Introduction To Sampling

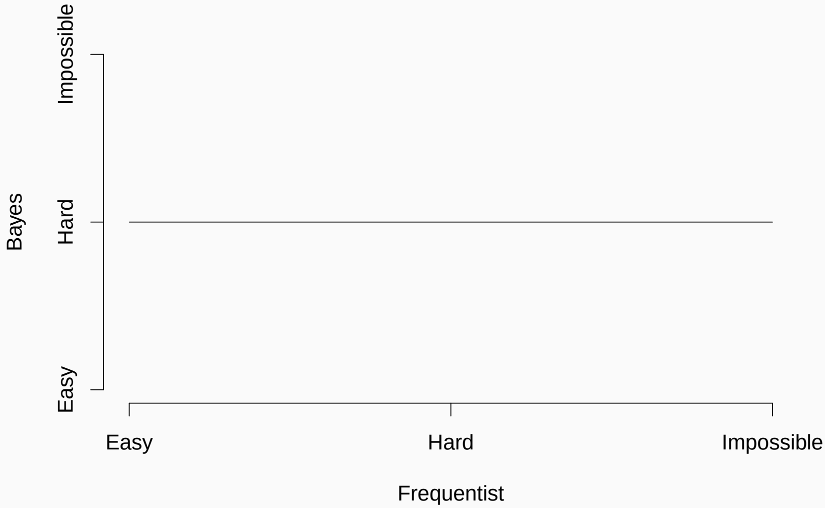
Jeff Rouder

May, 2022

Estimate the parameters in a normal

- I mean, how hard is this
- $\hat{\mu} = \frac{\sum y_i}{n}$
- $\hat{\sigma^2} = \frac{\sum (y_i - \bar{y})^2)}{n-1}$
- $\text{CI} = \pm t(n-1, .975) \times \sqrt{n} \times \frac{\hat{\mu}}{\hat{\sigma}}$

Bayesian. . .


Bayes
Easy
Hard
Impossible
Frequentist
Easy
Hard
Impossible

Sampling as Knowledge

Who Wakes Up First?

Let X denote when Becky wakes up. It is minutes after 7a.

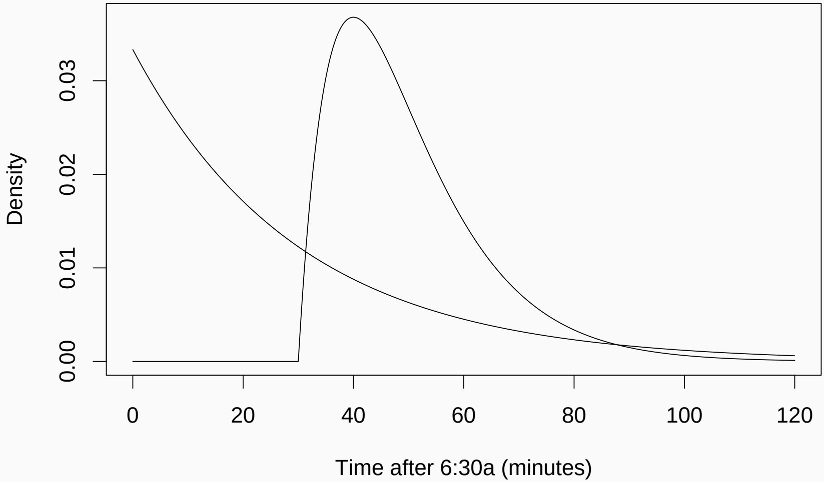
$$X \sim \text{Gamma}(2, 10)$$

Let Y denote when Jeff wakes up. It is minutes after 6:30a.

$$Y \sim \text{Gamma}(1, 30)$$

What is the probability Becky wakes up before Jeff, $Pr(X < Y)$?

Setup

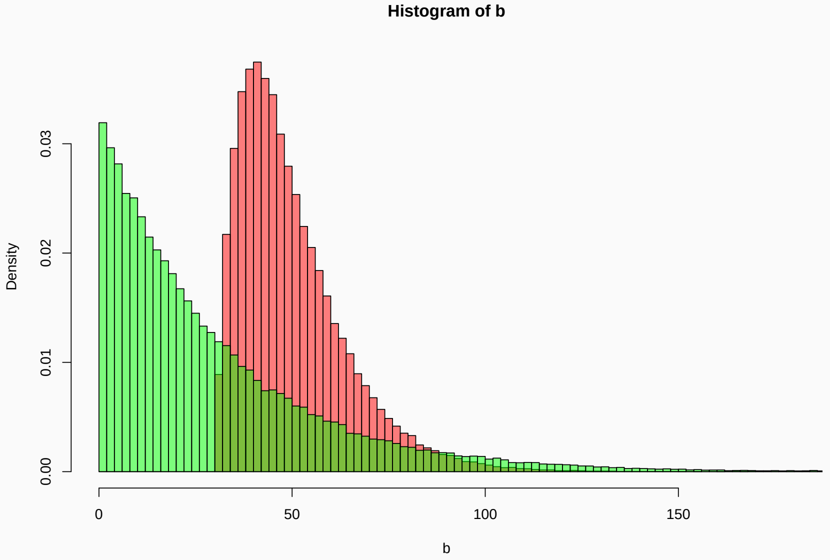


Density
0.00
0.01
0.02
0.03
0
20
40
60
80
100
120
Time after 6:30a (minutes)

Simulation Code

```
M=100000
j=rgamma(M,shape=1,scale=30)
b=rgamma(M,shape=2,scale=10)+30
hist(b,breaks=seq(0,500,2),col=rgb(1,0,0,.5),
     xlim=c(0,180),prob=T)
hist(j,add=T,col=rgb(0,1,0,.5),breaks=seq(0,500,2),prob=T)
print(mean(b<j))
```

Simulations




```
## [1] 0.2064
```

MCMC techniques

- Samples from posterior
- e.g. $(\mu, \sigma^2 | \boldsymbol{Y})$
- Samples are enough to learn

How To?

- Program Your Own Sampler To Learn About MCMC
    - fast to execute
    - you feel smart
    - insightful
- JAGS
    - fast but autocorrelation
    - great manual
    - stable
- stan
    - best samplers,
    - changing, info scattered
    - very popular (cool kids)

Model + Data

$$
\begin{aligned}
Y_i \sim & \mathrm{N}(\mu, \sigma^2) \\
\mu \sim & \mathrm{N}(a, b) \\
\sigma^2 \sim & \text{Inverse Gamma}(q, s)
\end{aligned}
$$

$Y =$ 92, 115, 100, 98

Demo, Roll My Own

$$\mu \mid \sigma^2, \boldsymbol{Y} \sim \text{Normal}(vc, v), \quad v = \left(\frac{N}{\sigma^2} + \frac{1}{b}\right)^{-1}, \quad c = \frac{N\bar{Y}}{\sigma^2} + \frac{a}{b}$$

$$\sigma^2 \mid \mu, \boldsymbol{Y} \sim \text{IG}\left(q + \frac{N}{2}, s + \frac{SSE}{2}\right)$$

]

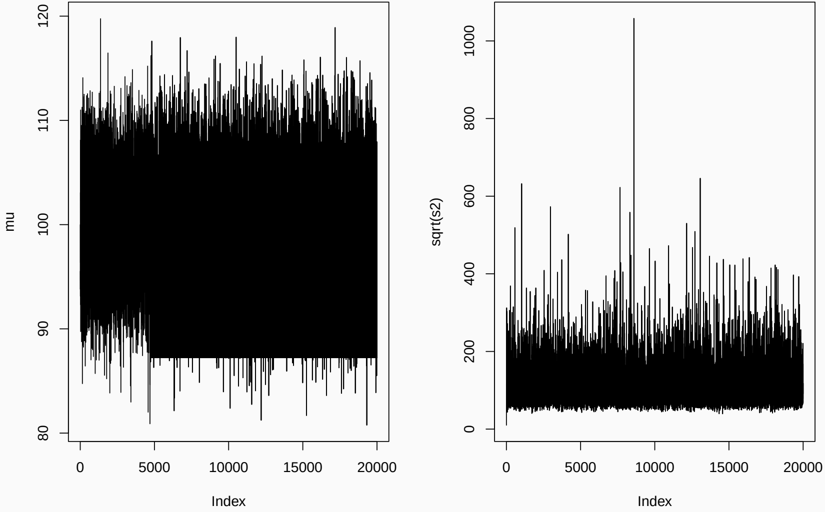
Demo, Roll My Own

```
Y=c(92,115,100,98)
ybar=mean(Y)
N=length(Y)
a=100
b=25
q=.5
s=.5*115^2
M=20000 # number of samples
mu=1:M
s2=1:M
mu[1]=100 #initial value
s2[1]=100 #initial value
```

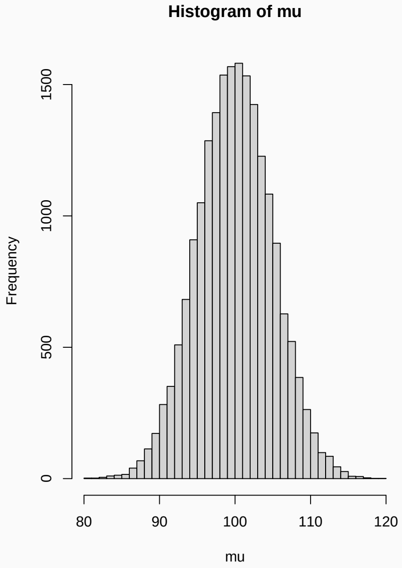
Demo, Roll My Own

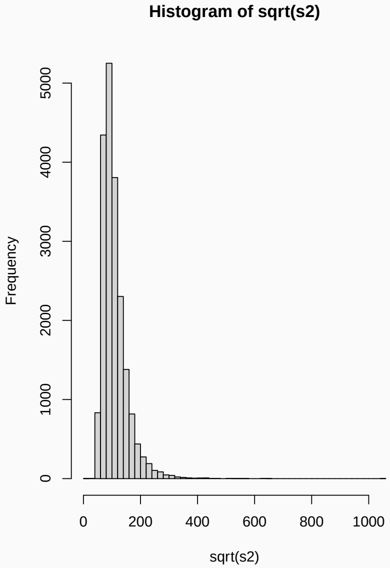
```
for (m in 2:M){
  v=1/(N/s2[m-1]+1/b)
  c=N*ybar/s2[m-1]+a/b
  mu[m]=rnorm(1,c*v,sqrt(v))
  SSE=sum((y-mu[m])^2)
  s2[m]=rinvgamma(1,N/2+q,SSE/2+s)
  }
```

Demo Roll My Own



Demo Roll My Own





JAGS

Consider This:

1. We wish to estimate how fast each of you are to react to a tone. When you hear a tone, hit a button. Pretty boring.
2. Let's assume each person has a *true score.* This is your speed in the limit of infinitely many trials.
3. If you know the ith person's true score, denoted μ_i, then RT on the jth trial, denoted Y_{ij}, is distributed as

$$Y_{ij}|\mu_i \sim \mathrm{N}(\mu_i, .15^2),$$

   where Y is measured in seconds.

4. Suppose we sample people from a population. Then it is reasonable to think each person's true score is also distributed

$$\mu_i \sim \mathrm{N}(.5, .1^2)$$

Consider This

- We wish to estimate μ_i.
- Some ppl recommend sample mean: (J trials)

$$\hat{\mu}_i = \bar{y}_i = \sum_{j=1}^{J} y_{ij}/J$$

- Other ppl recommend the median: Order values, pick middle one.

Jeff's Estimator

- I propose the following *regularized* estimator:
  - $\bar{y}_i$ is usual sample mean
  - s_i^2 is usual sample variance
  - $s^2 = \sum_{i=1}^{I} s_i^2 / I$
  - $\bar{y}$ is grand mean; mean of $\bar{y}_i$:
    - $\bar{y} = \sum_{i=1}^{I} \bar{y}_i / I = \sum_{ij} y_{ij} / IJ$
  - $s_{\bar{y}}^2$ is standard error squared; variance of $\bar{y}_i$
    - $s_{\bar{y}}^2 = \sum_i (\bar{y}_i - \bar{y})^2 / (I - 1)$

$$\hat{\mu}_i = vc,$$

where

$$v = \left(\frac{J}{s^2} + \frac{1}{s_{\bar{y}}^2} \right)^{-1}$$

$$c_i = \left(\frac{J\bar{y}_i}{s^2} + \frac{\bar{y}}{s_{\bar{y}}^2} \right)$$

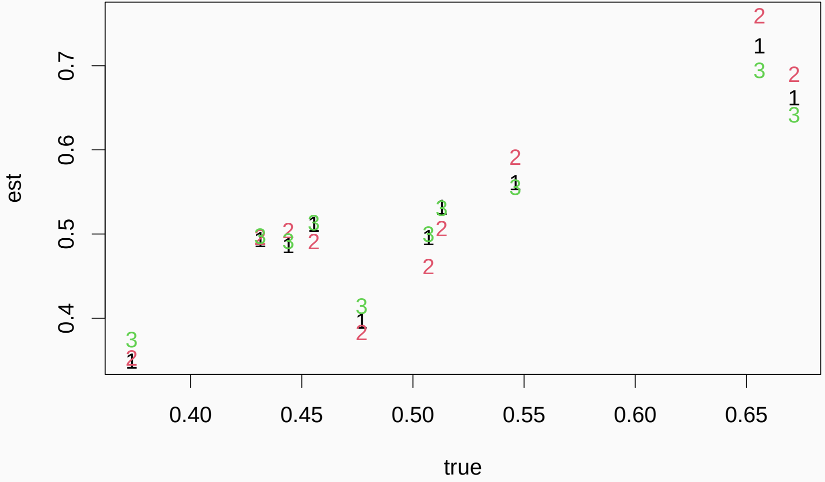
Jeff's Estimator

- Whacky
- Legitimate as it uses nothing but the data
- Which of the three is best

Defining Best

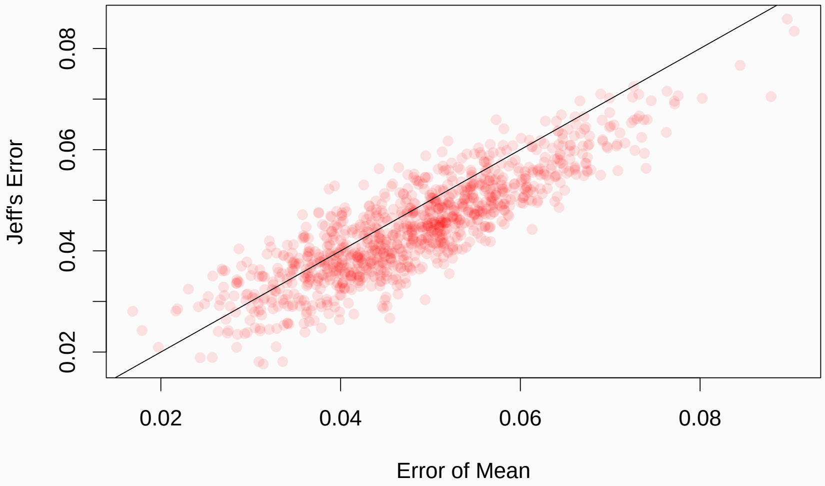
- Generate ground truth $(\mu_1, \ldots, \mu_I)$
- Generate data $\boldsymbol{Y}$
- Compute estimates $(\hat{\mu}_1, \ldots, \hat{\mu}_I)$
- Compute estimation error: $e_i = \hat{\mu}_i - \mu_i$
- Compute RMS error: $\text{rmse} = \sqrt{\sum_i e_i^2 / I}$

Example




```
##    Mean Median Jeff's
##   0.046  0.058  0.041
```

Do it 10,000 Times




```
## [1] "Proportion of Jeff's Wins: 0.738"
```

How Did I Do That

My estimator was based on the fact that the data were hierarchical

- People sampled from a population
- Trials nested in people
- Two sources of variation

Case 1

- Each person reads 10 nouns.
- How good of a reader is each person.

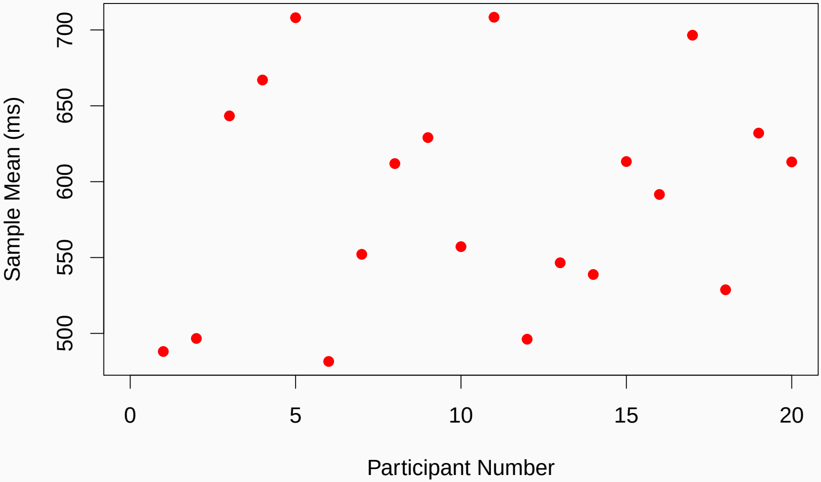
True Values

- Let μ_i denote the true value for the ith person

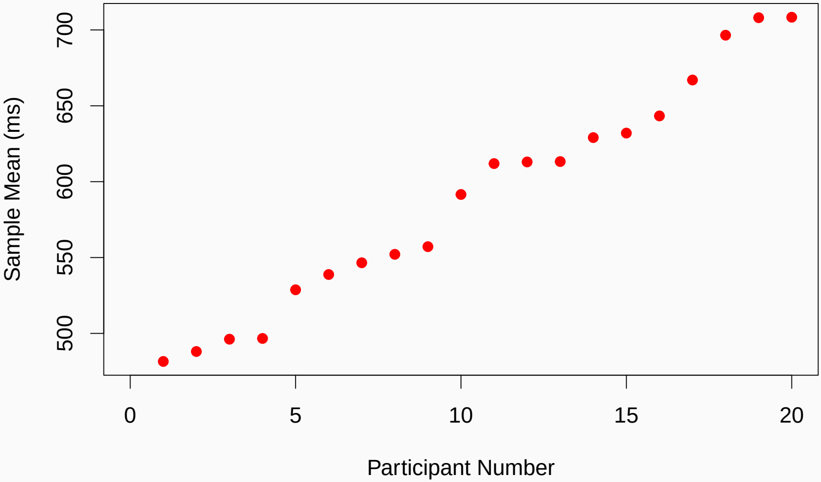
Data + Model

- Data, Y_{ij}, j denotes replicate for ith person
- $Y_{ij} \sim \mathrm{N}(\mu_i, \sigma^2)$

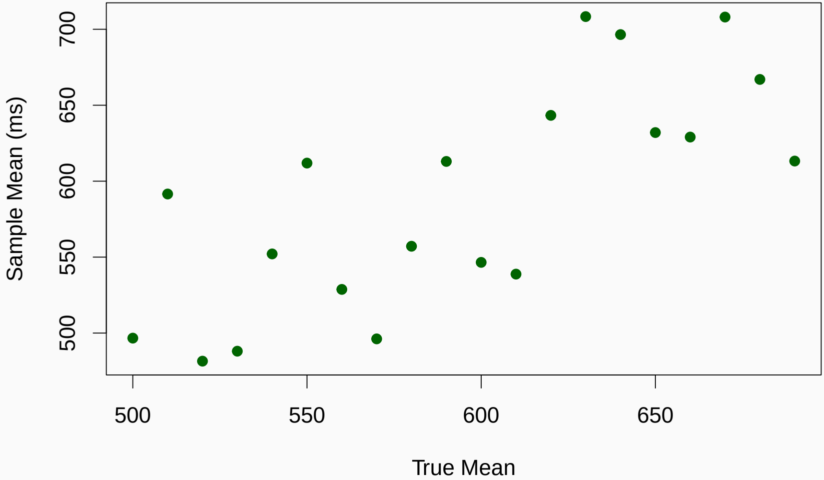
Mean score per person



Mean score per person (sorted)



Sample Mean vs. True Mean



Non-Hierarchical Bayes Model

$$Y_{ij} \sim \mathrm{N}(\mu_i, \sigma^2)$$
$$\mu_i \sim \mathrm{N}(a, b)$$
$$\sigma^2 \sim \mathrm{IG}(q, s)$$

- $a = 600$
- $b = 400^2$

Non-Hierarchical Bayes Model

Conditional Posteriors via Bayes Rule:

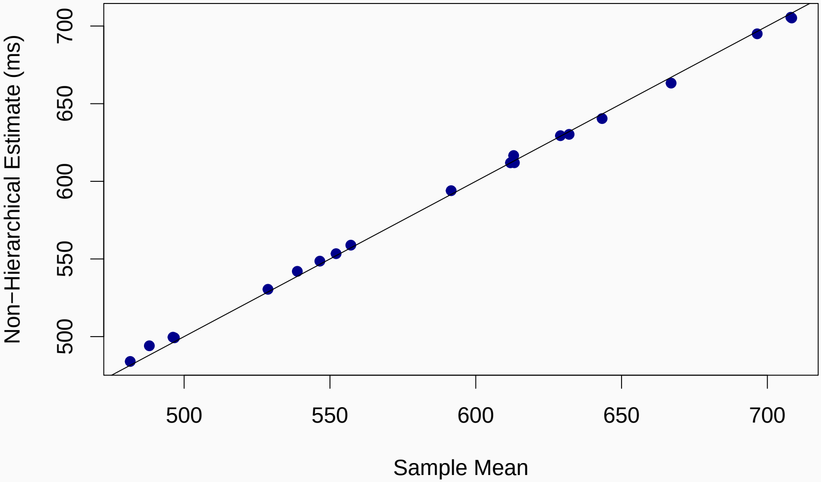
$$\mu_i \sim \mathrm{N}(c_i v, v)$$
$$v = \left(\frac{J}{\sigma^2} + \frac{1}{b}\right)$$
$$c_i = \left(\frac{J\bar{Y}_i}{\sigma^2} + \frac{a}{b}\right)$$

$$\sigma^2 \sim \mathrm{IG}(q + IJ/2, s + \mathrm{SSE}/2)$$

Plan of Chain

- Sample all μ's given σ^2
- Sample σ^2 given allμ's.
- Repeat

Results



Non−Hierarchical Estimate (ms)
500
550
600
650
700
Sample Mean
500
550
600
650
700

Hierarchical

These people can be treated as fixed or random.

- Fixed. I am interested in John Q. Doe as a unique individual. Whatever results I find cannot be generalized.
- Random. I sample John Q. Doe from a population and would like to generalize to the population as well.

We treated each person as fixed in the previous case.

Hierarchical

People as random effects

Hierarchical

$$
\begin{aligned}
Y_{ij} &\sim \mathrm{N}(\mu_i, \sigma^2) \\
\mu_i &\sim \mathrm{N}(\eta, \delta) \\
\sigma^2 &\sim \mathrm{IG}(q_1, s_1) \\
\eta &\sim \mathrm{N}(a, b) \\
\delta &\sim \mathrm{IG}(q_2, s_2)
\end{aligned}
$$

Hierarchical

- Stage 1: Sample people from a population model

$$\mu_i \sim \mathrm{N}(\eta, \delta)$$

- Stage 2: Sample observations from each person.

$$Y_{ij} \sim \mathrm{N}(\mu_i, \sigma^2)$$

Hierarchical

We all act as prior for each other.

Hierarchical

Conditional Posteriors

$$\mu_i \sim \mathrm{N}(c_i v, v)$$
$$v = \left(\frac{J}{\sigma^2} + \frac{1}{\delta}\right)$$
$$c_i = \left(\frac{J\bar{Y}_i}{\sigma^2} + \frac{\eta}{\delta}\right)$$

$$\sigma^2 \sim \mathrm{IG}(q_1 + IJ/2, s_1 + \sum_{ij}(Y_{ij} - \mu_i)^2/2)$$

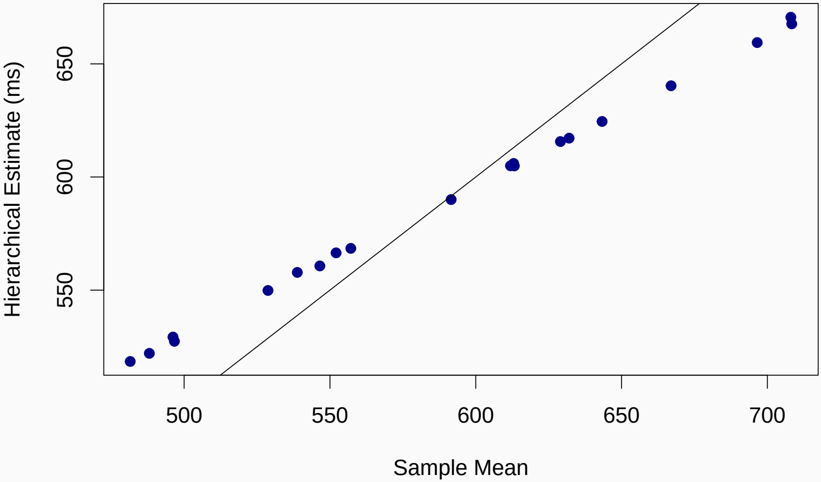
Hierarchical

Conditional Posteriors

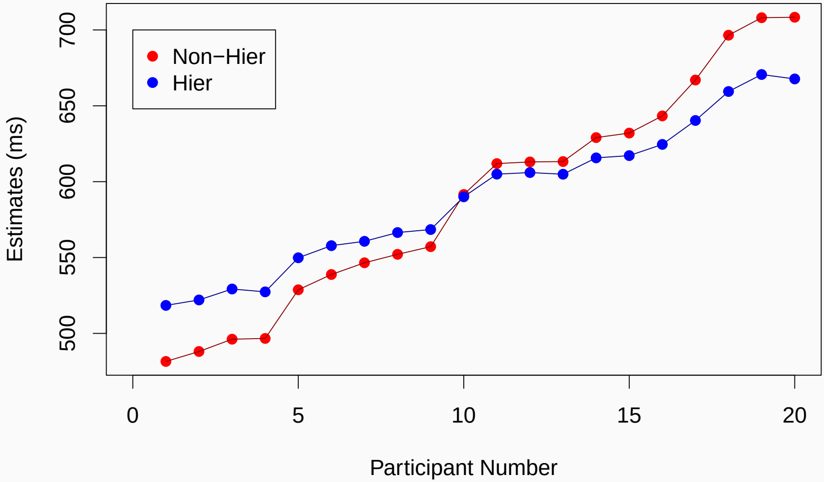
$$\eta \sim \mathrm{N}(c_0 v_0, v_0)$$
$$v_0 = \left(\frac{I}{\delta} + \frac{1}{b}\right)$$
$$c_0 = \left(\frac{I\bar{\mu}}{\delta} + \frac{a}{b}\right)$$

$$\delta \sim \mathrm{IG}(q_2 + I/2, s_1 + \sum_i (\mu_i - \eta)^2/2)$$

Hierarchical Results



Sorted by People



Versus True Values