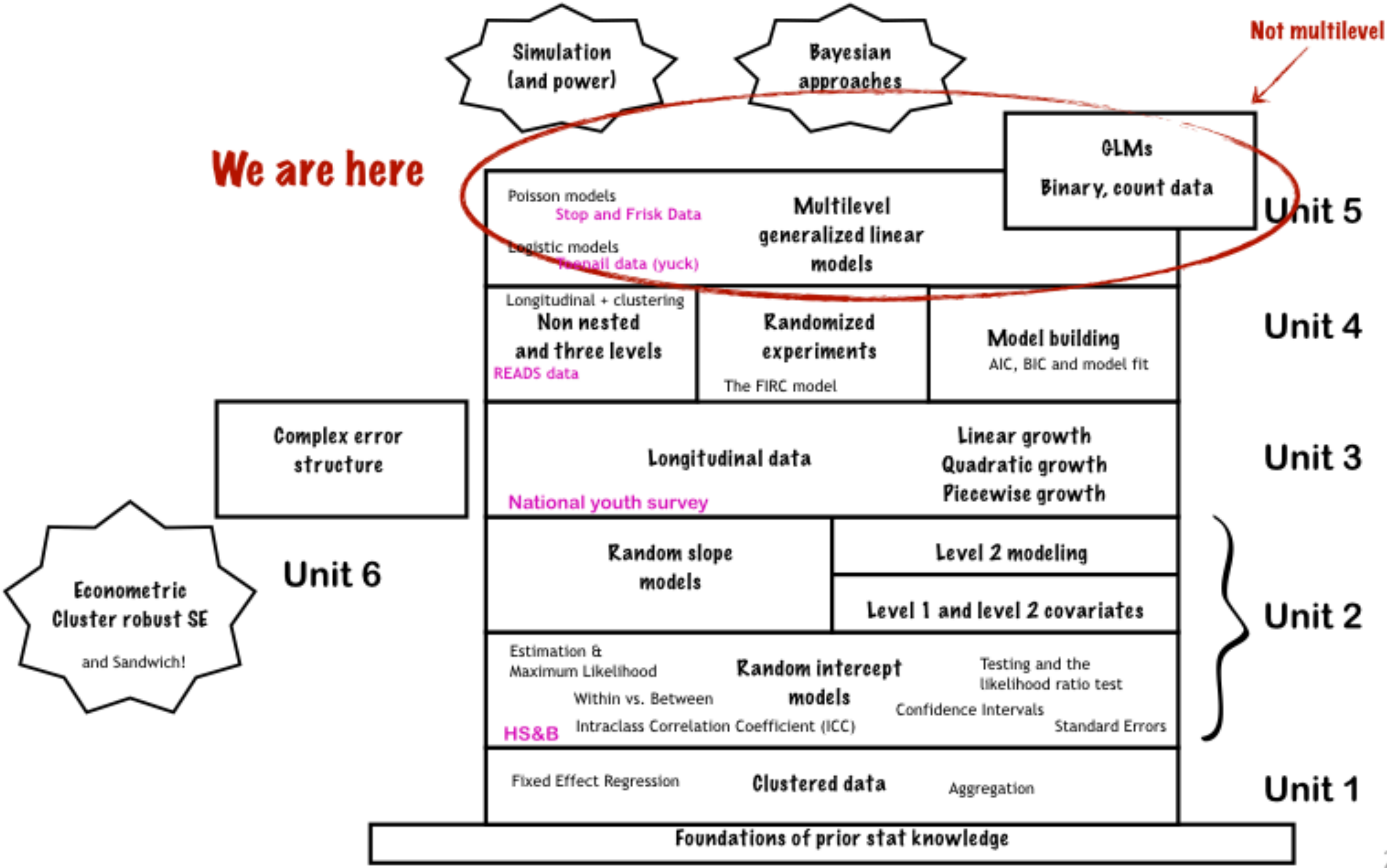
S-043/Stat-151
Analysis for Clustered and Longitudinal Data
(Multilevel & Longitudinal Models)

Multilevel Logistic Regression

R&B Chapter 10

Also Chapter 14 (mainly 14.2 and 14.3) of G&H

Unit 5: Where are we?



Today's Goals

Quickly review some highlights of Poisson regression

Move on to Logistic regression… but with multiple levels.

- ★ We will look at a new dataset about kids being retained a year
- ★ We will see how we can describe variation in retention rates across schools
- ★ We will also see how we can model individual differences (e.g., between girls and boys)

By the end you will be able to

- ★ Fit a multilevel generalized model
- ★ Understand that the estimated coefficients need to be transformed to make them interpretable
- ★ Use predict() to make the results of a logistic model more interpretable.

Back to the Poisson GLM

Super Cinder (aka Cinder) wants us to (1) look awesome for Halloween, and (2) revisit last lecture for a moment…

"I love wearing a costume."

The anatomy of GLMs via Poisson

Our Covariates

$$X_i = (1, Eth_i, Pcnt_i)$$

Each observation is a bunch of covariates plus an observed outcome (a count) plus an exposure.

Our Linear Predictor

$$\eta_i = X_i'\beta$$

"eta"

Our rate is a transform of our linear predictor

$$\theta_i = g(\eta_i) = exp(X_i'\beta)$$

This is our inverse "link function"

exposure X rate gives how many events we *expect* to see

$$\hat{Y}_i = u_i\theta_i = exp(log(u_i) + X_i'\beta)$$

We can shove our exposure into the linear predictor, making it look like a covariate.

We **actually** observe a Poisson draw around what we expected

$$Y_i = Poisson(u_i\theta_i)$$

$$E[Y_i] = \hat{Y}_i, sd(Y_i) = \sqrt{\hat{Y}_i}$$

This step is where the randomness is!

(Our residual is then the difference)

$$resid_i = Y_i - \hat{Y}_i$$

Our fit model revisited

```
> fit.4 <- glm (stops ~ factor(eth) + factor(precinct),
          family=quasipoisson,
          offset=log(past.arrests), data=stops )
> display(fit.4)
```

Just think of quasipoisson as Poisson, fixed.

```
                    coef.est coef.se
(Intercept)          -1.38     0.24
factor(eth)2          0.01     0.03
factor(eth)3         -0.42     0.04
factor(precinct)2    -0.15     0.35
factor(precinct)3     0.56     0.27
...
factor(precinct)74    1.15     0.27
factor(precinct)75    1.57     0.35
---
  n = 225, k = 77
  residual deviance = 3427.1, null deviance = 46120.3
(difference = 42693.1)
  overdispersion parameter = 21.9
```

We can believe these standard errors now.

(And they are substantially larger.)

Pop! Quiz

$$\text{logit} = \text{log odds} = \log\left(\frac{p}{1-p}\right) = \alpha$$

In reverse: $$logit^{-1}(\alpha) = \frac{1}{1+exp(-\alpha)} = \frac{exp(\alpha)}{1+exp(\alpha)}$$

If the log odds is 3, what is the chance of winning?

Multilevel & Logistic, here we come!

Today's Data: Thailand Retention Data

From R&B Chapter 10

Students in schools, as per usual.

Students are either retained or not (*binary outcome*)

Variables:

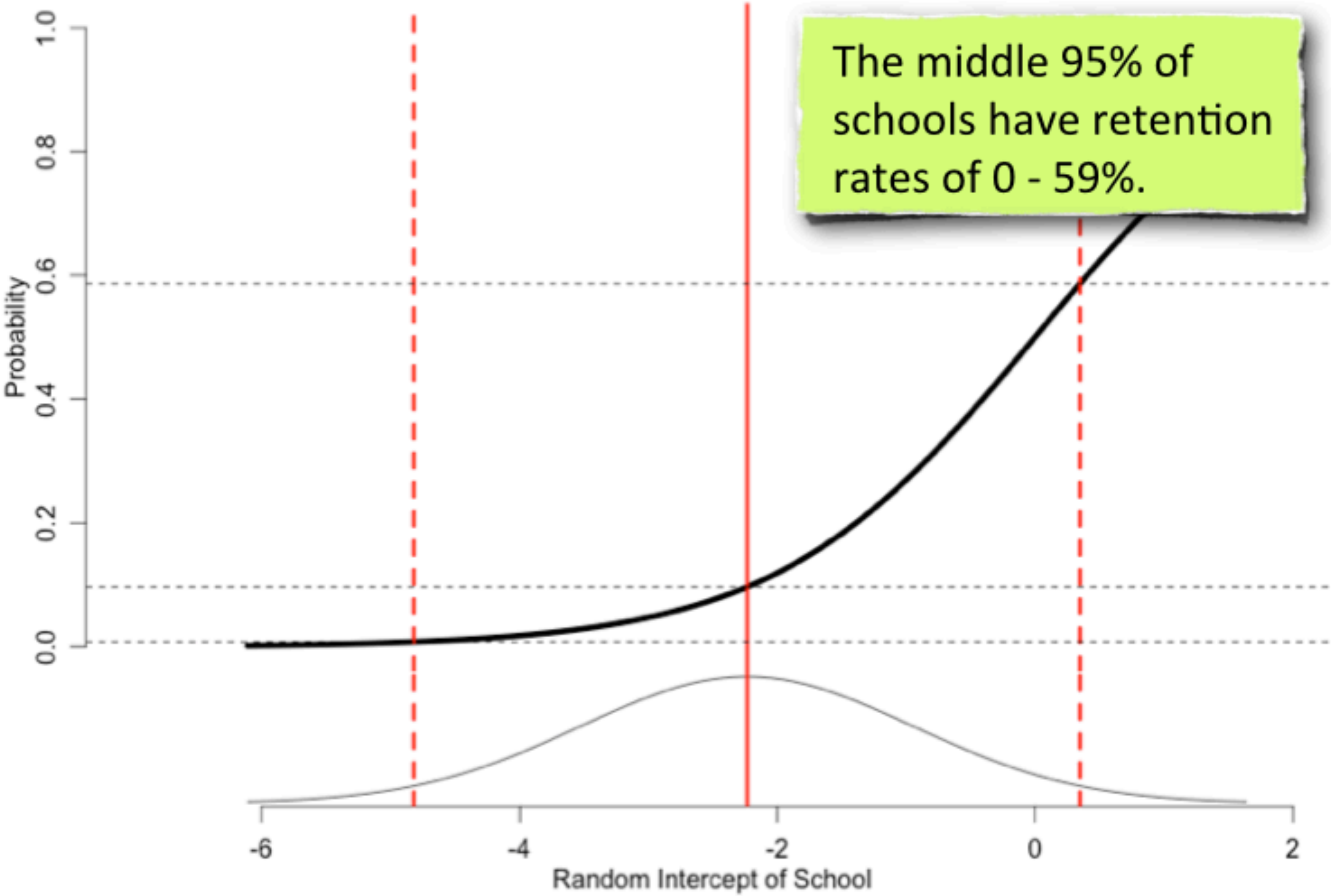
- ★Retained (1067 of 7516 retained)   This is our outcome.
- ★Gender
- ★Pre-primary school experience. (Whether they have prior experience with school.)
- ★Mean SES of school

Goal: (1) describe variation in retention rates across schools, and

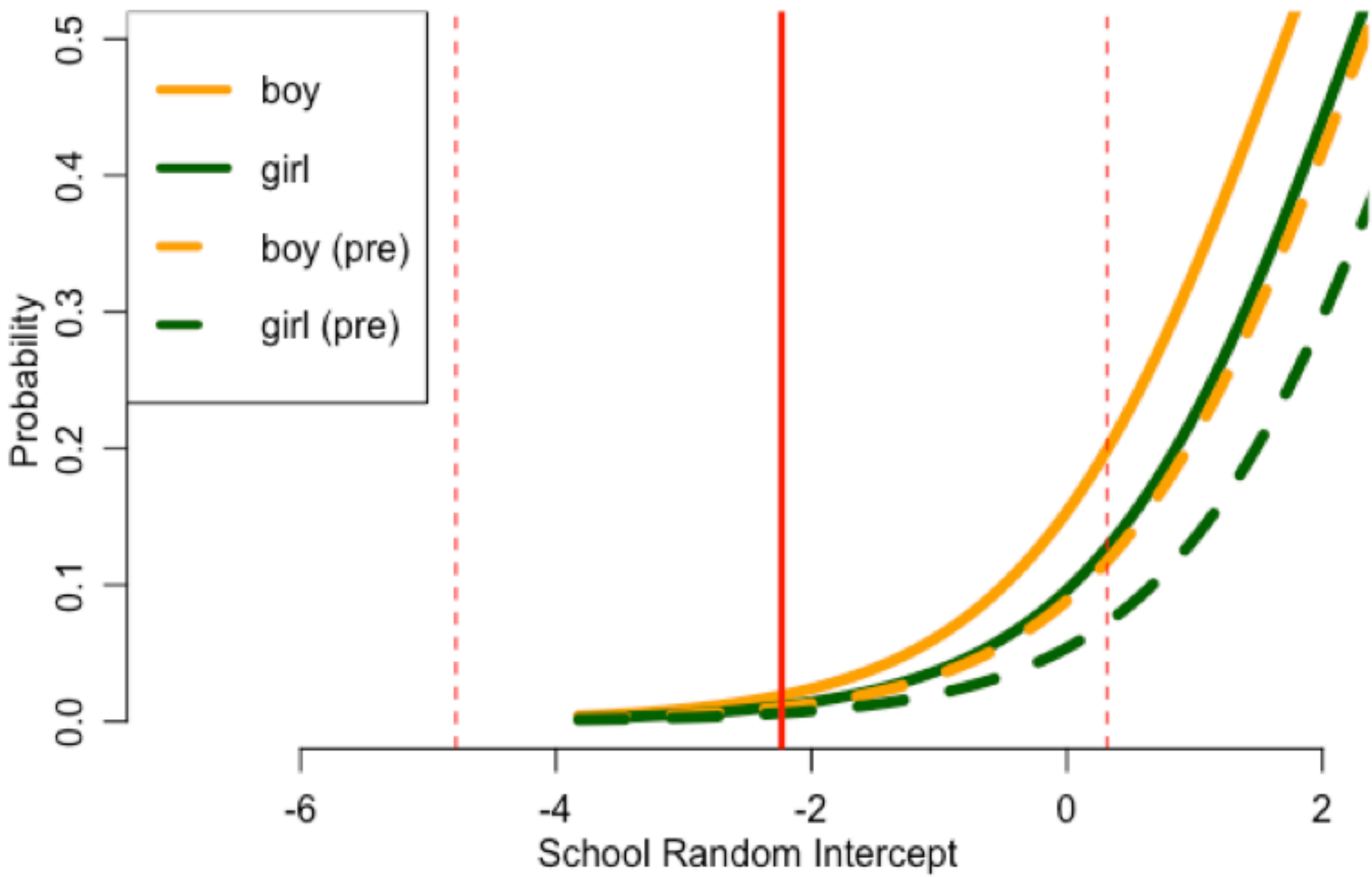
(2) describe how students with different prior experiences differ.

(3) see if school mean SES is associated with rate of retention.

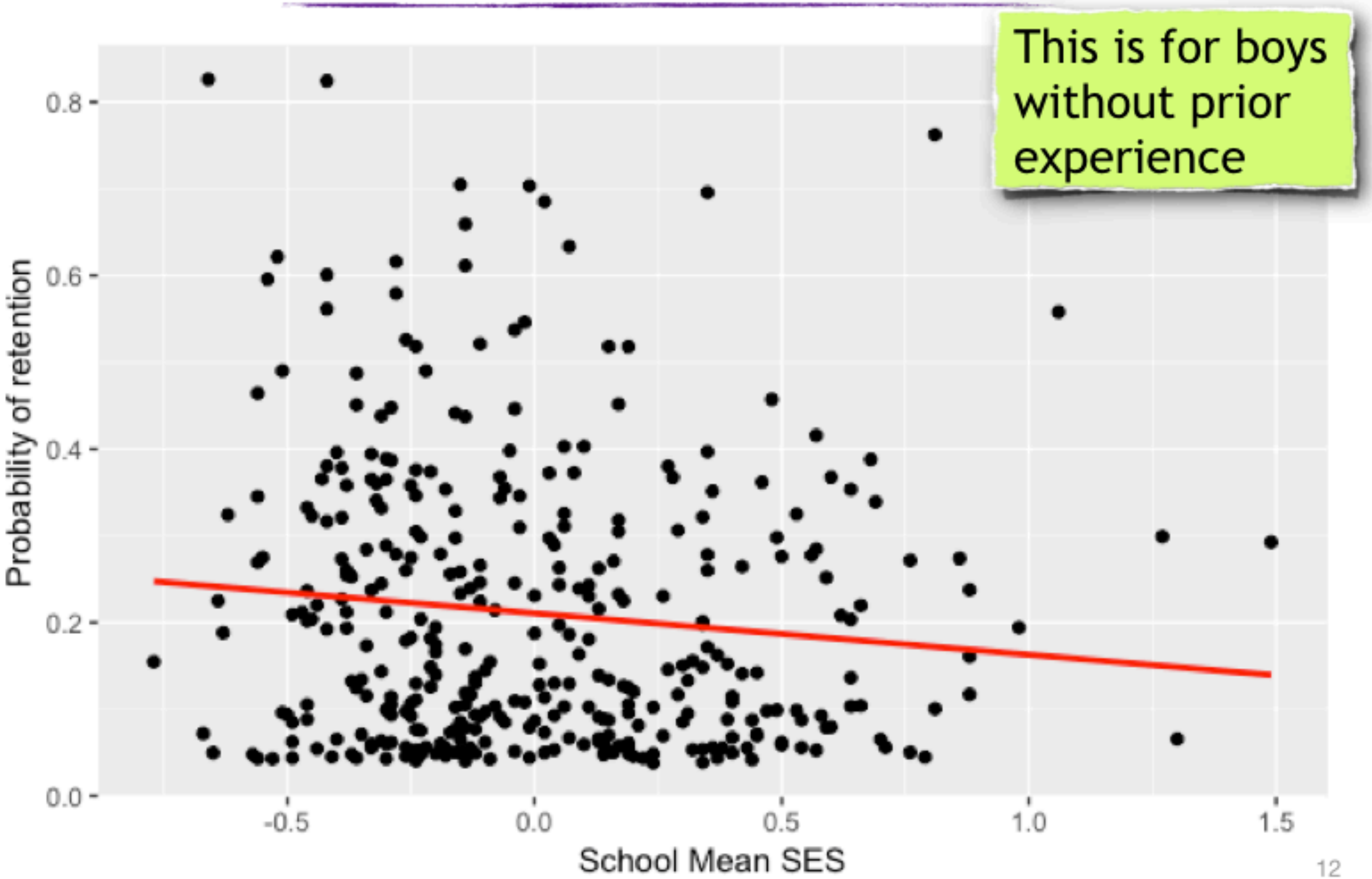
Distribution of schools and their retention rates



Retention probabilities for four groups



Higher Mean SES is associated with lower retention



Components of a Generalized Linear Model

$$Y_{ij} \sim F(\mu_{ij}, \psi)$$
$$E[Y_{ij}] = \mu_{ij}$$

Each observation has some predicted mean (the μ_{ij}, often called $\hat{Y}_{ij}$)

We specify some **sampling distribution** for our actual outcomes Y_{ij} around the μ_{ij}

F is a distribution (e.g., Normal, Poisson, Binomial, etc., that takes the mean and possibly other parameters such as a variance (the ψ).

Component: Level-1 Link Function

"Eta" $\eta_{ij} = g(\mu_{ij}) \text{ with } Y_{ij} \sim F(\mu_{ij}, \psi)$

Eta is our **linear predictor**

We tie our linear predictor to the *actual* prediction with our **link function**

So far, we don't have any multilevel anything going on...

We have a linear model to predict our eta for each unit. Then we transform eta to get our actual prediction mu.

Component:
Level 1 Structural Model

Note how eta is our linear predictor—and it looks just like classic OLS

$$\eta_{ij} = \beta_{0j} + \beta_{1j} X_{1ij} + \beta_{2j} X_{2ij}$$

Note lack of residual term. (Where did it go?)

Now we see grouping structure: Each coefficient could vary by group, as shown by the *j* subscripts.

To denote pooling across all groups, omit the *j*.

$$\eta_{ij} = \beta_{0j} + \beta_1 X_{1ij} + \beta_2 X_{2ij}$$

Q: What kind of multilevel model is this?

Component: The Level 2 Model

Each level 1 coefficient that varies by group has a level-2 equation:

$$\beta_{qj} = \gamma_{q0} + \gamma_{q1} W_{1j} + \ldots + u_{qj}$$

Our level-2 is (usually) in our classic OLS framework (i.e., this is exactly the same as we have seen before).

Each coefficient in our level-1 model is modeled as

★ a linear combination of level-2 predictors with level-2 coefficients.

★ a random level-2 noise term (the random effect)

Logistic Binomial Model

If we have m_{ij} observations for unit i in group j, and Y_{ij} is the number of successes, we have a random intercept model of

$$Y_{ij} \sim Binom(m_{ij}, \mu_{ij})$$

m_{ij} coin flips, each with μ_{ij} chance of heads

$$E[Y_{ij}] = m_{ij}\mu_{ij}$$

We have m_{ij} * μ_{ij} because m observations each with μ_{ij} chance
(The m_{ij} is like the exposure from the Poisson model.)

$$\mu_{ij} = g^{-1}(\eta_{ij}) = logit^{-1}(\eta_{ij}) = \frac{1}{1+\exp(-\eta_{ij})}$$

logistic link function

$$\eta_{ij} = \beta_{0j} + \beta_1 X_{ij}$$

linear predictor has an intercept varying by group

$$\left.\begin{aligned}\beta_{0j} &= \gamma_{00} + u_{0j}\\ u_{0j} &\sim N(0, \tau)\end{aligned}\right\}$$

Level 2 is same-old, same-old. Only level 1 has the links and fancy connections to outcomes

This is a random intercept model with a logistic model for level 1

The Simplest Model

```
> M0 = glmer( REPEAT ~ 1 + (1|SCHOOLID), data=dat,
+              family=binomial(link="logit") )
Warning messages:
1: In checkConv(attr(opt, "derivs"), opt$par, ctrl = control
$checkConv,  :
  Model failed to converge with max|grad| = 0.184618 (tol = 0.001,
component 1)
2: In checkConv(attr(opt, "derivs"), opt$par, ctrl = control
$checkConv,  :
  Model is nearly unidentifiable: very large eigenvalue
 - Rescale variables?
> display( M0 )
glmer(formula = REPEAT ~ 1 + (1 | SCHOOLID), data = dat, family =
binomial(link = "logit"))
coef.est  coef.se
   -2.23     0.00

Error terms:
 Groups   Name        Std.Dev.
 SCHOOLID (Intercept) 1.29
 Residual             1.00
---
number of obs: 7516, groups: SCHOOLID, 356
AIC = 5547, DIC = 4087
deviance = 4815.1
```

Ignore this.

We are not converging. We are also getting different answers from R&B. Not sure what is going on.

What does the intercept mean?

A “Typical” school has a random intercept of 0, meaning we have:

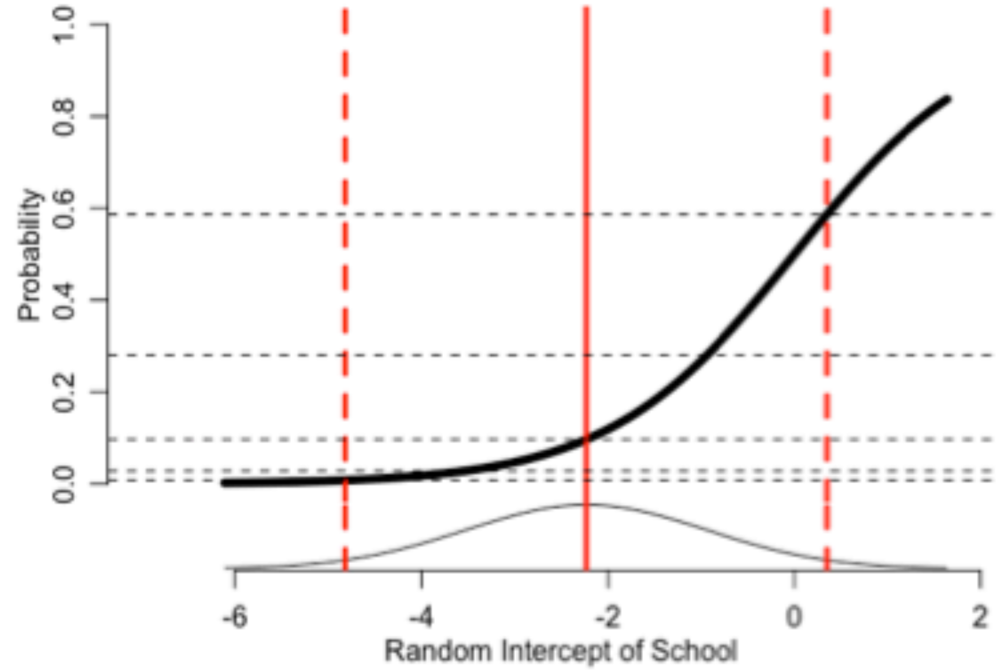
- ★ Expected log-odds of retention = -2.23
- ★ I.e., odds of exp(-2.23) = 0.109 or about 1 to 9
- ★ I.e., probability of 1/(1+exp(2.22)) = 0.097

Compare with overall rate of retention of 14%

KEY POINT:

“Typical” is not the same as population average due to non-linearity!

Assessing cross-school variation (Also, ICC? Not so useful.)



Variation is **immense.**

Some schools have retention rates of near 0% and some are above 50%!

What about ICC?

- ★ ICC is ratio of level-2 variance to total variance.
- ★ But different schools have different kids, and so different variances due to Binomial.
- ★ There is no natural ratio here.
- ★ Instead draw pictures and show people **how much schools vary.**

Mean SES (Centered)

```
> M1 = glmer( REPEAT ~ 1 + SEX + PPED + MSESC + (1|SCHOOLID),
+             data=dat,
+             family=binomial(link="logit") )

> display( M1 )
            coef.est coef.se
(Intercept) -2.24     0.11
SEXboy       0.53     0.08
PPEDyes     -0.63     0.10
MSESC       -0.30     0.21

Error terms:
 Groups   Name        Std.Dev.
 SCHOOLID (Intercept) 1.27
 Residual             1.00
---
number of obs: 7516, groups: SCHOOLID, 356
AIC = 5456.9, DIC = 4014.4
deviance = 4730.7
```

What is the probability of a boy with no prior experience in an average school being retained?

Our covariates:
Gender
pre-school experience
mean school SES

What levels?

Convert the log to odd multipliers

(And examining our impact of Mean SES)

```
> exp( fixef( M1 ) )
(Intercept)      SEXboy      PPEDyes      MSESC
      0.107       1.706        0.534      0.743
```

The odds of a boy being retained is 1.7 times those of a girl, all other things being equal (including school).

For schools, each unit increase in Mean SES corresponds to a reduction of x0.74

> sd(sdat$mean.ses)
[1] 0.381

so 1 SD above average gives:
> exp(0.381 * 0.743) = 0.89
I.e., schools 1 SD above average have 90% of the retention rate (on average)

Predicting individual retention

Two ways, same thing

```
> dat$eta.hat = predict( M1 )
> dat$eta.hat = predict( M1, type = "link" )
> sample( dat$eta.hat, 7 )
[1] -0.91311147 -3.98446224 -0.86143874 -0.05038648
-3.16325990 -2.55520422 -2.68551484
```

This is the linear predictor

```
> dat$mu.hat = predict( M1, type="response" )
> sample( dat$mu.hat, 7 )
[1] 0.11416657 0.45092846 0.46885009 0.02925308
0.15426398 0.20458749 0.10572866
```

These are probabilities!

We can predict our linear predictors for our individuals, or predict our outcome (response).

This is using the link function or not.

Recap of our research questions

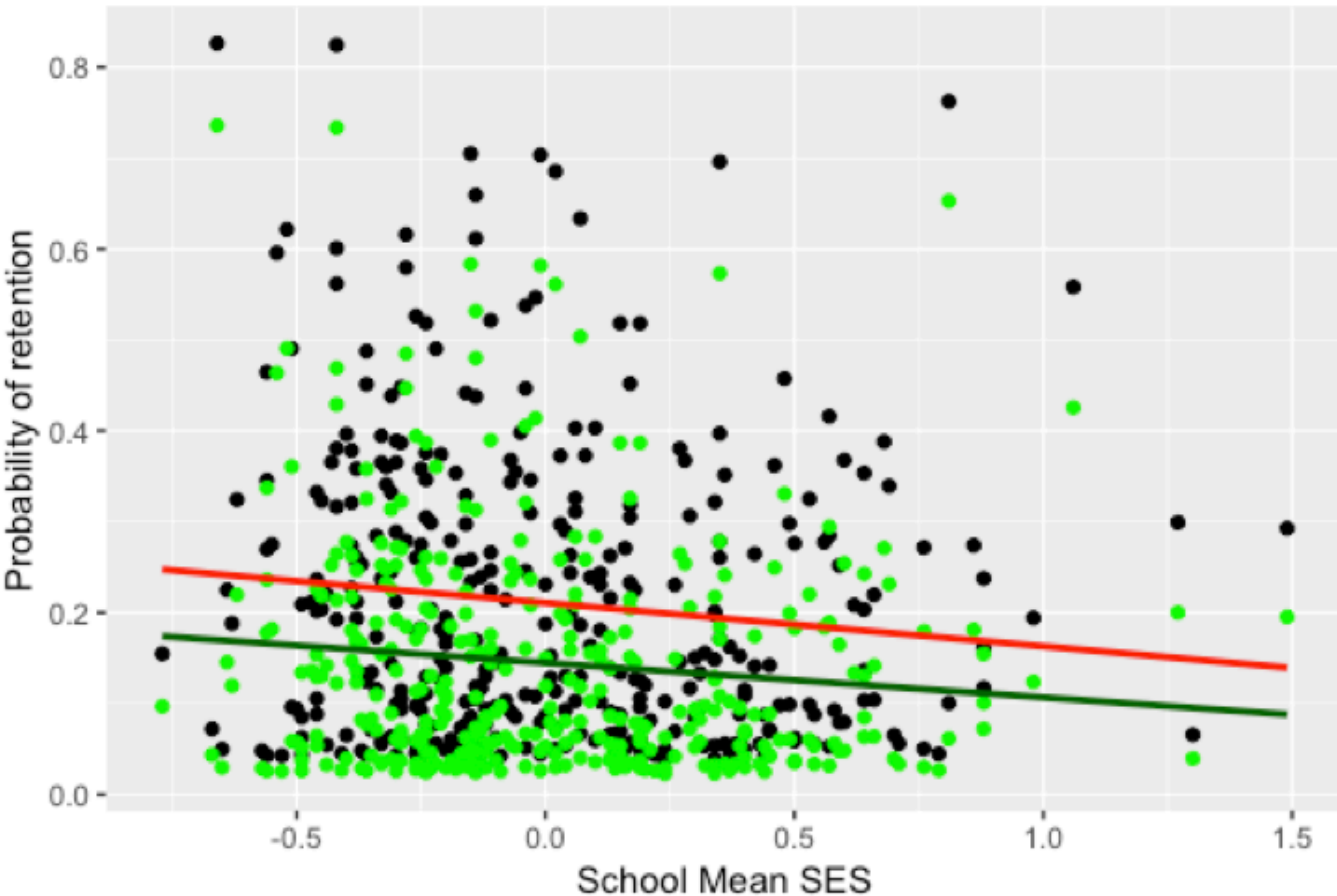
RQ1: describe variation in retention rates across schools, and

RQ2: describe how students with different prior experiences differ.

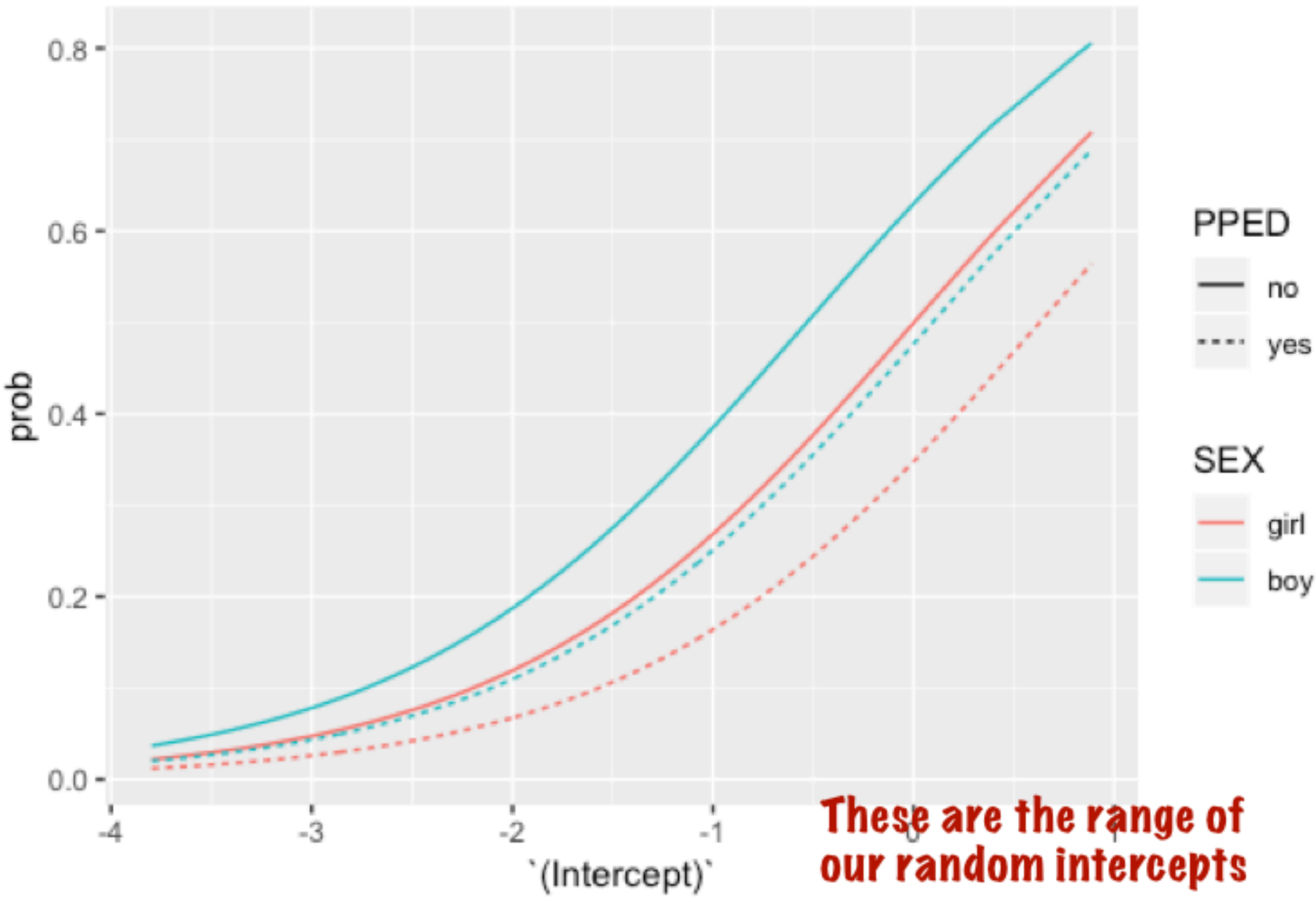
RQ3: see if school mean SES is associated with rate of retention.

- ★ Look at coefficient
- ★ Make plot to assess impact

Boys and girls across schools

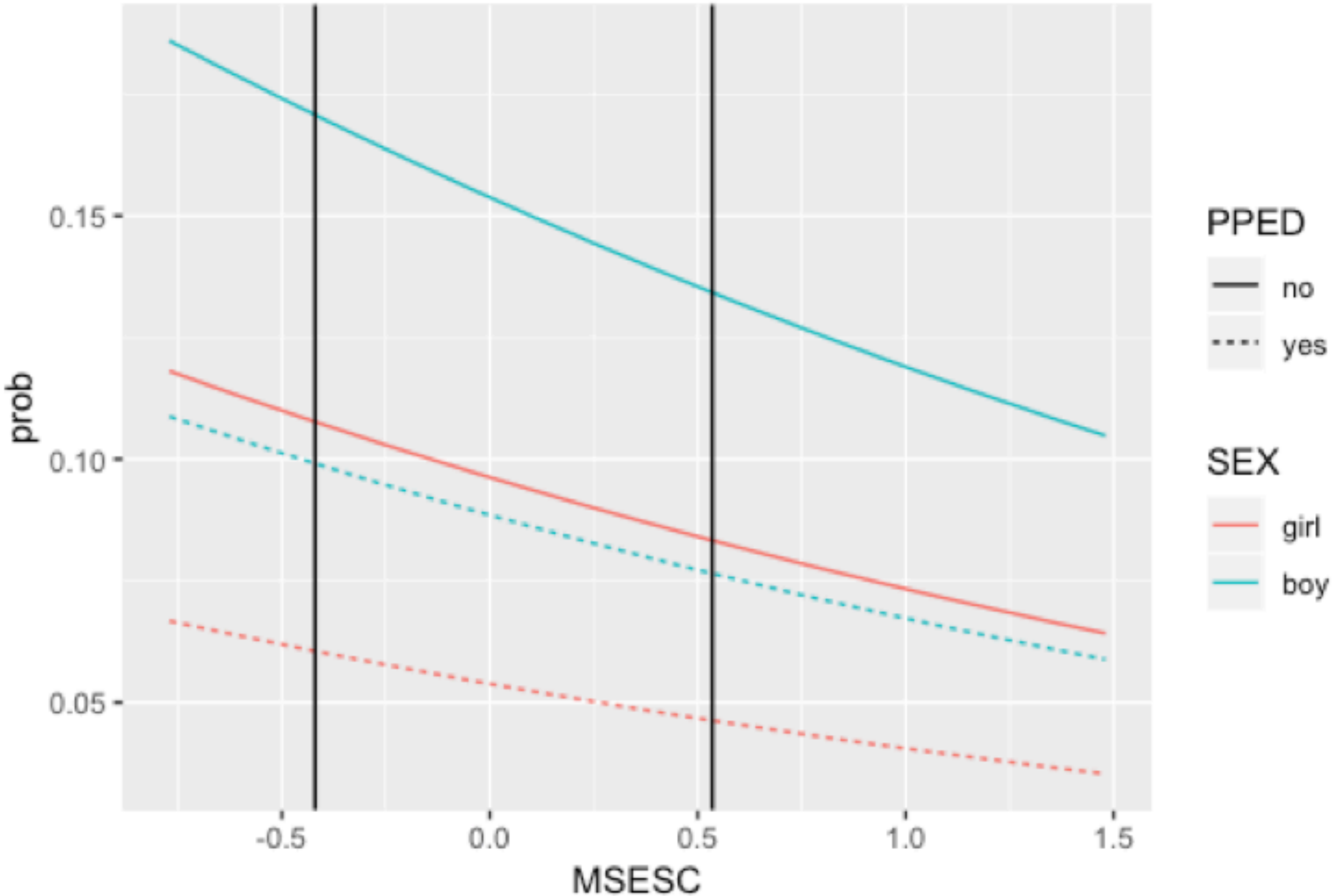


Our four groups of children (for different schools)



These are the range of our random intercepts

Impact of Mean SES



Population Average vs. Individual School

So far we are predicting for individuals in individual schools.

As things are non-linear, this is not the same as estimating averages across the population.

E.g., if we are investigating pre-primary experience and repetition we have two possible questions:

A: What is the association between pre-primary experience and risk of repetition at a given school?

B: What is the overall association of these two things averaged across schools?

*The coefficient in our model is **NOT** the answer to the second question.*

Fitting a Linear Model instead...

```
> M1.lin = lmer( as.numeric(REPEAT =="yes") ~ 1 + SEX +
PPED + MSESC + (1|SCHOOLID), data=dat )
> display( M1.lin )
lmer(formula = as.numeric(REPEAT == "yes") ~ 1 + SEX + PPED
+
    MSESC + (1 | SCHOOLID), data = dat)
            coef.est coef.se
(Intercept)  0.15     0.01
SEXboy       0.05     0.01
PPEDyes     -0.06     0.01
MSESC       -0.03     0.02

Error terms:
 Groups   Name        Std.Dev.
 SCHOOLID (Intercept) 0.14
 Residual             0.32
---
number of obs: 7516, groups: SCHOOLID, 356
AIC = 4724.2, DIC = 4655
deviance = 4683.5
```

...has some trouble

```
# our predictions for all students
pd = predict( M1.lin )

skimr::skim( pd )

Skim summary statistics

── Variable type:numeric ──────────────────────
 variable missing complete    n mean   sd     p0   p25  p
       pd       0     7516 7516 0.14 0.13 -0.065 0.045 0.

> mean( pd <= 0 )
[1] 0.05361895
```

some predictions are below 0, indicating negative chance of retention.

about 5% of our predictions are negative!

Also, the sd of the random intercept is 0.14, indicating entire schools with negative retention

Recap

Pepper the Black Cat

What we have seen today

http://cs179.org/lec52

Multilevel data with binary outcomes can be thought of as:

- ★ A logistic regression framework at level 1
- ★ A vanilla random-effects world at level 2 (no logistic stuff here!)

We can do this because level 2 equations are predicting the *coefficients* of the level 1 regression models, and these coefficients are continuous.

Interpreting models with nonlinear links is tricky.

Solution: make plots and describe trends and variation.

Appendix:
Red States vs. Blue States
Another example of multilevel logistic regression

See Chapter 14.2 of G&H

Interesting case study, worth a careful read for interpreting models, etc.

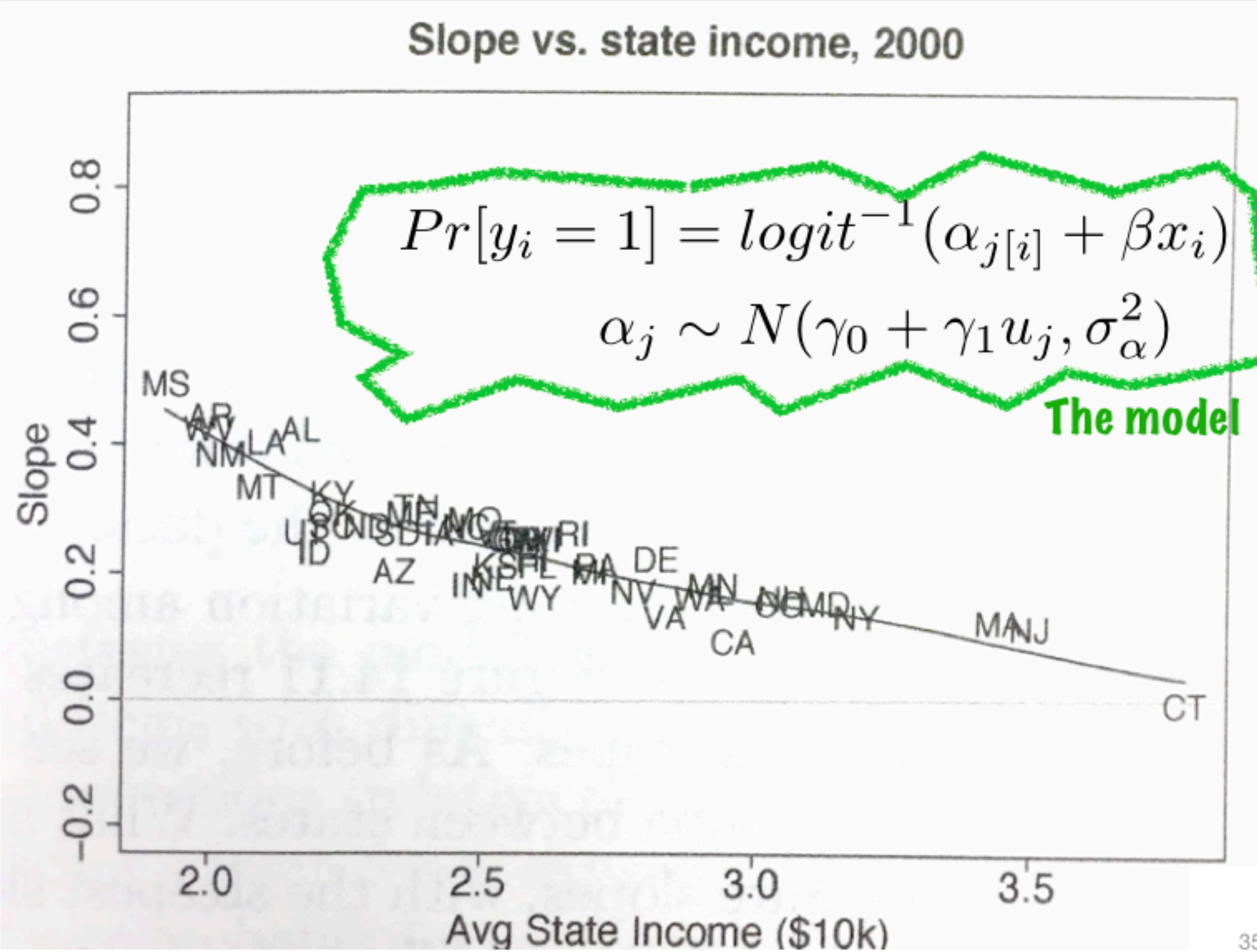
Red States vs. Blue States

The two following things are true:

- ★ Richer states tend to have a higher vote share for Democratic candidates
- ★ Richer people tend to have a higher vote share for Republican candidates

We can model this to understand what is going on with this seeming "paradox."

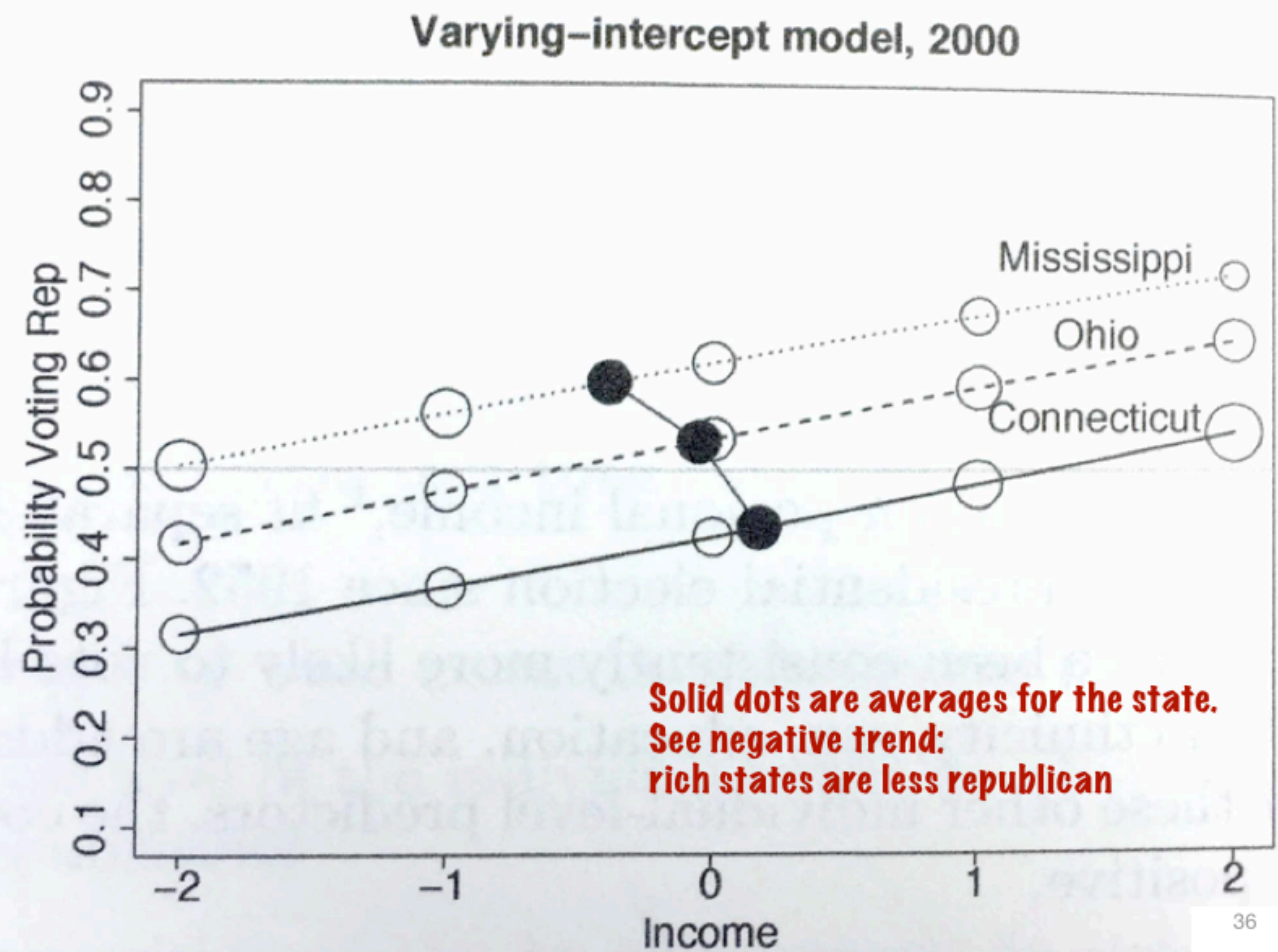
(See A Gelman's "Red State, Blue State, Rich State, Poor State: Why Americans Vote the Way They Do")



$$Pr[y_i = 1] = logit^{-1}(\alpha_{j[i]} + \beta x_i)$$

$$\alpha_j \sim N(\gamma_0 + \gamma_1 u_j, \sigma_\alpha^2)$$

The model


Varying–intercept model, 2000
Probability Voting Rep
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Mississippi
Ohio
Connecticut
Solid dots are averages for the state.
See negative trend:
rich states are less republican
−2
−1
0
1
2
Income

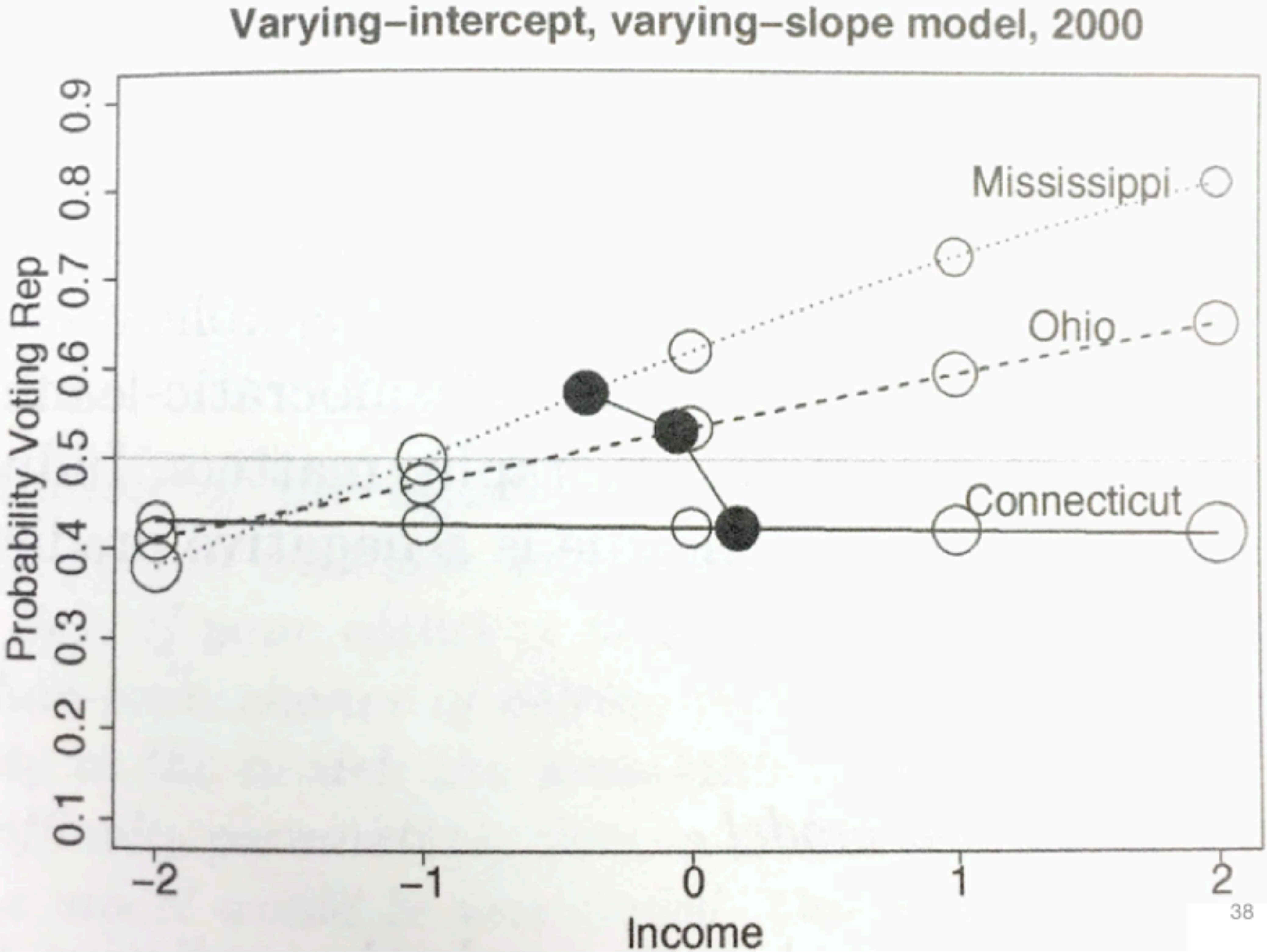
Varying intercept, varying slope

$$Pr[y_i = 1] = logit^{-1}(\alpha_{j[i]} + \beta_{j[i]} x_i)$$
$$\alpha_j = \gamma_{00} + \gamma_{01} u_j + e_{0j}$$
$$\beta_j = \gamma_{10} + \gamma_{11} u_j + e_{1j}$$

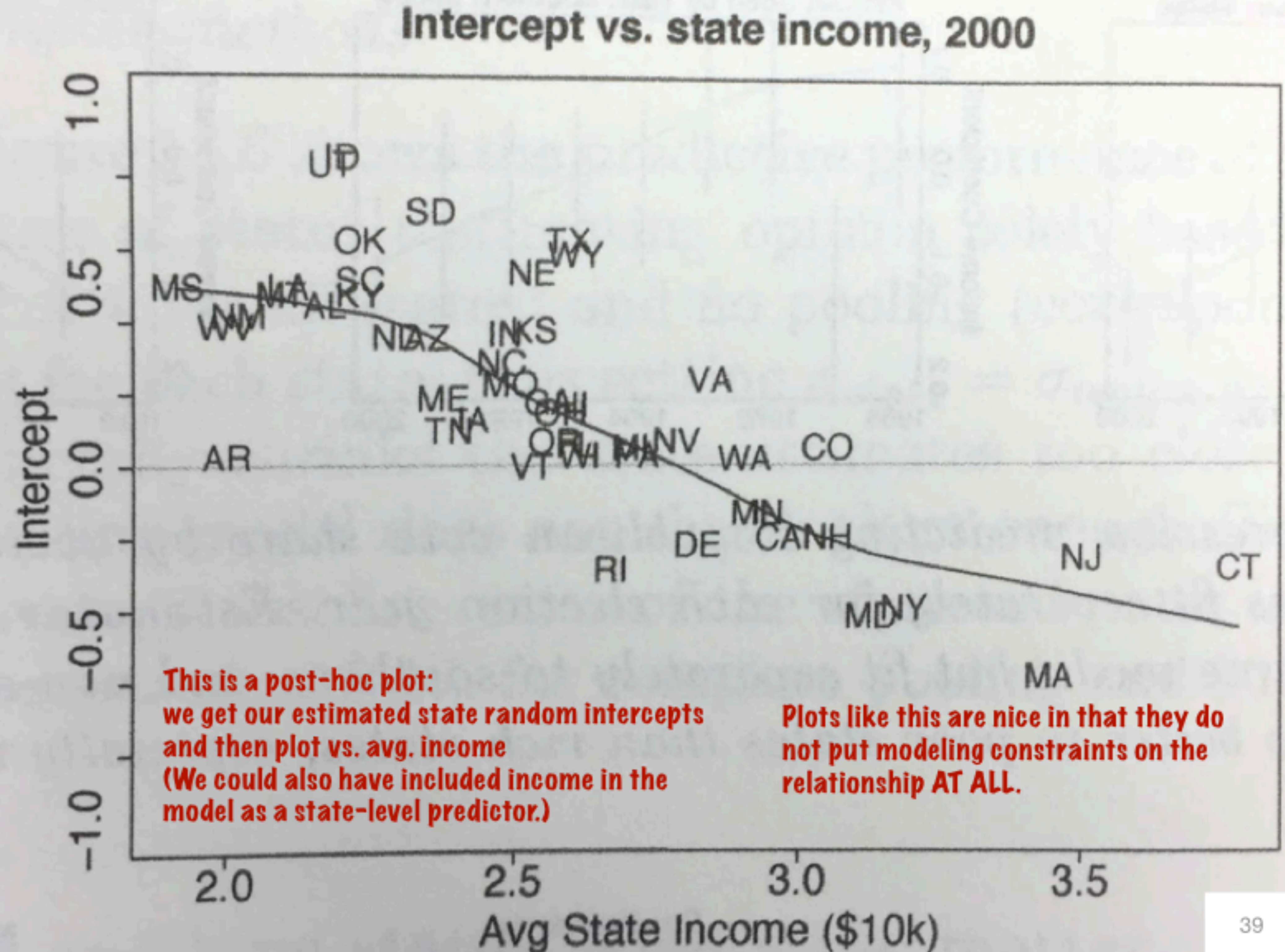
Level 2 is the same old, same old.

We have joint normally distributed residuals at level 2 as per usual.

It is just level 1 that gets the extra structure with links, etc.



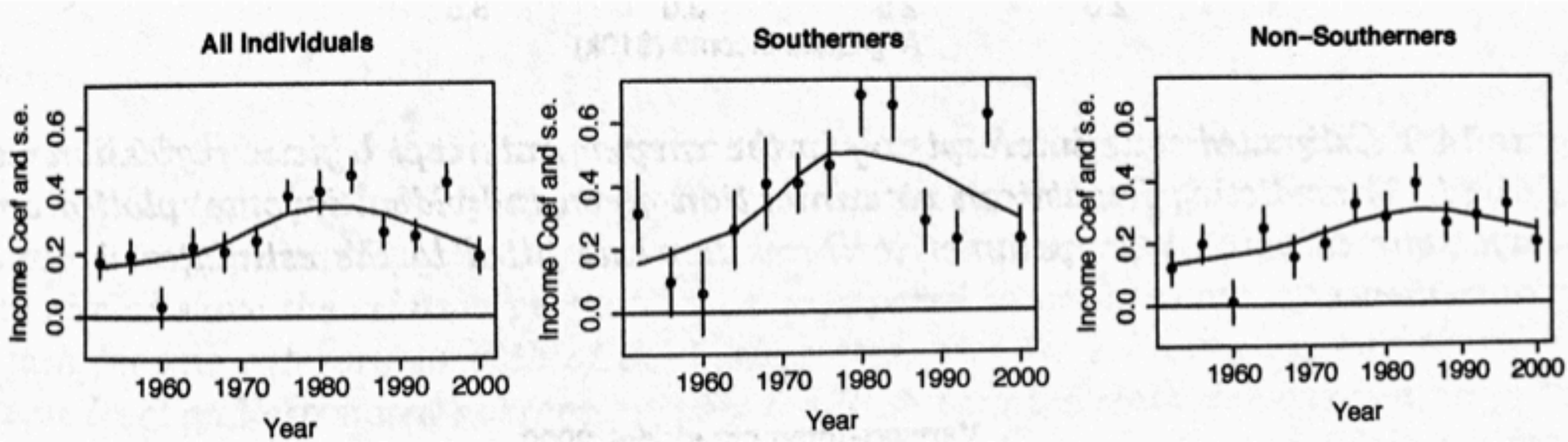
Varying–intercept, varying–slope model, 2000
Probability Voting Rep
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Mississippi
Ohio
Connecticut
−2
−1
0
1
2
Income



This is a post-hoc plot:
we get our estimated state random intercepts and then plot vs. avg. income
(We could also have included income in the model as a state-level predictor.)

Plots like this are nice in that they do not put modeling constraints on the relationship AT ALL.

Rich people vote republican

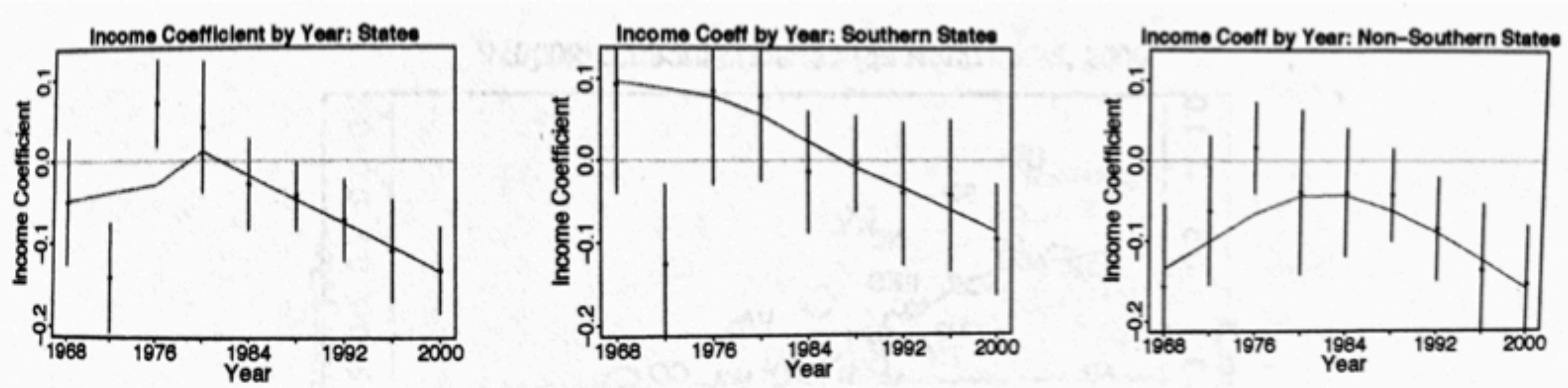


Each year vote for republican regressed onto income.

We see the coefficient (slope) of the model for each year.

0 would be no relationship.

Rich states vote democrat



Regression of republican vote share on average state income.

Right two plots broken down by north & south

Model fit separately for each election year.

Take-away messages

Interpreting the model is where the work is

You should rely on vizualizations and descriptions of different “typical cases”

(E.g., a rich, middle, and poor state)

You should plot the individual random effects against other level-2 covariates to look for trends

Describe across the groups, but avoid making strong claims about specific groups.