

2. R による統計分析例

2.1 分散分析

分散分析とは

- ・「分散」(バラツキ)を分析する
- ・データの「バラツキ」が、何らかの理由(要因)によるものなのか、偶然の誤差なのかを判定する。
- ・「要因」の数により、「一要因」から「三要因」まである。
- ・集めたデータが、一人の被験者からとった複数のデータ(「被験者内」)なのか、複数の被験者からとったデータ(「被験者間」)なのかを区別する。
- ・要因が複数ある場合は、「混合」(被験者内と被験者間)の場合もある。

分散分析の流れ

1. 分散分析をする
2. 交互作用があるか
 1. ない場合 => 主効果が有意か
 2. ある場合 => 交互作用の分析
 1. 主効果は意味がなくなる
 2. 単純主効果が有意か
3. 3水準(条件)以上の有意な主効果もしくは交互作用があるか
 1. ない場合 => End
 - 1.2 水準で有意な場合、差があるのはどこなのかは明白なので、多重比較する必要はない。
 2. ある場合 => 多重比較(二つずつ条件を組み合わせ、実際に有意な差があるのはどこを見つける)
 1. 主効果が有意であった場合 => 主効果の多重比較
 2. 交互作用が有意であった場合 => 単純主効果の多重比較

2.2 R で分散分析は ANOVA 君できまり

- ・ js-STAR と同じ使い勝手・説明
- ・ 田中・山際(1992)『ユーザーのための教育・心理統計と実験計画法 方法の理解から論文の書き方まで』で勉強できる。
- ・ 中野・田中(2012)『フリーソフト js-STAR でかんたん統計データ分析』

anova 君のインストール(Rで書かれたプログラムを読み込んで使う)

1. anovakun というソースコードをダウンロードする
2. R のメニューからソースコードを読み込む

```
source("http://riseki.php.xdomain.jp/index.php?plugin=attach&refer=ANOVA%E5%B&openfile=anovakun_471.txt") %90 %9
```

分散分析のタイプと分析の手順

分散分析のタイプ一覧

```
A s ( 1 要因参加者間 )
s A ( 1 要因参加者内 )
A B s ( 2 要因参加者間 )
A s B ( 2 要因混合 )
s A B ( 2 要因参加者内 )
A B C s ( 3 要因参加者間 )
```

A B s C (3 要因混合)
 A s B C (3 要因混合)
 s A B C (3 要因参加者内)

1. 分析のタイプを決める。

1. 要因はいくつあるか
2. それぞれの要因は「被験者内」か「被験者間」か（「対応のある」データか「対応のない」データか）
3. それぞれの要因に「水準」はいくつあるか
4. 例
 1. トレーニングをする前と後で成績が上がるか（一要因・被験者内・二水準）
 2. トレーニングをする前と後、そして、しばらくしてからも効果が残るか（一要因・被験者内・三水準）
 3. 文学部、工学部、理学部で、英語の成績に差があるか（一要因・被験者間・三水準）
 4. 漢字の画数が多い熟語と少ない熟語で読む時間に差があるか（一要因・被験者内・二水準）
 5. 漢字の画数が多い熟語と少ない熟語で、実在する語と実在しない語とで、読む時間に差があるか（二要因・被験者内・ 2×2 ）
 6. 漢字の画数が多い熟語と少ない熟語を読むときに、日本語母語話者と学習者とで差があるか（二要因・混合・ 2×2 ）
 7. 漢字の画数が多い熟語と少ない熟語で、実在する語と実在しない語とで、日本語母語話者と学習者とで読む時間に差があるか（三要因・混合・ $2 \times 2 \times 2$ ）

2。

3. データを読み込む。

4. 分析のタイプに合わせて、コマンドを実行する。

ANOVA 君の使い方

3 . データファイルの作成

・ポイント

- ・被験者内は、横に並べる
- ・被験者間は、縦に並べる

・js-STAR の形式を参考に

```
Error in ci.calc(dat = dat, design = design, factnames = factnames, cilmd = cilmd, :
could not find function "anyNA"
```

古い R(3.0.1) で古い anovakun(4.3.3) だと大丈夫なんだけど、、

この一行をコピーして anyNA() が使えるようにしてください。（Thanks to 阿部君）

```
anyNA = function(x) {any(is.na(x))}
```

anovakun のコマンド

anovakun(データ , " 分析のタイプ ", 各要因の水準数 , , オプション)

1. 要因を A から Z で
2. 小文字の s は「被験者」の意味

1. 被験者間は s の左側
 - 1.As
2. 被験者内は s の右側
 - 1.sA
3. 被験者間と被験者内の両方の要因があるときは
 1. 被験者間要因の方から先に割り当てる
4. 同じ要因が複数ある場合は
 1. より包括的な要因から順にならべる

- ・ 1 要因被験者間で、水準が 3 つだとすると

anovakun(データ , "As", 3)

- ・ 2 要因混合タイプで、水準が、それぞれ、2 と 3 だとすると

anovakun(データ , "AsB", 2, 3)

- ・ オプションとしては、多重比較など。
 - ・ 「Bonferroni の方法」が有名だが、検出力が低くなるので、その改良版の「Holm の方法」を使うようになっている。

anovakun(データ , "AsB", 2, 3, holm=T)

2.3 分析の実例「三つのクラスで、成績に差があるか？」

対応のサンプルの場合：を例に

- ・ ポイントは、データをどういうフォーマットで並べた表を用意するか。
 - ・ 対応のない三つのクラス間なので、被験者間
 - ・ 縦に並べる
 - ・ 左側にクラスの違いを示す記号を付けておく
 - ・ 要因は一つだけ
 - ・ 水準は三つ（三つのクラス）
- ・ データの形

```
a 76
a 85
a 81
a 41
a 73
a 66
a 88
a 78
a 85
a 47
a 71
a 62
a 89
a 84
a 74
a 78
a 78
a 75
a 93
a 57
a 61
```

```

a 79
b 56
b 44
b 46
b 73
b 74
b 58
b 72
b 74
b 85
b 68
b 67
b 68
b 73
b 68
b 83
b 80
b 76
b 81
b 81
b 77
b 53
b 57
c 54
c 67
c 73
c 57
c 74
c 74
c 65
c 82
c 74
c 82
c 66
c 80
c 60
c 78
c 69
c 82
c 71
c 63
c 60
c 75
c 72
c 68

```

- ・データを R に読み込んでおく
 - ・「クリップボード」にコピーしておいて
 - ・読み込むコマンド

```
three <- read.table("clipboard")
```

- ・anovakun の命令

```
anovakun(three, "As", 3)
```

- ・その結果

[As-Type Design]

This output was generated by anovakun 4.7.1 under R version 3.2.1.
It was executed on Tue Jul 21 19:20:51 2015.

<< DESCRIPTIVE STATISTICS >>

	A	n	Mean	S.D.
a1	22	73.6818	13.3502	
a2	22	68.8182	11.7823	

```
a3 22 70.2727 8.1951
```

```
<< ANOVA TABLE >>
```

Source	SS	df	MS	F-ratio	p-value
A	274.2121	2	137.1061	1.0706	0.3490 ns
Error	8068.4091	63	128.0700		
Total	8342.6212	65	128.3480		

+p < .10, *p < .05, **p < .01, ***p < .001

```
output is over -----///
```

2.4 分析の実例「事前テストと事後テスト、遅延テストで差があるか」

対応のサンプルの例（練習のための作例です）

- ・あるクラスで学習実験をしました。
- ・事前テスト => 学習 => 事後テスト => 1週間後に遅延テスト
 - ・同じ人から3種類のデータを取っているので「対応のあるデータ」(被験者内)
- ・学習の効果はあったでしょうか？
- ・また、その効果は維持されたでしょうか？

1. 要因は、タイミングだけなので、1 要因（水準は三つ）
 2. タイプとしては（1 要因参加者内）タイプ。
- ・分散分析の場合、対応のあるデータについては、球面性の検定をする必要がある。
 - ・各水準間の「差」の分散が等しいことを確認するため
 - ・その際に、もし有意だった場合（p 値が 0.05 より小さい場合）球面性が保たれていないことになるが、自動で自由度の補正を行うオプションをつけるようにすればよい

- ・コマンド
 - ・データの読み込み

```
prepostdelay <- read.table("clipboard", head=T)
```

- ・実行

```
anovakun(prepostdelay, "sA", 3, auto=T)
```

- ・グラフで見る

```
boxplot(prepostdelay)
```

2.5 やってみよう

杉浦・岩崎（2003）日本語学習者のための擬音語・擬態語学習用マルチメディア CALL 教材の改善に向けて

- ・20 人の被験者に CALL 教材を使ってもらいました。
 - ・学習してもらう項目「実験項目群」と学習してもらわない項目「統制項目群」の二種

類の項目群からなるテストを受けてもらいました。

- ・テストは、事前・事後・遅延の3種類を受けてもらいました。
- ・教材に含まれていて学習した項目（実験項目群）と教材に含まれていなかった項目（統制項目群）とで、差があるかどうか。

・データ

・分析

1. 実験項目群について、事前・事後・遅延の成績に差があるか？
2. 統制項目群について、事前・事後・遅延の成績に差があるか？
3. 実験項目群と統制項目群とで、事前・事後・遅延の成績に差があるか？

杉浦担当の授業についての課題

上の「分析」をRを使ってやってみること。

その手順と結果と考察をレポートにまとめること。

評価のポイントは、

- 1) 分析手順がわかっているか、
 - 2) 結果のどこを見てどう判断すればよいかわかっているか、
 - 3) 結果について、自分で考察することができるか、
- という点を評価しますので、「わかっている」ということが、わかるように書いてください。
枚数の制限は特に設けませんが、出てきた結果を、どこが必要か、どこは不要か、ということを考えずにただコピペして枚数を増やすというのはよくありません。
必要なことを十分書いてください。

授業の感想も、書いてくださいね。よろしく。

締め切りは、7月30日木曜日午後5時

PDF ファイルにして、杉浦にメールで送ってください。（24 時間以内に受け取ったという返事をもらってください。ない場合は催促してください。）