

# Regression Discontinuity Designs

Mastering Metrics Ch.4 (後半) :  
Fuzzy RD

### 【背景】

- ボストンとニューヨークの公立校は**選別試験**を行っている.
- 他のアメリカの公立校とは異なり、出願者を試験によって選別する.
  - ボストンの場合、出願者の半数未満しかBoston Latin School (BLS) などの高校に入学できない.
  - ニューヨークの場合、Big Appleと呼ばれる高校 (Stuyvesantなど) に入学できるのは出願者のうちの6分の1に過ぎない.

- このような激しい競争を経て、入学したがることは理解可能。
  - 選抜試験のある高校を卒業した多くの生徒は優れたキャリアへと進んでいく.
  - これらの公立校の生徒は他の公立校の生徒と比べて、どんな指標で見ても優れている.
- 経済学者や社会学者はこのような学校に入学することの結果に関心がある。
  - 生徒は先生から学ぶのと同様に、彼らのピア（同級生）から多くを学んでいる.

## 【分析上の課題】

- これらの学校に入学することの結果は、選抜入試そのものによる影響を受けている可能性がある。
  - もし学校が成績優秀者を求めている場合、学校が付加価値を与えるかどうかによらず、もともと成績優秀な生徒が受験してくる可能性がある（**選択バイアス**）。

## 【対策・方針】

- 選択バイアスを乗り越える手がかりとして、先に見たOregon Health Authorityの医療保険のくじの方法を参考にする。
  - 医療保険のくじと同様、StuyvesantやBLSでは一部の生徒について、試験の結果ではなく、ランダムに入学させる生徒を決めている。
  - この実験的な状況をいかして、選抜試験のある高校が付加価値を与えているか否かについて分析を行っていく（**自然実験**）。

### 【選抜試験の特性】

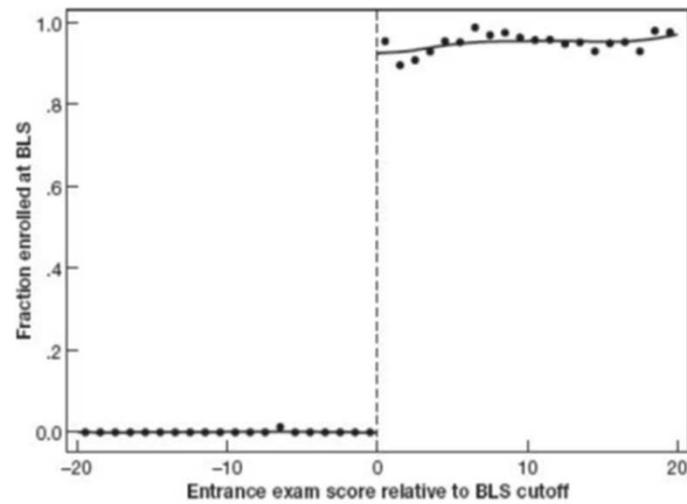
- 合格最低点に近い点数の受験者は、閾値の右側にいるか左側にいるかにかかわらず、ランダムに合格させる.
- 今回の実験デザインが難しいのは、単純に閾値が介入をうけるか・受けないかを定めるスイッチとなっていないことである。
  - 閾値で学校に入学するか否かが不連続に決まってしまうという性質による.
  - ある合格した生徒は違う高校に通うことにするかもしれないし、落ちた生徒は必ずどこか違う学校に行ってしまう.

## 【Fuzzy RD】

- このようにある閾値が、介入を受けるか否かのスイッチとして機能するのではなく、介入を受ける確率や平均的な特徴を変える場合、RDデザインはfuzzyであるという。

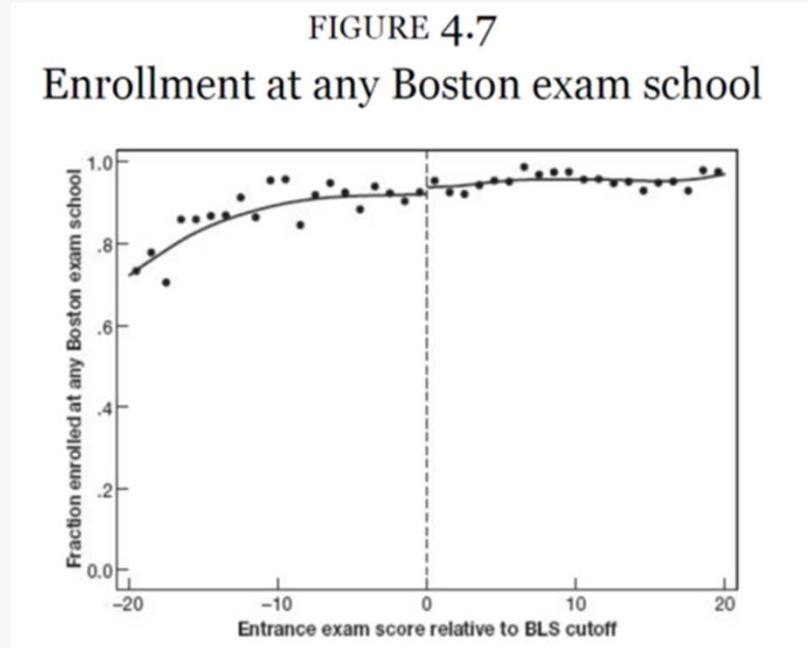
- 何が選抜試験を課す高校の介入なのか？
  - Figure 4.6 – 4.8では、BLSの受験者に着目して上記に対する回答を示す.
- 選抜試験を受けるBLSなどの受験者はIndependent School Entrance Exam (ISEE)という試験を受ける.
- Figure 4.6 – 4.8では、このISEEがBLSの閾値（合格最低点）付近の受験者を示している.

FIGURE 4.6  
Enrollment at BLS



### 【BLSの入学率】

- 全員ではないが、閾値より右側の受験者のほとんどが入学.



### 【選抜試験があるいずれかの公立校への入学率】

- BLSの閾値を下回った受験者のほとんどが、他の選抜試験のある公立校に入学している。
- ある高校に入学することのオッズ\*はBLSの閾値に対して不変。

$$* \text{ オッズ (Odds)} = \frac{p}{1-p}$$

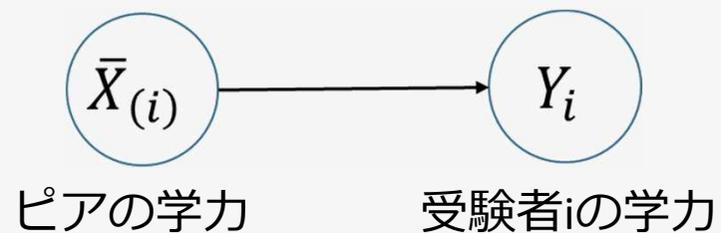
### 【ピア効果】

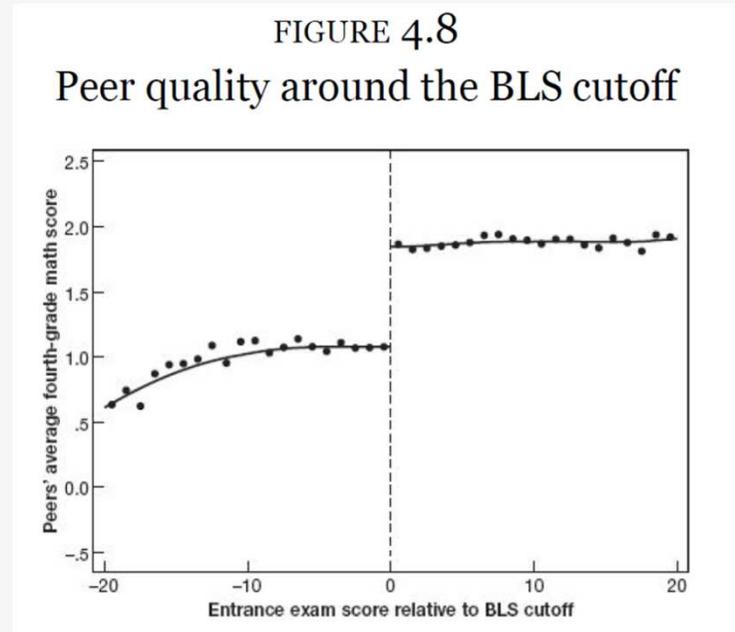
- クラスメートの能力が生徒の学習に因果効果をもたらすか？
  - 成績の良い（悪い）生徒と同じ学校に入学することで、成績が上昇する（下がる）か。

### 【なぜピア効果が問題となるのか？】

- アメリカでは、生徒は家の近くの学校に入学することが多い。
- 貧しく・非白人系で・成績の低い生徒はよくできる生徒（白人）から遠いところに住んでいるケースが多い。
- そのため、近場の学校に生徒を割り当ててしまうと貧しいマイノリティの生徒の成績が伸びる機会を奪っている可能性がある。

- 今回のボストンの選抜試験は、ピアの質についての自然実験となりうる。
  - ボストンの選抜試験校のいずれかに入学資格のある生徒は、閾値を下回った受験者よりもより成績の良いピアのいる学校に通うことになる。





### 【ピアの質】

- ピアの質を4<sup>th</sup>-gradeの数学のスコアで測る（選抜試験に出願する2年前の点数）。
- テストスコアは標準偏差に変換している。
- ピアのスコアは閾値を境にして、 $0.8\sigma$ 高くなっている。
  - 介入の強さ（**treatment intensity**）に大きな変化がある。

### 【ピア効果の分析？】

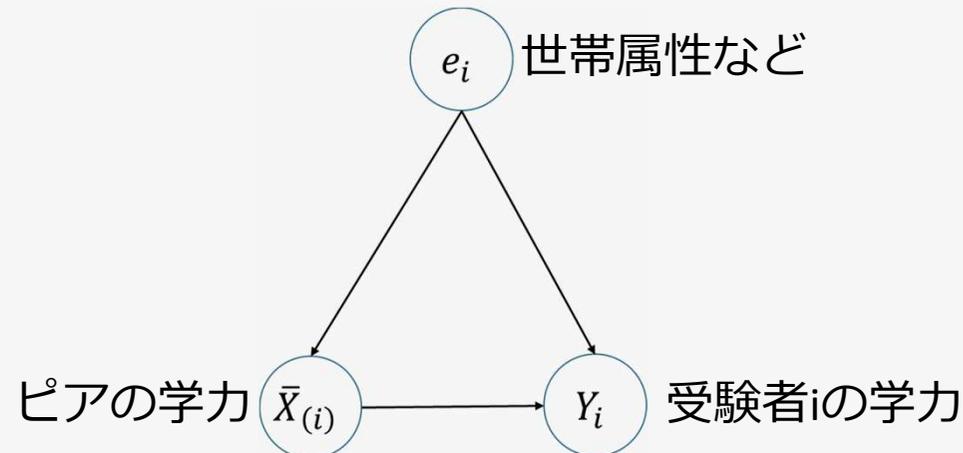
- ピアの質と学力の関係を見るために単純な回帰分析を行う。
- 選抜試験校を受験した生徒の7<sup>th</sup>-gradeでの数学の得点に、クラスメートの4<sup>th</sup>-gradeの平均スコアを回帰する。

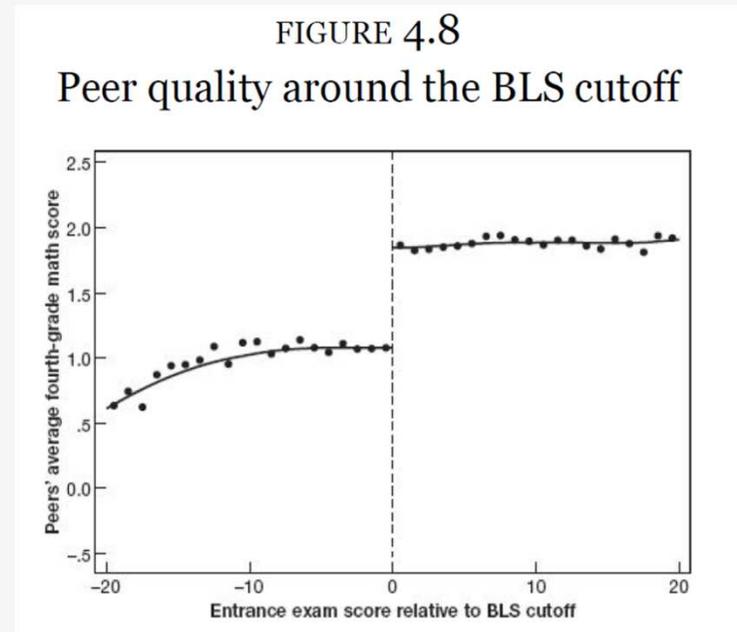
$$Y_i = \theta_0 + \theta_1 \bar{X}_{(i)} + \theta_2 X_i + u_i \quad (4.6)$$

- OLS推定量 $\hat{\theta}_1$ の推定値は0.25。
  - ピアのテストスコアの標準偏差が1変化することによって、 $0.25\sigma$ テストスコアが上昇する（？）。
- この結果をもって、ピアの質が上昇することによる効果ととらえてよいのか？

**【問題点】**

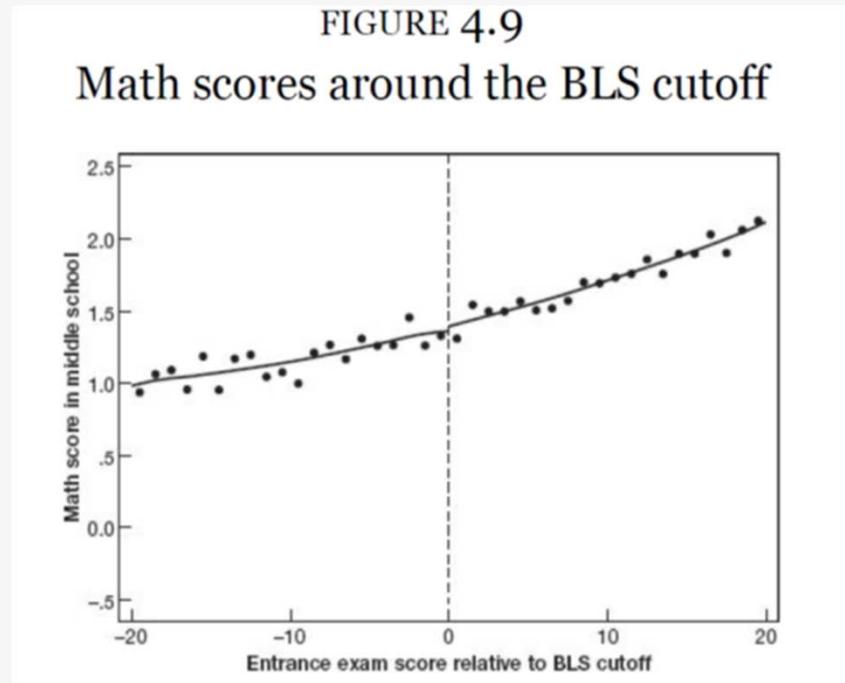
- 一見、回帰によって生徒の学力とクラスメートの学力に相関があるように見える.
- しかし、この結果を因果的に解釈することは危険.
  - 一緒に教育を受けた生徒は様々な理由により同様の傾向を持ちうる.  
例：世帯属性の類似
- OLSによる分析では、世帯属性が固定されていない。
  - 因果関係をとらえるためには、異なるクラスメートのグループにランダムに割り振らなくてはならない.





### 【対応】

- BLSの出願者のピアの能力に顕著な差がみられることを利用する。
- ピアの質の閾値でのジャンプは、ランダム割り当てによるもの。
  - この閾値付近では出願者自身の能力や、動機、世帯属性については同様のジャンプは起こっていないと考えられる。



### 【7<sup>th</sup>- & 8<sup>th</sup>-gradeでの数学の得点の分布】

- 閾値を境にして介入は強くなるが、受験者自身の入学後のスコアについて明確なジャンプはみられない。

**【ピア効果の推定①】**

- (4.2) 式と同様に、Figure 4.9でのジャンプの大きさは以下の推定式によって推定できる。

$$Y_i = \alpha_0 + \rho D_i + \beta_0 R_i + e_{0i} \quad (4.7)$$

- ここで $D_i$ は入学資格の有無を示すダミー変数であり、 $R_i$ は入学資格を決めるrunning variable (ISEEスコア)。
- これを推定すると $\hat{\rho} = -0.02$  (s.e. = 0.10)を得る (有意差なし)。
- (4.7) 式は内生変数をクラスメートの平均スコア ( $\bar{X}_{(i)}$ ) とする、2SLSの誘導形として考えることができる。

**【ピア効果の推定②】**

- 2SLSで考えると、1<sup>st</sup>-stageは以下のようにあらわすことができる。

$$\bar{X}_{(i)} = \alpha_1 + \phi D_i + \beta_1 R_i + e_{1i} \quad (4.8)$$

- ここでパラメータ $\phi$ は、入学資格を得たことで得られた平均的なピアの質として考えることができる。
- $\hat{\phi} = 0.80$ として推定される。
- つまり、入学資格を得ることによって $0.80\sigma$ ピアのスコアが上昇することを意味している。

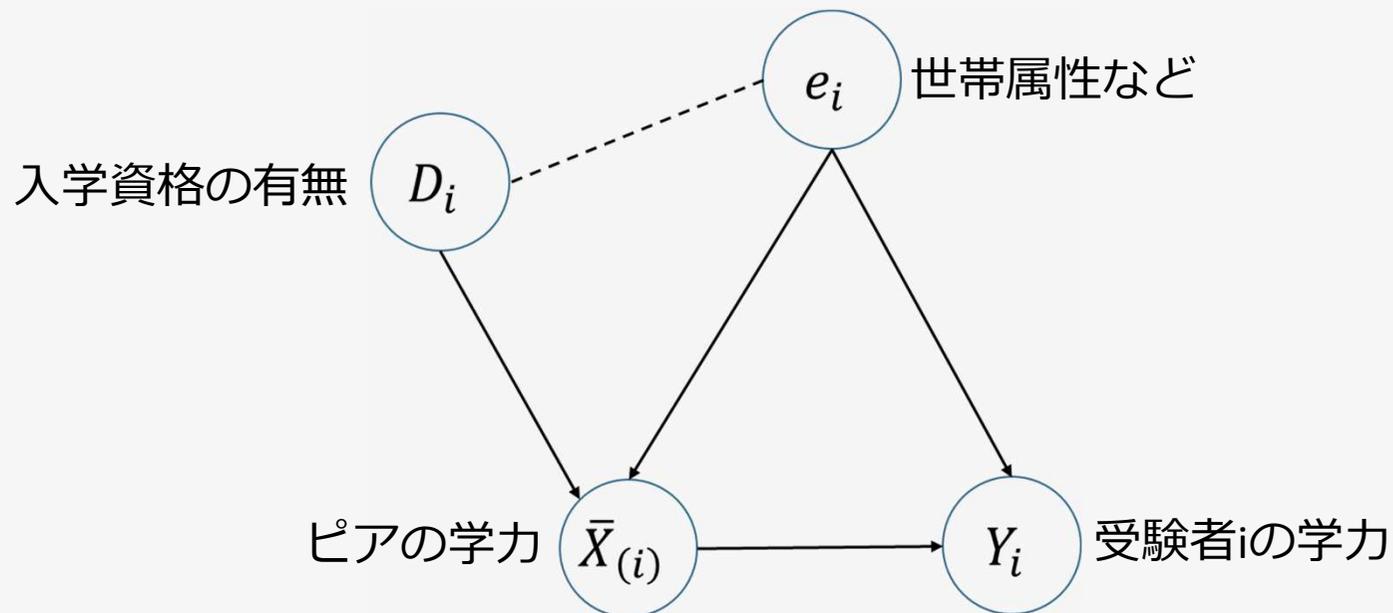
- 2<sup>nd</sup>-stageはピアの質が数学のスコアに与える効果をとらえている.
- 通常、2<sup>nd</sup>-stageでは1<sup>st</sup>-stageと同じコントロール変数を含めて推定を行う.
- 2SLSの2<sup>nd</sup>-stageについては、以下のように定式化できる.

$$Y_i = \alpha_2 + \lambda \hat{X}_{(i)} + \beta_2 R_i + e_{2i} \quad (4.9)$$

- ここで $\lambda$ はピアの質による因果効果を表している.
- また、 $\hat{X}_{(i)}$ は(4.8)式の1<sup>st</sup>-stageより得られた予測値である.

## 【因果経路】

- (4.9) 式では、1<sup>st</sup>-stage同様running variable  $R_i$ を含めている。
- しかし、ジャンプを示すダミー変数である $D_i$ については含まれていない。
  - これは $D_i$ が操作変数として扱われているから。
- 実質的に、 $R_i$ でコントロールしたうえで、閾値付近で学校の質による直接の効果がないと仮定しており、もしあるとしてもピア効果を経由する間接的なものであるとしている（除外制約）。



### 【推定結果まとめ】

- (4.9) 式の2SLS推定量 $\hat{\lambda}$ は **-0.23** (s.e.=0.132)である.
  - 誘導形の推定量 $\hat{\rho}$ はほぼ0であり、統計的に有意ではなかった.
  - 今回、2SLS推定量も統計的な有意性はなく、誘導形の推定量と同様の結果となる.
- これらの結果は、OLS推定量の推定値0.25とは異なる.
- 選抜試験のある学校による効果はピア効果だけではない.
  - 除外制約は特定の因果経路を必要とするが、実際にそれがただ一つのものである必要はない.

### 【別の観点からの選抜試験校の効果】

- 選抜試験のある高校のピア効果以外の特徴は人種の偏りである。
  - ボストンのほとんどの公立高校では、白人が多い。
  - ピアの質を白人の比率に置き換えて、2SLS推定量を求めた場合も効果はみられない。
- 良い先生やより進んだ大学レベルのカリキュラムがある可能性。
  - これらは生徒の学力に対してプラスの効果を持つと考えられる。
  - 欠落変数バイアスにより、今回の2SLS推定量は過大推定している可能性がある。

### 【注意点】

- 操作変数法での留意点同様、fuzzy RDについても操作変数が結果変数に影響を与える因果経路を精査することが必要。