

## データ解析実習 ( 1 回目 ) ガイダンス



2011.10.03 月曜、4・5限

担当教員：高木英至、田端章明

### 今日のスケジュール

- ▶ (事前質問紙の実施・回収)
- ▶ I. この授業のテーマ
- ▶ II. なぜ統計が必要か？
- ▶ III. この授業の進め方

### ■ I. 授業のテーマ

- ▶ データをいじって統計（特に基礎的な多変量解析）を理解する
- ▶ データに関する感覚を養う



### 注意：「社会調査士資格」対応について

- ▶ 現在、教養学部と経済学部で共同で、社会調査士対応の授業の開設を始めました
  - ▶ 経済学部の世話人：松本先生（政治学）
  - ▶ 教養学部の世話人：高木、水野の予定

## 【参考】社会調査士資格の 標準カリキュラム

- ▶ 【A】社会調査の基本的事項に関する科目
- ▶ 【B】調査設計と実施方法に関する科目
- ▶ 【C】基本的な資料とデータの分析に関する科目
- ▶ 【D】社会調査に必要な統計学に関する科目
- ▶ **【E】量的データ解析の方法に関する科目**
- ▶ 【F】質的な分析の方法に関する科目
- ▶ 【G】社会調査の実習を中心とする科目
- ▶ ※EとFはどちらかひとつを選択
- ▶ この授業 → E中心でC、Dも少し

## 埼玉大学での対応

- ▶ 【A】社会調査の基本的事項に関する科目
  - ▶ 社会調査法Ⅰ（経済学部授業）
- ▶ 【B】調査設計と実施方法に関する科目
  - ▶ 社会調査法Ⅱ（教養学部授業）
- ▶ 【C】基本的な資料とデータの分析に関する科目
  - ▶ 統計学Ⅰ（経済学部授業）
- ▶ 【D】社会調査に必要な統計学に関する科目
  - ▶ 統計学Ⅱ（経済学部授業）
- ▶ **【E】量的データ解析の方法に関する科目**
  - ▶ **データ解析実習（この授業）**
- ▶ 【F】質的な分析の方法に関する科目
  - ▶ 質的分析法（経済学部授業）
- ▶ 【G】社会調査の実習を中心とする科目
  - ▶ 社会調査実習（教養学部授業）

▶ 6

## 従来、この授業は評判が悪かった

- ▶ なぜか？
  - ▶ ステップを追って習得するよう、複数の授業が設計されていなかった
- ▶ いきなりデータ解析をする、という無理があった
  - ▶ この授業でも同様
- ▶ 社会調査士資格対応
  - ▶ 順を追って習得させる工夫あり
- ▶ 統計の習得には次のステップが必要
  - ▶ 「記述統計学」で統計になれる（C科目）
  - ▶ 推測統計学の基礎（D科目）
  - ▶ 多変量解析（E科目）
  - ▶ 以上の実習

## 無理はあるが...

- ▶ この授業では次を扱う
  - ▶ 多変量解析の基礎（E科目）
  - ▶ 以上の実習
  - ▶ 必要に応じてD科目の内容
    - ▶ 本日の質問紙の結果による
- ▶ ここまで、質問

## ■ II. なぜ統計が必要か？

- ▶ 行動科学、社会科学の勉強には統計の知識が不可欠
- ▶ なぜなら
  - ▶ 研究自体が統計分析を前提にデザインされる
  - ▶ 知らないと、研究論文を理解できない
- ▶ 基礎的教養
  - ▶ [例] 比率 (e.g., 内閣支持率) の信頼区間
    - ▶  $P_0$ : 標本の比率       $N$ : サンプル数
    - ▶  $Z_{\alpha/2}$ : 正規偏差       $\varepsilon$ : 比率の標本誤差
    - ▶  $\varepsilon = Z_{\alpha/2} \cdot [P_0 (1 - P_0) / N]^{1/2}$ .

## 例示：データ解析の使われ方

- ▶ 卒論、2005年度卒、生澤さん
- ▶ 『東京都における少年犯罪の外的要因』
- ▶ ネットでデータを取得
  - ▶ 東京の指標データ
    - ▶ 「統計からみる東京のすがた」
  - ▶ 平成15年度の刑法犯罪データと少年犯罪のデータ
    - ▶ 「警視庁ホームページ」の「統計」
  - ▶ 分析単位＝行政区
    - ▶ 東京23区、八王子市、立川・国立市、武蔵野市、三鷹市、青梅市・奥多摩町、府中市、昭島市、調布・狛江市、町田市、小金井・国分寺市、小平市、日野市、東村山・清瀬市、東大和・武蔵村山市、東久留米・西東京市、多摩・稲城市、大島町・利島村、新島・神津島村、三宅村・御蔵島村、八丈島・青ヶ島村、小笠原村

## データ

- ▶ 都市指標データ (28データ)
  - ▶ 1. 15歳未満人口
  - ▶ 2. 15～64歳人口
  - ▶ 3. 65歳以上人口
  - ▶ 4. 外国人人口
  - ▶ 5. 世帯数
  - ▶ 6. 完全失業者数
  - ▶ 7. 第一次産業従事者数
  - ▶ 8. 第二次産業従事者数
  - ▶ 9. 第三次産業従事者数
  - ▶ 10. 事業所数
  - ▶ 11. ....
  - ▶ 28. 人口密度
- ▶ 少年犯罪補導データ (12データ)
  - ▶ 1. 占有離脱物横領補導
  - ▶ 2. 万引き補導
  - ▶ 3. 非侵入窃盗 (万引き含まず) 補導
  - ▶ 4. 侵入窃盗補導
  - ▶ 5. 恐喝補導
  - ▶ 6. 傷害補導
  - ▶ 7. 放火補導
  - ▶ 8. 殺人補導
  - ▶ 9. 暴行補導
  - ▶ 10. 脅迫補導
  - ▶ 11. 窃盗補導
  - ▶ 12. 猥褻補導

表1：少年犯罪補導件数 (15歳未満人口比) の因子分析結果

成分行列\*

|            | 成分    |       |
|------------|-------|-------|
|            | 1     | 2     |
| 占有補導／15人口  | .829  | .187  |
| 万引き補導／15人口 | .496  | -.030 |
| 非侵入補導／15人口 | .736  | .204  |
| 侵入補導／15人口  | -.198 | .709  |
| 恐喝補導／15人口  | .700  | -.146 |
| 傷害補導／15人口  | .263  | -.054 |
| 放火補導／15人口  | -.042 | .765  |
| 殺人補導／15人口  | .042  | -.212 |
| 暴行補導／15人口  | .218  | .667  |
| 脅迫補導／15人口  | -.151 | .037  |
| 盗品補導／15人口  | -.002 | .402  |
| 猥褻補導／15人口  | .739  | -.177 |

因子抽出法: 主成分分析

a. 2 個の成分が抽出されました

- 第1因子は「窃盗系補導」
- 第2因子は「暴力系補導」

表2：窃盗系補導（15歳未満人口比）の重回帰分析結果

| 係数* |          |            |        |       |        |      |
|-----|----------|------------|--------|-------|--------|------|
|     |          | 非標準化係数     |        | 標準化係数 | t      | 有意確率 |
|     |          | B          | 標準誤差   | ベータ   |        |      |
| モデル | (定数)     | -1.718     | .486   |       | -3.532 | .001 |
|     | 完全失業者／人口 | 72.654     | 19.792 | .493  | 3.671  | .001 |
| 2   | (定数)     | -1.836     | .467   |       | -3.934 | .000 |
|     | 完全失業者／人口 | 103.618    | 23.263 | .703  | 4.454  | .000 |
|     | 人口密度     | -6.291E-05 | .000   | -.359 | -2.277 | .028 |

a. 従属変数: REGR factor score 1 for analysis 1窃盗系補導／15人口

- 「完全失業者」 → 少年の窃盗系犯罪を促進
- 「人口密度」 → 少年の窃盗系犯罪を阻害

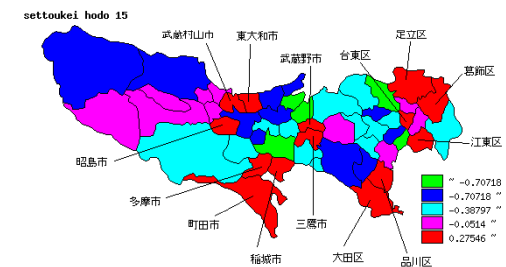


図1：区市町村における窃盗系補導率（15歳未満人口比）

表3：暴力系補導（15歳未満人口比）の重回帰分析結果

| 係数* |         |         |         |       |        |      |
|-----|---------|---------|---------|-------|--------|------|
|     |         | 非標準化係数  |         | 標準化係数 |        |      |
|     |         | B       | 標準誤差    | ベータ   |        |      |
| モデル |         |         |         | t     | 有意確率   |      |
| 1   | (定数)    | -.435   | .238    |       | -1.829 | .074 |
|     | 都市公園／人口 | 939.746 | 408.993 | .334  | 2.298  | .027 |

a. 従属変数: REGR factor score 2 for analysis 1暴力系補導／15人口

- 「都市公園」 → 少年の窃盗系犯罪を促進

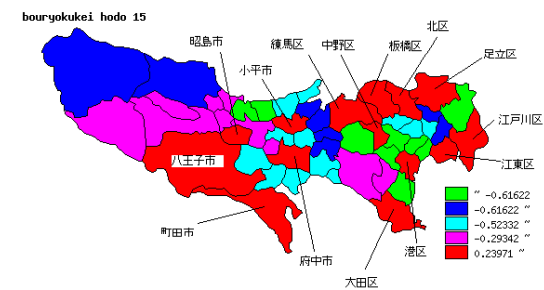


図2：区市町村における暴力系補導率（15歳未満人口比）

### ■Ⅲ. この授業の進め方

#### ▶ 参考書

- ▶ 南風原朝和 (2002) 『心理統計学の基礎』、有斐閣 (アルマ) ¥2,200+税
- ▶ 統計の基礎的考え方にに関する授業
  - ▶ ロウントリー 『新・涙なしの統計学』 新世社
- ▶ 統計の入門的授業
  - ▶ 田畑吉雄 『やさしい統計学』 現代数学社
  - ▶ 東京大学教養学部統計学教室 (編) 『統計学入門』 東京大学出版会
- ▶ 分野別の統計利用方法
  - ▶ ボーンシュテット、ノーキ 『社会統計学』 ハーベスト社

### 授業の内容 (おおまかに)



- ▶ 1. 平均値の推測・検定：10月
  - ▶ 平均値の信頼区間
  - ▶ 平均値の条件間比較 (t 検定、F 検定)
- ▶ 2. クロス表での関連の確認：11月
  - ▶ 単純クロス表 (分割表)
  - ▶ カイ二乗検定
- ▶ 3. 因子分析とクラス分析：11～12月
- ▶ 4. 重回帰分析：12月～1月
- ▶ 5. 対数線形モデル (多重クロス表分析)：1月

### 1. 平均値の推測・検定：10月

#### ▶ 平均値の信頼区間

- ▶ 例：日ごとの (大阪の気温－東京の気温) を集計すると、+0.7だった。東京の気温と大阪の気温に差があるといえるか？ (検定)

- ▶ 例：内閣支持率の信頼区間

#### ▶ 平均値の条件間比較 (t 検定、F 検定)

- ▶ 実験条件と対照条件間で、実験参加者の行動に差があるか？

### 2. クロス表での関連の確認：11月

#### ▶ 単純クロス表 (分割表)

#### ▶ カイ二乗検定

居住と血液B型のクロス表

|    |     | 血液B型   |       | 合計   |       |      |
|----|-----|--------|-------|------|-------|------|
|    |     | Bでない   | B型    |      |       |      |
| 居住 | 自宅  | 度数     | 83    | 14   |       |      |
|    |     | 期待度数   | 74.3  | 22.7 |       |      |
|    |     | 調整済み残差 | 2.6   | -2.6 |       |      |
|    | 自宅外 | 度数     | 117   | 47   |       |      |
|    |     | 期待度数   | 125.7 | 38.3 |       |      |
|    |     | 調整済み残差 | -2.6  | 2.6  |       |      |
|    |     | 合計     |       | 度数   | 200   | 61   |
|    |     |        |       | 期待度数 | 200.0 | 61.0 |

カイ二乗検定

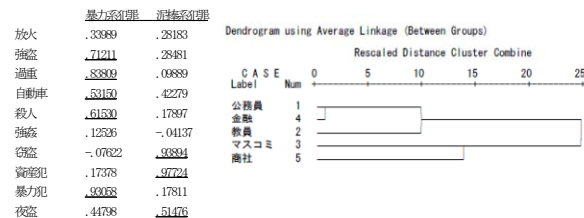
|                   | 値                  | 自由度 | 漸近有意確率 (両側) | 正確有意確率 (両側) | 正確有意確率 (片側) |
|-------------------|--------------------|-----|-------------|-------------|-------------|
| Pearson のカイ二乗     | 6.887 <sup>a</sup> | 1   | .009        |             |             |
| 連続修正 <sup>b</sup> | 6.116              | 1   | .013        |             |             |
| 尤度比               | 7.260              | 1   | .007        |             |             |
| Fisher の直接法       |                    |     |             | .010        | .006        |
| 線型と線型による連関        | 6.861              | 1   | .009        |             |             |
| 有効なケースの数          | 261                |     |             |             |             |

a. 2x2 表に対してのみ計算

b. 0セル (.0%) は期待度数が 5 未満です。最小期待度数は 22.67 です。

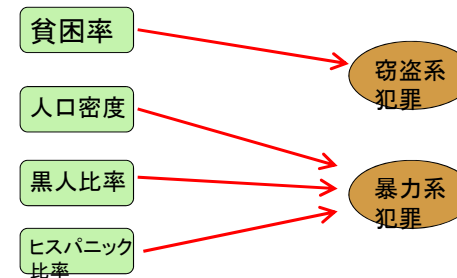
### 3. 因子分析とクラスタ分析：11～12月

犯罪数値の因子分析（主因子法・ヴァリマックス回転）



### 4. 重回帰分析：12月～1月

▶ 例：地区ごとの犯罪率を説明する要因は何か？



### 授業の方法



- ▶ 講義＋実習
  - ▶ 参考書で関連箇所を事前に読んでおく
  - ▶ 一定の頻度で課題を出す
- ▶ 成績
  - ▶ 課題レポート、授業中の質問紙(one-minute paper)の評価点の合計
  - ▶ 提出遅れ＝減点
  - ▶ 課題を期限内に出すことが重要
- ▶ 高木への連絡
  - ▶ etakagi@mail.saitama-u.ac.jp

### 次回（10/17）までにしておくこと

- ▶ 10/10は休み
- ▶ 10/17までに、参考書を入手しておくこと
  - ▶ 南風原朝和（2002）『心理統計学の基礎』、有斐閣（アルマ）
- ▶ 10/17までに次の事項を確認しておく
  - ▶ 1. 統計の基礎概念にあたる箇所
    - ▶ 第1、2章（特に2章の3）、第4章（4章の4は分からなくてよいです。）
  - ▶ 2. 検定と推定の考え方 テキストの第5章
  - ▶ 3. 相関係数の概念（第3章）



今日は  
おしまい