Inferencia bayesiana de la hipótesis nula

Dr. Héctor Nájera
Dr. Curtis Huffman

Nipótesis nula y descubrimiento

```
. regress bwght cigs male parity faminc faminc2

      Source |       SS           df       MS      Number of obs   =     1,387
-------------+----------------------------------   F(5, 1381)      =     12.54
       Model |  20057634.7         5  4011526.94   Prob > F        =    0.0000
    Residual |   441644355     1,381  319800.402   R-squared       =    0.0434
-------------+----------------------------------   Adj R-squared   =    0.0400
       Total |   461701990     1,386  333118.319   Root MSE        =    565.51

------------------------------------------------------------------------------
       bwght |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        cigs |  -13.25496   2.588155    -5.12   0.000     -18.3321   -8.177817
        male |   91.00571   30.45073     2.99   0.003     31.27103    150.7404
      parity |   47.51602   17.07493     2.78   0.005     14.02041    81.01163
      faminc |   8.660125   3.013622     2.87   0.004     2.748353     14.5719
     faminc2 |  -.0826576   .0414853    -1.99   0.047    -.1640386   -.0012765
       _cons |   3115.172   57.80237    53.89   0.000     3001.782    3228.562
------------------------------------------------------------------------------
```

¿De qué hablamos cuando hablamos de esto?

La clase pasada

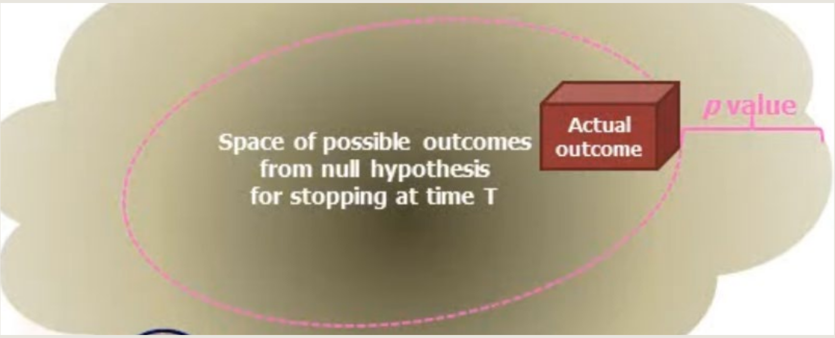
Rechazando la nula (Meehl paradox):

1. La teorías deberían ser más difíciles de confirmar a medida que tenemos más precisión (Todas las teorías están mal)
2. ¿La teoría se confirma rechazando la nula?
3. La nula no puede ser exactamente verdadera
4. Rechazar la nula es más fácil a medida que la precisión aumenta!

Solución a la paradoja de Meehl

- Buscar aceptar el valor que se predice y no buscar rechazar un valor nulo no predicho

- Esto significa aumentar la precisión de la predicción del valor predicho



AH!

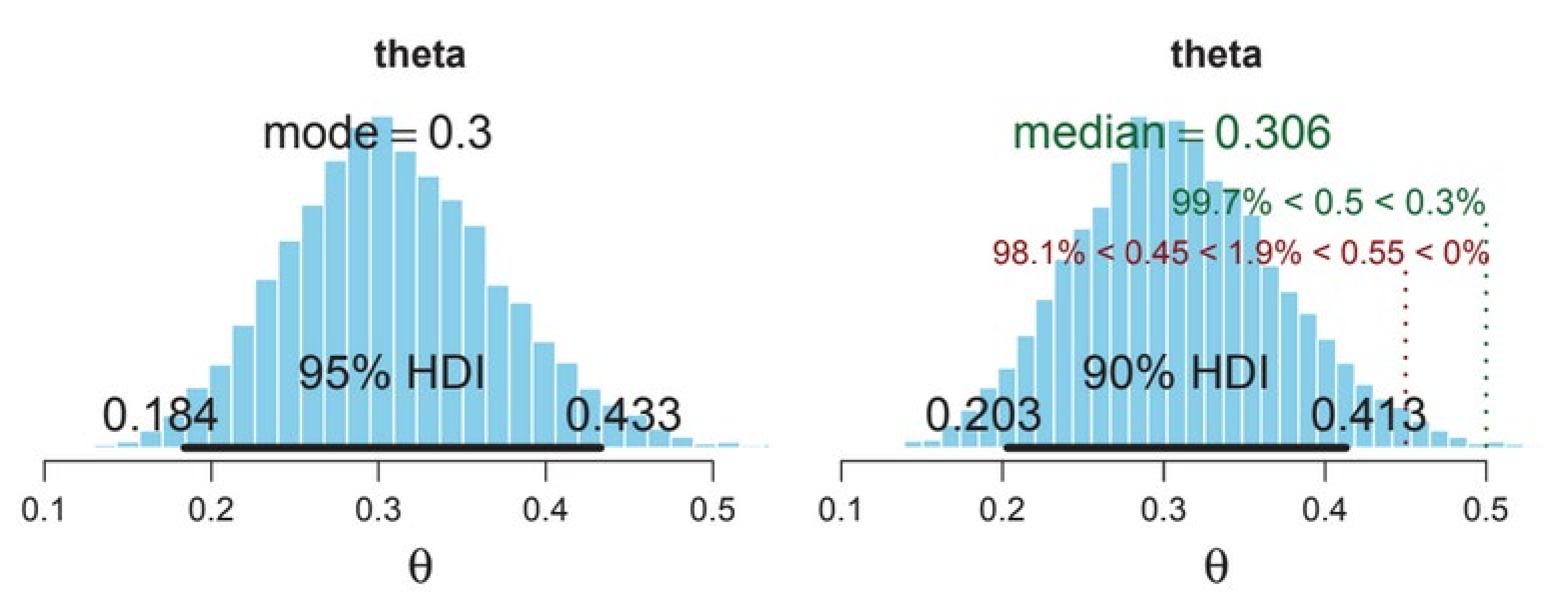
- New statistics: P-values<.001, Tamaños de efecto, intervalos de confianza y meta análisis como medios para evitar los problemas asociados a pruebas de significancia de las hipótesis nula (NHST).

- Pero como vimos la clase pasada, esto no lo resuelve.

- P-value: Es la probabilidad de que el estadístico (t) su hubiera obtenido de la nula (dada cierta regla respecto al tamaño de los resultados posibles).

- La precisión es la meta

- HDI: Medición de la precisión del valor posterior del parámetro



Bayes e hipótesis nulas. Dos caminos

- ¿Es creíble el valor "nulo" de un parámetro?

  1. Estimación de parámetros: ¿el valor de interés "cae" dentro del HDI?

  2. Comparación de modelos con diferentes a prioris (según los diferentes valores que admitan del parámetro)

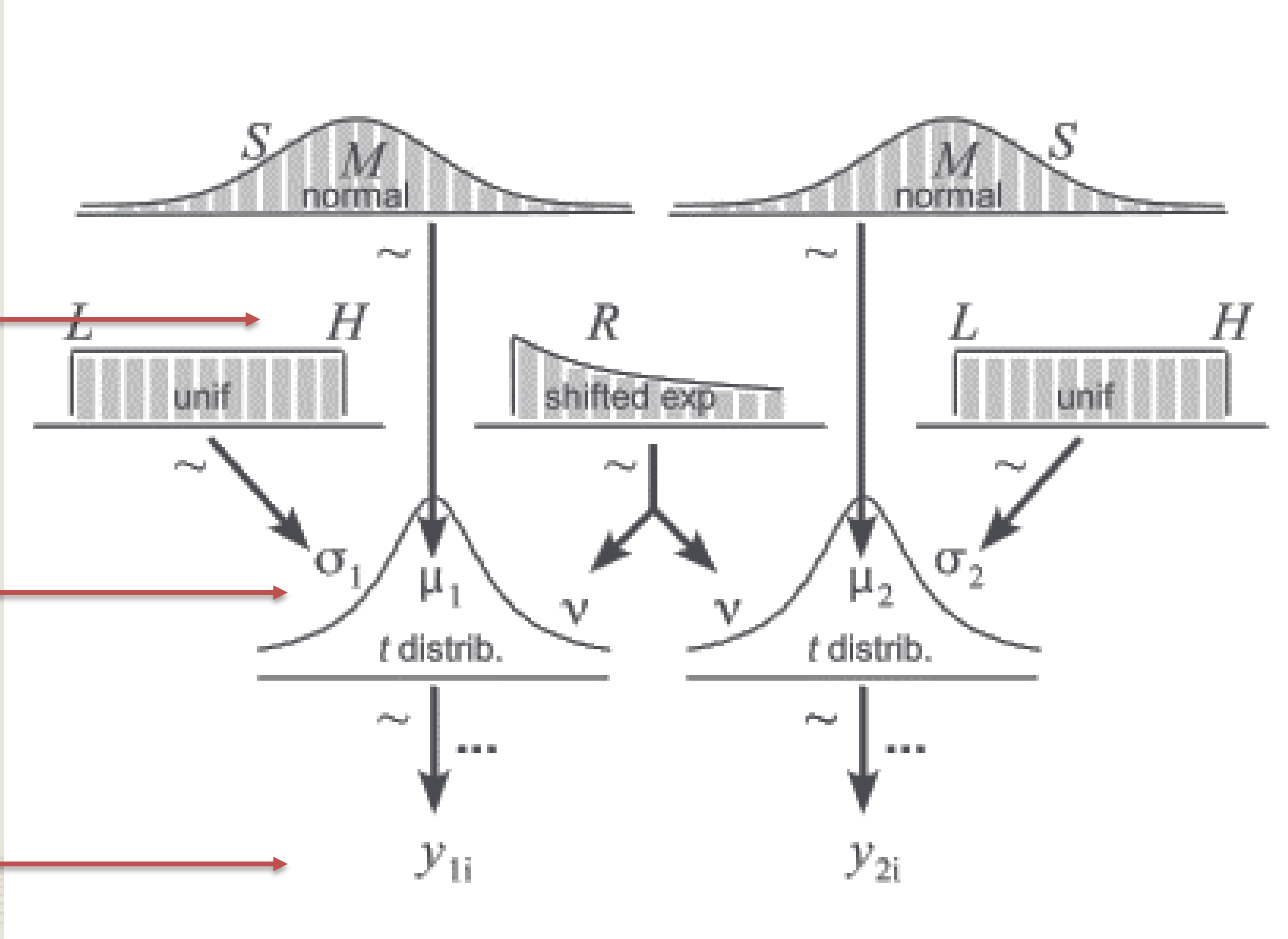
Algunas consideraciones de la inferencia bayesiana

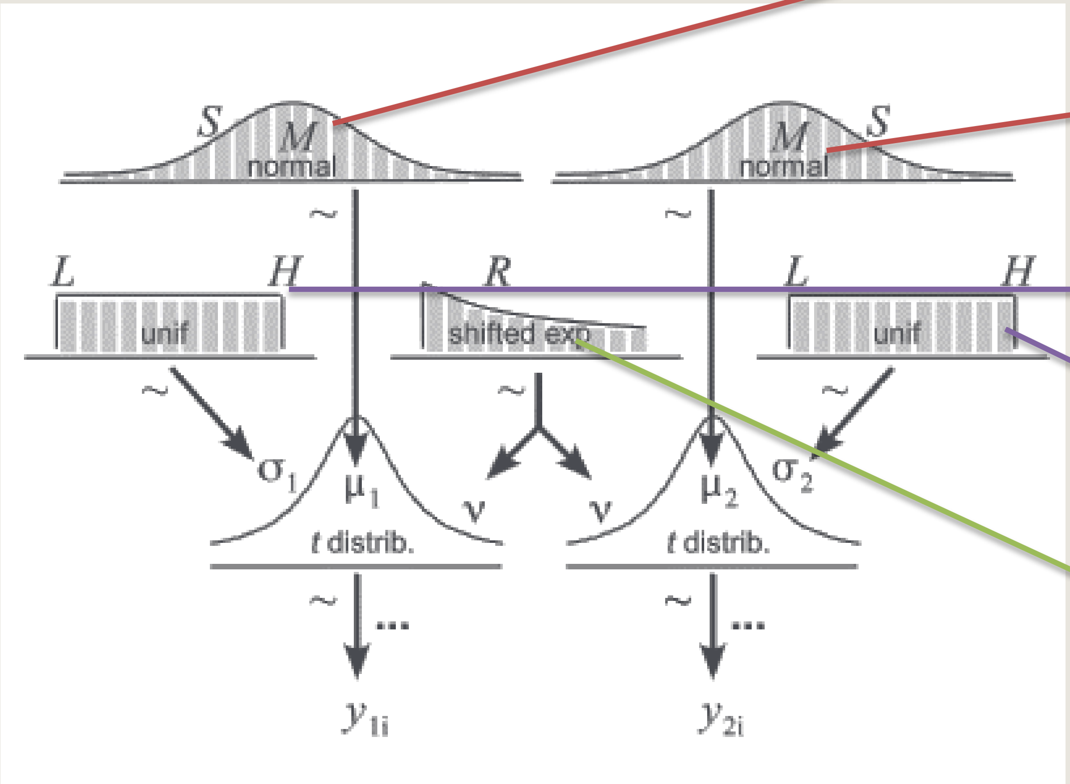
Hipótesis: Diferencias entre dos grupos

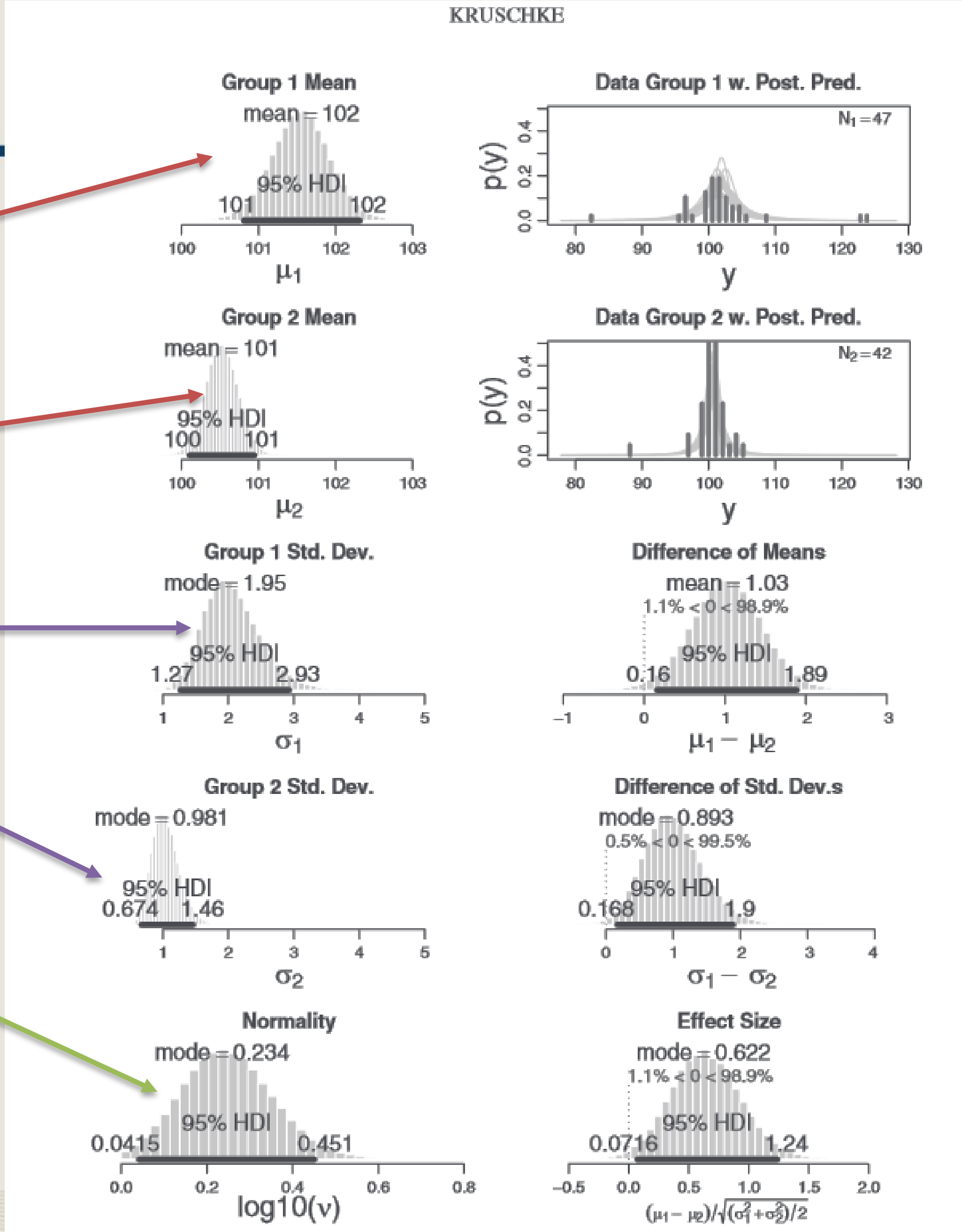
Priors: Qué sabemos de cada parámetro antes de estos datos

Los datos de cada grupo se describen con distribuciones t usando parámetros: mu, nu y sigma

Los datos “fijos”




S
M
normal
M
normal
S
L
H
R
L
H
unif
shifted exp
unif
σ1
μ1
ν
ν
μ2
σ2
t distrib.
t distrib.
y1i
y2i


KRUSCHKE
Group 1 Mean
mean = 102
95% HDI
101
102
μ1
Data Group 1 w. Post. Pred.
N1 = 47
p(y)
y
Group 2 Mean
mean = 101
95% HDI
100
101
μ2
Data Group 2 w. Post. Pred.
N2 = 42
p(y)
y
Group 1 Std. Dev.
mode = 1.95
95% HDI
1.27
2.93
σ1
Difference of Means
mean = 1.03
1.1% < 0 < 98.9%
95% HDI
0.16
1.89
μ1 − μ2
Group 2 Std. Dev.
mode = 0.981
95% HDI
0.674
1.46
σ2
Difference of Std. Dev.s
mode = 0.893
0.5% < 0 < 99.5%
95% HDI
0.168
1.9
σ1 − σ2
Normality
mode = 0.234
95% HDI
0.0415
0.451
log10(ν)
Effect Size
mode = 0.622
1.1% < 0 < 98.9%
95% HDI
0.0716
1.24

Combinando las posteriores

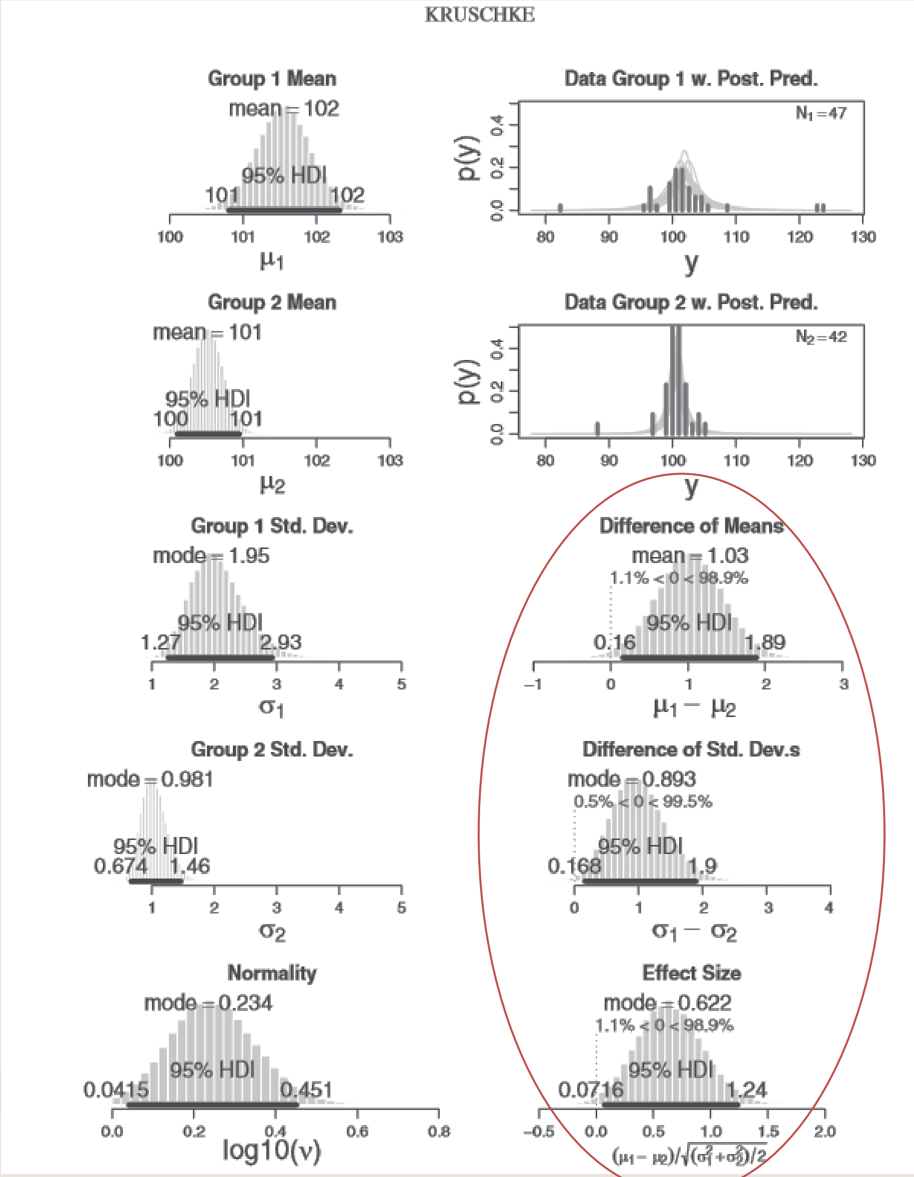
El computo bayesiano ofrece distribuciones de los valores posibles de los parámetros de interés:

Diferencia de medias, de escala (dispersión) y del tamaño del efecto

$$\mu_1 - \mu_2$$

$$\sigma_1 - \sigma_2$$

$$\left(\mu_1 - \mu_2 \sqrt{\frac{\sigma_1^2 + \sigma_2^2}{2}}\right)$$



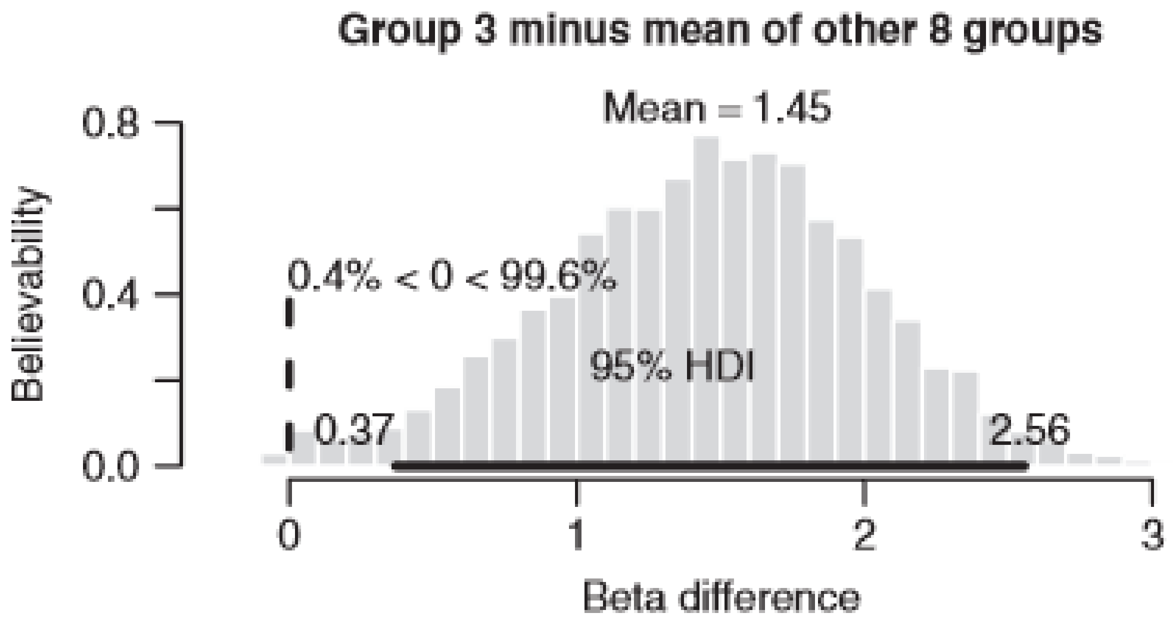


FIGURE 6 | Posterior estimate for difference of groups in Solari et al.[36] The third group is credibly different from the mean of the other eight groups.

HDI Bayesianos

El HDI de 95 % son los valores de Θ que tienen al menos ese nivel mínimo de credibilidad posterior.

- A diferencia del IC
  - El HDI tiene una interpretación directa en términos de la credibilidad de los valores de Θ
  - Es explícitamente acerca de $p(\Theta|D)$, que es exactamente de lo que queremos hablar
  - El HDI no depende del diseño de la investigación (el número de pruebas)
  - El HDI responde a las creencias a priori de las investigadoras, como debe ser.

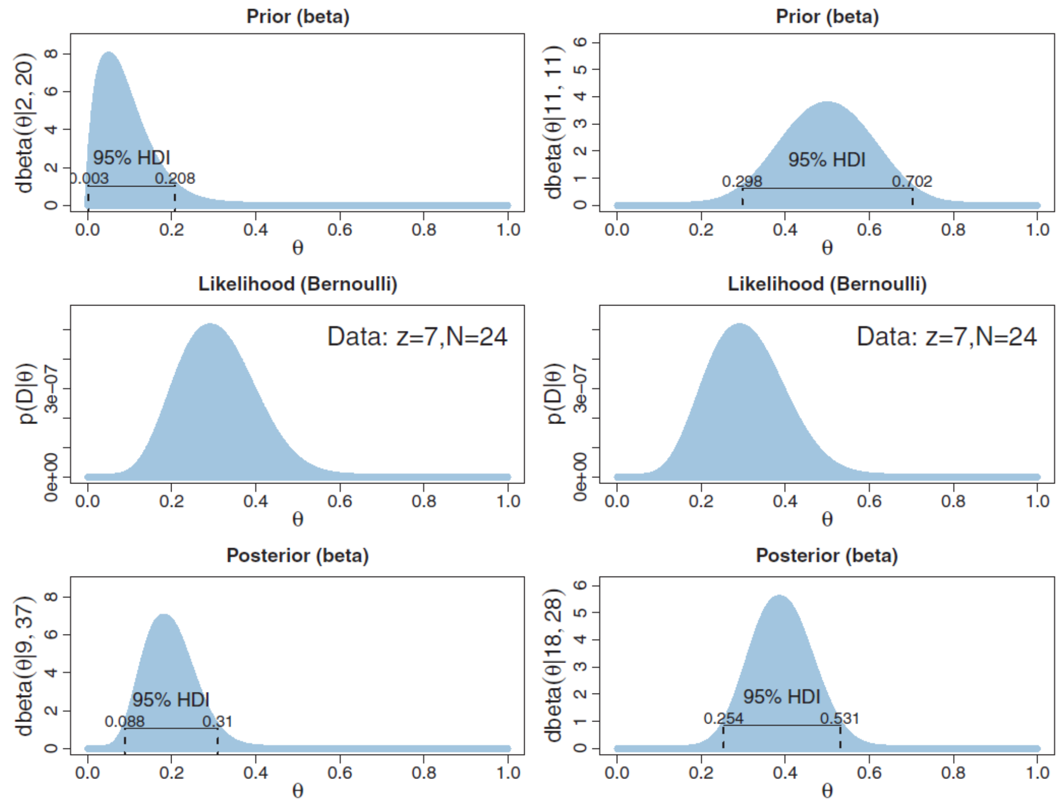


Figure 11.7 Posterior HDI for the bias of a Bernoulli process, when the prior assumes a tail-strong nail (left column) or a fair coin (right column).

Bayes vs frecuencias

Tenemos distribuciones completas de los valores de los parámetros

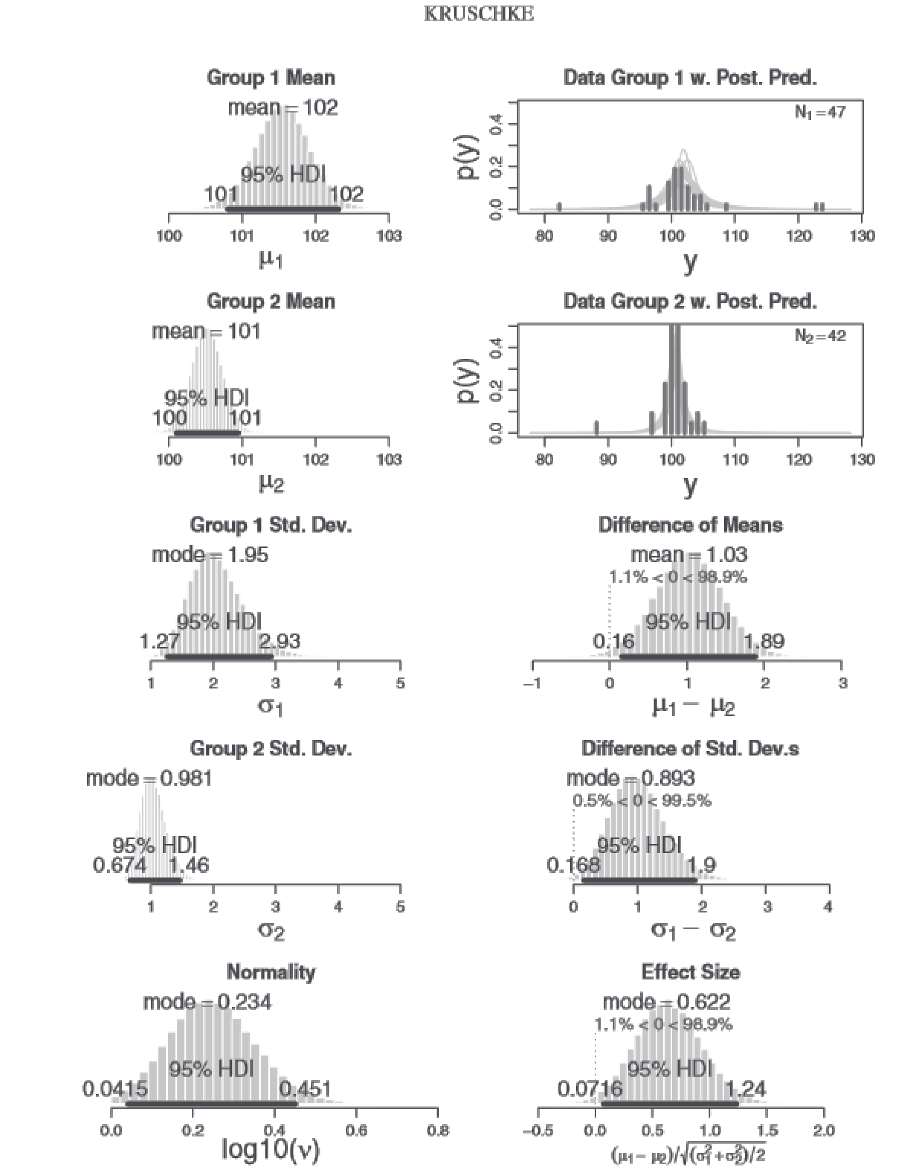
- No solo un estimador punctual

Tenemos precision explícita de los parámetros de interés

- No afectadas por las intenciones de muestreo y de prueba

Modelo flexible que es robusto a valores extremos

- No hay supuesto de normalidad



Evaluación de hipótesis nula en Bayes

¿Qué hacer si quieres evaluar la probabilidad de un valor nulo?

Dos caminos:

1. Estimación vía parámetros: Región de Equivalencia Práctica (ROPE) y el Intervalo de más alta densidad (HDI)

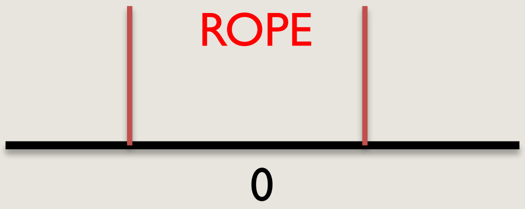
2. Comparación bayesiana de modelos: Decisión basada en factores bayesianos (Bayes factors)

1. Estimación de parámetros

Usa la distribución posterior para discernir los valores creíbles del parámetro.

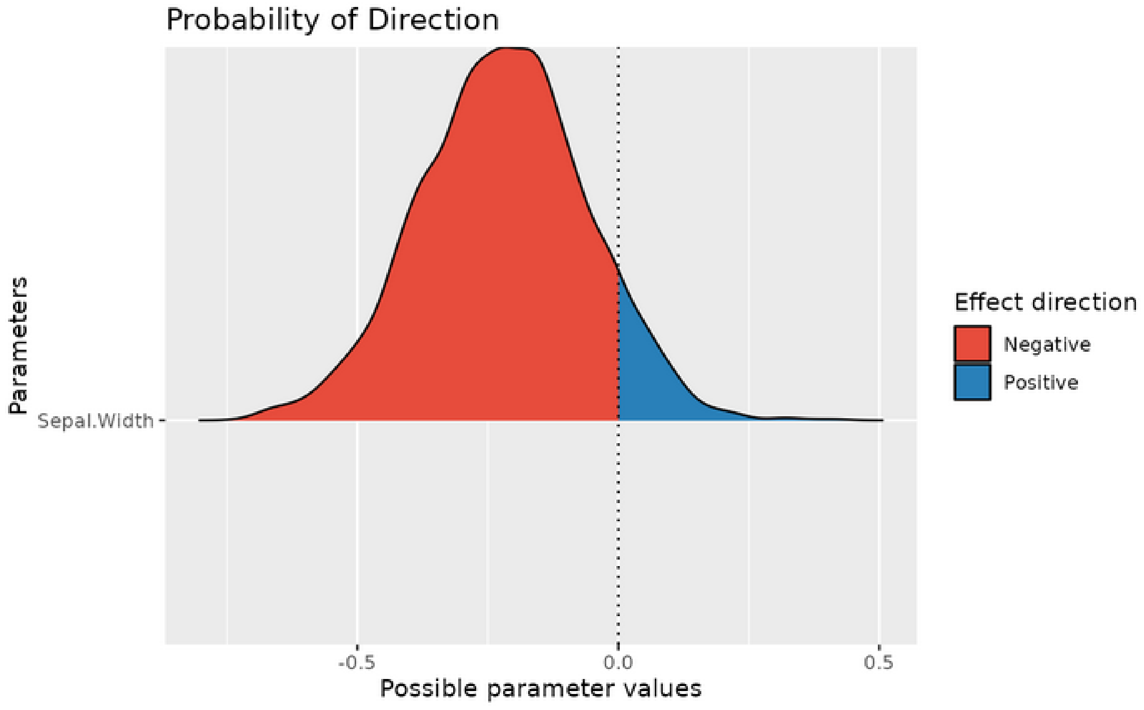
Considere un valor de referencia a evaluar del parámetro de interés: 0, -.1 - .1, 100 – 200.

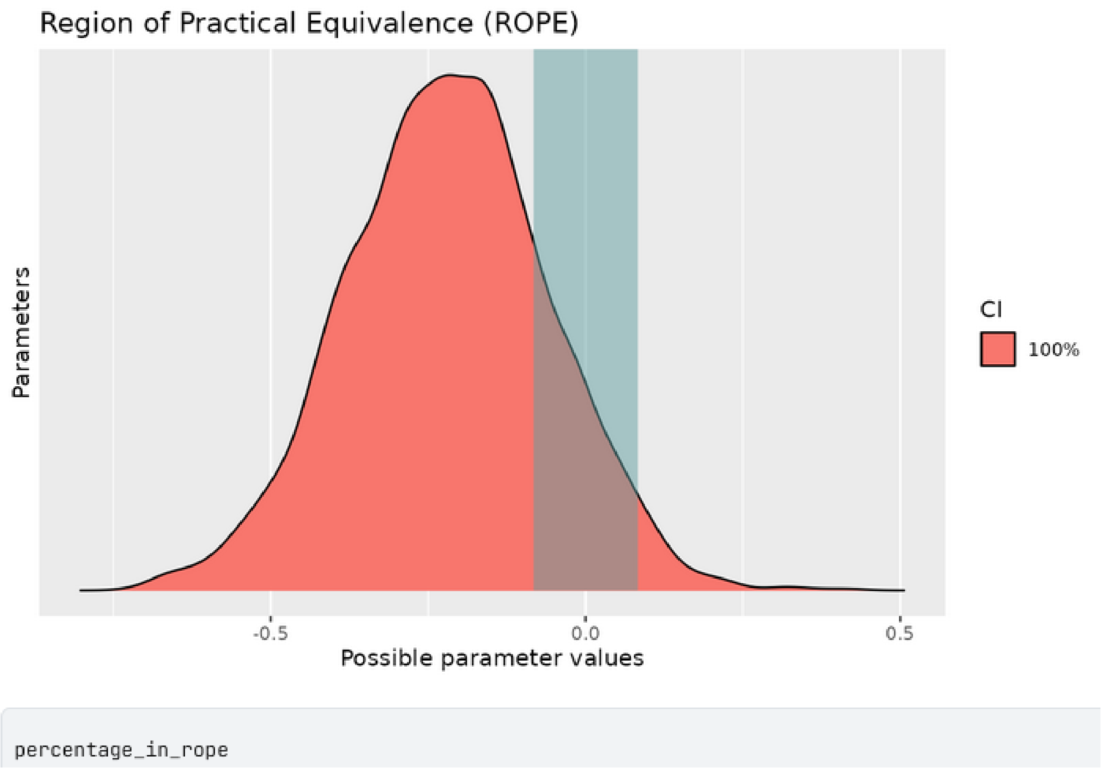
- Región de equivalencia en la práctica (ROPE): rango (del parámetro) considerado como el mismo en términos prácticos.



- Pero, ¿cómo se determina la ROPE?
  - Estadística vs Teóricamente significativo

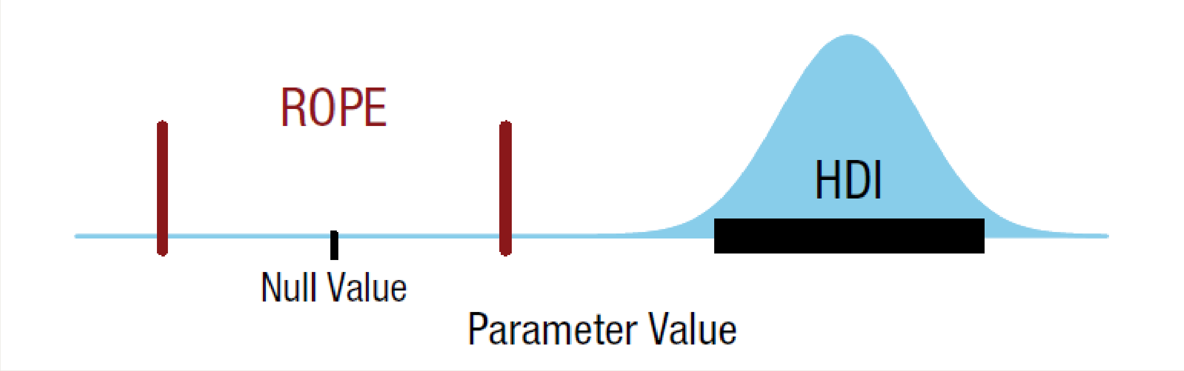
Region of practical equivalence (ROPE)





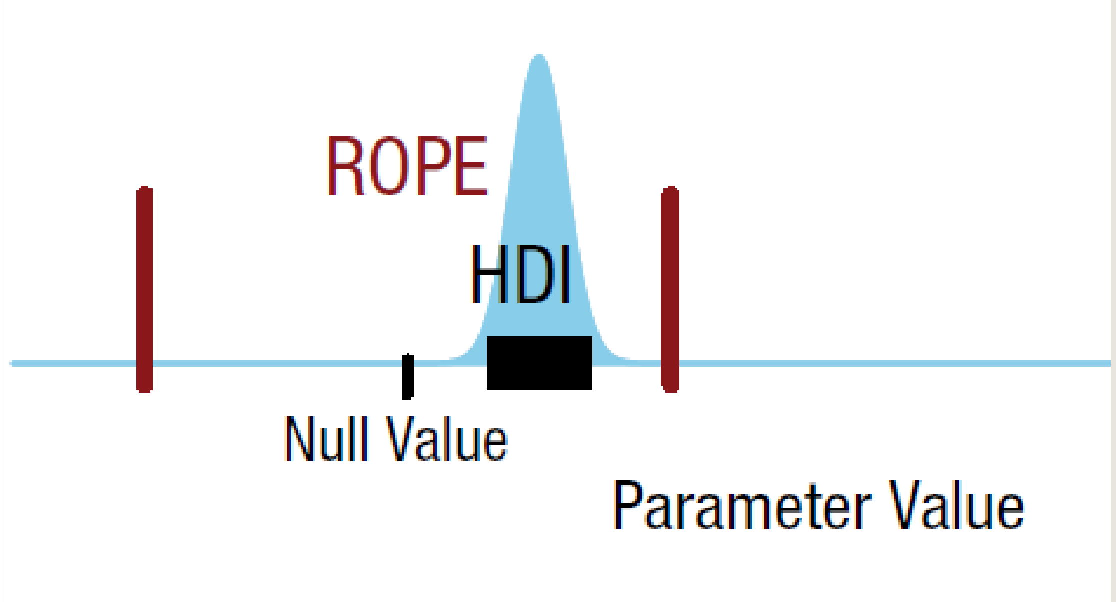
Estimación de parámetros

- Regla de decisión
  - El valor de un parámetro es declarado como **no creíble**, o rechazado, si su ROPE yace fuera del HDI relevante (típicamente el de 95 %) de su distribución posterior.



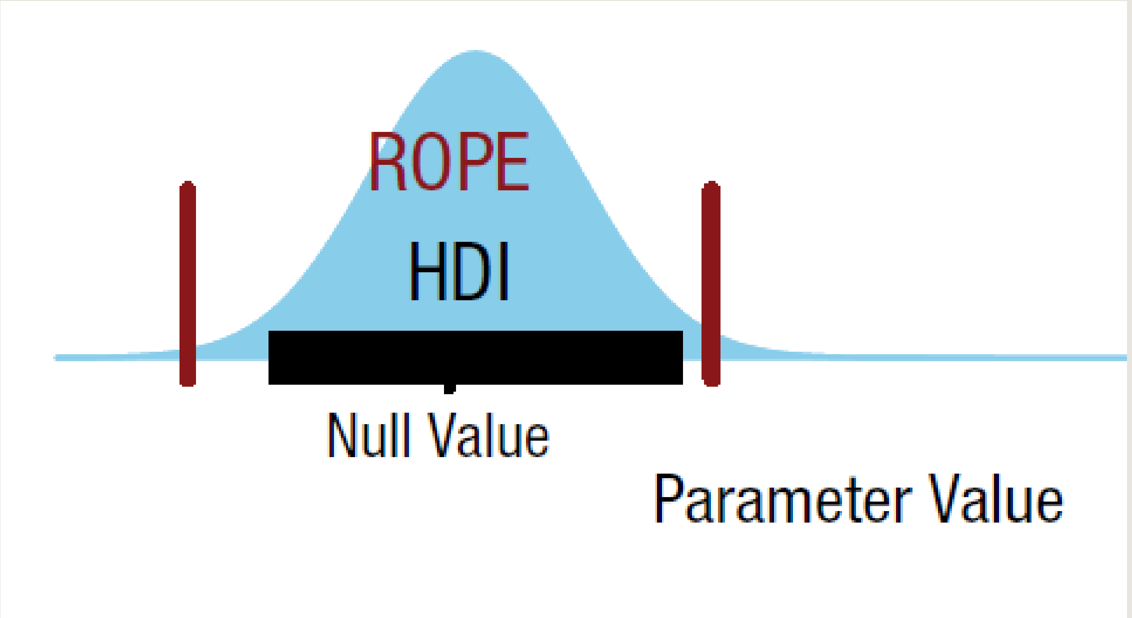
Estimación de parámetros

- Regla de decisión
  - El valor de un parámetro es declarado **aceptado** si su ROPE contiene completamente el HDI relevante (típicamente el de 95 %) de su distribución posterior.



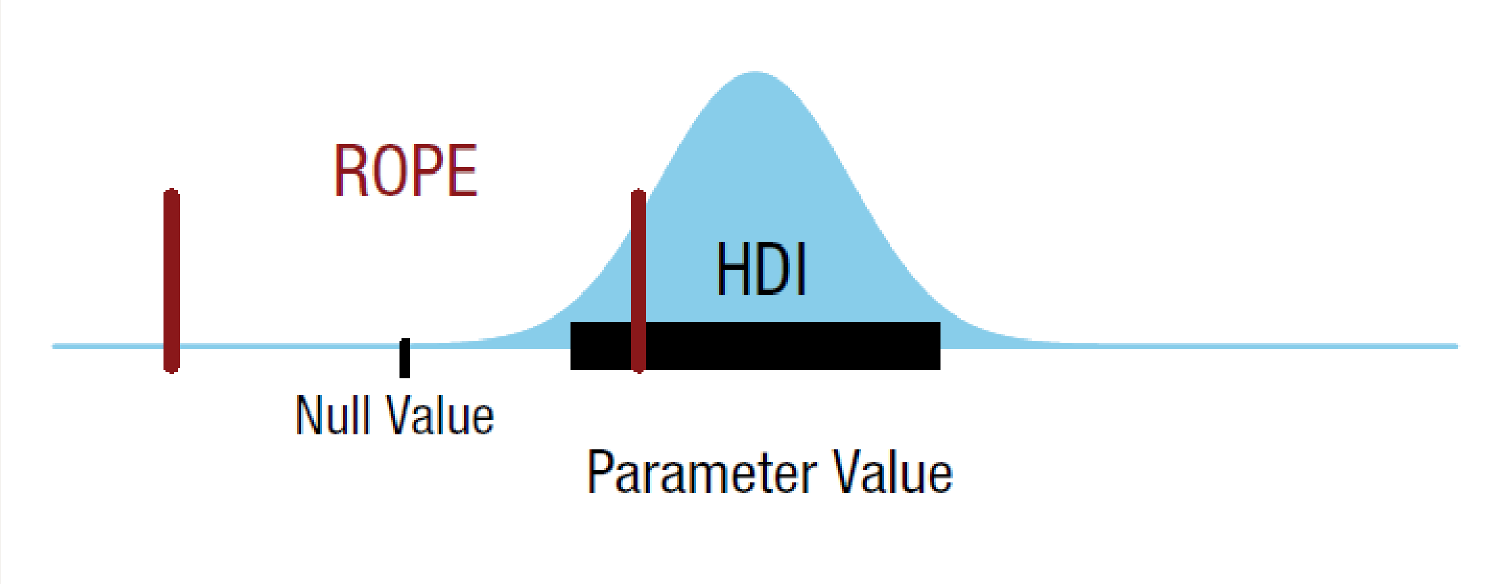
Estimación de parámetros

- Regla de decisión
  - El valor de un parámetro es declarado **aceptado** si su ROPE contiene completamente el HDI relevante (típicamente el de 95 %) de su distribución posterior.



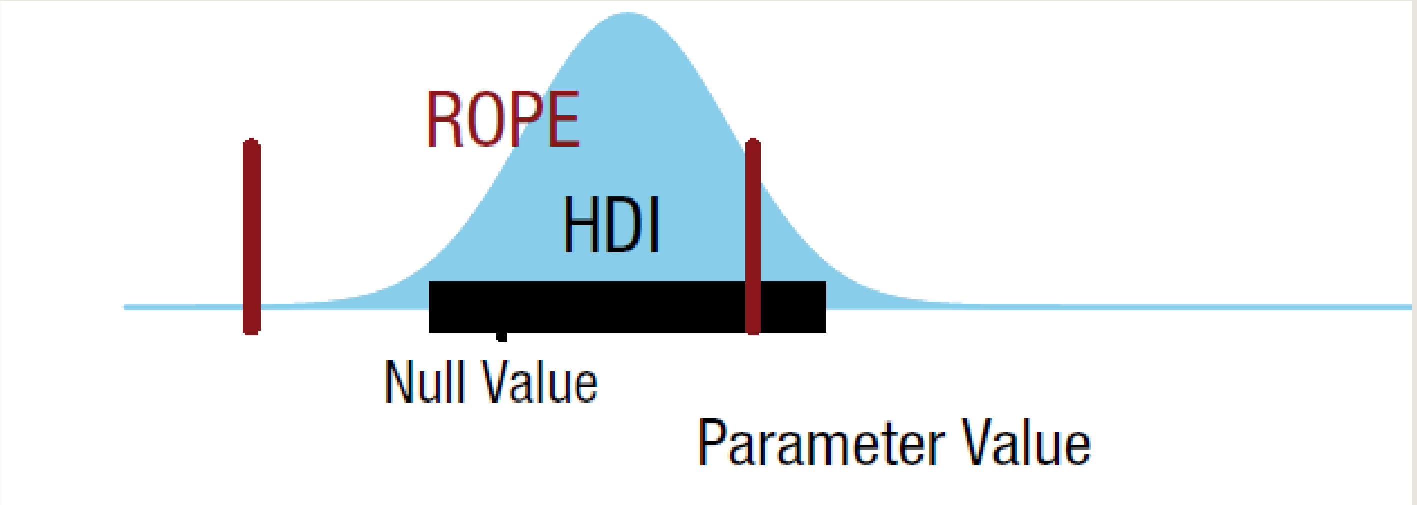
Estimación de parámetros

- Regla de decisión
  - ¿Qué hacer si el HDI y la ROPE traslapan, sin que la ROPE cubra la HDI por completo (ninguna de las condiciones anteriores se satisface?



Estimación de parámetros

- Regla de decisión
  - ¿Qué hacer si el HDI y la ROPE traslapan, sin que la ROPE cubra la HDI por completo (ninguna de las condiciones anteriores se satisface?



No siempre el 95% es simétrico

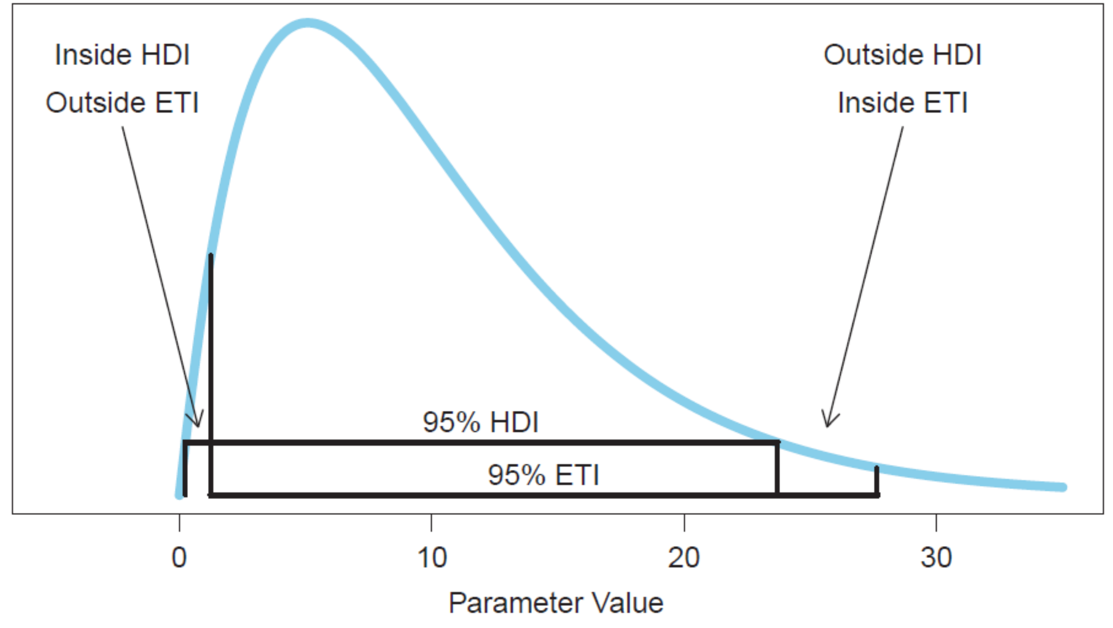
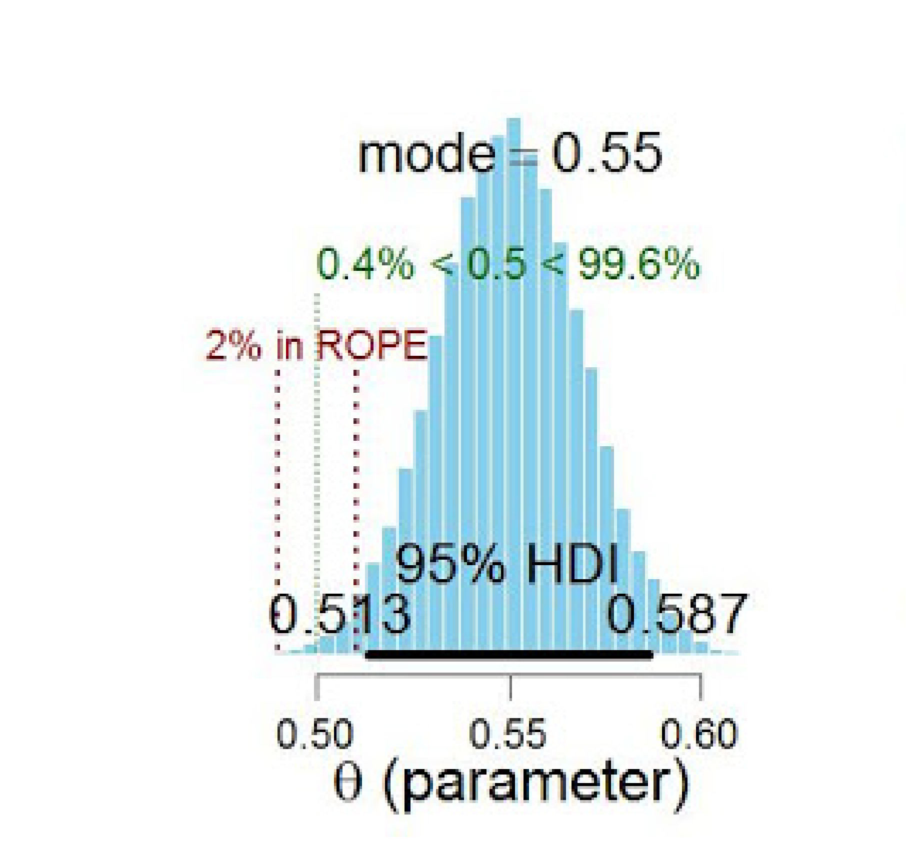


Figure 12.2: A skewed distribution has different 95% highest density interval (HDI) than 95% equal-tailed interval (ETI). Copyright © Kruschke, J. K. (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan. 2nd Edition.* Academic Press / Elsevier.

- Esto es importante para las desviaciones estándar
- Pero también para cierto tipo de parámetros en modelos espaciales o en la estimación de valores atípicos

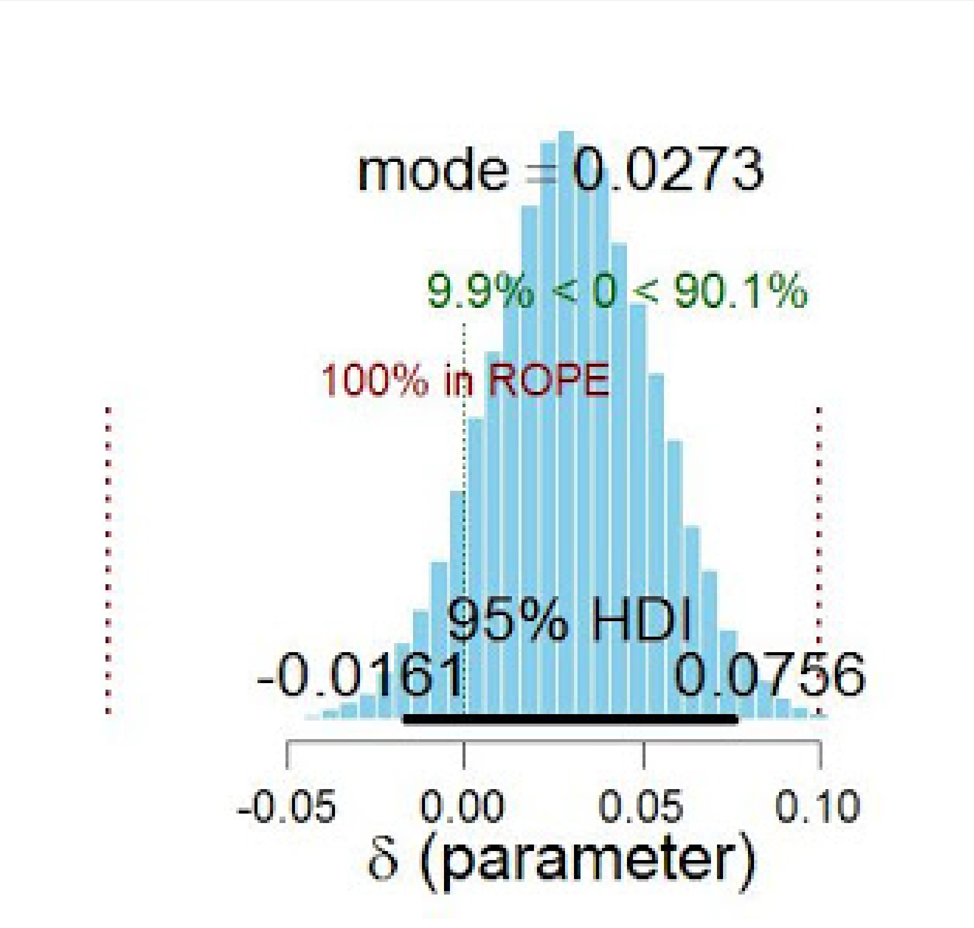
Estimación de parámetros

- Regla de decisión
  - El valor de un parámetro es declarado como **no creíble**, o rechazado, si su ROPE yace fuera del HDI relevante (típicamente el de 95 %) de su distribución posterior.



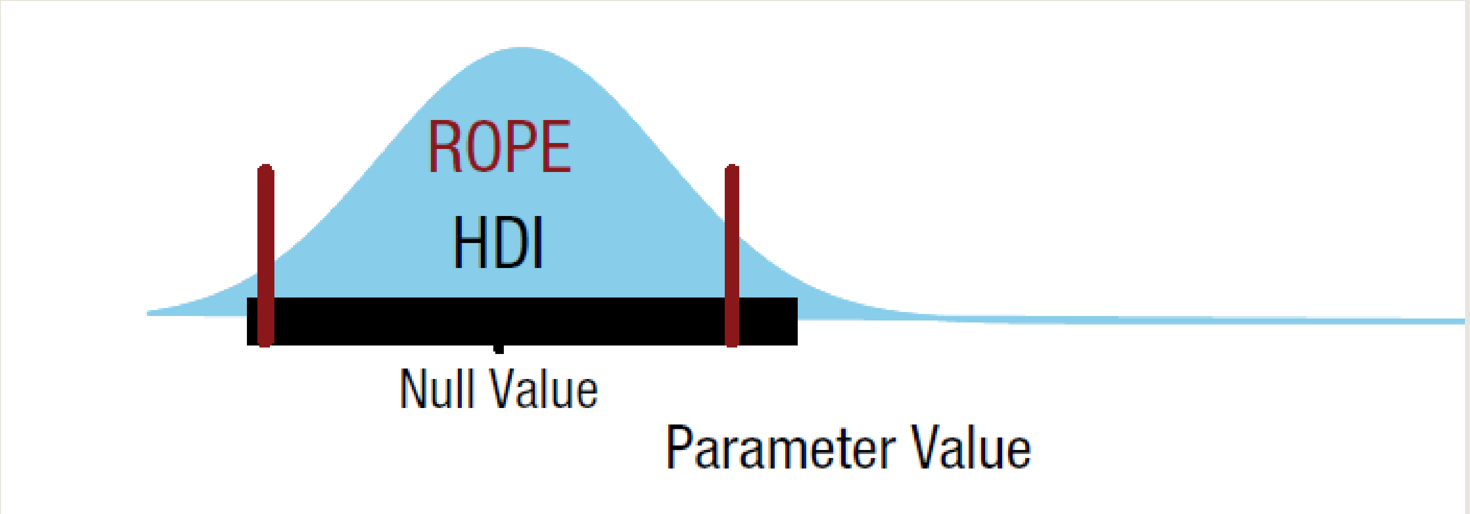
Estimación de parámetros

- Regla de decisión
  - El valor de un parámetro es declarado **aceptado** si su ROPE contiene completamente el HDI relevante (típicamente el de 95 %) de su distribución posterior.



Estimación de parámetros

- Regla de decisión
  - ¿Qué hacer si el HDI y la ROPE traslapan, sin que la ROPE cubra la HDI por completo (ninguna de las condiciones anteriores se satisface?



Estimación de parámetros

- No olvidar la inestabilidad de las fronteras de las HDI debido a la aleatoriedad del MCMC
- Es importante advertir que la regla de decisión es independiente de la inferencia bayesiana (la parte bayesiana se acaba en la estimación de la distribución posterior)

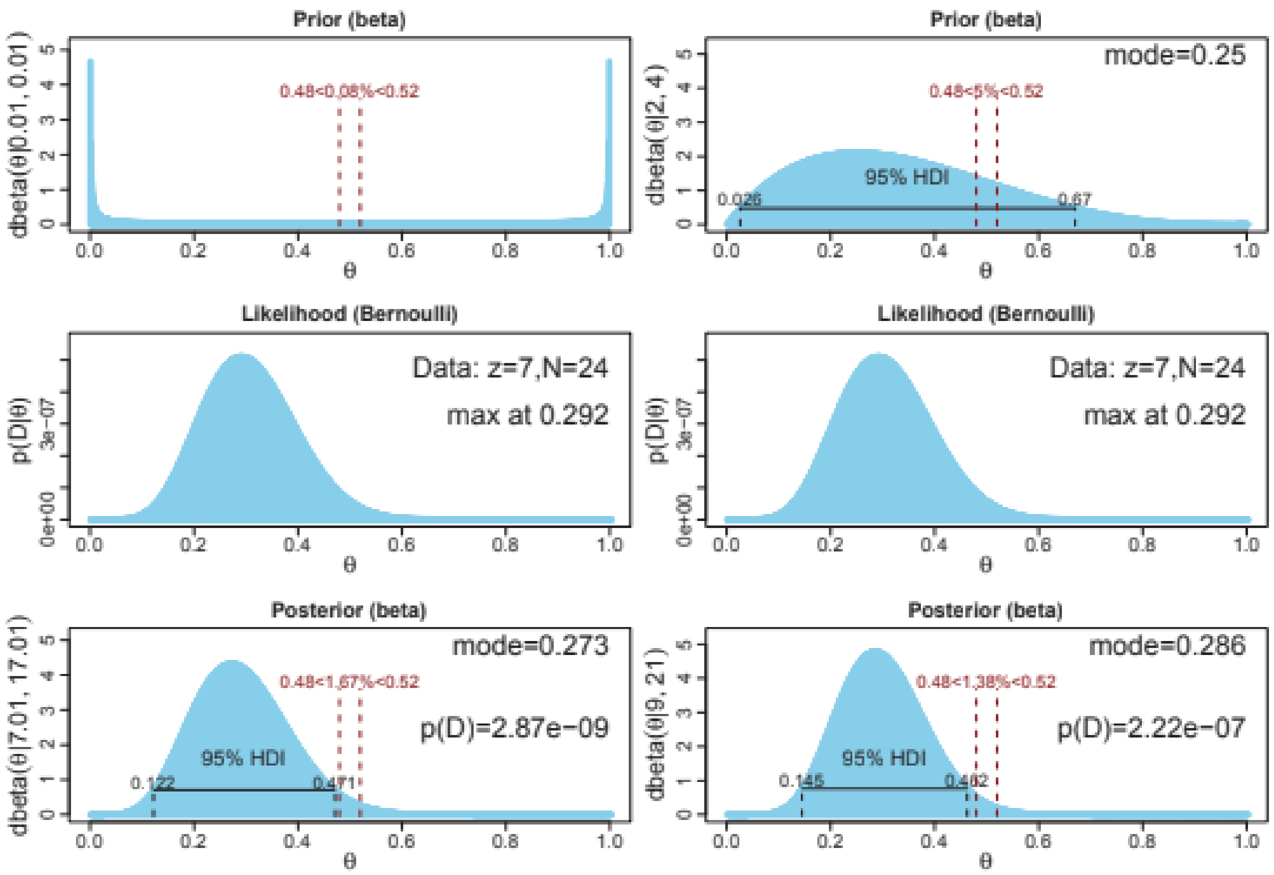
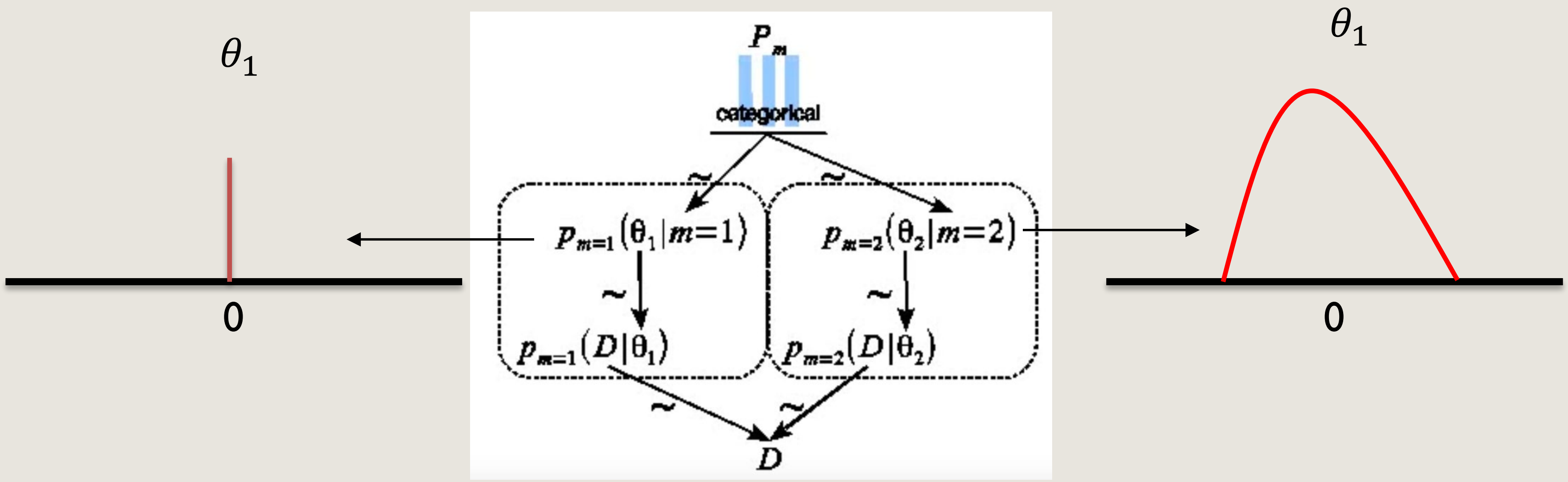


Figure 12.3: Left column: Haldane prior. Right column: Mildly informed prior. Vertical dashed lines mark a ROPE from 0.48 to 0.52. Annotation above the dashed lines indicates the percentage of the distribution within the ROPE. Copyright © Kruschke, J. K. (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan. 2nd Edition.* Academic Press / Elsevier.

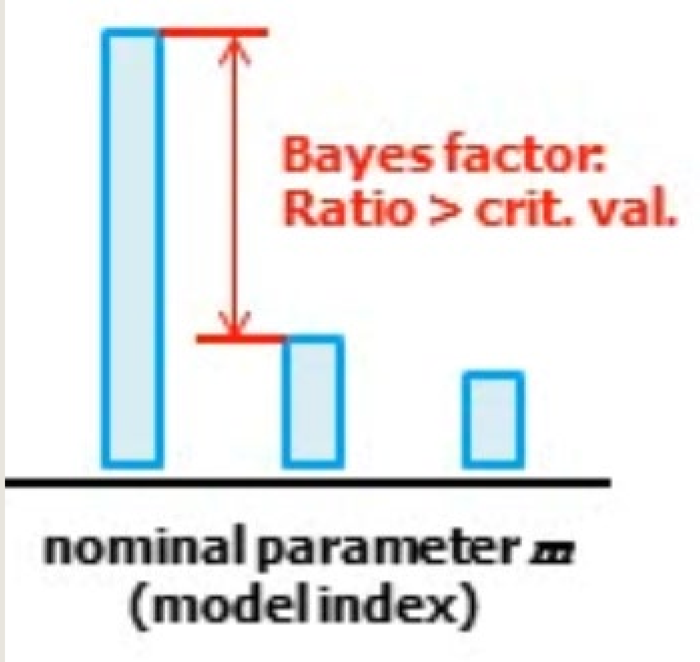
2. Estimación de la nula vía modelos

- Planteamos la pregunta:
  - Estimación puntual del valor nulo v distribución de posibilidades



Que debemos de pensar de estos modelos

- *Marginal likelihood*
- $P(D|\theta_1)p(\theta_1|m = 1)d\ \theta_1$ vs $P(D|\theta_2)p(\theta_2|m = 2)d\ \theta_2$
- Si tomamos la razón de las marginales: Bayes factor

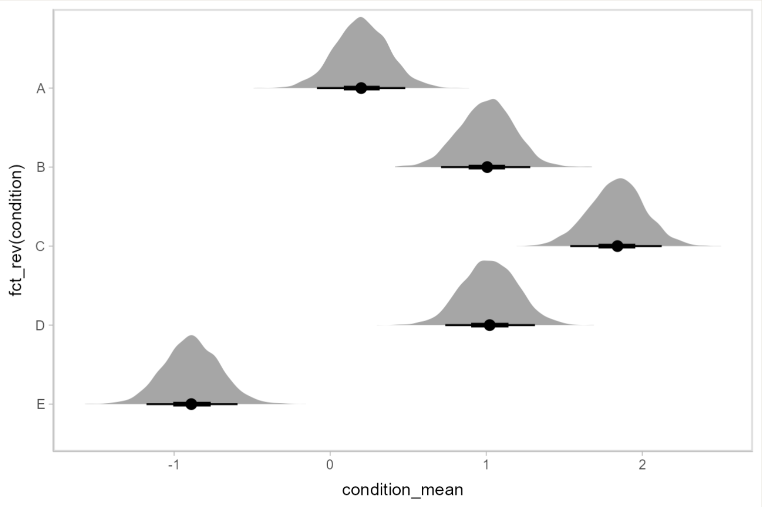


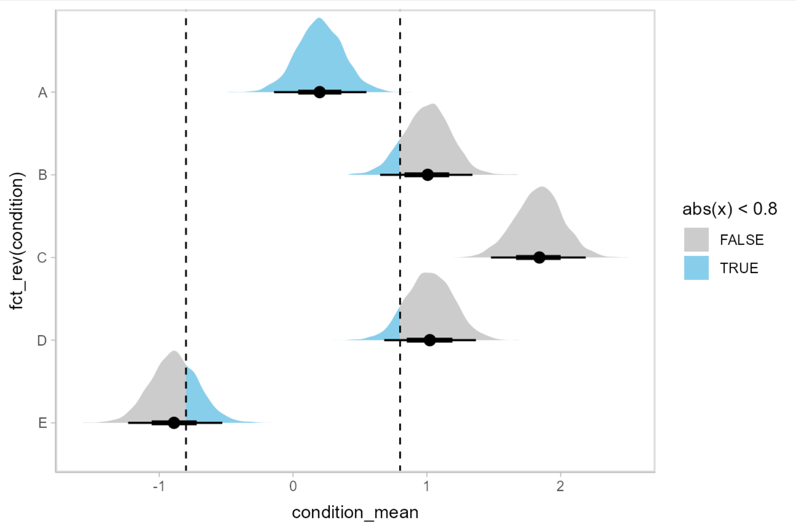
Clase de hoy: Inferencia con las posteriores

```
. regress bwght cigs male parity faminc faminc2

      Source |       SS           df       MS      Number of obs   =     1,387
-------------+----------------------------------   F(5, 1381)      =     12.54
       Model |  20057634.7         5  4011526.94   Prob > F        =    0.0000
    Residual |   441644355     1,381  319800.402   R-squared       =    0.0434
-------------+----------------------------------   Adj R-squared   =    0.0400
       Total |   461701990     1,386  333118.319   Root MSE        =    565.51

------------------------------------------------------------------------------
       bwght |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        cigs |  -13.25496   2.588155    -5.12   0.000     -18.3321   -8.177817
        male |   91.00571   30.45073     2.99   0.003     31.27103    150.7404
      parity |   47.51602   17.07493     2.78   0.005     14.02041    81.01163
      faminc |   8.660125   3.013622     2.87   0.004     2.748353     14.5719
     faminc2 |  -.0826576   .0414853    -1.99   0.047    -.1640386   -.0012765
       _cons |   3115.172   57.80237    53.89   0.000     3001.782    3228.562
------------------------------------------------------------------------------
```





Consideraciones: Estimación de parámetros

- No olvidar la inestabilidad de las fronteras de las HDI debido a la aleatoriedad del MCMC
- Es importante advertir que la regla de decisión es independiente de la inferencia bayesiana (la parte bayesiana se acaba en la estimación de la distribución posterior)

Próxima clase