日本教育心理学会・研究委員会企画セミナー

マルチレベル分析入門

分寺 杏介

神戸大学大学院　経営学研究科

bunji@bear.kobe-u.ac.jp

※本スライドは，クリエイティブ・コモンズ 表示-非営利 4.0 国際 ライセンス（CC BY-NC 4.0）に従って利用が可能です。

よろしくお願いします。

- **分寺 杏介（ぶんじ・きょうすけ）**

  神戸大学大学院経営学研究科

- **専門：心理統計学**

  因子分析とかSEMとか項目反応理論とか，そのあたりの近く

  実はマルチレベルモデルを自分の分析では使ったことはないのですが……

  （階層ベイズモデリングはよく使います）

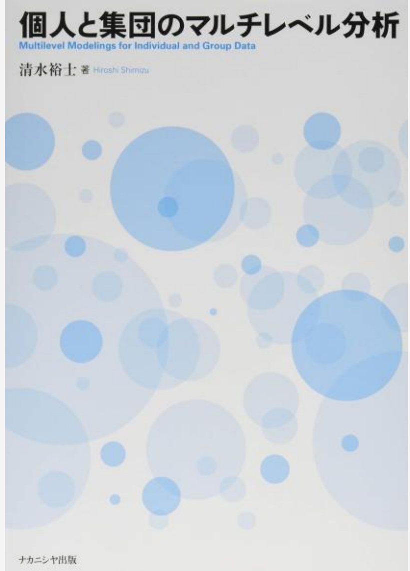
  この機会に基礎をお勉強したので，その成果発表ということで

別の機会に作成した講義資料をもとにしています

時間の都合でかなり端折っています ▶ 講義資料の完全版はこちら

講義資料を見ても分からないことがあればお気軽に質問をください！
調べて追記するかもしれません。

資料作成時に最も参考にした書籍



これらを読んでおけばいいのでは

…と思いつつ，自分なりに再構築しました。

やること・やらないこと

やること

マルチレベルモデルの基本的な考え方

- そもそもなぜマルチレベルなどという考え方が必要なのか

Rでの実装（glmmTMBパッケージ）

- とりあえず基本的なマルチレベルモデルはできるようになる
- 一般化線形モデル（GLMM）までできるようになりたいという強い気持ち

やらないこと

Rの基本

マルチレベル構造方程式モデリング

- R (lavaan) ではまだ十分に実行できないのと，SEMの基礎知識が必要なため

階層ベイズモデリング

- これもベイズ統計（とstanなど）の基礎知識が必要になってしまうため

1

マルチレベルなデータ

「データに階層性がある」について

大抵のデータに対して，階層性を考えることもできるし無視できることもある

学校	生徒	学級	学校	地域	
実験	試行	参加者			
企業	従業員	課	部	本部	企業

基本的には入れ子（ネスト）の関係にあると考える

【学校の場合】

上位カテゴリが複数（e.g., 部署と支社）ある場合や
1つの個体が複数の上位カテゴリに所属する（e.g., SNSのグループ）場合は
また別のモデルが必要になりますが
ここでは最もシンプルな
◀のような形式のみを考えていきます

データの構造のイメージ

【個人レベル】

生徒ID	学校ID	模試の点数	勉強時間
1	A	76	5.2
2	A	48	4.7
⋮	⋮	⋮	⋮
10	A	91	4.1
11	B	60	1.2
12	B	77	2.6
⋮	⋮	⋮	⋮

【集団レベル】

学校ID	生徒の総数
A	600
B	400
⋮	⋮

**レベルごとに
変数が存在するデータ**

本セミナーで使用するデータ

PISA2018「読解力リテラシー」の公開データを加工したもの

ダウンロード

```
dat <- readRDS("PISA_JPN.rds")
```

変数(本セミナーで使用するもののみ抜粋)

変数名	説明
student_id	回答者(生徒)のID
school_id	回答者(生徒)が所属する学校のID
read_score	PISA読解力の得点
wants_univ	予想する最終学歴(進学希望の代理指標として)
escs	回答者(生徒)の社会経済的地位
sch_escs_mean	学校ごとのescsの平均値
s_t_ratio	ST比(教員1人あたりの生徒の数)

分析の目的

読解力に影響を与える要因はなにか?

個人レベルの変数

集団レベルの変数

PISAデータの構造

（層化）二段階抽出によってランダムサンプリングが行われている

1 集団レベルのランダムサンプリング

学校レベルでは一斉に試験を実施できるので
オペレーションがかなり楽になります

（公立・私立などの区分による生徒数のバランスを考慮して）
対象となる学校をランダムに抽出する

2 個人レベルのランダムサンプリング

選ばれた学校から
対象となる生徒をランダムに抽出する

個人も学校もランダムサンプリング

▶ **個人の得点や，学校の平均点には関心がない**

統計的推定の精度(標準誤差)を見誤ってしまうため

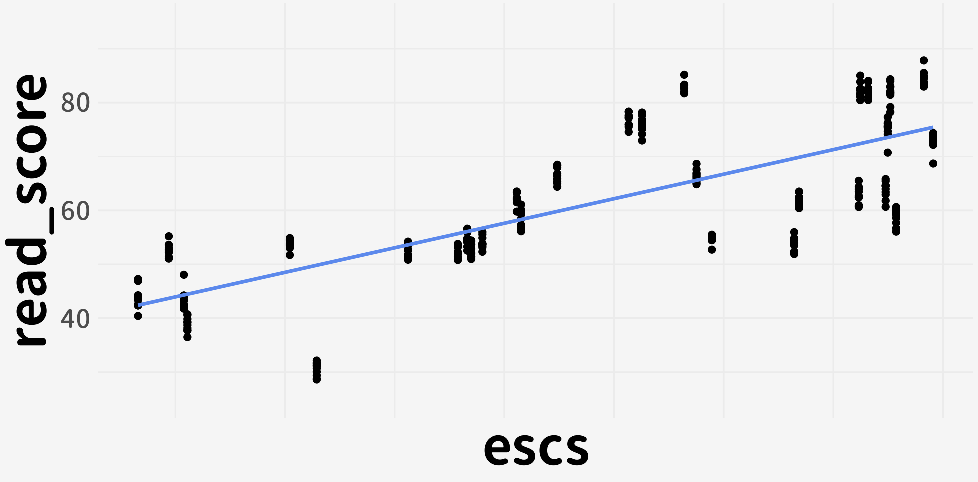
例

escsが高い生徒ほど read_score が高いかを統計的に検証するため,
escsを説明変数とした単回帰分析を行うことにしました。
ただし,read_scoreはテストによって変動するため,1人100回テストを受けてもらいました。

反復測定のイメージです

	ID	生徒ID	read_score	escs
1人分	1	1	76	0.8
	2	1	74	0.8
	⋮	⋮	⋮	⋮
	100	1	77	0.8
	101	2	60	-0.4
	102	2	58	-0.4
	⋮	⋮	⋮	⋮

【散布図】30 × 100 = 3000 個



統計的推定の精度(標準誤差)を見誤ってしまうため

例
escsが高い生徒ほど read_score が高いかを統計的に検証するため，
escsを説明変数とした単回帰分析を行うことにしました。
ただし，read_scoreはテストによって変動するため，1人100回テストを受けてもらいました。

反復測定のイメージです

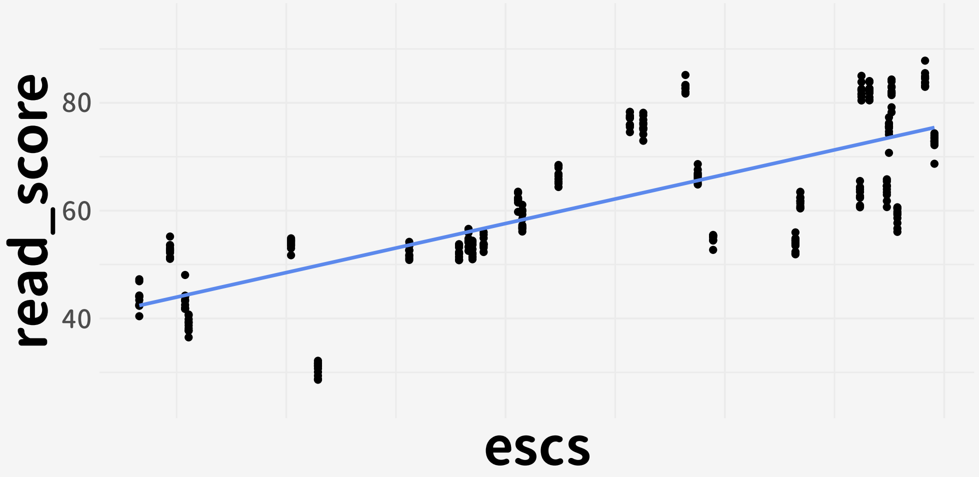
escsの回帰係数が有意かを検定しよう!

一般的なテキスト(やRのlm( )関数)で使用する
回帰係数の標本分布は

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma_e^2}{n\sigma_x^2}\right)$$

サンプルサイズが大きくなるほど
有意になりやすくなる

【散布図】$30 \times 100 = 3000$ 個



統計的推定の精度(標準誤差)を見誤ってしまうため

例 escsが高い生徒ほど read_score が高いかを統計的に検証するため，escsを説明変数とした単回帰分析を行うことにしました。ただし，read_scoreはテストによって変動するため，1人100回テストを受けてもらいました。

反復測定のイメージです

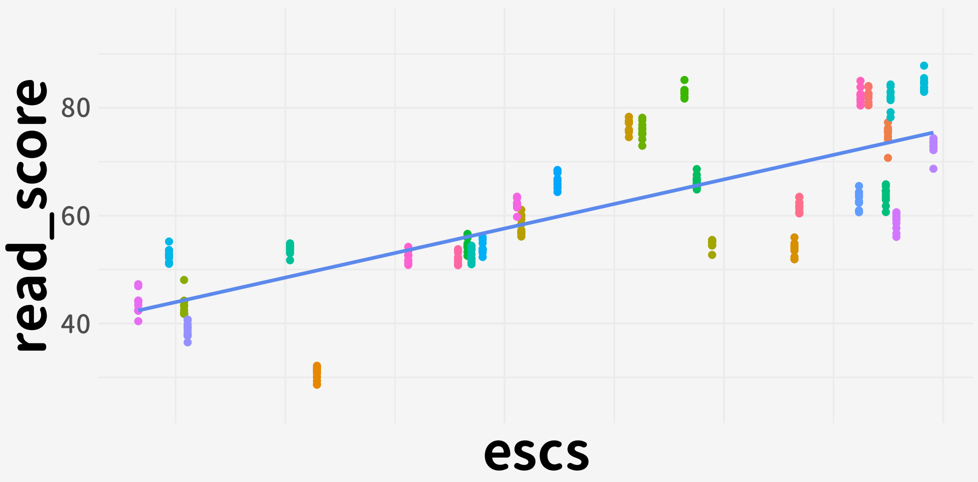
でもいいの?ホントにそれで

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma_e^2}{n\sigma_x^2}\right)$$

$n = 3000$ と考えていいのか?

でも同じ人のテストの点数はだいたい同じだよ?

【散布図】$30 \times 100 = 3000$ 個



統計的推定の精度(標準誤差)を見誤ってしまうため

例 escsが高い生徒ほど read_score が高いかを統計的に検証するため,
escsを説明変数とした単回帰分析を行うことにしました。
ただし,read_scoreはテストによって変動するため,1人100回テストを受けてもらいました。

反復測定のイメージです

でもいいの?ホントにそれで

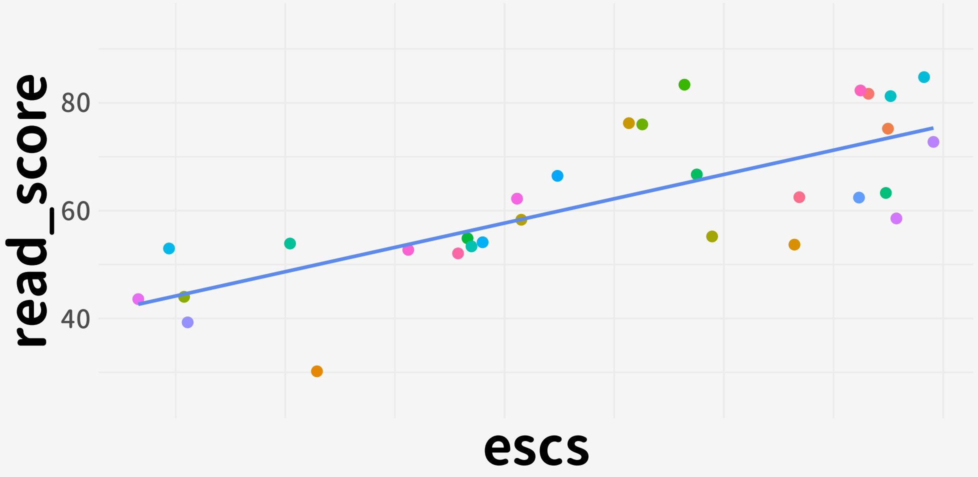
$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma_e^2}{n\sigma_x^2}\right)$$

$n = 3000$ と考えていいのか?

ぎゅっとまとめたら $n = 30$ じゃない?

read_scoreを平均値で置き換え

【散布図】$30 \times 100 = 3000$ 個



統計的推定の精度(標準誤差)を見誤ってしまうため

例 escsが高い生徒ほど read_score が高いかを統計的に検証するため,
escsを説明変数とした単回帰分析を行うことにしました。
ただし,read_scoreはテストによって変動するため,1人100回テストを受けてもらいました。

反復測定のイメージです

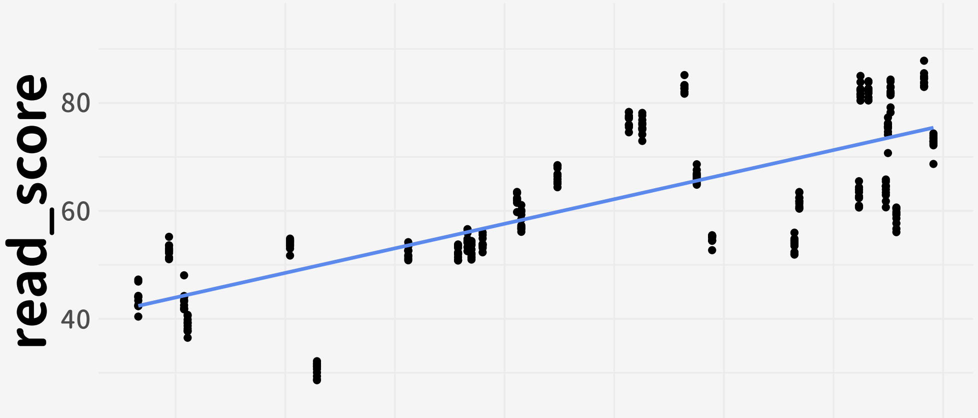
というわけで……

一般的なテキスト(やRのlm( )関数)で使用する
回帰係数の標本分布は

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma_e^2}{n\sigma_x^2}\right)$$

回帰直線からの誤差が互いに独立である場合に,
第1種の過誤の確率がαになる

【散布図】$30 \times 100 = 3000$ 個

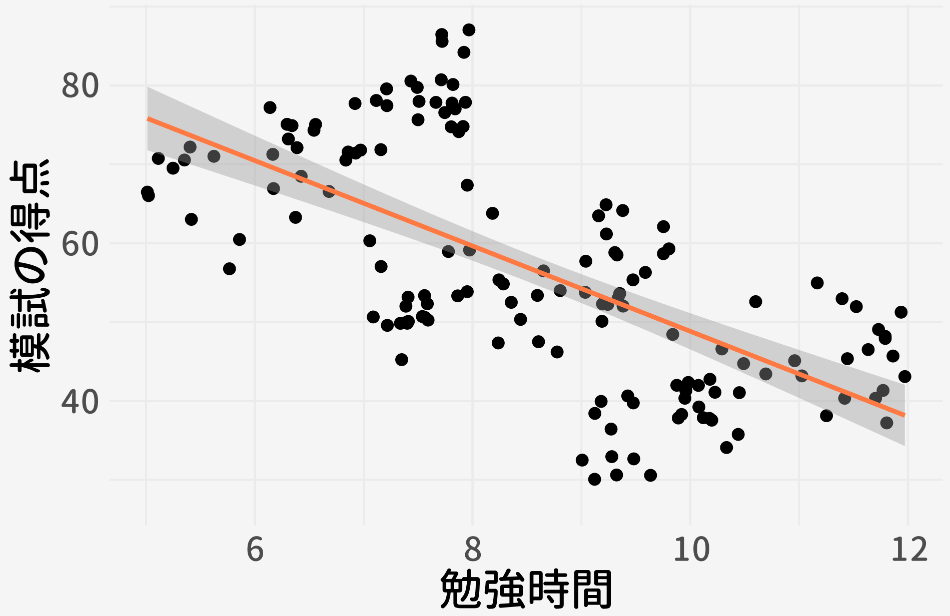


階層データでは,第1種の過誤の確率が大きくなってしまう

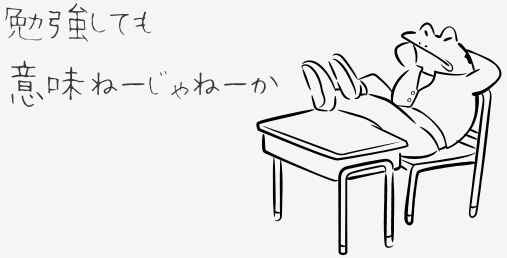
不用意に有意になってしまう

個体の効果と集団の効果が区別できなくなってしまうため

例 勉強時間が長い学生ほど模試の成績が高いのかを調べることにしました。
以下の図は,勉強時間と模試の得点の散布図です。

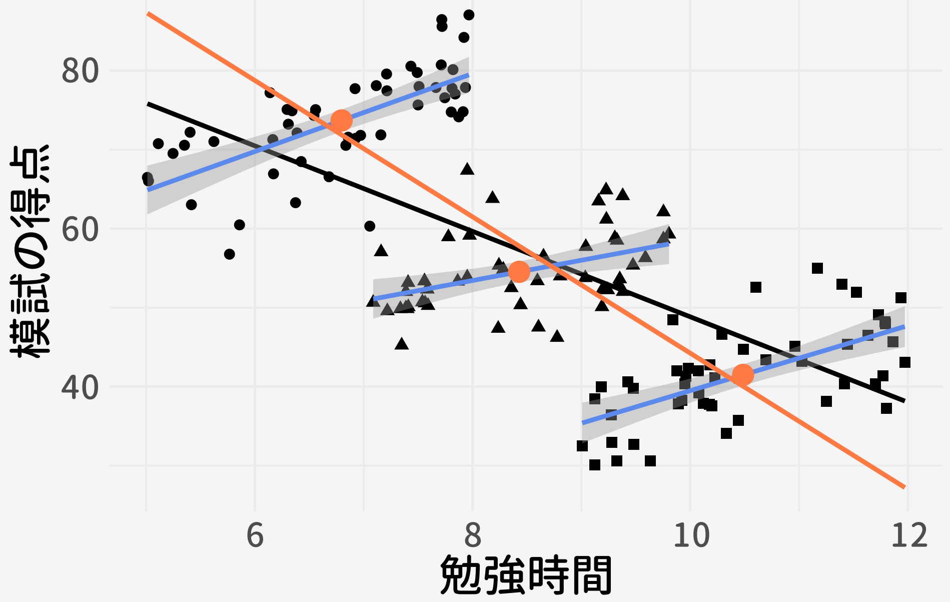


◀の図では明らかに
勉強時間が長いほど成績が低下する
ことが示されている



個体の効果と集団の効果が区別できなくなってしまうため

例 勉強時間が長い学生ほど模試の成績が高いのかを調べることにしました。
以下の図は，勉強時間と模試の得点の散布図です。
ちなみにこのデータは，地元の3つの高校から得られたものです。



グループごとに見てみると
勉強時間が長いほど成績が向上する
ことが示される

高校	学力	勉強時間
A	屈指の進学校	短め
B	普通の公立	ふつう
C	…	長め

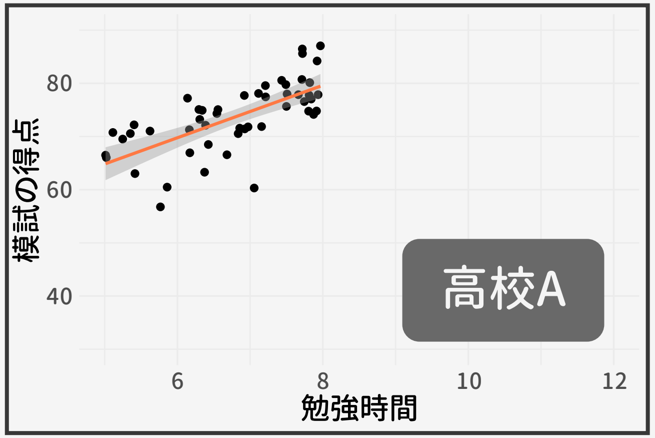
シンプソンのパラドックス

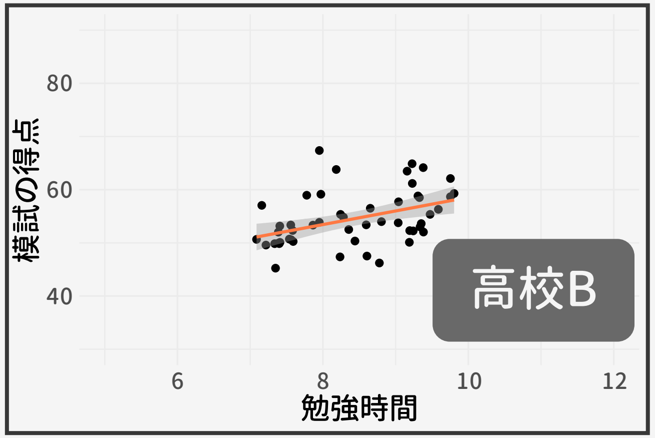
グループ間のベースラインの違いのせいで
全体では結果が違って見えてしまった

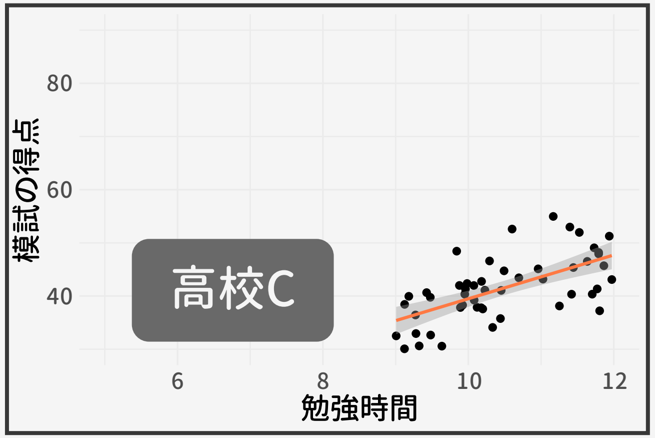
集団レベルの効果と個人レベルの効果が混ざって算出されてしまう

階層性にどう対処するか

集団ごとに分析する?







- ▶ 結果をまとめるのがめんどくさそう

グループの数が100とかなったらどうするの?

- ▶ **「全体的な傾向」があるかを確認しづらい**

e.g., (高校ごとに違いはあれど) **基本的には「勉強時間が長いほど得点が高い」か?**

そして「全体的な傾向」を確認するなら**全データを使って推定したほうが精度も良い**

データを分けるという行為は
多くの場合,単純に勿体ない

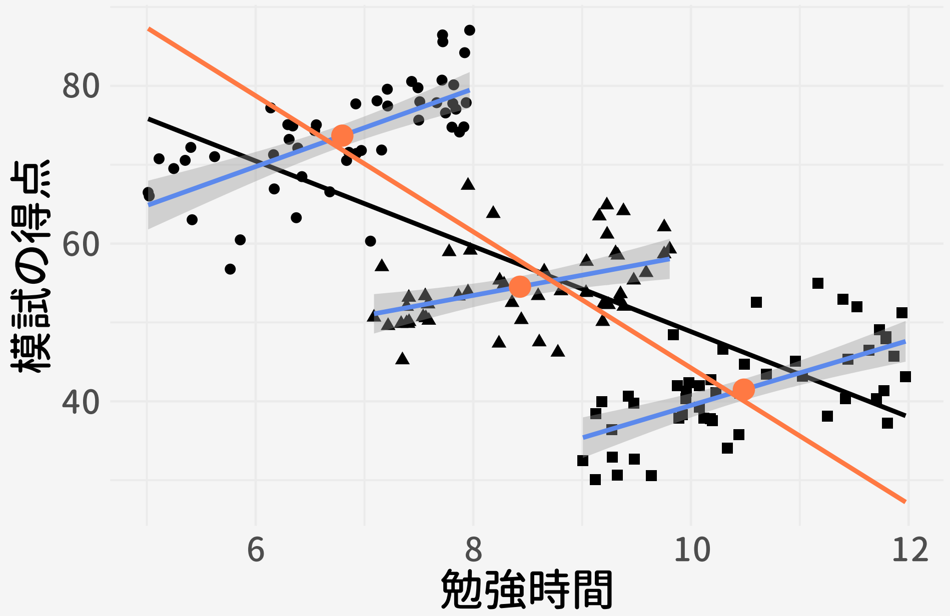
ということでマルチレベルモデルを使いましょう

2

マルチレベルな分析の基本的な考え方

変動を分解しよう

先程の「勉強時間」と「模試の得点」の例

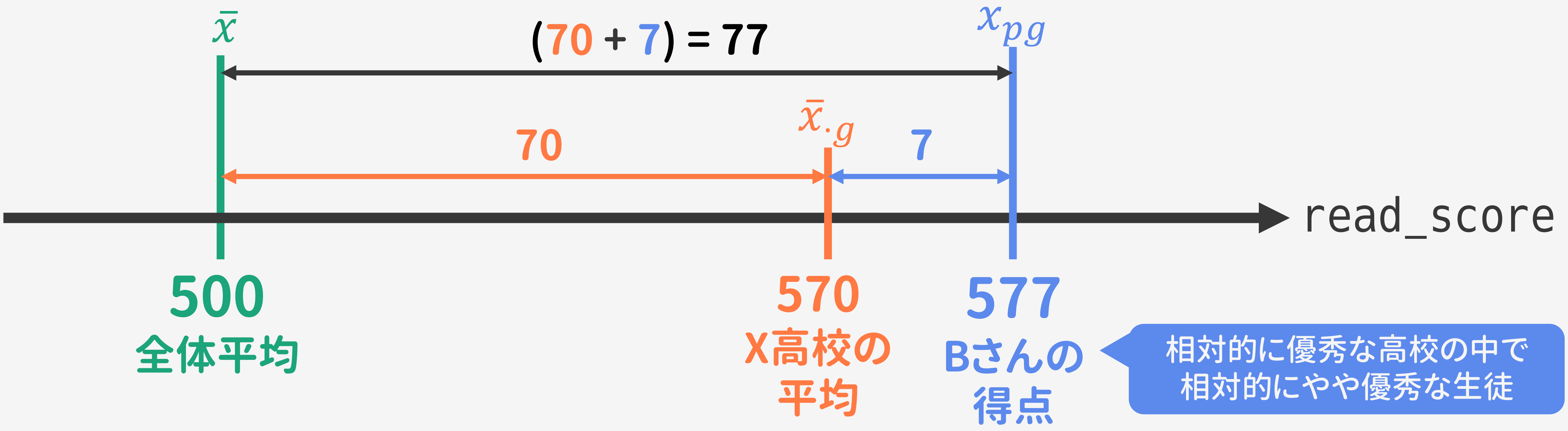


【個人レベルでは】 集団内での相対的な程度，という意味です。
勉強時間が長いほど
模試の得点も高い傾向 ＋

【集団レベルでは】
平均的に勉強時間が長い学校ほど
模試の得点は低い傾向 －

**この2つの効果を
分けて考えていきます**

例 Bさんのread_scoreは577点でした。
ちなみに，PISAの平均点は500点，またBさんが所属しているX高校の平均点は570点でした。



数式で表すと…

$$x_{pg} = \bar{x}_{\cdot g} + \left[x_{pg} - \bar{x}_{\cdot g}\right]$$
$$= \bar{x} + \left[\bar{x}_{\cdot g} - \bar{x}\right] + \left[x_{pg} - \bar{x}_{\cdot g}\right]$$

g は集団 (group) の添字
p は個人 (person) の添字

数式で分離

■ $x_{pg} = \bar{x}_{\cdot g} + [x_{pg} - \bar{x}_{\cdot g}]$ を使って回帰分析を変形させると

P16の数理的な説明です

$$y_{pg} = \beta_0 + \beta_1 x_{pg} + u_{pg} \leftarrow \text{残差}$$
$$= \beta_0 + \beta_1(\bar{x}_{\cdot g} + [x_{pg} - \bar{x}_{\cdot g}]) + u_{pg}$$
$$= \beta_0 + \beta_1 \bar{x}_{\cdot g} + \beta_1 [x_{pg} - \bar{x}_{\cdot g}] + u_{pg}$$

x の平均値が高い集団にいるほど y の値が大きくなる ＋ 集団内で相対的にxが高い人ほど y の値が大きくなる

➡ **いっそ回帰係数を別々に設定してしまえば良いのでは?**

※あとで詳しく説明します。

変動を「個人レベル」と「集団レベル」に分けたこと，ありますか？

そもそも変動を分解して考える必要がある理由

▶「集団レベル」の変動が無視できないほど大きいため

「集団レベル」の変動が小さい
＝「個人レベル」の変動のみ
なので普通に分析したら良い

「個人レベル」に比べて「集団レベル」が大きいかを評価する統計的手法

▶ それが**分散分析** (analysis of variance: ANOVA)

【学校ごとにread_scoreの平均値が異なるかの分散分析】

```
summary(aov(read_score ~ school_id, data = dat))
```

```
              Df   Sum Sq Mean Sq F value Pr(>F)
school_id    182 21708859  119279   21.38 <2e-16 ***
Residuals   5583 31148843    5579
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

簡単に言うと

$$\frac{\text{集団レベルの分散}}{\text{個人レベルの分散}}$$

$$\frac{\text{Sum Sq / Df (school\_id)}}{\text{Sum Sq / Df (Residuals)}}$$

分散分析のポイント

- 帰無仮説は「すべての集団で平均値が等しい」

  $H_0: \mu_1 = \mu_2 = \cdots = \mu_G$

  対立仮説は「どこかの群間には差がある」

- 有意ならば「集団ごとに差がありそう」なのでマルチレベルかも

- ただし以下のことまではまだわからない

  具体的にどこの集団間に差があるか ▶ 多重比較をしてみたり

  集団ごとのばらつきがどの程度であるか

  出力の `Mean Sq` を見ればだいたいはわかりますが

- 得られた結果は**データに含まれる学校における結果である**

  ✕ `read_score` の平均値は学校ごとに差がある

  ○ データに含まれる185校の間の少なくとも1ペアの間に差がある

関心があるのは?

「学校ごとに差が(どの程度)あるか」に関心があるとして……

PISAデータに含まれる学校は無作為に選ばれている

「その学校」である必然性はない,というか**もっと一般的な話がしたい**

【本当に知りたいこと】

母集団に含まれる各群(e.g., 日本の学校)の間には
どの程度のバラつきがあるのか

↑

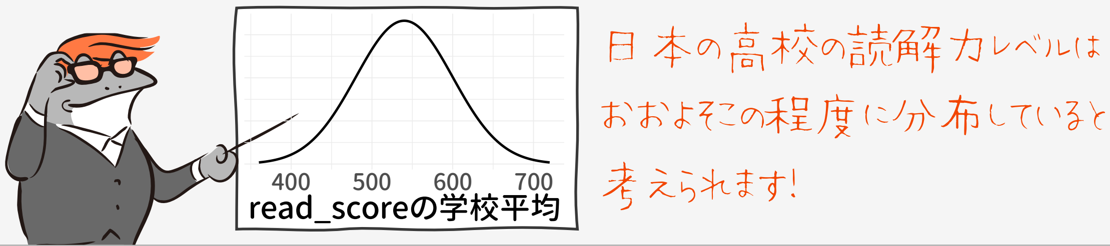
「すべての群間に差がない」などとは
最初から考えていない,**重要なのはその「程度」**

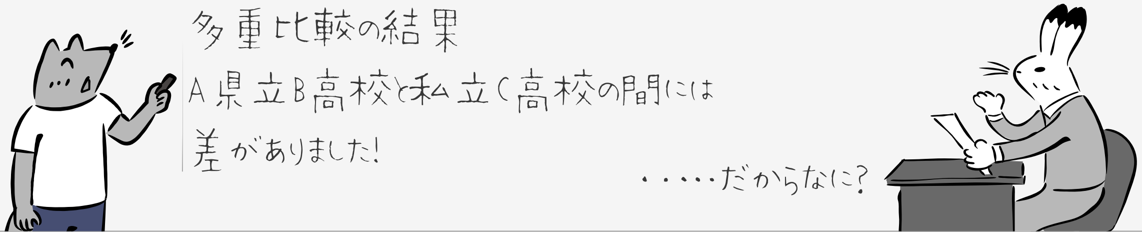
【cf. 分散分析から言えること】

標本に含まれる各群の間には
有意なバラつきがあるのか

↓ 推論くらいは可能ですが

標本に含まれない群に関しては
何も言うことができない





分散分析によって「学校ごとに差がある」と検討することは変なアプローチ

例 ある農園のみかんは，例年平均70gになるように栽培されています。
今年のみかんの重さをチェックするため，ランダムに40個のみかんの重さを測りました。

○ 正しい（ふつうの）アプローチ

1. 標本平均を計算する
2. **標本平均が**70と有意に異なるかを検定する
3. 有意であれば「問題あり」と判断する

× 変なアプローチ

1. 一つ一つの重さをチェックする
2. **70g±2gの範囲に入らないものが1つでもあったら**「問題あり」と判断する

↓

【変だと言える理由】

- たまたま標本に「ズレ」が見られたからと言って母集団レベルで「有意なズレ」があるとは言えないはず
- 「特定の標本」のズレには関心はない

実は分散分析と回帰分析は同じです

ダミー変数を使えば同じになる

$$y_{pg} = \beta_0 + \beta_{002}D_{002} + \beta_{003}D_{003} + \cdots + \beta_{185}D_{185} + u_{pg}$$

student_id	school_id	read_score	学校ダミー			
			D_{002}	D_{003}	…	D_{185}
0462	001	704.541	0	0	…	0
0850	001	569.687	0	0	…	0
⋮	⋮	⋮	⋮	⋮	…	⋮
6323	001	663.052	0	0	…	0
0030	002	407.067	**1**	0	…	0
0071	002	516.066	**1**	0	…	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
6182	185	564.172	0	0	…	**1**

【ダミー変数】

- school_idが002なら$D_{002} = 1$
- school_idが003なら$D_{003} = 1$

…という要領で設定する

（すべてのダミーが0ならschool_idは001）

(school_id=001) $y_{pg} = \beta_0 + u_{pg}$

(school_id=002) $y_{pg} = \beta_0 + \beta_{002} + u_{pg}$

⋮

(school_id=185) $y_{pg} = \beta_0 + \beta_{185} + u_{pg}$

説明変数を数字以外で与えた場合は，自動的にダミー変数化される

【学校ごとのread_scoreの差を評価する回帰分析】

```
summary(lm(read_score ~ school_id, data = dat))
```

```
Coefficients:
               Estimate Std. Error t value Pr(>¦t¦)
(Intercept)    621.2408    12.6256  49.205  < 2e-16 ***
school_id002 -148.3601    17.9861  -8.249  < 2e-16 ***
school_id003 -159.6712    18.9384  -8.431  < 2e-16 ***
school_id004 -115.5121    18.1239  -6.373 2.00e-10 ***
school_id005 -212.7595    18.4223 -11.549  < 2e-16 ***
school_id006   -9.3313    18.1239  -0.515 0.606669
 [ reached 'max' / getOption("max.print") -- omitted 177 rows ]
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 74.69 on 5583 degrees of freedom
Multiple R-squared:  0.4107,    Adjusted R-squared:  0.3915
F-statistic: 21.38 on 182 and 5583 DF,  p-value: < 2.2e-16
```

$y_{pg} = \beta_0 + \beta_{002}D_{002} + \beta_{003}D_{003} + \cdots + \beta_{185}D_{185} + u_{pg}$

(Intercept)がβ_0

school_id002から順に$\beta_{002}, \beta_{003}, \cdots$

【各学校の平均値】

school_id	パラメータ	平均値
001	β_0	621.2408
002	$\beta_0 + \beta_{002}$	472.8807
003	$\beta_0 + \beta_{003}$	461.5696
⋮	⋮	⋮

P22 (aov関数) のF値と完全に一致 ▶ 本質的に同じ分析方法なので当然

マルチレベルモデルの適用における重要な区分

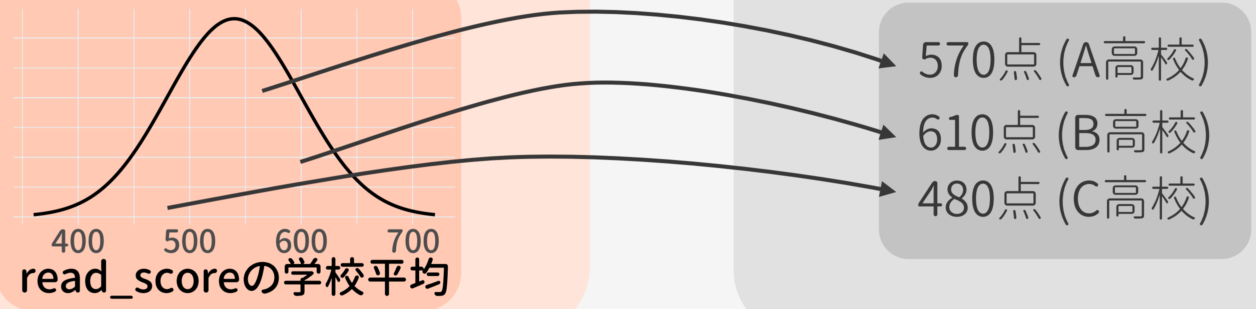
【変量効果】

一つ一つの値はランダムサンプリングの実現値

値そのものではなく，その背後の分布に関心がある

例 朝の読書タイムを行っている高校では，
していない高校と比べて読解力が高いのか？

変量効果の分布



【固定効果】

一つ一つの値そのものに関心がある

特定の群を選んでデータを取っているはず

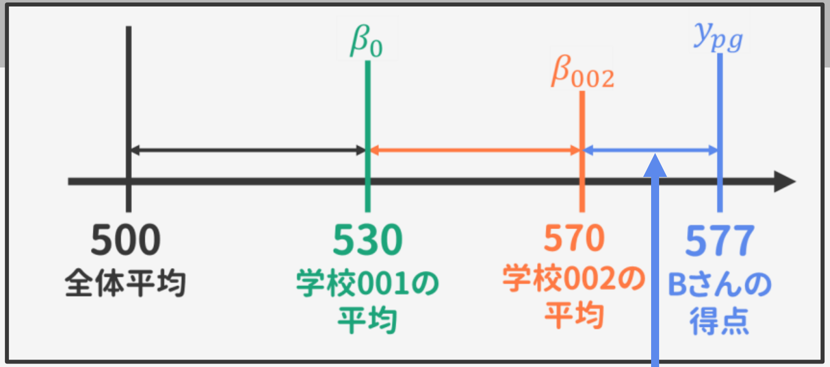
例 ○○市内で，朝の読書タイムを行っている
A高校は，B高校よりも読解力が高いのか？

【要注意】
変量効果の分布の平均値は
その値（母平均）に関心が
あるため，固定効果です

3

ランダム切片モデル

分散分析から先に進む



▶ ランダム切片モデル (random intercept model)

ポイントは「変量効果」の置き方

（固定効果ver.） $y_{pg} = \beta_0 + \beta_{002}D_{002} + \beta_{003}D_{003} + \cdots + \beta_{185}D_{185} + u_{pg}$ ← 群平均からの差

学校001の平均値

一つ一つの係数（群間の差）には関心がないので

$$y_{pg} = \beta_0 + u_{0g} + u_{pg}$$

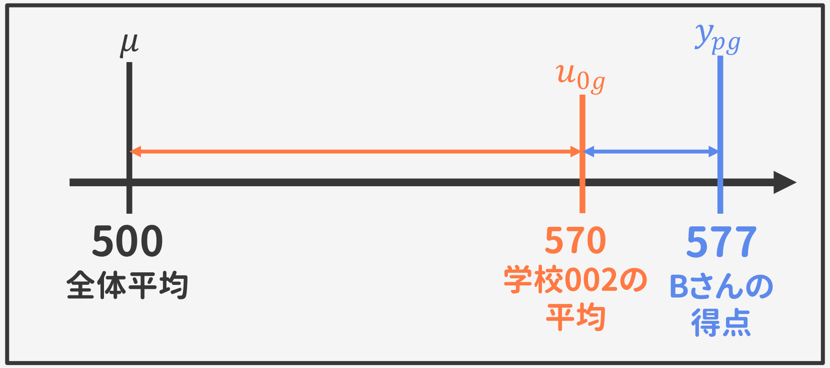
学校ごとに異なる値を取る確率変数

特定の群を基準にするのはなんだか気持ち悪いので

全体平均

（変量効果ver.） $y_{pg} = \mu + u_{0g} + u_{pg}$

観測値には {集団レベルの変量効果 / 個人レベルの変量効果} が含まれている

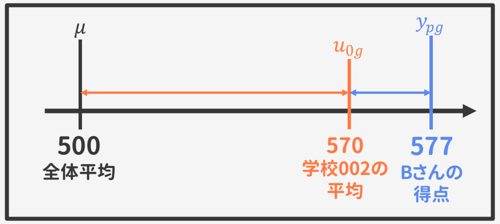


$$y_{pg} = \mu + u_{0g} + u_{pg}$$

レベル別に式を書くとわかりやすい

（レベル1：個人） $y_{pg} = \beta_{0g} + u_{pg}$

（レベル2：集団） $\beta_{0g} = \mu + u_{0g}$



通常の回帰分析でも標本分布の導出のためには「誤差は正規分布に従う」という仮定が必要でした

$$y_p = \beta_0 + \beta_1 x_p + u_p,$$
$$u_p \sim N(0, \sigma^2)$$

【仮定1】変量効果は通常正規分布に従う（ものとみなす）

$$u_{pg} \sim N(0, \sigma_{pg}^2)$$
$$u_{0g} \sim N(0, \sigma_{0g}^2)$$

群（集団） g の平均値を β_{0g} と表すと，

$$y_{pg} \sim N(\beta_{0g}, \sigma_{pg}^2)$$
$$\beta_{0g} \sim N(\mu, \sigma_{0g}^2)$$

【仮定2】異なるレベル間の変量効果は独立である

$$\mathrm{Cov}(u_{0g}, u_{pg}) = 0 \Rightarrow \sigma_y^2 = \sigma_{0g}^2 + \sigma_{pg}^2$$

ある変数の値の変動を
「個人レベル」と「集団レベル」に分解
▼
「集団レベル」の分散説明率を
簡単に計算できるようになる

モデルの式・推定するパラメータの比較

【ランダム切片モデル】(変量効果)

$$y_{pg} = \mu + u_{0g} + u_{pg}$$

または

(レベル1:個人)　$y_{pg} = \beta_{0g} + u_{pg}$

(レベル2:集団)　$\beta_{0g} = \mu + u_{0g}$

$$u_{pg} \sim N\left(0, \sigma_{pg}^2\right)$$
$$u_{0g} \sim N\left(0, \sigma_{0g}^2\right)$$

全体平均	μ	1個
変量効果の分散	$\sigma_{0g}^2, \sigma_{pg}^2$	2個

【分散分析モデル】(固定効果)

$$y_{pg} = \beta_0 + \beta_{002} D_{002} + \cdots + \beta_G D_G + u_{pg}$$

$$u_{pg} \sim N\left(0, \sigma_{pg}^2\right)$$

切片	β_0	1個
回帰係数	$\beta_{002}, \cdots, \beta_G$	G-1個
変量効果の分散	σ_{pg}^2	1個

マルチレベルモデルでは,変量効果の値を直接推定はしない

▶ パラメータ数が大幅に削減できるため,推定が安定しやすい

推定の違いについては,後ほどもう少し説明します

本セミナーでは，glmmTMBというパッケージを使用します。

一般的にRでマルチレベルモデルというと lme4（lmerTest）が紹介されがちですが glmmTMBのほうがいろいろなことができるのでおすすめです

> lme4とglmmTMBは，推定アルゴリズムが異なる（おかげでglmmTMBのほうが幅広く適用可能な）だけです。
> 嬉しいことに関数の記法はほぼ同じなので，lme4を用いた書籍を読み替える場合は，関数名を変えるだけでOKです。

【必要なパッケージのインストールと読み込み】

```
# Windowsの場合は事前に別のパッケージのインストールが必要
install.packages("TMB", type = "source")
install.packages("glmmTMB")
library(glmmTMB)
```

使用する関数名はパッケージ名と同じglmmTMB( )

わかりやすい!

【ランダム切片モデル】

```
model0 <- glmmTMB(read_score ~ (1 ¦ school_id), data = dat)
```

lm( )などでも登場していましたが【引数formula】

- y ~ xの形で表される
- 左右の対応は回帰モデルの式と同じ
  左辺 y には被説明変数
  右辺 x には説明変数
- 変量効果はカッコを使って表す
  カッコ内の右には集団を表す変数の名前
  カッコ内の左には変量効果(1は切片)
- レベル1の変量効果は自動的に入る
  回帰分析における残差なので

厳密に対応させて書くと

$$y_{pg} = \mu + u_{0g} + u_{pg}$$

```
read_score ~ 1 + (1 ¦ school_id) + (1 ¦ person_id)
```

書かなくても自動的に入る

もともと入っているので書いてしまうと二重になる

（レベル1：個人）　$y_{pg} = \beta_{0g} + u_{pg}$

（レベル2：集団）　$\beta_{0g} = \mu + u_{0g}$

$u_{pg} \sim N(0, \sigma_{pg}^2)$

$u_{0g} \sim N(0, \sigma_{0g}^2)$

他の多くの関数と同様に，summaryでOK

【ランダム切片モデルの出力を見る】

```
summary(model0)
```

【変量効果】

```
Random effects:

Conditional model:
 Groups    Name        Variance Std.Dev.
 school_id (Intercept) 3624     60.2
 Residual              5579     74.7
Number of obs: 5766, groups:  school_id, 183
```

各変量効果の分布の分散・標準偏差

$u_{pg} \sim N(0, 74.7^2)$

$u_{0g} \sim N(0, 60.2^2)$

【固定効果】

```
Conditional model:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  503.699      4.562   110.4   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

固定効果の推定値と検定

$\beta_{0g} = 503.699 + u_{0g}$

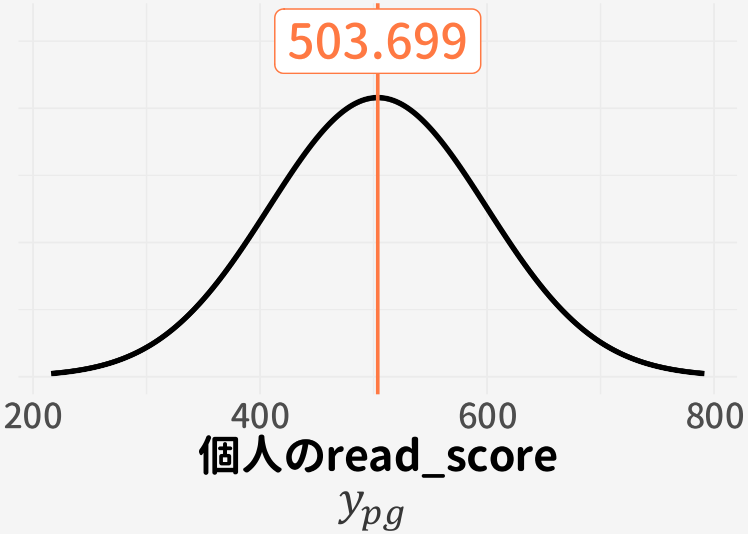
係数の意味

（レベル1：個人）　$y_{pg} = \beta_{0g} + u_{pg}$　　$u_{pg} \sim N(0, 74.7^2)$

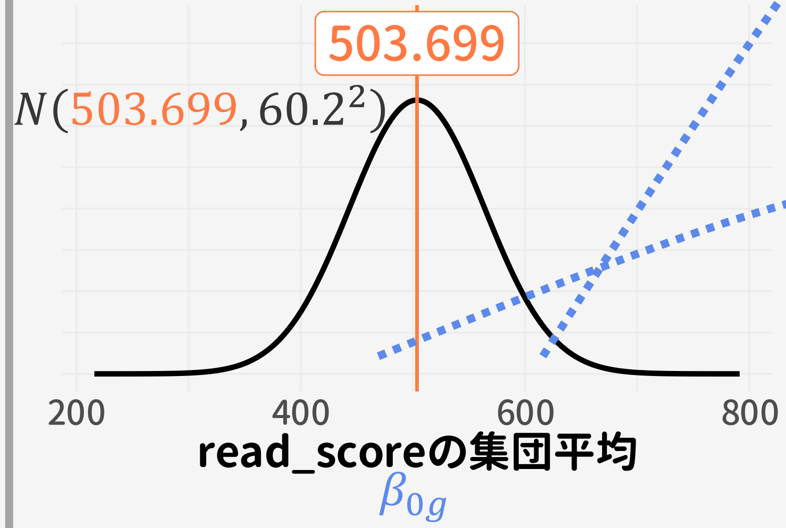
（レベル2：集団）　$\beta_{0g} = 503.699 + u_{0g}$　　$u_{0g} \sim N(0, 60.2^2)$

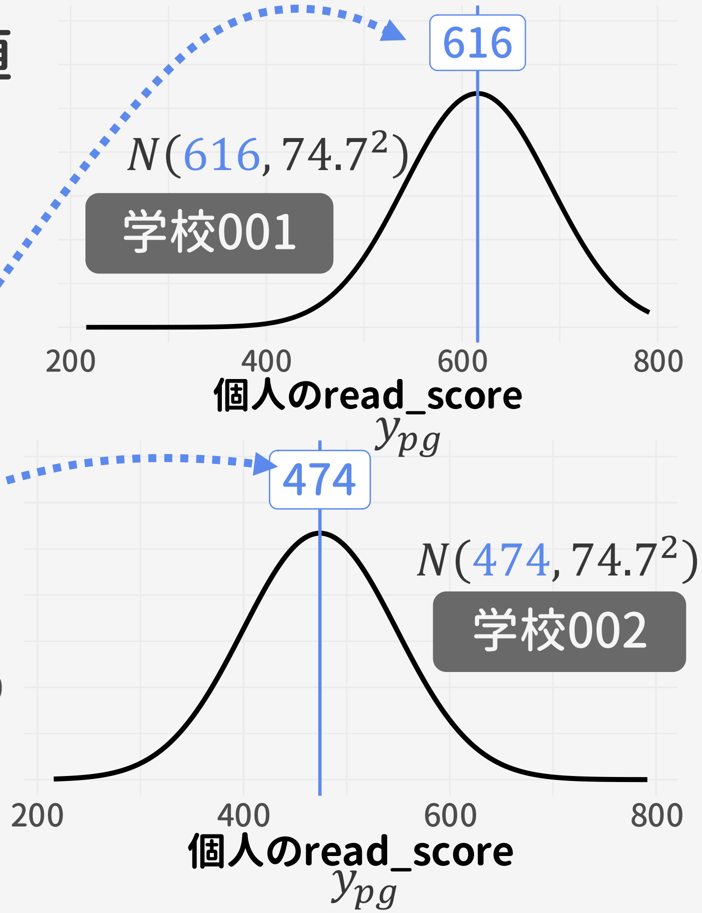
全体平均 μ は個人レベルではなく「集団レベル」の平均値

× 正しくない解釈



○ 正しい解釈





事後的に計算することは可能です

【集団の平均パラメータを事後的に計算する】

```
coef(model0)$cond$school_id
```

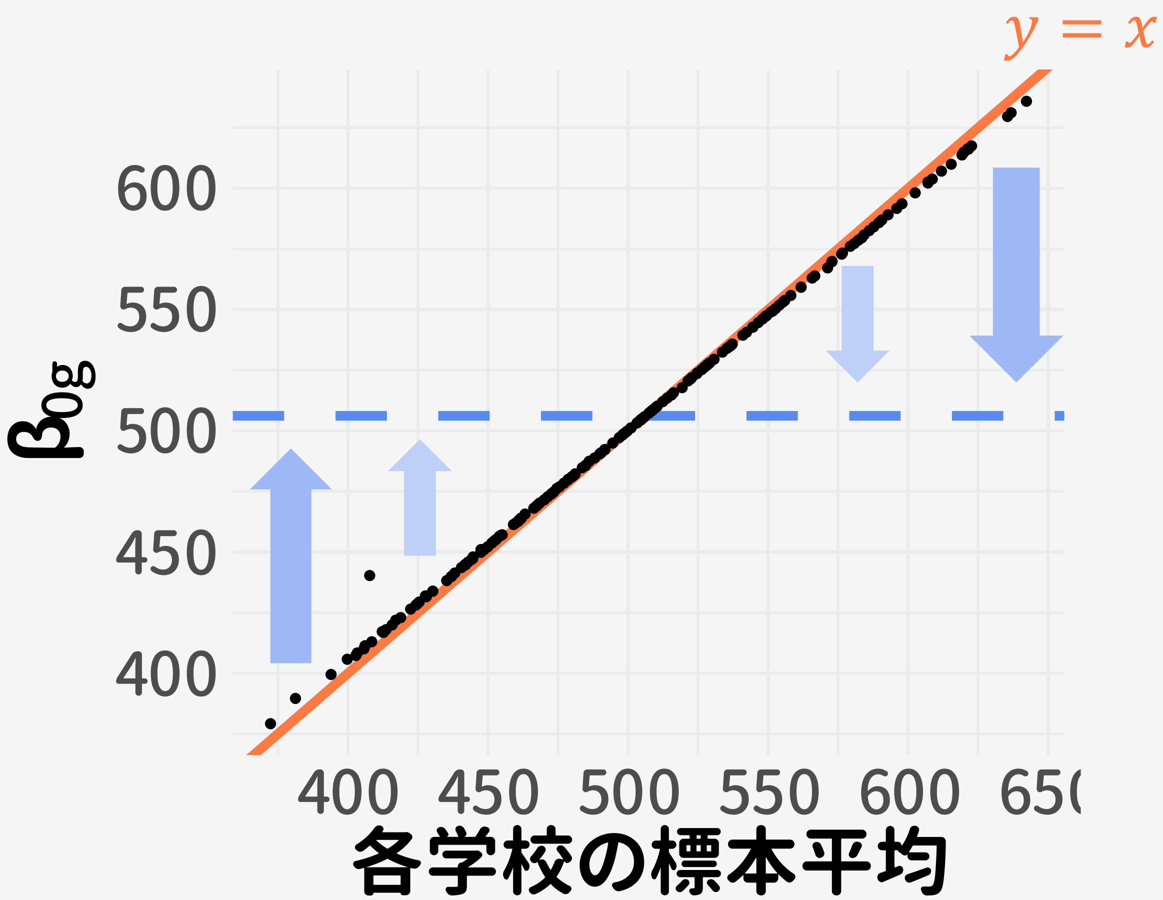
```
 (Intercept)
001    616.2887
002    474.2156
003    463.7651
004    505.6382
005    412.9860
006    607.0865
007    584.1363
008    416.8365
----(以下省略)----
```

【cf. 各学校の標本平均(=固定効果分散分析の推定値)】

school_id	パラメータ	平均値
001	β_0	621.2408
002	$\beta_0 + \beta_{002}$	472.8807
003	$\beta_0 + \beta_{003}$	461.5696
004	$\beta_0 + \beta_{004}$	505.7287
005	$\beta_0 + \beta_{005}$	408.4813
006	$\beta_0 + \beta_{006}$	611.9095
007	$\beta_0 + \beta_{007}$	587.7783
008	$\beta_0 + \beta_{008}$	412.6578
⋮	⋮	⋮

だいたい同じ?

「標本平均」と「事後的に求めた β_{0g}」の散布図を描いてみると



基本的にはどちらでも同じような結果だが
傾きが少しだけ小さくなっている

全体平均（青い点線）に近づくように
（縮退）推定が行われている

サンプルサイズが小さいときに有効となる可能性があります

変量効果にするメリット：推定の安定性

本来変量効果の特定の値に関心はないのですが，
ここではメカニズムの理解のためにあえて考えてみます。

例 ある学校では，たまたま試験日にインフルエンザが流行していたせいで，
3人しか受験できませんでした。ただし，再試験の時間もないのでこのまま分析します。

school_id == "001"の学校からは以下の3人のデータのみが得られた

1	2	3	標本平均
770.257	710.876	722.404	**734.5123**

【変量効果で推定する場合】

他の学校のデータから，
学校平均の分布が推定される
▶ **標本平均とこの分布の間を取って**
調整された値が出てくる(657.2806)

【固定効果で推定する場合】

他の学校のデータがあろうとなかろうと
標本平均がそのまま推定値になる
▶ 学校の平均値は734.5123

(イメージ)推定を相互にサポートしている

ベイズ統計で言うところの「尤度」と「事前分布」の関係に非常によく似ています

各学校の観測値から分布を得る

school_id	人数	平均値
001	3	734
002	33	472
003	27	461
004	31	505
005	130	408
006	30	611
007	27	587
008	31	412
⋮	⋮	⋮

分布をもとに学校ごとの平均が調整される

school_id	β_{0g}
001	657
002	478
003	470
004	504
005	409
006	600
007	575
008	422
⋮	⋮

人数が少ないと，標本平均の偏りが「偶然」である可能性が高い ▶ 分布によって強く調整される

人数が多いと，標本平均の偏りは「必然」かもしれない ▶ 分布による調整は弱い

場合によるわけですが，統計的にも判断する必要があります

概念的な階層性と統計的な階層性

多くのデータは，**何かしらの意味で「階層性」がある**と考えられる

【例】サンプリングの都合で複数の高校の生徒に質問紙調査を行った場合

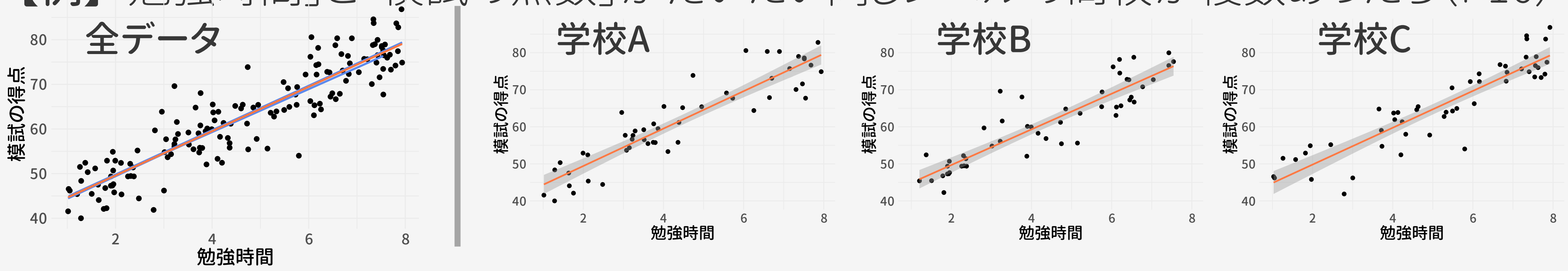
学内の掲示板で実験の参加者を募集した場合

場合によるわけですが，統計的にも判断する必要があります

概念的な階層性と統計的な階層性

ただし，**それらの階層性を常に考える必要があるわけでもない**

【例】「勉強時間」と「模試の点数」がだいたい同じレベルの高校が複数あったら（P16）



▶ わざわざ分けても意味なさそう？

PP10-16の説明は「データの集団凝集性」によるもの
▶ 集団レベルのばらつきが大きい場合にはマルチレベルにすべき

級内相関係数 (interclass correlation coefficient [ICC])

変数の分散のうち，集団レベルによって説明できる割合（分散説明率）

P31

【仮定2】異なるレベル間の変量効果は独立である

$$\mathrm{Cov}\left(u_{0g}, u_{pg}\right) = 0 \Rightarrow \sigma_y^2 = \sigma_{0g}^2 + \sigma_{pg}^2$$

ある変数の値の変動を
「個人レベル」と「集団レベル」に分解
▼
「集団レベル」の分散説明率を
簡単に計算できるようになる

$$\Rightarrow \mathrm{ICC} = \frac{\sigma_{0g}^2}{\sigma_y^2} = \frac{\sigma_{0g}^2}{\sigma_{0g}^2 + \sigma_{pg}^2}$$

ICCの目安

$\mathrm{ICC} \geq 0.1$ あたりからはマルチレベルモデルにすべきと言われているらしい

ICCの計算

P35の推定値から手計算も可能です

$$\mathrm{ICC} = \frac{\sigma_{0g}^2}{\sigma_{0g}^2 + \sigma_{pg}^2} = \frac{60.2^2}{60.2^2 + 74.7^2} \approx 0.394$$

Rで計算させる場合

performanceパッケージがおすすめです

ICCの計算だけなら他のパッケージでも問題ないのですが，
このパッケージは回帰分析に関するいろいろな診断ができて便利です

【パッケージのインストールと読み込みとICCの計算】

```
install.packages("performance")
library(performance)
icc(model0)
```

```
# Intraclass Correlation Coefficient

    Adjusted ICC: 0.394
  Unadjusted ICC: 0.394
```

ここまででは，まだ回帰分析とは言えません

学校ごとの切片の違いをモデリングしただけで，説明変数を投入していない

説明変数を追加して，回帰分析らしいことをしていきましょう

【問】社会経済的地位（escs）が高い生徒ほど，読解力も高い傾向にあるのか？

$$y_{pg} = \mu + \beta_{\mathrm{ESCS}}\mathrm{ESCS}_{pg} + u_{0g} + u_{pg}$$

ESCS_{pg}は個人レベル の変数なので

（レベル1：個人） $y_{pg} = \beta_{0g} + \beta_{\mathrm{ESCS}}\mathrm{ESCS}_{pg} + u_{pg}$ $u_{pg} \sim N\left(0, \sigma^2_{pg}\right)$

（レベル2：集団） $\beta_{0g} = \mu + u_{0g}$ $u_{0g} \sim N\left(0, \sigma^2_{0g}\right)$

説明変数はformulaの右辺に足せば良い

【ランダム切片回帰分析モデル】

```
model1 <- glmmTMB(read_score ~ escs + (1 ¦ school_id), data = dat)
summary(model1)
```

【変量効果】

```
Random effects:

Conditional model:
 Groups    Name        Variance Std.Dev.
 school_id (Intercept) 3313     57.56
 Residual              5558     74.56
Number of obs: 5766, groups:  school_id, 183
```

各変量効果の分布の分散・標準偏差

$$u_{pg} \sim N(0, 74.56^2)$$
$$u_{0g} \sim N(0, 57.56^2)$$

【固定効果】

```
Conditional model:
            Estimate Std. Error z value Pr(>¦z¦)
(Intercept)  504.791      4.375  115.38  < 2e-16 ***
escs           9.478      1.564    6.06 1.36e-09 ***
```

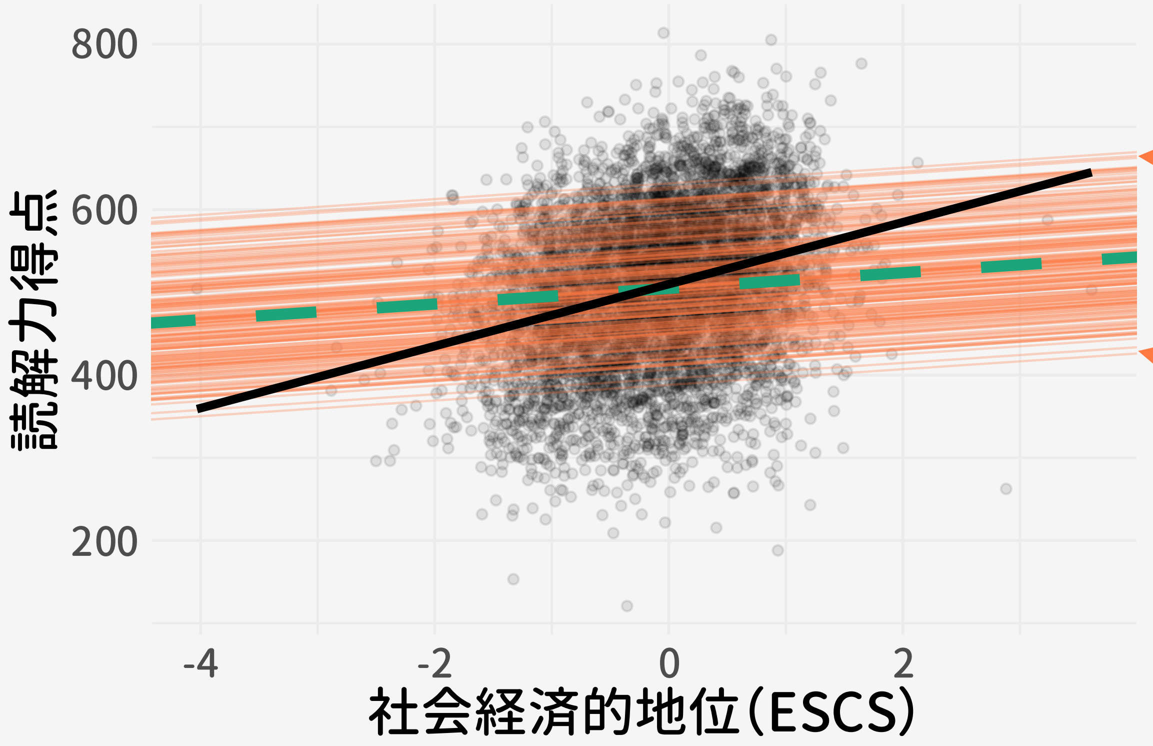
固定効果の
推定値と検定

$$y_{pg} = \beta_{0g} + 9.478\mathrm{ESCS}_{pg} + u_{pg}$$
$$\beta_{0g} = 504.791 + u_{0g}$$

改めて係数の解釈を

$$y_{pg} = 504.791 + 9.478\mathrm{ESCS}_{pg} + u_{0g} + u_{pg}$$

- 傾きはすべての集団で共通，切片が異なるだけ



（学校001）

$$\hat{y}_{pg} = \overbrace{504.791 + 106.9}^{611.691} + 9.478\mathrm{ESCS}_{pg}$$

（全体平均）

$$\hat{y}_{pg} = 504.791 + 9.478\mathrm{ESCS}_{pg}$$

（学校010）

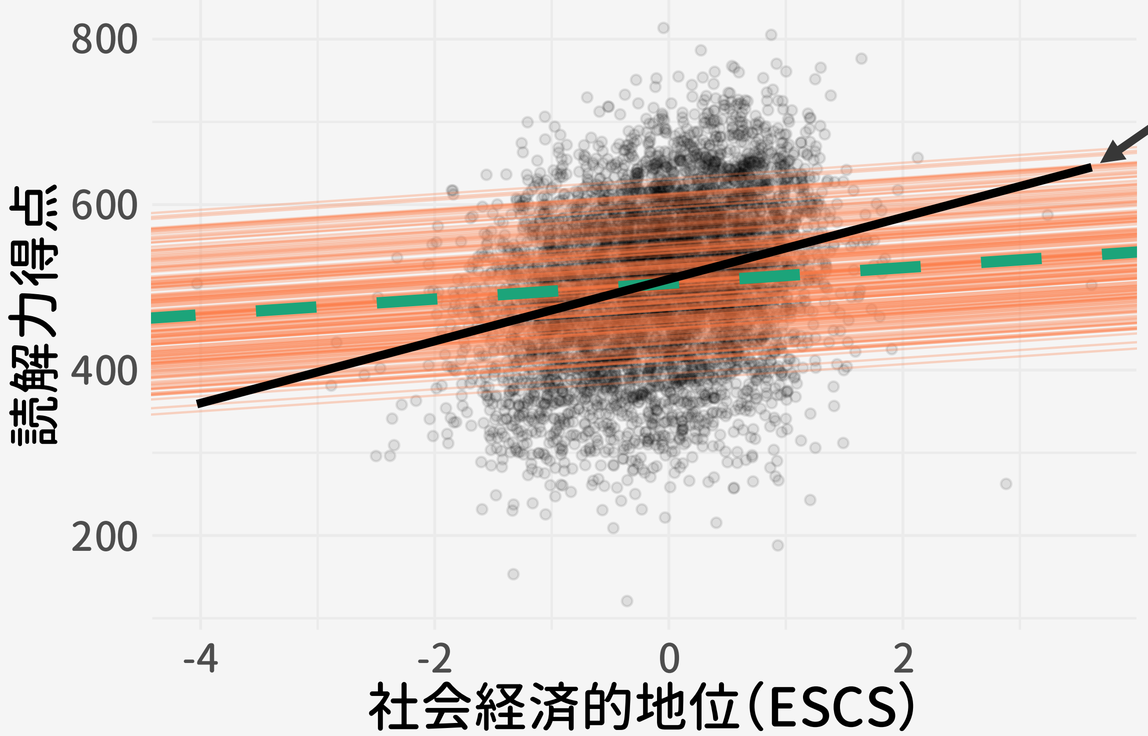
$$\hat{y}_{pg} = \underbrace{504.791 - 116.6}_{388.191} + 9.478\mathrm{ESCS}_{pg}$$

```
coef(model1)$cond$school_id
```

の出力をもとに計算した値です。

$$
\begin{aligned}
y_{pg} &= 504.791 + 9.478\mathrm{ESCS}_{pg} + u_{0g} + u_{pg} \\
&= 504.791 + 9.478\left[\overline{\mathrm{ESCS}}_{.g} + \left(\mathrm{ESCS}_{pg} - \overline{\mathrm{ESCS}}_{.g}\right)\right] + u_{0g}
\end{aligned}
$$

係数には2つの効果が混ざっている



マルチレベルではない回帰分析と比べると
傾きは小さくはなっている

▶ ランダム切片によって分離できている

推定された係数9.478は，それでも

- 集団レベル効果（ $\overline{\mathrm{ESCS}}_{.g}$ ）
- 個人レベル効果（ $\mathrm{ESCS}_{pg} - \overline{\mathrm{ESCS}}_{.g}$ ）

が混ざったもの

このままでは解釈が難しい！

$$
\begin{aligned}
y_{pg} &= \mu + \beta_{\text{ESCS}}\text{ESCS}_{pg} + u_{0g} + u_{pg} \\
&= \mu + \beta_{\text{ESCS}}\left[\overline{\text{ESCS}}_{.g} + \left(\text{ESCS}_{pg} - \overline{\text{ESCS}}_{.g}\right)\right] + u_{0g} + u_{pg} \\
&= \mu + \beta_{\text{ESCS}}\overline{\text{ESCS}}_{.g} + \beta_{\text{ESCS}}\left(\text{ESCS}_{pg} - \overline{\text{ESCS}}_{.g}\right) + u_{0g}
\end{aligned}
$$

係数を分けてしまおう

集団平均に対する回帰係数
▶ESCSが高い学校にいる人ほど…

集団平均からの偏差に対する回帰係数
▶学校内で,相対的にESCSが高い人ほど…

$$y_{pg} = \mu + \beta_{\overline{\text{ESCS}}}\overline{\text{ESCS}}_{.g} + \beta_{\text{ESCS}}\left(\text{ESCS}_{pg} - \overline{\text{ESCS}}_{.g}\right) + u_{0g}$$

(レベル1:個人) $y_{pg} = \beta_{0g} + \beta_{\text{ESCS}}\left(\text{ESCS}_{pg} - \overline{\text{ESCS}}_{.g}\right) + u_{pg}$

(レベル2:集団) $\beta_{0g} = \mu + \beta_{\overline{\text{ESCS}}}\overline{\text{ESCS}}_{.g} + u_{0g}$

あとは2つの説明変数を
別々に用意してあげるだけ

中心化した変数の用意

$$y_{pg} = \mu + \beta_{\overline{\text{ESCS}}}\overline{\text{ESCS}}_{\cdot g} + \beta_{\text{ESCS}}\left(\text{ESCS}_{pg} - \overline{\text{ESCS}}_{\cdot g}\right) + u_{0g}$$

(P8より) sch_escs_mean

集団平均中心化 (centering within cluster [CWC])

集団平均からの偏差に変換する ▶「集団内での相対的な値」として解釈できる

【集団平均中心化】

```
dat$escs_cwc <- dat$escs - dat$sch_escs_mean
```

以後，$\text{ESCS}_{pg} - \overline{\text{ESCS}}_{\cdot g} = \text{CWC}_{pg}$ と表します

準備が整いました

【2つのレベルの効果を分離したランダム切片回帰モデル】

もとの講義資料に準拠したナンバリングです

```
model3 <- glmmTMB(read_score ~ escs_cwc + sch_escs_mean + (1 ¦ school_id), data = dat)
```

【2つのレベルの効果を分離したランダム切片回帰モデルの推定結果】

```
summary(model3)
```

【変量効果】

```
Random effects:

Conditional model:
 Groups    Name        Variance Std.Dev.
 school_id (Intercept) 1461     38.23
 Residual              5557     74.54
Number of obs: 5766, groups:  school_id, 183
```

各変量効果の分布の分散・標準偏差

$u_{pg} \sim N(0, 74.54^2)$

$u_{0g} \sim N(0, 38.23^2)$

【固定効果】

```
Conditional model:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    518.060      3.132  165.43  < 2e-16 ***
escs_cwc         7.446      1.563    4.77 1.88e-06 ***
sch_escs_mean  126.569      8.153   15.53  < 2e-16 ***
```

固定効果の
推定値と検定

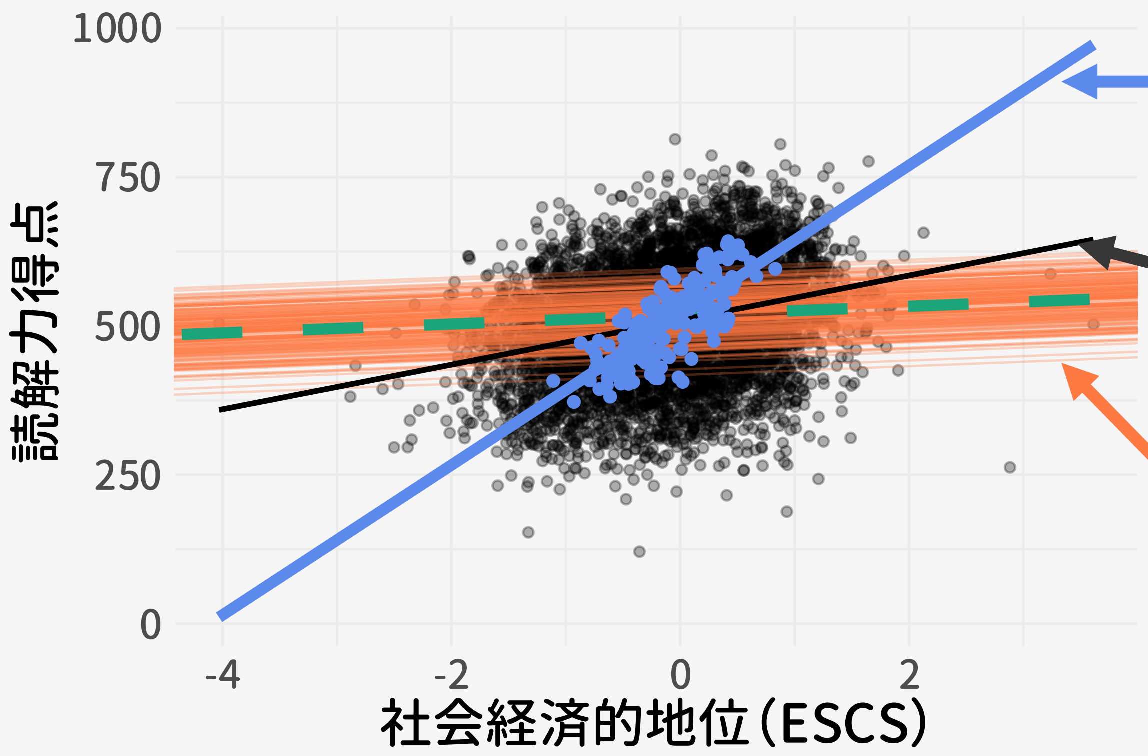
(レベル1:個人) $$y_{pg} = \beta_{0g} + 7.446\left(\mathrm{ESCS}_{pg} - \overline{\mathrm{ESCS}}_{.g}\right) + u_{pg}$$

(レベル2:集団) $$\beta_{0g} = 518.060 + 126.569\overline{\mathrm{ESCS}}_{.g} + u_{0g}$$

2つのレベルの効果をまとめて図示する

（レベル1：個人） $y_{pg} = \beta_{0g} + 7.446\mathrm{CWC}_{pg} + u_{pg}$ $u_{pg} \sim N(0, 74.54^2)$

（レベル2：集団） $\beta_{0g} = 518.060 + 126.569\overline{\mathrm{ESCS}}_{.g} + u_{0g}$ $u_{0g} \sim N(0, 38.23^2)$



【集団レベル効果】
≒集団ごとに計算した標本平均をもとに
回帰分析を行った結果

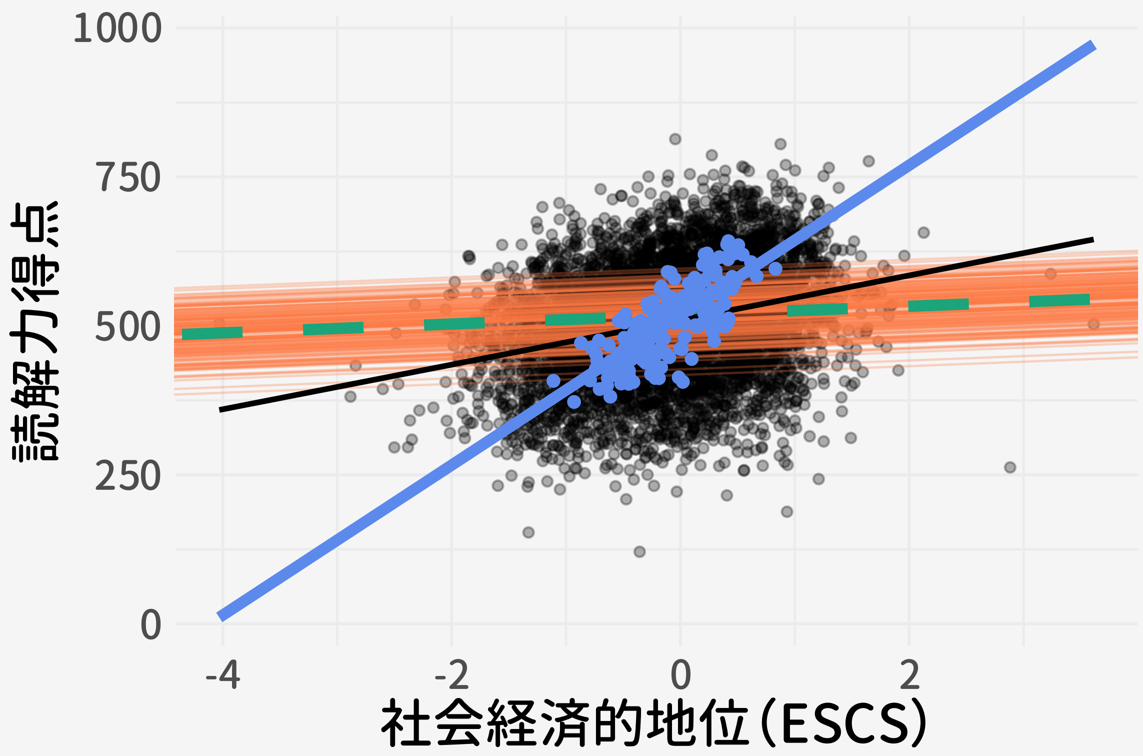
【非マルチレベルの効果】
＝集団レベルと個人レベルが混ざるため
傾きはその間に来る

【個人レベル効果】
＝集団平均中心化した値をもとに
回帰分析を行った結果

2つのレベルの効果の解釈&比較

（レベル1：個人） $y_{pg} = \beta_{0g} + 7.446\mathrm{CWC}_{pg} + u_{pg}$ $u_{pg} \sim N(0, 74.54^2)$

（レベル2：集団） $\beta_{0g} = 518.060 + 126.569\overline{\mathrm{ESCS}}_{.g} + u_{0g}$ $u_{0g} \sim N(0, 38.23^2)$



【集団レベル効果】
ESCSが平均的に高い学校ほど
読解力も平均的に高い傾向にある

【個人レベル効果】
学校内で相対的にESCSが高い生徒ほど
読解力も高い傾向にある

2つの係数を比べると……
【集団レベル効果】の係数のほうが
【個人レベル効果】の係数よりも明らかに大きい

学校の文脈効果が強く現れている

集団レベルの説明変数の追加

例 生徒の読解力には，ST比（教員1人あたり生徒数）も影響するのではないかと考えました。
少人数教育の効果を検証しようと思います。

とりあえず式に追加してみたら良い

$$y_{pg} = \mu + \beta_{\overline{\mathrm{ESCS}}} \overline{\mathrm{ESCS}}_{\cdot g} + \beta_{\mathrm{ESCS}} \mathrm{CWC}_{pg} + \beta_{\mathrm{ST}} \mathrm{ST}_{\cdot g} + u_{0g}$$

（レベル1：個人） $$y_{pg} = \beta_{0g} + \beta_{\mathrm{ESCS}} \mathrm{CWC}_{pg} + u_{pg}$$

（レベル2：集団） $$\beta_{0g} = \mu + \beta_{\overline{\mathrm{ESCS}}} \overline{\mathrm{ESCS}}_{\cdot g} + \beta_{\mathrm{ST}} \mathrm{ST}_{\cdot g} + u_{0g}$$

【集団レベル説明変数を追加したランダム切片回帰モデル】

```
model4 <- glmmTMB(read_score ~ escs_cwc + sch_escs_mean + s_t_ratio + (1 ¦ school_id),
                  data = dat)
summary(model4)
```

結果は省略します

4

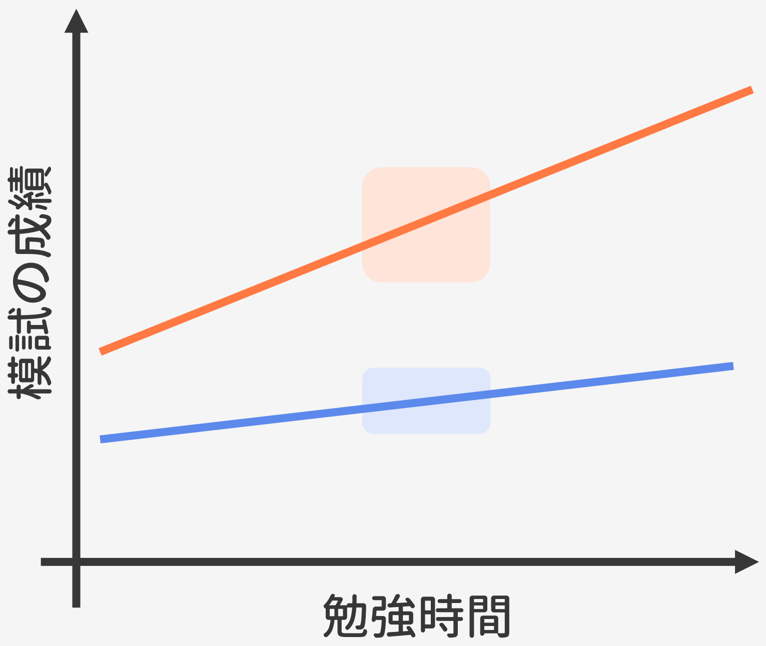
ランダム傾きモデル

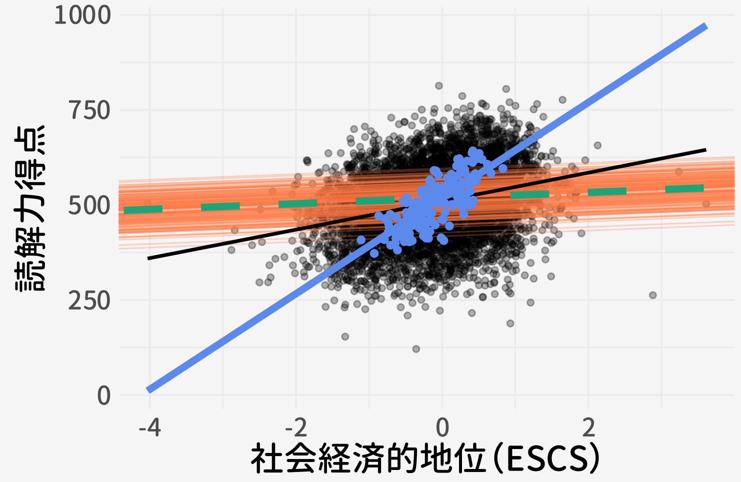
マルチレベルモデルの真骨頂へ

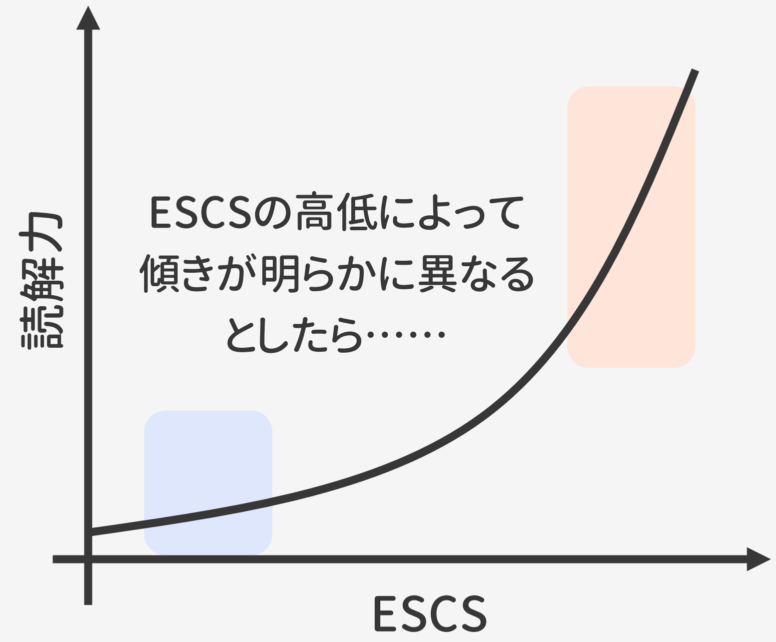
すべての集団で傾きが同じだと仮定していた

……が，もちろん常にそうとは限らない

【例】勉強時間と模試の成績







まずは固定効果として考えてみる

$$y_{pg} = \underbrace{\beta_0 + \beta_{\text{ESCS}}\text{ESCS}_{pg}}_{\text{学校001の係数（基準）}} + \underbrace{\beta_{0,002}D_{002} + \beta_{0,003}D_{003} + \cdots}_{\text{学校ダミーごとの切片の差}} + \underbrace{\beta_{\text{ESCS},002}D_{002}\text{ESCS}_{pg} + \beta_{\text{ESCS},003}D_{003}\text{ESCS}_{pg} + \cdots}_{\text{学校ダミーごとの傾きの差}} + u_{pg}$$

学校002の係数： $\beta_{0,002}D_{002}$, $\beta_{\text{ESCS},002}D_{002}\text{ESCS}_{pg}$

学校003の係数： $\beta_{0,003}D_{003}$, $\beta_{\text{ESCS},003}D_{003}\text{ESCS}_{pg}$

学校（ダミー）ごとに場合分けして書くならば

$$y_{pg} = \begin{cases} \beta_0 + \beta_{\text{ESCS}}\,\text{ESCS}_{pg} + u_{pg} & \text{（学校001）} \\ (\beta_0 + \beta_{0,002}) + (\beta_{\text{ESCS}} + \beta_{\text{ESCS},002})\,\text{ESCS}_{pg} + u_{pg} & \text{（学校002）} \\ (\beta_0 + \beta_{0,003}) + (\beta_{\text{ESCS}} + \beta_{\text{ESCS},003})\,\text{ESCS}_{pg} + u_{pg} & \text{（学校003）} \\ \vdots & \end{cases}$$

Rでダミー変数を用いた回帰分析として実行

説明変数は「（ダミー化された）学校 D」と「ESCS_{pg}」の2つ

$$+\beta_{\mathrm{ESCS},002} D_{002} \mathrm{ESCS}_{pg} +$$

学校ダミー

ここが交互作用

交互作用項は*で表すことができる

【学校ごとのread_scoreの切片とESCSの傾きの差を評価する回帰分析】

```
summary(lm(read_score ~ school_id * escs, data = dat))
```

```
（前略）
Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)        6.170e+02  1.657e+01  37.234  < 2e-16 ***
school_id002      -1.458e+02  2.225e+01  -6.554 6.13e-11 ***
school_id003      -1.552e+02  2.172e+01  -7.149 9.91e-13 ***
school_id004      -1.075e+02  2.266e+01  -4.744 2.15e-06 ***
（中略）
school_id185      -1.484e+02  2.295e+01  -6.468 1.08e-10 ***
escs               9.096e+00  2.317e+01   0.393 0.694693
school_id002:escs -1.346e+01  3.043e+01  -0.442 0.658330
school_id003:escs  4.823e+01  2.779e+01   1.736 0.082693 .
school_id004:escs -1.421e+00  2.890e+01  -0.049 0.960792
（後略）
```

$$y_{pg} = \beta_0 + \beta_{\mathrm{ESCS}} \mathrm{ESCS}_{pg} + \beta_{0,002} D_{002} + \beta_{0,003} D_{003} + \cdots + \beta_{\mathrm{ESCS},002} D_{002} \mathrm{ESCS}_{pg} + \beta_{\mathrm{ESCS},003} D_{003} \mathrm{ESCS}_{pg} + \cdots + u_{pg}$$

【各学校の切片と傾き】

school_id	切片		傾き	
001	β_0	617.0	β_{ESCS}	9.096
002	$\beta_0 + \beta_{002}$	471.2	$\beta_{\mathrm{ESCS}} + \beta_{\mathrm{ESCS},002}$	-4.364
003	$\beta_0 + \beta_{003}$	461.8	$\beta_{\mathrm{ESCS}} + \beta_{\mathrm{ESCS},003}$	13.919
⋮				

ランダム切片モデルのときと同じように考えたらOK

(固定効果ver.)

$$y_{pg} = \beta_0 + \beta_{\mathrm{ESCS}}\mathrm{ESCS}_{pg} + \beta_{0,002}D_{002} + \beta_{0,003}D_{003} + \cdots$$
$$+\beta_{\mathrm{ESCS},002}D_{002}\mathrm{ESCS}_{pg} + \beta_{\mathrm{ESCS},003}D_{003}\mathrm{ESCS}_{pg} + \cdots + u_{pg}$$

学校001の切片と傾き

一つ一つの係数(群間の差) には 関心がないので

残差

$$y_{pg} = \beta_0 + \beta_{\mathrm{ESCS}}\mathrm{ESCS}_{pg} + u_{0g} + u_{\mathrm{ESCS},g}\mathrm{ESCS}_{pg} + u_{pg}$$

特定の群を基準にするのはなんだか気持ち悪いので

切片と傾きの全体平均

(変量効果ver.)

$$y_{pg} = \mu_0 + \mu_{\mathrm{ESCS}}\mathrm{ESCS}_{pg} + u_{0g} + u_{\mathrm{ESCS},g}\mathrm{ESCS}_{pg} + u_{pg}$$
$$= (\mu_0 + u_{0g}) + (\mu_{\mathrm{ESCS}} + u_{\mathrm{ESCS},g})\mathrm{ESCS}_{pg} + u_{pg}$$

切片と傾きをそれぞれ変量効果にしたら良いだけ

もう少し詳しく定式化

$$y_{pg} = (\mu_0 + u_{0g}) + (\mu_{\mathrm{ESCS}} + u_{\mathrm{ESCS},g})\mathrm{ESCS}_{pg} + u_{pg}$$

レベル別に式を書くと

(レベル1：個人) $y_{pg} = \beta_{0g} + \beta_{\mathrm{ESCS},g}\mathrm{ESCS}_{pg} + u_{pg}$

(レベル2：集団) $\beta_{0g} = \mu_0 + u_{0g}$

$\beta_{\mathrm{ESCS},g} = \mu_{\mathrm{ESCS}} + u_{\mathrm{ESCS},g}$

【仮定】変量効果は通常(多変量)正規分布に従う(ものとみなす)

(レベル1：個人) $u_{pg} \sim N(0, \sigma_{pg}^2)$

(レベル2：集団) $\begin{bmatrix} u_{0g} \\ u_{\mathrm{ESCS},g} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{0g}^2 & \sigma_{(0g)(\mathrm{ESCS},g)} \\ \sigma_{(0g)(\mathrm{ESCS},g)} & \sigma_{\mathrm{ESCS},g}^2 \end{bmatrix}\right)$

- 同じレベルの変量効果間は相関を認める
- 異なるレベルの変量効果間は独立

群(集団) g の切片と傾きを
それぞれ $\beta_{0g}, \beta_{\mathrm{ESCS},g}$ と表すと，

(レベル1：個人) $y_{pg} \sim N(\beta_{0g} + \beta_{\mathrm{ESCS},g}\mathrm{ESCS}_{pg}, \sigma_{pg}^2)$

(レベル2：集団) $\begin{bmatrix} \beta_{0g} \\ \beta_{\mathrm{ESCS},g} \end{bmatrix} \sim MVN\left(\begin{bmatrix} \mu_0 \\ \mu_{\mathrm{ESCS}} \end{bmatrix}, \begin{bmatrix} \sigma_{0g}^2 & \sigma_{(0g)(\mathrm{ESCS},g)} \\ \sigma_{(0g)(\mathrm{ESCS},g)} & \sigma_{\mathrm{ESCS},g}^2 \end{bmatrix}\right)$

$$y_{pg} = (\mu_0 + u_{0g}) + (\mu_{\mathrm{ESCS}} + u_{\mathrm{ESCS},g})\mathrm{ESCS}_{pg} + u_{pg}$$

回帰係数に個人レベルと集団レベルの効果が混ざっているのがイヤなら

説明変数から「集団レベル」の成分を取り除こう ▶ 集団平均中心化

$$y_{pg} = (\mu_0 + u_{0g}) + (\mu_{\mathrm{ESCS}} + u_{\mathrm{ESCS},g})\underbrace{\mathrm{CWC}_{pg}}_{\texttt{escs\_cwc}} + u_{pg}$$

$$= \mu_0 + \mu_{\mathrm{ESCS}}\underbrace{\mathrm{CWC}_{pg}}_{\texttt{escs\_cwc}} + u_{0g} + u_{\mathrm{ESCS},g}\underbrace{\mathrm{CWC}_{pg}}_{\texttt{escs\_cwc}} + u_{pg}$$

formulaに書き下すと

```
read_score ~ 1 + escs_cwc + (1 + escs_cwc | school_id)
```

【ランダム傾き回帰モデル】 もとの講義資料に準拠したナンバリングです

```
model5 <- glmmTMB(read_score ~ escs_cwc + (escs_cwc | school_id), data = dat)
```

結果をみる

【ランダム傾き回帰モデルの推定結果】

```
summary(model5)
```

【変量効果】

```
Random effects:

Conditional model:
 Groups    Name        Variance Std.Dev. Corr
 school_id (Intercept) 3626.8   60.22
           escs_cwc     106.9   10.34    -0.09
 Residual              5512.9   74.25
Number of obs: 5766, groups:  school_id, 183
```

(Corr)

$$\sigma_{(0g)(\mathrm{ESCS},g)} = r^2_{(0g)(\mathrm{ESCS},g)} \sigma_{0g} \sigma_{\mathrm{ESCS},g} \approx -56.05$$

各変量効果の分布の分散・標準偏差

$$u_{pg} \sim N(0, 74.25^2)$$

$$\begin{bmatrix} u_{0g} \\ u_{\mathrm{ESCS},g} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 60.22^2 & -56.05 \\ -56.05 & 10.34^2 \end{bmatrix} \right)$$

【固定効果】

```
Conditional model:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  503.695      4.563  110.40  < 2e-16 ***
escs_cwc       7.396      1.751    4.22  2.4e-05 ***
```

固定効果の推定値と検定

(レベル1:個人) $y_{pg} = \beta_{0g} + \beta_{\mathrm{ESCS},g}\mathrm{CWC}_{pg} + u_{pg}$

(レベル2:集団) $\beta_{0g} = 503.695 + u_{0g}, \quad \beta_{\mathrm{ESCS},g} = 7.396 + u_{\mathrm{ESCS},g}$

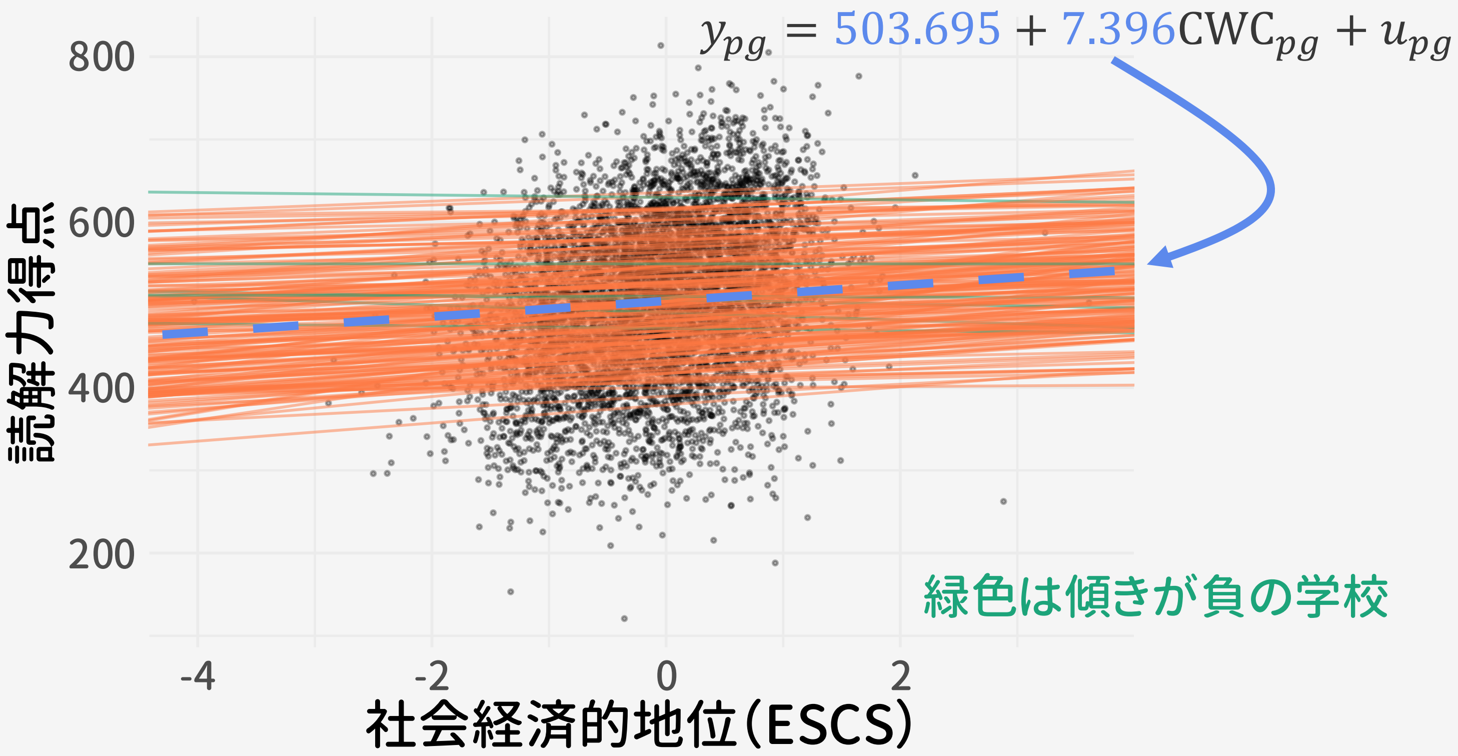
事後的に計算することはもちろん可能です

【集団の平均パラメータを事後的に計算する】

```
coef(model5)$cond$school_id
```

```
    (Intercept)  escs_cwc
001    616.3353  6.187427
002    474.2600  5.242309
003    463.3115 23.355328
004    505.6372  7.446572
005    413.0458  4.324954
006    607.0751  8.663129
007    584.1511  7.371261
008    416.7438 10.494779
009    503.3077  6.174089
010    379.0714 10.940293
------(以下省略)------
```

【集団ごとの回帰直線を引いてみると……】



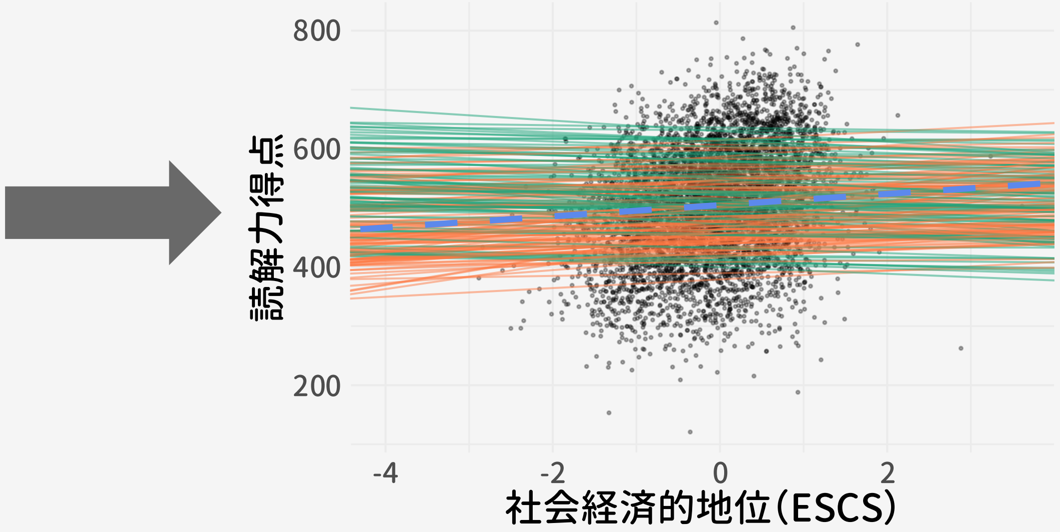
glmmTMB (lme4)では,変量効果は平均0として推定される

- 傾きの全体平均 μ_{ESCS} を推定する場合には,**書き忘れないように注意!**

> ※特に切片は書かなくても良いために,傾きのほうは忘れやすいかも

【傾きの全体平均を入れ忘れたランダム傾き回帰モデル】

```
model5_wrong <- glmmTMB(read_score ~ (escs_cwc | school_id), data = dat)
```

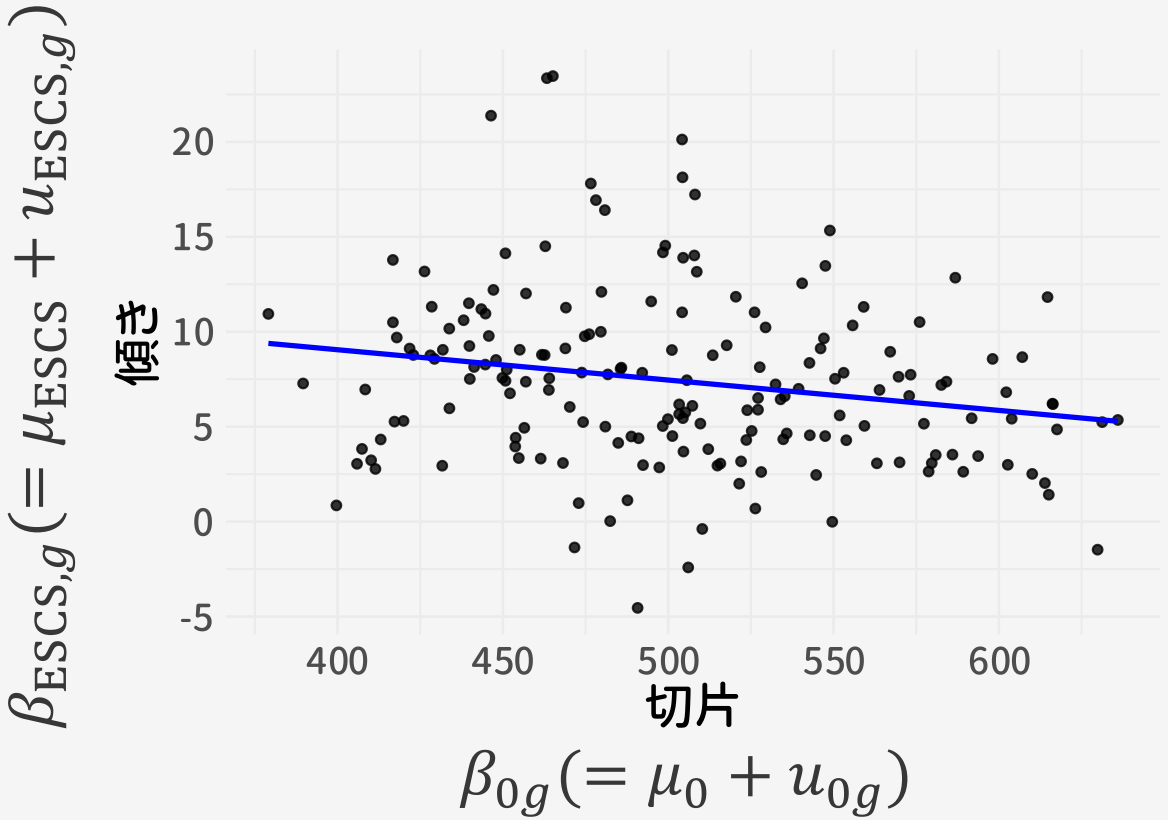


前ページと比べると
明らかに傾きが負の学校(緑色)が増えている

変量効果どうしの相関を解釈する

■ 切片と傾きの変量効果には負の相関が見られた $\begin{bmatrix} u_{0g} \\ u_{\mathrm{ESCS},g} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 60.22^2 & -56.05 \\ -56.05 & 10.34^2 \end{bmatrix}\right)$

【変量効果どうしの散布図と回帰直線を引いてみると……】



切片が大きい集団ほど，
escs_cwcがread_scoreに与える正の効果が
小さくなっていく

【考えられるメカニズム】　異論は認めます

- 切片が大きい＝学力が高い学校は，選抜効果によってESCSの効果が小さくなる。
- 切片が小さい＝学力が高くない集団は選抜効果が弱いためにESCSの効果がまだ残っている。

集団レベルの説明変数は学校ごとの切片や傾きの違いを説明できるか?

前ページの想像にのっとれば「学校ごとの平均学力」あるいは「学校ごとの平均ESCS」

▲これらが高いほど,切片は大きくなる一方で,傾きは小さくなると言えるのか?

回帰式に説明変数を組み込んでみよう

学校 g のESCSの平均値 $\overline{\text{ESCS}}_{\cdot g}$ を傾きと切片の説明変数にする

sch_escs_mean

P60の式とよく見比べて,
どこがどう変化したかを確認してください

(レベル1:個人) $y_{pg} = \beta_{0g} + \beta_{\text{ESCS},g}\text{CWC}_{pg} + u_{pg}$

(レベル2:集団) $\beta_{0g} = \mu_0 + \beta_{0,\overline{\text{ESCS}}}\overline{\text{ESCS}}_{\cdot g} + u_{0g}$

$\beta_{\text{ESCS},g} = \mu_{\text{ESCS}} + \beta_{1,\overline{\text{ESCS}}}\overline{\text{ESCS}}_{\cdot g} + u_{\text{ESCS},g}$

係数$\beta_{0,\overline{\text{ESCS}}}$ が正ならば
$\overline{\text{ESCS}}_{\cdot g}$ が高い学校ほど
切片 β_{0g} が高いことになる

ただちょっと交互作用が厄介です

一旦落ち着いて,レベル別の式を1つにまとめていきましょう

（レベル1：個人） $y_{pg} = \beta_{0g} + \beta_{\mathrm{ESCS},g}\mathrm{CWC}_{pg} + u_{pg}$

（レベル2：集団） $\beta_{0g} = \mu_0 + \beta_{0,\overline{\mathrm{ESCS}}}\overline{\mathrm{ESCS}}_{.g} + u_{0g}$

$\beta_{\mathrm{ESCS},g} = \mu_{\mathrm{ESCS}} + \beta_{1,\overline{\mathrm{ESCS}}}\overline{\mathrm{ESCS}}_{.g} + u_{\mathrm{ESCS},g}$

$$y_{pg} = \mu_0 + \beta_{0,\overline{\mathrm{ESCS}}}\overline{\mathrm{ESCS}}_{.g} + u_{0g} + \left(\mu_{\mathrm{ESCS}} + \beta_{1,\overline{\mathrm{ESCS}}}\overline{\mathrm{ESCS}}_{.g} + u_{\mathrm{ESCS},g}\right)\mathrm{CWC}_{pg} + u_{pg}$$

$$= \underbrace{\mu_0}_{\text{切片}} + \underbrace{\beta_{0,\overline{\mathrm{ESCS}}}\overline{\mathrm{ESCS}}_{.g}}_{\text{集団レベル}} + \underbrace{\mu_{\mathrm{ESCS}}\mathrm{CWC}_{pg}}_{\text{個人レベル}} + \underbrace{\beta_{1,\overline{\mathrm{ESCS}}}\overline{\mathrm{ESCS}}_{.g}\mathrm{CWC}_{pg}}_{\text{レベル間交互作用}} + \underbrace{u_{0g} + u_{\mathrm{ESCS},g}\mathrm{CWC}_{pg} + u_{pg}}_{\text{変量効果（残差）}}$$

→ **sch_escs_meanとescs_cwcの交互作用項**を固定効果に含めたら表現できる！

formulaを指定していよいよ実行してみる

$$y_{pg} \boxed{= \mu_0} + \boxed{\beta_{0,\overline{\mathrm{ESCS}}}\overline{\mathrm{ESCS}}_{.g}} + \boxed{\mu_{\mathrm{ESCS}}\mathrm{CWC}_{pg}} + \boxed{\beta_{1,\overline{\mathrm{ESCS}}}\overline{\mathrm{ESCS}}_{.g}\mathrm{CWC}_{pg}} + \boxed{u_{0g} + u_{\mathrm{ESCS},g}\mathrm{CWC}_{pg} + u_{pg}}$$

切片　集団レベル　個人レベル　レベル間交互作用　変量効果（残差）

formulaに
書き下すと

```
read_score ~ 1 + sch_escs_mean + escs_cwc + sch_escs_mean:escs_cwc + (1 + escs_cwc | school_id)
```

主効果と交互作用をすべて推定する場合
以下のように書ける

```
read_score ~ 1 + sch_escs_mean * escs_cwc + (1 + escs_cwc | school_id)
```

【レベル間交互作用を含めたランダム傾き回帰モデル】

```
model6 <- glmmTMB(read_score ~ sch_escs_mean * escs_cwc + (escs_cwc | school_id),
                  data = dat)
```

結果をみる

```
summary(model6)
```

【レベル間交互作用を含めたランダム傾き回帰モデルの推定結果】

【変量効果】

```
Random effects:

Conditional model:
 Groups    Name        Variance Std.Dev. Corr
 school_id (Intercept) 1463     38.24
           escs_cwc     107     10.34    -0.17
 Residual              5513     74.25
Number of obs: 5766, groups:  school_id, 183
```

各変量効果の分布の分散・標準偏差

$$u_{pg} \sim N(0, 74.25^2)$$

$$\begin{bmatrix} u_{0g} \\ u_{\mathrm{ESCS},g} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 38.24^2 & -67.22 \\ -67.22 & 10.34^2 \end{bmatrix}\right)$$

【固定効果】

```
Conditional model:
                     Estimate Std. Error z value Pr(>|z|)
(Intercept)          518.0581     3.1314  165.44  < 2e-16 ***
escs_cwc               7.4769     1.8728    3.99 6.54e-05 ***
sch_escs_mean        126.5833     8.1517   15.53  < 2e-16 ***
escs_cwc:sch_escs_mean 0.4573     4.8602    0.09    0.925
```

固定効果の推定値と検定

$\overline{ESCS}_{.g}$ は傾きの違いをほぼ説明できない

（レベル2の切片） $$\beta_{0g} = 518.0581 + 126.5833\overline{\mathrm{ESCS}}_{.g} + u_{0g}$$

（レベル2の傾き） $$\beta_{\mathrm{ESCS},g} = 7.4769 + 0.4573\overline{\mathrm{ESCS}}_{.g} + u_{\mathrm{ESCS},g}$$

5

モデル比較

モデルはつながっていた，かも

ランダム傾きを含めるべきかどうかをどう判断する?

ICCではちょっと厳しそう……

モデルが「入れ子」になっていることに注目

【model6】
(レベル1:個人) $y_{pg} = \beta_{0g} + \beta_{\mathrm{ESCS},g}\mathrm{CWC}_{pg} + u_{pg}$
(レベル2:集団) $\beta_{0g} = \mu_0 + \beta_{0,\overline{\mathrm{ESCS}}}\overline{\mathrm{ESCS}}_{.g} + u_{0g}$
$\beta_{\mathrm{ESCS},g} = \mu_{\mathrm{ESCS}} + \beta_{1,\overline{\mathrm{ESCS}}}\overline{\mathrm{ESCS}}_{.g} + u_{\mathrm{ESCS},g}$

$\beta_{0,\overline{\mathrm{ESCS}}} = \beta_{1,\overline{\mathrm{ESCS}}} = 0$

【model5】
(レベル1:個人) $y_{pg} = \beta_{0g} + \beta_{\mathrm{ESCS},g}\mathrm{CWC}_{pg} + u_{pg}$
(レベル2:集団) $\beta_{0g} = \mu_0 + u_{0g}$
$\beta_{\mathrm{ESCS},g} = \mu_{\mathrm{ESCS}} + u_{\mathrm{ESCS},g}$

ランダム傾きモデル

$\beta_{1,\overline{\mathrm{ESCS}}} = u_{\mathrm{ESCS},g} = 0$
$(\beta_{\mathrm{ESCS},g} = \beta_{\mathrm{ESCS}} = \mu_{\mathrm{ESCS}})$

【model3】
(レベル1:個人) $y_{pg} = \beta_{0g} + \beta_{\mathrm{ESCS}}\mathrm{CWC}_{pg} + u_{pg}$
(レベル2:集団) $\beta_{0g} = \mu + \beta_{\overline{\mathrm{ESCS}}}\overline{\mathrm{ESCS}}_{.g} + u_{0g}$

ランダム傾きを含めるべきかどうかをどう判断する?

ICCではちょっと厳しそう……

モデルが「入れ子」になっていることに注目

ランダム切片モデル

【model3】（レベル1：個人） $y_{pg} = \beta_{0g} + \beta_{\mathrm{ESCS}}\mathrm{CWC}_{pg} + u_{pg}$

（レベル2：集団） $\beta_{0g} = \mu + \beta_{\overline{\mathrm{ESCS}}}\overline{\mathrm{ESCS}}_{\cdot g} + u_{0g}$

$\beta_{\mathrm{ESCS}} = \beta_{\overline{\mathrm{ESCS}}}$

$\mathrm{CWC}_{pg} + \overline{\mathrm{ESCS}}_{\cdot g} = \mathrm{ESCS}_{pg}$ のため

【model1】（レベル1：個人） $y_{pg} = \beta_{0g} + \beta_{\mathrm{ESCS}}\mathrm{CWC}_{pg} + u_{pg}$

（レベル2：集団） $\beta_{0g} = \mu + \beta_{\mathrm{ESCS}}\overline{\mathrm{ESCS}}_{\cdot g} + u_{0g}$

$\beta_{\mathrm{ESCS}} = 0$

【model0】（レベル1：個人） $y_{pg} = \beta_{0g} + u_{pg}$

（レベル2：集団） $\beta_{0g} = \mu + u_{0g}$

「入れ子」の関係にあるモデル同士を比較しよう

尤度比検定(あるいは情報量規準)による比較

「入れ子」の関係にあるモデルどうしの比較に使える

統計モデルは，パラメータ(回帰係数)が増えるほどデータの当てはまりが良くなる

▶ **追加したパラメータ数に見合うだけの尤度の改善が見られるか**を検定する方法

H_0: 2つのモデルの尤度に差はない
▶ 棄却されない場合はシンプルな方にする

Rだと簡単に実行可能

【尤度比検定(ついでに情報量規準も)】

```
anova(model0, model1, model3, model6)
```

```
Models:
model0: read_score ~ (1 ¦ school_id), zi=~0, disp=~1
model1: read_score ~ escs + (1 ¦ school_id), zi=~0, disp=~1
model3: read_score ~ escs_cwc + sch_escs_mean + (1 ¦ school_id), zi=~0, disp=~1
model6: read_score ~ sch_escs_mean * escs_cwc + (escs_cwc ¦ school_id), zi=~0, disp=~1
       Df   AIC   BIC logLik deviance    Chisq Chi Df Pr(>Chisq)
model0  3 66671 66691 -33332    66665
model1  4 66636 66663 -33314    66628  36.6917      1  1.384e-09 ***
model3  5 66498 66531 -33244    66488 139.9982      1  < 2.2e-16 ***
model6  8 66498 66551 -33241    66482   5.9989      3     0.1117
```

この場合はmodel3が良い?

尤度比検定の注意点

複数のパラメータが一気に追加された場合は「総合的な尤度の改善」を見る

model3とmodel6の比較で言えば，**「有効なパラメータ」と「ほぼ意味のないパラメータ」が同時に投入された**結果，総合的には有意にならなかった可能性

> ただ今回のデータでmodel3にパラメータを1つずつ追加すると，そもそも推定が安定しない

正確な「尤度」が必要なので，最尤法で推定する必要がある

glmmTMBではデフォルトで最尤法 (ML) を用いるのであまり気にする必要はない

lme4では**デフォルトで制限付き最尤法 (REML) を用いる**ので，そのままではダメ

変量効果の分散をより頑健に推定するための方法。
集団の数が少ない場合には，通常の最尤法では
変量効果の分散を過小推定してしまう傾向がある。

【REMLではなくMLで推定する】

```
lme4::lmer(formula, REML = FALSE, …)
```

別の視点：レベルごとの説明変数の分散説明率

最後には様々な視点から総合的に判断しましょう

ランダム効果の分散の減少率によって評価できる

		σ^2_{pg}		σ^2_{0g}		$\sigma^2_{ESCS,g}$	
モデル	formula	推定値	減少率	推定値	減少率	推定値	減少率
model0	(1 ¦ school_id)	74.70^2	(BL)	60.20^2	(BL)	—	—
model1	escs + (1 ¦ school_id)	74.56^2	0.37%	57.56^2	8.58%	—	—
model3	escs_cwc + sch_escs_mean + (1 ¦ school_id)	74.54^2	0.43%	38.23^2	59.67%	—	—
model5	escs_cwc + (escs_cwc ¦ school_id)	74.25^2	1.20%	60.22^2	0.00%	10.34^2	(BL)
model6	sch_escs_mean * escs_cwc + (escs_cwc ¦ school_id)	74.25^2	1.20%	38.24^2	59.65%	10.34^2	0.00%

(レベル1：個人) $y_{pg} = \beta_{0g} + \beta_{\mathrm{ESCS},g}\mathrm{CWC}_{pg} + u_{pg}$

(レベル2：集団) $\beta_{0g} = \mu_0 + \beta_{0,\overline{\mathrm{ESCS}}}\overline{\mathrm{ESCS}}_{\cdot g} + u_{0g}$

【model6】 $\beta_{\mathrm{ESCS},g} = \mu_{\mathrm{ESCS}} + \beta_{1,\overline{\mathrm{ESCS}}}\overline{\mathrm{ESCS}}_{\cdot g} + u_{\mathrm{ESCS},g}$

レベルごとに分けた各回帰式における
決定係数のようなもの

▶ **分散が大きく減少する説明変数は投入する価値がある！**

今回で言えばsch_escs_meanは有効

6

マルチレベル一般化線形モデル

またの名をGLMM

被説明変数のタイプによってはトンチンカンな回帰直線を引くことがある

例 escsが高い生徒ほど大学進学を希望するかを調べることにしました。
被説明変数のwants_univは，進学希望者は1，それ以外は0を取る二値変数です。

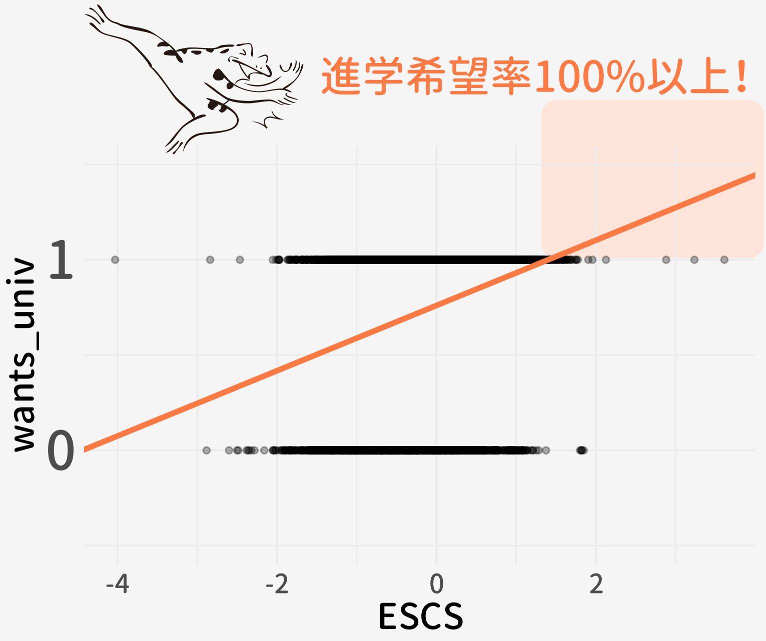
【ふつうの回帰分析】

```
summary(lm(wants_univ ~ escs, data = dat))
```

【結果の係数の部分だけ】

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.759881   0.005570   136.4  <2e-16 ***
escs        0.171119   0.007607    22.5  <2e-16 ***
```

$$\hat{y}_p = 0.760 + 0.171\mathrm{ESCS}_p$$



回帰分析の改良

$$\hat{y}_p = \beta_0 + \beta_{\text{ESCS}}\text{ESCS}_p$$

- ただの直線ではみ出してしまう
  - ▶ [0,1] からはみ出さないようにいい感じの変換をしよう!

たいてい「こういう場合にはこの変換」という定石があります

- このケースで有名なのはロジスティック変換 $g(x) = \dfrac{\exp(x)}{1+\exp(x)}$

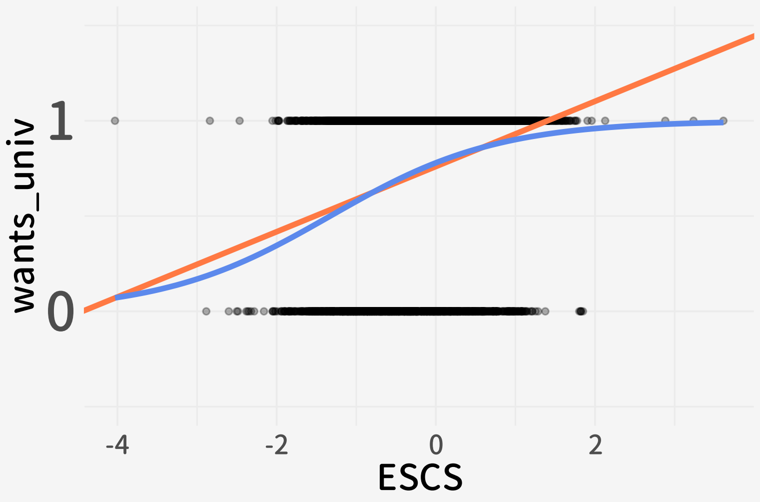
$$\hat{y}_p = g\left(\beta_0 + \beta_{\text{ESCS}}\text{ESCS}_p\right) = \frac{\exp(\beta_0 + \beta_{\text{ESCS}}\text{ESCS}_p)}{1+\exp\left(\beta_0 + \beta_{\text{ESCS}}\text{ESCS}_p\right)}$$

式変形させると

$$\log\frac{\hat{y}_p}{1-\hat{y}_p} = \boxed{\beta_0 + \beta_{\text{ESCS}}\text{ESCS}_p}$$

対数オッズ

もとの回帰分析と同じ!



回帰分析の一般化

説明変数による
予測値の変動

「大体」の予測値 ＋ 予測とのズレ

回帰分析における $\beta_0 + \beta_1 x_p$ の形を残しつつ…

ふつうの回帰分析

- 被説明変数 y_p が予測値と誤差の和

$$\hat{y}_p = \beta_0 + \beta_1 x_p$$

一般線形モデルとも呼ばれます

$$y_p \sim N(\hat{y}_p, \sigma^2)$$

【回帰係数 β_1 の意味】
▶ 説明変数 x_p が大きいほど
被説明変数 y_p も大きくなるか

ロジスティック回帰分析

- 予測確率 $\hat{y}_p$ について

$$\hat{y}_p = \frac{\exp(\beta_0 + \beta_1 x_p)}{1 + \exp(\beta_0 + \beta_1 x_p)}$$

$$y_p \sim Bernoulli(\hat{y}_p)$$

【回帰係数 β_1 の意味】
▶ 説明変数 x_p が大きいほど
予測確率 $\hat{y}_p$ も大きくなるか

ポアソン回帰分析

- 期待値 $\hat{y}_p$ について

$$\hat{y}_p = \exp(\beta_0 + \beta_1 x_p)$$

$$y_p \sim Poisson(\hat{y}_p)$$

【回帰係数 β_1 の意味】
▶ 説明変数 x_p が大きいほど
期待値 $\hat{y}_p$ も大きくなるか

全部まとめて **一般化線形モデル** と呼ばれます
Generalized Linear Model（GLM）

もちろんこれら以外にも
様々な確率分布に基づくGLMがあります

まずは固定効果として考えてみる

P26と比べると，左辺が変わっただけで基本的な考え方は同じ

$$\log \frac{\hat{y}_{pg}}{1-\hat{y}_{pg}} = \underbrace{\beta_0}_{\text{学校001の係数（基準）}} + \underbrace{\beta_{002}D_{002} + \beta_{003}D_{003} + \cdots + \beta_{185}D_{185}}_{\text{学校ダミーごとの切片の差}}$$

学校（ダミー）ごとに場合分けして書くならば

$$\log \frac{\hat{y}_{pg}}{1-\hat{y}_{pg}} = \begin{cases} \beta_0 & \text{（学校001）} \\ (\beta_0 + \beta_{0,002}) & \text{（学校002）} \\ (\beta_0 + \beta_{0,003}) & \text{（学校003）} \\ \vdots & \end{cases}$$

または

$$\hat{y}_{pg} = \begin{cases} \dfrac{\exp(\beta_0)}{1+\exp(\beta_0)} & \text{（学校001）} \\ \dfrac{\exp(\beta_0 + \beta_{0,002})}{1+\exp(\beta_0 + \beta_{0,002})} & \text{（学校002）} \\ \dfrac{\exp(\beta_0 + \beta_{0,003})}{1+\exp(\beta_0 + \beta_{0,003})} & \text{（学校003）} \\ \vdots & \end{cases}$$

なのでRでもだいたいおなじ

左辺がどのように変形しているかを示す引数

$$\log \frac{\hat{y}_{pg}}{1-\hat{y}_{pg}}$$

lm( )の代わりにglm( )を使えば良いだけ

【学校ごとの進学率の差を評価する一般化線形モデル】

```
summary(glm(wants_univ ~ school_id, data = dat, family=binomial(link="logit")))
```

```
Coefficients:
              Estimate Std. Error z value Pr(>¦z¦)
(Intercept)   1.757e+01  6.687e+02   0.026    0.979
school_id002 -1.817e+01  6.687e+02  -0.027    0.978
school_id003 -1.698e+01  6.687e+02  -0.025    0.980
school_id004 -1.775e+01  6.687e+02  -0.027    0.979
school_id005 -1.776e+01  6.687e+02  -0.027    0.979
school_id006 -1.068e-07  9.599e+02   0.000    1.000
school_id007 -1.479e+01  6.687e+02  -0.022    0.982
school_id008 -1.719e+01  6.687e+02  -0.026    0.979
（後略）
```

$$\log \frac{\hat{y}_{pg}}{1-\hat{y}_{pg}} = \beta_0 + \beta_{002}D_{002} + \beta_{003}D_{003} + \cdots + \beta_{185}D_{185}$$

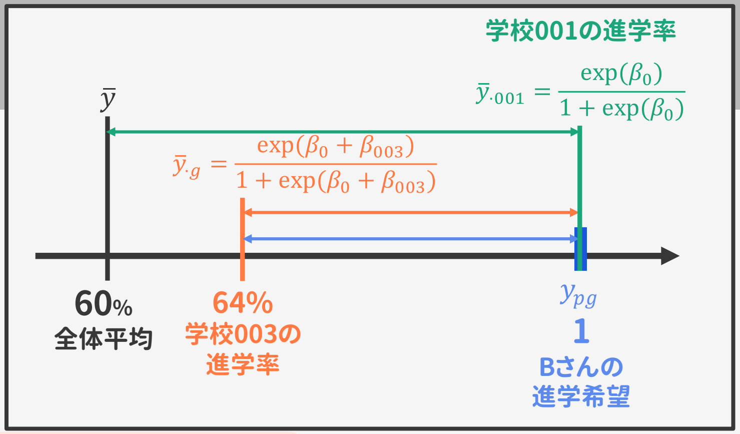
【各学校の推定値と進学率】

school_id	パラメータ	推定値	進学率
001	β_0	17.57	1.000
002	$\beta_0 + \beta_{002}$	-0.60	0.354
003	$\beta_0 + \beta_{003}$	0.59	0.643
⋮	⋮	⋮	

（学校002の場合） $\hat{y}_{pg} = \dfrac{\exp(\beta_0 + \beta_{002})}{1+\exp(\beta_0 + \beta_{002})} = \dfrac{\exp(17.57 - 18.17)}{1+\exp(17.57 - 18.17)} = \dfrac{\exp(-0.60)}{1+\exp(-0.60)} \approx 0.354$

変量効果に置き換えていく（ランダム切片モデル）

注意ポイントは「変量効果の分散」



（固定効果ver.） $$\log\frac{\hat{y}_{pg}}{1-\hat{y}_{pg}} = \beta_0 + \beta_{002}D_{002} + \beta_{003}D_{003} + \cdots + \beta_{185}D_{185}$$

学校001の平均値

一つ一つの係数（群間の差）には関心がないので

$$\log\frac{\hat{y}_{pg}}{1-\hat{y}_{pg}} = \beta_0 + u_{0g}$$

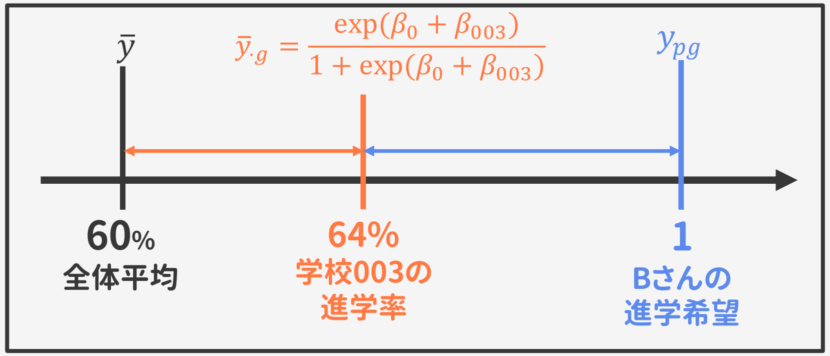
特定の群を基準にするのはなんだか気持ち悪いので

全体平均

（変量効果ver.） $$\log\frac{\hat{y}_{pg}}{1-\hat{y}_{pg}} = \mu + u_{0g}$$

観測値には $\left\{ \begin{array}{l} \text{集団レベルの変量効果} \\ \text{個人レベルの変量効果} \end{array} \right.$ が含まれている

こいつはどこに行った？



尤度から考えるとわかりやすい，かも?

進学希望は0か1の二値変数 ▶ 尤度はベルヌーイ分布に従う

あるいは試行回数1の二項分布

$$y_{pg} \sim Bernoulli(\hat{y}_{pg})$$

観測値には，確かに個人レベルの変量効果（というか誤差）が含まれている

$$\log \frac{\hat{y}_{pg}}{1-\hat{y}_{pg}} = \mu + u_{0g} \Rightarrow \hat{y}_{pg} = \frac{\exp(\mu + u_{0g})}{1+\exp(\mu + u_{0g})} \Rightarrow y_{pg} = \frac{\exp(\mu + u_{0g})}{1+\exp(\mu + u_{0g})} + u_{pg}$$

ただし，取りうる値が$\hat{y}_{pg}$によって完全に決定する

観測値y_{pg}が0か1である以上，

$$u_{pg} = \begin{cases} \hat{y}_{pg} & (y_{pg} = 0\text{のとき}) \\ 1-\hat{y}_{pg} & (y_{pg} = 1\text{のとき}) \end{cases}$$

ベルヌーイ分布$Bernoulli(p)$に従う確率変数は期待値がp，分散が$p(1-p)$でした

「個人レベル」と「集団レベル」の分散成分を独立には推定できない!

以上を踏まえてきちんと定式化

$$y_{pg} = \frac{\exp(\mu + u_{0g})}{1 + \exp(\mu + u_{0g})} + u_{pg}$$

- レベル別に式を書くと

（レベル1：個人）
$$y_{pg} = \frac{\exp(\beta_{0g})}{1 + \exp(\beta_{0g})} + u_{pg}$$

（レベル2：集団）
$$\beta_{0g} = \mu + u_{0g}$$

- 【仮定】集団レベル変量効果は通常正規分布に従う（ものとみなす）

$$u_{0g} \sim N\left(0, \sigma_{0g}^2\right)$$

群（集団）g の平均値を β_{0g} と表すと，

$$y_{pg} \sim Bernoulli\left(\frac{\exp(\beta_{0g})}{1 + \exp(\beta_{0g})}\right)$$

$$\beta_{0g} \sim N\left(\mu, \sigma_{0g}^2\right)$$

- 個人レベル変量効果について推定するパラメータは存在しない

> 通常の回帰分析（尤度が正規分布）では$y_{pg} \sim N\left(\beta_{0g}, \sigma_{pg}^2\right)$ でしたが
> ベルヌーイ分布では独立に推定できる分散パラメータσ_{pg}^2がないのです

Rでやるのはとってもかんたん

【ランダム切片GLM】

glm( )のときと同じ指定方法

```
model1_glm <- glmmTMB(wants_univ ~ (1 | school_id),
                      data = dat, family = binomial(link = "logit"))
summary(model1_glm)
```

【変量効果】

```
Random effects:

Conditional model:
 Groups    Name        Variance Std.Dev.
 school_id (Intercept) 2.164    1.471
Number of obs: 5766, groups:  school_id, 183
```

変量効果の分布の分散・標準偏差

$u_{0g} \sim N(0, 1.471^2)$

【固定効果】

```
Conditional model:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)    1.449      0.118   12.28   <2e-16 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

固定効果の
推定値と検定

$\beta_{0g} = 1.449 + u_{0g}$

平均的な進学率の学校の進学率

（個人レベルの）進学率の平均値，ではありません！

$\mu = 1.449$なので $\hat{y}_{pg} = \frac{\exp(1.449)}{1+\exp(1.449)} \approx 0.810$

進学率の学校ごとのばらつき

$u_{0g} \sim N(0, 1.471^2)$ ▶ $\beta_{0g} \sim N(1.449, 1.471^2)$

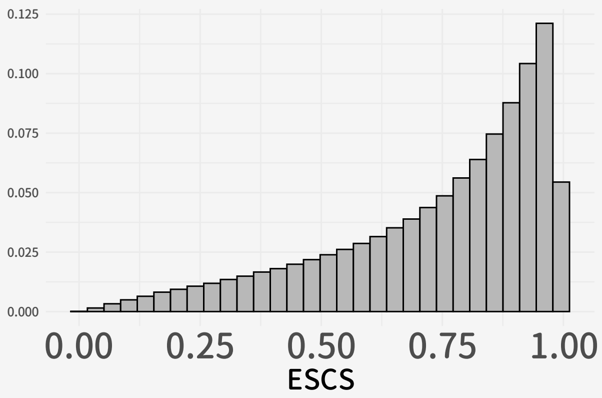
【95%区間を出してみたり】

```
plogis(qnorm(c(0.025, 0.975),
       mean = 1.449, sd = 1.471))
```

```
[1] 0.1924596 0.9870306
```

【乱数を用いてヒストグラムを作ったり】

```
g_y_r <- rnorm(1000000,
               mean = 1.449,
               sd = 1.471)
p <- plogis(g_y_r)
hist(p)
```



どのような学校では進学率が高いのか？

最初から個人レベル・集団レベルの説明変数をそれぞれ投入します。

集団平均に対する回帰係数
▶ESCSが高い学校にいる人ほど…

集団平均からの偏差に対する回帰係数
▶学校内で，相対的にESCSが高い人ほど…

$$\log\frac{\hat{y}_{pg}}{1-\hat{y}_{pg}} = \mu + \beta_{\overline{\mathrm{ESCS}}}\overline{\mathrm{ESCS}}_{.g} + \beta_{\mathrm{ESCS}}\mathrm{CWC}_{pg}$$

レベル別に書くと

（レベル1：個人）
$$\log\frac{\hat{y}_{pg}}{1-\hat{y}_{pg}} = \beta_{0g} + \beta_{\mathrm{ESCS}}\mathrm{CWC}_{pg}$$

（レベル2：集団）
$$\beta_{0g} = \mu + \beta_{\overline{\mathrm{ESCS}}}\overline{\mathrm{ESCS}}_{.g} + u_{0g}$$

【ランダム切片GLMの推定結果】

```
model2_glm <- glmmTMB(wants_univ ~ escs_cwc + sch_escs_mean + (1 ¦ school_id),
                      data = dat, family = binomial(link = "logit"))
summary(model2_glm)
```

【変量効果】

```
Random effects:

Conditional model:
 Groups    Name        Variance Std.Dev.
 school_id (Intercept) 0.5417   0.736
Number of obs: 5766, groups:  school_id, 183
```

各変量効果の分布の分散・標準偏差

$$u_{0g} \sim N(0, 0.736^2)$$

【固定効果】

```
Conditional model:
              Estimate Std. Error z value Pr(>¦z¦)
(Intercept)    1.88638    0.07958  23.703  < 2e-16 ***
escs_cwc       0.38946    0.05521   7.054 1.74e-12 ***
sch_escs_mean  3.58849    0.20254  17.717  < 2e-16 ***
```

固定効果の
推定値と検定

(レベル1：個人)
$$\log \frac{\hat{y}_{pg}}{1-\hat{y}_{pg}} = \beta_{0g} + 0.389\mathrm{CWC}_{pg}$$

(レベル2：集団)
$$\beta_{0g} = 1.886 + 3.588\overline{\mathrm{ESCS}}_{\cdot g} + u_{0g}$$

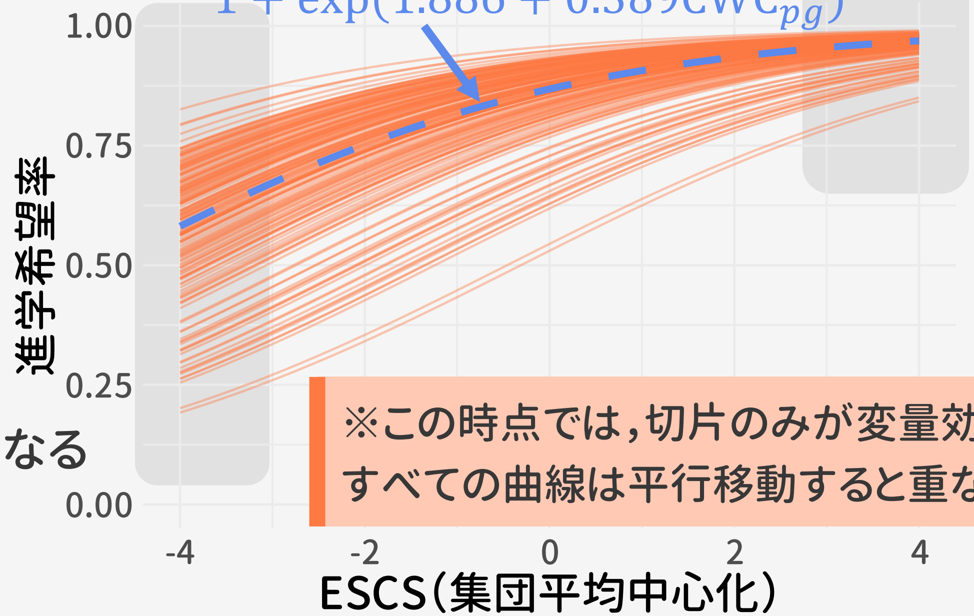
（レベル1：個人） $\log \frac{\hat{y}_{pg}}{1-\hat{y}_{pg}} = \beta_{0g} + 0.389\mathrm{CWC}_{pg}$

（レベル2：集団） $\beta_{0g} = 1.886 + 3.588\overline{\mathrm{ESCS}}_{.g} + u_{0g}$

CWC_{pg} の係数が正の値 ▶ **相対的にESCSが高い生徒ほど進学率がやや高い**

$\overline{\mathrm{ESCS}}_{.g}$ の係数が正の値 ▶ **ESCSが平均的に高い学校ほど進学率がかなり高い**

$$\hat{y}_{pg} = \frac{\exp(1.886 + 0.389\mathrm{CWC}_{pg})}{1 + \exp(1.886 + 0.389\mathrm{CWC}_{pg})}$$



どの学校でも
上位層は大学にほぼ進学する

下位層の進学希望は
学校のレベルによって異なる

※この時点では，切片のみが変量効果のため
すべての曲線は平行移動すると重なります

まずは固定効果として考えてみる

P57と比べると，左辺が変わっただけで基本的な考え方は同じ

$$\log\frac{\hat{y}_{pg}}{1-\hat{y}_{pg}} = \underbrace{\beta_0 + \beta_{\mathrm{ESCS}}\mathrm{ESCS}_{pg}}_{\text{学校001の係数（基準）}} + \beta_{0,002}D_{002} + \beta_{0,003}D_{003} + \cdots + \beta_{\mathrm{ESCS},002}D_{002}\mathrm{ESCS}_{pg} + \beta_{\mathrm{ESCS},003}D_{003}\mathrm{ESCS}_{pg} + \cdots$$

学校002の係数

学校003の係数

学校ダミーごとの切片の差

学校ダミーごとの傾きの差

学校（ダミー）ごとに場合分けして書くならば

$$\log\frac{\hat{y}_{pg}}{1-\hat{y}_{pg}} = \begin{cases} \beta_0 + \beta_{\mathrm{ESCS}}\,\mathrm{ESCS}_{pg} & \text{（学校001）} \\ (\beta_0 + \beta_{0,002}) + (\beta_{\mathrm{ESCS}} + \beta_{\mathrm{ESCS},002})\,\mathrm{ESCS}_{pg} & \text{（学校002）} \\ (\beta_0 + \beta_{0,003}) + (\beta_{\mathrm{ESCS}} + \beta_{\mathrm{ESCS},003})\,\mathrm{ESCS}_{pg} & \text{（学校003）} \\ \vdots \end{cases}$$

なのでRでもだいたいおなじ

左辺がどのように変形しているかを示す引数

$$\log \frac{\hat{y}_{pg}}{1-\hat{y}_{pg}}$$

lm( )の代わりにglm( )を使えば良いだけ

【学校ごとのread_scoreの差を評価する一般化線形モデル】

```
summary(glm(wants_univ ~ school_id * escs, data = dat, family=binomial(link="logit")))
```

```
（前略）
Coefficients:
                    Estimate Std. Error z value Pr(>¦z¦)
(Intercept)        1.957e+01  2.400e+03   0.008    0.993
school_id002      -2.030e+01  2.400e+03  -0.008    0.993
school_id003      -1.885e+01  2.400e+03  -0.008    0.994
school_id004      -1.936e+01  2.400e+03  -0.008    0.994
（中略）
school_id185      -1.854e+01  2.400e+03  -0.008    0.994
escs              -2.981e-07  3.356e+03   0.000    1.000
school_id002:escs -3.045e-01  3.356e+03   0.000    1.000
school_id003:escs  1.004e+00  3.356e+03   0.000    1.000
school_id004:escs  8.362e-01  3.356e+03   0.000    1.000
（後略）
```

$$\log \frac{\hat{y}_{pg}}{1-\hat{y}_{pg}} = \beta_0 + \beta_{\mathrm{ESCS}}\mathrm{ESCS}_{pg} + \beta_{0,002}D_{002} + \beta_{0,003}D_{003} + \beta_{\mathrm{ESCS},002}D_{002}\mathrm{ESCS}_{pg} + \beta_{\mathrm{ESCS},003}D_{003}\mathrm{ESCS}_{pg}$$

【各学校の切片と傾き】

school_id	切片		傾き	
001	β_0	19.57	β_{ESCS}	0.000
002	$\beta_0 + \beta_{002}$	-0.73	$\beta_{\mathrm{ESCS}} + \beta_{\mathrm{ESCS},002}$	-0.304
003	$\beta_0 + \beta_{003}$	0.72	$\beta_{\mathrm{ESCS}} + \beta_{\mathrm{ESCS},003}$	1.004
⋮				

今までの説明をそのまま適用したら簡単

（固定効果ver.）

$$\log\frac{\hat{y}_{pg}}{1-\hat{y}_{pg}} = \beta_0 + \beta_{\mathrm{ESCS}}\mathrm{ESCS}_{pg} + \beta_{0,002}D_{002} + \beta_{0,003}D_{003} + \cdots$$
$$+\beta_{\mathrm{ESCS},002}D_{002}\mathrm{ESCS}_{pg} + \beta_{\mathrm{ESCS},003}D_{003}\mathrm{ESCS}_{pg} + \cdots$$

学校001の切片と傾き

一つ一つの係数（群間の差） には 関心がないので

$$\log\frac{\hat{y}_{pg}}{1-\hat{y}_{pg}} = \beta_0 + \beta_{\mathrm{ESCS}}\mathrm{ESCS}_{pg} + u_{0g} + u_{\mathrm{ESCS},g}\mathrm{ESCS}_{pg}$$

特定の群を基準にするのはなんだか気持ち悪いので

切片と傾きの全体平均

（変量効果ver.） $$\log\frac{\hat{y}_{pg}}{1-\hat{y}_{pg}} = \mu_0 + \mu_{\mathrm{ESCS}}\mathrm{ESCS}_{pg} + u_{0g} + u_{\mathrm{ESCS},g}\mathrm{ESCS}_{pg}$$
$$= (\mu_0 + u_{0g}) + (\mu_{\mathrm{ESCS}} + u_{\mathrm{ESCS},g})\mathrm{ESCS}_{pg}$$

```
model3_glm <- glmmTMB(wants_univ ~ escs_cwc + (escs_cwc ¦ school_id),
                      data = dat, family = binomial(link = "logit"))
summary(model3_glm)
```

【変量効果】

```
Random effects:

Conditional model:
 Groups    Name        Variance Std.Dev. Corr
 school_id (Intercept) 2.21612  1.4887
           escs_cwc    0.07018  0.2649   -0.28
Number of obs: 5766, groups:  school_id, 183
```

$$\sigma_{(0g)(\mathrm{ESCS},g)} = r^{2}_{(0g)(\mathrm{ESCS},g)}{}^{(\mathrm{Corr})}\sigma_{0g}\sigma_{\mathrm{ESCS},g} \approx -0.110$$

各変量効果の分布の分散・標準偏差

$$\begin{bmatrix} u_{0g} \\ u_{\mathrm{ESCS},g} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1.489^2 & -0.110 \\ -0.110 & 0.265^2 \end{bmatrix}\right)$$

【固定効果】

```
Conditional model:
            Estimate Std. Error z value Pr(>¦z¦)
(Intercept)  1.46466    0.11965  12.241  < 2e-16 ***
escs_cwc     0.35556    0.07242   4.909 9.13e-07 ***
```

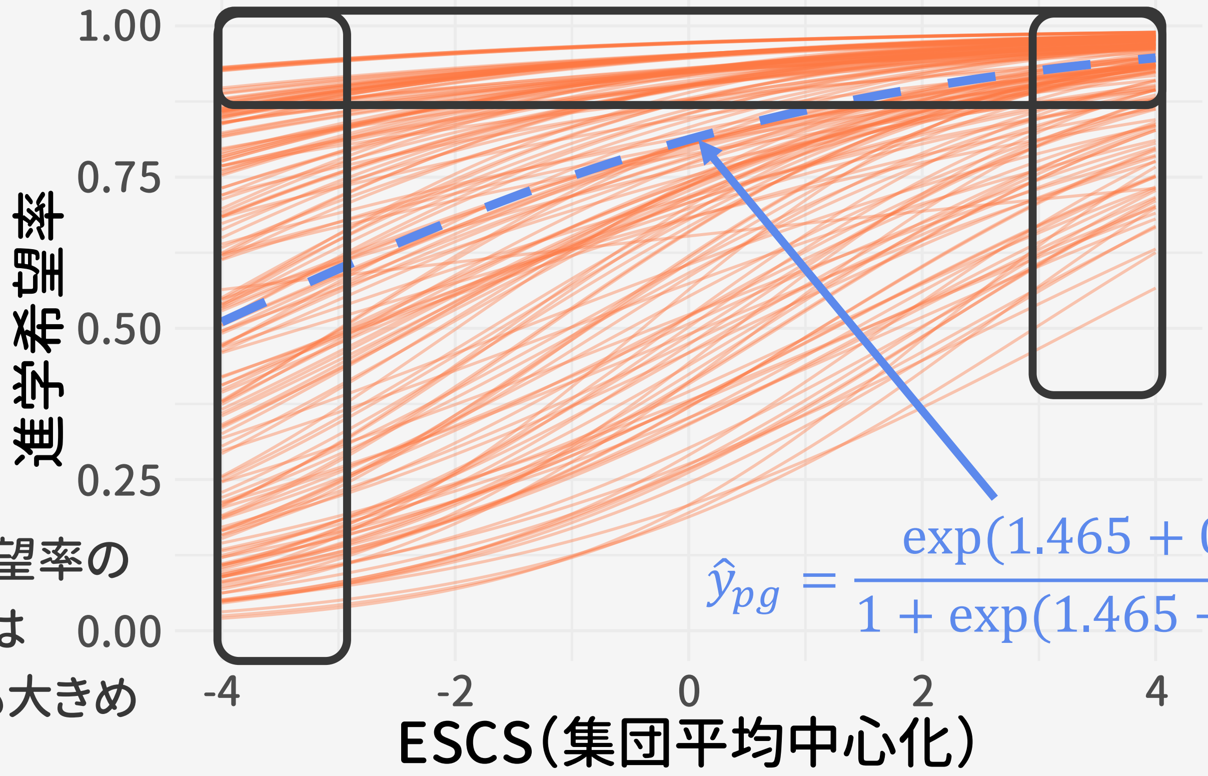
固定効果の
推定値と検定

(レベル1:個人) $\log\frac{\hat{y}_{pg}}{1-\hat{y}_{pg}} = \beta_{0g} + \beta_{\mathrm{ESCS},g}\mathrm{CWC}_{pg}$

(レベル2:集団) $\beta_{0g} = 1.465 + u_{0g},\ \ \beta_{\mathrm{ESCS},g} = 0.356 + u_{\mathrm{ESCS},g}$

集団ごとの回帰曲線を引いてみる

学校によってはESCSの上位層も下位層も
みんな大学に進学を希望している



ESCS最上位層の進学希望率は
低いところでは50%くらい

ESCS最下位層の進学希望率の
学校ごとのばらつきは
最上位層のばらつきよりも大きめ

$$\hat{y}_{pg} = \frac{\exp(1.465 + 0.356\mathrm{CWC}_{pg})}{1 + \exp(1.465 + 0.356\mathrm{CWC}_{pg})}$$

係数間に負の相関があったので

切片が大きい学校ほど傾きが小さい傾向がある

優秀な学校では，CWCに関係なく
みんな進学希望するために傾きが小さい?

個人レベルと集団レベルを分離する

同じ流れなので細かいところは省略

PP66-68とやっていることは同じです

学校 g のESCSの平均値 $\overline{\text{ESCS}}_{.g}$ を傾きと切片の説明変数にする

sch_escs_mean

（レベル1：個人） $\log \frac{\hat{y}_{pg}}{1-\hat{y}_{pg}} = \beta_{0g} + \beta_{\text{ESCS},g}\text{CWC}_{pg}$

（レベル2：集団） $\beta_{0g} = \mu_0 + \beta_{0,\overline{\text{ESCS}}}\overline{\text{ESCS}}_{.g} + u_{0g}$

$\beta_{\text{ESCS},g} = \mu_{\text{ESCS}} + \beta_{1,\overline{\text{ESCS}}}\overline{\text{ESCS}}_{.g} + u_{\text{ESCS},g}$

係数 $\beta_{0,\overline{\text{ESCS}}}$ が正ならば $\overline{\text{ESCS}}_{.g}$ が高い学校ほど切片 β_{0g} が高いことになる

$$\log \frac{\hat{y}_{pg}}{1-\hat{y}_{pg}} = \underbrace{\mu_0}_{\text{切片}} + \underbrace{\beta_{0,\overline{\text{ESCS}}}\overline{\text{ESCS}}_{.g}}_{\text{集団レベル}} + \underbrace{\mu_{\text{ESCS}}\text{CWC}_{pg}}_{\text{個人レベル}} + \underbrace{\beta_{1,\overline{\text{ESCS}}}\overline{\text{ESCS}}_{.g}\text{CWC}_{pg}}_{\text{レベル間交互作用}} + \underbrace{u_{0g} + u_{\text{ESCS},g}\text{CWC}_{pg}}_{\text{変量効果（残差）}}$$

```
read_score ~ 1 + sch_escs_mean + escs_cwc + sch_escs_mean:escs_cwc + (1 + escs_cwc | school_id)
```

実行してみる

```
model4_glm <- glmmTMB(wants_univ ~ escs_cwc * sch_escs_mean + (escs_cwc | school_id),
                      data = dat, family = binomial(link="logit"))
summary(model4_glm)
```

【変量効果】

```
Conditional model:
 Groups    Name        Variance Std.Dev. Corr
 school_id (Intercept) 0.54961  0.7414
           escs_cwc    0.05358  0.2315   0.50
Number of obs: 5766, groups:  school_id, 183
```

$$\sigma_{(0g)(\mathrm{ESCS},g)} = \overset{(\mathrm{Corr})}{r^2_{(0g)(\mathrm{ESCS},g)}} \sigma_{0g} \sigma_{\mathrm{ESCS},g} \approx 0.086$$

各変量効果の分布の分散・標準偏差

$$\begin{bmatrix} u_{0g} \\ u_{\mathrm{ESCS},g} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0.741^2 & 0.086 \\ 0.086 & 0.232^2 \end{bmatrix}\right)$$

【固定効果】

```
Conditional model:
                       Estimate Std. Error z value Pr(>|z|)
(Intercept)             1.88564    0.08027  23.493  < 2e-16 ***
escs_cwc                0.26730    0.08712   3.068  0.00215 **
sch_escs_mean           3.57660    0.20388  17.543  < 2e-16 ***
escs_cwc:sch_escs_mean -0.50243    0.20912  -2.403  0.01628 *
```

固定効果の推定値と検定

$\overline{\mathrm{ESCS}}_{.g}$ は傾きの違いを少し説明できる

（レベル2の切片） $\beta_{0g} = 1.886 + 3.577\overline{\mathrm{ESCS}}_{.g} + u_{0g}$

（レベル2の傾き） $\beta_{\mathrm{ESCS},g} = 0.267 - 0.502\overline{\mathrm{ESCS}}_{.g} + u_{\mathrm{ESCS},g}$

結果の解釈

（レベル2の切片）　$\beta_{0g} = 1.886 + 3.577\overline{\mathrm{ESCS}}_{.g} + u_{0g}$

（レベル2の傾き）　$\beta_{\mathrm{ESCS},g} = 0.267 - 0.502\overline{\mathrm{ESCS}}_{.g} + u_{\mathrm{ESCS},g}$

$$\log\frac{\hat{y}_{pg}}{1-\hat{y}_{pg}} = \beta_{0g} + \beta_{\mathrm{ESCS},g}\mathrm{CWC}_{pg}$$

- **ESCSの学校平均が高いと，切片は大きくなっていく**
  - ▶ ESCS上位層の学校ほど，大学進学希望率は高くなる
- **ESCSの学校平均が高いと，CWCの傾きが小さくなっていく**
  - ▶ ESCS上位層の学校では，相対的なESCSと進学希望率の関係がなくなる
    ESCS下位層の学校では，相対的なESCSが進学希望率に強く関係してくる
- **可視化して確認していこう！（単純傾斜分析）**

  modelbasedパッケージで行っていきます

```
install.packages("modelbased")
library(modelbased)
```

調整変数が特定の値のときの傾き（限界効果）

簡単に言うと，escs_cwcが1増えると $\hat{y}_{pg}$ がどれだけ大きくなるか（の期待値）

今回は$\overline{\mathrm{ESCS}}_{.g}$ が特定の値のときの，CWC_{pg}の効果 $\beta_{\mathrm{ESCS},g}$が知りたい

【単純傾斜分析】

```
estimate_slopes(model4_glm, trend = "escs_cwc", by = "sch_escs_mean")
```

sch_escs_mean	Slope	SE	95% CI	z	p
-1.11	0.09	0.02	[0.04, 0.13]	3.76	< .001
-0.89	0.12	0.03	[0.07, 0.17]	4.73	< .001
-0.68	0.14	0.02	[0.09, 0.18]	5.98	< .001
-0.46	0.11	0.02	[0.08, 0.14]	7.08	< .001
-0.25	0.07	0.01	[0.05, 0.09]	6.02	< .001
-0.03	0.03	0.01	[0.01, 0.05]	2.98	0.003
0.18	9.24e-03	8.86e-03	[-0.01, 0.03]	1.04	0.297
0.40	2.20e-04	6.29e-03	[-0.01, 0.01]	0.03	0.972
0.61	-2.11e-03	4.00e-03	[-0.01, 0.01]	-0.53	0.597
0.83	-2.05e-03	2.37e-03	[-0.01, 0.00]	-0.86	0.387

デフォルトでは
データ内の範囲を
10等分した値

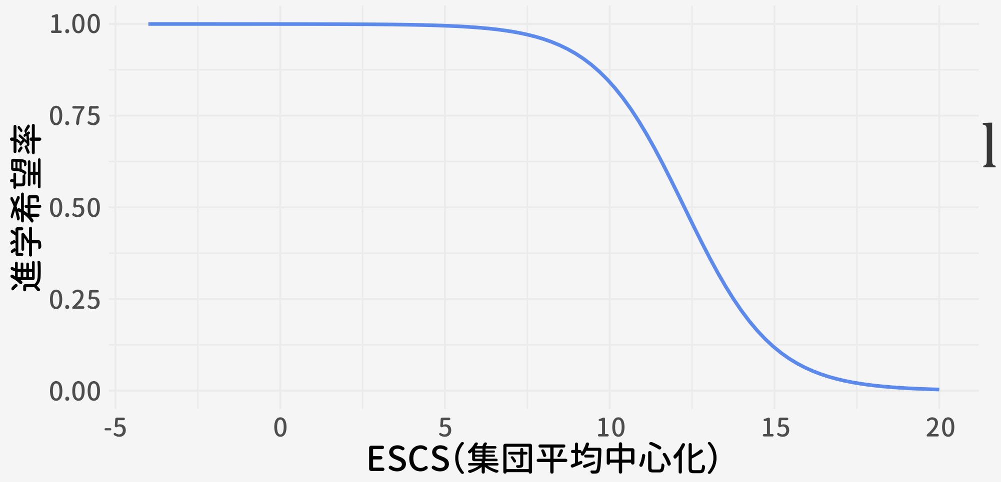
平均値が
低い学校では
CWC_{pg}の効果が
有意になっている

「傾き」で評価すると，変な解釈になってしまうことがある

(レベル2の切片) $\beta_{0g} = 1.886 + 3.577\overline{\mathrm{ESCS}}_{.g} + u_{0g}$

(レベル2の傾き) $\beta_{\mathrm{ESCS},g} = 0.267 - 0.502\overline{\mathrm{ESCS}}_{.g} + u_{\mathrm{ESCS},g}$

▶ $\overline{\mathrm{ESCS}}_{.g} = 2$の学校では，傾きの期待値は $\beta_{\mathrm{ESCS},g} = -0.737$ となる(有意)



$$\log\frac{\hat{y}_{pg}}{1-\hat{y}_{pg}} = 9.04 - 0.737\mathrm{CWC}_{pg}$$

「ESCSが平均的に高い学校では，相対的なESCSが高い生徒ほど大学進学を希望しなくなる」

「実質的な効果」を考えてみましょう

(おまけ)「限界効果」って何?②

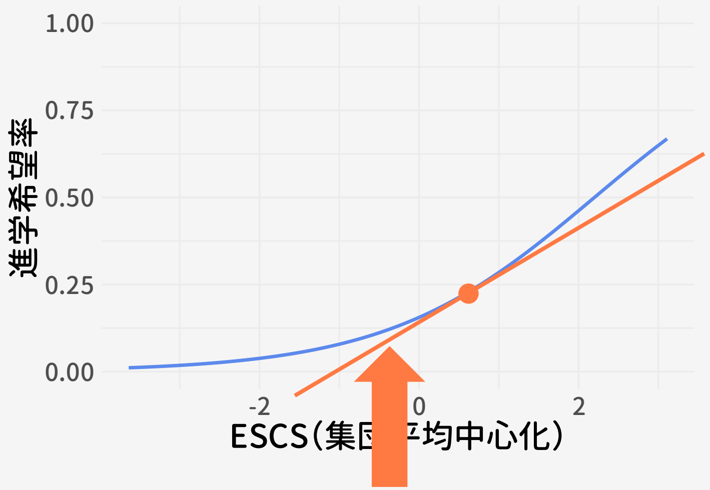
（切片） $\beta_{0g} = 1.886 + 3.577\overline{\text{ESCS}}_{.g} + u_{0g}$

（傾き） $\beta_{\text{ESCS},g} = 0.267 - 0.502\overline{\text{ESCS}}_{.g} + u_{\text{ESCS},g}$

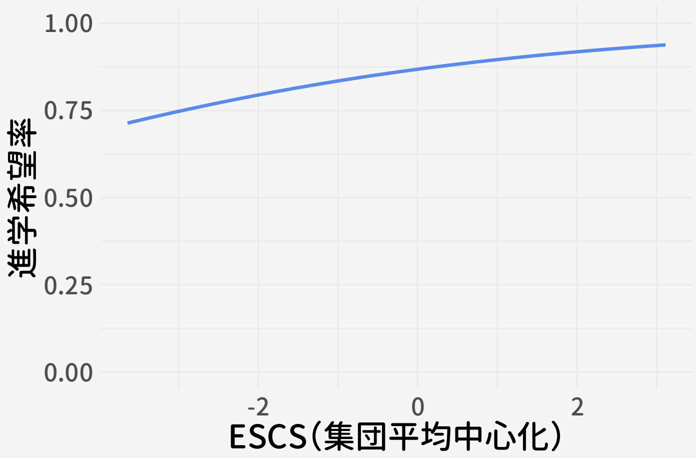
データに出現するCWC_{pg} の範囲は$[-3.64, 3.11]$

この範囲に限定していくつかプロットしてみると…

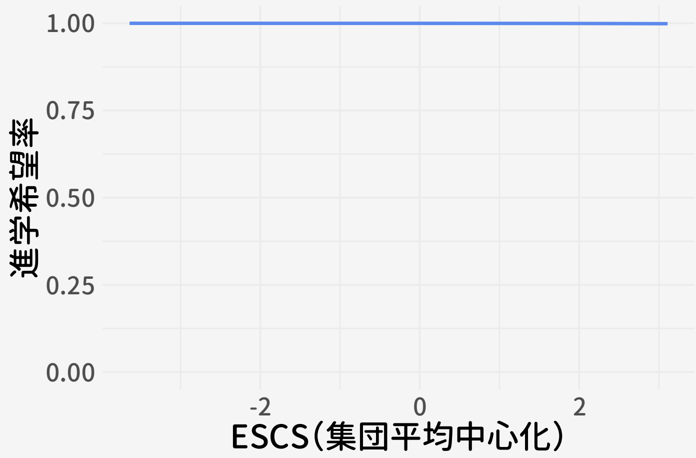
$$\log\frac{\hat{y}_{pg}}{1-\hat{y}_{pg}} = -1.691 + 0.769\text{CWC}_{pg}$$
$$\left(\overline{\text{ESCS}}_{.g} = -1\right)$$



$$\log\frac{\hat{y}_{pg}}{1-\hat{y}_{pg}} = 1.886 + 0.267\text{CWC}_{pg}$$
$$\left(\overline{\text{ESCS}}_{.g} = 0\right)$$



$$\log\frac{\hat{y}_{pg}}{1-\hat{y}_{pg}} = 9.04 - 0.737\text{CWC}_{pg}$$
$$\left(\overline{\text{ESCS}}_{.g} = 2\right)$$



特定の位置での傾きが

限界効果 (marginal effect) ▶ データ全体で期待値をとったものが **平均限界効果** (average ME)

estimate_slopes( )の出力のデフォルト

（おまけ）「特定の値」を指定したい場合は

明示的に与える

```
estimate_slopes(model4_glm, trend = "escs_cwc", by = "sch_escs_mean = c(-0.5, 0, 0.5)")
```

```
sch_escs_mean ¦     Slope ¦        SE ¦         95% CI ¦     z ¦      p
-----------------------------------------------------------------------
-0.50         ¦      0.12 ¦      0.02 ¦ [ 0.09, 0.15] ¦  6.95 ¦ < .001
0.00          ¦      0.03 ¦      0.01 ¦ [ 0.01, 0.05] ¦  2.61 ¦  0.009
0.50          ¦ -1.36e-03 ¦ 5.15e-03 ¦ [-0.01, 0.01] ¦ -0.26 ¦  0.791
```

特別な指示を与える

平均±1*SD*

```
estimate_slopes(model4_glm, trend = "escs_cwc", by = "sch_escs_mean = [sd]")
```

```
sch_escs_mean ¦    Slope ¦       SE ¦         95% CI ¦    z ¦      p
--------------------------------------------------------------------
-0.46         ¦     0.11 ¦     0.02 ¦ [ 0.08, 0.14] ¦ 7.08 ¦ < .001
-0.10         ¦     0.04 ¦     0.01 ¦ [ 0.02, 0.06] ¦ 3.86 ¦ < .001
0.26          ¦ 4.72e-03 ¦ 7.94e-03 ¦ [-0.01, 0.02] ¦ 0.60 ¦  0.552
```

(おまけ)回帰式に基づく「傾き」の計算が必要な場合は

$\hat{\beta}_{\mathrm{ESCS},g} = 0.267 - 0.502\overline{\mathrm{ESCS}}_{.g}$
リンク関数のレベルでの期待値を出す

引数predictを設定する

```
estimate_slopes(model4_glm, trend = "escs_cwc", by = "sch_escs_mean", predict = "link")
```

```
sch_escs_mean ¦ Slope ¦   SE ¦         95% CI ¦     z ¦      p
-------------------------------------------------------------
-1.11         ¦  0.82 ¦ 0.18 ¦ [ 0.46, 1.17] ¦  4.53 ¦ < .001
-0.89         ¦  0.71 ¦ 0.14 ¦ [ 0.44, 0.98] ¦  5.10 ¦ < .001
-0.68         ¦  0.60 ¦ 0.10 ¦ [ 0.40, 0.80] ¦  5.98 ¦ < .001
-0.46         ¦  0.49 ¦ 0.07 ¦ [ 0.36, 0.63] ¦  7.01 ¦ < .001
-0.25         ¦  0.39 ¦ 0.06 ¦ [ 0.26, 0.51] ¦  6.20 ¦ < .001
-0.03         ¦  0.28 ¦ 0.08 ¦ [ 0.12, 0.44] ¦  3.36 ¦ < .001
0.18          ¦  0.17 ¦ 0.12 ¦ [-0.06, 0.40] ¦  1.43 ¦  0.151
0.40          ¦  0.06 ¦ 0.16 ¦ [-0.25, 0.37] ¦  0.38 ¦  0.703
0.61          ¦ -0.05 ¦ 0.20 ¦ [-0.44, 0.34] ¦ -0.24 ¦  0.811
0.83          ¦ -0.16 ¦ 0.24 ¦ [-0.63, 0.32] ¦ -0.64 ¦  0.521
```

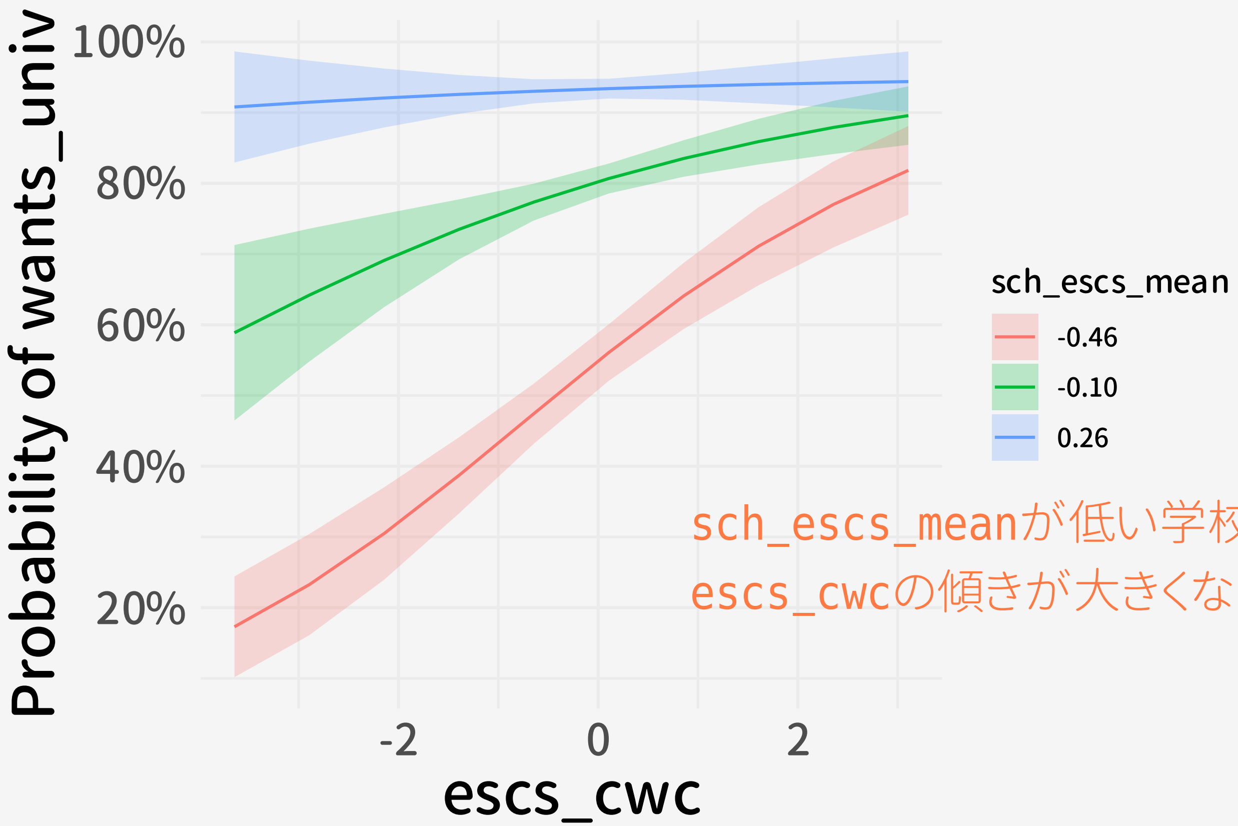
交互作用の可視化

もちろん一般化線形モデルでなくても実行可能です

データ内の平均±1SDについて表示せよ

【単純傾斜分析】

```
means <- estimate_means(model4_glm, by = c("escs_cwc", "sch_escs_mean = [sd]"))
plot(means)
```



sch_escs_meanが低い学校ほど
escs_cwcの傾きが大きくなっている

Johnson-Neyman区間

もちろん一般化線形モデルでなくても実行可能です

精度は上がるが計算時間がかかる

sch_escs_meanの値がどこからどこまで，効果は有意と言えるのか？

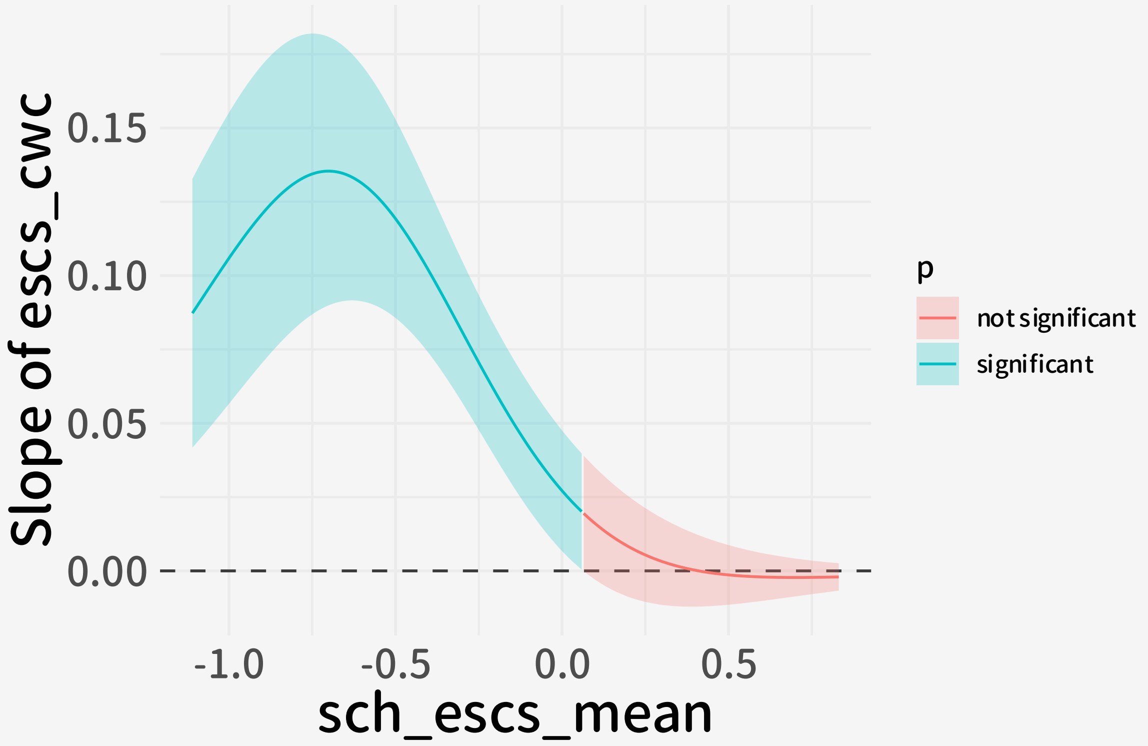
```
slopes_jn <- estimate_slopes(model4_glm, trend = "escs_cwc", by = "sch_escs_mean", length = 300)
```

```
summary(slopes_jn)
```

```
Start ¦  End ¦ Direction ¦ Confidence
--------------------------------------------
-1.11 ¦ 0.06 ¦ positive  ¦ Significant
0.06  ¦ 0.41 ¦ positive  ¦ Not Significant
0.42  ¦ 0.83 ¦ negative  ¦ Not Significant
```

- -1.11から0.06までは有意な正の効果
- 0.06から0.41までも正の効果だが有意ではない
- 0.42から0.83までは負の効果・有意ではない

```
plot(slopes_jn)
```



7

まとめ

マルチレベルモデルを完全に理解した

本セミナーのまとめ

- **マルチレベルモデルの基本的な考え方を紹介しました**
  - 分散分析からステップアップしたもの,と考えるととっつきやすい気がする
- **マルチレベルモデルをRで実行する方法を紹介しました**
  - これから始めるなら`glmmTMB`がおすすめ(`lme4`から乗り換えても良いくらい)
  - 結果の解釈の方法も色々あるので,うまく使えると面白い仮説が立てられる気がする
  - 最近は便利なパッケージも開発されているんですねぇ
- **ついでに一般化線形モデルも紹介しました**
  - マルチレベルモデルに限らず知っておいたほうが良いモデリング

（おまけ）マルチレベルGLMはもっと柔軟に

- **回帰分析よりも複雑な相関関係にも「集団レベル」と「個人レベル」があるのでは?**

  ▶ マルチレベル相関係数を用いたマルチレベルSEMに進みましょう (現状Mplusが必要)

- **ロジスティック回帰・ポアソン回帰では分散がデータに合わない事がある**

  ベルヌーイ分布・ポアソン分布は期待値と分散が一緒に変動してしまうため

  ▶ 過分散な状況に対応したモデルなどもglmmTMBはできる

- **変量効果が正規分布ではないと考えられる場合**

  たまに極端に離れた集団がある場合 ▶ 変量効果の分布を t 分布にする

  分布が左右対称ではない場合 ▶ 変量効果をガンマ分布や対数正規分布などにする

  ▶ **`glmmTMB`パッケージでは不可能なので，ベイズ統計モデリング(stan)等が必要です**

日本教育心理学会・研究委員会企画セミナー

マルチレベル分析入門

分寺 杏介

神戸大学大学院　経営学研究科

bunji@bear.kobe-u.ac.jp

※本スライドは，クリエイティブ・コモンズ 表示-非営利 4.0 国際 ライセンス（CC BY-NC 4.0）に従って利用が可能です。