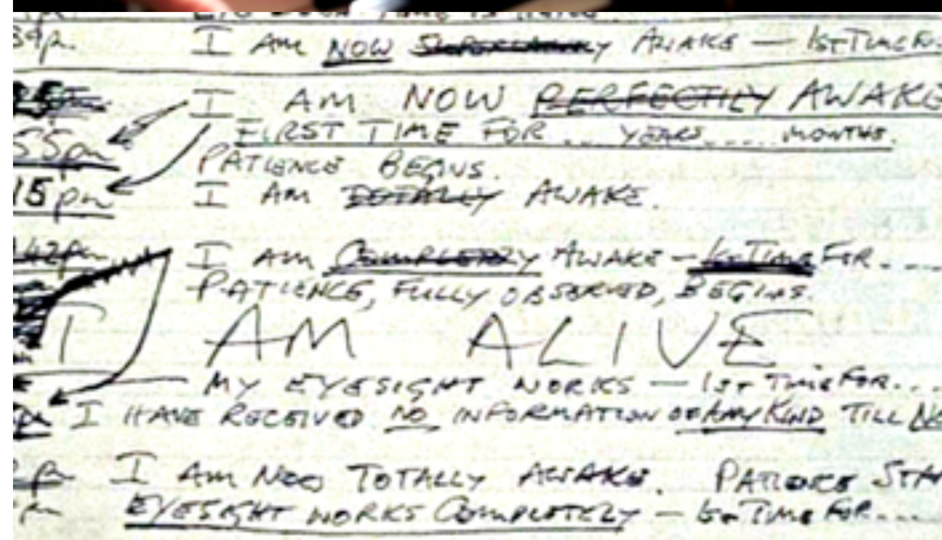
Statistical Rethinking

Week 8: Multilevel Models

Richard McElreath

Anterograde amnesia

- Musicologist and conductor Clive Wearing
- Lost parts of prefrontal and hippocampus
- Can still play piano
- Can't remember what happened 1 min ago

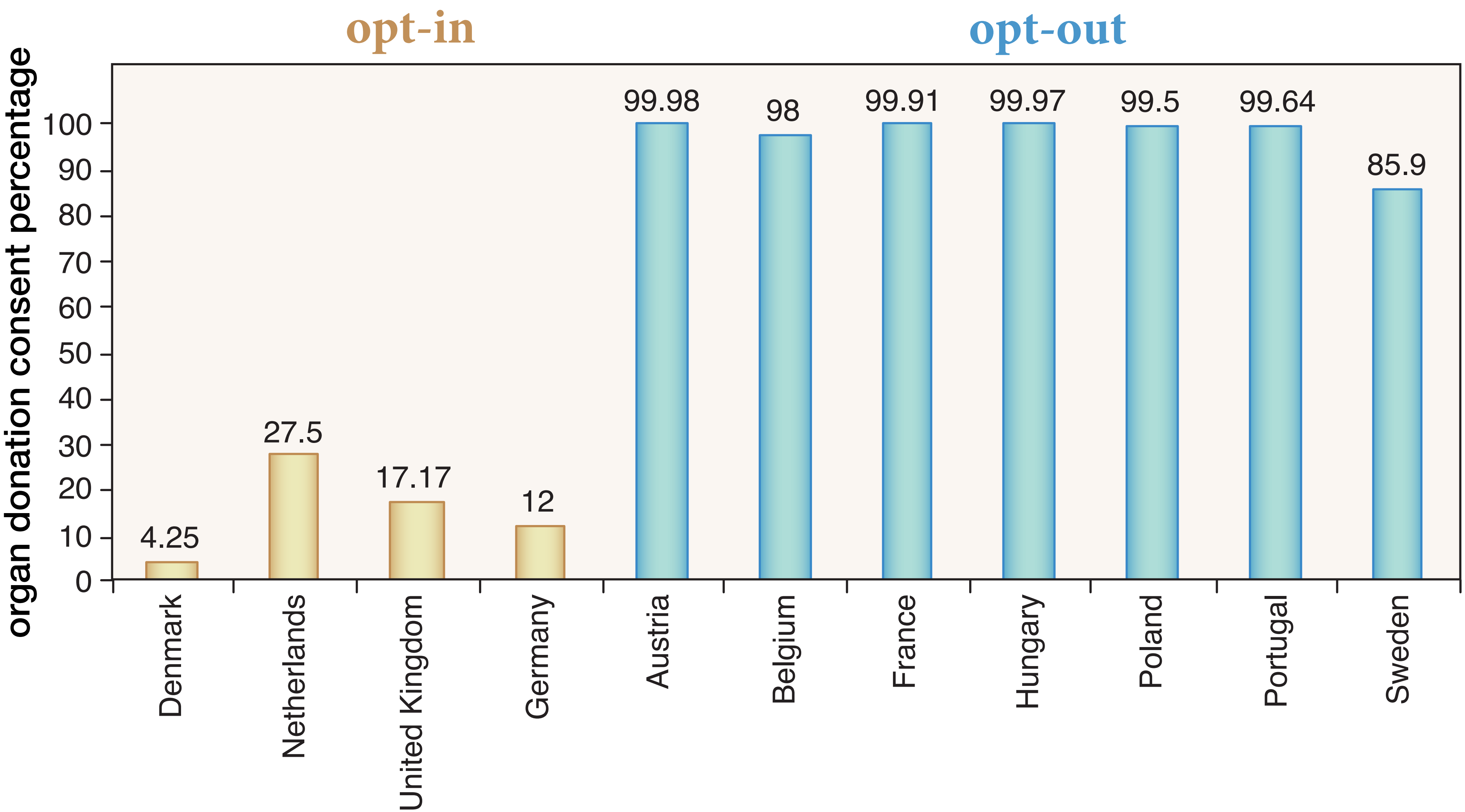


Anterograde amnesia

- "Fixed effects" models have anterograde amnesia
  - Every new cluster (individual, pond, road, classroom) is a new world
  - No information passed among clusters
- Multilevel models remember and pool information
  - Properties of clusters come from a "population"
  - Inferred population defines pooling
  - If previous clusters improve your guess about a new cluster, you want to use pooling

Learning, forward and back


Depends upon variation


opt-in
opt-out
organ donation consent percentage
0
10
20
30
40
50
60
70
80
90
100
4.25
27.5
17.17
12
99.98
98
99.91
99.97
99.5
99.64
85.9
Denmark
Netherlands
United Kingdom
Germany
Austria
Belgium
France
Hungary
Poland
Portugal
Sweden

Multilevel should be default

- Defaults are powerful things
- Single-level regression is default
  - People justify multilevel models
- This is backwards
  - Multilevel estimates usually better
  - Should have to justify not using multilevel model

Goals

- Introduce multilevel models
- How *shrinkage* and *pooling* work
- Why they produce better estimates
- How to fit with `map2stan`
- Methods of plotting and comparing
- Advanced: Continuous categories and Gaussian process regression

Multilevel models

- Usual use is to model clustering
  - Classrooms within schools
  - Students within classrooms
  - Grades within students
  - Questions within exams
- Repeat measures of units
- Imbalance in sampling
- "pseudoreplication"

Multilevel models

- Examples from earlier:
  - !Kung individuals in families
  - Species in clades
  - Nations in continents
  - Applicants in departments

Example: Tadpole predation

```
library(rethinking)
data(reedfrogs)
d <- reedfrogs
```

- Numbers of surviving tadpoles
- Different densities/sizes
- With and without predators
- We'll focus on variation across tanks

Tadpole models

- Structure:
  - Tadpoles in tanks, different densities
  - Outcome: number surviving
- Can fit two basic models:
  1. Dummy variable for each tank
  2. Multilevel model with ***varying intercepts*** by tank

Regularized intercepts

number surviving, tank i

$$
\begin{aligned}
s_i &\sim \text{Binomial}(n_i, p_i) \\
\text{logit}(p_i) &= \alpha_{\text{TANK}[i]} \\
\alpha_{\text{TANK}} &\sim \text{Normal}(0, 5)
\end{aligned}
$$

regularizing prior

Regularized intercepts

$$s_i \sim \text{Binomial}(n_i, p_i)$$
$$\text{logit}(p_i) = \alpha_{\text{TANK}[i]}$$
$$\alpha_{\text{TANK}} \sim \text{Normal}(0, 5)$$

R code 12.2

```
library(rethinking)
data(reedfrogs)
d <- reedfrogs

# make the tank cluster variable
d$tank <- 1:nrow(d)

# fit
m12.1 <- map2stan(
    alist(
        surv ~ dbinom( density , p ) ,
        logit(p) <- a_tank[tank] ,
        a_tank[tank] ~ dnorm( 0 , 5 )
    ),
    data=d )
```

Adaptive regularization

$$
\begin{aligned}
s_i &\sim \mathrm{Binomial}(n_i, p_i) \\
\mathrm{logit}(p_i) &= \alpha_{\textsc{tank}[i]} \\
\alpha_{\textsc{tank}} &\sim \mathrm{Normal}(\alpha, \sigma) \\
\alpha &\sim \mathrm{Normal}(0, 1) \\
\sigma &\sim \mathrm{HalfCauchy}(0, 1)
\end{aligned}
$$

Adaptive regularization

$$
\begin{aligned}
s_i &\sim \text{Binomial}(n_i, p_i) \\
\text{logit}(p_i) &= \alpha_{\text{TANK}[i]} \\
\textit{varying intercepts} \longrightarrow \alpha_{\text{TANK}} &\sim \text{Normal}(\alpha, \sigma) \\
\alpha &\sim \text{Normal}(0, 1) \\
\sigma &\sim \text{HalfCauchy}(0, 1)
\end{aligned}
$$

Terminology

- ***Varying intercepts*** also called ***random intercepts***
- Neither of these terms makes much sense
    - "random"? Sometimes associated with research design, but design irrelevant
    - Ordinary dummy variables also "vary" across clusters
- Distinctive because individual intercepts learn from one another
    - *mnestic*: opposite of *amnestic*

Adaptive regularization

$$
\begin{aligned}
s_i &\sim \text{Binomial}(n_i, p_i) \\
\text{logit}(p_i) &= \alpha_{\text{TANK}[i]} \\
\alpha_{\text{TANK}} &\sim \text{Normal}(\alpha, \sigma)
\end{aligned}
$$

varying intercepts

mean

standard deviation

Adaptive regularization

$$
\begin{aligned}
s_i &\sim \text{Binomial}(n_i, p_i) \\
\text{logit}(p_i) &= \alpha_{\text{TANK}[i]} \\
\textit{varying intercepts} \longrightarrow \alpha_{\text{TANK}} &\sim \text{Normal}(\alpha, \sigma) \\
\alpha &\sim \text{Normal}(0, 1) \\
\sigma &\sim \text{HalfCauchy}(0, 1)
\end{aligned}
$$

Adaptive regularization

$$
\begin{aligned}
s_i &\sim \text{Binomial}(n_i, p_i) \\
\text{logit}(p_i) &= \alpha_{\text{TANK}[i]} \\
\textit{varying intercepts} \longrightarrow \alpha_{\text{TANK}} &\sim \text{Normal}(\alpha, \sigma) \\
\alpha &\sim \text{Normal}(0, 1) \\
\sigma &\sim \text{HalfCauchy}(0, 1)
\end{aligned}
$$

Survival across tanks has a *distribution*.
This *distribution* is the prior for each tank.
Distribution needs its own prior.

$$
\begin{aligned}
s_i &\sim \text{Binomial}(n_i, p_i) \\
\text{logit}(p_i) &= \alpha_{\text{TANK}[i]} \\
\alpha_{\text{TANK}} &\sim \text{Normal}(\alpha, \sigma) \\
\alpha &\sim \text{Normal}(0, 1) \\
\sigma &\sim \text{HalfCauchy}(0, 1)
\end{aligned}
$$

R code 12.3

```
m12.2 <- map2stan(
    alist(
        surv ~ dbinom( density , p ) ,
        logit(p) <- a_tank[tank] ,
        a_tank[tank] ~ dnorm( a , sigma ) ,
        a ~ dnorm(0,1) ,
        sigma ~ dcauchy(0,1)
    ),
    data=d , iter=4000 , chains=4 )
```

R code 12.3

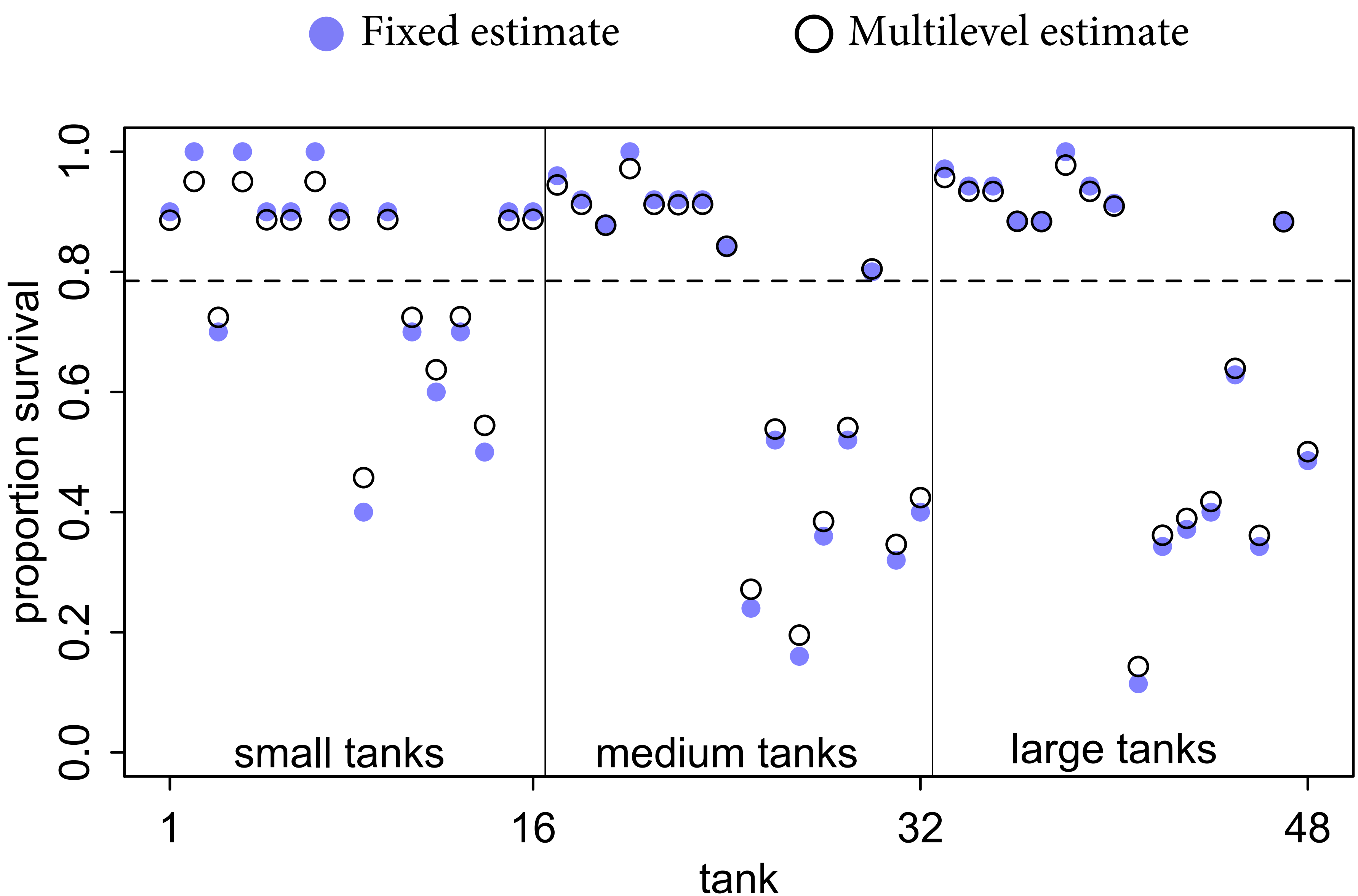
```
m12.2 <- map2stan(
    alist(
        surv ~ dbinom( density , p ) ,
        logit(p) <- a_tank[tank] ,
        a_tank[tank] ~ dnorm( a , sigma ) ,
        a ~ dnorm(0,1) ,
        sigma ~ dcauchy(0,1)
    ),
    data=d , iter=4000 , chains=4 )
```

R code 12.4

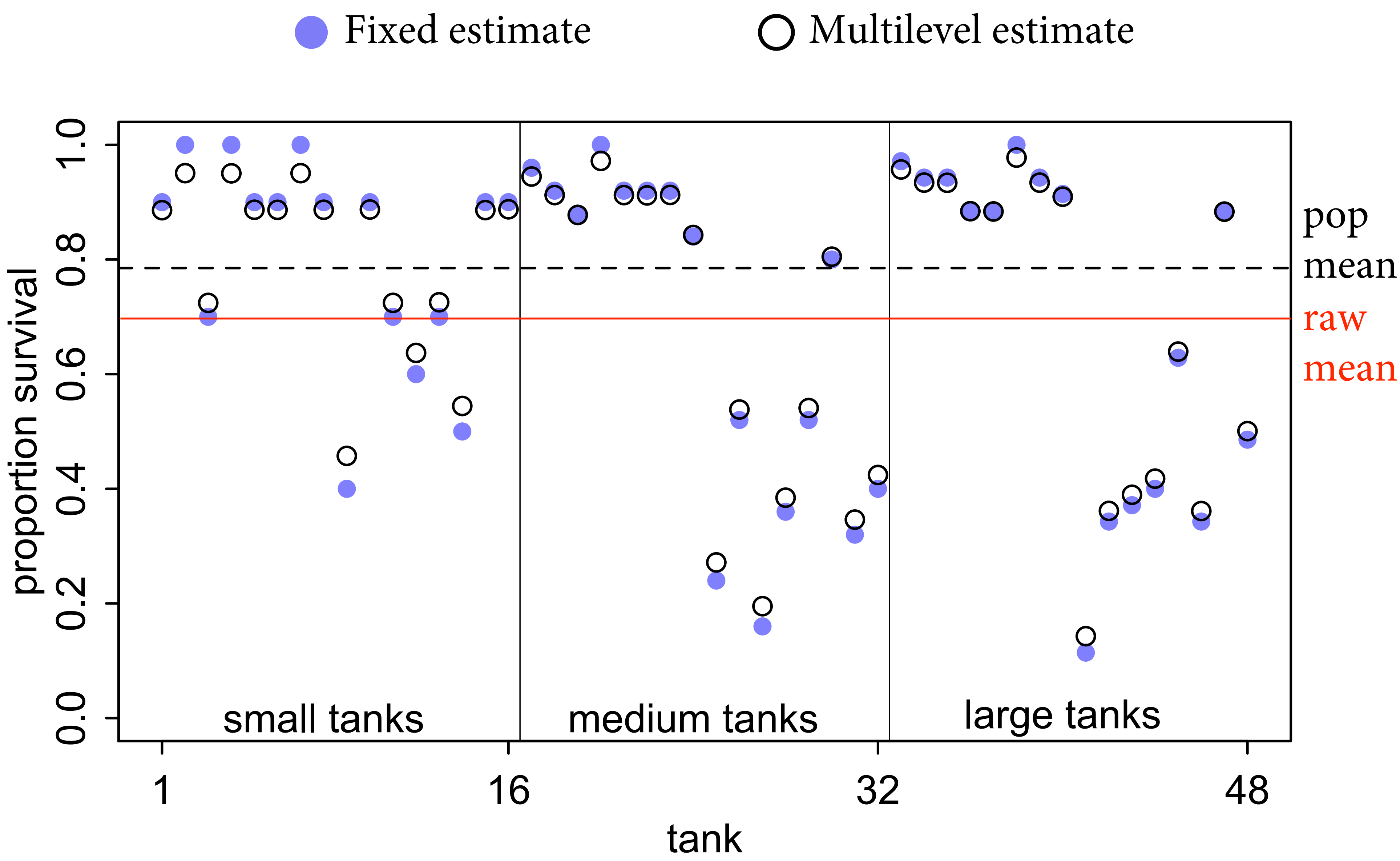
```
compare( m12.1 , m12.2 )
```

```
        WAIC pWAIC dWAIC weight    SE  dSE
m12.2 1010.2  38.0   0.0      1 37.94   NA
m12.1 1023.3  49.4  13.1      0 43.01 6.54
```

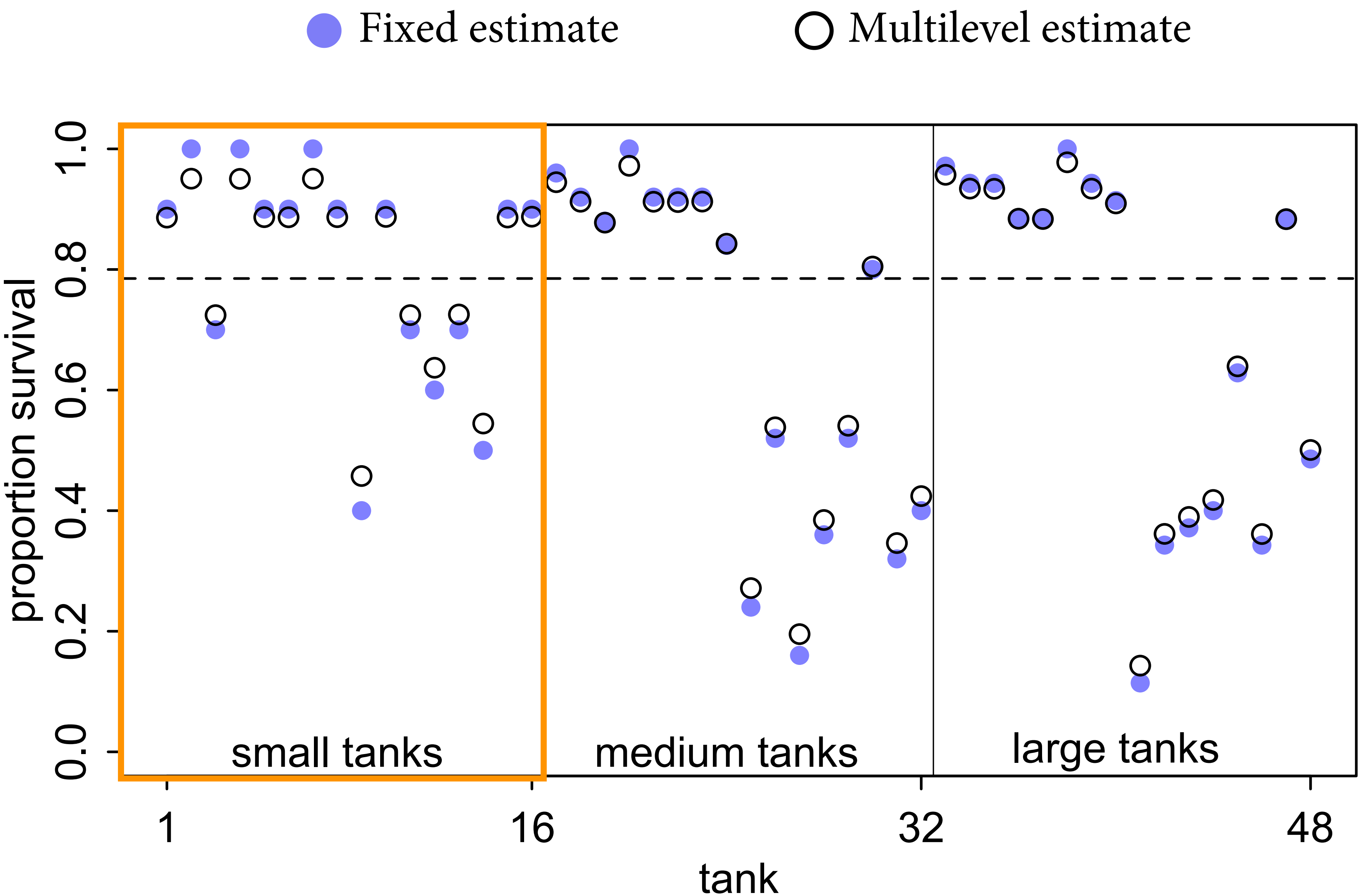
48 tanks + a + sigma => 50 parameters



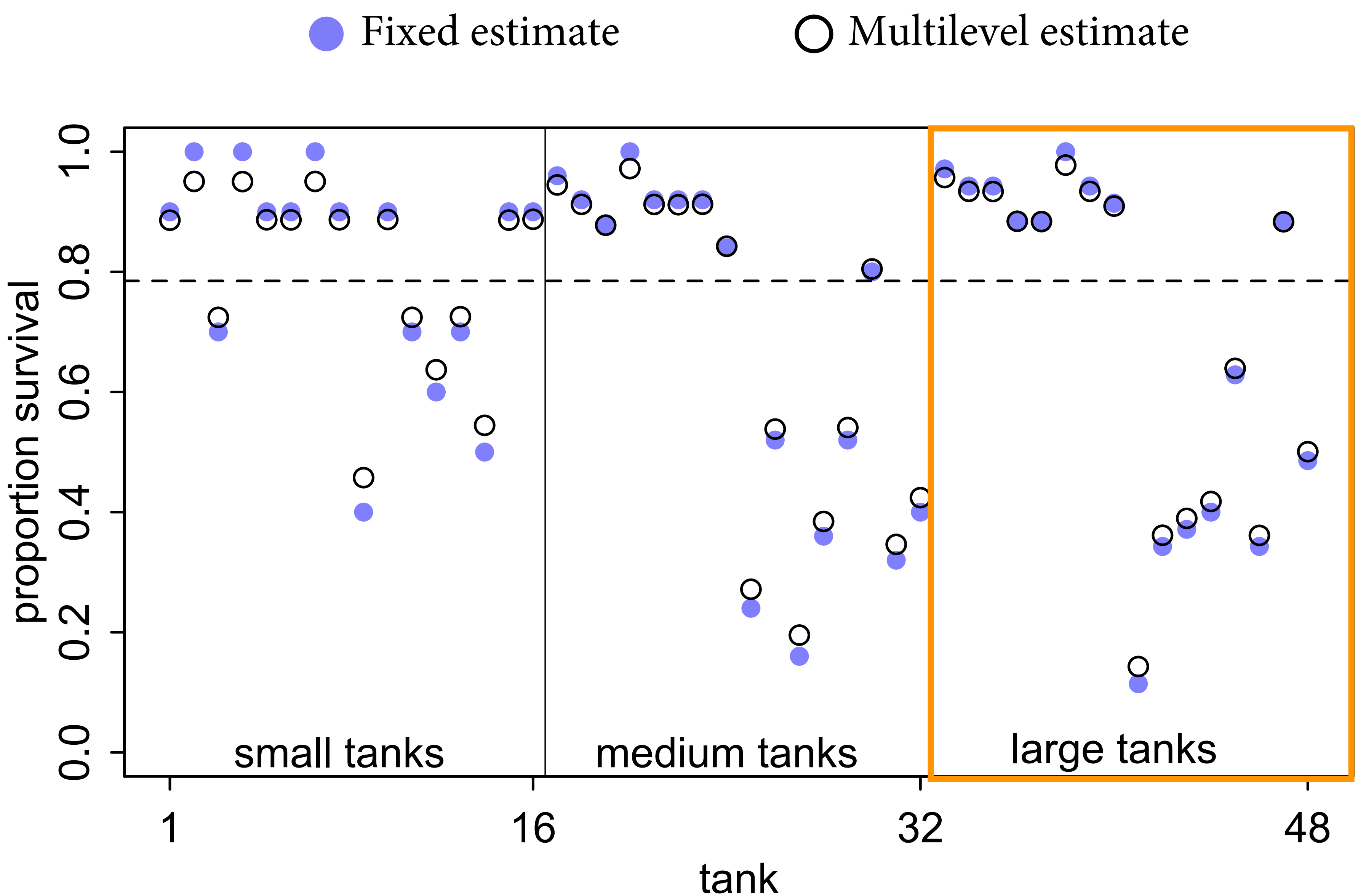
Don't expect predictions to match observations exactly.
Instead expect ***shrinkage***.



Population mean not equal to raw empirical mean. Why?
Imbalance in amount of evidence across tanks.



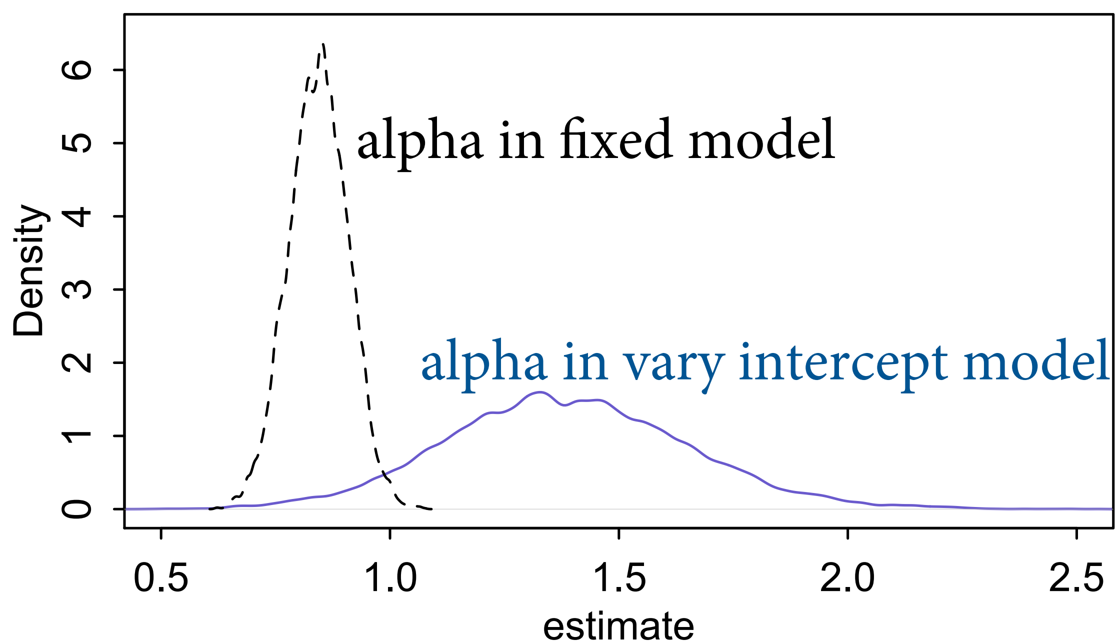
Small tanks => high sampling variation. More shrinkage towards mean. Further from mean => more shrinkage.



Large tanks => low sampling variation. Less shrinkage towards mean at all distances from mean.

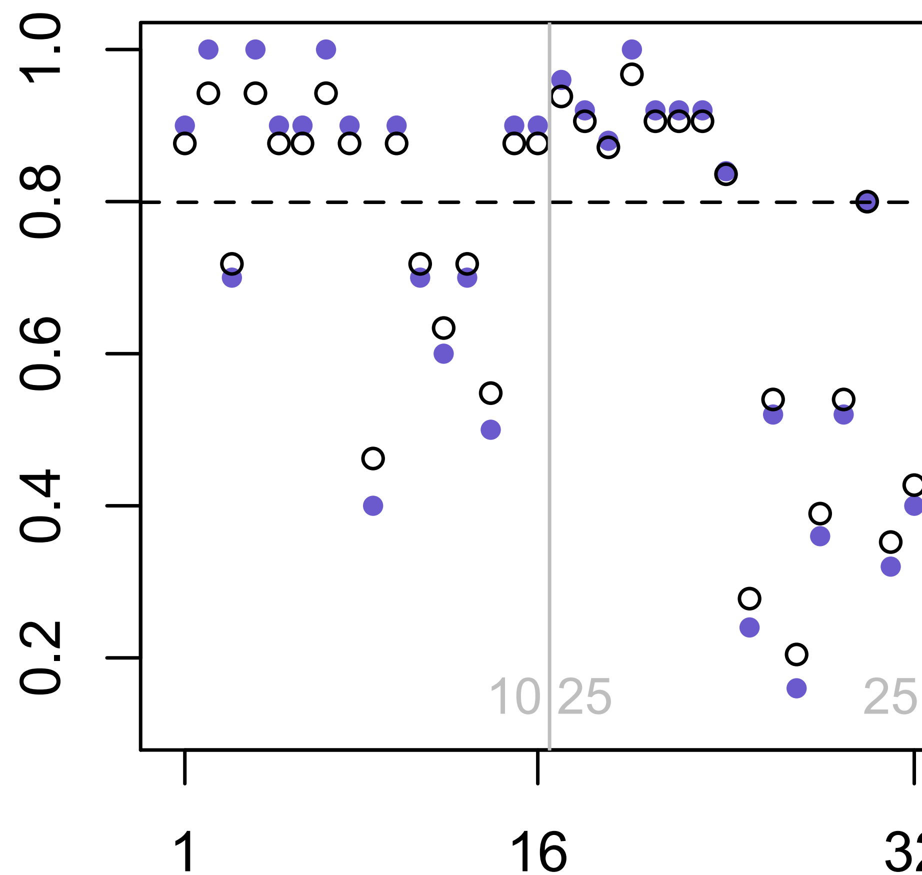
Careful comparing estimates

- Typical for intercept ("fixed effect") to change and become more uncertain
  - Meaning of parameter changes: no longer mean of data, but rather mean of distribution of intercepts
  - Uncertainty larger, because many combinations of alpha, sigma, a[tank]'s can produce same empirical mean of data



Shrinkage

- Varying effect estimates *shrink* towards mean (alpha)
- Further from mean, more shrinkage
- Fewer data in cluster, more shrinkage
- Same as regression to the mean, really



Pooling

- Shrinkage arises from ***pooling***
- Each tank informs estimates of other tanks
- The model doesn't have amnesia!
- Effect of pooling influenced by
  - amount of data in cluster
  - amount of variation among clusters (sigma)

Pool, or the bad guys win

Ulysses' Compass again

- Why are ***varying effects*** (partial pooling) more accurate than ***fixed effects*** (no pooling)?
- Grand mean: maximum underfitting
- Fixed effects: maximum overfitting
- Varying effects: adaptive regularization