データ解析実習（４回目）

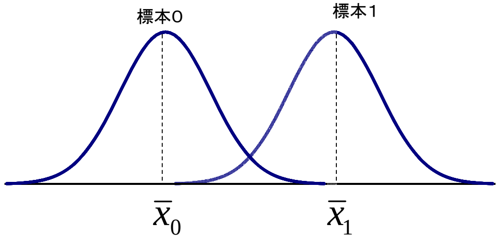
2011.10.31　月曜、４・５限

担当教員：高木英至、田端章明

■本日の授業

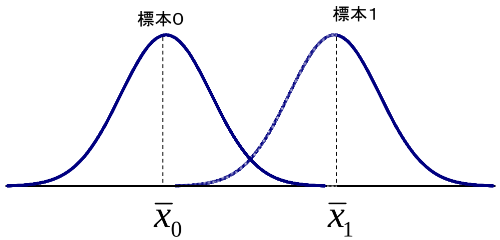
- ４限：分散分析
  - もう１つの「平均値の差の検定」
  - 考え方
    - 平方和の分割、など
    - 一元配置の分散分析表
    - 多重比較検定
    - 多元配置
    - 主効果と交互作用効果
- ５限：クロス表分析
  - ＳＰＳＳをいじる

集団間の平均値の差の検定



- 帰無仮説(H_0)：母集団では平均値に差がない
- 考え方１：標本の平均値の分布（正規分布、ｔ分布）を利用する考え方　→　ｔ検定

集団間の平均値の差の検定



- 考え方２：集団による値のバラツキが有意であるかどうか、を検定する
  - 帰無仮説(H_0)：母集団では平均値に差がない
  - 分散分析（Ｆ検定）

平方和の分割

全体の平方和は

$$SS_{total} = \sum_{j=1}^{a}\sum_{i=1}^{n_j}(y_{ij}-\bar{y})^2 \qquad (9\cdot1) \qquad p.273$$

群間の平方和は

$$SS_A = \sum_{j=1}^{a} n_j(\bar{y}_j-\bar{y})^2 \qquad (9\cdot2)$$

- 前提
  - 集団数 ＝ a
  - 各集団の標本数：n_j

群内の平方和(残差の平方和)は

$$SS_e = \sum_{i=1}^{n_1}(y_{i1}-\bar{y}_1)^2 + \sum_{i=1}^{n_2}(y_{i2}-\bar{y}2)^2 + \ldots$$

$$= \sum_{j=1}^{a} n_j s_j{}^2 \qquad (9\cdot3)$$

$$SS_{total} = SS_A + SS_e$$

平方和の分割

- 全体の平方和に占める群間の平方和の割合が大きいほど、全体の値のバラツキは集団差によって説明できる
- 偶然からでも、標本ではある程度の群間の平方和は生じる
- 群間の平方和が有意に（偶然から生じ得る以上に）大きければ、集団間の平均値の差は有意だと考える

$$\eta = \sqrt{\frac{SS_A}{SS_{total}}} \quad ： \quad 相関比$$

分散分析表

平方和／自由度

グループ間平均平方／グループ内平均平方

要因	平方和	自由度	平均平方	F 値	有意確率
グループ間	SS_A	$a-1$	$SS_A/(a-1)$	$[SS_A/(a-1)]$ $/[SS_e/(n-a)]$	p
グループ内	SS_e	$n-a$	$SS_e/(n-a)$		
全体	SS_{total}	$n-1$	SS_{total} $/(n-1)$		

- 自由度(degrees of freedom)
  - 平均値が固定されたとき、自由に動ける値の数
  - 自由度が大きければ自動的に平方和は大きくなる
- Ｆ値の検定：　Ｆ検定、Ｆ分布表
  - 帰無仮説が正しければＦ値は小さい

分散分析表

分散分析

平方和／自由度

グループ間平均平方／グループ内平均平方

開放性

	平方和	自由度	平均平方	F 値	有意確率
ｸﾞﾙｰﾌﾟ間	1108.626	4	277.156	2.601	.037
ｸﾞﾙｰﾌﾟ内	20887.593	196	106.569		
合計	21996.219	200			

- 自由度(degrees of freedom)
  - 平均値が固定されたとき、自由に動ける値の数
  - 自由度が大きければ自動的に平方和は大きくなる
- Ｆ値の検定：　Ｆ検定、Ｆ分布表
  - 帰無仮説が正しければＦ値は小さい

多重比較検定

- 多数の平均値があるときの、相互の差の検定
  - t 検定は使えない（p.279-）
- → 多重比較検定を使う
  - S-N-K検定、テューキー検定、ダンカン検定、…

開放性

Student-Newman-Keuls

学部	度数	α = .05 のｻﾌﾞｸﾞﾙｰﾌﾟ 1	2
教養学部	50	46.5000	
工学部	52	49.6731	49.6731
教育学部	30	50.2000	50.2000
理学部	33	50.7273	50.7273
経済学部	36		53.6389
有意確率		.281	.337

等質なｻﾌﾞｸﾞﾙｰﾌﾟのｸﾞﾙｰﾌﾟ平均値が表示されています。

a 調和平均ｻﾝﾌﾟﾙｻｲｽﾞ = 38.272 を使用

b ｸﾞﾙｰﾌﾟ ｻｲｽﾞが等しくありません。ｸﾞﾙｰﾌﾟ ｻｲｽﾞの調和平均が使用されます。ﾀｲﾌﾟ I 誤差水
ん。

多元配置の分散分析

- 要因と水準
  - 水準：要因での条件数
- 一元配置：要因が１つ
- 多元配置：要因が複数
- 例：２元配置の分散分析

前提：各条件（セル）の観測数が等しい（最も単純な場合）

2要因：A, B － 水準数はそれぞれ、a, b

全体の平方和は

$$SS_{total} = \sum_{j=1}^{a}\sum_{i=1}^{n_j}(y_{ij} - \bar{y})^2$$

$$= SS_A + SS_B + SS_{AB} + SS_e \quad (9\text{-}14) \qquad p.288$$

分散分析表

平方和／自由度

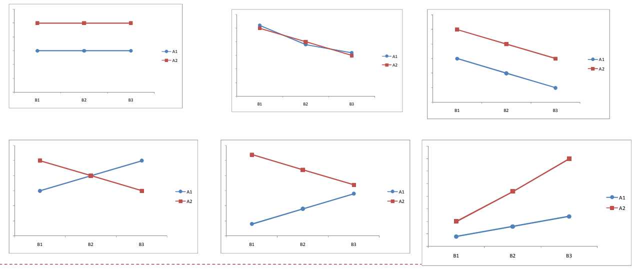
グループ間平均平方／グループ内平均平方

要因	平方和	自由度	平均平方	F 値	有意確率
A	SS_A	$a-1$	MS_A	MS_A/MS_e	p_A
B	SS_B	$b-1$	MS_B	MS_B/MS_e	p_B
AxB	SS_{AB}	$(a-1)(b-1)$	MS_{AB}	MS_{AB}/MS_e	p_{AB}
誤差	SS_e	$n-ab$	MS_e		
全体	SS_{total}	$n-1$			

- 主効果(main effect):A,B
- 交互作用効果(interaction effect):AxB

グラフによる例

2要因：A、B
A：2水準
B：3水準
条件内でバラツキは小さいと仮定



4限は
おしまい