

Resultados de matemática

El problema y los datos

Los datos que vamos a utilizar provienen de un estudio longitudinal realizado en escuelas de Estados Unidos en 2002. Los datos constituyen una muestra de escuelas y dentro de cada escuela una muestra de estudiantes. Hay niños de ‘quinto año’ de 100 escuelas grandes (con al menos 400 estudiantes enrolados en quinto)

```
library(tidyverse); library(rstan)
source("http://www.stat.washington.edu/~hoff/Book/Data/data/chapter8.r")
d <- as_data_frame(Y.school.mathscore)
```

1. Dibuja el score contra la escuela. Usa `reorder()` (u otra función) para ordenar las escuelas según el score medio.
2. Dibuja un histograma de las medias por escuela, los desvíos por escuela.
3. Dibuja un scatter-plot de la media por escuela contra la cantidad de estudiantes seleccionados en cada escuela.

Modelo jerarquico

Consideramos el modelo

$$\begin{aligned} (y_{j,1}, \dots, y_{j,n_j}) &\stackrel{\text{ind}}{\sim} N(\mu_j, \sigma^2) \\ \mu_j &\stackrel{\text{ind}}{\sim} N(\mu, \tau^2) \\ \mu &\sim N(m, t_0) \quad \tau^2 \sim IG(a/2, at/2) \quad \sigma^2 \sim IG(b/2, bt/2) \end{aligned}$$

1. Es razonable este modelo ? Cual puede ser una debilidad ?
2. Cual es la ‘posterior completa’ de μ_j , $p(\mu_j | \mu_{-j}, \mu, \tau, \sigma, y)$?
3. Como se interpretan τ y σ en este problema ?
4. Que rol tienen los parámetros en $p(\tau^2)$ y $p(\sigma^2)$?

Resultados del modelo

```
dt.list <- with(d, list(
  N = nrow(d), S=max(school), y=mathscore, scl=school,
  m0 = 50, t0=5, a=1, t=100, b=1, s=100
))

### mod = stan_model(file = 'school.stan') ###
```

```
res = sampling(mod, data=dt.list, iter=1e3 )
```

El resultado (objeto `res`) esta en el archivo `resultado.Rdata`, utiliza `load()` para cargarlo.

1. Que comentarios podemos hacer sobre los resultados de la inferencia de los hyperparametros del modelo?

```
# Inferencia para hyper-parametros  
print(res, pars = c('mu', 'tau', 'sigma'))
```

El coeficiente de correlación intra-clase (ICC) se define como

$$ICC = \frac{\tau^2}{\sigma^2 + \tau^2}$$

mide la correlación del resultado de dos estudiantes en la misma escuela.

2. Dibuja un histograma para ICC y para los hyper-parámetros del modelo. Comienza con

```
extract(res, pars = c('mu', 'tau', 'sigma')) %>% as_data_frame()
```

3. Dibuja un scatter-plot de: \bar{y}_j vs $E(\mu_j|y)$, interpreta los resultados.