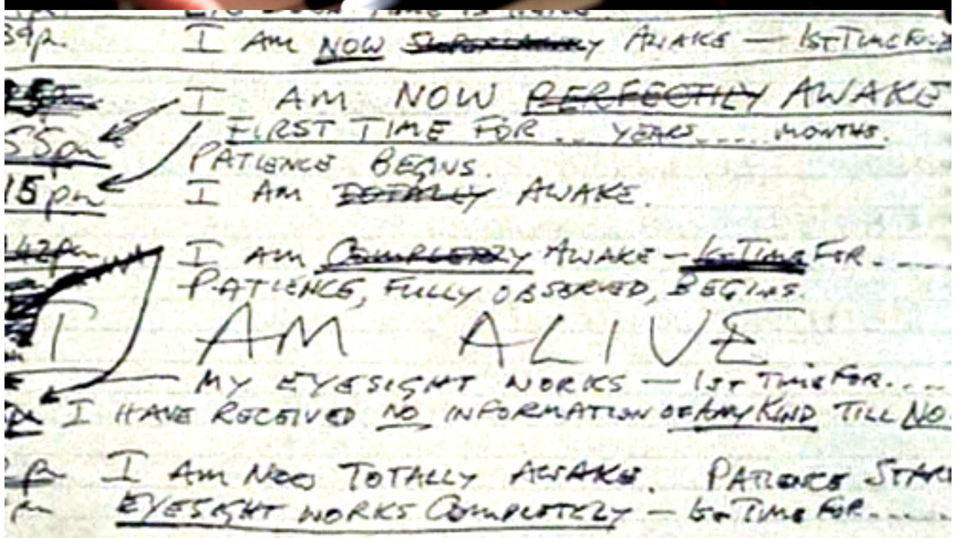
Statistical Rethinking
Winter 2019

Lecture 15 / Week 8

Models With Memory

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

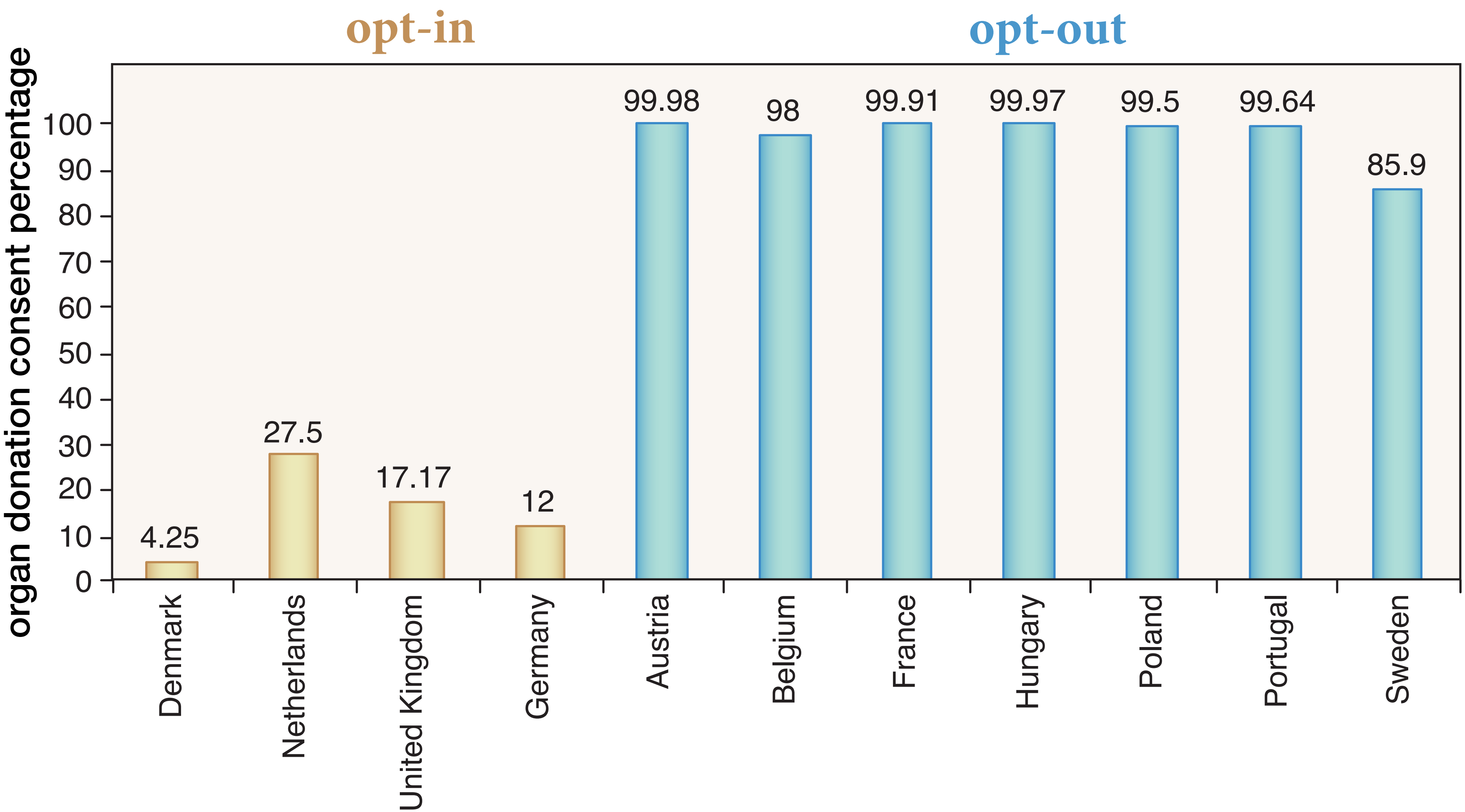
- "Fixed effects" models have anterograde amnesia
  - Every new cluster (individual, pond, road, classroom) is a new world
  - No information passed among clusters
- Multilevel models remember and pool information
  - Properties of clusters come from a "population"
  - Inferred population defines pooling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

Learning, forward and back

Depends upon variation

Cause & Reconciliation

- Fighting a battle on two fronts:
- (1) Causal inference — don't make causal salad
- (2) Functional inference — estimation not trivial
- No unique solutions, but lots of good options


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 program with `ulam`
- Methods of plotting and comparing
- Open up more options

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

$$
\begin{aligned}
S_i &\sim \text{Binomial}(N_i, p_i) \\
\text{logit}(p_i) &= \alpha_{\text{TANK}[i]} \\
\alpha_j &\sim \text{Normal}(0, 1.5)
\end{aligned}
$$

regularizing prior

Regularized intercepts

R code 13.2

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

dat <- list(
    S = d$surv,
    N = d$density,
    tank = d$tank )

# approximate posterior
m13.1 <- ulam(
    alist(
        S ~ dbinom( N , p ) ,
        logit(p) <- a[tank] ,
        a[tank] ~ dnorm( 0 , 1.5 )
    ), data=dat , chains=4 , log_lik=TRUE )
```

$$
\begin{aligned}
S_i &\sim \text{Binomial}(N_i, p_i) \\
\text{logit}(p_i) &= \alpha_{\text{TANK}[i]} \\
\alpha_j &\sim \text{Normal}(0, 1.5)
\end{aligned}
$$

Adaptive regularization

$$
\begin{aligned}
S_i &\sim \text{Binomial}(N_i, p_i) \\
\text{logit}(p_i) &= \alpha_{\text{TANK}[i]} \\
\alpha_j &\sim \text{Normal}(\bar{\alpha}, \sigma) \\
\bar{\alpha} &\sim \text{Normal}(0, 1.5) \\
\sigma &\sim \text{Exponential}(1)
\end{aligned}
$$

Adaptive regularization

$$
\begin{aligned}
S_i &\sim \text{Binomial}(N_i, p_i) \\
\text{logit}(p_i) &= \alpha_{\text{TANK}[i]} \\
\textit{varying intercepts} \longrightarrow \alpha_j &\sim \text{Normal}(\bar{\alpha}, \sigma) \\
\bar{\alpha} &\sim \text{Normal}(0, 1.5) \\
\sigma &\sim \text{Exponential}(1)
\end{aligned}
$$

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

$$S_i \sim \text{Binomial}(N_i, p_i)$$
$$\text{logit}(p_i) = \alpha_{\text{TANK}[i]}$$
varying intercepts $\longrightarrow$ $$\alpha_j \sim \text{Normal}(\bar{\alpha}, \sigma)$$

mean *standard deviation*

Adaptive regularization

$$
\begin{aligned}
S_i &\sim \text{Binomial}(N_i, p_i) \\
\text{logit}(p_i) &= \alpha_{\text{TANK}[i]} \\
\alpha_j &\sim \text{Normal}(\bar{\alpha}, \sigma) \\
\bar{\alpha} &\sim \text{Normal}(0, 1.5) \\
\sigma &\sim \text{Exponential}(1)
\end{aligned}
$$

varying intercepts

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

$$
\begin{aligned}
S_i &\sim \text{Binomial}(N_i, p_i) \\
\text{logit}(p_i) &= \alpha_{\text{TANK}[i]} \\
\alpha_j &\sim \text{Normal}(\bar{\alpha}, \sigma) \\
\bar{\alpha} &\sim \text{Normal}(0, 1.5) \\
\sigma &\sim \text{Exponential}(1)
\end{aligned}
$$

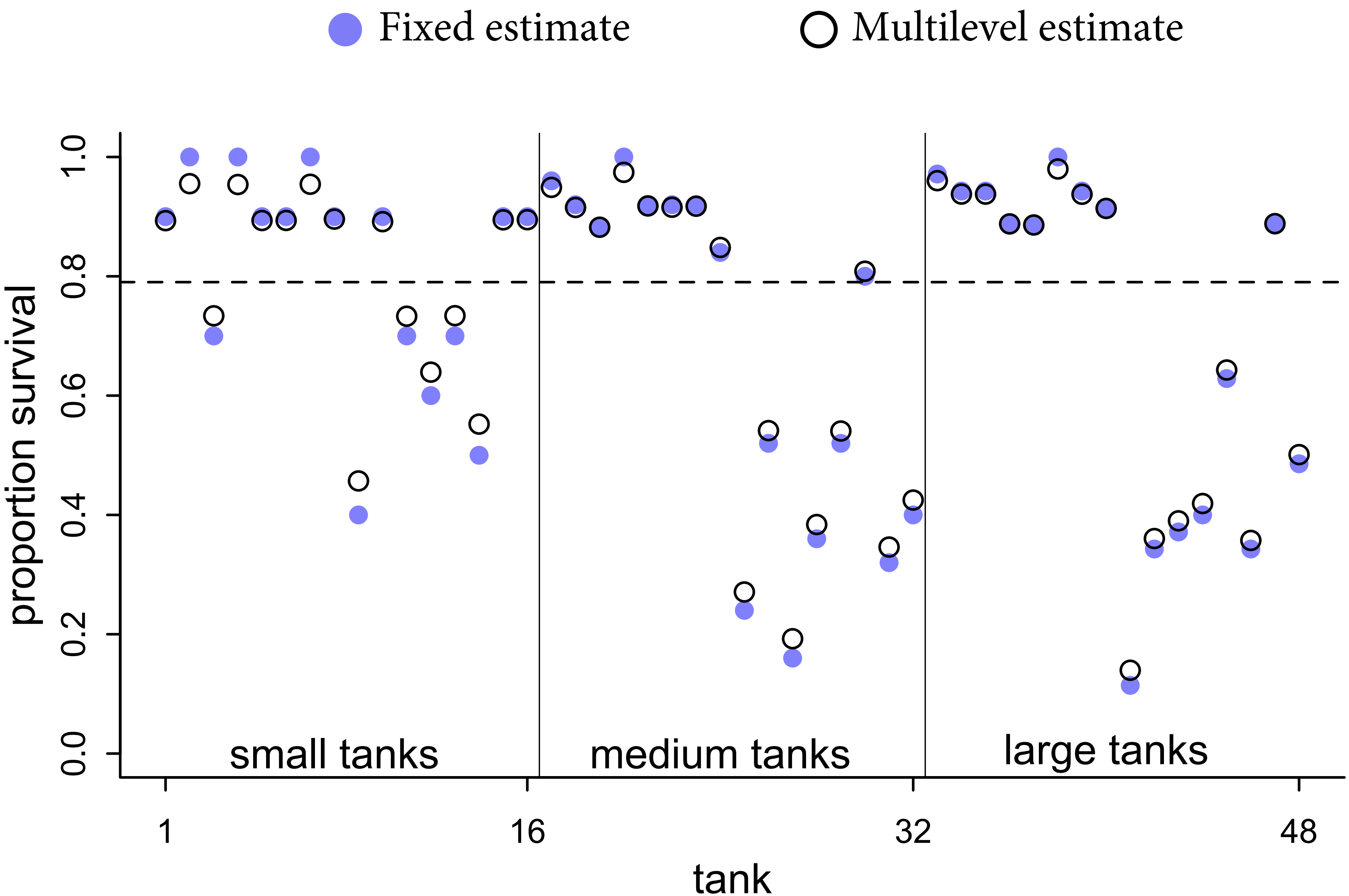
```
m13.2 <- ulam(
    alist(
        S ~ dbinom( N , p ) ,
        logit(p) <- a[tank] ,
        a[tank] ~ dnorm( a_bar , sigma ) ,
        a_bar ~ dnorm( 0 , 1.5 ) ,
        sigma ~ dexp( 1 )
    ), data=dat , chains=4 , log_lik=TRUE )
```

R code 13.4

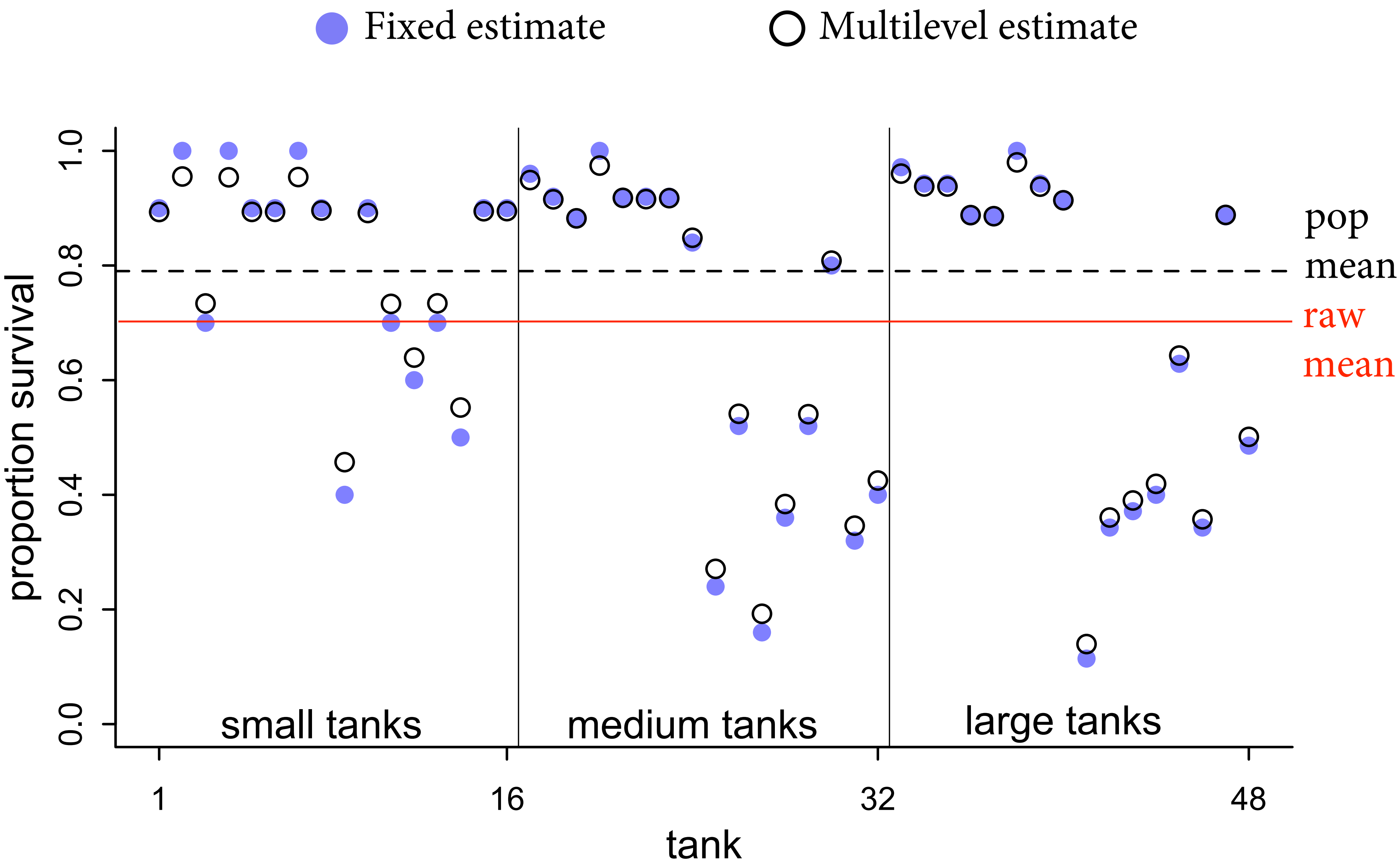
```
compare( m13.1 , m13.2 )
```

```
      WAIC pWAIC dWAIC weight   SE  dSE
m13.2  202  21.7     0      1 7.35   NA
m13.1  213  24.8    11      0 4.71 3.58
```

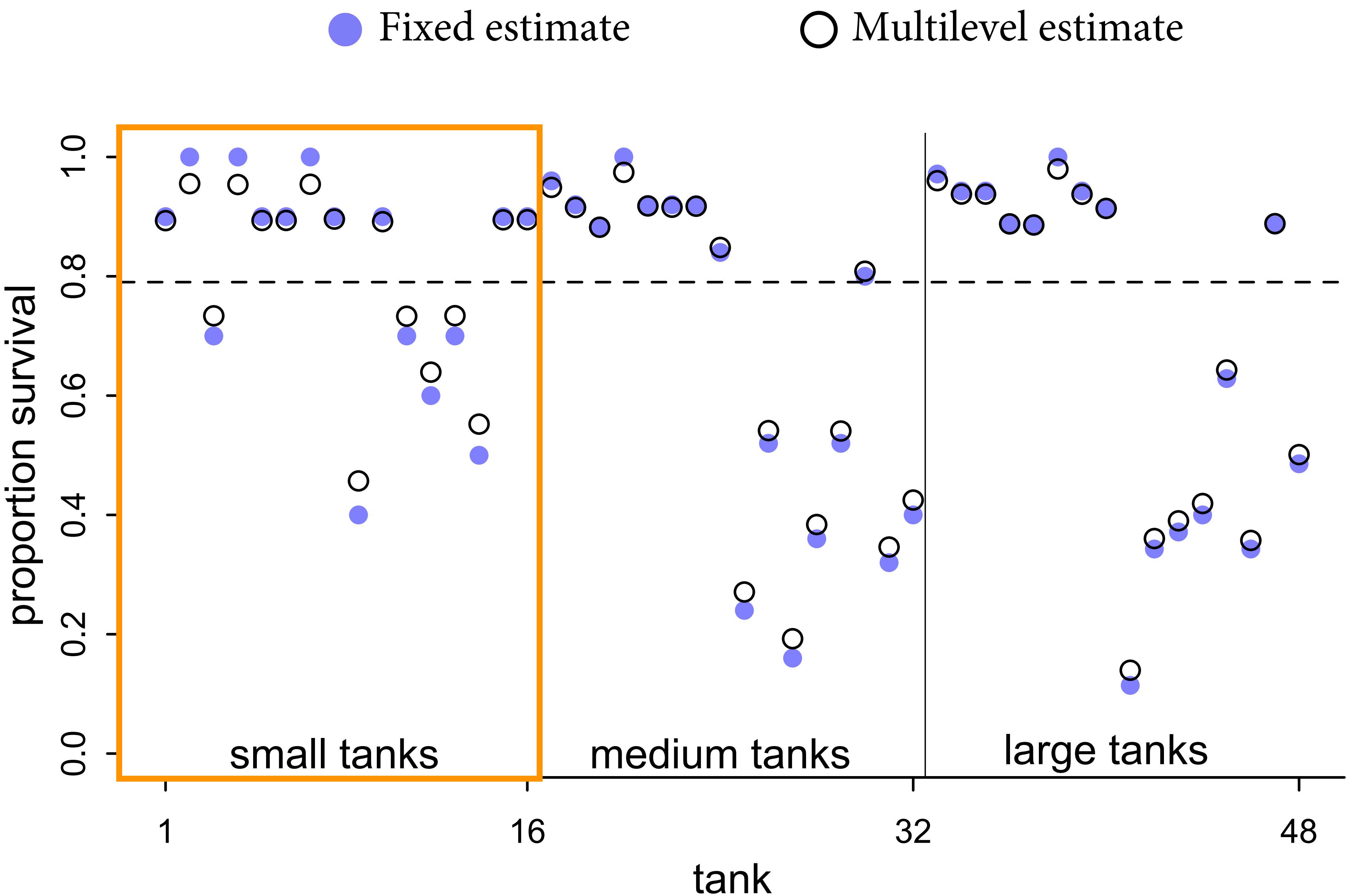
- m13.1: 48 parameters vs 25 effective
- m13.2: 50 parameters vs 22 effective
- Model with more parameters has fewer effective parameters
- Why? Ended up with stronger prior.



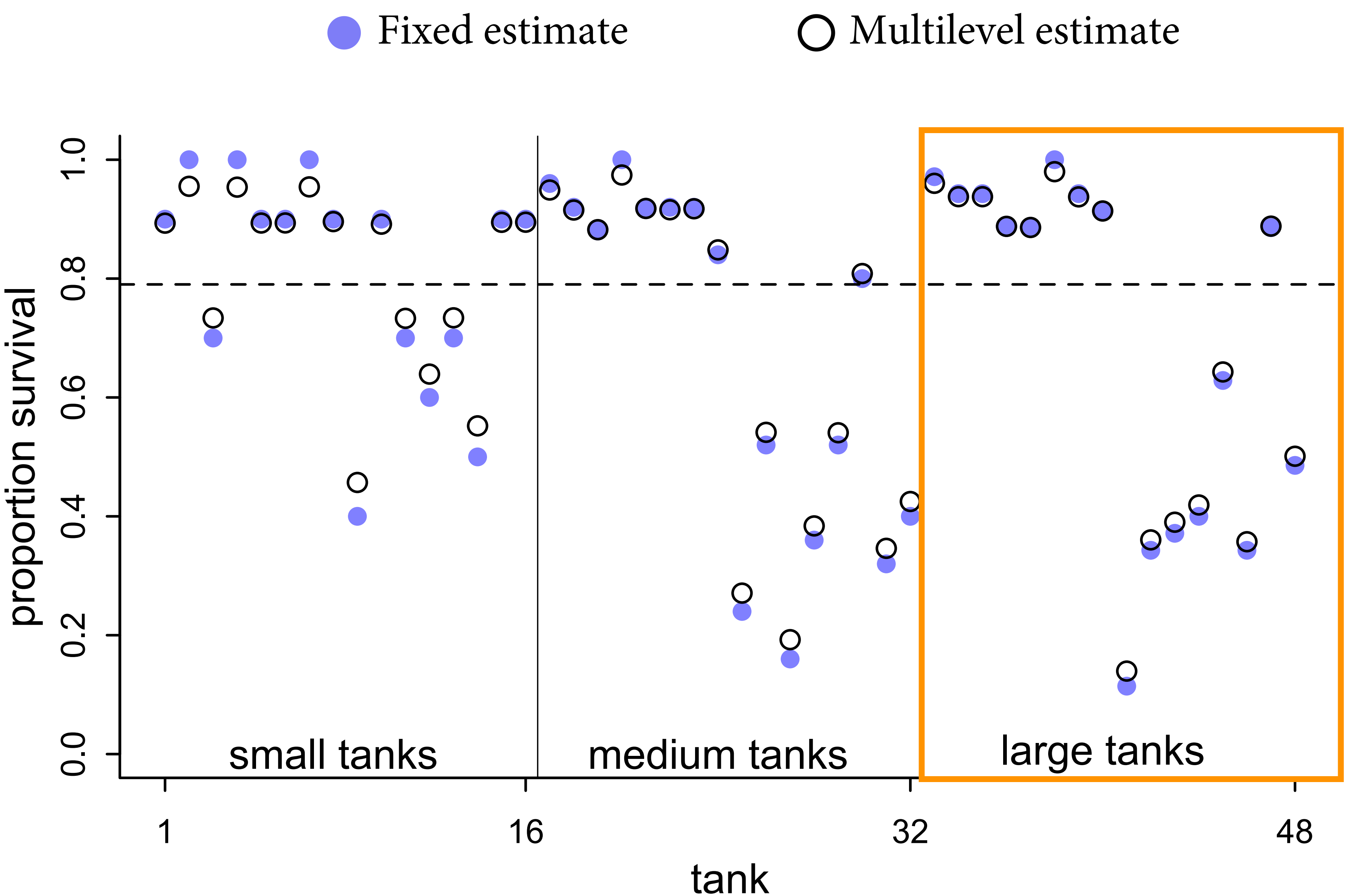
Don't expect predictions to match observations exactly. Instead expect ***shrinkage***.



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



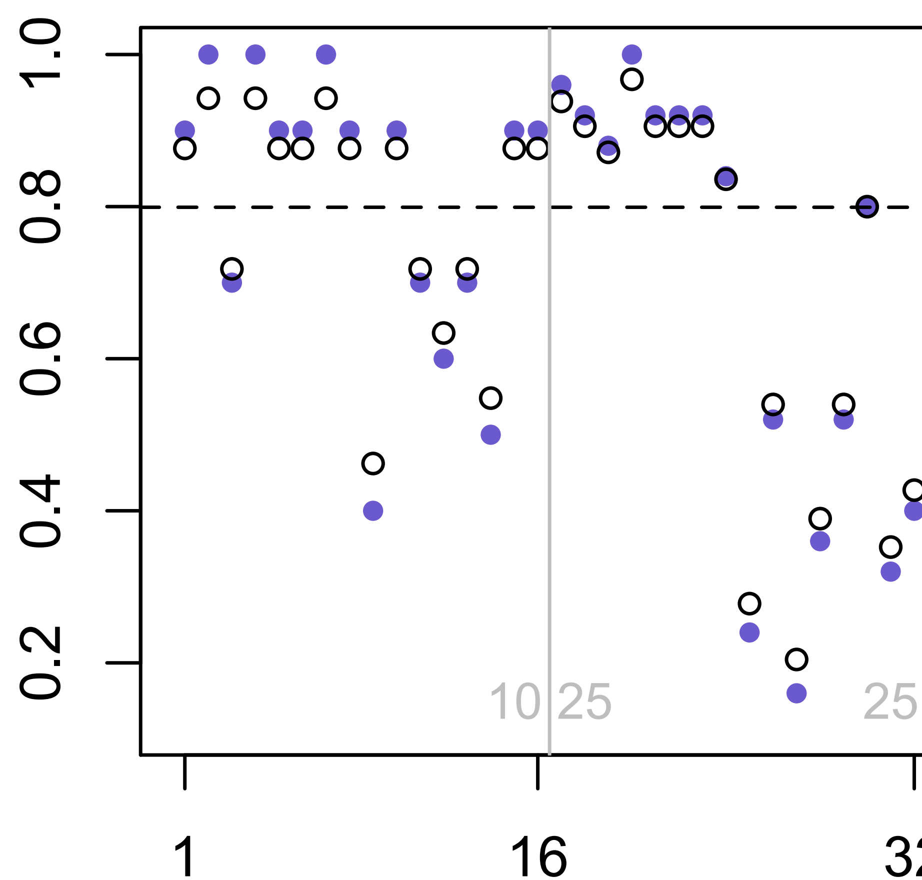
Small tanks => high sampling variation. More shrinkage towards mean. Further from mean => more shrinkage.



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

Shrinkage

- Varying effect estimates *shrink* towards mean (a_bar)
- Further from mean, more shrinkage
- Fewer data in cluster, more shrinkage
- Same as regression to the mean, really
- Shrinkage results from *pooling* of information

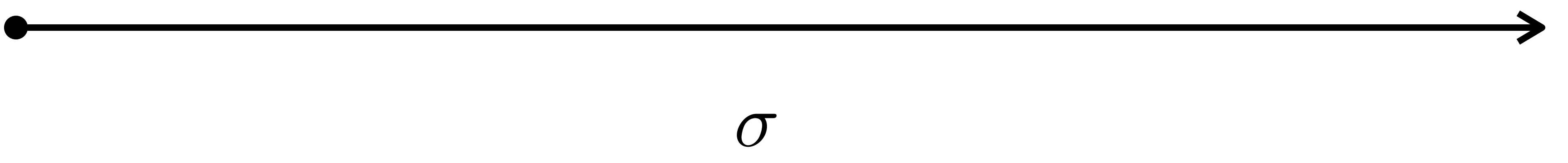


Ulysses' Compass again

- 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

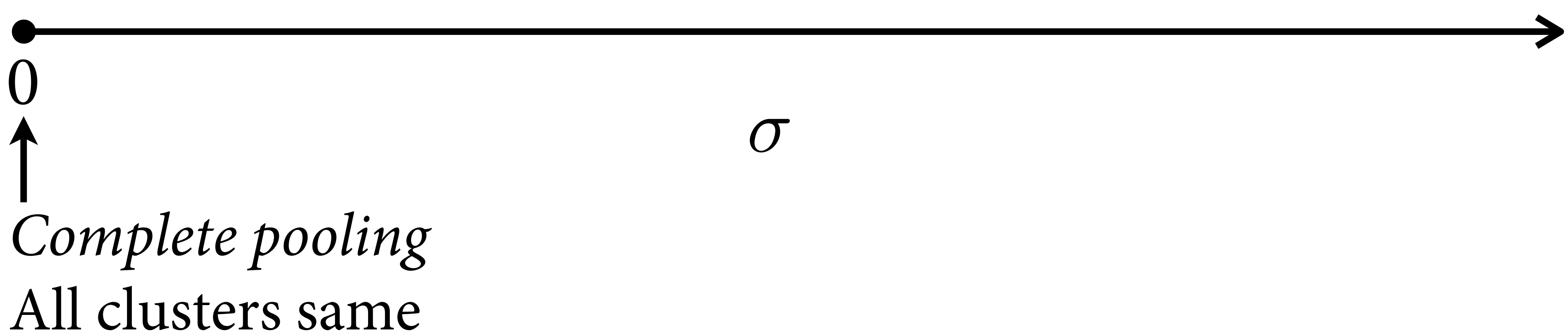
Ulysses' Compass again

$$
\begin{aligned}
S_i &\sim \text{Binomial}(N_i, p_i) \\
\text{logit}(p_i) &= \alpha_{\text{TANK}[i]} \\
\alpha_j &\sim \text{Normal}(\bar{\alpha}, \sigma) \\
\bar{\alpha} &\sim \text{Normal}(0, 1.5) \\
\sigma &\sim \text{Exponential}(1)
\end{aligned}
$$



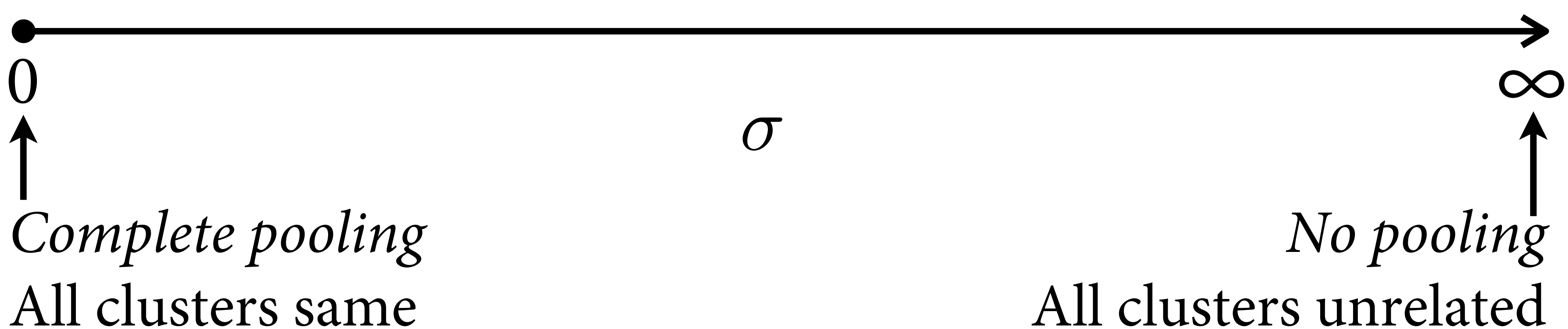
Ulysses' Compass again

$$
\begin{aligned}
S_i &\sim \text{Binomial}(N_i, p_i) \\
\text{logit}(p_i) &= \alpha_{\text{TANK}[i]} \\
\alpha_j &\sim \text{Normal}(\bar{\alpha}, \sigma) \\
\bar{\alpha} &\sim \text{Normal}(0, 1.5) \\
\sigma &\sim \text{Exponential}(1)
\end{aligned}
$$



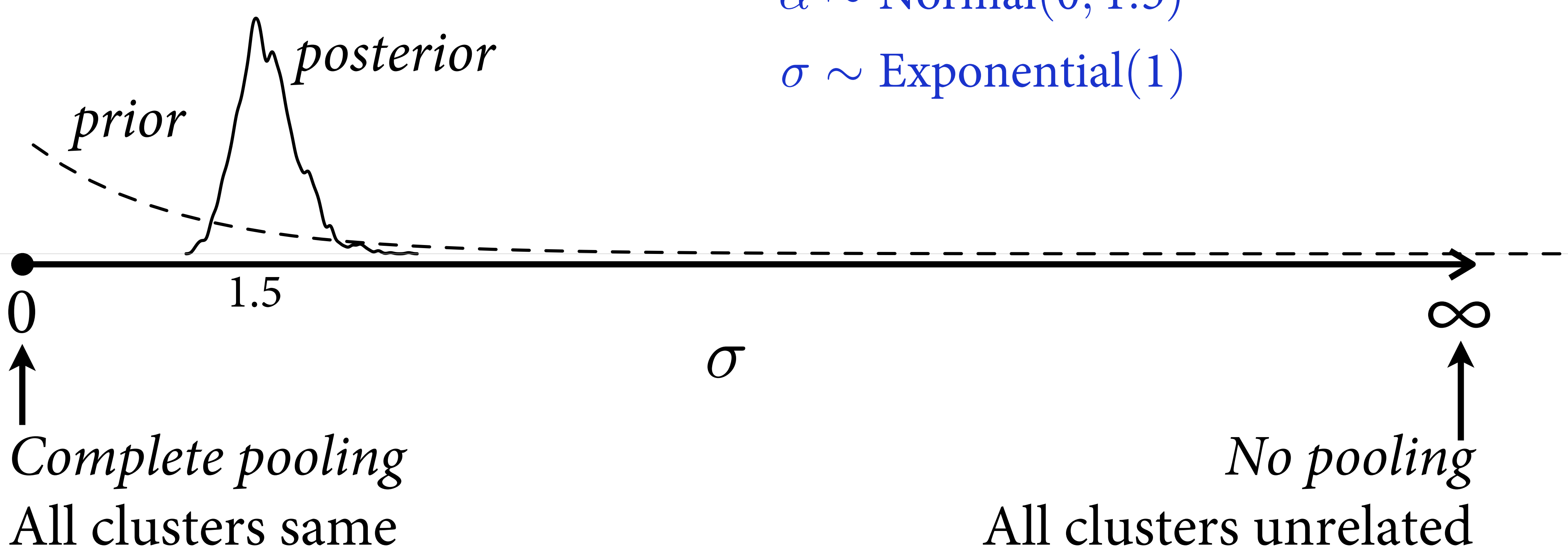
Ulysses' Compass again

$$
\begin{aligned}
S_i &\sim \text{Binomial}(N_i, p_i) \\
\text{logit}(p_i) &= \alpha_{\text{TANK}[i]} \\
\alpha_j &\sim \text{Normal}(\bar{\alpha}, \sigma) \\
\bar{\alpha} &\sim \text{Normal}(0, 1.5) \\
\sigma &\sim \text{Exponential}(1)
\end{aligned}
$$

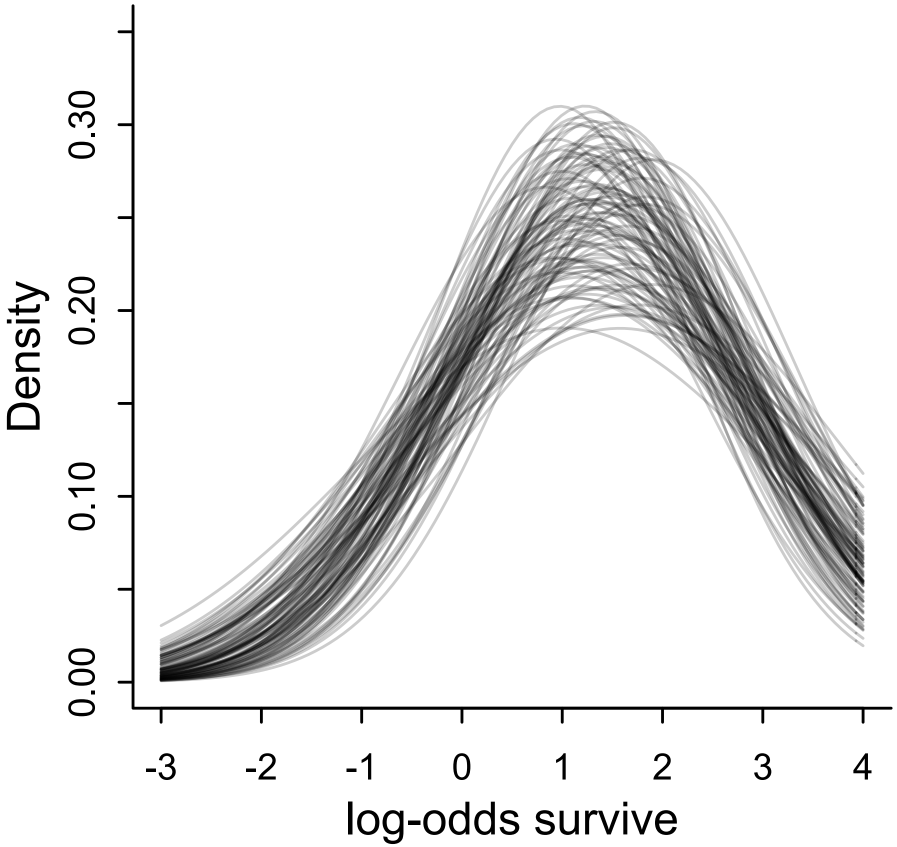


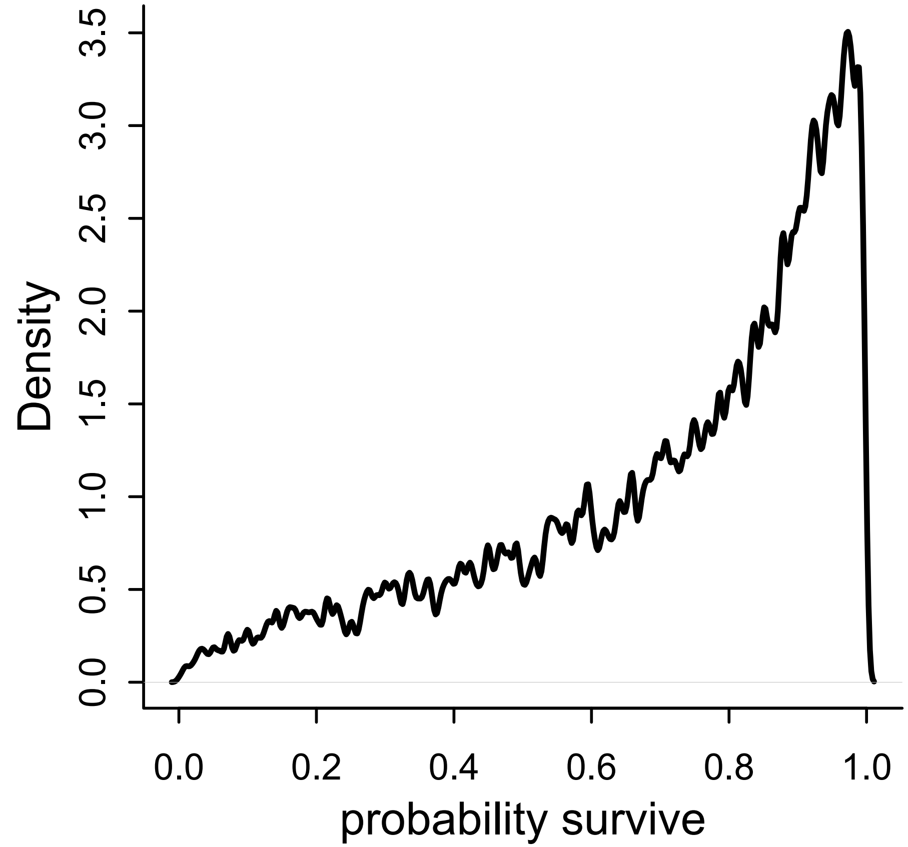
Ulysses' Compass again

$$
\begin{aligned}
S_i &\sim \text{Binomial}(N_i, p_i) \\
\text{logit}(p_i) &= \alpha_{\text{TANK}[i]} \\
\alpha_j &\sim \text{Normal}(\bar{\alpha}, \sigma) \\
\bar{\alpha} &\sim \text{Normal}(0, 1.5) \\
\sigma &\sim \text{Exponential}(1)
\end{aligned}
$$



Regularizing distribution

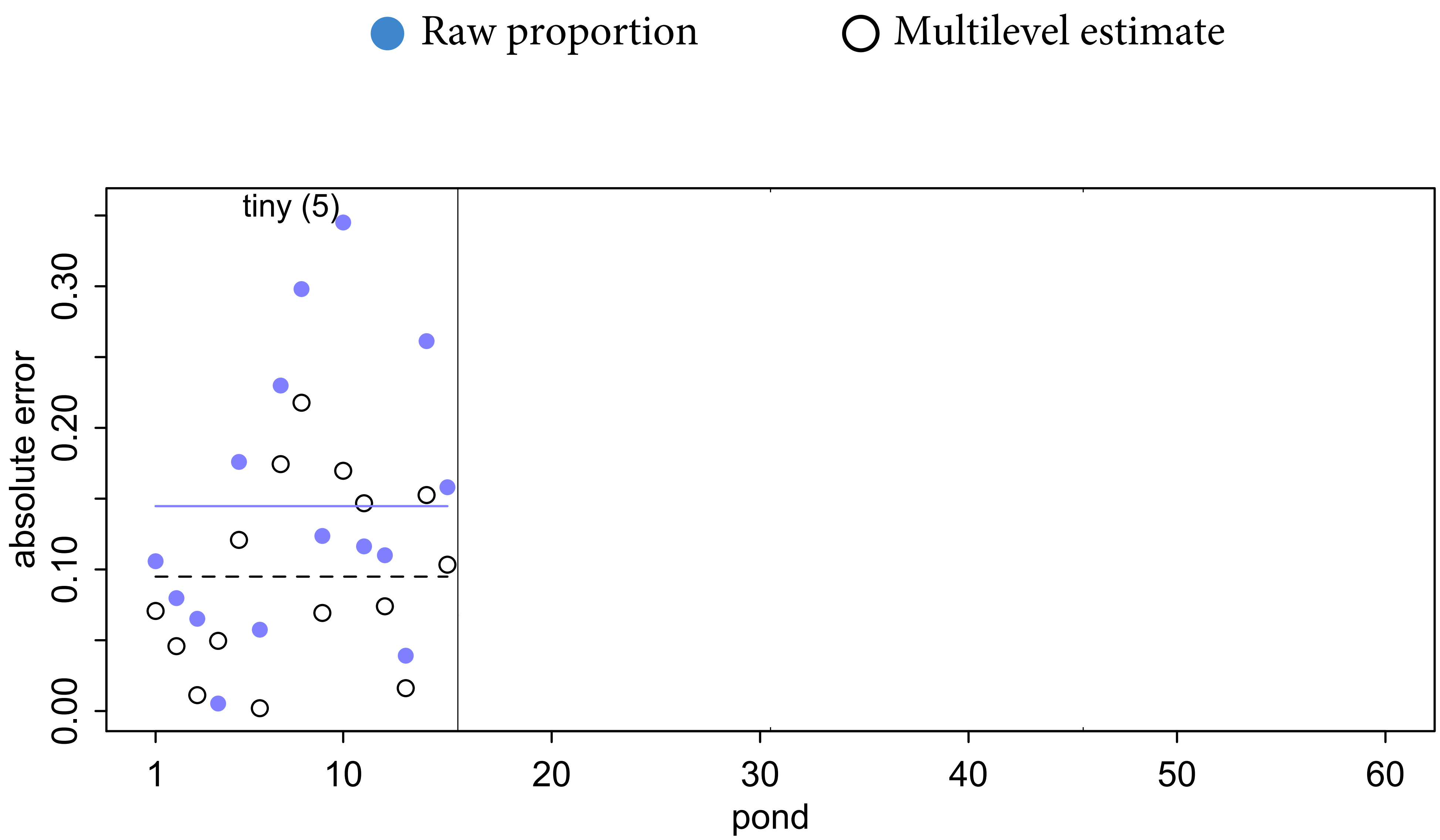


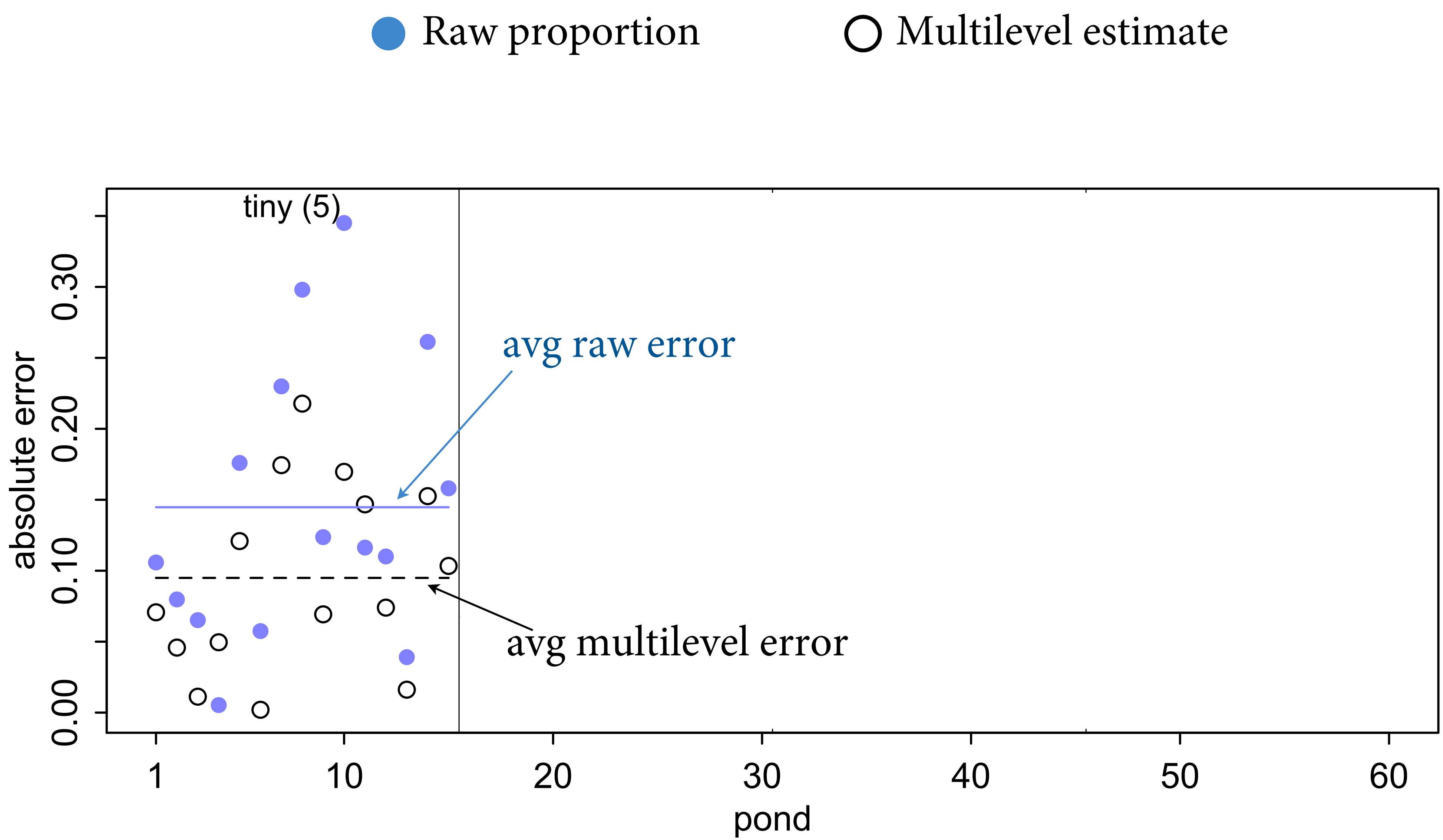


Ulysses' Compass again

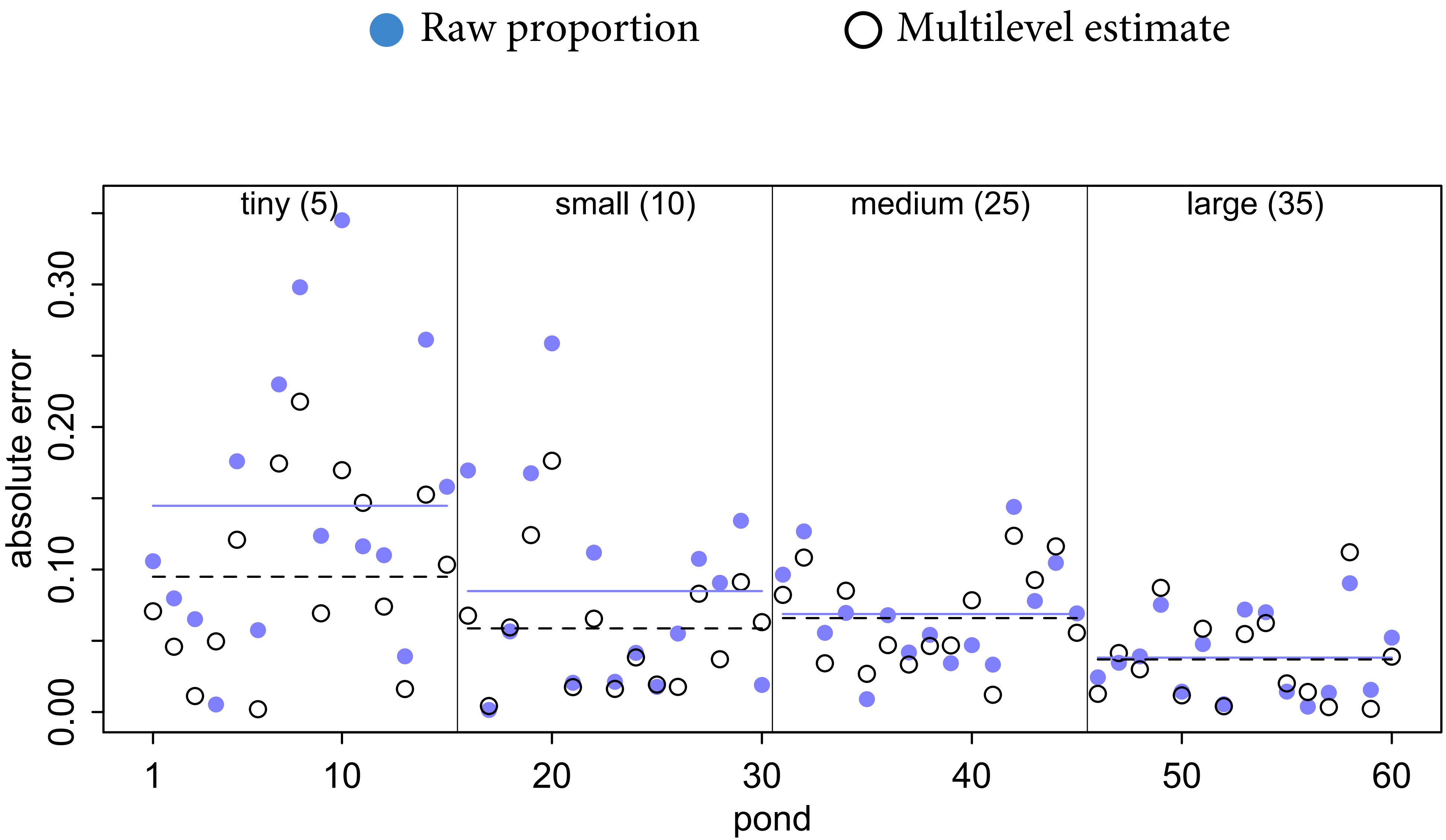
- Simulate to demonstrate accuracy advantage
  - 60 ponds
  - 5, 10, 25, 35 tadpoles each of 15

	pond	n	true.a	s	p.nopool	p.partpool	p.true
1	1	5	-3.089936132	1	0.2000000	0.32173203	0.04352429
2	2	5	0.267290817	5	1.0000000	0.91305884	0.56642768
3	3	5	0.896554101	4	0.8000000	0.79164823	0.71024085
4	4	5	1.934806220	5	1.0000000	0.91276066	0.87378044
5	5	5	-0.758682067	0	0.0000000	0.17692527	0.31893247
6	6	5	3.904836388	5	1.0000000	0.91337140	0.98025353
7	7	5	2.271914139	4	0.8000000	0.79349508	0.90652411
8	8	5	2.886101619	4	0.8000000	0.79557800	0.94715510
9	9	5	1.436457877	3	0.6000000	0.64219989	0.80790553
10	10	5	1.156079068	3	0.6000000	0.64414477	0.76061953


Raw proportion
Multilevel estimate
tiny (5)
absolute error
0.00
0.10
0.20
0.30
1
10
20
30
40
50
60
pond



Raw proportion
Multilevel estimate
tiny (5)
avg raw error
avg multilevel error
absolute error
0.00
0.10
0.20
0.30
1
10
20
30
40
50
60
pond


Raw proportion
Multilevel estimate
tiny (5)
small (10)
medium (25)
large (35)
absolute error
0.00
0.10
0.20
0.30
1
10
20
30
40
50
60
pond