

要因と主効果

要因 (Factor)

- 研究参加者間要因 (対応のない要因) (Between Subject Factor)

異なる集団において、同一変数を測定するときの、**集団の違い**
入院患者のうつ傾向得点に対する 年代 (40代, 50代, 60代)
の影響

各研究参加者はいずれか1つの年代群に所属する

- 研究参加者内要因 (対応のある要因) (Within Subject Factor)

1つの集団において、同一変数を、異なる条件下で測定するときの、**条件の違い**

生徒の充実感得点に対する 授業方法 (TT, 習熟度別, 協同学習) の違いの影響

各生徒はすべての授業方法を経験する

主効果(Main Effect)

- ある要因において、群（集団，条件）の違いにより平均値に差異が認められるとき，その要因の**主効果**があると言う
- 主効果の例
入院患者のうつ傾向得点の平均値について，年代間で差がある
生徒の充実感得点の平均値について，授業方法間で差がある
- 群の違いにより，母平均に差があると言えるかどうかを判断する分析方法として**分散分析 (Analysis of Variance: ANOVA)** がある
- 2群の平均値の比較を分散分析で行っても間違いではないが，通常は t 検定を用いる。結果は同一となる

多群の平均値の比較の考え方

分析例：1 研究参加者間要因

- 入院患者のうつ傾向得点に対する年代(40代, 50代, 60代)の影響を知りたい
- 年代群の違いによって、うつ傾向得点の平均値に違いがあると言えるか？

年齢群	群内番号	うつ得点
40	1	29
40	2	32
⋮	⋮	⋮
40	73	24
50	1	24
50	2	28
⋮	⋮	⋮
50	81	36
60	1	23
60	2	36
⋮	⋮	⋮
60	84	20

group	N	M	SD
40代	73	26.56	5.71
50代	81	29.26	7.09
60代	84	27.51	6.69
全体	238	27.82	6.62

基本的な考え方

群間で母平均に差がある場合

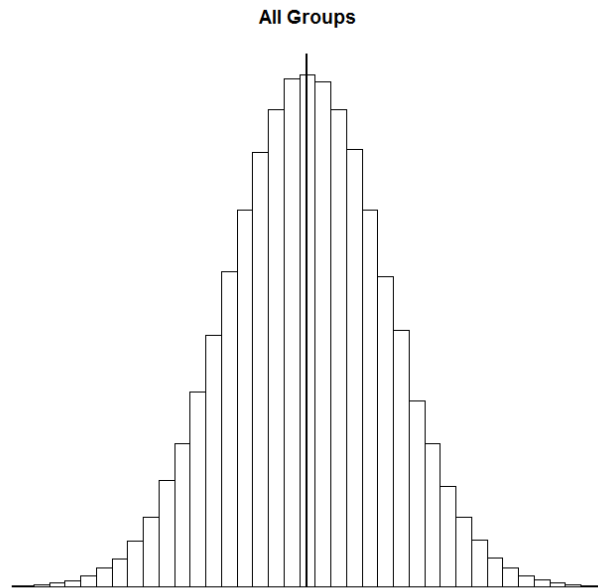
- 各群のデータ分布は，それぞれの群の標本平均の周りに集まり，各群の分布の位置にズレが見られるだろう

群間で母平均に差がない場合

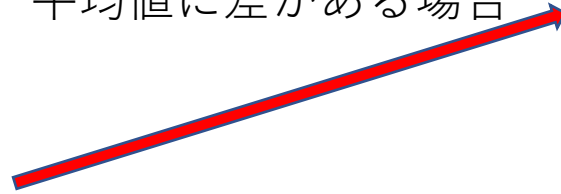
- 各群のデータ分布は，それぞれの群の標本平均の周りになだらかに集まり，各群の分布の位置はだいたい同じだろう

基本的な考え方のイメージ

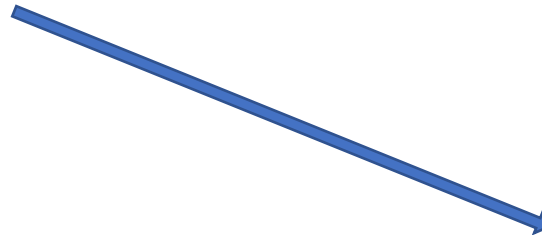
全体の分布



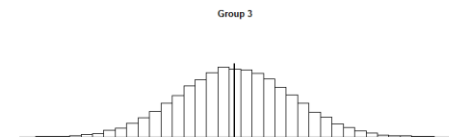
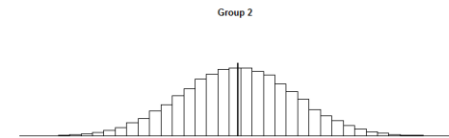
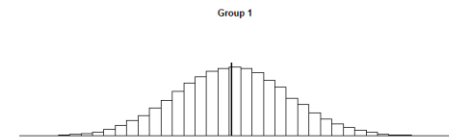
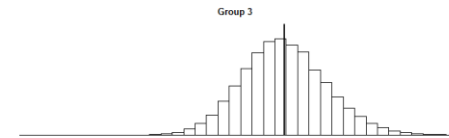
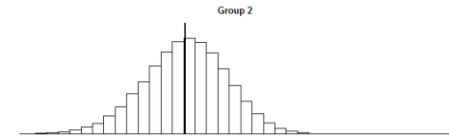
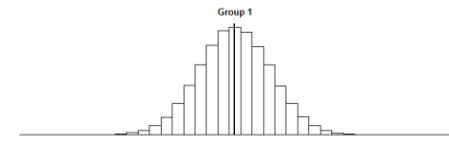
平均値に差がある場合



平均値に差がない場合



群別の分布



データ・群平均・全体平均

- 群 j に所属する被験者 i のデータ : x_{ij}
29, 32, ..., 24, 28, ..., 23, 36, ..., 20
- 群 j の平均 : $\bar{x}_{.j}$
26.56, 29.26, 27.51
- 全体平均 : $\bar{x}_{..} = 27.82$
- 少しトリッキーな式を考える
$$x_{ij} - \bar{x}_{..} = x_{ij} - \bar{x}_{.j} + \bar{x}_{.j} - \bar{x}_{..}$$
- 各辺を2乗したものを i, j について合計する
$$\sum (x_{ij} - \bar{x}_{..})^2 = \sum (x_{ij} - \bar{x}_{.j} + \bar{x}_{.j} - \bar{x}_{..})^2$$

平方和 (Sum of Squares: SS)

- 右辺を整理すると次のようになる

$$\sum (x_{ij} - \bar{x}_{..})^2 = \sum (\bar{x}_{.j} - \bar{x}_{..})^2 + \sum (x_{ij} - \bar{x}_{.j})^2$$

$a^2 = b^2 + c^2$ のような式 (三平方の定理)

$\sum (x_{ij} - \bar{x}_{..})^2$ 全体のデータの散らばり (SST: Total Sum of Squares)

$\sum (\bar{x}_{.j} - \bar{x}_{..})^2$ 各群の平均値の散らばり (SSA: Sum of Squares of Factor A)

$\sum (x_{ij} - \bar{x}_{.j})^2$ 群内のデータの散らばり (SSR: Residual Sum of Squares)

- $$\text{SST} = \text{SSA} + \text{SSR}$$

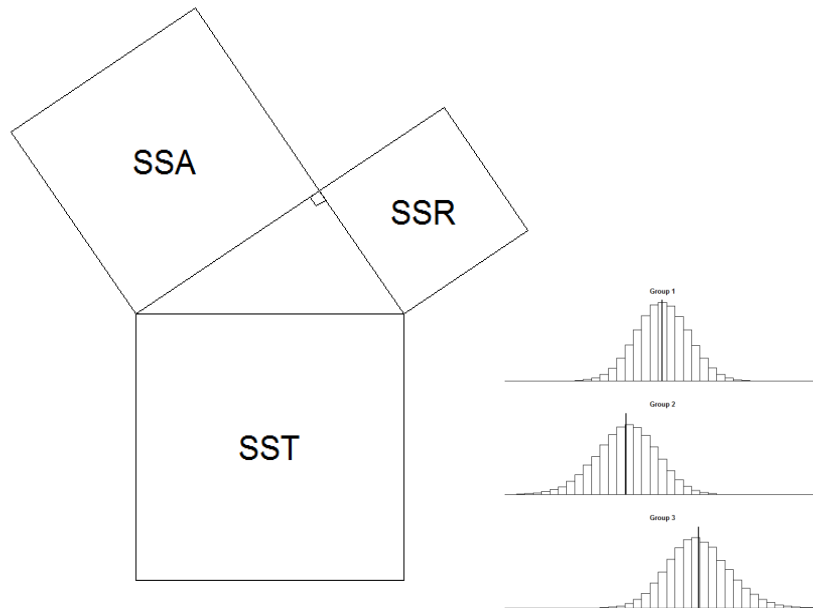
全体平方和 = 群間平方和 + 残差平方和

平方和の分割

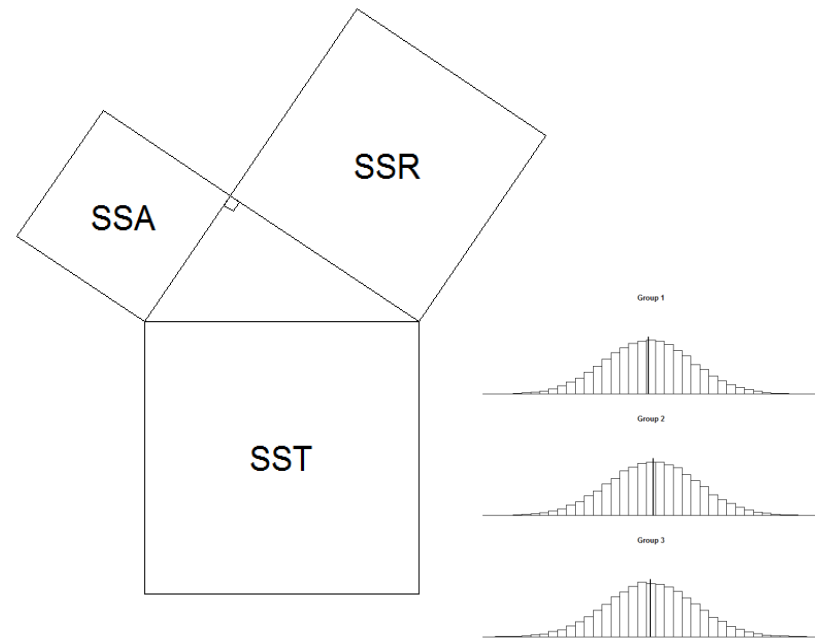
- 全体のデータの散らばりは
全体平均に対する各群の平均値の散らばりと
群内における群平均からのデータの散らばりに分割
できる
- 群間の平均値の差が大きければ
群間平方和は大きく
残差平方和は小さい
- 群間の平均値の差が小さければ
群間平方和は小さく
残差平方和は大きい

平方和の分割のイメージ

各群の平均値が散らばっている場合



各群の平均値が似通っている場合



直角は、統計学では「無相関」を表す
データ全体の散らばりは、各群の平均値の散らばりと、
各群内のデータの散らばりに分けて考えることができる

F 統計量

- $F = \frac{\text{群間平方和} / \text{自由度}_1}{\text{残差平方和} / \text{自由度}_2} = \frac{SSA / (a-1)}{SSR / (N-a)}$

a : 群数

N : 全体の被験者数

- 自由度1 : 群の自由度 $a-1$
第1～第 $a-1$ 群でなければ第 a 群と分かる
- 自由度2 : 各群におけるデータの自由度の和 $N-a$
 $(n_1 - 1) + (n_2 - 1) + \cdots + (n_a - 1)$

F 統計量

- F の式に $\frac{SSA}{SSR}$ という構造があるのがポイント
- 群間の平均値の差が大きい $\rightarrow$ SSA 大きい $\rightarrow$ F 値が大きい
- 群間の平均値の差が小さい $\rightarrow$ SSA 小さい $\rightarrow$ F 値が小さい

仮説の設定

- 帰無仮説「 $H_0: \mu_1 = \mu_2 = \mu_3$ 」

母集団において、各年代の入院患者のうつ得点の母平均はすべて等しい

- 対立仮説「 $H_1: \text{Not } H_0$ 」

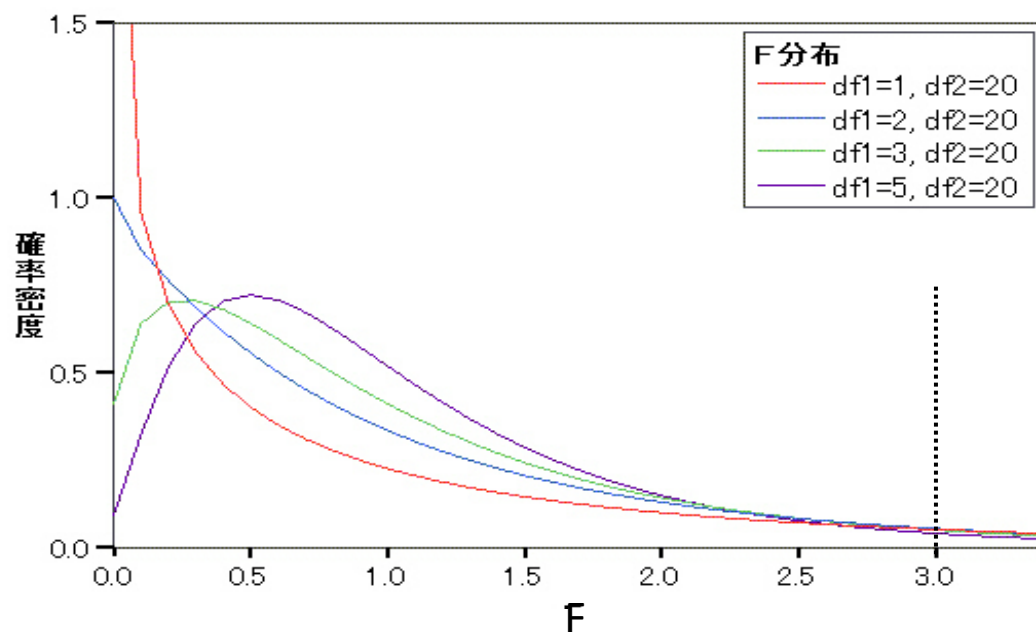
母集団において、各年代の入院患者のうつ得点の母平均はすべて等しい訳ではない

少なくともどれか1つの群の平均値は他の群と異なる（全部がバラバラとまでは求めている）

- F 値が限界値を超えて大きいとき、群間の平均値に差があると判断する

分散分析の F 検定は両側検定か片側検定か？

- F 値は0以上の値になる
- F 値が 0 = 分子が 0 = 群間平方和が 0 = 群間の平均値が等しい
- F 値が大 = 分子が大 = 群間平方和が大 = 群間の平均値に差がある
- 分散分析の F 検定は片側検定



なぜ「分散」分析？

- 多群の平均値を比較する検定は分散分析
- 各群の平均値の散らばりを検討するにあたって
全体のデータの散らばり（分散）を
各群の平均値の散らばり（分散）と
各群内におけるデータの散らばり（分散）に
分割して評価しているので
「分散」分析