



データ解析実習 ガイダンス

2012.10.01 月曜、4・5限

担当教員：高木英至

1

今日のスケジュール

▶ この授業は

- ▶ 3年生以下 → 講義番号 F12051
- ▶ 4年生以上 → 講義番号 L22061

▶ 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

▶ 2

■ 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



▶ 3

【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

▶ 4

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
 - ▶ 消えた
- ▶ 【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）

▶ 5

いわゆる「統計学」として学ぶべきこと

領域	内容	社会調査士科目
記述統計学	データをまとめる	C科目
推測統計学 (数理統計学)	確率論の応用 推測、検定の基礎	D科目
多変量解析	多数の変数の処理 複雑なデータを読み解く	E科目

- 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C科目、D科目の内容



▶ 6

■ 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 : 標本サイズ (サンプル数)
 - $Z_{\alpha/2}$: 正規偏差
 - ε : 比率の標本誤差

$$\varepsilon = Z_{\alpha/2} \sqrt{\frac{P_0(1-P_0)}{N}}$$

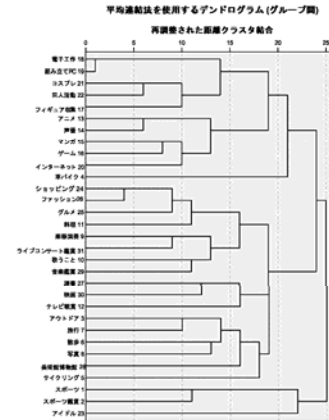
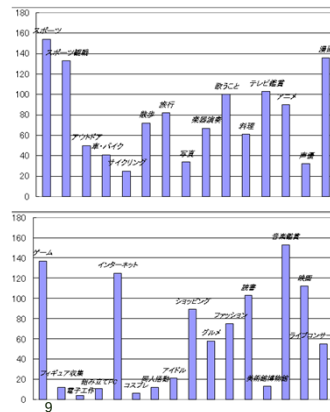
▶ 7

例示：データ解析の使われ方

- ▶ 昨年度の卒論、山下かほりさん
- ▶ 『趣味とおたく態度および社会的スキルの関連』
- ▶ 学生対象の調査
 - ▶ F1 性別 (1.男性 2.女性)
 - ▶ F2 学年 (1.1年次 2.2年次 3.3年次 4.4年次以降)
 - ▶ Q1 趣味のジャンルに関する質問
 - ▶ Q2 一日の平均趣味時間
 - ▶ Q3 月の平均趣味金額
 - ▶ Q4 サブカルチャーの趣味に関する質問
 - ▶ Q5 おたく態度尺度
 - ▶ Q6 Big five 尺度
 - ▶ Q7 KISS 尺度—社会的スキル
 - ▶ Q8 大学の友人に求める機能
 - ▶ Q9 趣味で知り合った友人に求める機能
 - ▶ Q10 友人関係尺度
 - ▶ Q11 早稲田シャイネス尺度
 - ▶ Q12 シャイ行動尺度

▶ 8

趣味のジャンルの度数分布



8つのグループに分類

- ▶ 「能動のおたく系趣味」(18~17)
 - ▶ →電子工作、組み立てP C、コスプレ、同人活動、フィギュア収集
- ▶ 「受動のおたく系趣味」(13~20)
 - ▶ →アニメ、声優、マンガ、ゲーム、インターネット
- ▶ 「車・バイク」(4) →車・バイク
- ▶ 「満足系行動の趣味」(24~29)
 - ▶ →ショッピング、ファッション、グルメ、料理、楽器演奏、ライブ・コンサート鑑賞、歌うこと、音楽鑑賞
- ▶ 「鑑賞系趣味」(27~12) →読書、映画、テレビ観賞
- ▶ 「アクティブ系趣味」(3~5)
 - ▶ →アウトドア、旅行、散歩、写真、美術館博物館、サイクリング
- ▶ 「スポーツ系趣味」(1~2) →スポーツ、スポーツ鑑賞
- ▶ 「アイドル」(23) →アイドル

▶ 10

おたく態度尺度の因子分析

因子行列 ^{a)}	因子	
	1	2
お金、時間	.433	.580
手元に置く	.386	.342
語りたい	.799	-.006
情報を発信する	.448	.112
一緒に過ごす	.763	-.173
共感してほしい	.706	-.290
趣味を貫く	.244	.263

- ▶ 1. 「趣味の群れ」因子
 - ▶ 趣味を通して人と交流したいという因子
- ▶ 2. 「趣味没頭」因子
 - ▶ 趣味そのものに対する熱中度合いの因子
- ▶ 3. 「情報を発信する」

▶ 11

おたく系趣味の程度を態度要因で予測する 重回帰分析、変数選択

▶ 能動のおたく系趣味

- ▶ 能動のおたくに走るのは趣味に没頭するタイプの人

モデル	平方和	自由度	平均平方	F 値	有意確率
1 回帰	2.769	1	2.769	8.372	.004(a)
残差	95.248	288	.331		
全体	98.017	289			

モデル	非標準化係数	標準化係数	t	有意確率
1 (定数)	-.310	.164	-1.887	.060
趣味没頭	.055	.019	.168	2.893

▶ 受動のおたく系趣味

- ▶ 受動のおたくに走るのは、会話スキルに乏しく、趣味没頭、趣味で人と接する、という嗜好のある人

モデル	平方和	自由度	平均平方	F 値	有意確率
3 回帰	49.846	3	16.615	7.195	.000(c)
残差	633.513	283	2.309		
全体	703.359	286			

モデル	非標準化係数	標準化係数	t	有意確率
3 (定数)	.856	.559	1.531	.127
会話スキル	-.073	.024	-.176	3.066
趣味没頭	.114	.054	.129	2.119
趣味の群れ	.089	.044	.122	2.009

▶ 12

- ▶ 満足系行動の趣味
- ▶ 満足系行動に走るのは、会話スキルがあって情報発信に関心がある人

平方和	自由度	平均平方	F 値	有意確率
回帰	2	67.794	16.315	.000(b)
残差	284	4.155		
全体	286			

因子 2	(定数)	非標準化係数	標準化係数		t	有意確率
		B	標準誤差	β		
		-.057	.429		-.132	.89
	会話ス キル	.138	.032	.243	4.296	.000
	情報発 信	.388	.118	.186	3.293	.001

- ▶ アイドル
- ▶ アイドルに走るのは、趣味で群れたい嗜好のある人

平方和	自由度	平均平方	F 値	有意確率
回帰	1	.405	6.142	.014(a)
残差	289	.066		
全体	290			

モデル		非標準化係数		標準化係数	t	有意確率
		B	標準誤差	ベータ		
1	(定数)	-.081	.064		-1.274	.204
	趣味の群れ	.017	.007	.144	2.478	.014

▶ 13

■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

▶ 14

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. 因子分析とクラスター分析：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

▶ 15

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？

▶ 16

2. クロス表での関連の確認：11月

▶ 単純クロス表（分割表）

▶ カイニ乗検定

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	97
	期待度数		74.3	22.7	97.0
	調整済み残差		2.6	-2.6	
自宅外	度数		117	47	164
	期待度数		125.7	38.3	164.0
	調整済み残差		-2.6	2.6	
合計	度数		200	61	261
	期待度数		200.0	61.0	261.0

カイニ乗検定

	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 のカイニ乗	6.887 ^a	1	.009		
連続修正 ^b	6.116	1	.013		
尤度比	7.260	1	.007		
Fisher の直接法	6.861	1	.009	.010	.006
線型と線型による連関有効なケースの数	261				

a. 2x2 表に対してのみ計算

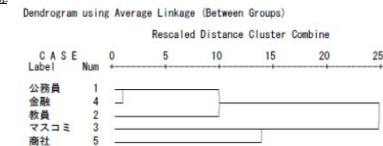
b. 0 セル (.0%) は期待度数が 5 未満です。最小期待度数は 22.67 です。

▶ 17

3. 因子分析とクラスタ分析：11～12月

犯罪類型の因子分析（主成分法-ヴァリマックス回転）

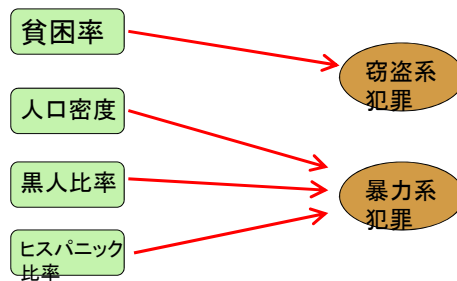
	暴力系犯罪	経済犯罪
放火	.33989	.28183
強盗	.71211	.28481
過重	.83802	.09889
自動車	.53130	.42279
殺人	.61530	.17897
強姦	.12536	-.04137
窃盗	-.07622	.93894
資信犯	.17378	.97724
暴力犯	.93058	.17811
夜盗	.44798	.51426



▶ 18

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？



▶ 19

授業の方法



- ▶ 受講者数にもよるが...
- ▶ 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
 - ▶ 課題レポート評価点の合計
 - ▶ 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

▶ 20

次回（10/15）までにしておくこと

- ▶ 来週10/8は休み
- ▶ 10/15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）→他でもよい
- ▶ 10/15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）

▶ 21



今日はおしまい

▶ 22