

対応のある/ない分散分析の Excel での実行

樋口さぶろお <http://hig3.net>

龍谷大学理工学部数理情報学科

生活の中の統計技術 L10(2018-12-10 Mon)

最終更新: Time-stamp: "2018-12-10 Mon 13:00 JST hig"

今日の目標

- 多群のデータを Excel の箱ひげ図で可視化して比較できる
- 対応のない/ある 1 要因, 対応のない 2 要因の分散分析を Excel で実行できる



L09-Q2

Quiz 解答:分散分析

水準の数 $\ell = 3$, 繰り返しの数 $r = 5$.

$$\bar{y}_{1\bullet} = 10, \bar{y}_{2\bullet} = 15, \bar{y}_{3\bullet} = 23, \bar{y}_{\bullet\bullet} = 16.$$

級間平方和 (級間変動)

$$S_A = \sum_j \sum_i (\bar{y}_{j\bullet} - \bar{y}_{\bullet\bullet})^2 = 5 \times \sum_j (\bar{y}_{j\bullet} - \bar{y}_{\bullet\bullet})^2 = 430.$$

$$\text{残差平方和 (誤差変動)} S_E = \sum_i \sum_j (y_{ji} - \bar{y}_{j\bullet})^2 = 106.$$

$$\text{全平方和 (全変動)} S_T = \sum_j \sum_i (y_{ji} - \bar{y}_{\bullet\bullet})^2 = 430 + 106 = 536.$$

分散分析表は次の通り.

要因	平方和	自由度	平均平方	F_0
級間 A	430	$3 - 1 = 2$	$430/2 = 215$	$215/8.833 = 24.34$
残差 E	106	$14 - 2 = 12$	$106/12 = 8.833$	
全 T	536	$15 - 1 = 14$		

$24.34 > F_{0.05}(2, 12) = 3.885$ より, 全水準の母平均値が等しいという帰無仮説は棄却される.

対応のない 1 要因の分散分析

いろいろやった後、最終的に分散の大きさを比較するために、F 検定という統計的仮説検定を行う手続き。

与えられたデータ 各組で異なる教え方をした。ランダムに何人かよんできてテストをした。

j	学級	点数
1	A 組	78 79 79 80
2	B 組	78 86 81 83 82
3	C 組	86 85 87

問 母集団の点数の平均値は、学級によって異なるか？

$$\text{成績} = \text{平均値} + \text{教え方} + \text{誤差}$$

$$\text{全平方和} = \text{級間平方和} + \text{級内平方和}$$

級間と級内の大小の比較をするのが「対応のない」1 要因の分散分析

対応のある 1 要因の分散分析

十分にそろばんの訓練を積んだ 4 人の生徒に、昇級試験の模擬試験を 3 回受けさせた。問題は同じだが、3 回とも違うタイプのそろばんを使った。

i	そろばんの種類	生徒 1	2	3	4	個数	級内平均	不偏標準
1	そろばん A	78	79	79	80	4	79	
2	そろばん B	78	81	83	86	5	82	
3	そろばん C	85	86	86	87	3	86	
計						12	82	

問 1 そろばんの違いは点数に影響するか？

生徒のスキルによる差は、級内平均和に入れず、ブロック間平方和として分離する処理。

$$\text{成績} = \text{平均値} + \text{そろばん} + \text{生徒のスキル} + \text{誤差}$$

$$\text{全平方和} = \text{級間平方和} + \text{ブロック間平方和} + \text{級内平方和}$$

級間と級内の大小の比較をするのが「対応のある」1 要因の分散分析「対応のない」より試験紙の感度が高い。

対応のない 2 要因の分散分析

そろばんの検定で、そろばんのタイプ (ハードウェア) 3 種, 読み上げ音声 2 種, で $3 \times 2 \times 4 = 24$ 名の同じ実力の生徒に試させた。

	音声 1	音声 2
そろばん A	78 78 80 80	79 92 90 79
そろばん B	78 82 84 78	81 85 89 85
そろばん C	85 84 84 85	80 83 82 83

問 そろばんのタイプの違いは点数に影響するか? 主効果 1

問 音声の違いは点数に影響するか? 主効果 2

問 このそろばんとこの音声を組み合わせると、とくに効果を発揮するよ
うなことがあるか? '食べ合わせ' はあるか? 交互作用

3 種類の試験紙の組.

成績 = 平均値 + そろばん + 音声 + 交互作用 + 誤差

全平方和 = 級間平方和 1 + 級間平方和 2 + 交互作用平方和 + 級内平方和

ここまで来たよ

8 略解:分散分析

9 対応のある/ない分散分析の Excel での実行

- Excel で分散分析

準備:統計ツールを有効化

動画

ファイル > オプション > アドイン > Excel のアドイン > 設定 > 分析ツール に
チェックを入れて OK する.

用語

Excel の (翻訳された) 用語は, 統計学の標準的な用語でないことがあるので
ご注意ください.

データ方向: 横に並んでたら **行**, 縦に並んでたら **列**

ラベル: 行 (や列) の先頭にある, 級の名前を文字列で入力したセル
表記 $3.0E-15 = 3.0 \times 10^{-15}$.

可視化

Excel で一方の級で **層別** して簡単に可視化するには,

挿入 > グラフ > 箱ひげ図

Excel の入出力例:対応のない 1 要因の分散分析

データ > データ分析 > 分散分析: 一元配置

分散分析: 一元配置						
概要						
グループ	標本数	合計	平均	分散		
A組	4	316	79	0.66666667		
B組	4	328	82	11.33333333		
C組	4	344	86	0.66666667		
分散分析表						
変動要因	変動	自由度	分散	測された分散	P-値	F 境界値
グループ間	98.6666667	2	49.33333333	11.6842105	0.0031517	4.25649473
グループ内	38	9	4.22222222			
合計	136.666667	11				

A組	78	79	79	80
B組	78	86	81	83
C組	86	85	87	86

「グループ間 (級間)P-値 < あらかじめ設定した有意水準」なら試験紙変色。「平均値は級による」

Excel の入出力例:対応のある 1 要因の分散分析

データ > データ分析 > 分散分析: くり返しのない二元配置

分散分析: 繰り返しのない二元配置						
概要	標本数	合計	平均	分散		
そろばんA	4	316	79	0.66666667		
そろばんB	4	328	82	11.33333333		
そろばんC	4	344	86	0.66666667		
生徒1	3	242	80.66666667	21.33333333		
生徒2	3	250	83.33333333	14.33333333		
生徒3	3	247	82.33333333	17.33333333		
生徒4	3	249	83	9		
分散分析表						
変動要因	変動	自由度	分散	測された分散	P-値	F 境界値
行	98.6666667	2	49.33333333	11.6842105	0.00852731	5.14325285
列	12.6666667	3	4.22222222	1	0.45472475	4.75706266
誤差	25.33333333	6	4.22222222			
合計	136.666667	11				

	生徒1	生徒2	生徒3	生徒4
そろばんA	78	79	79	80
そろばんB	78	86	81	83
そろばんC	86	85	87	86

「行 (級間)P-値 < あらかじめ設定した有意水準」なら試験紙変色. 「平均値は級 (そろばん) による」

Excel の入出力例:対応のない 2 要因の分散分析

データ > データ分析 > 分散分析: くり返しのある二元配置

1 標本あたりの行数 は, データの繰り返し (主要因 1, 主要因 2 が同一であるデータ, 同一条件の重複するデータ) の個数を指定する (この例なら 4). 縦に並べる.

データの選択範囲は, 必ずラベルを含む必要.

	そろばんA	そろばんB	そろばんC
音声1	78	78	85
音声1	78	82	84
音声1	80	84	84
音声1	80	78	85
音声2	79	81	80
音声2	92	85	83
音声2	90	89	82
音声2	79	85	83

結果 (次ページ) の解読方法

「標本 (音声)P-値 < あらかじめ設定した有意水準」なら試験紙 1 変色.
平均値は音声による.

「列 (そろばん)P-値 < あらかじめ設定した有意水準」なら試験紙 2 変色.
平均値はそろばんによる.

分散分析: 繰り返しのある二元配置						
概要	そろばんA	そろばんB	そろばんC	合計		
音声1						
標本数	4	4	4	12		
合計	316	322	338	976		
平均	79	80.5	84.5	81.3333333		
分散	1.33333333	9	0.33333333	8.78787879		
音声2						
標本数	4	4	4	12		
合計	340	340	328	1008		
平均	85	85	82	84		
分散	48.6666667	10.6666667	2	18.9090909		
合計						
標本数	8	8	8			
合計	656	662	666			
平均	82	82.75	83.25			
分散	31.7142857	14.2142857	2.78571429			
分散分析表						
変動要因	変動	自由度	分散	測された分散	P-値	F 境界値
標本	42.6666667	1	42.6666667	3.55555556	0.0755919	4.41387342
列	6.33333333	2	3.16666667	0.26388889	0.770979	3.55455715
交互作用	82.3333333	2	41.1666667	3.43055556	0.05467235	3.55455715
繰り返し誤差	216	18	12			
合計	347.333333	23				

お知らせ

- 来週 2018-12-17 月 2 は (たぶん)5-203 で
- 図書館ミニ講義「確率を学ぶ～年末ジャンボ宝くじが当たる確率は!?～」by 樋口
 - ▶ 2018-12-20 木 12:45-13:15
 - ▶ 生協コンビニ地下スチューデント commons (瀬田) ミーティングスペース
- レポート 1(長くない)
 - ▶ Manaba で振り返りの作文的なもの <https://manaba.ryukoku.ac.jp>
 - ▶ 2018-12-17 月夜 まで
- 期末試験計画
 - ▶ 30 ピーナッツ/科目 100 ピーナッツ
 - ▶ 60 分
 - ▶ 2019-01-28 月
- レポート計画