Agenda

1. Roadmap
2. Learning from categories using pooled and unpooled models
3. Partial pooling
4. Specifying nested models
5. Indexing in models in R

Roadmap—the course so far

1. Redefining linear regression in a Bayesian framework

Outcome as draws from a probability distribution

$$Y_i \sim \text{Norm}(\mu_i, \sigma)$$

$$\mu_i = X_i \beta$$

Center of distribution modelled as linear combination of predictors

Roadmap—the course so far

1. Redefining linear regression in a Bayesian framework
2. Dealing with discrete outcome variables using GLM

Changing the outcome distribution to fit the data

$$Y_i \sim \text{Pois}(\lambda_i)$$

$$\log(\lambda_i) = X_i \beta$$

Using link functions to map linear combination of predictors to an appropriate scale

Roadmap—the course so far

1. Redefining linear regression in a Bayesian framework
2. Dealing with discrete outcome variables using GLM
3. **Dealing with structured predictors using hierarchical linear models**

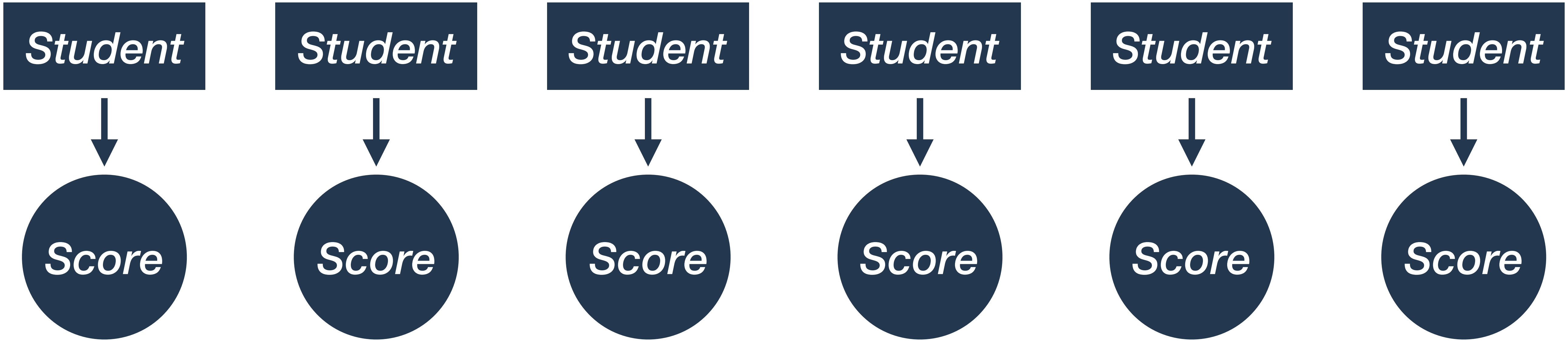
$$Y_i \sim \mathrm{Norm}(\mu_i, \sigma)$$
$$\mu_i = \boxed{X_i \beta}$$

Adding structure to the linear model to account for relationships between observations

Framing the problem

Student performance on a standardized test

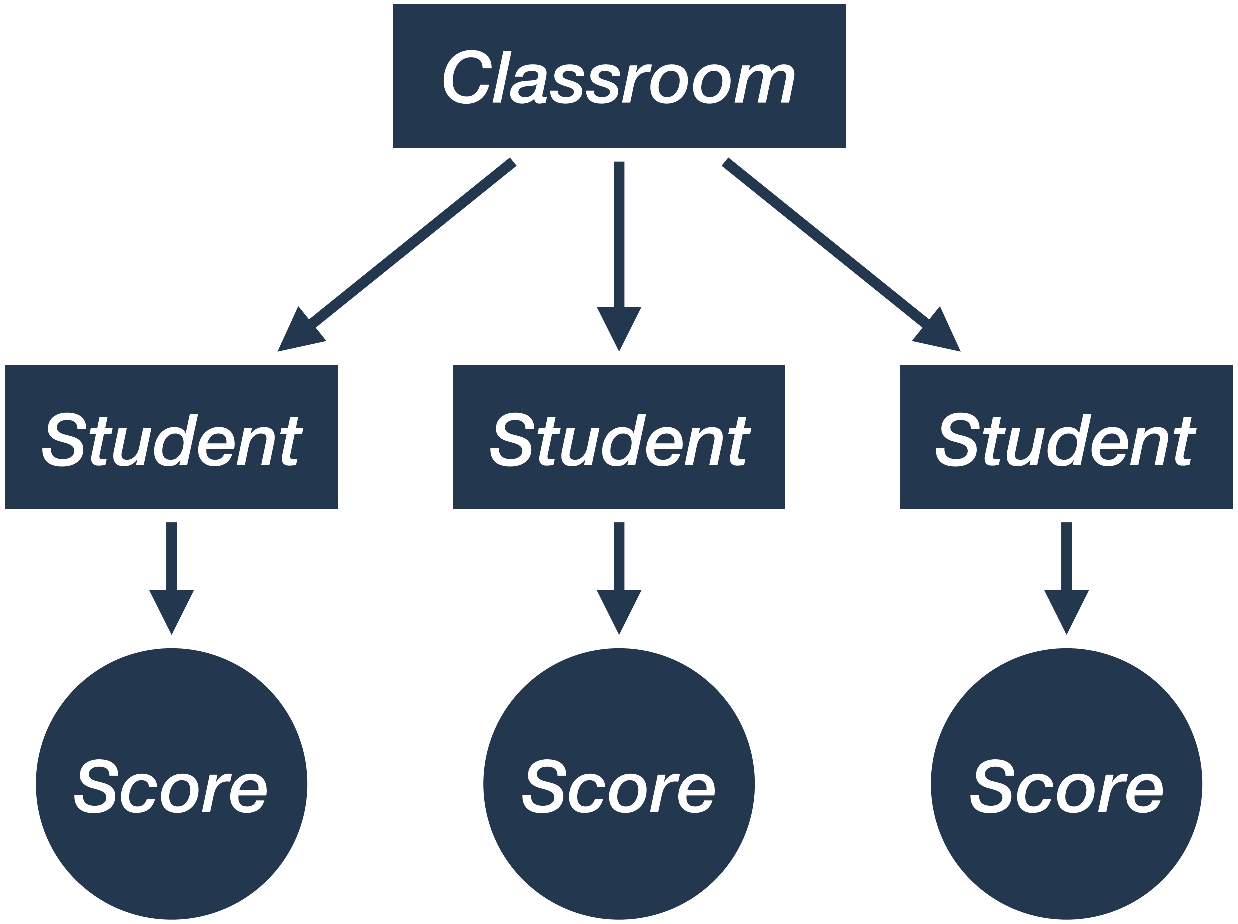
Each student has a score, which we can model using characteristics of the student, their school, etc.



Framing the problem

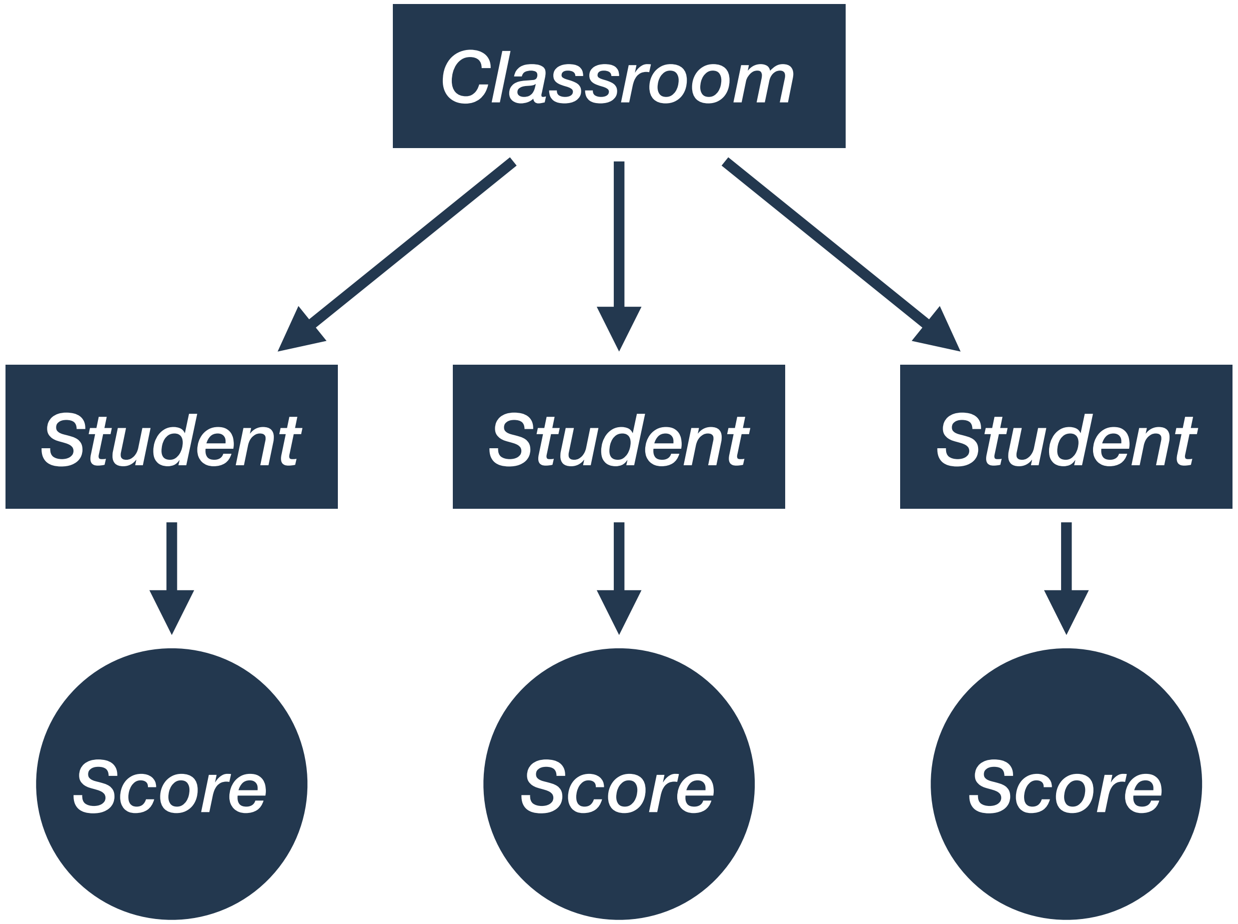
Student performance on a standardized test

Each student has a score, which we can model using characteristics of the student, their school, etc.



Students' performance is not independent

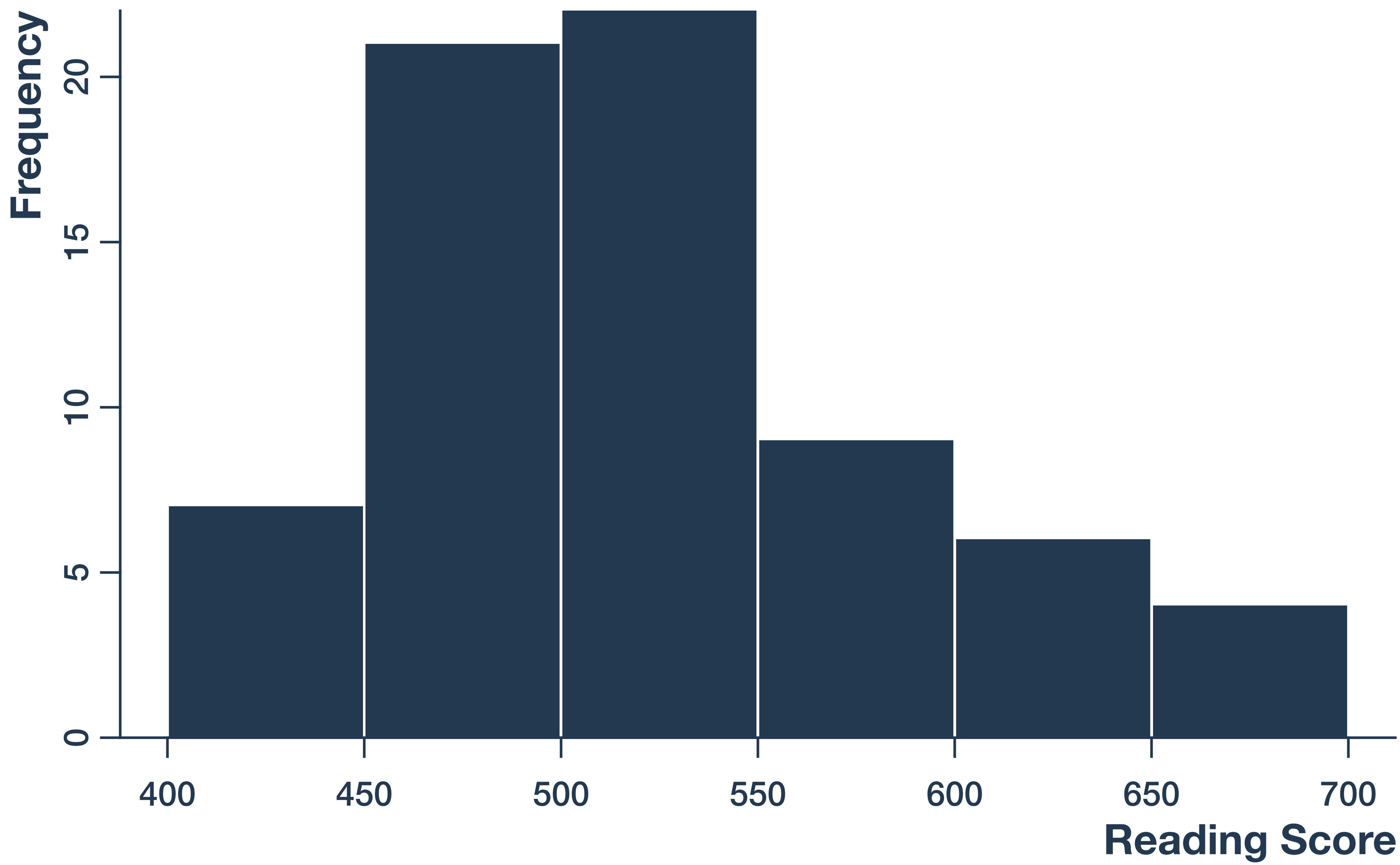
Students in the same classroom share many relevant characteristics (teacher, school, funding, classroom environment, etc.)



Tennessee STAR study

Sample | 69 students in 16 grade-one classrooms from Tennessee schools in 1986

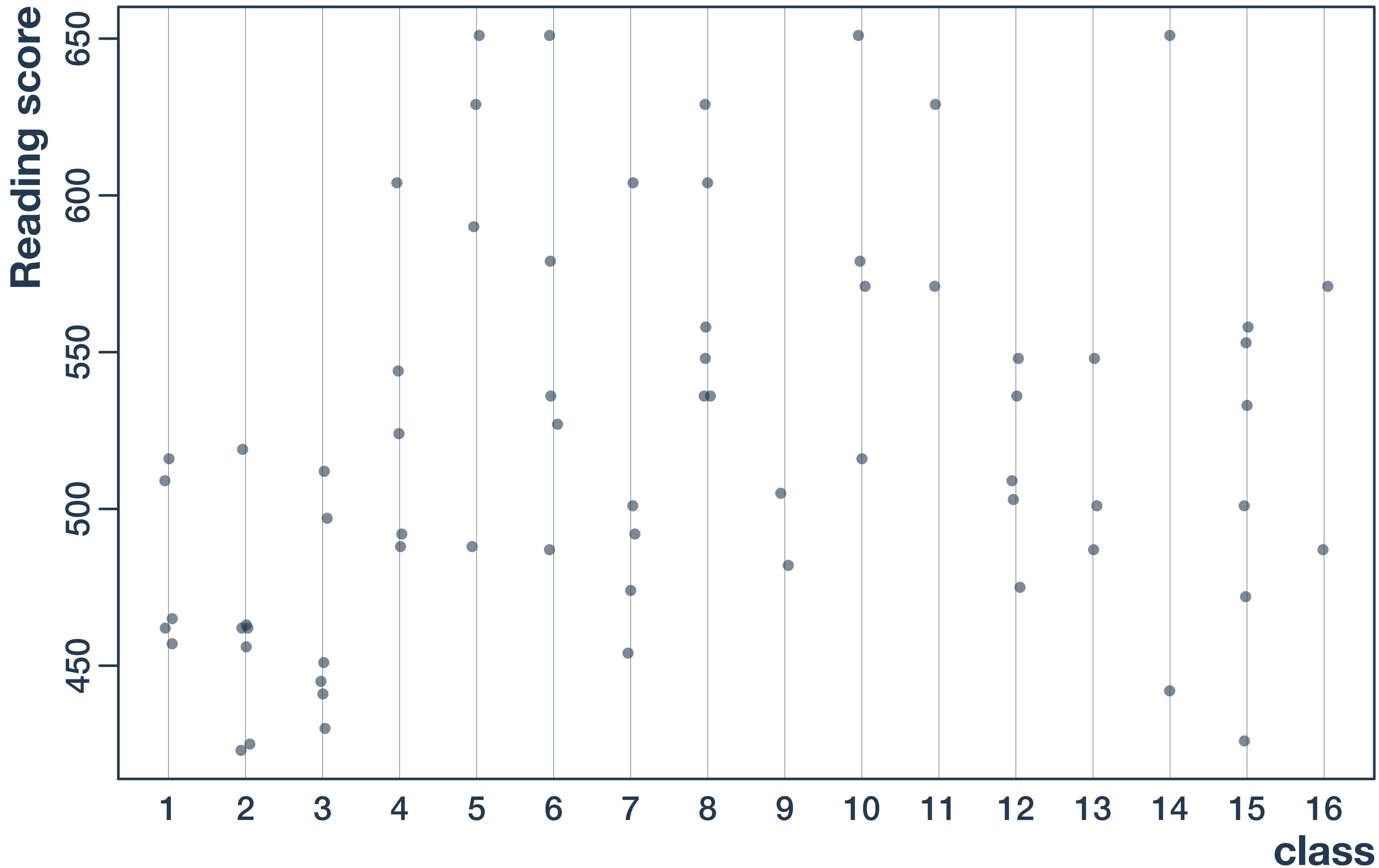
Outcome | Standardized reading score on Stanford Achievement Test (SAT)



Data

Sample | 69 students in 16 grade-one classrooms from Tennessee schools in 1986

Outcome | Standardized reading score on Stanford Achievement Test (SAT)



Complete pooling

All variation attributed to global standard deviation σ

Strategy 1 (pooling)

Ignore hierarchical structure of data and pool all students' scores together.

$$S_i \sim \text{Norm}(\mu_i, \sigma)$$

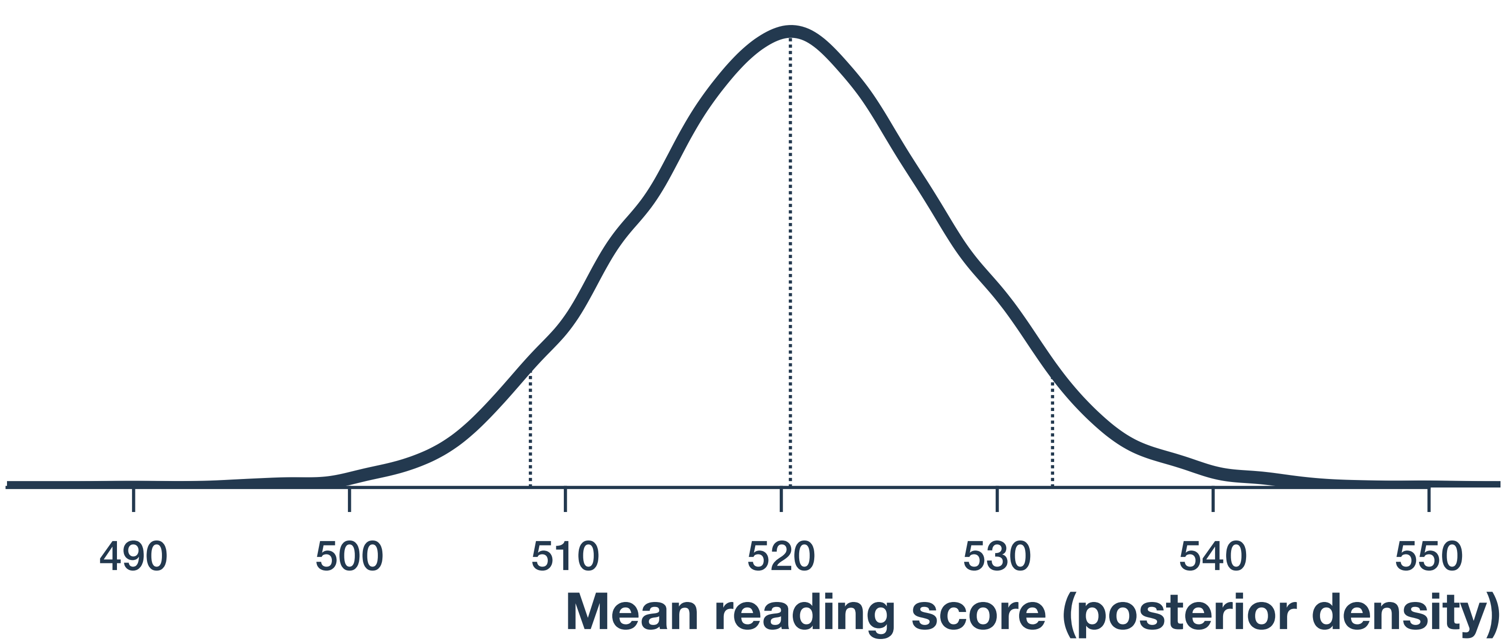
$$\mu_i = \alpha$$

Single average score α for all students

$$\alpha \sim \text{Norm}(500, 100)$$

$$\sigma \sim \text{Unif}(0, 100)$$

	Mean	*90% credible interval*	
α	520.54	508.46	532.56
σ	61.02	52.49	69.55



Incorporating categories

Indicator variables

$$Y_i \sim \text{Norm}(\mu_i, \sigma)$$

$$\mu_i = \alpha + \beta X_i$$

$$E(Y_i | X_i = 0) = \alpha$$

$$E(Y_i | X_i = 1) = \alpha + \beta$$

Pick a reference category and construct indicator variables for all other categories.

Intercept α captures reference mean, other means measured as difference from reference.

Can be fit with OLS in standard matrix specification of linear regression.

Fixed effects

$$Y_i \sim \text{Norm}(\mu_i, \sigma)$$

$$\mu_i = \begin{cases} \alpha_0 & \text{if } X_i = 0 \\ \alpha_1 & \text{if } X_i = 1 \end{cases}$$

$$E(Y_i | X_i = 0) = \alpha_0$$

$$E(Y_i | X_i = 1) = \alpha_1$$

Omit global mean α and give each category its own mean α_k.

Numerically identical to indicator variables, but harder to specify statistically.

No pooling

Standard deviation σ explains variation around each α_k

Strategy 2 (no pooling)

Include a separate parameter for each classroom's average score (fixed effects).

$$S_{ik} \sim \mathrm{Norm}(\mu_{ik}, \sigma)$$

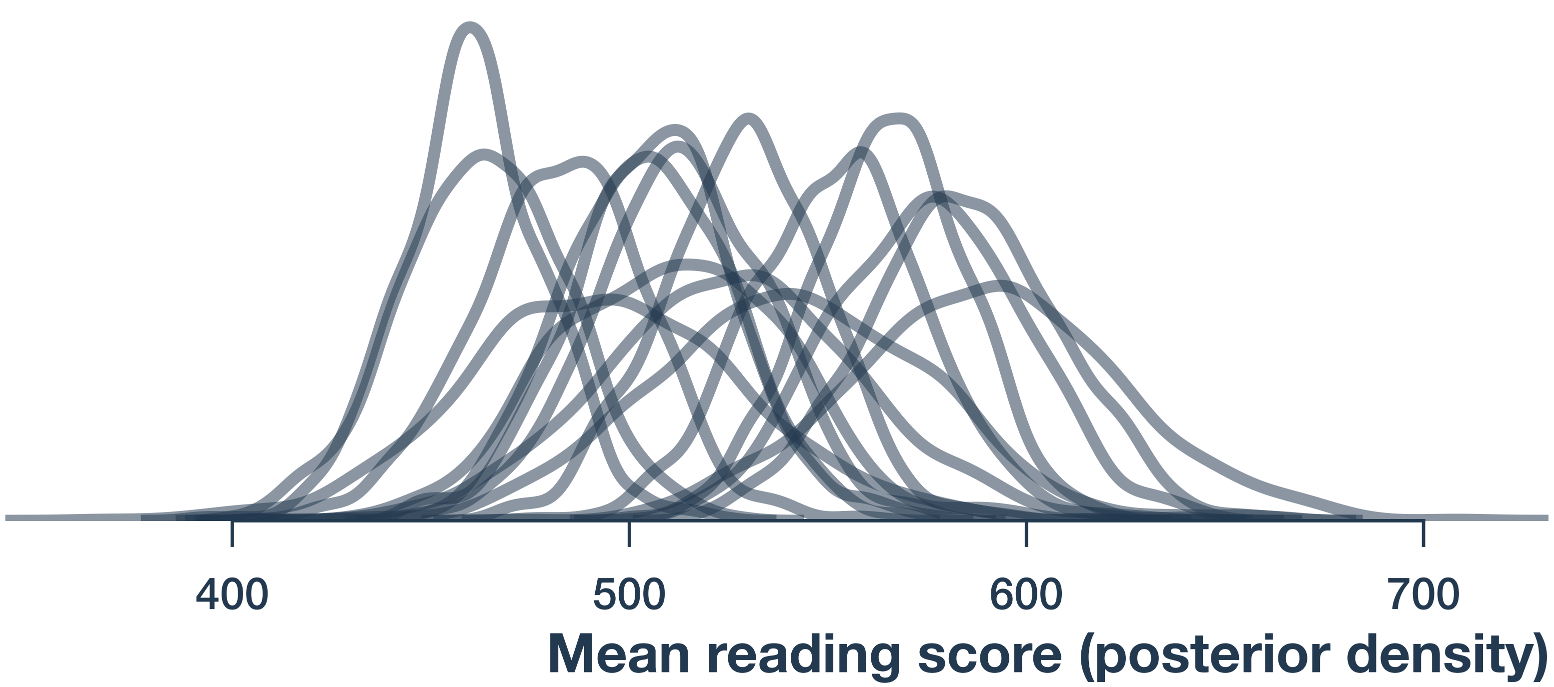
$$\mu_{ik} = \alpha_k$$

Each classroom k has its own average score α_k

$$\alpha_k \sim \mathrm{Norm}(500, 100)$$

$$\sigma \sim \mathrm{Unif}(0, 100)$$

Variability in α_k accounts for some inter-student variation



Partial pooling

Strategy 3 (partial pooling)

Include a separate parameter for each classroom's average score, but model those averages as random draws from a normal distribution.

$$S_{ik} \sim \text{Norm}(\mu_{ik}, \sigma)$$

$$\mu_{ik} = \alpha_k$$

Each classroom k still has its own average score α_k

$$\alpha_k \sim \text{Norm}(\gamma, \eta)$$

The prior distribution for classroom averages is estimated from the data

$$\sigma \sim \text{Unif}(0, 100)$$

$$\gamma \sim \text{Norm}(500, 100)$$

$$\eta \sim \text{Unif}(0, 100)$$

No pooling versus partial pooling

No pooling

$$S_{ik} \sim \text{Norm}(\mu_{ik}, \sigma)$$

$$\mu_{ik} = \alpha_k$$

$$\alpha_k \sim \text{Norm}(500, 100)$$

$$\sigma \sim \text{Unif}(0, 100)$$

Partial pooling

$$S_{ik} \sim \text{Norm}(\mu_{ik}, \sigma)$$

$$\mu_{ik} = \alpha_k$$

$$\alpha_k \sim \text{Norm}(\gamma, \eta)$$

Parameters γ and η describe 'typical' classrooms, allowing information to be shared among all of the α_k estimates.

$$\sigma \sim \text{Unif}(0, 100)$$

σ measures variability within each classroom

$$\gamma \sim \text{Norm}(500, 100)$$

$$\eta \sim \text{Unif}(0, 100)$$

η measures variability between all classrooms

Partial pooling

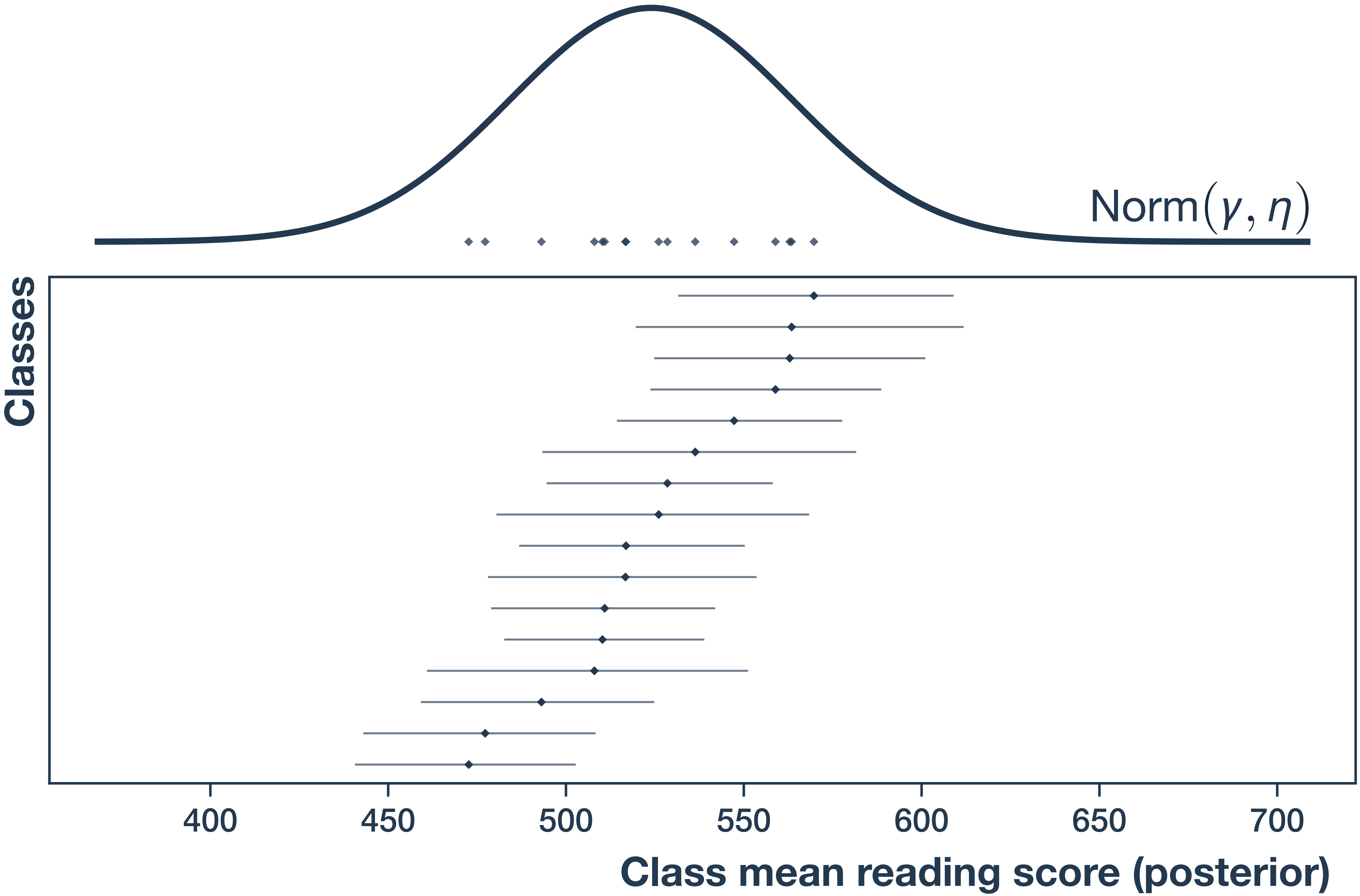
$$S_{ik} \sim \text{Norm}(\mu_{ik}, \sigma)$$
$$\mu_{ik} = \alpha_k$$

$$\alpha_k \sim \text{Norm}(\gamma, \eta)$$

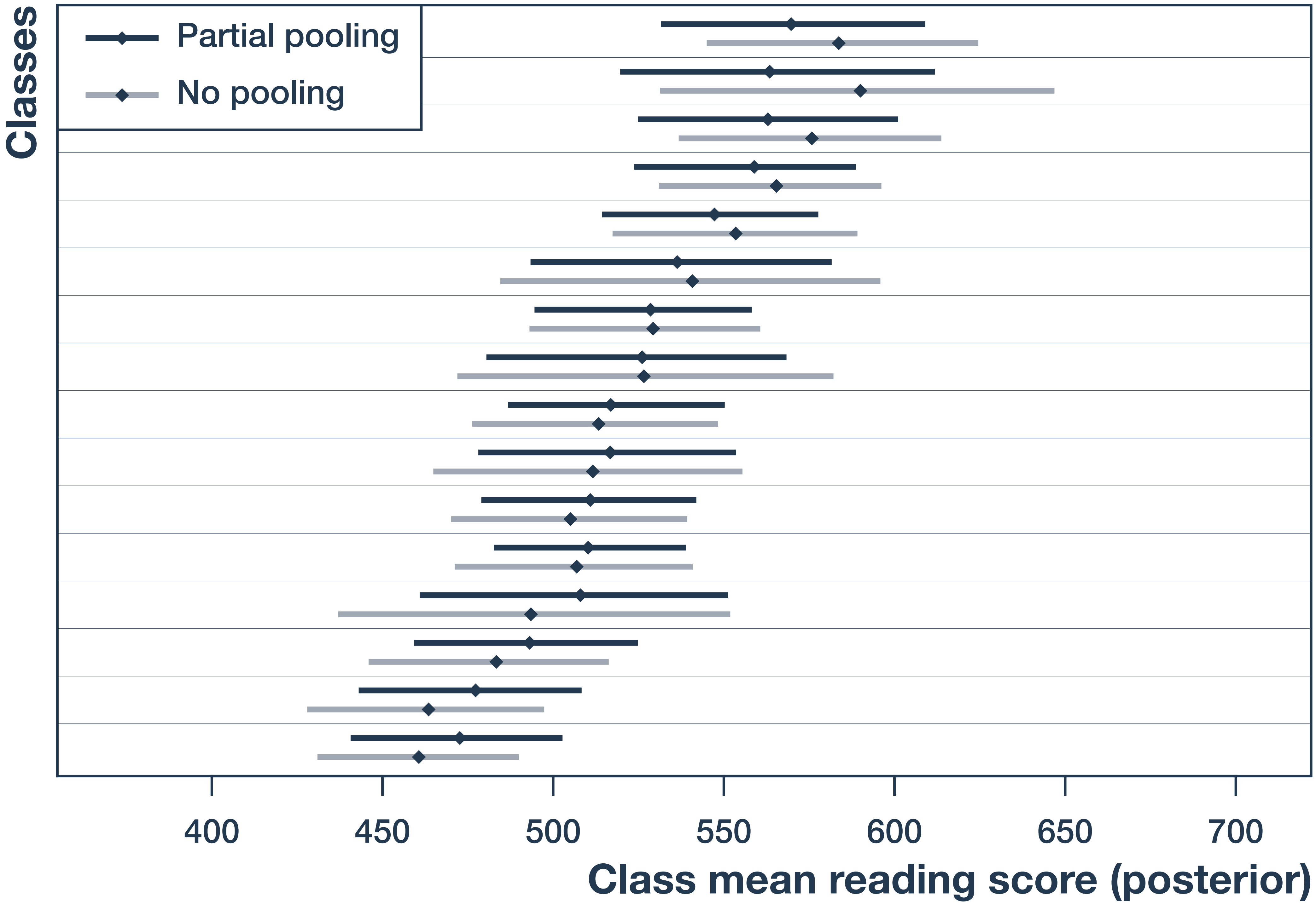
$$\sigma \sim \text{Unif}(0, 100)$$
$$\gamma \sim \text{Norm}(500, 100)$$
$$\eta \sim \text{Unif}(0, 100)$$

	Mean		***90% credible interval***
$\boldsymbol{\sigma}$	50.80	42.73	58.21
$\boldsymbol{\gamma}$	523.95	503.52	545.22
$\boldsymbol{\eta}$	39.78	22.90	59.29
$\boldsymbol{\alpha_1}$	492.93	461.80	526.99
$\boldsymbol{\alpha_2}$	472.53	441.40	503.23
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\boldsymbol{\alpha_{16}}$	525.66	481.70	569.47

Partial pooling



Partial pooling



Summary

Complete pooling

- Disregards nested levels
- Pools all data into same group
- Precise estimate of mean
- **Underfit:** Errs systematically in prediction

WAIC: 767.0

Eff. Param: 1.7

Partial pooling (random effects)

- Group-level estimates in population context
- Mutual distribution allows information sharing
- Small groups take cues from the rest of the groups
- **"Just right" fit:** Optimal balance of information pooling

WAIC: 751.0

Eff. Param: 12.5

No pooling (fixed effects)

- Independent estimate for each group
- No information shared between groups
- Imprecise estimates for smaller groups
- **Overfit:** Does poorly in out-of-sample prediction

WAIC: 754.4

Eff. Param: 15.6