Agenda

1. **Indexing in models using R (from Tuesday)**
2. **Adding student-level predictors**
3. **Adding class-level predictors**
4. **Random intercepts in R**

Intercept-only model

Partial pooling (random effects)

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

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

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

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

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

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

Accounting for student race

Number of participants by race/ethnicity

White	4222
Black	2126
Asian	19
Hispanic	9
Native American	4
Other	11
Total	**6391**

Number of classes by experimental condition

Small	122
Large	114
Large + Aide	98
Total	**334**

Accounting for student race

But first, some notation

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

$$\mu_{ik} = \boxed{\beta_{0k}} + \boxed{\beta_1 B_i}$$

β_0 is the intercept.

Coefficient β_1 measures difference in test scores for Black and white students.

$$\beta_{0k} \sim \text{Norm}(\boxed{\gamma_0, \phi_0})$$

Subscripts on γ and ϕ to remind us which coefficient they refer to.

Accounting for student race

But first, some notation

$$S_{ik} \sim \text{Norm}(\mu_{ik}, \sigma)$$
$$\mu_{ik} = \beta_{0k} + \beta_1 B_i$$

$$\boxed{\beta_{0k} \sim \text{Norm}(\gamma_0, \phi_0)}$$

$$S_{ik} \sim \text{Norm}(\mu_{ik}, \sigma)$$
$$\mu_{ik} = \beta_{0k} + \beta_1 B_i$$

$$\boxed{\begin{aligned} \beta_{0k} &= \gamma_0 + \eta_{0k} \\ \eta_{0k} &\sim \text{Norm}(0, \phi_0) \end{aligned}}$$

Equivalent ways to describe the same distribution for β_{0k}.

Accounting for student race

Full model

$$
\begin{aligned}
S_{ik} &\sim \text{Norm}(\mu_{ik}, \sigma) \\
\mu_{ik} &= \beta_{0k} + \beta_1 Black_i + \beta_2 Asian_i \\
&\quad + \beta_3 Hispanic_i + \beta_4 NativeAmerican_i \\
&\quad + \beta_5 Other_i \\
\\
\beta_{0k} &= \gamma_0 + \eta_{0k} \\
\eta_{0k} &\sim \text{Norm}(0, \phi_0) \\
\\
\beta_1, \ldots, \beta_5 &\sim \text{Norm}(0, 50) \\
\sigma &\sim \text{Unif}(0, 100) \\
\gamma_0 &\sim \text{Norm}(500, 100) \\
\phi_0 &\sim \text{Unif}(0, 100)
\end{aligned}
$$

Accounting for student race

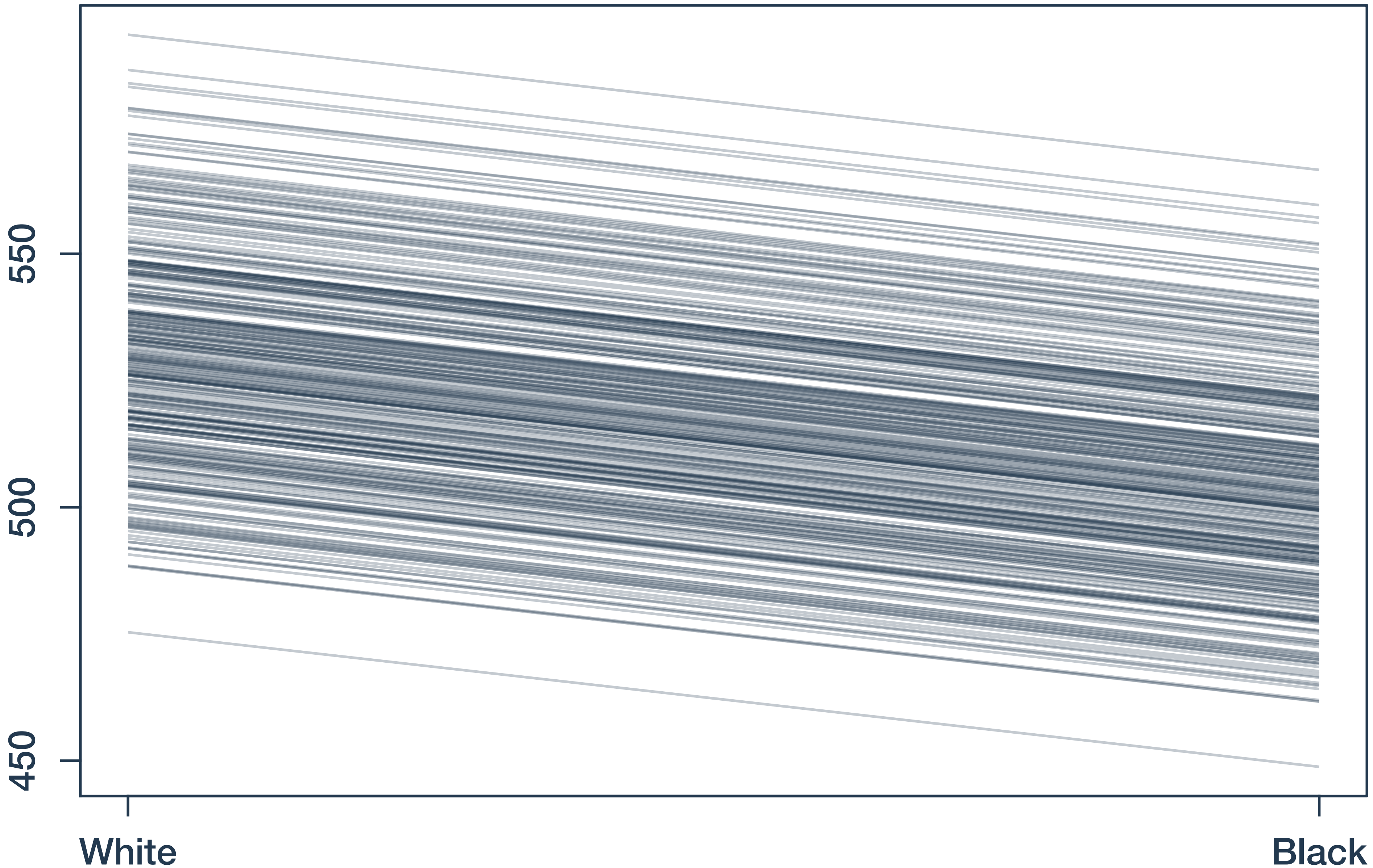
$$S_{ik} \sim \text{Norm}(\mu_{ik}, \sigma)$$
$$\mu_{ik} = \beta_{0k} + \beta_1 Black_i + \beta_2 Asian_i + \beta_3 Hispanic_i + \beta_4 NativeAmerican_i + \beta_5 Other_i$$

$$\beta_{0k} = \gamma_0 + \eta_{0k}$$
$$\eta_{0k} \sim \text{Norm}(0, \phi_0)$$

$$\beta_1, \ldots, \beta_5 \sim \text{Norm}(0, 50)$$
$$\sigma \sim \text{Unif}(0, 100)$$
$$\gamma_0 \sim \text{Norm}(500, 100)$$
$$\phi_0 \sim \text{Unif}(0, 100)$$

	Mean	90% credible interval	
γ_0	530.65	528.21	533.29
β_1	-26.78	-29.52	-23.17
β_2	6.70	-9.03	25.07
β_3	16.39	-10.53	41.08
β_4	-13.10	-49.60	20.18
β_5	14.81	-9.17	33.92
σ	47.03	46.37	47.75
ϕ_0	23.98	22.01	25.83
β_{1k}	$\vdots$	$\vdots$	$\vdots$

Accounting for student race



$$\gamma_0 = 530.65$$

$$\beta_1 = -26.78$$

Comparing to pooled model

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

$$\mu_{ik} = \beta_0 + \beta_1 Black_i + \beta_2 Asian_i + \beta_3 Hispanic_i + \beta_4 NativeAmerican_i + \beta_5 Other_i$$

$$\beta_0, \ldots, \beta_5 \sim \text{Norm}(0, 50)$$

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

	Mean	*90% credible interval*	
$\boldsymbol{\beta_0}$	533.00	531.59	534.29
$\boldsymbol{\beta_1}$	-36.86	-39.11	-34.71
$\boldsymbol{\beta_2}$	12.61	-7.84	29.98
$\boldsymbol{\beta_3}$	7.61	-21.58	31.06
$\boldsymbol{\beta_4}$	-28.47	-66.14	8.46
$\boldsymbol{\beta_5}$	13.83	-10.70	35.93
$\boldsymbol{\sigma}$	52.37	51.49	53.07

Comparing to pooled model

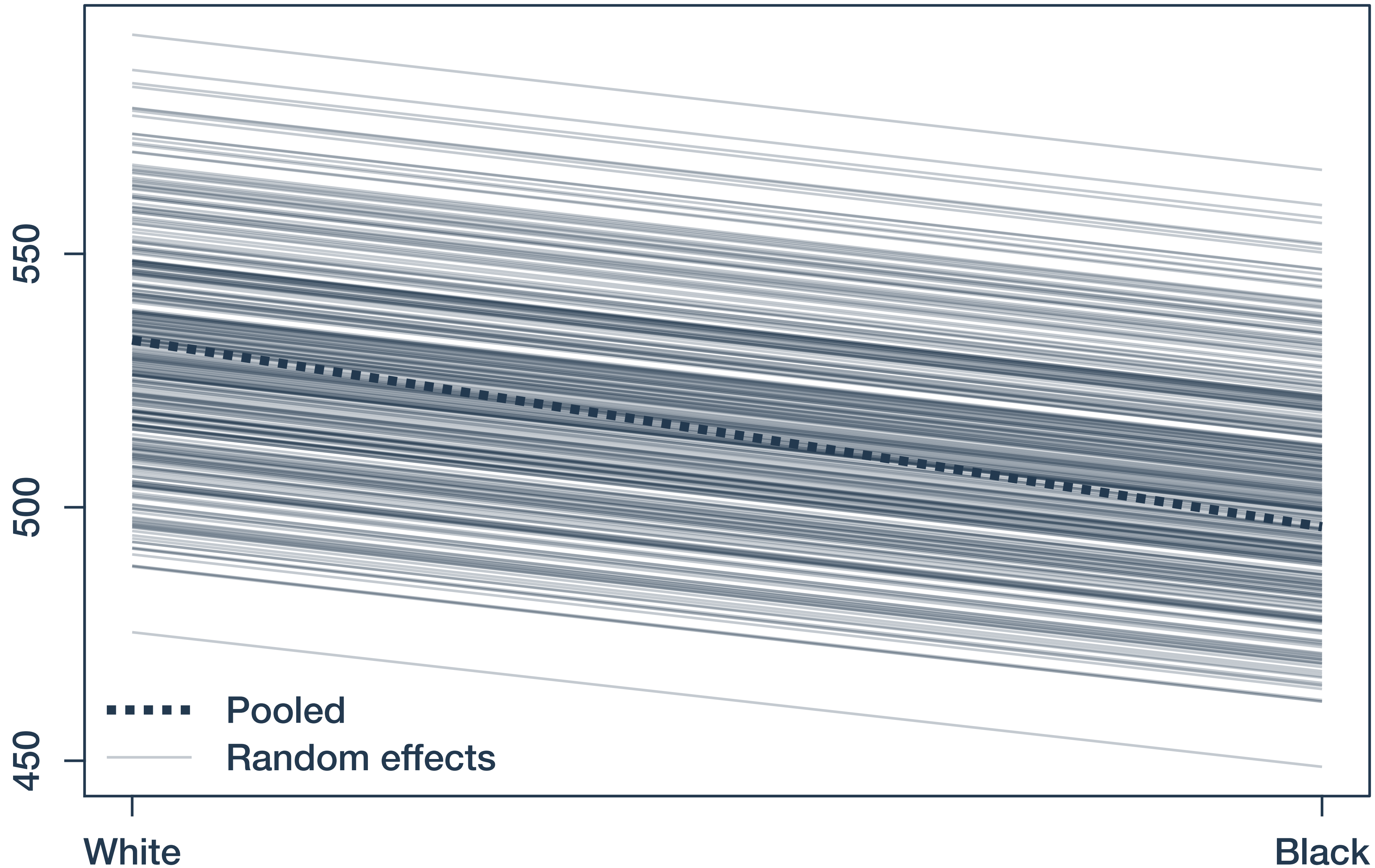
Random effects

	Mean	90% credible interval	
γ_0	530.65	528.21	533.29
β_1	-26.78	-29.52	-23.17
β_2	6.70	-9.03	25.07
β_3	16.39	-10.53	41.08
β_4	-13.10	-49.60	20.18
β_5	14.81	-9.17	33.92
σ	47.03	46.37	47.75
ϕ_0	23.98	22.01	25.83
β_{1k}	⋮	⋮	⋮

Pooled

	Mean	90% credible interval	
β_0	533.00	531.59	534.29
β_1	-36.86	-39.11	-34.71
β_2	12.61	-7.84	29.98
β_3	7.61	-21.58	31.06
β_4	-28.47	-66.14	8.46
β_5	13.83	-10.70	35.93
σ	52.37	51.49	53.07

Accounting for student race



Random effects | $\gamma_0 = 530.65$, $\beta_1 = -26.78$

Pooled | $\beta_0 = 533.00$, $\beta_1 = -36.86$

Accounting for class size

Number of participants by race/ethnicity

White	4222
Black	2126
Asian	19
Hispanic	9
Native American	4
Other	11
Total	**6391**

Number of classes by experimental condition

Small	122
Large	114
Large + Aide	98
Total	**334**

Accounting for class size

Full model

$$
\begin{aligned}
S_{ik} &\sim \text{Norm}(\mu_{ik}, \sigma) \\
\mu_{ik} &= \beta_{0k} + \beta_1 Black_i + \beta_2 Asian_i \\
&\quad + \beta_3 Hispanic_i + \beta_4 NativeAmerican_i \\
&\quad + \beta_5 Other_i \\
\\
\beta_{0k} &= \gamma_0 + \boxed{\gamma_1 Small_k} + \eta_{0k} \\
\eta_{0k} &\sim \text{Norm}(0, \phi_0) \\
\\
\beta_1, \ldots, \beta_5 &\sim \text{Norm}(0, 50) \\
\sigma &\sim \text{Unif}(0, 100) \\
\gamma_0 &\sim \text{Norm}(500, 100) \\
\gamma_1 &\sim \text{Norm}(0, 50) \\
\phi_0 &\sim \text{Unif}(0, 100)
\end{aligned}
$$

Coefficient γ_1 measures average test score for classes in the "small" experimental condition.

Accounting for class size

Note For computational efficiency and other pragmatic reasons, second-level terms like γ_1 sometimes need to be implemented as first-level components.

In random-intercept models, this is a trivial matter, but it gets more complex with models we will cover soon.

$$
\begin{aligned}
S_{ik} &\sim \text{Norm}(\mu_{ik}, \sigma) \\
\mu_{ik} &= \beta_{0k} + \beta_1 Black_i + \beta_2 Asian_i \\
&\quad + \beta_3 Hispanic_i + \beta_4 NativeAmerican_i \\
&\quad + \beta_5 Other_i
\end{aligned}
$$

$$
\begin{aligned}
\beta_{0k} &= \gamma_0 + \boxed{\gamma_1 Small_k} + \eta_{0k} \\
\eta_{0k} &\sim \text{Norm}(0, \phi_0)
\end{aligned}
$$

Think about γ_1 as a second-level variable.

Mathematically, γ_1 can be included at the first level.

$$
\begin{aligned}
S_{ik} &\sim \text{Norm}(\mu_{ik}, \sigma) \\
\mu_{ik} &= \beta_{0k} + \boxed{\gamma_1 Small_k} + \beta_1 Black_i + \beta_2 Asian_i \\
&\quad + \beta_3 Hispanic_i + \beta_4 NativeAmerican_i \\
&\quad + \beta_5 Other_i
\end{aligned}
$$

$$
\begin{aligned}
\beta_{0k} &= \gamma_0 + \eta_{0k} \\
\eta_{0k} &\sim \text{Norm}(0, \phi_0)
\end{aligned}
$$

Accounting for class size

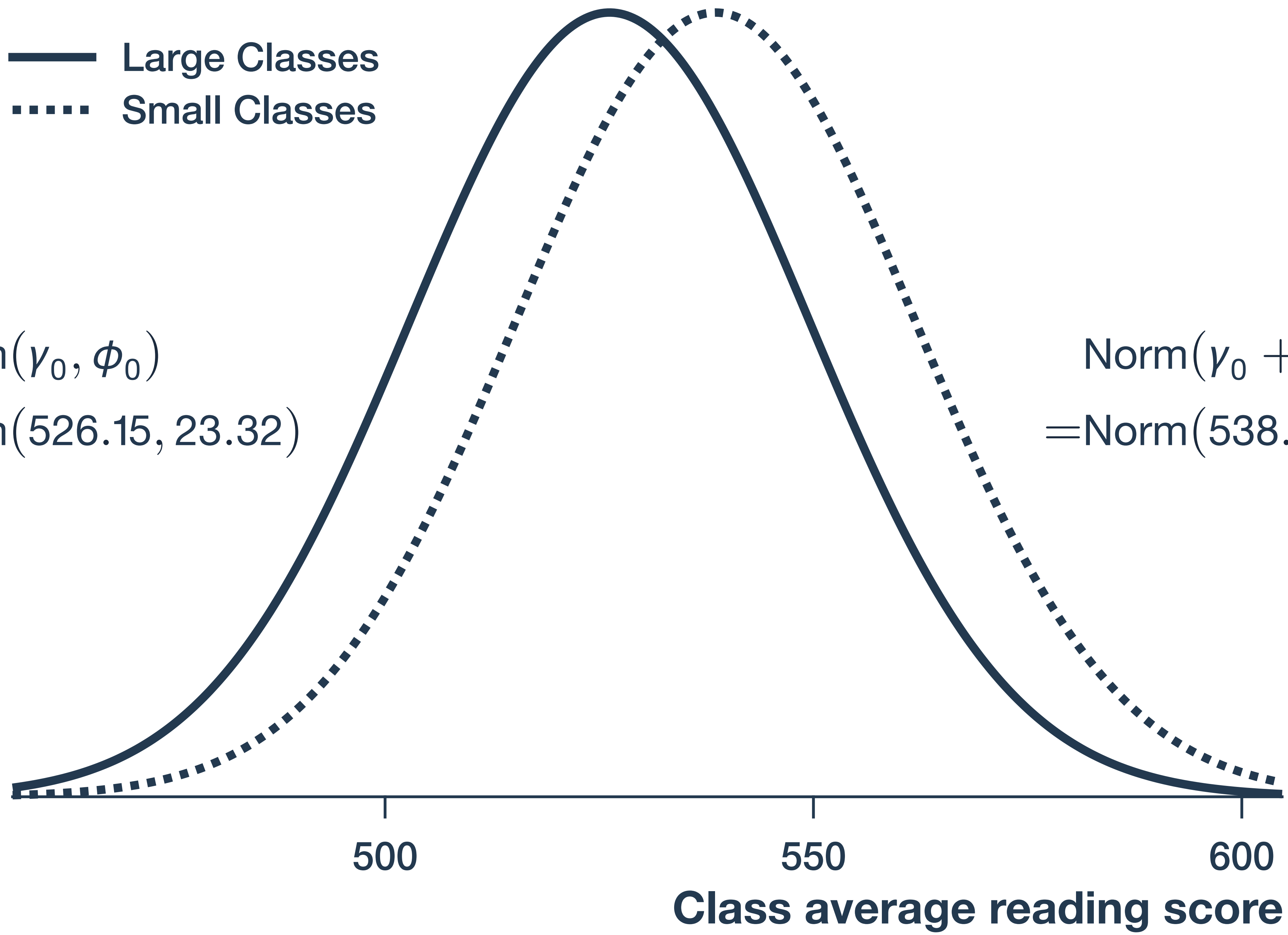
$$S_{ik} \sim \text{Norm}(\mu_{ik}, \sigma)$$

$$\mu_{ik} = \beta_{0k} + \beta_1 Black_i + \beta_2 Asian_i$$
$$+ \beta_3 Hispanic_i + \beta_4 NativeAmerican_i$$
$$+ \beta_5 Other_i$$

$$\beta_{0k} = \gamma_0 + \gamma_1 Small_k + \eta_{0k}$$
$$\eta_{0k} \sim \text{Norm}(0, \phi_0)$$

	Mean	***90% credible interval***	
γ_0	526.15	523.34	529.04
γ_1	12.37	8.26	16.82
β_1	-26.70	-29.88	-23.51
β_2	7.34	-8.59	24.19
β_3	16.37	-6.66	42.92
β_4	-14.17	-48.66	23.74
β_5	14.83	-8.74	35.25
σ	47.03	46.33	47.69
ϕ_0	23.32	21.32	25.31
β_{1k}	$\vdots$	$\vdots$	$\vdots$

Accounting for class size



$\text{Norm}(\gamma_0, \phi_0)$
$=\text{Norm}(526.15, 23.32)$

$\text{Norm}(\gamma_0 + \gamma_1, \phi_0)$
$=\text{Norm}(538.52, 23.32)$