

# 英語能力テストの分析に まつわる統計解析手法

---

成蹊大学理工学部情報科学科

岩崎 学

# 本日の話の流れ

- 成蹊大学で実施されている TOEIC テストの分析を例に,
  - 統計手法が実際のデータ解析にどのように用いられるか(実際家の視点)
  - 実際のデータが統計手法にどのような問題提起をするか(理論家の視点)
- 成蹊大学における TOEIC 試験の(初歩的な)結果分析
- 取り上げる統計手法
  - 統計的因果推論 (statistical causal inference)
  - 平均への回帰 (regression towards the mean)
  - 共分散分析 (analysis of covariance)
  - 回帰分断デザイン (regression discontinuity design)
  - マルチレベル分析 (multilevel analysis)
  - 集計データに基づく解析 (ecological inference)
  - 欠測データ解析 (missing data analysis)
- おわりに(「統計検定」の紹介と簡単なまとめ)

# 成蹊大学における TOEIC テスト

- 2010年度入学生から新たに開始 (TOEIC-IP)
- 1年4月, 1年12月, 2年終了時の計3回実施
  - 1年4月: 1年次の英語のクラス分けに利用
  - 1年12月: 2年次の英語のクラス分けに利用
  - 2年終了時: 確認のため実施. TOEFL を TOEIC に換算
- クラス分けの基準(理工学部情報科学科の場合)
  - 上級クラス=1, 普通クラス=4
  - TOEIC の合計点が  $c$  以上  $\Rightarrow$  上級クラス
  - TOEIC の合計点が  $c$  未満  $\Rightarrow$  普通クラス
- 本稿での利用データ: 2010年度入学生

# テスト結果の分析の論点

- 大きな目標: TOEIC テストの導入により学生の英語の力が向上したかの評価
- 個別の(小さな)目標
  - 4月のテストと12月のテストでスコアがどの程度伸びたか
  - 「上級—普通」のクラス分けは奏効したか(両クラスでの得点の伸びは?)
  - 2年次最後のテストの成績はどうか
    - 2年次最後のテストの非受験者の特質は?
  - 学科間, 学部間で相違はあるか

# 情報科学科の結果分析

- TOEIC の4月および12月の試験結果 (T-4, T-12)
- 性別, 前期の英語の成績(2科目)(Eng-T)
- 英語以外の科目の前期GPA (GPA)

	Eng-T	GPA	T-4	T-12
Eng-T	1.000			
GPA	0.636	1.000		
T-4	0.450	0.406	1.000	
T-12	0.518	0.337	0.700	1.000

# 平均への回帰

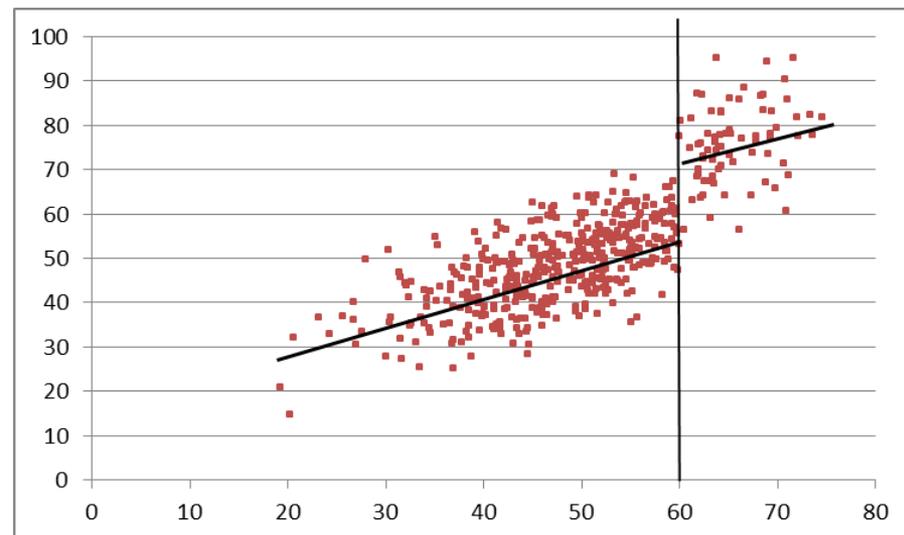
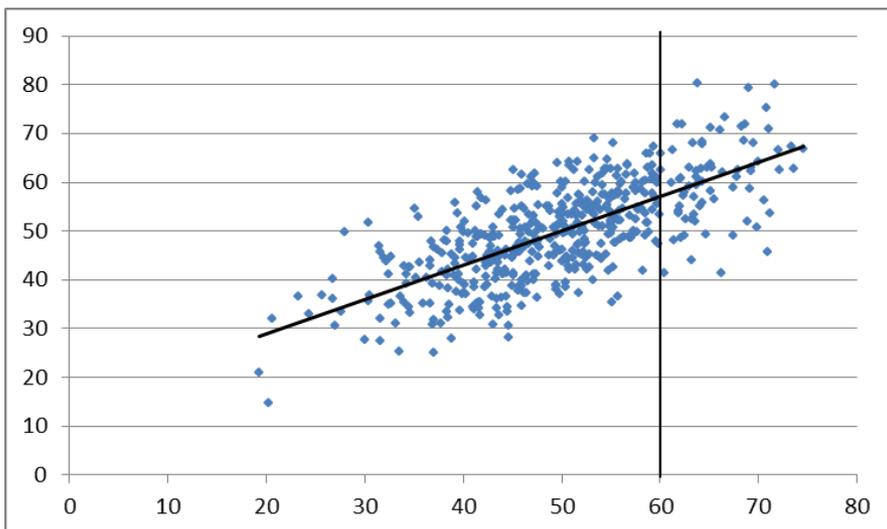
- Regression towards the mean
- $X$  の値が  $E[X]$  よりも大きな個体の  $Y$  の値は  $E[Y]$  よりも大きいものの  $Y - E[Y]$  は  $X - E[X]$  ほどは大きくない ( $E[Y]$  に近い)
- 説明変数  $X$  が  $X \geq c$  (もしくは  $X < 0$ ) でのみ観測されるとき、スクリーニングがあるという。
- 観測データの平均の差はバイアスを持つ
- 回帰モデル  $Y - \mu = \alpha + \beta(X - \mu)$  における定数項  $\alpha$  の推測に帰着 ( $\mu = E[X]$ ).
- 上級クラス:  $a = 15.35$  ( $P = 0.742$ )
- 普通クラス:  $a = 18.54$  ( $P = 0.026$ )
- 両クラスとも12月のほうが伸びているが上級クラスの伸びは統計的に有意でない。

# 回帰分断デザイン

- Regression discontinuity design (RD design)
- ある変数(割り当て変数 "assignment" variable)  $X$  およびある(既知の)定数  $c$  があり,

$$X \geq c \Rightarrow Z = 1, \quad X < c \Rightarrow Z = 0$$

- $Y(Z=1)$  と  $Y(Z=0)$  を比較(特に  $c$  の近傍)
- $\tau = E[Y(Z=1) | X = c] - E[Y(Z=0) | X = c]$



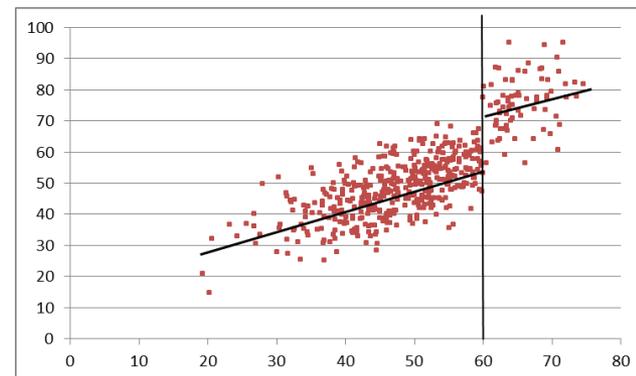
# 回帰分断デザインの例

- テストの成績が  $c$  未満の生徒をサマースクールに参加させ、そのサマースクールの効果をその後のテストで判断(補習授業の効果も同じ).
- 少人数教育の効果の判断のため、1学級の生徒数が  $c$  未満のクラスと  $c$  以上のクラスとで成績を比較.
- テストの点数が  $c$  以上の生徒に遊蕩(じゃない優等)奨学金を与えることで生徒のモチベーションは上がるか.
- ある年齢  $c$  を超えた年配者に対して博物館などの公共施設の利用料を安く設定することの効果判定.
- 年間所得が  $c$  以上の世帯には何とか手当を支給しないが、それが子供の成長に影響を与えるか.

などなど

# 回帰分断デザインの論点

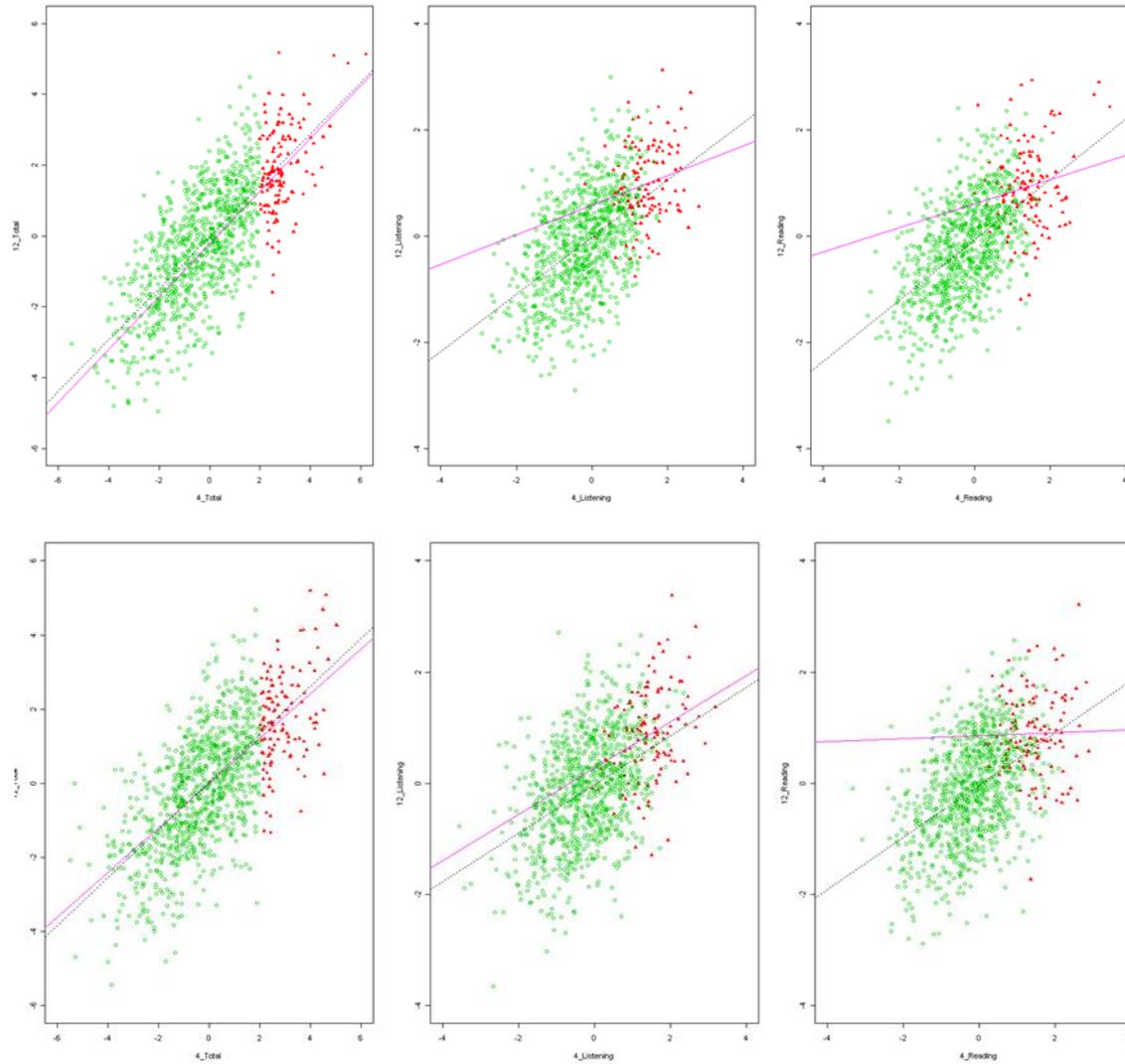
- Sharp RD or Fuzzy RD
  - $\Pr(Z = 1 | X \geq c) = 1$  or  $\Pr(Z = 1 | X \geq c) < 1$
- Manipulation of X ?
  - 各個体は自己の X の値を操作できるか.
  - Continuity on X
- Random assignment ?
  - $X = c$  の近傍では random assignment と見なせる  $\Rightarrow$  因果推論可能
  - しかし, overlap はない.
- 処置効果なしの検定は容易, 効果の大きさの推定は困難
- 局所スムージングを用いた処置効果の推定
  - $X = c$  から離れたところの個体は因果推論に関係が薄い
    - 成績があまりに悪い学生が優等奨学金を得たらどうなるか
    - 若者がシルバーパスをもらったならどう行動するか



# Composite Score による RD

- 4月のスコア  $(X^*) = \text{Listening } (X_1) + \text{Reading } (X_2)$
- 12月のスコア  $(Y^*) = \text{Listening } (Y_1) + \text{Reading } (Y_2)$
- $X^* \geq c \Rightarrow$  上級クラス,  $X^* < c \Rightarrow$  普通クラス
- 各クラスにおいて, Listening および Reading のスコアの推移  $(X_1 \rightarrow Y_1), (X_2 \rightarrow Y_2)$  が見たい.
- $X^*$  では両群で重なりはないが,  $X_1$  および  $X_2$  では重なりがある.

# Some Simulation Results



# マルチレベル分析

- マルチレベルモデル (multilevel model), 階層的モデリング (hierarchical model), 混合効果モデル (mixed effect model), 変量効果モデル (random effect model)
- 階層構造をもつデータに適用される分析法
  - 各生徒はそれぞれのクラスに属し, クラスは学校に, 学校は地域の学校群に属しているという具合
- モデル(切片変動モデル intercept-varying model) :

$$Y_i = \alpha_j + \beta x_i + \varepsilon_i \quad (j = 1, \dots, J; i = 1, \dots, n)$$

- 切片  $\alpha_j$  を一定値  $\alpha$  とみる  $\Rightarrow$  通常の回帰モデル (complete pooling)
- 切片  $\alpha_j$  を各群における定数とみる  $\Rightarrow$  共分散分析モデル (no pooling)
- 切片  $\alpha_j$  を確率変数とみる  $\Rightarrow$  マルチレベルモデル

# Ecological Regression

- 各学科の  $T(04)$  および  $T(12)$  の平均値のみのプロットおよびそれらに関する回帰直線
- 各個体間のばらつきが考慮されていないことから見かけ上の決定係数は極めて大きな値となっている
- Regression fallacy あるいは aggregation fallacy と呼ばれる現象

# 欠測データへの対処法

- 欠測のある個体は全部削除 (complete-case analysis)
- 欠測のある箇所はないものとする (available-value analysis)
- 欠測箇所を補完 (imputation, substitution)
  - 単一値代入 (single imputation)
    - 平均値, 最悪値, 回帰値, Hot Deck, Cold Deck
  - 多重代入 (multiple imputation)
- 欠測メカニズムを考慮した解析
  - MCAR = Missing Completely At Random
  - MAR = Missing At Random
  - NMAR = Not Missing At Random

# おわりに(統計検定の紹介)

- 日本統計学会公式認定
- 2011年より開始, 2012年が2回目
  - 2012年11月18日(日)実施(国際資格:2012年5月)

## 試験種目

- 1級:統計学(大学専門分野)
- 2級:統計学基礎(大学基礎科目)
- 3級:データの分析
- 4級:資料の活用
- 統計調査士:統計調査実務に関連する基本的知識
- 専門統計調査士:統計調査全般に関わる高度な専門的知識
- 国際資格:英国王立統計学会 (Royal Statistical Society =RSS) との共同認定(RSS/JSS 試験)

# おわりに(評価とは)

- 処置 (treatment), 介入 (intervention), 政策 (policy) の効果の定量的評価(統計的評価)の重要性の再認識
  - 「やりっぱなし」はもはや許されない
  - 統計的因果推論の発展
- 評価の方法論の見直しと新たな展開
  - 科研費申請中: 多種多様なデータに基づく統計的評価法の総合的研究
- 学習評価の視点(統計検定の経験から)
  - 評価は, 被評価者のためにある. 評価者のためではない.
  - 評価による, 現状への警鐘の意味
- まあ取り敢えず, 頑張りましょう!!

## 参考文献(最近の岩崎関連のもののみ)

- 岩崎 学 (2002) 不完全データの統計解析. エコノミスト社
- 岩崎 学 (2006) 統計的データ解析入門 単回帰分析. 東京図書.
- 岩崎 学・阿部貴行 (2006) 打ち切りおよびトランケーションの下でのパラメータ推定に及ぼす切断点の影響評価. 応用統計学, **35**, 1, 49-60.
- 岩崎 学・河田祐一 (2007) 処置前後研究における平均への回帰とその周辺. 日本統計学会誌シリーズJ, **36**, 2, 131-145.
- 岩崎 学・大道寺香澄 (2009) ゼロ過剰な確率モデルとそのテスト得点の解析への応用. 行動計量学, **36**, 1, 25-34.
- 岩崎 学 (2011a) 傾向スコア: その考え方と特性. 統計関連学会連合大会
- 岩崎 学 (2011b) Regression discontinuity design と因果推論. 統計関連学会連合大会.
- 岩崎 学 (2012a) 統計検定1級の目指すもの. ESTRELA, No. 216, 8-13.
- 岩崎 学 (2012b) 統計検定2級の結果分析. ESTRELA, No. 219, 6-11.
- 岩崎 学 (2012c) マルチレベル分析の考え方と実際. 統計関連学会連合大会
- 岩崎 学 (2012d) ビッグデータ時代に求められる統計解析と人材育成. IBM SPSS 統計フォーラム2012特別講演.
- 岩崎 学・吉田清隆 (2012) 統計検定: 出題傾向と結果分析. 統計関連学会連合大会市民講演会.