分割表 contingency table

名義尺度の変数　複数

該当するデータの個数

表

ほとんど、どの分野でも使う（よく使う）

目次

分割表

	生存	死亡
薬与える	120	115
与えない	118	22

個体数

	収入に 激しく不満	ややや不満	どちらでもない	やや満足	激しく満足
地域A	5200	3421	2988	1290	20
地域B	2466	143	235	456	12
地域C	8901	9806	7821	234	0

個体数（人数）

分割表　　ではない

	池からの サンプル	川からの サンプル
粒が大きい	120	759
粒が小さい	880	241

単位：グラム

120000	759000
880000	241000

単位：ミリグラム

分割表

大きさ

	生存	死亡
薬与える	120	115
与えない	118	22

2×2

	激しく不満	やや不満	どちらでもない	やや満足	激しく満足
地域A	5200	3421	2988	1290	20
地域B	2466	143	235	456	12
地域C	8901	9806	7821	234	0

3×5

次元の数＝２

分割表　　大きさ　変数の数

メス

	生存	死亡
薬与える	50	65
与えない	48	12

オス

	生存	死亡
薬与える	70	50
与えない	70	10

$2 \times 2 \times 2$

３次元、３元、3way

それぞれの名義変数ー順序がある／ない

	激しく不満	やや不満	どちらでもない	やや満足	激しく満足
地域A	5200	3421	2988	1290	20
地域B	2466	143	235	456	12
地域C	8901	9806	7821	234	0

３×５　　５の方に順序

３の方？

3カテゴリー以上（2以外）では

１つの変数内のカテゴリーの順番

意味がある場合／ない場合

傾向性仮説、自然な順序

ordered、order-restriction、monotone trend

分割表

次元　　　名義変数の数

各名義変数のカテゴリー数

３以上なら順序が意味を持つかも

	激しく不満	やや不満	どちらでもない	やや満足	激しく満足
地域A	15	21	20	19	10
地域B	18	19	23	31	15

2×k分割表なら、Mann-WhitneyのU検定

（Wilcoxon順位和検定）

２×２分割表

a	b
c	d

２×２分割表

	生存	死亡
要因あり	a	b
要因なし	c	d

	メス	オス
場所A	a	b
場所B	c	d

2×2分割表

	生存	死亡
要因あり	a	b
要因なし	c	d

	生存	死亡
要因あり	a	b
要因なし	c	

	生存	死亡
要因あり	a	b
要因なし		

	生存	死亡
要因あり	a	
要因なし	c	

	生存	死亡
要因あり	a	
要因なし		

２×２分割表

a	b
c	d

最小

カテゴリーの順序を考える
必要がない（２だから）

２×２分割表

カイ２乗検定　G検定　Fisherの正確確率検定

a	b
c	d

仮説（帰無仮説）

モデル

独立性

$$a:b=c:d$$
$$ad=bc$$
$$a:c=b:d$$

行や列を定数倍しても
変わらない

連関の指標

要約統計量

オッズ比　対数オッズ比

きわめて多種類

Pearsonのϕ

YuleのQ

CramerのV

C係数

テトラコリック（四分）相関係数

四分点相関係数

エントロピー指数

オッズ比 　　　対数オッズ比

odds ratio

独立性⇔1 　　　独立性⇔0

行や列を定数倍しても
変わらない

	生存	死亡
メス	a	b
オス	c	d

オッズーある結果はもう１つの結果
の何倍起こりやすいか

オッズの例ーメスでは生存は死亡の
の何倍起こりやすいか

a/b

オッズ比の例ーメスのオッズ／オスのオッズ

$$\frac{ad}{bc}$$

オッズ比　　　対数オッズ比

odds ratio

独立性⇔ 1　　　独立性⇔ 0

行列入れ替えてもよい

a	b
c	d

分子と分母のカテゴリーの入れ替え

オッズ	a/b	b/a	逆数
オッズ比	$\frac{ad}{bc}$	$\frac{bc}{ad}$	逆数
対数オッズ比	絶対値は同じ 正負が反対に		

2 × 2 分割表

	生存	死亡
要因あり	a	b
要因なし	c	d

相対リスク

＝（要因ありの死亡率）/（要因なしの死亡率）

絶対リスク

＝（要因ありの死亡率）ー（要因なしの死亡率）

超過リスク＝絶対リスク/（1 – 要因なしの死亡率）

sensitivity敏感度

		診断	
		陽性	陰性
本当は	あり	a	b
	なし	c	d

specificity特異度

分割表のカイ２乗検定

観察された数
$$\mathrm{O}$$

独立性
帰無仮説での期待値
$$\mathrm{E}$$

a	b
c	d

$$\frac{(a+b)(a+c)}{(a+b+c+d)}$$

カイ２乗と呼ばれる量 $$\frac{(\mathrm{O}\text{-}\mathrm{E})^2}{\mathrm{E}}$$ を全セルについて合計

分割表のカイ２乗検定

観察された数　　帰無仮説での期待値

O　　E

２つの「カイ２乗」

カイ２乗分布　**平均0で分散1の正規分布をする変数の２乗の分布**

自由度1

カイ２乗と呼ばれる量 $$\frac{(O-E)^2}{E}$$ を全セルについて合計

独立性（帰無仮説）

自由度(カテゴリー数-1)・(カテゴリー数-1)のカイ２乗分布

２×２なら自由度1のカイ２乗分布

分割表のカイ２乗検定

カイ２乗と呼ばれる量

$$\frac{(O-E)^2}{E}$$ の合計

カイ２乗分布

自由度1のカイ２乗分布

近似が悪い　よく言われる伝統的な制限条件

期待値１以下のセルはあってはいけない

期待値５以下のセルは20パーセント以下

サンプルサイズは200より大きい（2×2のとき）

Fisherの正確確率検定

exact probability test

帰無仮説

独立性

a	b
c	d

確率 $$\frac{(a+b)!(c+d)!(a+c)!(b+d)!}{a!b!c!d!(a+b+c+d)!}$$

２×２分割表の確率モデル

Fisherの正確確率検定　カイ２乗検定　G検定

独立性が成り立つとき　対数オッズ比=0

それ以外の状況

もう少し複雑な状況

確率モデルが必要

２×２分割表の確率モデル

データ１つずつ結果が独立に決まる

overdispersionなしと仮定

多項分布モデル

ポアソン分布モデル

２つの二項分布モデル

超幾何分布モデル

Fisherの正確確率検定　　カイ２乗検定　　G検定

overdispersionなしと仮定

２×２分割表の確率モデル

多項分布モデル　４つのできごとがある確率で起こる

確率の和＝１

ポアソン分布モデル　４つの結果が起こる回数が

ポアソン分布

２つの二項分布モデル

２項分布（例．生存or死亡）が２つ

超幾何分布モデル

Fisherの正確確率検定

２×２分割表の確率モデル

オッズ比（対数オッズ比）に関する推論では同じ

超幾何分布モデルに帰着する

多項分布モデル

ポアソン分布モデル

２つの二項分布モデル

超幾何分布モデル

対数線形モデル

log-linear models

一般化線形モデル

ポアソン回帰の一種

ポアソン分布モデル

カテゴリー数、次元が
大きくても使える

a	b
c	d

独立性　　交互作用

独立モデル（加法モデル）

log（目的変数の期待値）＝行の効果＋列の効果＋切片

この２モデルの比較で交互作用

飽和モデル

log（目的変数の期待値）＝行の効果＋列の効果＋交互作用の効果＋切片

2×2分割表

2つの名義変数は0-1

飽和モデル

a	b
c	d

β_0	$\beta_0 + \beta_1$
$\beta_0 + \beta_2$	$\beta_0 + \beta_1 + \beta_2 + \beta_{12}$

交互作用の効果

交互作用の項の係数β_{12}＝対数オッズ比

ロジスティック回帰？

	生存	死亡
要因あり	a	b
要因なし	c	d

交互作用あり＝対数オッズ比が0でない

要因の有無により生存率が異なる

決定論的な部分

目的変数の期待値　説明変数

関数形

対数線形モデル（ポアソン回帰）

$$e^{\beta_0+\beta_1}$$

（例．生存）

$$e^{\beta_0}$$

（例．死亡）

$$\frac{e^{\beta_0+\beta_1}}{e^{\beta_0}+e^{\beta_0+\beta_1}}=\frac{e^{\beta_1}}{1+e^{\beta_1}}$$

ロジスティック回帰と同じ

目的変数の分布

誤差構造

確率論的な部分

対数線形モデル（ポアソン回帰）

$$\frac{e^{-\mu_1}\mu_1^{y}}{y!}$$

（例．生存）

$$\frac{e^{-\mu_2}\mu_2^{n-y}}{(n-y)!}$$

（例．死亡）

（例．n個体のうち、生存y個体、死亡n-y個体）

$$\frac{n!}{y!(n-y)!}\left(\frac{\mu_1}{\mu_1+\mu_2}\right)^{y}\left(\frac{\mu_2}{\mu_1+\mu_2}\right)^{n-y}$$

二項分布

ロジスティック回帰と同じ

シンプソンのパラドクス

E.H.Simpson 多様度指数

G.U.Yule

	要因あり	要因なし
生存	970	9000
死亡	30	1000

対数オッズ比 **1.279**

要因ありの方が生存率高

	要因あり	要因なし
生存	5000	20
死亡	5000	80

対数オッズ比 **1.386**

要因ありの方が生存率高

プール

	要因あり	要因なし
生存	5970	9020
死亡	5030	1080

対数オッズ比 **-1.95**

要因ありの方が生存率低

似た傾向の２つの分割表をプールして１つの分割表
ちがった傾向

80	16
16	160

対数オッズ比1.609

50	452
10	452

対数オッズ比1.609

プール

130	468
170	612

対数オッズ比0

同じ関係がある２つを、
プールすると関係が無くなる

70	30
70	30

対数オッズ比0

10	66
170	1122

対数オッズ比0

プール

80	96
240	1152

対数オッズ比1.386

それぞれ関係がない（独立性）２つを、
プールすると関係が生じる

シンプソンのパラドクス

E.H.Simpson

似た傾向の２つの分割表をプールして１つの分割表

→ ちがった傾向

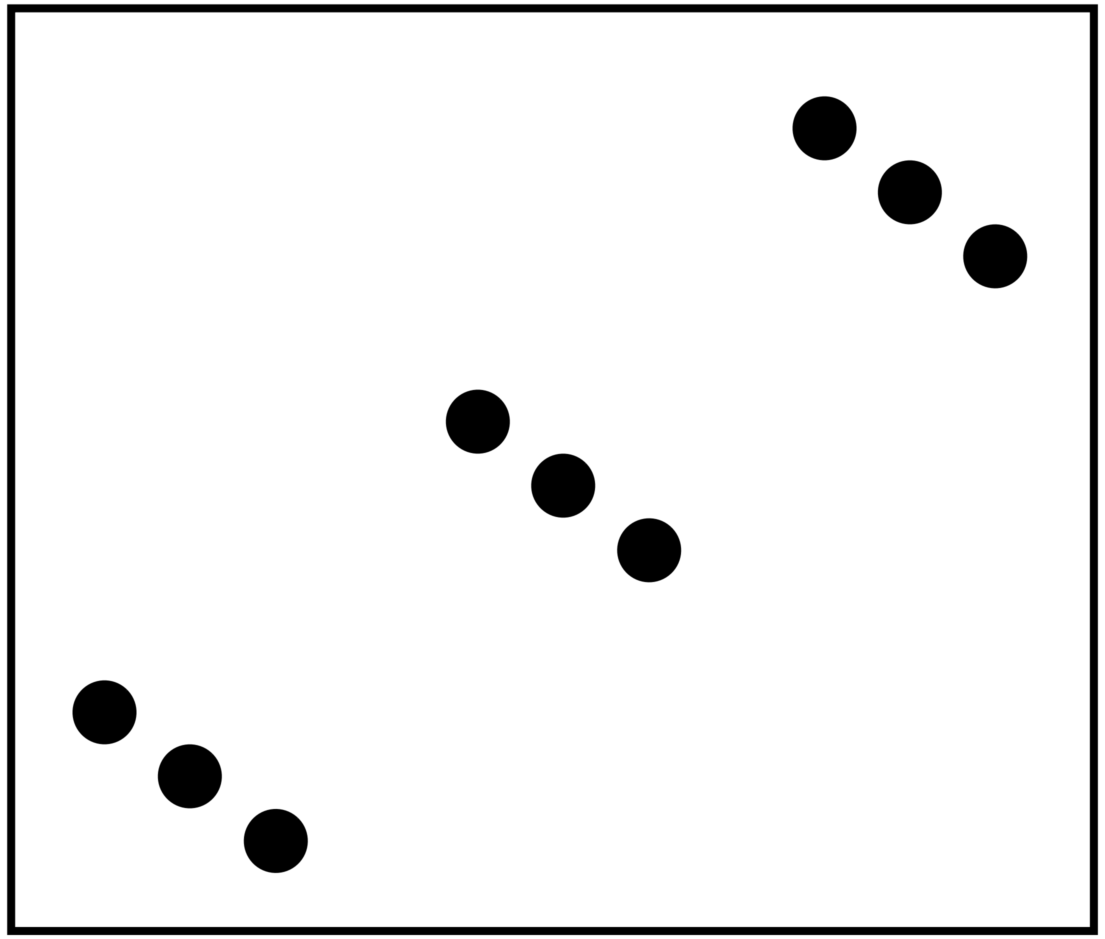
例

同じ関係　同じ関係 $\xrightarrow{\text{プール}}$ 逆の関係

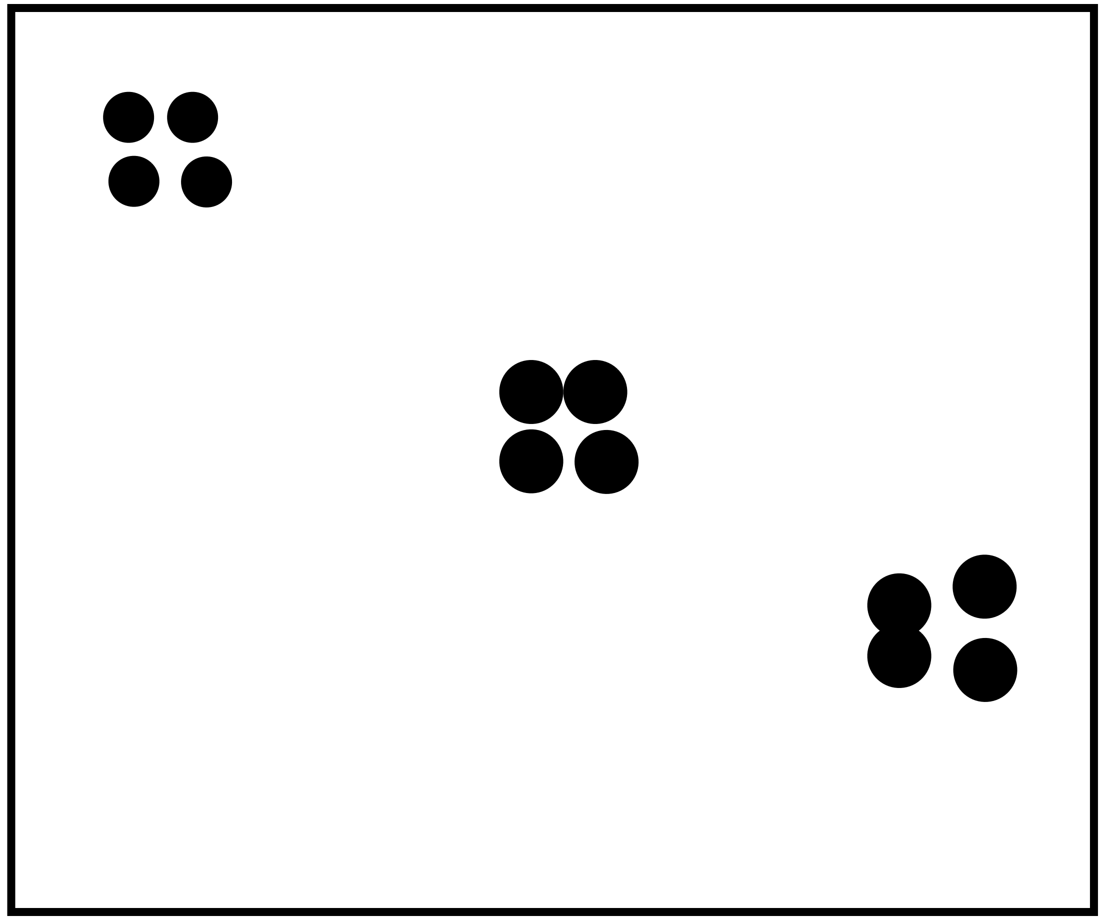
同じ関係　同じ関係 $\xrightarrow{\text{プール}}$ 関係なし

関係なし　関係なし $\xrightarrow{\text{プール}}$ 関係あり

グループ内とグループ間



グループ内で負、プールすると正



グループ内で関係なし、プールすると負

層stratum　ブロック

層の中で要因（説明変数）の効果を見る

ロジスティック回帰

対数線形モデル（ポアソン回帰）

Mantel-Haenszel検定

共通オッズ比

Cochran-Mantel-Haenszel検定

層（ブロック）を認識しそこなうと、結論が大きく変わる可能性がある