



データ解析実習

2012.10.29 月曜、4・5限

担当教員：高木英至

1

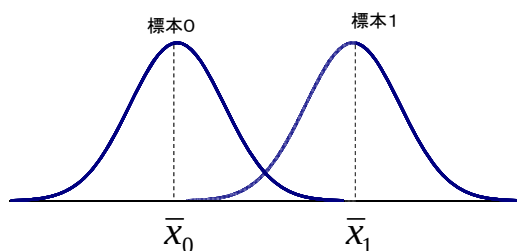
■ 本日の授業

- ▶ 基礎概念導入
- ▶ 前半(4限)
 - ▶ 一元配置と二元配置の分散分析
 - ▶ 交互作用効果
- ▶ 後半(5限)SPSS解説



▶ 2

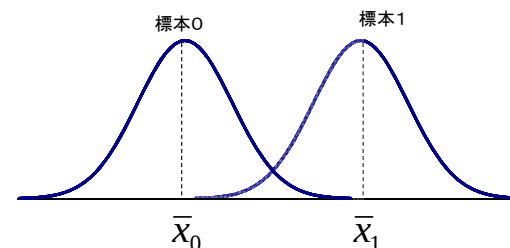
集団間の平均値の差の検定 (t 検定)



- ▶ 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

▶ 3

集団間の平均値の差の検定 (F 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)

▶ 4

平方和の分割

全体の平方和は

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

群間の平方和は

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

群内の平方和(残差の平方和)は

$$\begin{aligned} SS_e &= \sum_{i=1}^{n_1} (y_{i1} - \bar{y}_1)^2 + \sum_{i=1}^{n_2} (y_{i2} - \bar{y}_2)^2 + \dots \\ &= \sum_{j=1}^a n_j s_j^2 \quad (9\cdot3) \\ SS_{total} &= SS_A + SS_e \end{aligned}$$

▶ 5

▶ 前提

- ▶ 集団数 = a
- ▶ 各集団の標本数 : n_j

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

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

▶ 6

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、自由に動ける個の数
 - ▶ 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

▶ 7

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
グループ間	1108.626	4	277.156	2.601	.037
グループ内	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、自由に動ける個の数
 - ▶ 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

▶ 8

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、...

▶ 9

開放性

Student-Newman-Keuls

学部	度数	$\alpha = .05$ のサブグループ	
		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

等質なサブグループのグループ平均値が表示されています。

a 調和平均サンプルサイズ = 38.272 を使用

b グループ サイズが等しくありません。グループ サイズの調和平均が使用されます。タイプ I 誤差が α 。

▶ 10

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

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

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

▶ 11

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

▶ 12

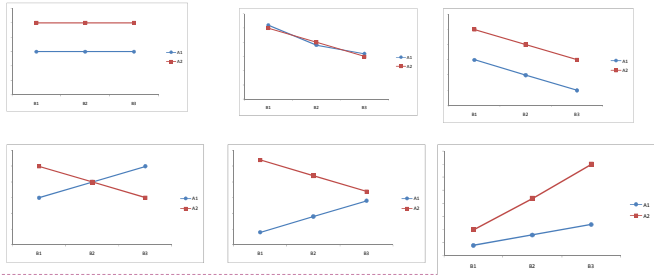
グラフによる例

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



▶ 13



4限は おしまい

▶ 14