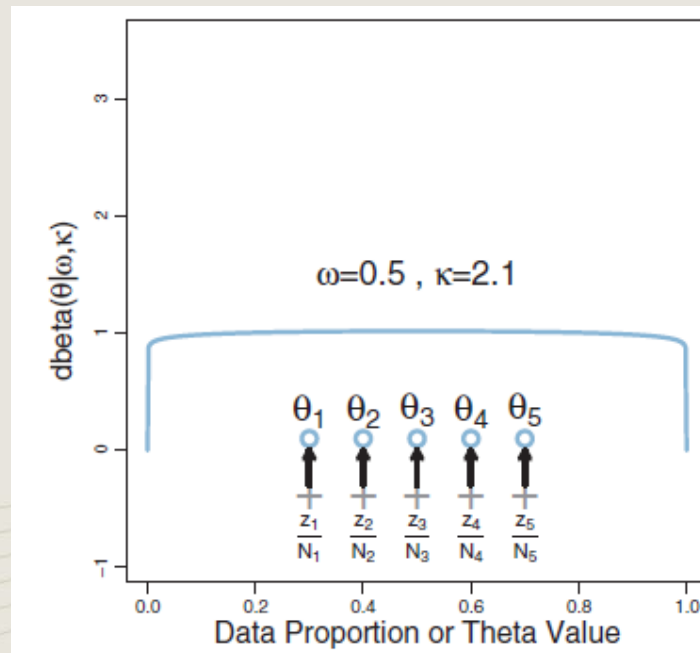
| | mean | sd | 5.5% | 94.5% | n_eff | Rhat4 |
|-------|------|------|------|-------|-------|-------|
| a_bar | 1.35 | 0.26 | 0.93 | 1.78 | 2746 | 1 |
| sigma | 1.61 | 0.21 | 1.32 | 1.96 | 1557 | 1 |





Shrinkage

- En los modelos multinivel, las estimaciones de los parámetros de los niveles inferiores están más cerca unos de otros de lo que lo estarían si no modeláramos los niveles superiores.
 - Los parámetros de niveles inferiores son atraídos hacia las modas de la distribución del nivel superior debido a la (nueva) estructura de dependencia entre ellos.




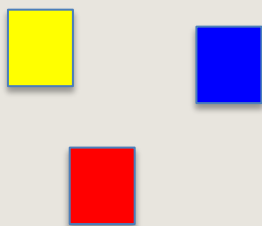
Viejos malentendidos

- ¿Cuándo deberíamos usar modelos multinivel?
 - ~~Cuando los grupos sean muestreados aleatoriamente~~ **X**
- El modelaje multinivel es una técnica de “aprendizaje” (estimación de la posterior), no una descripción del proceso que dio origen a los datos
 - Los datos pueden tener cualquier origen
 - Siempre que haya información respecto a cómo fueron acopiados los datos debería incorporarse al modelo, pero esa es otra historia.



Viejos malentendidos

- ¿Cuándo deberíamos usar modelos multinivel?
 - ~~Cuando se tengan muchos grupos (muestra grande)~~ 
- El único límite son los grados de libertad (#ecuaciones en relación con #incógnitas; no exactamente)



- Cuando el objetivo es estimar la varianza misma (entre/inter), la incertidumbre es función del tamaño de muestra como siempre, pero esa es otra historia.



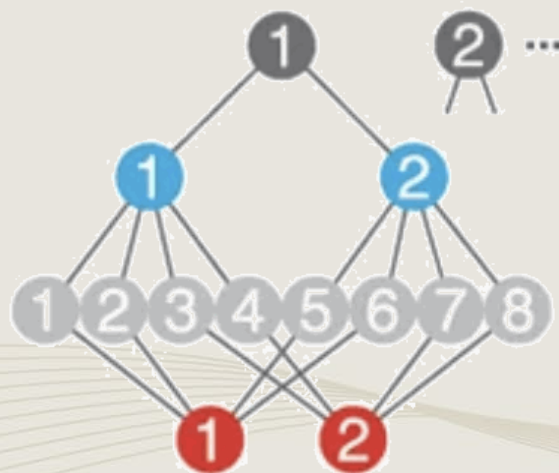
Viejos malentendidos

- ¿Cuándo deberíamos usar modelos multinivel?
 - ~~Cuando el supuesto de normalidad tenga sentido~~ **X**
 - ~~Cuando la distribución de clusters tenga sentido~~ **X**
 - ¿Sexo?
- Los priors no son una afirmación, compromiso o hipótesis de cómo se distribuyen los datos, sino una descripción del estado del conocimiento actual.
- La distribución normal es la de máxima entropía entre las que cuentan con media y varianza finitas.
- Priors normales no implican posteriores normales (es siempre posible aprender no-normalidades de los datos)



Viejos malentendidos

- ¿Cuándo deberíamos usar modelos multinivel?
 - La respuesta correcta es: siempre que se pueda (arrancar siempre ahí por omisión)
 - Noten que la anidación puede ser cruzada y tener tantos niveles como sean necesarios (tenga sentido)





Ejemplo: Estimación de la inseguridad alimentaria a nivel municipal

Objetivo: Tener estimaciones de inseguridad alimentaria a nivel municipal

Descripción del problema:

- La ENIGH no es representativa a nivel municipal
- Alta variabilidad geográfica
- Nos gustaría estimar un modelo predictivo con la ENIGH que después pudiéramos aplicar al Censo
- Hay muchos modelos candidatos. ¿Qué se les ocurre?



CONEVAL: Modelo logit por regiones

- La variación regional es muy alta
- Un modelo nacional quizá sería demasiado general
- Un modelo por estado no es muy parsimonioso (32 MODELOS)
- CONEVAL: ¡Dividir al país en seis regiones y ajustar seis modelos logísticos!

CONEVAL no reproduce el valor de diseño –ENIGH- para prácticamente ningún estado.

Las diferencias rondan el 3%.

¿Es mucho? Miren que la inseguridad alimentaria no cambió ni con la pandemia entre 2018 y 2020!

Cuadro 4. Comparación del porcentaje de carencia por acceso a la alimentación obtenido de forma directa y a través de modelos logísticos, según entidad federativa, México, 2020

| Clave | Entidad | Porcentaje de carencia por acceso a la alimentación | | Diferencia absoluta entre estimación directa y por modelos |
|-------|---------------------|---|--------------------|--|
| | | Estimación por modelos | Estimación directa | |
| 00 | Nacional | 23.8 | 20.8 | 3.0 |
| 01 | Aguascalientes | 20.2 | 17.4 | 2.8 |
| 02 | Baja California | 16.2 | 12.7 | 3.5 |
| 03 | Baja California Sur | 26.2 | 22.7 | 3.5 |
| 04 | Campeche | 28.3 | 24.5 | 3.8 |
| 05 | Coahuila | 18.4 | 15.2 | 3.2 |
| 06 | Colima | 19.2 | 16.6 | 2.6 |
| 07 | Chiapas | 24.9 | 21.1 | 3.8 |
| 08 | Chihuahua | 16.4 | 13.0 | 3.4 |
| 09 | Ciudad de México | 18.5 | 15.9 | 2.6 |
| 10 | Durango | 21.8 | 18.7 | 3.1 |
| 11 | Guanajuato | 25.3 | 22.8 | 2.5 |
| 12 | Guerrero | 35.3 | 33.3 | 2.0 |
| 13 | Hidalgo | 28.5 | 25.3 | 3.2 |
| 14 | Jalisco | 17.9 | 14.3 | 3.6 |
| 15 | México | 25.0 | 21.4 | 3.6 |
| 16 | Michoacán | 24.8 | 21.4 | 3.4 |
| 17 | Morelos | 25.7 | 23.1 | 2.6 |
| 18 | Nayarit | 22.3 | 19.8 | 2.5 |
| 19 | Nuevo León | 16.6 | 14.1 | 2.5 |
| 20 | Oaxaca | 31.6 | 28.4 | 3.2 |
| 21 | Puebla | 30.4 | 27.3 | 3.1 |
| 22 | Querétaro | 19.9 | 17.0 | 2.9 |
| 23 | Quintana Roo | 29.3 | 27.3 | 2.0 |
| 24 | San Luis Potosí | 21.4 | 17.8 | 3.6 |
| 25 | Sinaloa | 23.9 | 22.0 | 1.9 |
| 26 | Sonora | 24.1 | 22.1 | 2.0 |
| 27 | Tabasco | 41.8 | 40.9 | 0.9 |
| 28 | Tamaulipas | 16.7 | 13.9 | 2.8 |
| 29 | Tlaxcala | 29.2 | 27.2 | 2.0 |
| 30 | Veracruz | 26.0 | 23.2 | 2.8 |
| 31 | Yucatán | 27.3 | 22.9 | 4.4 |
| 32 | Zacatecas | 19.6 | 16.4 | 3.2 |

Fuente: elaboración del CONEVAL con base en el Modelo Estadístico 2020 para la continuidad del MCS-ENIGH.





¿Qué supuestos tiene la aproximación del CONEVAL?

- Al partir la muestra partimos la variabilidad total
- Toda la variabilidad se busca explicar mediante características del hogar
- Bueno... CONEVAL viola este principio porque incluye efectos municipales –pero en realidad no es así-
 - Fíjense que el modelo no tiene forma de partir la varianza en individual y contextual
 - Este error es más común de lo que creen
- ¿Qué pasaría si le aplicamos este modelo al Censo?
 - Se propagaría el error hacia abajo –los municipios-

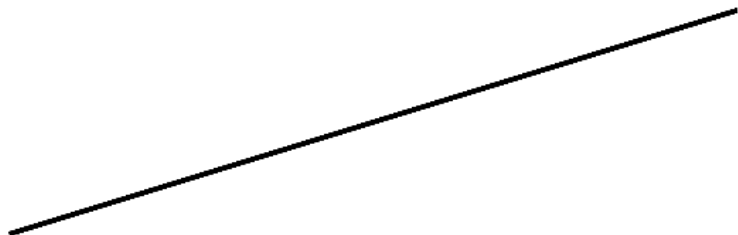
Modelación multinivel –jerárquica–

- Queremos aprovechar toda la variabilidad de la muestra pero también incorporar variación más “local”
- La modelación multinivel busca hacer esto:
 - Se usa toda la muestra y se agregan efectos aleatorios para refinar la modelación de lo local
 - Por ejemplo, permitir que cada estado y cada municipio tenga su propia ordenada al origen
 - Esto además **SÍ** permite incorporar efectos fijos con variables medidas a nivel estatal y municipal

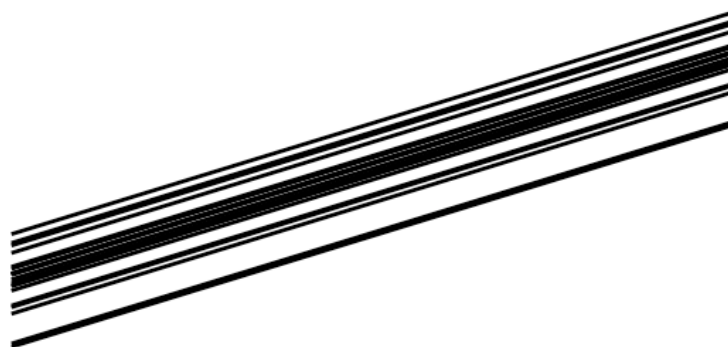


¿Ordenadas al origen particulares?

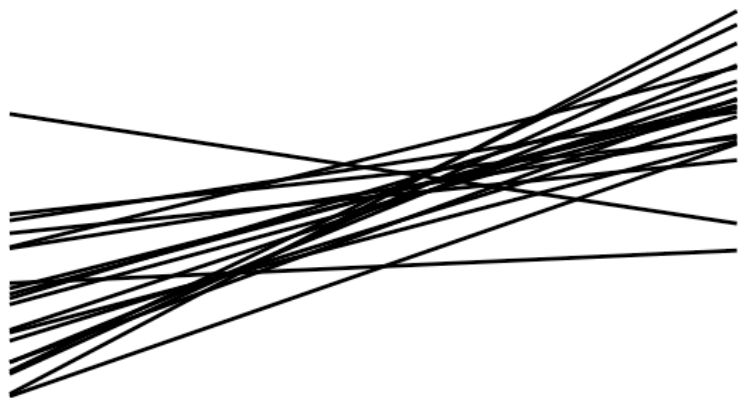
No-pooling



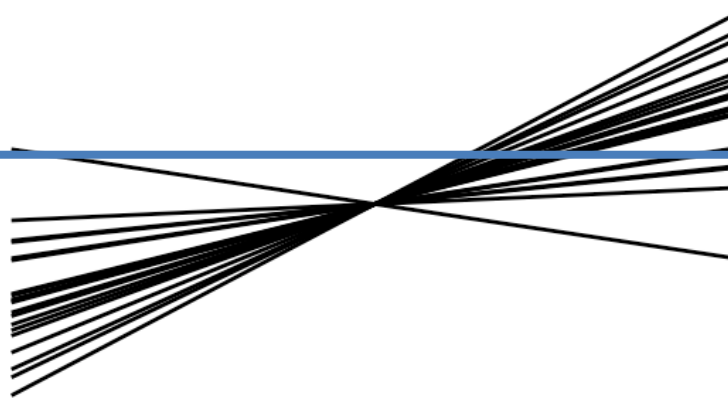
Varying intercept



Varying slope and intercept



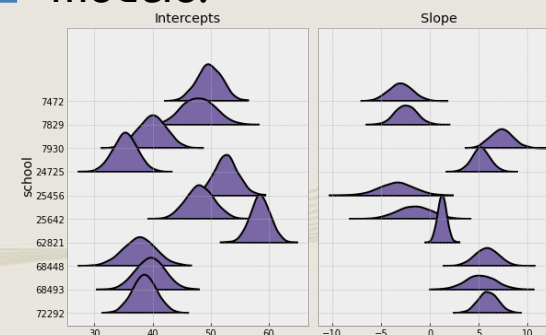
Varying Slopes



Cada estado tiene la suya y cada municipio también.

En un mismo modelo.

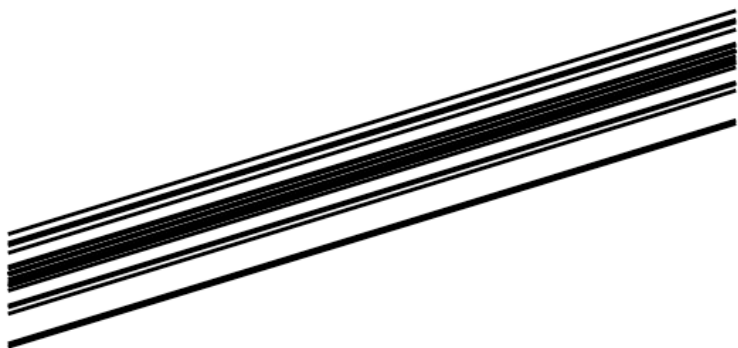
Incluso podrías hacer algo así en un mismo modelo.



¿Multinivel = Jerárquico?

- Los modelos jerárquicos lo son en el sentido de los parámetros, no de la estructura de los datos
- En bayes cualquier modelo es jerárquico –híper parámetros-
- Estructuras multinivel llevan a modelos jerárquicos (¿Por qué?)

Varying intercept



Puedo resumir las ordenadas al origen de los estado considerando la varianza de las mismas

$$\eta_k \sim \text{half-normal}(\mu_k, \sigma_k)$$

$$\mu_k \sim \text{normal}(\mu_{ent}, \sigma_{ent})$$



Contents lists available at ScienceDirect

Social Science Research

journal homepage: www.elsevier.com/locate/ssresearch



Bridging the gap between multilevel modeling and economic methods

Aleksey Oshchepkov^{a,*}, Anna Shirokanova^b

^a Centre for Labour Market Studies and Department of Applied Economics, HSE University, Moscow, Russian Federation

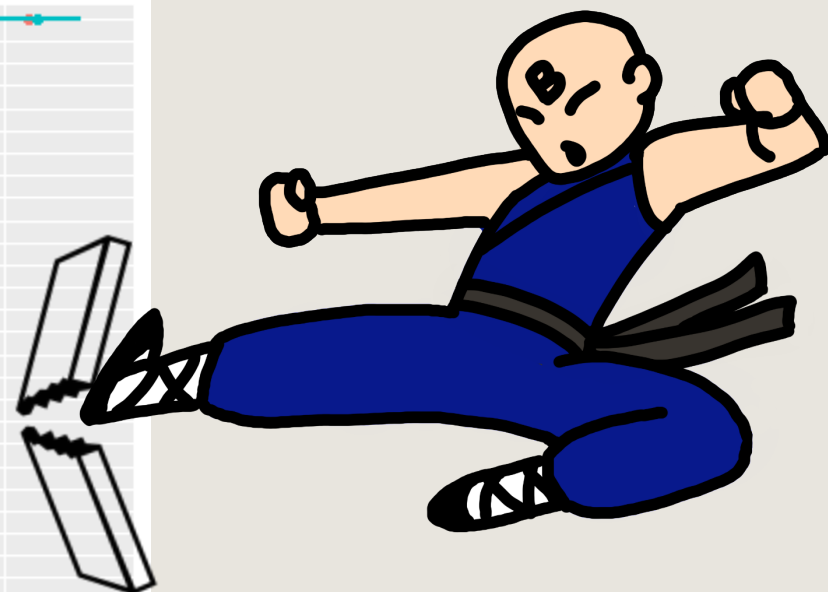
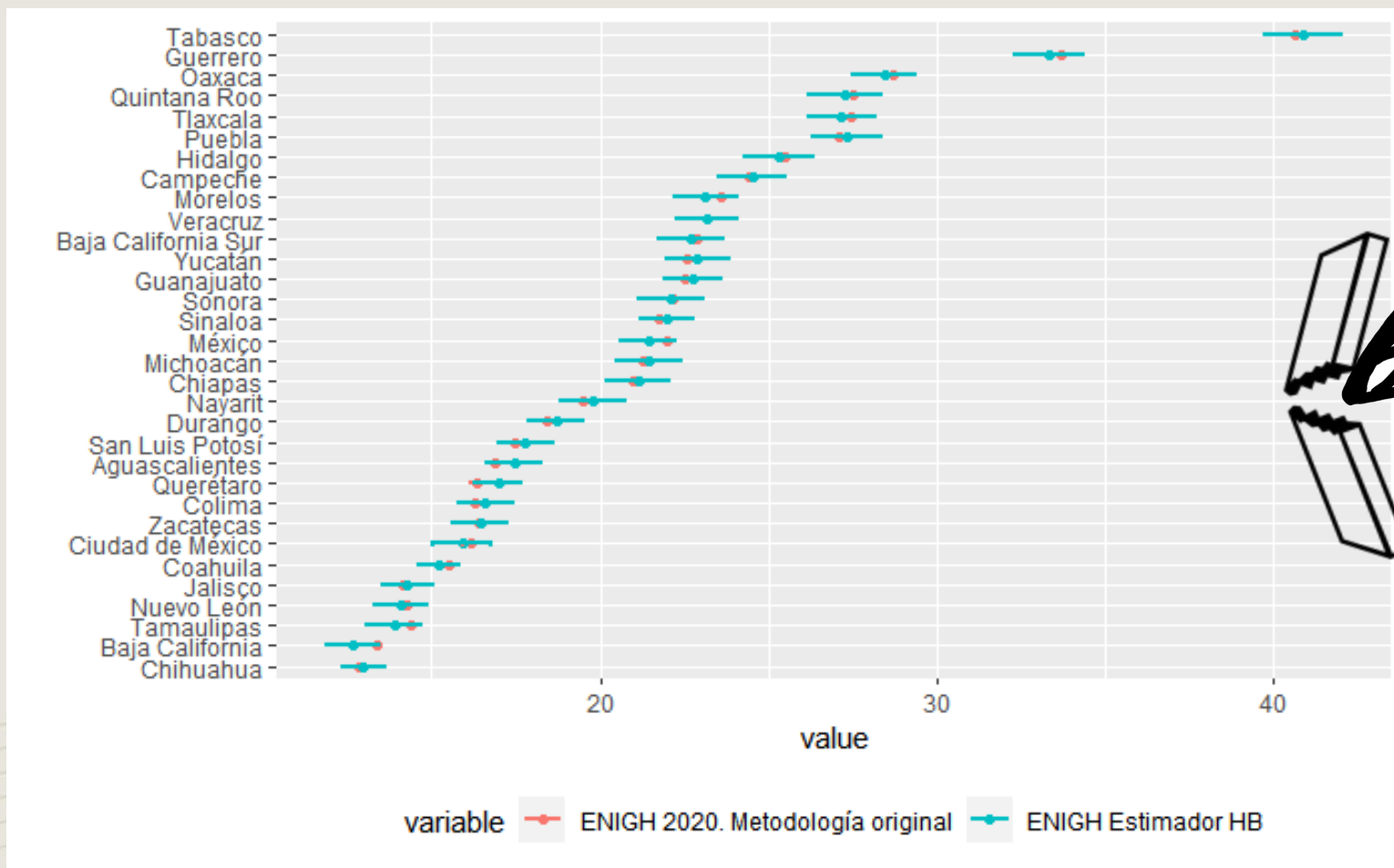
^b Ronald F. Inglehart Laboratory for Comparative Social Research and Department of Sociology, HSE University, Saint-Petersburg, Russian Federation

“In sum, MLM generally can do all the things that ‘economic methods’ should do in a nested data setting and do the even better”

Con modelación jerárquica

El modelo reproduce para todos los estados el valor del estimador de diseño. El error es seis veces menor al del modelo de CONEVAL!

¡La magia del shrinkage de los modelos multinivel!





Referencias

Kruschke, J. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*.

McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan*.
CRC press.

<https://youtu.be/iwVqiiXYeC4?si=fLQcm9tqqJFI2hBS>

<https://youtu.be/sgqMkZeslxA?si=cd3mplPKbNwhS0sL>