



データ解析実習

2012.11.05 月曜、4・5限

担当教員：高木英至

1

■本日の授業

- ▶ 初歩的な分析方法のまとめ
 - ▶ カテゴリー変数：度数分布、クロス表
 - ▶ 数値型変数：基本統計量、相関係数、分散分析（一元、二元）
- ▶ 4限
 - ▶ 講義：クロス表分析
- ▶ 5限
 - ▶ 実習



▶ 2

次回までの宿題

- ▶ 各自に割り当てられたデータを用い、SPSSで「初歩的な分析」を実施し、その結果をレポート（ワードファイル）にまとめよ。
- ▶ 目標：データが持っている情報を、この分析手法の範囲で把握する
- ▶ 次回の授業の折にファイルで提出
 - ▶ USBで持参のこと
- ▶ 次回ないし次々回に（必要なら）解説します。
- ▶ 次回の予定
 - ▶ 因子分析：講義（と実習？）

▶ 3

■単純クロス表分析

- ▶ 例：2004年度の卒論用の調査
 - ▶ 「社会心理学入門」受講者（主に1年生）を対象とした質問紙調査
- ▶ 用いる変数＝カテゴリー変数
 - ▶ 性別（f1）： 1＝男、2＝女
 - ▶ 学年（f2）： 1＝1年、2＝2年以上
 - ▶ 住居（f3）： 1＝自宅、2＝自宅外
 - ▶ 血液型（q13）：
 - ▶ 1＝A型、2＝B型、3＝O型、4＝AB型
- ▶ これらの変数間の単純クロス表をとる

▶ 4

学年と性別は関連する。
男は1年が多く、女は2年以上が多い傾向がある。

性別と学年のクロス表

		学年		合計
		1年	2年以上	
性別 男	度数	122	16	138
	期待度数	115.2	22.8	138.0
女	度数	100	28	128
	期待度数	106.8	21.2	128.0
合計	度数	222	44	266
	期待度数	222.0	44.0	266.0

カイ2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2乗	5.084 ^a	1	.024		
連続修正 ^b	4.367	1	.037		
尤度比	5.122	1	.024		
Fisher の直接法				.031	.018
線型と線型による連関	5.065	1	.024		
有効なケースの数	266				

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 21.17 です。

5%水準
で有意

▶ 5

カイ二乗値 (χ^2)

▶ 観測度数と期待度数

- ▶ 期待度数：2 変数間に関連がない（確率的に独立である）ときの度数の期待値

$$\text{カイ二乗値}(\chi^2) = \sum \frac{(\text{観測度数} - \text{期待度数})^2}{\text{期待度数}}$$

$$0 \leq \chi^2 < +\infty$$

▶ 6

例：性別と血液型

性別と血液型のクロス表

		血液型				合計
		A	B	O	AB	
性別 男	度数	52	29	35	13	129
	期待度数	49.3	31.0	37.1	11.7	129.0
女	度数	45	32	38	10	125
	期待度数	47.7	30.0	35.9	11.3	125.0
合計	度数	97	61	73	23	254
	期待度数	97.0	61.0	73.0	23.0	254.0

カイ2乗検定

	値	自由度	漸近有意確率(両側)
Pearson のカイ2乗	1.105 ^a	3	.776
尤度比	1.106	3	.776
線型と線型による連関	.072	1	.788
有効なケースの数	254		

a. 0 セル (.0%) は期待度数が 5 未満です。最小期待度数は 11.32 です。

有意に
ならない

▶ 7

注意点 1：期待度数 5

- ▶ カイ二乗検定では、期待度数が 5 未満のセルがあるのはまずい。

居住と血液型のクロス表

		血液型				合計
		A	B	O	AB	
居住 自宅	度数	42	14	27	9	92
	期待度数	35.1	22.1	26.4	8.3	92.0
自宅外	度数	55	47	46	14	162
	期待度数	61.9	38.9	46.6	14.7	162.0
合計	度数	97	61	73	23	254
	期待度数	97.0	61.0	73.0	23.0	254.0

カイ2乗検定

	値	自由度	漸近有意確率(両側)
Pearson のカイ2乗	6.856 ^a	3	.077
尤度比	7.147	3	.067
線型と線型による連関	.409	1	.522
有効なケースの数	254		

a. 0 セル (.0%) は期待度数が 5 未満です。最小期待度数は 8.33 です。

▶ 8

なぜ5以上が必要か？

- ▶ カイ二乗検定に使うカイニ乗分布は漸近分布であり、期待度数5未満のセルがあると漸近であることが保証されない
- ▶ Fisherの直説法(exact test)を使えば問題ない
 - ▶ ただし 2×2のクロス表のときだけ適用可能

性別と学年のクロス表					カイ二乗検定				
性別	学年	1年		合計	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)
		1年	2年以上						
男	度数	122	16	138	Pearson のカイ二乗	5.030 ^a	1	.024	
	期待度数	115.2	22.8	138.0	連続修正 ^b	4.367	1	.037	
女	度数	100	28	128	尤度比	5.122	1	.024	
	期待度数	106.8	21.2	128.0	Fisher の直接法			.031	.018
合計	度数	222	44	266	線型と線型による連関	5.085	1	.024	
	期待度数	222.0	44.0	266.0	有効なケースの数	266			

a. 0セル(.0%)は期待度数が5未満です。最小期待度数は21.17です。

▶ 9

注意点 2：残差分析

- ▶ どのセルで有意差が生じたか？
- ▶ 残差分析
- ▶ 調整済み残差（正規分布）
 - ▶ 絶対値が1.96以上ならそのセルで有意

性別と学年のクロス表				
性別	学年	1年		合計
		1年	2年以上	
男	度数	122	16	138
	期待度数	115.2	22.8	138.0
女	度数	100	28	128
	期待度数	106.8	21.2	128.0
合計	度数	222	44	266
	期待度数	222.0	44.0	266.0

▶ 10

居住と血液型のクロス表

居住	血液型	血液型				合計
		A	B	O	AB	
自宅	度数	42	14	27	9	92
	期待度数	35.1	22.1	26.4	8.3	92.0
自宅外	度数	55	47	46	14	162
	期待度数	61.9	38.9	46.6	14.7	162.0
合計	度数	97	61	73	23	254
	期待度数	97.0	61.0	73.0	23.0	254.0

カイ二乗検定

	値	自由度	漸近有意確率 (両側)
Pearson のカイ二乗	6.856 ^a	3	.077
尤度比	7.147	3	.067
線型と線型による連関	.409	1	.522
有効なケースの数	254		

a. 0セル(.0%)は期待度数が5未満です。最小期待度数は8.33です。

▶ 11

居住と血液B型のクロス表

居住	血液B型	血液B型		合計
		Bでない	B型	
自宅	度数	83	14	97
	期待度数	74.3	22.7	97.0
自宅外	度数	117	47	164
	期待度数	125.7	38.3	164.0
合計	度数	200	61	261
	期待度数	200.0	61.0	261.0

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 ^a	1	.009		
連続修正 ^b	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。

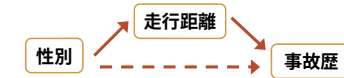
▶ 12

■ 3重クロス表をとる

- ▶ クロス表のエラボレイション（精緻化）
- ▶ ここでの例示
 - ▶ 媒介関係を見出す
 - ▶ 疑似関係を見出す
 - ▶ 交互作用効果を見出す
- ▶ 3重（以上の）クロス表での検定は、対数線形モデルなどで行う必要あり。

▶ 13

媒介関係



- ▶ 性別×事故歴有無のクロス表（右表）
 - ▶ 男性に事故歴が多い
 - ▶ 男性が事故が多いと言えるべきか？
- ▶ 走行距離の長短を層変数（第3変数）とした3重クロス表（右表）
 - ▶ 走行距離の長短が性別×事故歴有無の関連を媒介しているの分かる

	事故歴あり	事故歴なし	合計	
男性	63	137	200	$\phi=0.156$
女性	36	164	200	
全体	99	301	400	

	事故歴あり	事故歴なし	合計	
男性	53	57	110	$\phi=-0.003$
女性	17	18	35	
全体	70	75	145	

	事故歴あり	事故歴なし	合計	
男性	10	80	90	$\phi=-0.006$
女性	19	146	165	
全体	29	226	255	

▶ 14

疑似関係



- ▶ 婚別×キャンディ嗜好のクロス表（右表）
 - ▶ 既婚者はキャンディを食べない
 - ▶ 結婚するとキャンディを食べなくなる？
- ▶ 年齢を層変数（第3変数）とした3重クロス表（右表）
 - ▶ 若い方が未婚が多い
 - ▶ 若い方がキャンディを食べる
- ▶ 婚別×キャンディ嗜好の関連は、年齢から生じた疑似関係

	キャンディ食べる	キャンディ食べない	合計	
未婚	103	35	138	$\phi=0.125$
既婚	163	99	262	
全体	266	134	400	

	キャンディ食べる	キャンディ食べない	合計	
未婚	83	20	103	$\phi=0.041$
既婚	46	13	59	
全体	129	33	162	

	キャンディ食べる	キャンディ食べない	合計	
未婚	20	15	35	$\phi=0.012$
既婚	117	86	203	
全体	137	101	238	

▶ 15

交互作用効果(interaction effect)

- ▶ 友人数×携帯利用のクロス表（右表）
 - ▶ 一見すると関連がない
 - ▶ 変ではないか？
- ▶ 年齢を層変数（第3変数）とした3重クロス表（右表）
 - ▶ 交互作用があった
 - ▶ 若年層では正の関連
 - ▶ 中高年層では負の関連
 - ▶ 疑似無関連

	携帯利用多	携帯利用少	合計	
友人多	100	80	180	$\phi=0.010$
友人少	120	100	220	
全体	220	180	400	

	携帯利用多	携帯利用少	合計	
友人多	80	20	100	$\phi=0.218$
友人少	60	40	100	
全体	140	60	200	

	携帯利用多	携帯利用少	合計	
友人多	20	60	80	$\phi=-0.250$
友人少	60	60	120	
全体	80	120	200	

▶ 16



4限は おしまい