

Montgomery の *Design and Analysis of Experiments*
(2005) に対する R コンパニオン

Christophe Lalanne

訳 荒木 孝治

平成 23 年 11 月 29 日

はじめに

本稿は、Douglas C. Montgomery の優れた著書 *Design and Analysis of Experiments* (以降、DAE と表記) に対する補助的な参照マニュアルである。DAE は、現在第 6 版であるが¹、実験計画法で利用するさまざまな手法を取り扱っている。古典的な比較実験 (独立または非独立の場合の 2 標本の検定・推定)、 k 個の母平均の場合へのその自然な拡張 (1 元配置分散分析)、ブロック化のさまざまな方法 (乱塊法、ラテン方格、それらの派生方法)、要因配置実験 (特に、 2^k)、一部実施法、回帰モデルへの適合、応答曲面法、ロバストパラメータ設計、変量効果²を持つ標準的な計画の導出 (変量要因、枝分かれ、分割法) 等を扱っている。

このようなコンピュータ向けの文書を作成しようと思った動機は、Agresti の著名な本 *Categorical Data Analysis* に対する Laura Thompson の文書³を読んだことにある。実際、このような取り組みは、統計学者の活動の他の面、つまり、コンピュータを利用することに焦点を当てる優れた活動となる。しかし、splusdiscrete ほどは、網羅的ではない。

本稿は、Montgomery の原著と同じ素材を含むが、部分的に電子化したものでもなければ完全に代替するものでもない。データ分析に対する現代的なコンピュータ利用に主眼を置いたものとなっている。本稿の各章の構成は、次のようになる。まず、Montgomery により提示された主概念の要約を与える。次に、最も重要と思われる (私の主観であるが) ものを R を用いて例示する。Montgomery が用いた例は全て、R を用いて再分析する。しかし、DAE の各章の終わりにある練習問題に対しては解答を与えない。関心のある読者には、目的とする解析を “R 流” で行う方法に関するアドバイスを必要に応じて与えることにする。

R について

Montgomery は、各章で提示したデータセットの分析において、主に商用のソフトウェアを利用している。こうしたソフトウェアは統計解析の非常に良いパッケージではあるが、実験室で働く人たちがそれらを利用するには、一定額の投資が必要となる。そこで、R のようなオープンソースのソフトウェアの利用可能性が、商用のソフトウェア (学生にとってはたいてい利用不可能) に代わる選択肢として浮上する。

R は、プログラミング言語 S とソフトウェア S-PLUS に基づいて開発されてきたが、S-PLUS を完全に書き直したフリーのクローンではない。実際、両者には異なる面があり、R の開発思想に関しては、次のアドレスで参照できる。

R は、CRAN のウェブサイト (<http://www.cran.r-project.org>) から自由にダウンロードことができ、本サイトには多くの文書やチュートリアルもある。Montgomery

¹日本で入手できるものは国際学生バージョンの第 7 版。

²(訳注)原著では、fixed effects

³彼女の本の改訂版は次で取得可：<https://home.comcast.net/~lthompson221/Splusdiscrete2.pdf>

が彼の本で利用しているような商用のソフトウェアに対して、R が良い選択となる理由は、組み込まれた統計ルーチンのソースコードが公開されており、検証することができることにある。これに加えて、特定の分析に適した機能を利用者自身が追加することができ、また、分析プロセスをバッチ処理化することにより一度に実行することができる利点がある。

スクリプト

Montgomery の本で行われた全ての分析は、Linux 上の R 2.4 を利用してスタートし、Mac OS X 上の R, version 2.7 で再分析した。全ての例のソースコードは

<http://www.aliquote.org/articles/stat/dae/www.aliquote.org/articles/stat/dae/>

より取得できる。本稿で利用したデータセットもこのサイトにある。スクリプトは、 $R \geq 2.0$ のどのバージョンでも問題なく動くはずである。問題が生じる場合は、発見したバグに関する詳細な情報とともにその内容を e メールアドレス (<mailto:christophe.lalanne@gmx.net>) に送ってほしい。

本稿を作成するのに Sweave を利用しなかった。その理由は、書き始める際に、利用しなくても快適であったからである。また、シミュレートしたデータやパッケージに強く依存する内容がないため、 $\text{T}_{\text{E}}\text{X}$ の verbatim 環境のみで充分であったからである。第 2 章の一部を除いて、含めるコードは全て静的であり、コンパイルには dvips + ps2pdf のみを利用している。さらに、ソースファイルは“巨大”である。tex のファイルに関心のある読者は、<http://www.aliquote.org/articles/stat/dae/www.aliquote.org/articles/stat/dae/dae.tex> よりダウンロードすることができる。

目次

第1章	はじめに	1
第2章	簡単な比較実験	3
2.1	第2章の要約	3
2.2	標本分布	3
2.3	仮説検定	6
2.4	2標本 t 検定	8
2.5	1つの母平均 μ の基準値との比較	10
2.6	対応のあるサンプルへの適用	10
2.7	ノンパラメトリック法	11
第3章	1因子の実験：分散分析	13
3.1	第3章の要約	13
3.2	母数効果モデルの分析	13
3.3	モデルパラメータの推定	14
3.4	モデルのチェック	17
3.5	処理平均の比較	19
3.6	検出力とサンプル数	21
3.7	ノンパラメトリック分散分析	23
第4章	乱塊法，ラテン方格，および関連する計画	26
4.1	第4章の要約	26
4.2	乱塊法	26
4.3	ラテン方格計画	30
4.4	グレコラテン方格計画	32
4.5	釣り合い型不完備ブロック計画	32
第5章	要因計画入門	38
5.1	第5章の要約	38
5.2	2因子要因計画	38
5.3	一般の要因計画，応答曲線・応答曲面	45
5.4	要因計画におけるブロック化	50

第 6 章	2^k 要因実験	54
6.1	第 6 章の要約	54
6.2	2^2 計画	54
6.3	2^3 計画	58
第 7 章	2^k 要因実験におけるブロック化と交絡化	62
第 8 章	2 水準系直交表	63
第 9 章	3 水準系, 混合系直交表	64
第 10 章	回帰モデルへのあてはめ	65
第 11 章	応答曲面法と設計	66
第 12 章	ロバストパラメータ設計とロバストプロセス研究	67
第 13 章	変量因子の実験	68
第 14 章	枝分かれ法と分割法	69
第 15 章	計画と分析に関する他のトピック	70
Appendix		71

第1章 はじめに

Montgomery の本 *Design and Analysis of Experiments* の第 6 版は、生物医学や生産工学における研究で一般に利用されるさまざまな環境に関連して、観測された結果から意味のある結論を導くための多くの手法が取り扱われている。これは「技」であり、実験計画 (DOE) と呼ばれている。Montgomery の本の構成は、基礎となる統計理論の深い理解と実験の幅広い設定、つまり、釣り合い型不完備ブロック計画、分割法、応答曲面法を取り扱っているという点において他の類書とは異なっている。こうした DOE 手法が一つの統計的フレームワークのもとで取り扱われることはあまりなく、このテキストは、基本的な ANOVA モデルに共通の基礎情報を提供していることは有益である。

出版元である Wiley のウェブサイトには、

Douglas Montgomery は、工程や製品のパフォーマンスを最適化するための実験を計画し、実行し、分析するための最も効率的な武器を読者に与える。彼は、システムの特性と最適化のための情報を得るための実験を統計的に計画する方法、製造工程を改良し、新しい工程・製品を設計し、開発するための方法を提示する。また、製品の設計における代替材料の評価方法、製造における特性、実験を有効に、効率的に行うための方法を読者は知ることができる。

と記されている。

現代の統計ソフトウェアは、ますます強力になっており、コンピュータを集約的に利用する手法（ブートストラップ、ジャックナイフ、並べ替え検定等）を、一晩以上コンピュータから離れていなくても実行することができる。さらに、探索的な多変量解析手法は、複数の変量間の関係をグラフ化する新しい、エキサイティングな手法を生み出した。これらは結果を報告するための手法であるが、観測されたデータに対する統計モデルを検証するための有益な補助的な手法でもある。

オープンソースの統計ソフトウェアである R を用いて本で提示されたデータを分析する。公的なウェブサイトである www.r-project.org には、R やそのパッケージに対する国際的な貢献者が著した情報や文書がある。私見では、R は S 言語の初期の開発から恩恵を受けており、統計解析環境であることとデータセットの取り扱いに柔軟性を持つ。現在では、グラフ化能力と推定エンジンとして、より柔軟な統計フレームワークとして設計されている。

本書を通じて用いた R のパッケージを、アルファベット順に示す。簡単な説明をつけるが、関数の利用法等の詳細についてはオンラインヘルプを参照のこと (`help(package="xx")`)。

パッケージリスト CRAN のウェブサイトでは、2007 年以降、タスクビュー (Task Views) という形でパッケージの一部が整理されている。幸いなことに、タスクビューには Experimental Design (実験計画) がある。本稿を作成しはじめた頃、一部実施法やラテン超方格計画といった複雑な計画を作成することに関する文書はほとんどなく、R を用いて DOE を掘り下げて取り扱うものもなかった。例外として、ブロック化と要因配置実験に言及している [26] と、J. Faraway の文書 *Practical Regression and Anova using R* [15] (CRAN で公開されている文書¹)、Berger & Maurer[4] のコンパニオン文書である G. Vikneswaran の An R companion to “Experimental Design” (<http://cran.r-project.org/doc/contrib/Vikneswaran-ED.companion.pdf>) がある。

`car` は、ANOVA の計画と回帰モデルに関する有益な関数を提供する。

`lattice` は、伝統的な R のグラフィックスを補強し、多変量のグラフ化の機能を持つ。

Trellis Displays に関しては、<http://stat.bell-labs.com/project/trellis/> を参照のこと。

`lme4` は、パッケージ `nlme` の新しい強化されたバージョンで、枝分かれや階層モデルといったデータ構造を扱うことができる

`nlme` では、Pinheiro & Bates [20] が開発した混合モデルを扱うことができる。

`npmc`² では、ノンパラメトリックな多重比較法の 2 つを利用することができる。

さらに学ぶために 参考文献を各章の終わりに記載しているが、一般的な参考文献として [14, 6, 9] (英語)[3, 7, 13, 1] (フランス語) がある。

¹Faraway は (非) 線形モデル, GLM, 混合モデルに関する 2 冊の本を出版している [9, 10].

²(訳注) 2011 年 11 月現在, CRAN から取り除かれている。旧バージョンは, アーカイブ (<http://cran.r-project.org/src/contrib/Archive/npmc/>) より取得可能。

第2章 簡単な比較実験

2.1 第2章の要約

Montgomery は、簡単な比較実験を計画する方法（処理や条件，因子水準）を定義した後，このような計画の分析に関連する基本的な統計概念を簡単に説明する．これには，標本分布の考え方や仮説の設定を含む．2 標本に関連した問題を，特定の分布を仮定する手法やノンパラメトリックな手法も取り扱っている．より複雑な問題に取りかかる前に理解すべきなのは，中心となる概念である t 検定である．実際には，検定統計量の構成，帰無仮説（これは常に処理平均間に差がないということを述べる）の下でのこの統計量の分布に関する仮定，結果から結論を得る方法である．統計的推測に関する入門書を求める全ての科学者にとって必読である．

2.2 標本分布

R では，いくつかの確率分布を扱うことができる．関数名の頭につけられた記号， d , p , q , r は，密度関数（density function），確率の値（probability value），パーセント点（quantile value），乱数（random number）に対応する．例えば，標準正規分布（ $\mathcal{N}(0, 1)$ ¹または Z 分布と記す）からのランダムに抽出された大きさ 10 のサンプルは，次のコマンドにより得ることができる．

```
x <- rnorm(10)
```

R においては，乱数の発生器（RNG：random number generator）を呼ぶたびに，異なる種²が用いられる．そのため，後に同じ値を再現できるようにその値を指定しておくと便利である．これには次のコマンドを利用する．

```
set.seed(891)
```

関数 `set.seed` を用いて，特定の状態に RNG を設定することができる．引数として，1 から 1023 までの整数を与える．乱数の生成は統計理論の一部であり，これらの技術はシミュレーションにおいて広く用いられている．さらに，乱数は，ブートストラップ推定や

¹（訳注）原著では $\mathcal{N}(0; 1)$ と表記されているが，本稿では $\mathcal{N}(0, 1)$ とする．

²乱数の発生器としては，元々，再帰合同式，例えば， $x_{k+1} = a_0 + b \cdot x_k \pmod{c}$ に基づく方法が採用されていた．ここで， a_0 は初期種である．今では，もっと洗練されたアルゴリズムを利用することができる．詳細については，コマンド `?RNGkind` によりヘルプを参照．

モンテカルロシミュレーションのようなコンピュータ集約的なアルゴリズムの基礎である。これらの話題へのエレガントな入門として [12] (R の RNG の利用法に関するヒントについては第 8 章参照) がある。

R を用いてさまざまなグラフを作成することができる。グラフは、様々な目的に対して最も有益な統計ツールである。とりわけ、点グラフとヒストグラムは連続変量を可視化するための有益なツールである。図 2.1 は、次のコマンドを用いて作成することができる。

```
# 結合部の引っ張り強度データ (表 2-1, p. 24)
y1 <- c(16.85, 16.40, 17.21, 16.35, 16.52, 17.04, 16.96, 17.15, 16.59, 16.57)
y2 <- c(16.62, 16.75, 17.37, 17.12, 16.98, 16.87, 17.34, 17.02, 17.08, 17.27)
y <- data.frame(Modified=y1, Unmodified=y2)
y.means <- as.numeric(apply(y, 2, mean))
opar <- par(mfrow=c(2, 1), mar=c(5, 7, 4, 2), las=1)
stripchart(y, xlab=expression("Strength (kgf/cm^2)"), pch=19)
arrows(y.means, rep(1.5, 2), y.means, c(1.1, 1.9), length=.1)
text(y.means, c(1.2, 1.8), round(y.means, 2), pos=4, cex=.8)
# 乱数 (本で利用されている金属回収データの代わりに)
rd <- rnorm(200, mean=70, sd=5)
hist(rd, xlab="パーセント点", ylab="相対頻度", main="正規乱数\n N(70,5)")
par(opar)
```

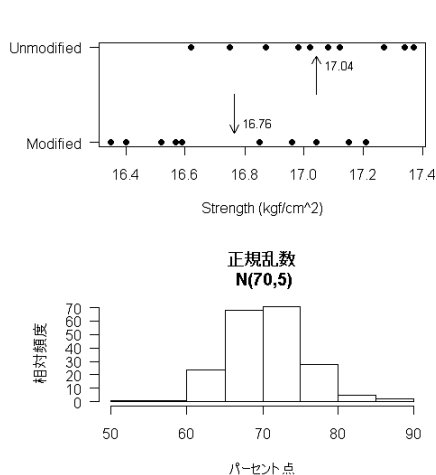


Figure 2.1: 引っ張り強度の点グラフ (上図) と 200 個の正規乱数のヒストグラム (下図)

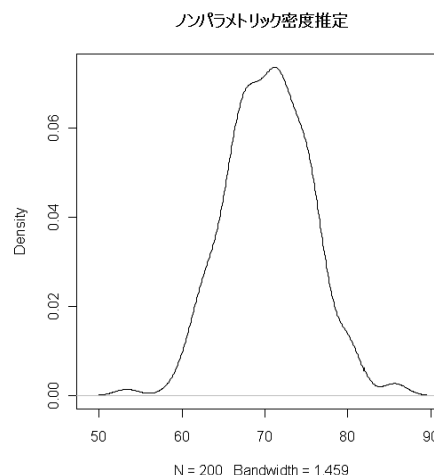


Figure 2.2: ヒストグラムで用いたのと同じ 200 個の正規乱数の密度推定

このコードを見てわかるように、R では、比較的少ないコマンドを用いて強力なグラフィックスを作成することができる。実際、本やウェブサイトで、探索的多変量グラフィックスを取り扱っているものがある。関心のある読者は、次を参照。

- S. Deepayan (2008). *Lattice. Multivariate Data Visualization with R*³. Springer.
<http://www.springer.com/statistics/computational/book/978-0-387-75968-5>
- R Graph Gallery, <http://addictedtor.free.fr/graphiques/>
- Trellis Display, <http://stat.bell-labs.com/project/trellis/>
- P. Murrel (2005). *R Graphics*⁴. Chapman & Hall/CRC.

もちろん、必携の一冊は Venables & Ripley による S 言語に関する本 [26] であり、これは第 4 版になっている。

ヒストグラムは、データ数が多いとき（例えば $n > 20$ ）、その利用はより適切となる。ヒストグラムを描き、そこに、データから推定した平均と分散を持つ正規分布の密度関数を重ね合わせることがよく行われる。しかし、探求している分布を強調するためのより良い方法は、特にデータが連続的な場合、図 2.2 のようなノンパラメトリックな密度関数を描くことである⁵。母集団分布の明確な姿を得ることができるが、ヒストグラムを作成するための区間の数を適切に決定することは容易ではない。これに関する付加的な議論に関しては、[26] (pp. 126–130) を参照のこと。

他の方法として、箱ひげ図を作成することがある。図 2.3 に示したように、多くの情報を箱ひげ図より取り出すことができる。第 1 に、長方形の箱は全体のデータの半数を表し、メディアンは箱の中に線分として表示されている。だから、箱の上部は 75%点であり、25%点は箱の下部である。ひげの端の線は最小値と最大値を表すが、メディアンからの $\pm 1.5 \times$ 四分位範囲外にデータがある場合は、それを点で表示する。このような外れ値の処理方法は Tukey の方法として知られている。箱ひげ図を作成するには関数 `boxplot()` (引数として式、または、+ 因子を与える) を利用する。図 2.3 は、次のコマンドで作成することができる。

```
boxplot(y,ylab="強度 (kgf/cm^2)", las=1)
```

平均 0、分散 1 の Laplace-Gauss 分布—正規分布のこと—の例を図 2.4 に示す。これは密度関数なので、領域の面積は 1 であり、現代のコンピュータソフトウェアを用いると、 x の 2 つの値の間の確率を簡単に求めることができる。例えば、グレーの網掛け領域の面積、つまり確率 $P(1.2 \leq y < 2.4)$ は、0.107 となる。R では、`pnorm(2.4)-pnorm(1.2)` として求めることができる。

```
x <- seq(-3.5, 3.5, by=0.01)
y <- dnorm(x)
plot(x,y,xlab="", ylab="", type="l", axes=F, lwd=2)
```

³R コードと図は次のサイトにある。<http://dsarkar.fhcrc.org/lattice/book/figures.html>.

⁴R コードと図は次のサイトにある。<http://www.stat.auckland.ac.nz/~paul/RGraphics/rgraphics.html>.

⁵(訳注) 次のコマンドにより作成可能。

```
plot(density(rd, bw=1.459), main="ノンパラメトリック密度推定")
```

John Tukey (1915–2000) は、時系列データのスペクトル推定の現代的な方法、高速フーリエ変換 (FFF) を提案した。

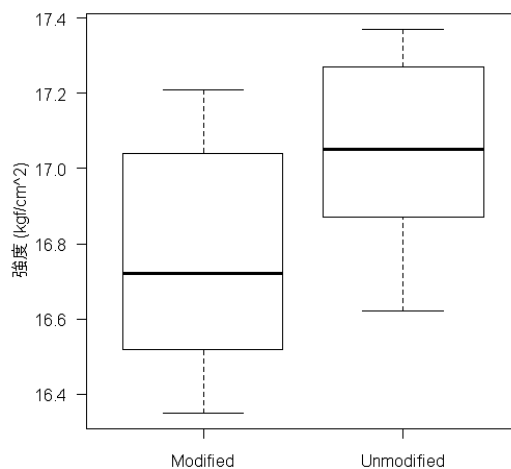


Figure 2.3: ポートランドセメント引っ張り強度実験データの箱ひげ図 .

```
axis(side=1, cex.axis=.8); axis(2, pos=0, las=1, cex.axis=.8)
mtext(expression(mu), side=1, line=2, at=0)
mtext(expression(paste(frac(1, sigma*sqrt(2 * pi)), " ",
    plain(e)^{-frac(-(x - mu)^2,
    2 * sigma^2)})), side=3, line=0)
# 特定の領域をハイライトする (描くのは左から右へ, そして右から左へ)
polygon(c(x[471:591], rev(x[471:591])), c(rep(0,121), rev(y[471:591])),
    col="lightgray", border=NA)
```

2.3 仮説検定

統計的仮説は、一般に、与えられたモデルに基づいて2つの対立する主張の形で定式化される。帰無仮説は、処理効果は他のものと比べて異なるということに関する知識を反映する。2標本に関する結果を記述する分析モデル

$$y_{ij} = \mu_i + \varepsilon_{ij} \quad i = 1, 2; j = 1, 2, \dots, n_i, \quad (2.1)$$

を考える。ここで、 y_{ij} はグループ i の (統計的) 単位 j からの観測値であり、 μ_i はグループ i の平均である。すると、これに対応する仮説として、

$$\begin{aligned} H_0: & \mu_1 = \mu_2 \\ H_1: & \mu_1 \neq \mu_2 \end{aligned} \quad (2.2)$$

がある。ここで、 H_0 は効果が無いという帰無仮説であり、 H_1 (H_A と表記することもある) は H_0 の論理的な否定である。

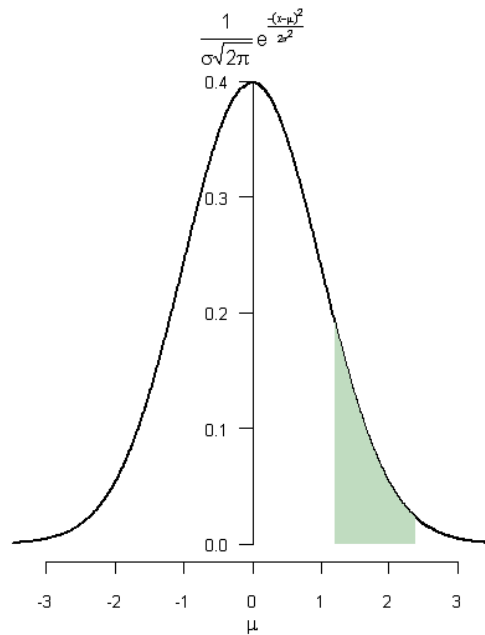


Figure 2.4: 正規分布の密度関数

検定のこのフレームワークは、2種類の誤り（過誤）の可能性を導く。第1種の誤り（その確率を α とする）は、帰無仮説が正しいにも関わらずこれを棄却する誤りであり、第2種の誤り（その確率 β ）は、帰無仮説を棄却すべきであるにも関わらずそうしない誤りである。式で表すと、

$$\alpha = \Pr(\text{第1種の誤り}) = \Pr(H_0 \text{を棄却} \mid H_0 \text{は真}) \quad (2.3)$$

$$\beta = \Pr(\text{第2種の誤り}) = \Pr(H_0 \text{を棄却しない} \mid H_0 \text{は偽}) \quad (2.4)$$

となる。

α は一般に有意水準と呼ばれるが、検定を実行したときに統計ソフトウェアはその結果を出力する。これら2つの誤りはどちらも重要であるが、多くの研究において、第2種の誤りは無視されがちである。図 2.5 は、2つの仮説における2つの量（ α と β ）の関係を分布で表したものである。この図を作成するためのスクリプトは、CRAN のウェブサイトより取得できる（いくつかのコマンドで作ることはそれほど難しくはないが）

（参考）次のコマンドで作成可能。

```
x <- seq(-4,8, by=0.01)
y <- dnorm(x)
z <- dnorm(x, 3, 1)
plot(x, y, xlab="", ylab="", type="l", axes=F)
```

```

lines(x, z)
axis(side=1, cex.axis=.8)
polygon(c(x[601:1200], rev(x[601:1200])), c(rep(0,600), rev(z[601:1200])),
  col="lightgray", border="black")
polygon(c(x[601:1200], rev(x[601:1200])), c(rep(0,600), rev(y[601:1200])),
  col="black", border=NA)
polygon(c(x[1:200], rev(x[1:200])), c(rep(0,200), rev(y[1:200])),
  col="black", border=NA)
text(3, 0.2, expression(H[0] :mu[1] == mu[2] + 3), cex=0.75)
text(0, 0.2, expression(H[0] :mu[1] == mu[2]), cex=0.75)
mtext(expression(Z == frac(mu[1] - mu[2], sigma / sqrt(n))), side=1,
  line=3, cex=0.75)
text(6, 0.4, expression(P(abs(Z)>1.96, H[1])=0.85), cex=0.75)
text(6, 0.38, expression(P(abs(Z)>1.96, H[0])=0.05), cex=0.75)

```

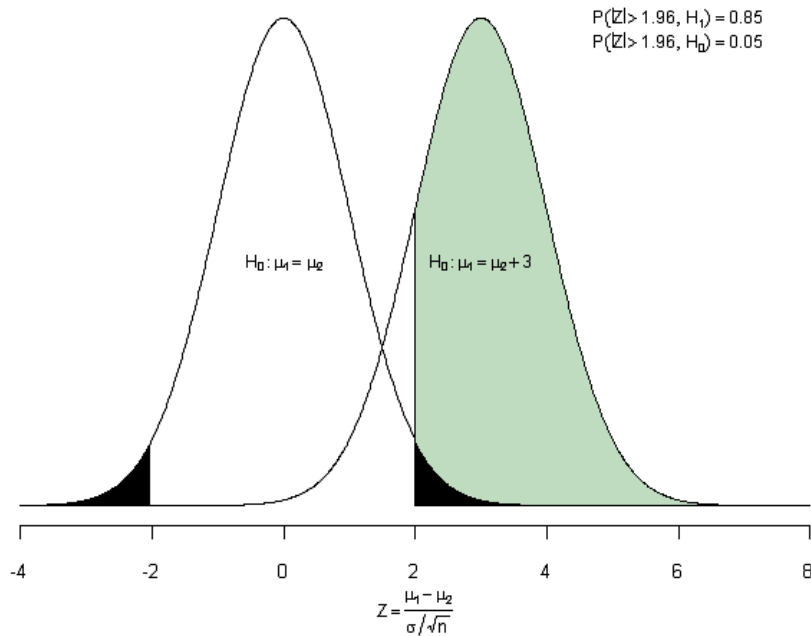


Figure 2.5: 第 1 種の誤りの確率 α と第 2 種の誤りの確率 β .

2.4 2 標本 t 検定

特性に関する 2 組の観測値を比較するとき, (1) 検定統計量を構成し, (2) その標本分布を求め, (3) 関連する p 値を計算, という 3 つの手順で実行する. 既に述べたように, p 値は, 統計量が, データを用いて計算したその値以上となる確率を示す. これは明らかに純粋に頻度論者のアプローチであり, たいていの場合, これで充分である.

検定統計量は、

$$t_0 = \frac{\bar{y}_1 - \bar{y}_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \quad (2.5)$$

となる。ここで、 \bar{y}_1, \bar{y}_2 はグループ平均、 n_1, n_2 はサンプルの大きさ、 S_p はプールされた分散である。 $n_1 = n_2$ のとき、釣り合い型計画という。プールされた分散は単に、群内分散の平均であり、一般に、

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \quad (2.6)$$

で求める。

$n_1 + n_2 - 2$ は、検定統計量の自由度で、自由に変化することができる観測値の数である。

推定のパラダイムにおける2つのアプローチを区別する必要がある。Neyman & Pearsonの観点では、統計的検定は純粹に2分決定（帰無仮説を棄却するか採択するか）に対する回答を与えるのみであり、実験⁶を始める前に決定しておく有意水準、例えば5%、と比較して p 値の位置をさらに解釈することはない。これに対して、Fisher[11] は、 p 値自体が帰無仮説に対する結果の強さの判定の指標となるという考え方を擁護した。

Sir Ronald Aylmer Fisher (1890–1962) は、小標本に対して適切な手法と標本分布の発展に顕著に貢献し、分散分析の父と考えられている。

これらの2つの方法と統計的な結果を解釈する方法については非常に長い間の論争がある。本稿では前者のアプローチ（2分決定ルール）を採用するが、 p 値も与えることにする。 p 値は、一般に漸近理論に基づいて計算される。

信頼区間（CI: Confidence Interval）は、検定統計量の標本分布に基づいて簡単に計算することができる。スチューデントの $T(\nu)$ 分布と呼ばれ、このパーセント点は R で計算することができる（?qt 参照）。2つの平均の差、つまり $\bar{y}_1 - \bar{y}_2$ の $100(1 - \alpha)\%$ 信頼区間は、次式により容易に求めることができる。

$$(\bar{y}_1 - \bar{y}_2) \pm t_{\alpha/2, n_1 + n_2 - 2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}. \quad (2.7)$$

$\alpha = 0.05$ のとき、95%信頼区間となる。信頼区間に関する興味深い議論が Lecoutre 他の文献、例えば、[16, 17] にある。

引っ張り強度データに関数 `t.test()` を適用すと、

```
t.test(y1, y2, var.equal=TRUE)
```

結果は次のようになる。

```
Two Sample t-test
```

```
data: y1 and y2
t = -2.1869, df = 18, p-value = 0.0422
```

⁶Neyman-Pearson の観点では、false alarm の確率がある値 α 以下であることを保証するなかで、検出の確率を最大とする決定ルールを構成する。尤度比検定は大きさ α の最強検定であるが、通常、この検定が利用されることはない。

```

alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.54507339 -0.01092661
sample estimates:
mean of x mean of y
 16.764    17.042

```

R は、 t_0 値、自由度、 p 値、式 2.7 を用いる 95% 信頼区間を出力する。検定では、一般に用いられる 5% の有意水準で有意性を判定する。あるいは、 p 値が帰無仮説に対する証拠の強さを与える。95% 信頼区間が 0 を含むかどうかで判定することも可能である。全体として、2 つの処理間の差は 0.278 kgf/cm^2 ある。

```
as.numeric(diff(apply(y, 2, mean)))
```

`var.equal=TRUE` オプションを省略すると、R は Welch の t 検定を行う。この場合、分散のプールした推定値を用いず、少し保守的ではない p 値を得る近似自由度を用いる。この手法は、Satterthwaite の近似 p 値法と呼ばれる [21, 27]。自由度を計算する式は、

$$\nu = \frac{2(w_1 + w_2)}{w_{12}/(n_1 - 1) + w_{22}/(n_2 - 1)} \quad (2.8)$$

である。

前の例に適用すると、 t 値は -2.187 、自由度は 17.025 、 p 値は 0.043 となる。

```
t.test(y1, y2)
```

整数値でない自由度を報告するとわかりにくいので、無視されることが多い。2 群の分散はそれほど異なっていないので、検定統計量の自由度をそれほど大きく調整する必要が無いので、よく似た p 値を得た ..

2.5 1 つの母平均 μ の基準値との比較

2.6 対応のあるサンプルへの適用

何らかの形で 2 標本が関連しているとき、別の状況が生まれる。例えば、いくつかの検査サンプルを用いてチップ 1 とチップ 2 の両方で試験を行う場合を考える。データは、`hardness.txt` にある。

```

tmp <- scan("hardness.txt", sep=",")
hardness <- data.frame(y=tmp, tip=gl(2,10))
t.test(y ~ tip, data=hardness, paired=TRUE)

```

2 つのチップの間に有意な違いがあると結論づけることはできない ($t(9) = -0.26$, p 値 = 0.798)。2 つの評価値をグラフ化すると (図 2.6 の左)、両データは傾きが 1 の直線に沿って分布していることがわかる。これは、チップ 1 とチップ 2 の間に関係があることを示し

ている．このような特定のコンテキストにおいて，動きの一致性を検証するためのより有効な方法は，2つの観測値の合計の関数としてチップ1とチップ2の差をプロットすることである（図 2.6 の右）．これは，生物医学の分野で [5] により提案された方法である．

```
with(hardness, plot(y[tip==1], y[tip==2], xlab="Tip 1", ylab="Tip 2"))
abline(0,1)
with(hardness, plot(y[tip==1] + y[tip==2], y[tip==1] - y[tip==2],
                    xlab="Tip 1 + Tip 2", ylab="Tip 1 - Tip 2", ylim=c(-3, 3)))
abline(h=0)
```

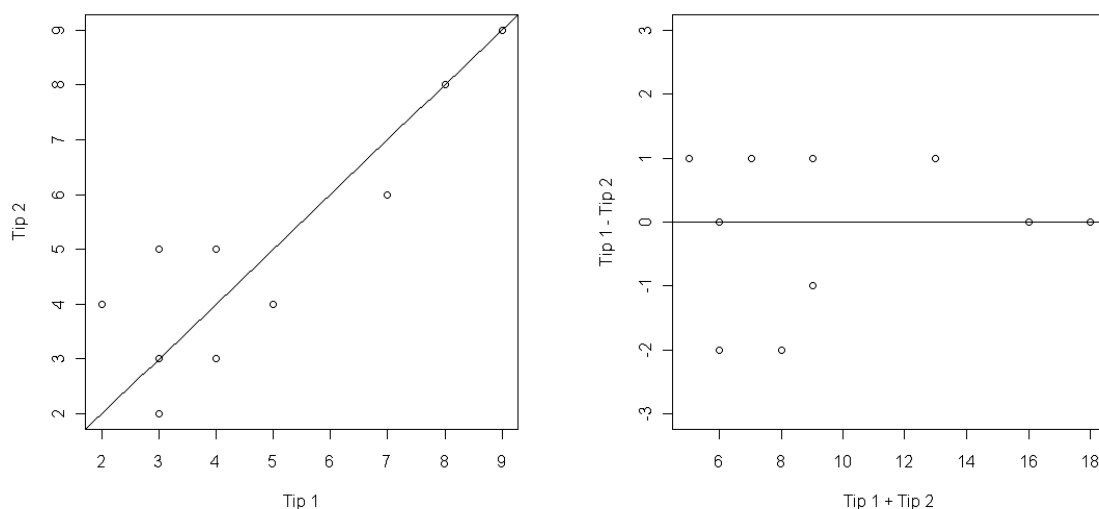


Figure 2.6: 強度の検証実験

対応があることを無視するとどうなるかを見てみよう．

```
t.test(y ~ tip, data=hardness, var.equal=TRUE)
```

期待されるように，自由度は前のものと比べて2倍になる ($n_1 = n_2 = n$ のとき， $n_1 + n_2 - 2 = 2(n - 1)$)． t 値が大きくなったことは追加的なばらつきを説明していないことを意味する．

2.7 ノンパラメトリック法

2標本の検定に対して，データを集めた方法に応じて，2つのノンパラメトリック検定を利用することができる．2つのサンプルが独立なら，Mann-Whitney-Wilcoxon 順位検定を利用し，対応があるデータなら，Wilcoxon 符号付き順位検定を利用する．両者は，R の関数 `wilcox.test` を用いて実行でき，オプションとして `paired={TRUE|FALSE}` を与える．前の例に対しては，

```
wilcox.test(y1, y2)  
wilcox.test(y ~ tip, data=hardness, paired=TRUE)
```

とする。

第3章 1因子の実験：分散分析

3.1 第3章の要約

Montgomery は、1元配置分散分析の基本となる原理を、“古典的”アプローチ（平方和を用いていわゆる分散分析表と呼ばれる表を利用する方法）と回帰アプローチ（モデルパラメータの推定と正規方程式の解を求める方法）の両者により概説する。完全モデルを評価した後で、通常、どの処理平均が他のものと比較して真に異なるかを決定する必要が生じる。だから、多重比較が必要となる。このとき、多数の仮説検定を行うことから、第1種の誤りに注意する必要がある。他のアプローチとして、実験毎の誤差率を膨らませないようにする直交対比の実験がある。最後に、ノンパラメトリック法（Kruskal-Wallisの分散分析）およびその多重比較版を提示する。

3.2 母数効果モデルの分析

エッチング速度データは、ファイル `etchrate.txt` にある。分析を始める前に、特性値の変化の仕方を図で確認しておく（図 3.1）。

```
etch.rate <- read.table("etchrate.txt", header=T)
grp.means <- with(etch.rate, tapply(rate,RF,mean))
with(etch.rate, stripchart(rate~RF, vert=T, method="overplot", pch=1))
stripchart(as.numeric(grp.means)~as.numeric(names(grp.means)), pch="x",
           cex=1.5, vert=T, add=T)
title(main="エッチング速度データ", ylab=expression(paste("エッチング速度 (",
  ring(A), "/min)")), xlab="RF 電力 (W)")
legend("bottomright", "群平均", pch="x", bty="n")
```

この散布図から、RF 電力が増加するとエッチング速度も増加することが明白である。実際、平均エッチング速度は、160W での $551.2 \text{ \AA}/\text{min}$ から、220W では $707.0 \text{ \AA}/\text{min}$ に増加している。さらに、この増加の仕方は線形のようなのであるが、これについては後に議論する。

1元配置分散分析の最も基本的なモデル式は、

$$y_{ij} = \mu_i + \varepsilon_{ij} \quad i = 1, \dots, a; j = 1, \dots, n, \quad (3.1)$$

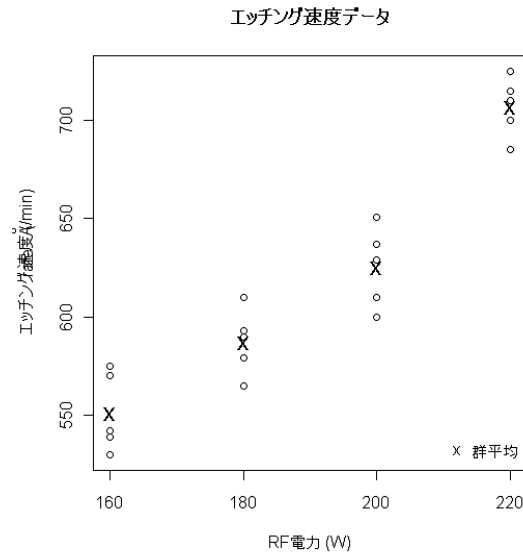


Figure 3.1: エッチング速度データ

である．ここで， y_{ij} は，処理（または群） i の j 番目の観測値であり， ε_{ij} は，NIID¹を仮定する誤差である．式 3.1 を平均モデルという． μ_i を一般平均 μ との関係で考えると， $\mu_i = \mu + \tau_i$ と表すことができ，式 3.1 は

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \quad i = 1, \dots, a; j = 1, \dots, n \quad (3.2)$$

となる．

τ_i を効果といい，このモデルを効果モデルという．

3.3 モデルパラメータの推定

分散分析表（表 3.1）は，次のコマンドを用いて作成することができる．関数 `aov()` と `lm()` は，分散分析モデルを作成するときに特に重要であるが，関数 `lm()` を利用するとき，変数のコーディングが重要であることを強調しておく．質的変量は R の用語でいう因子（factor）である必要があり，そうでないと回帰分析が実行される．

```
# 先ず，各変数を因子に変換する
etch.rate$RF <- as.factor(etch.rate$RF)
etch.rate$run <- as.factor(etch.rate$run)
# 次に，モデルを作る
etch.rate.aov <- aov(rate ~ RF, etch.rate)
summary(etch.rate.aov)
```

Table 3.1: エッチング速度実験に対する分散分析モデルの結果 .

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
RF	3	66870.55	22290.18	66.80	0.0000
Residuals	16	5339.20	333.70		

エッチング速度 RF の平均平方 (Mean Sq: Mean Square) は残差 (residuals) (誤差ともいう) の平均平方 (MS: Mean Square) これは群内分散の平均として求められるが、と比べて遙かに大きい。いい換えると、群内変動に比べて群間変動が遙かに大きく、このような結果は偶然だけでは生じにくい。R は総平方和を表示しないが、これはたいていの場合、重要ではないからである。ちなみに、総平方和は、 $66870.55 + 5339.20 = 72209.75$ (自由度は $3 + 16 = 19$) である。

分散分析の結果に戻ると、一般平均と処理効果 $\hat{\tau}_i = \bar{y}_{1.} - \bar{y}_{..}$ を次により推定することができる。

```
# 一般平均
(erate.mean <- mean(etch.rate$rate))
# 処理効果
with(etch.rate, tapply(rate, RF, function(x) mean(x) - erate.mean))
```

これにより

RF	160	180	200	220
$\hat{\tau}_i$	-66.55	-30.35	7.65	89.25

となる。

関数 `model.tables()` を用いて同じ結果を得ることができる。この関数は、複雑な計画に対して効果の大きさに関する有益な情報を与える。

```
model.tables(etch.rate.aov)
```

最後に、処理効果の $100(1 - \alpha)\%$ 信頼区間を求める。これには、次の公式を用いて計算することができる。

$$\bar{y}_{i.} \pm t_{\alpha/2, N-a} \sqrt{\frac{MS_e}{n}}. \quad (3.3)$$

任意の 2 つの処理を比較するには、次式を用いる。

$$(\bar{y}_{i.} - \bar{y}_{j.}) \pm t_{\alpha/2, N-a} \sqrt{\frac{2MS_e}{n}}. \quad (3.4)$$

R を直接用いる場合、プールした標準偏差を求めてこれに対応する t 分布のパーセント点を求めるだけでよい。

¹Independent and Identically Distributed Normal: 独立で同一の正規分布に従うこと。

```
MSe <- summary(etch.rate.aov)[[1]][2,3]
SD.pool <- sqrt(MSe/5)
t.crit <- c(-1, 1) * qt(.975, 16)
```

よって、 $\bar{y}_i \pm 17.3$ により1つの平均の95%信頼区間を求めることができ、 $(\bar{y}_i - \bar{y}_j) \pm 24.5$ により平均の差の信頼区間を求めることができる。後者に対しては、独立サンプルの $\bar{y}_1 - \bar{y}_2$ に対する t 検定の結果と比較することができる。

```
with(etch.rate, t.test(rate[RF==160], rate[RF==180], var.equal=TRUE))
```

これに対する平均の差の95%信頼区間は、 $[-63.11, -9.29]$ であり、上記の結果 $[-60.7, -11.7]$ と少し異なる。これは、分散分析のフレームワークでは、全てのサブサンプルに基づいて計算したプールした標準偏差を利用しているからである。

```
mean(tapply(etch.rate$rate, etch.rate$RF, var))/5
mean(c(var(etch.rate$rate[etch.rate$RF==160]),
        var(etch.rate$rate[etch.rate$RF==180])))
```

しかし、 t 検定では、関連する2つのサンプルのみを利用する。この場合、プールされた標準偏差は大きくなり、13.05である。それに応じて、信頼区間は広くなる。

```
mean(c(var(etch.rate$rate[etch.rate$RF==160]),
        var(etch.rate$rate[etch.rate$RF==180])))
```

一般に、分散分析の F 検定が有意のとき、少なくとも1対の平均が異なる（あるいは、 $(i \neq j | \mu_i - \mu_j \neq 0)$ なる i, j が存在する）ことを意味する。これを、対応するグループ間の t 検定により確認することができる。しかし、同時に全ての平均の差を t 検定するのは良くない。なぜなら、多くの比較を行うと第1種の誤りが大きくなるからである。この方法では、有意の差を検出する確率は $\alpha = 0.05$ より大きくなってしまふ。実験全体の有意水準が $\alpha = 0.05$ であることを保証するには、前の表現における $\alpha/2$ を $\alpha/2r$ で置き換える必要がある。ここで、 r は同時に計算する区間の数である。これは、Bonferroniの定理に基づく。

Rはオブジェクト指向言語なので、オブジェクトetch.rate.aovに関数confint()を直接適用したくなるだろう。しかし、Rの出力は次のようになる。

```
> confint(etch.rate.aov)
                2.5 %    97.5 %
(Intercept) 533.88153 568.51847
RF180        11.70798  60.69202
RF200        49.70798  98.69202
RF220       131.30798 180.29202
```

結果は、処理効果が群平均とエッチング速度の全体平均との差(τ_i)で表現されているならば、期待される95%信頼区間とは全く異なっている。ここで計算されているのは、規

Carlo Bonferroni (1892–1960) は、多重推定にとどまらず、確率論に貢献した。Bonferroniの調整は、Booleの不等式に依存するが、これに関する論文を2つ書いている。

準とする水準，ここでは RF の第 1 水準（つまり，160W），と処理効果との差に対する 95% 信頼区間である．そのため，差 $\bar{y}_4 - \bar{y}_1$ の 95% 信頼区間が [131.3, 180.3] となっている²．式 3.4，例えば RF220 というラベルのついた最終行を用いて結果の正しさを確認することができる．

```
as.numeric(grp.means[4] - grp.means[1]) + c(-1, 1) * qt(.975, 16) * sqrt(2 * MSE / 5)
```

3.4 モデルのチェック

モデルチェックは，次の仮定の確認を含む（重要性の順）：

1. 独立性，
2. 等分散性（群内の分散の等しさ），
3. 残差の正規性．

要するに， $e_{ij} = y_{ij} - \hat{y}_{ij}$ で定義される残差の分散に構造がなく，処理間でバランスしていることである．

モデルの検証は，図を用いて行うことができる．これがしばしば推奨される方法であるが，上記の各仮説を検定する方法もある．図 3.2 にいくつかの診断プロットを示す．

```
opar <- par(mfrow=c(2,2), cex=.8)
plot(etch.rate.aov)
par(opar)
```

`aov()` の結果に対して `plot` メソッドを適用すると，オプション `which=` でコントロールできるグラフをいくつか表示することができる．R がデフォルトで作成するのは（`?plot.lm` で確認せよ），予測値に対する残差のプロット，予測値に対する $\sqrt{e_{ij}}$ の尺度-位置プロット，正規 QQ プロット，レベレッジ（てこ比）に対する残差のプロットである．最初の 2 つの図は，分散の均一性と正規性の仮定からの乖離をチェックするのに有益である．最後のプロット（残差 vs. レベレッジ）により，影響力の大きい観測値の存在を調べることができる．これは，別々に作成しなければならない診断プロットに対する便利なラッパー関数である．例えば，`plot(fitted(etch.rate.aov), residuals(etch.rate.aov))` により最初の図（残差 vs. 予測値）を作成することができ，`qqnorm(residuals(etch.rate.aov))`；`qqline(residuals(etch.rate.aov))` により正規 QQ プロットを作成することができる．`derived plan` や予測変数の一部に関心がある場合は，これらのコマンドを用いることは有益である．

外れ値を大まかにチェックするために，標準化残差を用いることができる．標準化残差は，

$$d_{ij} = \frac{e_{ij}}{\sqrt{MS_e}} \quad (3.5)$$

²ランダムなのは真の（母集団）のパラメータではなく，信頼区間であることを思いだそう．

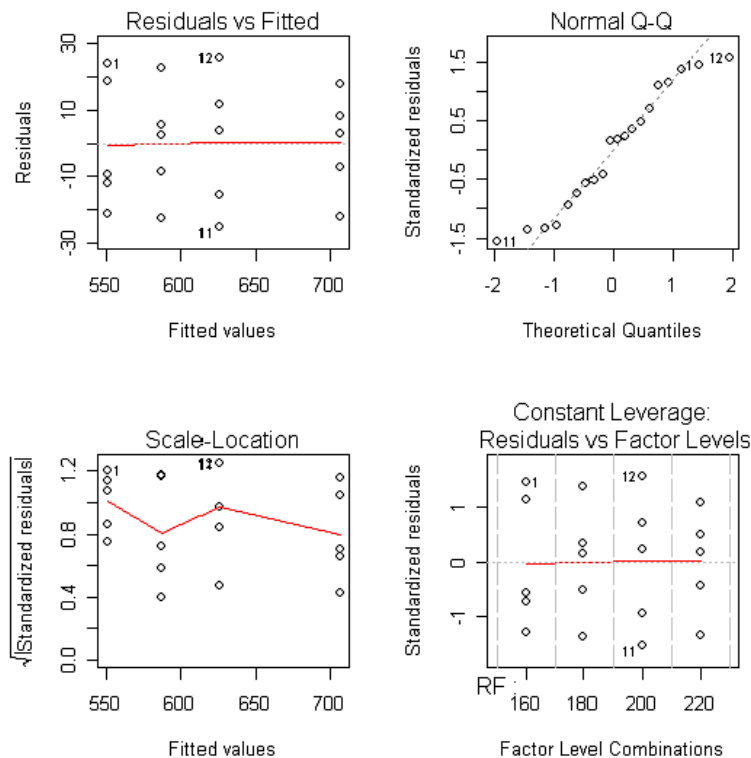


Figure 3.2: ANOVA モデルのチェック .

で定義されるが, $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$ のとき, $\mathcal{N}(0, 1)$ に従うことがわかる .

独立性の問題は, 主に実験の計画やデータを集める方法から生じる . おそらく, 独立性のチェックのための最も簡単なグラフ化の手法は, 実験順序や時間のインデックスに対して残差をプロットすることである (図 3.3)³ . この手法により, 等分散の仮定をチェックすることができる . なぜなら, 分散が一定でないと, 平均特性が異なる観測値の集合が存在したり, 外れ値のシステマティックなパターンが表れたりするからである .

図 3.3 を見ると, このようなパターンはなく, 独立性の仮説を棄却する理由はない . グラフ化の手法に対する歴史的な検定手法として, Durbin-Watson 検定がある . これは, 誤差の系列相関を検定することを目的とし, デフォルトでは, ラグ 1 を用いる . パッケージ `car` や `lmtest` で利用することができる .

```
require(car)
durbin.watson(etch.rate.aov)
```

³ (訳注) 次のコマンドで作成できる .
`plot(etch.rate.aov$residuals, ylab="残差")`
`abline(h=0, lty=2)`

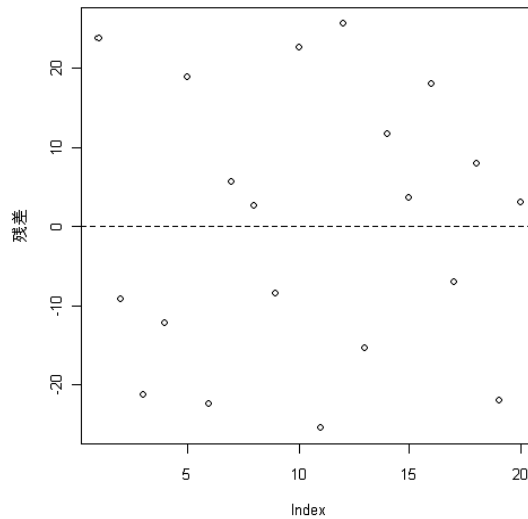


Figure 3.3: 分散分析モデル (2) のチェック

均一分散，または等分散の仮定が実際には最も重要である．というのは，群内分散の平均を用いてプールした分散を推定するからである．この仮定からの逸脱は，分散が異なる群がいくつか存在することを意味する．有意に“大きい”あるいは“小さい”ということは，何を計測するかということに依存するため，均一分散の仮定を検定により棄却できても，この疑問に答えることはできない．実際， $H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_a^2$ を棄却したとき，何がいえるだろうか．最も広く利用されている等分散の検定は，Bartlett の検定である．

```
bartlett.test(rate ~ RF, data=etch.rate)
```

正規性からの大きな乖離が疑われる場合，等分散の検定の代わりに Levene (ルビーン) の検定を用いてもよい．この検定はパッケージ `car` で利用できる．

```
levene.test(etch.rate.aov)
```

最後に，残差の正規性は，図 3.2 に示すような QQ プロット (フランス語でいう，*droite de Henry*)，そこでは，図の対角位置の近くに点が存在することが期待される，や Shapiro-Wilk 検定を用いて直接検証することができる．後者の場合，各サブサンプルに対して検定しなければならないため，群内のデータ数が少ない場合，問題となる．

```
shapiro.test(etch.rate$rate[etch.rate$RF==160])
```

Maurice Stevenson Bartlett (1910–2002) は空間的，時間的パターンを持つデータの分析の研究を行った．彼は，また，統計的推測と多変量解析の研究において貢献した．

3.5 処理平均の比較

処理数が $a = 4$ のとき，比較の組合せ数は $4(4-1)/2$ となり，帰無仮説は $H_0: (i, j)$ に対して $\mu_i = \mu_j$ ，となる．R を用いてパラメトリックな多重比較を行う方法はいくつかあ

る．おそらく最も一般的でわかりやすいのは，全ての処理平均の対での比較を行うことだろう．第 1 種の誤りが膨らまないようにするため，いくつかの方法が提案されている．これらの中で，最も保守的な方法は Bonferroni の修正法であり，これは，比較回数で α の名目値を調整する（これに関しては既にページ 16 で議論した）．

先ず，Bonferroni または Hochberg の修正を用いて，対での t 検定を行うと，処理平均が等しいという全ての帰無仮説が棄却される（表 3.2 と 3.3）． p 値が少しずつ異なっているが，これは Hochberg の場合，修正項を適応的な方法で計算するからである．

```
pairwise.t.test(etch.rate$rate, etch.rate$RF, p.adjust.method="bonferroni")
pairwise.t.test(etch.rate$rate, etch.rate$RF, p.adjust.method="hochberg")
```

Table 3.2: Bonferroni の方法

	160	180	200
180	0.038	–	–
200	5.1e-05	0.028	–
220	2.2e-09	1.0e-07	1.6e-05

Table 3.3: Hochberg の方法

	160	180	200
180	0.0064	–	–
200	2.5e-05	0.0064	–
220	2.2e-09	8.5e-08	1.1e-05

他にも，多重検定により膨れあがる第 1 種の誤りを考慮するための修正した検定統計量を用いる方法がある．Tukey の HSD⁴検定がそれである [24]．R の関数 TukeyHSD() は，調整した p 値と 95%信頼区間を与える．さらに，洗練されたグラフによる要約（図 3.4）を与える．Tukey の HSD 検定を行うと，制約付 t 検定⁵と同じ結論を得る．結果を表 3.4 と図 3.4 に示す．95%信頼区間のどれも 0 を含まないことが分かる．

```
TukeyHSD(etch.rate.aov)
plot(TukeyHSD(etch.rate.aov), las=1)
```

Table 3.4: Tukey の HSD 法

$i - j$	δ	LB-CI	UP-CI	adj. p
180-160	36.2	3.145624	69.25438	0.0294279
200-160	74.2	41.145624	107.25438	0.0000455
220-160	155.8	122.745624	188.85438	0.0000000
200-180	38.0	4.945624	71.05438	0.0215995
220-180	119.6	86.545624	152.65438	0.0000001
220-200	81.6	48.545624	114.65438	0.0000146

処理平均のペア 160–180 と 200–180 の p 値は，前と同様，0.05 と 0.10 の間にあり，他の p 値と比べて大きい．これは，図 3.4 に示した 95%信頼区間の下限値からも明らかである．

⁴HSD は *Honest Significant Difference* を意味する．

⁵（訳注）フィッシャーの制約付 LSD 法のこと．

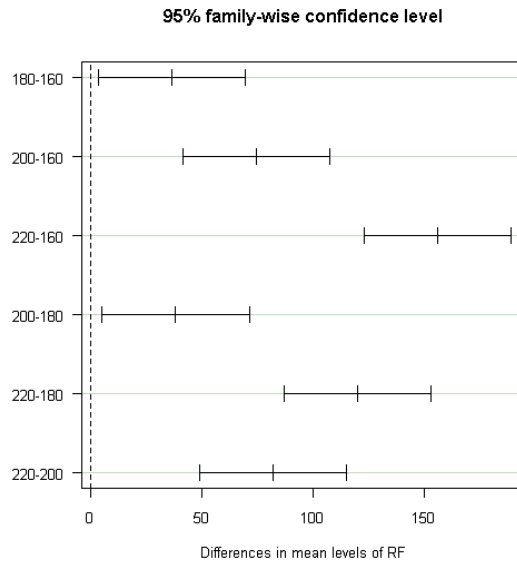


Figure 3.4: エッチング速度データに対する 1 元配置分散分析の HDS 検定の信頼区間

他の方法に関しては、ここでは議論しない。別の手法の長所や欠点に関心のある読者は、[6] (第 5 章) や [22] を参照のこと。多重比較用の R のパッケージとして `multcomp` (また、パッケージ `Design`) がある。これは、線形モデルの多重比較を取り扱うことができる。他の有益な参照文献として [8] がある。この添付パッケージに `multtest`⁶ がある。

前の手法の代替法として、いくつかの特定の処理平均と他のものを検定するための対比を構成する方法がある。もしこれらの対比、あるいは平均の差が互いに直交するように計画することができると、全体の有意水準を膨張させることなく、名目的な有意水準 0.05 で検定を実行することができる。

R でこのような対比を作る方法はいろいろあるが、ここでは 2 つのみ見る。前の方法で用いた行列計算に読者が混乱することがないことを期待するが。

3.6 検出力とサンプル数

検出力とサンプルサイズの決定は、2 つの関連した概念である。R では、関数 `power.t.test()` を用いて 1 標本、2 標本の t 検定の計算を行うことができる (母数効果の) 1 元配置分散分析の場合、パッケージ `QuantPsyc` 内の関数 `powerF()` または関数 `power.anova.test()` により実行することができる。

⁶さしあたり、その機能の一部のみを調べ、多重比較と p 値の調整という名前の非常に短いノートを作成した。これは、<http://www.aliquote.org/memos/2008/07/26/multiple-comparisons-and-p-value-adjustment/> より取得可能である。

この最後の関数は、 F 分布は、自由度を調整することにより操作できる（特に、分母の）ということ、あるいは、ANOVA や回帰分析で計算された F 検定の値に反映される効果は、サンプルサイズまたは所与の計画に対する検出力の推定を可能とするということに依存している。

一般に、検出力の計算は OC 曲線、これは第 2 種の誤りの確率をパラメータ Φ に対してプロットするものであるが、を用いて行う（図 3.6 参照）。例として、エッチング速度データの OC 曲線を図 3.5 に示す。

基本的に、2 つの非常に似た状況がある：1 つは、対立仮説のもとで期待処理平均を実験者が特定するものであり、もう 1 つは、処理平均の任意の 2 つの差の最小を実験者が指定するものである。

最初の場合を説明するために、プラズマエッチング実験を用いた応用を考える。4 つの処理平均が

$$\mu_1 = 575 \quad \mu_2 = 600 \quad \mu_3 = 650 \quad \text{and} \quad \mu_4 = 675,$$

で、 $\alpha = 0.01$ 、 $\sigma = 25 \text{ \AA}/\text{min}$ とする。そして、

$$\Phi^2 = \frac{n \sum_{i=1}^4 \tau_i^2}{a\sigma^2} = \frac{n(6250)}{4(25)^2} = 2.5n \quad (3.6)$$

の場合のみに、実験者は、検出力 0.90 で帰無仮説を棄却したいとする。

R を用いると、次のコードにより検出力を 0.90 とするのに必要なサンプルサイズを計算することができる。

```
grp.means <- c(575, 600, 650, 675)
power.anova.test(groups=4, between.var=var(grp.means), within.var=25^2,
  sig.level=.01, power=.90)
```

これにより次の結果を得る。

```
Balanced one-way analysis of variance power calculation

groups = 4
n = 3.520243
between.var = 2083.333
within.var = 625
sig.level = 0.01
power = 0.9
```

注：n は群内のサンプル数

上記の効果を検出力 0.90 で検出するには、サンプルの大きさを $n = 4$ とすればよい。固定したサンプルの場合（あらかじめ決定した）SD がどのように検出力に影響を与えるかを知りたい場合は、次のスクリプトを用いる。

```

sd <- seq(20, 80, by=2)
nn <- seq(4, 20, by=2)
beta <- matrix(NA, nr=length(sd), nc=length(nn))
for (i in 1:length(sd))
  beta[i,] <- power.anova.test(groups=4, n=nn, between.var=var(grp.means),
                              within.var=sd[i]^2, sig.level=.01)$power
colnames(beta) <- nn; rownames(beta) <- sd
opar <- par(las=1, cex=.8)
matplot(sd,beta, type="l", xlab=expression(sigma), ylab=expression(1-beta), col=1,
        lty=1)
grid()
text(rep(80,10),beta[length(sd),], as.character(nn), pos=3)
title("a=4 の処理平均\n に対する OC 曲線")
par(opar)

```

図 3.5 からわかるように (図の右), 標準偏差 $SD(\sigma)$ を大きくしたりサンプルサイズを小さくしたりすると検出力が小さくなる. 例えば, $\sigma = 50$ (前の場合の 2 倍) と $n = 6$ に対しては, 検出力はわずか 0.60 となる ($\alpha = 0.01$). 検出力を少なくとも 0.80 とするには, サンプルサイズを $n = 10$ まで上げる必要がある.

検出力の計算の 2 つ目の例として, 次の状況を考える. 実験者は, 2 つの処理平均の差が最大 $75\text{\AA}/\text{min}$ のとき, 有意水準 $\alpha = 0.01$, 確率 ≥ 0.90 で帰無仮説を棄却したいとする. SD が $\sigma = 25$ のとき, Φ^2 の最小値は次式で計算できる.

$$\Phi^2 = \frac{nD^2}{2a\sigma^2} = \frac{n(75)^2}{2(4)(25^2)} = 1.125n. \quad (3.7)$$

n の決定は, OC 曲線を用いて行うことができる. この計算を自動的に行う R の関数を見つけることができなかった.

3.7 ノンパラメトリック分散分析

1 元配置分散分析に対するノンパラメトリック手法は, Kruskal-Wallis の ANOVA である. これは, データの順位に基づいて計算するという点において異なるだけである. Kruskal-Wallis 検定は, a 個の処理はそれらの中心的傾向 (ここではメディアン) に関して等しいという帰無仮説を検定する. ノンパラメトリック検定は, パラメトリックなアプローチ同様, 群内のばらつきは等しいことを仮定することに注意. 実際, そうではない場合, 各群を区別することが難しくなる. なぜなら, それらの距離は部分母集団の分散の差によって説明されるからである.

```
kruskal.test(rate ~ RF, data=etch.rate)
```

により Kruskal - Wallis 検定を実行することができる. 結果として, 自由度 3 の検定統計量の値 16.91, および p 値として < 0.001 を得る.

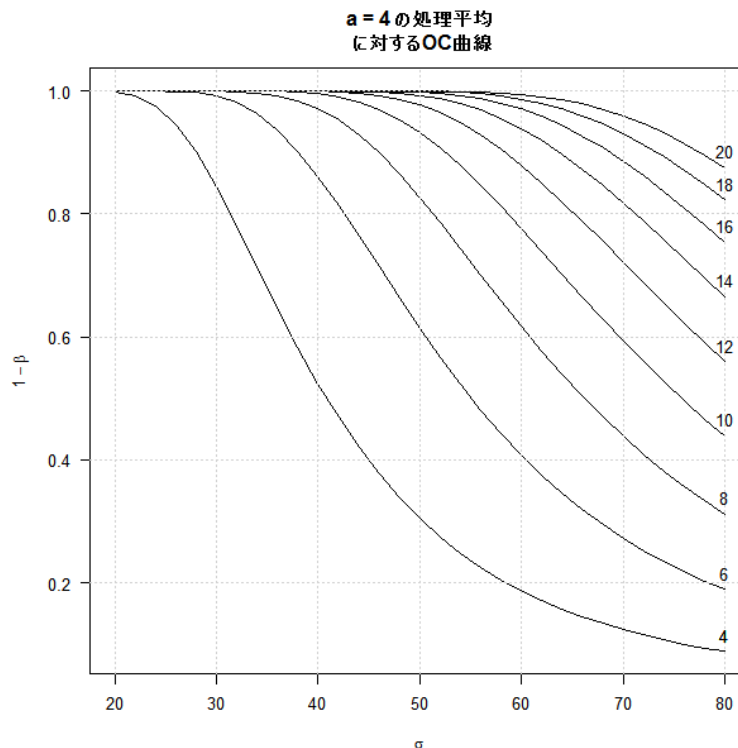


Figure 3.5: 1元配置分散分析に対するOC曲線の例(エッチング速度の例1)

この結果は、古典的なANOVAモデルを用いた前の結果を否定するものではない。ノンパラメトリックな多重比較法は、通常のパラメトリックな手法が持つ“容易さ”はない。実際、どのように検定するかという疑問を生じさせる。

パッケージ `npmc`⁷により、Behrens-Fisher と Steel に基づく非釣り合い型1元配置に対するノンパラメトリックな多重比較を行うことができる。これらの手法は [18] による。

```
library(npmc)
# we need to reformat the data.frame with var/class names
#
etch.rate2 <- etch.rate
names(etch.rate2) <- c("class", "run", "var")
summary(npmc(etch.rate2), type="BF")
```

結果を表 3.5 に示す。

パラメトリックな多重比較 (§ 3.5) と同様の結論を得る。つまり、全ての処理の対は有意に異なる。

⁷ (訳注) 現在、CRAN では公開されていない。

Table 3.5: エッチング速度データに対するノンパラメトリックな多重比較の適用

LB/UP-CI: 95%信頼区間の下限と上限; p 値: 片側, 両側 p 値

$i-j$	effect	LB-CI	UP-CI	p 値 1s	p 値 2s
Behrens-Fisher					
1-2	0.92	0.5764163	1.263584	0.011450580	0.020539156
1-3	1.00	0.9998842	1.000116	0.000000000	0.000000000
1-4	1.00	0.9998842	1.000116	0.000000000	0.000000000
2-3	0.94	0.6758851	1.204115	0.002301579	0.004440345
2-4	1.00	0.9998842	1.000116	0.000000000	0.000000000
3-4	1.00	0.9998842	1.000116	0.000000000	0.000000000
Steel					
1-2	0.92	0.4254941	1.414506	0.07123615	0.13078270
1-3	1.00	0.5054941	1.494506	0.02446374	0.04602880
1-4	1.00	0.5054941	1.494506	0.02417453	0.04631413
2-3	0.94	0.4469949	1.433005	0.05465670	0.10154286
2-4	1.00	0.5054941	1.494506	0.02412958	0.04654181
3-4	1.00	0.5054941	1.494506	0.02414774	0.04635531

分布の仮定に関するさらなるノート ノンパラメトリック検定であってもそのパラメトリックな対応手法と同じ仮定, 特に分散の均一性の仮定を必要とすることを指摘しておくことは重要である. これは並べ替え検定の場合も同じである. もし, 分散 (どんなばらつきかの測度を用いても) の大きさが互いに非常に異なる 2 つの分布を比較したい場合, パラメトリックあるいはノンパラメトリックな検定を用いることに意味はない. 実際, 観測値は 2 つの分布のどちらかから得られたものであるが, 分散の大きさが異なっていると, どちらから得られたものか, 判断できない. いい換えると, 入れ替え可能な仮説を失っているわけである.

しかし, Welch [28] が提案した検定統計量を少し修正すると, 分散の非均一性の場合に対して用いることができる. これをエッチング速度データに適用する.

```
oneway.test(rate ~ RF, etch.rate)
```

結果として F 値が 68.72, p 値が $< .001$ となる. Welch の修正 t 検定 (p. 10) では, 分母の自由度は調整されており, それが報告されることはめったにない.

第4章 乱塊法，ラテン方格，および関連する計画

4.1 第4章の要約

(訳注) 原文になし．

4.2 乱塊法

完備型乱塊法 (RCBD: Randomized Complete block Design) は、潜在的に存在する局外母数を制御しながら関心のある効果を調べるためのよく用いられる方法である．これを、局外母数を考慮するために事後的に調整する共分散分析と混同してはならない．

いわゆる効果モデルは、

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij} \quad (i = 1, 2, \dots, a; j = 1, 2, \dots, b) \quad (4.1)$$

$$\sum_{i=1}^a \tau_i = 0 \quad \text{and} \quad \sum_{j=1}^b \beta_j = 0 \quad (4.2)$$

となる．

RCBD における基本的な平方和の分解は次のようになる．

$$SS_T = SS_{\text{treat}} + SS_{\text{block}} + SS_E \quad (4.3)$$

ここで、*treat* は処理 (treatment)、*block* はブロック因子である．残差平方和 SS_E は、自由度は $(a-1)(b-1)$ を持ち、これは2つの因子 (*treat*, *block*) で説明できないばらつき大きさである．この計画のレイアウトは、2元配置分散分析で繰返しが無い場合 (セルのデータ数が1) に相当し、交互作用は推定できない．計画は直交しているので、モデルに記述する効果の順序は関係ない．このような加法性は常に成立するわけではなく、特に、ブロックと関心のある因子との間に交互作用の存在が期待される、あるいは、残差と予測値との関係を調べているときに交互作用が発見される場合はそうである．この場合、交互作用効果を見つけるためには要因配置実験 (5章, 6章) を用いる方が適切である．

次の例を考えよう．樹脂のバッチをブロックと考え、*flicks* に対する4水準の押し出し圧力の効果をRCBDにより調査することにした．データはファイル `vascgraft.txt` にあり、次表に示す．

樹脂のバッチ (ブロック)							計
PSI	1	2	3	4	5	6	
1	90.30	89.20	98.20	93.90	87.40	97.90	556.9
2	92.50	89.50	90.60	94.70	87.00	95.80	550.1
3	85.50	90.80	89.60	86.20	88.00	93.40	533.5
4	82.50	89.50	85.60	87.40	78.90	90.70	514.6
計	350.8	359.0	364.0	362.2	341.3	377.8	$y.. = 2155.1$

```
x <- scan("vascgraft.txt")
PSI.labels <- c(8500, 8700, 8900, 9100)
vasc.graft <- data.frame(PSI=gl(4, 6, 24), block=gl(6, 1, 24), x)
vasc.graft.aov <- aov(x ~ block + PSI, vasc.graft)
```

図 4.1 の a, b に応答の図 (箱ひげ図) を示す. 左の図 (a) は, ブロック (block) で応答を層別したものであり, 右 (b) は, 処理 (treat) で層別したものである.

```
boxplot(x ~ block, data=vasc.graft, xlab="ブロック",
        main="ブロックで層別した応答の箱ひげ図")
boxplot(x ~ PSI, data=vasc.graft, xlab="押し出し測度 (PSI)",
        main="処理で層別した応答の箱ひげ図")
```

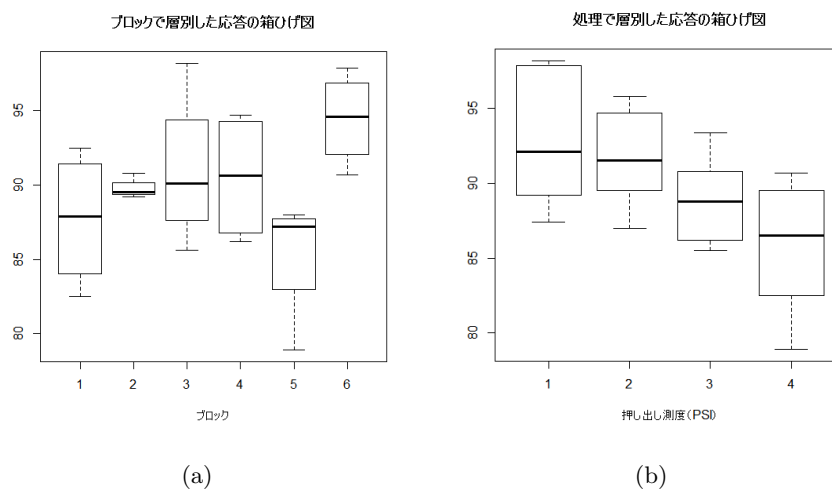


Figure 4.1: 代用血管実験の結果.

2つの因子に関する情報を表示したい場合, いわゆる交互作用プロットを利用することができる. 2つの因子の交互作用を考慮していない場合でもそうである. 図 4.2 は, 各ブロックに対する処理 (横軸) の関数として特性を色分けし, 線種を変えてプロットしたものである.

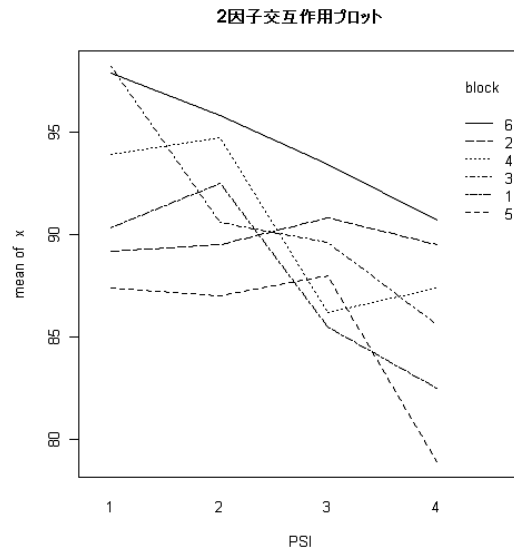


Figure 4.2: 代用血管実験の結果（続き）.

表 4.1 に, R シェルで `summary(vasc.graft.aov)` を実行することにより得られる古典的分散分析の結果を示す. 処理 (PSI) の効果は高度に有意であり, 処理効果がないという帰無仮説を棄却できる.

Table 4.1: モデル $y = \mu + \text{PSI}_i + \text{block}_j$ に対する結果.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
block	5	192.25	38.45	5.25	0.0055
PSI	3	178.17	59.39	8.11	0.0019
Residuals	15	109.89	7.33		

ブロック構造を無視することは不正確な結果を生み出すが, 依然として有意である. 分散分析モデルを作成した後, モデルの適切性をチェックするのは良い習慣である. それには (残差の) 正規性の仮定のみならず予測値と残差との間の関係 (*homoscedasticity*) をチェックする必要がある. 図 4.3 に (標準化した, または生の) 残差 vs. 予測値のプロット, QQ プロット, レベレッジ (テコ比) のプロットといったいくつかのプロットを示す.

```
opar <- par(mfrow=c(2, 2), cex=.8)
plot(vasc.graft.aov)
par(opar)
```

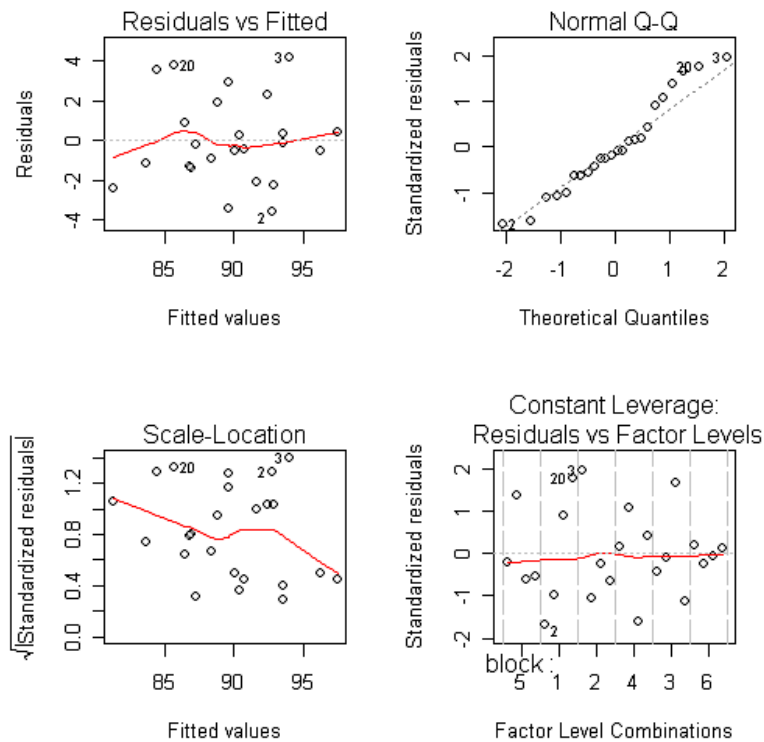


Figure 4.3: 代用血管データのモデルチェック

欠測値の取り扱い RCBd で欠測値が生じた場合，処理はブロックに対して直交しなくなる．一般的な解決法として2つある．

- 近似法: 欠測値を予測値を予測し，分析は通常通り行う．自由度は，残差平方和の自由度から1を引く；
- 正確法: 一般的な回帰の検定を行う（ここでは扱わない．Montgomeryの本，pp. 133–136 参照）．

10番目の観測値を削除する

```
x2 <- x
```

```
x2[10] <- NA
```

```
vasc.graft2 <- data.frame(PSI=gl(4, 6, 24), block=gl(6, 1, 24), x2)
```

欠測値 y_{10} を推定したい．このとき，誤差の平方和に対する寄与が最小になるようにしたい．これは，次を満たす \tilde{y}_{10} を推定することと等しい．

$$\min_y \sum_{i=1}^a \sum_{j=1}^b y_{ij}^2 - \frac{1}{b} \sum_{i=1}^a \left(\sum_{j=1}^b y_{ij} \right)^2 - \frac{1}{a} \sum_{j=1}^b \left(\sum_{i=1}^a y_{ij} \right)^2 + \frac{1}{ab} \left(\sum_{i=1}^a \sum_{j=1}^b y_{ij} \right)^2$$

$$\frac{d SS_E}{dy} = 0 \text{ より,}$$

$$\tilde{y}_{10} = \frac{ay'_i + by'_j - y'_{..}}{(a-1)(b-1)}$$

となる。 \tilde{y}_{10} の値は 91.08 である。

4.3 ラテン方格計画

ラテン方格計画は、実験計画においてブロック因子を考慮するためのもう 1 つの方法である。これにより、2 つの局外因子を取り扱うことができる。

ラテン方格は、処理列（例えば、 $\{A, B, C, D, E\}$ ）を行と列に関する円順列を統合することにより得ることができる。

ファイル `rocket.txt` に Montgomery のロケットプロペラ問題のデータがある。これは次のコマンドで読み込むことができる。

```
rocket <- read.table("rocket.txt", header=T)
```

処理の割り付けを次表に示す。

Batch	Operators				
	1	2	3	4	5
1	A	B	C	D	E
2	B	C	D	E	A
3	C	D	E	A	B
4	D	E	A	B	C
5	E	A	B	C	D

コマンド

```
plot(y ~ op + batch + treat, rocket)
```

により、3 つの因子のうちの 1 つの関数として y の箱ひげ図を連続的に作成して、調べることができる。

```
rocket.lm <- lm(y ~ factor(op) + factor(batch) + treat, rocket)
anova(rocket.lm)
```

計画は釣り合っているので、説明変数を記載する順序は結果に影響を与えない。分散分析の結果を表 4.2 に示す。処理効果 (`treat`) の F 値のみを解釈すればよい。

Table 4.2: ロケットデータの分析

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
factor(op)	4	150.00	37.50	3.52	0.0404
factor(batch)	4	68.00	17.00	1.59	0.2391
treat	4	330.00	82.50	7.73	0.0025
Residuals	12	128.00	10.67		

図 4.4 は、測定された特性に対する因子効果の構造を再現する試みである¹。図より、応答のばらつきの多くは処理条件 (treat) により説明することができ、オペレータ (op) はブロック因子 (batch) よりも大きな効果を持つ。

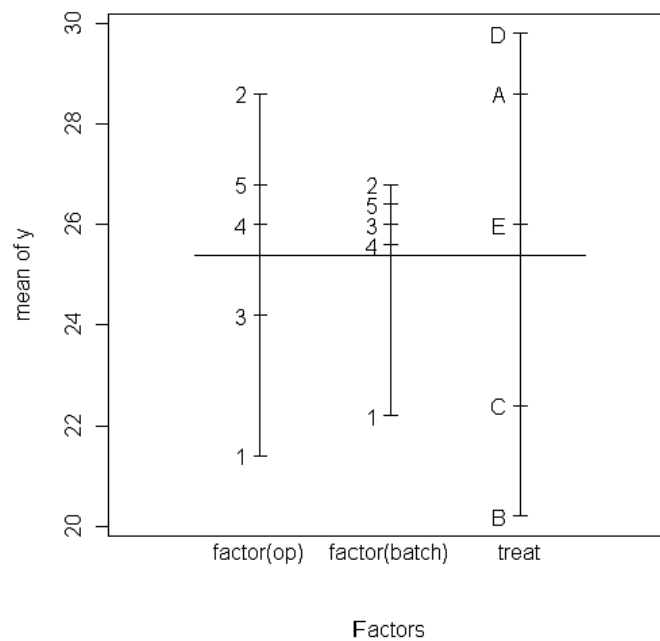


Figure 4.4: 効果プロット。

¹ (訳注) 次のコマンドで作成できる。
`plot.design(rocket.lm$model)`

4.4 グレコラテン方格計画

4.5 釣り合い型不完備ブロック計画

釣り合い型不完備ブロック計画 (BIBD: Balanced Incomplete Block Designs) は乱塊法の一つであるが、実験における各ブロックにおいて全ての処理が実行されない。処理数を a 、各ブロックにおける最大処理数を k ($k < a$) とすると、BIBD は $\binom{a}{k}$ 個の処理の組合せから構成される。

Douglas Montgomery は、教育評価や臨床試験において広く用いられているこの計画法に対する優れた説明を行っている。これに関する付加的な説明については、[14, 3] を参照のこと。しかし、教育的な視点からいうと、伝統的に BIBD と呼ばれている方法は、正式には真の BIBD ではない。実際、ブロックは因子として、因子はブロックとして取り扱われる (例えば、[25])。

触媒実験例 (次表) を考える。これは、4 つの異なる原料のバッチに対して触媒の種類の間数として反応時間を研究するものである。バッチはブロックとして考える。

処理	ブロック (原料のバッチ)				y_i
	1	2	3	4	
1	73	74	-	71	218
2	-	75	67	72	214
3	73	75	68	-	216
4	75	-	72	75	222
$y_{\cdot j}$	221	224	207	218	$y_{\cdot\cdot} = 870$

処理数を a 、ブロック数を b とする。各ブロックには k 処理が配置され、全実験数を r とする。すると、全データ数は $N = ar = bk$ であり、同じブロックに表れる各ペアの処理数は

$$\lambda = \frac{r(k-1)}{a-1}, \quad \lambda \in \{0, 1, 2, \dots\}$$

となる。

$a = b$ のとき、対称計画という。 λ は整数でないといけないので、許容できる解の空間は、かなり少なくなる。例えば、制約: $r = 4, t = 4, b = 8, k = 2$ の場合、BIBD は可能ではない。²

```
tab.4.21 <- matrix(c(73,NA,73,75,74,75,75,NA,NA,67,68,72,71,72,NA,75), nc=4)
tab.4.21.df <- data.frame(rep=as.vector(tab.4.21),
                        treat=factor(rep(1:4, 4)),
                        block=factor(rep(1:4, each=4)))
summary(aov(rep ~ treat + block + Error(block), tab.4.21.df))
```

²しかし、partially 釣り合い型不完備ブロック計画は可能である。

、残差と処理効果の調整した正確な推定値を得るためには、blockであることを指定する必要がある。本例では、触媒の種類は有意な主効果を持つ。しかし、Rにおいて、lm()に強制的に順序づけられた効果項を与えることにより、適切な平方和を計算することができる。

```
anova(lm(rep ~ block + treat, tab.4.21.df))
```

は、ブロック因子で調整した正しい処理効果を与える。

aovの出力の詳細を次に示す。Rは総平方和をブロック変動と残差変動に正しく分解していることがわかる。

```
Error: block
      Df Sum Sq Mean Sq
treat  3 55.000  18.333

Error: Within
      Df Sum Sq Mean Sq F value Pr(>F)
treat   3 22.7500  7.5833  11.667 0.01074 *
Residuals 5  3.2500  0.6500
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

このようにして、触媒の効果に対する調整した平方和 (adjusted MS) を計算することができる。ブロック効果に対する調整した平方和についても関心がある場合、適切な誤差項、つまり、Error(treat)を利用することにより可能であり、次の結果を得る。

```
Error: treat
      Df Sum Sq Mean Sq
treat  3 11.6667  3.8889

Error: Within
      Df Sum Sq Mean Sq F value Pr(>F)
block   3 66.083  22.028  33.889 0.0009528 ***
Residuals 5  3.250  0.650
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Minitabのように、両者の推定値を単一パスで得るには、aov()関数の2回の呼び出しをラップすればよい。表4.3は、両推定値(調整値と非調整値)と、それらのp値をまとめたものである。

他の解法として、パッケージagricolaeにある関数BIB.test()を利用することもできる。この関数の呼び出しに、式としてのインターフェースはないため、ブロック因子、固定効果の処理、応答変数を別々に指定する必要がある。

```
require(agricolae)
BIB.test(tab.4.21.df$treat, tab.4.21.df$treat, tab.4.21.df$rep,
         method="tukey", group=FALSE)
```

Table 4.3: BIB 分析の要約

Effect	df	MS	F	p-value
treat	3	3.889		
treat (Adj.)	3	7.583	11.667	0.01074
block	3	18.333		
block (Adj.)	3	22.028	33.889	0.00095

ノート この関数の機能，およびふるまい全てを調べることはしない（例えば，パラメータ group=）．さらに，上記のコードに対する正しい結果を得ることができない．

Tukey の対差（pairwise difference）は，次で得ることができる．

```
tab.4.21.lm <- lm(rep ~ block + treat, tab.4.21.df)
treat.coef <- tab.4.21.lm$coef[5:7]
# 触媒 4（基準値）の効果が表示されていないので，表示する．
treat.coef <- c(0, treat.coef)
pairwise.diff <- outer(treat.coef, treat.coef, "-")
```

summary(tab.4.21.lm) の出力より，標準誤差の大きさを 0.6982 と推定していることがわかる．より一般的に， $\sqrt{\frac{2k}{\lambda t}}\hat{\sigma}$ により SE を得ることができる．対応する Tukey の棄却限界（ $1 - \alpha = 0.95$ ）は，

```
crit.val <- qtkey(0.95, 4, 5)
```

で与えられる．

しかし，関数 BIB.test() では，処理水準を比較するための LSD，Tukey 検定または Waller - Duncan 検定を直接指定する．

表の形で要約する代わりに，対での比較の関数として信頼区間をグラフ化することができる（図 4.5）．

```
ic.width <- crit.val * 0.6982 / sqrt(2)
xx <- pairwise.diff[lower.tri(pairwise.diff)]
plot(xx, 1:6, xlab="対差
      (95%信頼区間)", ylab="", xlim=c(-5, 5), pch=19, cex=1.2, axes=F)
axis(1, seq(-5, 5))
mtext(c("4-1", "4-2", "4-3", "1-2", "1-3", "2-3"), side=2, at=1:6, line=2, las=2)
segments(xx-ic.width, 1:6, xx+ic.width, 1:6, lwd=2)
abline(v=0, lty=2, col="lightgray")
```

次に，この BIB が完全ランダム化計画（ブロックなし）より効率的かどうかに関心がある場合，[15] に従って，相対効率 RE を次式で計算することができる．

$$RE = \frac{\sigma_{CRD}^2}{\sigma_{RCBD}^2}$$

今の場合，次の出力より，RE は 13% である．

```
tab.4.21.lm.crd <- lm(rep ~ treat, tab.4.21.df)
(summary(tab.4.21.lm.crd)$sig / summary(tab.4.21.lm)$sig)^2
```

CRD が BIB と同じ精度を持つには，観測値を 13%多く取る必要があることになる．

ブロック間の情報を復元するには，R 内でトリッキーな操作が必要である．最適な結合推定量は， $\tau_i^* = \alpha_1 \hat{\tau}_i + \alpha_2 \tilde{\tau}_i$ として表現することができる．これはブロック内とブロック間の推定量の線形結合であり，その重みは $\hat{\tau}_i$ の分散の逆数に比例する．次式により，推定することができる．

$$\tau_i^* = \frac{kQ_i(\sigma^2 + k\sigma_\beta^2) + \left(\sum_{j=1}^b n_{ij}y_{.j} - kr\bar{y}_{..}\right)\sigma^2}{(r - \lambda)\sigma^2 + \lambda a(\sigma^2 + k\sigma_\beta^2)} \quad i = 1, 2, \dots, a \quad (4.4)$$

σ^2 と σ_β^2 は未知のパラメータなので，それらの推定値で置き換える．ブロック内分析から得られる誤差 MS を σ^2 とする．

図 4.6 はこの 4×4 計画の観測された応答を示す．最小 2 乗法による適合結果を，各“ランダムな”ブロックに対して重ね合わせている．次のコマンドを用いて，この trellis グラフィックス [2] を作成した．

```
require(lattice)
xyplot(rep ~ treat|block, tab.4.21.df,
  aspect="xy", xlab="触媒", ylab="応答",
  panel=function(x,y) {
    panel.xyplot(x,y)
    panel.lmline(x,y)
  })
```

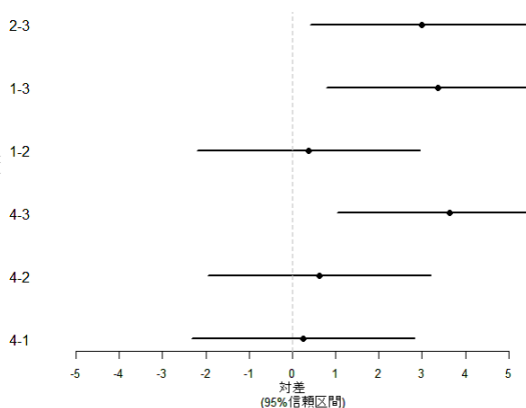


Figure 4.5: 触媒実験．触媒の種類別 (1, 2, 3, 4) の関数として各ブロックで測定された応答．

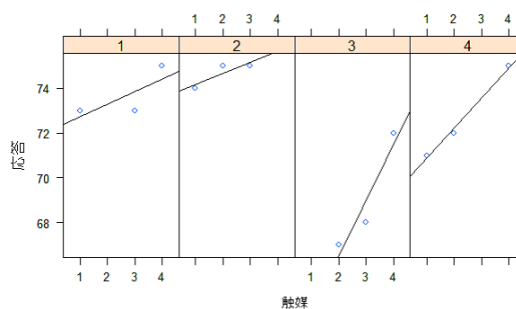


Figure 4.6: Tukey の 95%同時信頼区間

lme4 パッケージを用いても同じ結果を得ることができるが，この場合，REML 推定を行うことになる．

```
require(lme4)
print(tab.4.21.lm <- lmer(rep ~ treat + (1|block), tab.4.21.df), corr=F)
```

このときは、標準誤差とともに分散要素を含むより多くの出力を得る。

```
Linear mixed-effects model fit by REML
Formula: rep ~ treat + (1 | block)
Data: tab.4.21.df
AIC   BIC logLik MLdeviance REMLdeviance
44.22 46.64 -17.11      38.57      34.22
Random effects:
Groups   Name      Variance Std.Dev.
block    (Intercept) 8.00907  2.83003
Residual                    0.65035  0.80644
number of obs: 12, groups: block, 4

Fixed effects:
              Estimate Std. Error t value
(Intercept)  74.9704     1.4963   50.11
treat1       -3.5573     0.6973   -5.10
treat2       -3.3541     0.6973   -4.81
treat3       -2.9704     0.6973   -4.26
```

treat に対して別の線形対比を指定したい場合、前のモデルから定数項を単に削除するだけでよい。

```
print(tab.4.21.lm0 <- lmer(rep ~ -1 + treat + (1|block), tab.4.21.df))
```

上記の出力より、定数項は 74.97 であることがわかるが、これは treat=4 のときの観測値の平均に対応しない。実際には、これは 4 つのブロックに渡る予測平均の平均である。これは、ブロック (block) を変量効果として取り扱ったため ((1|block))、切片が別々に求められたことによる。

```
coef(tab.4.21.lm)[[1]]$'(Intercept)'  
mean(coef(tab.4.21.lm)[[1]][,1])
```

注記 BIB を構成するのは簡単ではない。1 つの方法は、多因子計画から出発することである (関心のある因子が多数あるが、主実験では効果のあるものだけを取り扱いたいという選別段階においてよく利用される)。[3] を参照して、第 1 列 (第 1 因子) を選択し、その要素に円順列を行うことにより $L_{12}(2^{11})$ 計画を構成することができる。次のコードはそのような構成法を簡単に実装するものであるが、最適ではない³。

³並べ替えのためのより効率的なアルゴリズムは、様々な R パッケージ (例えば、vegan, sna, dprep, e1071, gdata, magic) で実行可能である。

```

col <- c(1, 1, 0, 1, 1, 1, 0, 0, 0, 1, 0)
perm <- function(x) {
  s <- length(x)
  m <- matrix(nc=s, nr=s)
  y <- rep(x, 2)
  m[,1] <- x
  for (i in 2:s) { m[,i] <- y[i:(s+i-1)] }
  m
}
col.perm <- perm(col)
bib11 <- rbind(rep(0,11), col.perm)
# 計画がうまく釣り合っていることをチェックする
apply(bib11[-1,], 1, sum)
apply(bib11, 2, sum)

```

この特定の計画において、各ブロックに丁度6つの因子があり、各因子に6つのブロックがあることを確認することができる。読みやすくするために（少なくとも私から見て）、計画行列を表（表 4.7）ではなく図に表すことができる。これは、confusion matrix のように見える。

Figure 4.7: 10 ブロック × 10 因子の BIBD

R を用いたブロック計画の他の例が、[9] (Chapter 16) にある。

第5章 要因計画入門

5.1 第5章の要約

第5章では、釣り合い型2因子計画を取り扱う。適切に用いると、要因配置実験は計画の効率性を高め、1元配置実験を別々に行うよりも少ない実験回数で同じ精度を得ることができる。基本的なANOVAのモデル式は、新たな2番目の因子で説明される変動と、2つの因子間の交互作用によって説明される変動に拡張される。交互作用の概念が第1の関心事であり、科学的・統計的によく理解する必要がある。

5.2 2因子要因計画

一般に、効果モデルは次のようになる。

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk} \quad (5.1)$$

ここで、 i, j ($i = 1 \dots a, j = 1 \dots b$) は因子 A, B の水準で、 k ($k = 1 \dots n$) は観測値番号である。 abn 個のデータを得る順序をランダムにする。そのため、この計画を完全ランダム化計画という。

1つ以上の因子が量的であるとき、回帰モデルはより簡単に定式化できる。上記のモデルの正規方程式を書き下ろすとき、方程式体系には $a + b + 1$ 個の線形依存性が存在する。結果として、パラメータはユニークに決定されず、何らかの制約を課さない限り、モデルを直接推定することはできない。これは次により可能である： $\sum_{i=1}^a \hat{\tau}_i = 0, \sum_{j=1}^b \hat{\beta}_j = 0, \sum_{i=1}^a \hat{\tau}\hat{\beta}_{ij} = 0$ ($j = 1, 2, \dots, b$)、 $\sum_{j=1}^b \hat{\tau}\hat{\beta}_{ij} = 0$ ($i = 1, 2, \dots, a$)

少し計算すると、式5.1は平方和の(修正した)総和として表現することができる。ことがわかる。

$$\begin{aligned}
\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (y_{ijk} - \bar{y} \dots)^2 &= \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n [(\bar{y}_{i..} - \bar{y} \dots) + (\bar{y}_{.j.} - \bar{y} \dots) \\
&\quad + (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y} \dots) + (y_{ijk} - \bar{y}_{ij.})]^2 \\
&= bn \sum_{i=1}^a (\bar{y}_{i..} - \bar{y} \dots)^2 + an \sum_{j=1}^b (\bar{y}_{.j.} - \bar{y} \dots)^2 \\
&\quad + n \sum_{i=1}^a \sum_{j=1}^b (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y} \dots)^2 \\
&\quad + \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (y_{ijk} - \bar{y}_{ij.})^2
\end{aligned} \tag{5.2}$$

この分解を記号で表現すると，

$$SS_T = SS_A + SS_B + SS_{AB} + SS_E \tag{5.3}$$

となる．式 5.2 の右辺の最後の項より，誤差平方和を得るには最低 2 回の繰返し ($n \geq 2$) が必要であることがわかる．1 元配置法と同様，この項は残差項または誤差項という．

仮説検定は次に示す 3 つの手順で進められる．

- 行の処理効果が無い
 $H_0: \tau_1 = \tau_2 = \dots = \tau_a = 0$
- 列の処理効果が無い
 $H_0: \beta_1 = \beta_2 = \dots = \beta_b = 0$
- 行と列の処理に交互作用がない
 $H_0: (\tau\beta)_{ij} = 0$, 全ての i, j に対して

データ battery.txt に適用すると， 3^2 要因配置実験を簡単にセットアップすることができる (3 水準の因子が 2 つ)．データは，温度 (単位: F) と 3 水準の選択可能な計画パラメータから構成される．目的は，温度が極端に変動する機器で用いられるバッテリーを設計することにある．

```

battery <- read.table("battery.txt", header=TRUE)
battery$Material <- as.factor(battery$Material)
battery$Temperature <- as.factor(battery$Temperature)
summary(battery)

```

交互作用効果を含む 2 元配置分散分析モデルは，次により推定することができる．

```
battery.aov <- aov(Life ~ Material * Temperature, data=battery)
```

Life ~ Material * Temperature は,

```
Life ~ Material + Temperature + Material * Temperature
```

と同値である。後者は、各効果を明示的に記すものであり、`Life ~ . + .^2` とすると、データフレームに含まれる全ての因子が 2 次の交互作用までを含むことができる。

結果を表示するには `summary(battery.aov)` を利用する。結果を表 5.1 に示す。3 つの全ての効果は有意であり、特に、温度はバッテリーの寿命の約 50% を説明する。

Table 5.1: 3² バッテリー実験の ANOVA 表

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Material	2	10683.72	5341.86	7.91	0.0020
Temperature	2	39118.72	19559.36	28.97	0.0000
Material:Temperature	4	9613.78	2403.44	3.56	0.0186
Residuals	27	18230.75	675.21		

たいていの場合、応答変数の平均をプロットすることは、ANOVA 表に示された効果に関する洞察を得るのに非常に有益である。図 5.1 では、平均寿命 \bar{y}_{ij} を材料のタイプで層別して Temperature の関数として図示した。このグラフの各点は、4 つの観測地の平均を示す。これを交互作用プロットという。

```
with(battery, interaction.plot(Temperature, Material, Life, type="b",  
                               pch=19, fixed=T, xlab="温度 (F)", ylab="平均寿命"))
```

温度が高くなるにつれて、平均寿命が減少することが分かる。また、材料 3 は他の材料と比べて、特に高い温度のとき寿命が高くなるので、交互作用効果があることもわかる。

他の有益なプロットに、効果プロットがある。これは、関数 `plot.design()` を用いて作成することができるが、関数 `aov()` に与えたのと同じモデル式を引数としてとる。コマンド、

```
plot.design(Life ~ Material * Temperature, data=battery)
```

により図 5.2a を作成することができる。大きな温度効果がバッテリーの寿命の範囲に反映されている。

ここで、第 3 章と同じ手順により、モデルの診断とともに多重比較を行う。これらは基本的に、pp. 17, 19 で記述したものと同一原理に基づく。しかし、モデル診断は、各“処理”に対して行う必要があることに注意（すなわち、各因子水準を組み合わせる）。

このような計画において、テューキ (Tukey) の HSD は研究者により広く利用されている。

コマンド `TukeyHSD(battery.aov, which="Material")` により次の結果を得る。

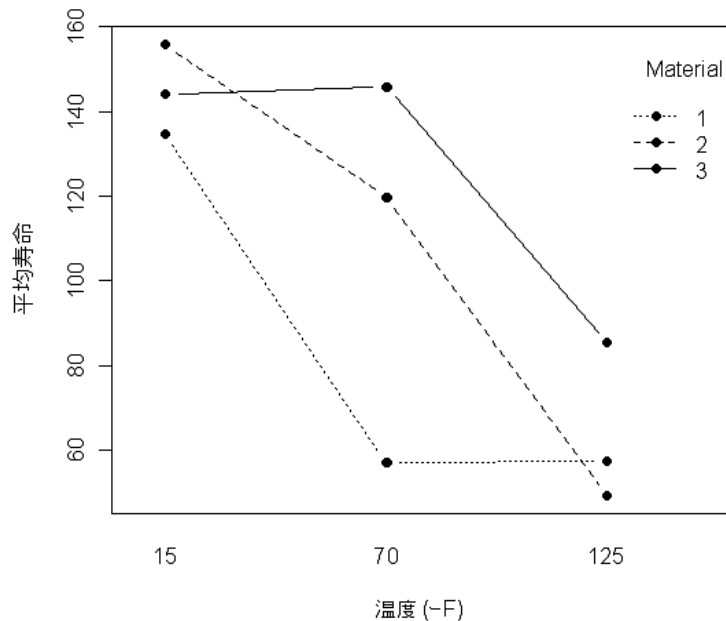


Figure 5.1: 材料の種類-バッテリー寿命試験の温度プロット.

Tukey multiple comparisons of means
95% family-wise confidence level

Fit: aov(formula = Life ~ . + .^2, data = battery)

```
$Material
      diff      lwr      upr      p adj
2-1 25.16667 -1.135677 51.46901 0.0627571
3-1 41.91667 15.614323 68.21901 0.0014162
3-2 16.75000 -9.552344 43.05234 0.2717815
```

交互作用が有意なので、このような計算を行ってはいけない。材料の種類の効果は、どの温度を考えているかによって異なるのである。70°Fでの材料の効果調べたい場合、似たような結果を得るが(Rによる適切な方法がわからなかったため、手で行った)、次に示すのが、交互作用が有意の場合の平均の対比を求める正しい方法である。

```
# 温度 70F で 3 つの平均を計算する
mm <- with(subset(battery, Temperature==70),
            aggregate(Life, list(M=Material), mean))
# 次に、t 分布のパーセント点と誤差 (ANOVA からプールした標準偏差 SD を用いて)
# の積を求める。
val.crit <- qtkey(.95, 3, 27) * sqrt(unlist(summary(battery.aov))["Mean Sq4"])/4
# 最後に、観測された平均の差と限界値とを比較する
diff.mm <- c(d.3.1=mm$x[3]-mm$x[1], d.3.2=mm$x[3]-mm$x[2], d.2.1=mm$x[2]-mm$x[1])
```

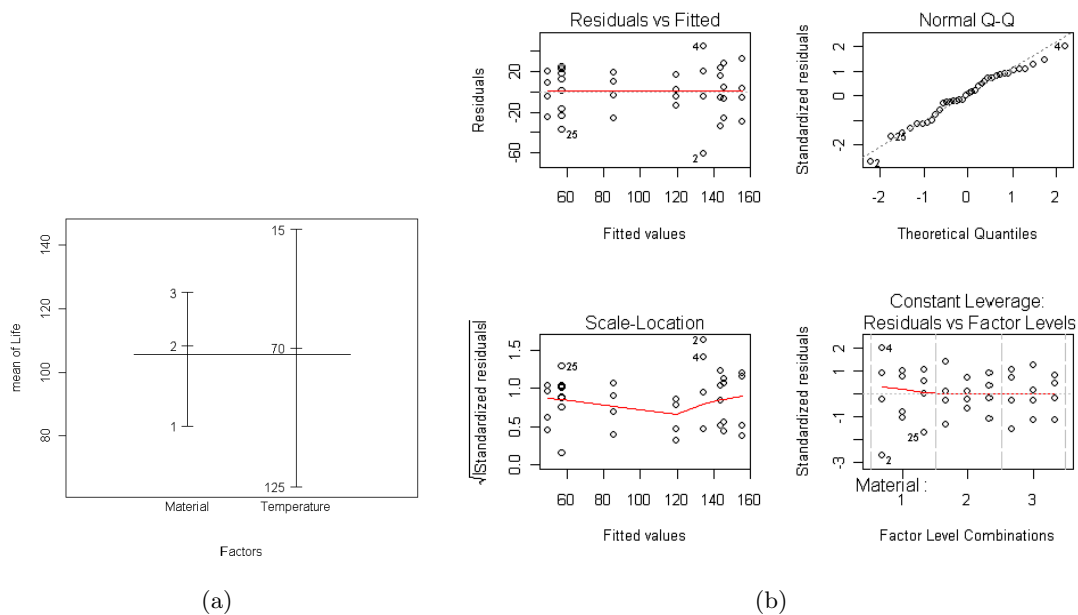


Figure 5.2: (a) 効果図 . (b) 診断プロット .

```
names(which(diff.mm > val.crit))
```

結論として、温度を 70°F に固定したとき、材料 1 と材料 3、材料 2 と材料 1 のみに有意な差がある。

モデルの診断、あるいは残差分析の結果を図 5.2b に示す。この図には、予測値に対する残差または標準化残差のプロット、Q-Q プロット、テコ比とクックの距離のプロットが含まれる。2 因子の要因配置モデルに対する残差の定義は、 $e_{ijk} = y_{ijk} - \hat{y}_{ijk}$ である。 $\hat{y}_{ijk} = \bar{y}_{ij}$ (ij 番目のセルの観測値の平均) なので、残差は

$$e_{ijk} = y_{ijk} - \bar{y}_{ij}. \quad (5.4)$$

と等しい。

予測値と残差の散布図より、大きな分散は大きな予測値に対応していることがわかる。図 5.2b では、2 つの観測値 (2 と 4) がハイライトされている。いいかえると、温度 15°F-材料 1 の組合せは、非常に大きな残差を持ち、これは分散の不均一性を示している。このことは `with(battery, tapply(Life,list(Material,Temperature),var))` といったコマンドを用いて、簡単に確認することができる。

	15	70	125
1	2056.9167	556.9167	721.0000
2	656.2500	160.2500	371.0000
3	674.6667	508.2500	371.6667

交互作用を含まないモデルを（それが適切なら）考えることもできる．これにはモデル式 5.1 の中で項 $(\tau\beta)_{ij}$ を削除すればよい．バッテリー寿命データに適用すると，

```
summary(battery.aov2 <- aov(Life ~ Material + Temperature, data=battery))
```

により次の結果を得る．

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Material	2	10684	5342	5.9472	0.006515 **
Temperature	2	39119	19559	21.7759	1.239e-06 ***
Residuals	31	27845	898		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

2つの主効果は，依然として有意である．しかし，この“縮小”モデルの残差分析を行うと（図 5.3），交互作用がないモデルは不適切であることがわかる．この図は，交互作用がないモデルの予測値 $\hat{y}_{ijk} = \bar{y}_{i..} + \bar{y}_{.j.} - \bar{y}_{...}$ を横軸に， $\bar{y}_{ij.}$ を縦軸とする散布図である．これは，観測されたセル平均と，交互作用がないモデルの推定されたセル平均との差として見ることができる．この図の中のどのパターンも，交互作用の存在を示唆している．

```
par(mar=c(2,5,1,1))
mm2 <- with(battery, tapply(Life, list(Material, Temperature), mean))
mm2 <- as.vector(mm2)
plot(fitted(battery.aov2)[seq(1,36, by=4)], mm2 - fitted(battery.aov2)[seq(1, 36, by=4)],
      xlab="", ylab=expression(bar(y)[ij.] - hat(y)[ijk]),
      pch=19, axes=FALSE, ylim=c(-30,30))
axis(1, at=seq(0, 200, by=50), pos=0)
text(155, 4, expression(hat(y)[ijk]), pos=4)
axis(2, at=seq(-30, 30, by=10), las=1)
```

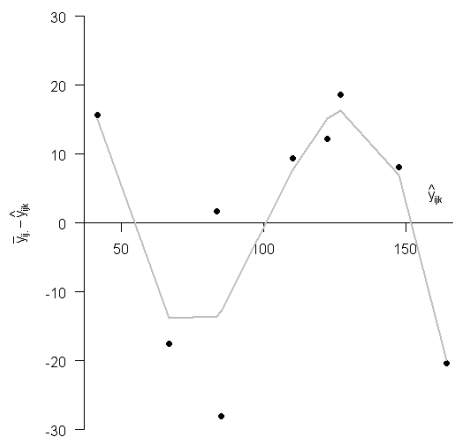


Figure 5.3: $\bar{y}_{ij.}$ 対 \hat{y}_{ijk} のプロット．

説明変数が増加すると予測値 \bar{y}_{ij} の変化のパターンが異なるという明白な傾向がある。このパターンを明瞭にするために、図 5.3 では次のコマンドにより loess 曲線を追記している。

```
yy <- order(fitted(battery.aov2)[seq(1, 36, by=4)])
xx.fit <- fitted(battery.aov2)[seq(1, 36, by=4)]
yy.fit <- mm2 - fitted(battery.aov2)[seq(1, 36, by=4)]
lines(xx.fit[yy], predict(loess(yy.fit[yy] ~ xx.fit[yy])), col="lightgray", lwd=2)
```

ここで、因子の組合せに対して、観測値が 1 つの場合を考える。式 5.2 より、誤差項は推定可能ではないことがわかる。なぜなら、それは交互作用と交絡しているからである。対応する期待平均平方は

$$\sigma^2 + \frac{\sum \sum (\tau\beta)_{ij}^2}{(a-1)(b-1)}.$$

となる。

結果として、交互作用効果がゼロ、すなわち $(\tau\beta)_{ij} = 0$ (全ての i, j に対して) でない限り、主効果に関する検定を行うことはできない。交互作用をプールした後、効果モデルが正しい場合、上記の残差 MS (平均平方) は σ^2 の不偏推定量である。

交互作用が存在するかどうかの検定手法がある。Tukey の非加法性、あるいは曲率検定 [23] は、残差平方和を自由度 1 を持つ非加法性 (交互作用) 効果と自由度 $(a-1)(b-1)-1$ の誤差項に分解する。分子の自由度が 1、分母の自由度が $(a-1)(b-1)-1$ 、有意水準 α の F 検定 $\frac{SS_N}{SS_E/[(a-1)(b-1)-1]}$ により、交互作用がないという帰無仮説を検定することができる。

この手法を、データ impurity.txt を用いて例示する。これは化学製品の研究で、不純物を特性とする 2 因子 (圧力と温度) のデータである。

```
impurity <- read.table("impurity.txt", header=TRUE)
impurity$Temperature <- as.factor(impurity$Temperature)
impurity$Pressure <- as.factor(impurity$Pressure)
```

交互作用項を持つ完全要因配置モデルを適用しても、予想通り主効果の検定は行われな

い。利用するコマンドは簡単である:

```
summary(aov(N ~ Temperature * Pressure, impurity))
```

	Df	Sum Sq	Mean Sq
Temperature	2	23.333	11.667
Pressure	4	11.600	2.900
Temperature:Pressure	8	2.000	0.250

アイデアは、Temperature×Pressure の項の平方和を 2 つの項に分解することにある。自分自身で実行することも可能であるが、a1r3 パッケージに興味深い関数がある。これを用いることにより、存在の可能性のある交互作用を図を用いて検証できるのみならず、Tukey の検定を行うことができる。実際、それはモデル $y \sim A + B$ に 2 次項を追加した検定に相当する。

```
library(alr3)
residualPlots(lm(N ~ Temperature + Pressure, impurity))
```

結果として検定統計量の値は 0.602 となり，その p 値は 0.547 である．Montgomery は，上記の F 検定を行ったが，その値は 0.36 である．これは正確には 0.602^2 である．なぜなら，`residual.plots()` は t 検定の結果を返すからである．この関数は，図 5.4 に示すように（下側左），予測値に対するスチューデント化残差の散布図に推定した 2 次項を点線で追加したものを作成する．

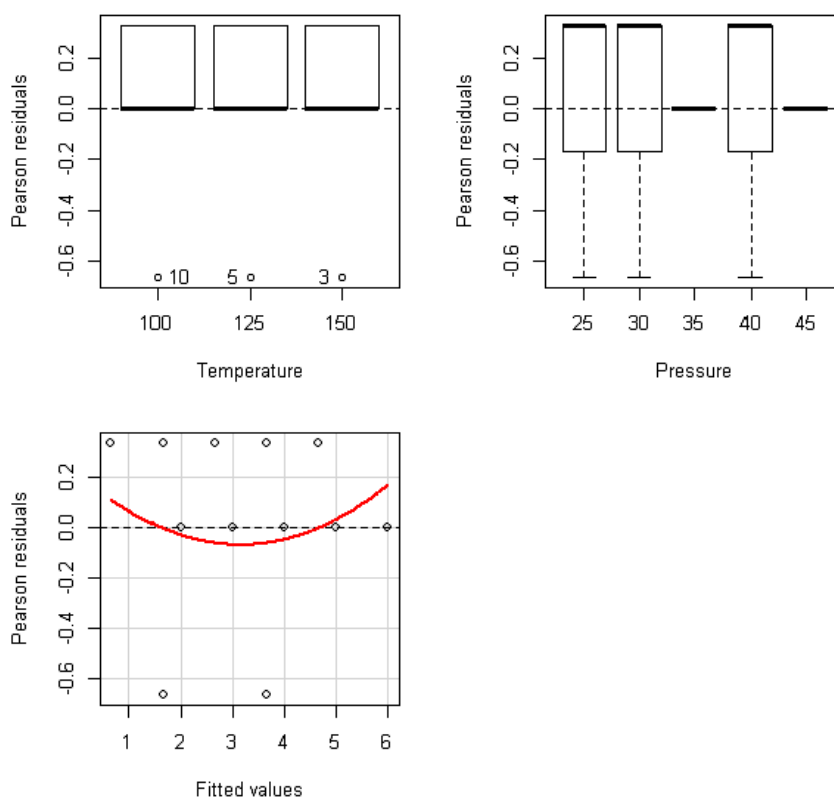


Figure 5.4: 不純物データに対する高次項の検証．

5.3 一般の要因計画，応答曲線・応答曲面

前節で説明したモデルは，任意の数の固定効果，それに応じた 2 因子交互作用，3 因子交互作用を持つモデルへと拡張することができる．

3 因子の計画の例として，ファイル `bottling.txt` のデータを考える．この研究では，製造プロセスで製造されるボトルへのソフトドリンクの充填高さをより一様にすることに関

心がある．充填プロセスにおいて，プロセスエンジニアは3つの制御因子，(A) 炭酸化率，(B) 操作圧，(C) ボトル製造スピード（分当たり），を持つ．モデルは

$$y \sim A + B + C + AB + AC + BC + ABC$$

と書くことができる．ここで， y は応答変数，つまり，充填高さの偏差である．

次の形でデータを読み込む：

```
bottling <- read.table("bottling.txt", header=TRUE,
                      colClasses=c("numeric", rep("factor", 3)))
summary(bottling)
```

結果の要約情報を図 5.5 に示す．これを表示するコマンド次のようにはかなり簡単であるが，もっと簡潔にすることもできるだろう．

```
opar <- par(mfrow=c(2, 2), cex=.8)
boxplot(Deviation ~ ., data=bottling, las=2, cex.axis=.8, ylab="偏差")
abline(h=0, lty=2)
par(las=1)
mm <- with(bottling, tapply(Deviation, Carbonation, mean))
ss <- with(bottling, tapply(Deviation, Carbonation, sd))
bp <- barplot(mm, xlab="炭酸化", ylab="偏差", ylim=c(-2,9))
arrows(bp, mm - ss / sqrt(4), bp, mm + ss / sqrt(4), code=3, angle=90, length=.1)
with(bottling, interaction.plot(Carbonation, Pressure, Deviation, type="b"))
with(bottling, interaction.plot(Carbonation, Speed, Deviation, type="b"))
par(opar)
```

これらの図は，高度に構造化されたデータセットを異なる視点からハイライトするために作成した．上左のパネルはセルで層別したプロットであり，セル当たり $n = 4$ の観測値を持つ3つの因子との組合せの図である．これにはあまり情報は無い．なぜなら我々の主要な関心は，因子自体の効果のみならず，交互作用効果にもあるからである．

上右の図は， B と C の集積データのプロットである．観測値の平均および標準誤差を炭酸化の水準に対してプロットしている．最後に，下部の2つの図は交互作用プロットであるが，同時に2因子のみを表示しているだけなので，依然として，主効果と交互作用の完全な分離はできていない．

3元配置分散分析モデルへの適用結果を，表 5.2 に示す．

```
summary(bottling.aov <- aov(Deviation ~ .^3, bottling))
```

分散分析表より，主効果は全て有意であるが，4つの交互作用は有意ではない．交互作用 $\text{Carbonation} \times \text{Pressure}$ （炭酸化 \times 圧力）は，慣習的な5%水準ではぎりぎり有意ではない．この結果より，交互作用項を削除すべきであり，これは次のステップで実行する（関数 `update()` を用いてこれを実行することができるが，今の場合，全ての交互作用効果を削除するのは面倒である）．

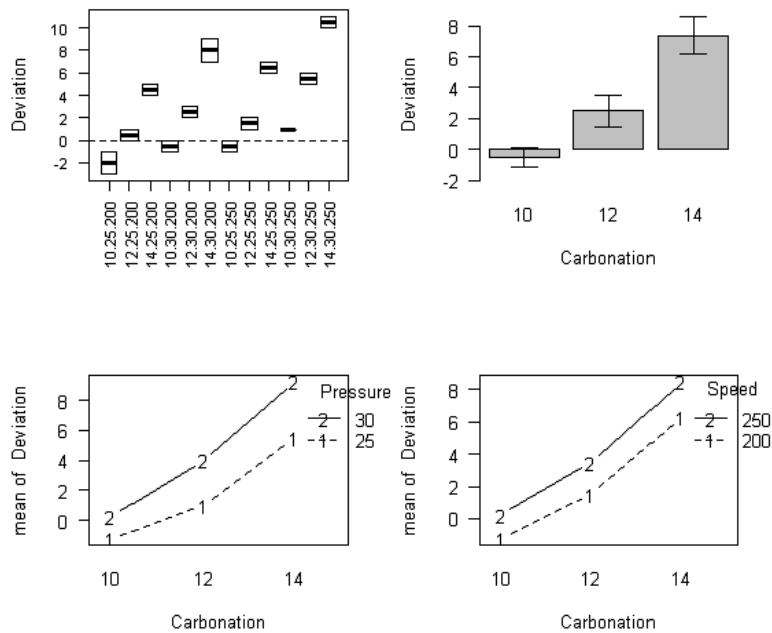


Figure 5.5: ボトル詰めデータセット.

```
bottling.aov2 <- aov(Deviation ~ ., bottling)
anova(bottling.aov2, bottling.aov)
```

表 5.3 に、主効果のみの場合の F 検定に関連する p 値を示す。上記のコードの最後の R コマンドは、モデルに交互作用効果を含むことの是非に関する形式的な検定を行うことができる。 F 検定は有意ではないので (p 値 = 0.225) 交互作用項を削除して良い。図 5.5 に示す交互作用プロットを見ることにより交互作用はないと簡単に判断することもできる。各ケースにおいて、2つの線は厳密に平行であり、要因モデルにおける交互作用の不在を意味する。

バッテリー寿命実験に戻る。量的変数である Temperature の 2 次項を含むとどうなるだろうか。

```
battery$Temperature.num <- as.numeric(as.character(battery$Temperature))
battery.aov3 <- aov(Life ~ Material + Temperature.num + I(Temperature.num^2)
+ Material:Temperature.num+Material:I(Temperature.num^2),
data=battery)
summary(battery.aov3)
```

上記のコマンドで、因子 Temperature の数値を記号に変換するために `as.character` を用いていることに注意。直接 `as.numeric(battery$Temperature)` を用いると、コード 1, 2, 3

Table 5.2: ボトル詰めデータに対する飽和モデルの結果 .

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Carbonation	2	252.75	126.37	178.41	0.0000
Pressure	1	45.37	45.37	64.06	0.0000
Speed	1	22.04	22.04	31.12	0.0001
Carbonation:Pressure	2	5.25	2.63	3.71	0.0558
Carbonation:Speed	2	0.58	0.29	0.41	0.6715
Pressure:Speed	1	1.04	1.04	1.47	0.2486
Carbonation:Pressure:Speed	2	1.08	0.54	0.76	0.4869
Residuals	12	8.50	0.71		

Table 5.3: ボトル詰めデータの縮小モデルの結果 . 有意な有効化のみを示している .

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Carbonation	2	252.75	126.37	145.89	0.0000
Pressure	1	45.37	45.37	52.38	0.0000
Speed	1	22.04	22.04	25.45	0.0001
Residuals	19	16.46	0.87		

を得ることになり、これは私たちが期待するものではない。なぜなら、 y の実際の値を予測したいからである（下記参照）。

今回、モデルを次のように書くことができる。

$$y \sim A + B + B^2 + A : B + A : B^2$$

ここで B は、因子 Temperature を示す。結果の要約情報を表 5.4 に示す。

Table 5.4: バッテリー寿命の気温の 2 次項を追加した当てはめ .

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Material	2	10683.72	5341.86	7.91	0.0020
Temperature.num	1	39042.67	39042.67	57.82	0.0000
I(Temperature.num^2)	1	76.06	76.06	0.11	0.7398
Material:Temperature.num	2	2315.08	1157.54	1.71	0.1991
Material:I(Temperature.num^2)	2	7298.69	3649.35	5.40	0.0106
Residuals	27	18230.75	675.21		

このモデルに基づく予測値を見ると、図 5.6 に表示した結果は、寿命は材料のタイプに

依存しており（つまり，表 5.4 における交互作用効果 $A : B^2$ が有意であること），バッテリー寿命が最大となる最適な温度が存在するという直感と一意していることがわかる．

```
new <- data.frame(Temperature.num=rep(seq(15, 125, by=5), 3), Material=gl(3, 23))
new$fit <- predict(battery.aov3, new)
opar <- par(las=1)
# 最初に予測値をプロットする
with(new, interaction.plot(Temperature.num, Material,fit, legend=FALSE,
                           xlab="温度", ylab="寿命", ylim=c(20, 190)))
txt.leg <- paste("原材料の種類", 1:3)
text(5:7,new$fit[new$Temperature.num==c(45,55,65)] - c(3, 3, -20), txt.leg, pos=1)
# 次に観測値を
points(rep(c(1, 15, 23), each=12), battery$Life, pch=19)
par(opar)
```

平滑線を得るには，温度の全ての定義域における数値因子として取り扱うとき，予測値を補完する必要がある．これは次のコマンドで可能である．

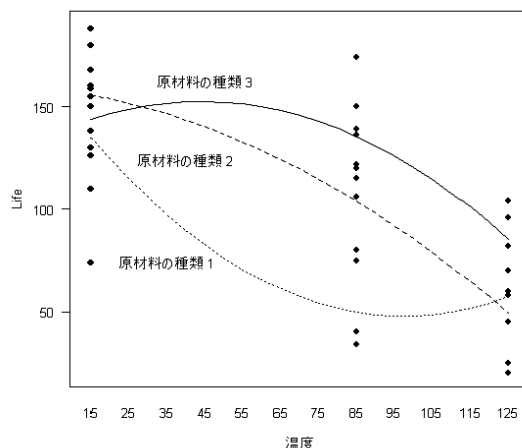


Figure 5.6: 非線形な温度効果を入れた場合のモデルのあてはめ．

2つ以上の数値因子があるとき，応答変数と計画因子との関係を調べるために応答曲面法を利用することもできる．これを `toollife.txt` にあるデータを用いて例示する．切削工具の寿命に対して切削速度（A）とツール角度（B）が与える影響を調べるためのデータである．推定するモデルは次のようになる．

$$y \sim A + B + A^2 + B^2 + A : B^2 + A^2 : B + A : B$$

私たちの主要な関心は2次項にあり，前の例のように，1次効果と2次効果は，階層原理（高次項を保持するときはその低次項も保持すべきである）を保証するために，交互作用効果で交差している．

```

tool <- read.table("toollife.txt", header=TRUE)
tool.lm <- lm(Life ~ Angle * Speed + I(Angle^2) * I(Speed^2) + Angle:I(Speed^2)
             + I(Angle^2):Speed, tool)
tmp.angle <- seq(15, 25, by=.1)
tmp.speed <- seq(125, 175, by=.5)
tmp <- list(Angle=tmp.angle, Speed=tmp.speed)
new <- expand.grid(tmp)
new$fit <- c(predict(tool.lm,new))

```

結果の一部を次に示す．これらは逐次的平方和なので直交している（ 3^2 計画自体の構造から）．

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-1.068e+03	7.022e+02	-1.521	0.1626
Angle	1.363e+02	7.261e+01	1.877	0.0932 .
Speed	1.448e+01	9.503e+00	1.524	0.1619
I(Angle^2)	-4.080e+00	1.810e+00	-2.254	0.0507 .
I(Speed^2)	-4.960e-02	3.164e-02	-1.568	0.1514
Angle:Speed	-1.864e+00	9.827e-01	-1.897	0.0903 .
I(Angle^2):I(Speed^2)	-1.920e-04	8.158e-05	-2.353	0.0431 *
Angle:I(Speed^2)	6.400e-03	3.272e-03	1.956	0.0822 .
Speed:I(Angle^2)	5.600e-02	2.450e-02	2.285	0.0481 *

これらの結果を解釈するよりも，この情報を有効にグラフ化する方法に関心がある．このために，2つの連続的な説明変数に対して補間した予測値を図化する．この図を等高線図という．図 5.7 は，lattice パッケージの関数 `contourplot()` を利用して作成した．

```

require(lattice)
contourplot(fit ~ Angle * Speed,data=new, cuts=8, region=T, col.regions=gray(7:16/16))

```

5.4 要因計画におけるブロック化

これまでは，完全ランダム化要因配置実験を取り扱ってきた．既に第 4 章で議論したように，制御できない攪乱要因が存在するときこの手法は利用できない．そのため，各ブロック内で繰返しのない完全要因実験を行うというアイデアが生じる．このときモデルは，

$$y_{ijk} = \mu + \tau_i + \beta_{ij} + (\tau\beta)_{ij} + \delta_k + \varepsilon_{ijk} \quad (5.5)$$

となる．モデルの記法は，モデル式 5.1 と同じである．

ここで，ブロックの効果を含めているが，ブロックと処理との間の交互作用は無視できると仮定している（RCBD と同じように）．交互作用 $(\tau\delta)_{ik}$ ， $(\beta\delta)_{jk}$ ， $(\tau\beta\delta)_{ijk}$ は誤差と交絡している．

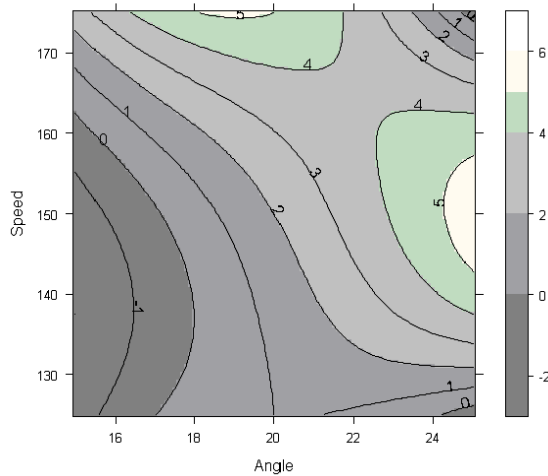


Figure 5.7: 切削ツール研究の等高線図 .

ファイル `intensity.txt` は、レーダースコープがターゲットを検出する能力を改良することを目的とする実験に関するデータである。背景ノイズ(“反射波”),スクリーンのフィルタのタイプという2つの因子を取りあげている。これらは固定効果として取り扱っている。オペレーターの利用可能性と知識程度の違いの基づいて複数のオペレーターを選択し、すべてのランが終了するまで1人のオペレーターをスコープの中に入れておく。このときオペレーターをブロックと考えることができる。だから、各ブロックに 3×2 回の処理組合せをランダムに配置した。データの要約を図 5.8 に示す。

```
intensity <- read.table("intensity.txt", header=TRUE,
                        colClasses=c("numeric", rep("factor", 3)))
require(lattice)
xyplot(Intensity ~ Ground|Operator, data=intensity, groups=Filter,
       panel=function(x, y,...){
         subs <- list(...)$subscripts
         panel.xyplot(x, y, pch=c(1,19),...)
         panel.superpose(x, y, panel.groups="panel.lmline", lty=c(1,2),...)
       }, key=list(text=list(lab=as.character(1:2)), lines=list(lty=1:2, col=1),
                  corner=c(1, .95), title="フィルタのタイプ", cex.title=.8), col=1)
```

応答のパターンを視覚的によりよく評価するために、フィルタのタイプとオペレータに対して最小2乗直線を当てはめている。図から分かるように、傾きと切片とはかなり変動していることがわかる。このように、ブロック化はオペレータの内在する効果を明らかにする良い方法である。

ANOVA モデルを適用するには、第4章で説明した手順に従い、`Error()` オプションを用いてブロックを別の階層として指定する必要がある。

```
intensity.aov <- aov(Intensity ~ Ground * Filter + Error(Operator), intensity)
```

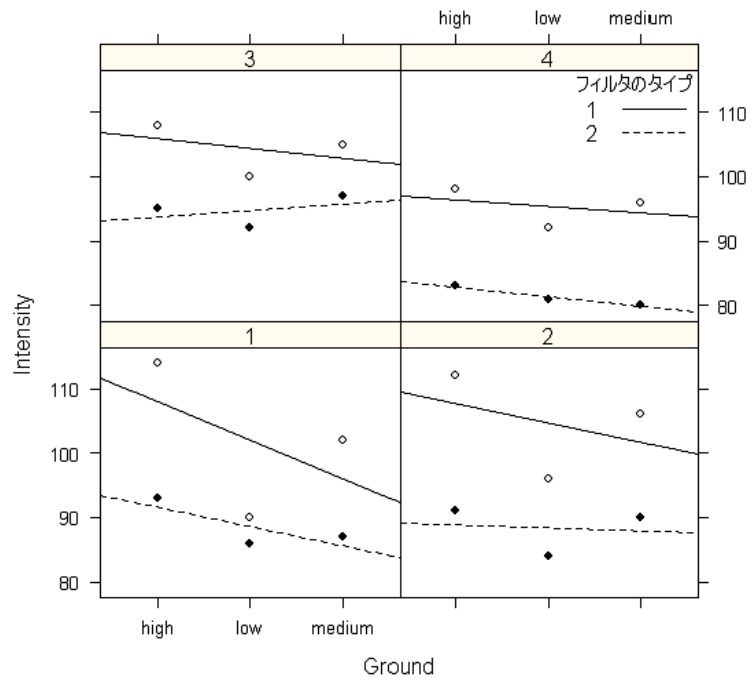


Figure 5.8: intensity データ.

```
summary(intensity.aov)
```

表 5.5 より, `summary()` コマンドでは, `R` はブロック効果を表示しないことが分かる. ブロックの平方和を得るには, `R` のコマンドプロンプトで `intensity.aov` を用いて, 直接 `aov` オブジェクトを呼ぶ必要がある.

Call:

```
aov(formula = Intensity ~ Ground * Filter + Error(Operator),
     data = intensity)
```

Grand Mean: 94.91667

Stratum 1: Operator

Terms:

	Residuals
Sum of Squares	402.1667
Deg. of Freedom	3

Residual standard error: 11.57824

(...)

注目すべきは、ブロック因子の平方和が、主効果の平方和（Ground の 335.58）または誤差項（166.33）と比べてかなり大きいことである。このことは、図 5.8 より応答変数に関して個人間の変動が大きいという直感的に把握した内容が正しいことを裏づける。ブロック因子の平方和の計算は、式 5.5 を用いて行うが、次から求めることもできる。

$$SS_{\text{blocks}} = \frac{1}{ab} \sum_{k=1}^n y_{..k}^2 - \frac{y_{...}^2}{abn} \quad (5.6)$$

Table 5.5: intensity データに適用した ANOVA モデルの結果。

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Residuals	3	402.17	134.06		
Ground	2	335.58	167.79	15.13	0.0003
Filter	1	1066.67	1066.67	96.19	0.0000
Ground:Filter	2	77.08	38.54	3.48	0.0575
Residuals	15	166.33	11.09		

第6章 2^k 要因実験

6.1 第6章の要約

6.2 2^2 計画

2^2 計画は、より一般的な 2^k 計画の中で最も簡単な計画である。2つの因子 A, B (“低い”と“高い”という2つの水準を持つ) を考える。繰返しを k 回とすると、 $2 \times 2 \times n$ 回の試行を行う完全ランダムイズ実験となる。

化学プロセスにおける収量に対する反応剤の濃度と触媒量の効果を調べたい。目的は、反応剤の濃度(15%または25%)と触媒量(1ポンドまたは2ポンド)が収量にどのように影響を与えるかを調べることにある(データファイルは yield.txt)。本実験における異なる処理の組合せの結果を図6.1に示す。ここで、“+”は高水準を、“-”は低水準を示す。

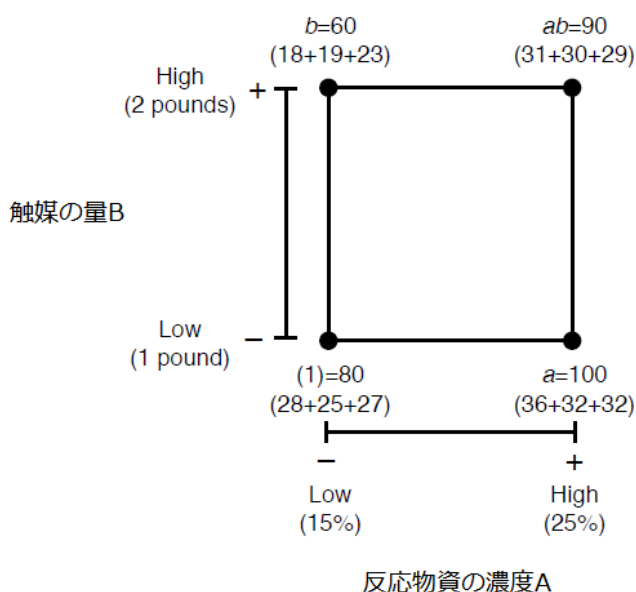


Figure 6.1: 2^2 計画における処理の組合せ。

因子の平均効果は、その因子の水準を変化させたことによる応答の、他の因子の水準全

体での平均変化として定義される。前の図では、記号 (1), a , b , ab が処理の組合せにおける n 回の繰返し全ての合計を表している。 B の低水準における A の効果を, $[a - (1)]/n$ と表し, B の高水準における A の効果を $[ab - b]/n$ と表す。これらの 2 つの量の平均が A の主効果である。

$$\begin{aligned} A &= \frac{1}{2n} \{[ab - b] + [a - (1)]\} \\ &= \frac{1}{2n} \{ab + a - b - (1)\}. \end{aligned} \quad (6.1)$$

B に対しても同様である。

$$\begin{aligned} B &= \frac{1}{2n} \{[ab - a] + [b - (1)]\} \\ &= \frac{1}{2n} \{ab - a + b - (1)\}; \end{aligned} \quad (6.2)$$

交互作用効果 AB を, B の高水準における A の効果と, B の低水準における A の効果の差で定義する。

$$\begin{aligned} AB &= \frac{1}{2n} \{[ab - b] - [a - (1)]\} \\ &= \frac{1}{2n} \{ab - a - b + (1)\}. \end{aligned} \quad (6.3)$$

他方, A の効果を次のようにして求めることもできる。

$$\begin{aligned} A &= \bar{y}_{A+} - \bar{y}_{A-} \\ &= \frac{ab + a}{2n} - \frac{b + (1)}{2n} \\ &= \frac{1}{2n} \{ab + a - b - (1)\}. \end{aligned} \quad (6.4)$$

これは式 6.1 と一致している。同じことが, B と AB にも適用できる。データで計算すると,

$$A = \frac{1}{2 \times 3} 90 + 100 - 60 - 80 = 8.33,$$

となる。

R を用いて計算するには次のようにする。

```
yield <- read.table("yield.txt", header=T)
attach(yield)
rm(yield)
yield.sums <- aggregate(yield, list(reactant=reactant, catalyst=catalyst), sum)
yield.sums
```

これが必要な情報全てを与える：

reactant	catalyst	x
1	high	high 90
2	low	high 60
3	high	low 100
4	low	low 80

A は正の効果を持つ。つまり，A を低水準（15%）から高水準（25%）に増加させると，収量も増える。B に関しては逆であるが，交互作用効果はあったとしてもわずかである。

分散分析モデルと対比の構成 分散分析により，因子の効果の方向と大きさを推定することができる。既に示したように，対比 $ab + a - b - (1)$ （式 6.1）を用いて A を推定することができる。この対比を，A の全効果という。上で導出した 3 つの対比は全て直交している。

任意の対比の平方和は，データ数と対比の係数の 2 乗の和を掛けたもので対比の 2 乗を割ったものに一致する。だから，

$$\left\{ \begin{array}{l} SS_A = \frac{[ab + a - b - (1)]^2}{4n} \\ SS_B = \frac{[ab - a + b - (1)]^2}{4n} \\ SS_{AB} = \frac{[ab - a - b + (1)]^2}{4n} \end{array} \right.$$

となる。

総平方和 SS_T は自由度 $4n - 1$ を持ち，通常の方法で計算できる。つまり，

$$SS_T = \sum_{i=1}^2 \sum_{j=1}^2 \sum_{k=1}^2 y_{ijk}^2 - \frac{y_{\dots}^2}{4n}. \quad (6.5)$$

これに対し，誤差平方和 SS_E は自由度 $4(n-1)$ を持ち， $SS_E = SS_T - SS_A - SS_B - SS_{AB}$ となる。処理の組合せは，次表のように示すことができる。

Effects	(1)	a	b	ab
A	-1	+1	-1	+1
B	-1	-1	+1	+1
AB	+1	-1	-1	+1

この順序を Yates の順序という。全ての対比は直交しているので， 2^2 （そして全ての 2^k ）計画は直交計画である。

`summary(aov(yield ~ reactant * catalyst))`

表 6.1 より，主効果は有意であるが，交互作用 A : B は有意ではない。

Frank Yates (1902.1994) は標本調査法と分析の研究を行った。要因実験の計画と分析に関する本を出版している。

Table 6.1: 収量実験に対する分散分析表.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
反応剤	1	208.33	208.33	53.19	0.0001
触媒	1	75.00	75.00	19.15	0.0024
反応剤:触媒	1	8.33	8.33	2.13	0.1828
Residuals	8	31.33	3.92		

回帰モデル 回帰モデルから推定された係数を用いて，次の予測式を設定できる．

$$\hat{y} = 18.333 + 0.833x_{\text{反応剤}} - 5.000x_{\text{触媒}},$$

ここで， $x_{\text{反応剤}}$ と $x_{\text{触媒}}$ については，これらの因子が取る実際の値を用いて \hat{y} を計算する．つまり，因子水準はそれに対応する数値として取り扱う（触媒量に対しては1または2，反応剤に対しては15または25）．これに対して分散分析表は，その水準にどんな値を与えようとも変わらない．しかし，モデルのパラメータは測定の単位に依存する．次のR スクリプトでは，2値変数を，適切な値を持つ順序変数に変換している．少しトリッキーな操作により，水準をそれらに対応する数値に正確に変換している．

```
reactant.num <- reactant
levels(reactant.num) <- c(25, 15)
reactant.num <- as.numeric(as.character(reactant.num))
catalyst.num <- catalyst
levels(catalyst.num) <- c(2, 1)
catalyst.num <- as.numeric(as.character(catalyst.num))
yield.lm <- lm(yield ~ reactant.num + catalyst.num)
yield.lm # 線型モデルLMの係数を与える
```

図 6.2 は，反応剤と触媒の関数として予測値をプロットした応答曲面と等高線図である．主効果のみを含む1次モデルを扱っているので，応答曲面は平面である．このような応答曲面は，プロセスの潜在的な改良の方向を発見するのに用いられる．このためによく用いられる手法が，第 11 章で説明する最急降下法である．

```
library(scatterplot3d)
s3d <- scatterplot3d(reactant.num, catalyst.num, yield, type="n",
angle=135, scale.y=1, xlab="反応剤", ylab="触媒", zlab="収量")
s3d$plane3d(yield.lm,lty.box="solid", col="darkgray")
tmp <- list(reactant.num=seq(15, 25, by=.5), catalyst.num=seq(1, 2, by=.1))
new.data <- expand.grid(tmp)
new.data$fit <- predict(yield.lm, new.data)
contourplot(fit ~ reactant.num + catalyst.num,new.data, xlab="反応剤",
ylab="触媒")
```

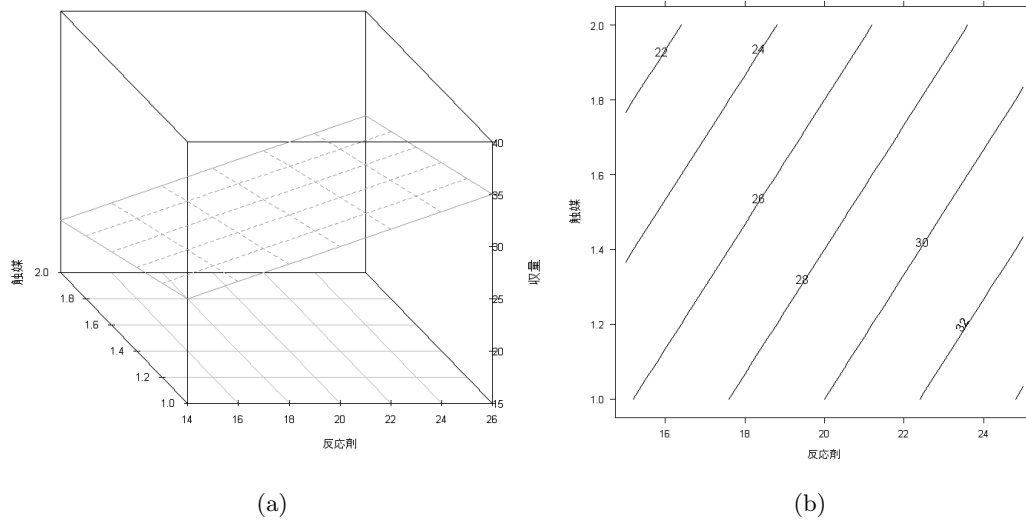


Figure 6.2: (a) 応答曲面 . (b) 等高線図 .

6.3 2^3 計画

2^k 計画で3つの因子 A, B, C を考えるとき, 2^3 要因計画となる. このとき, 処理の組合せは8つあり, これを立方体で表示することができる (図 6.3). これを計画行列という. 2^3 計画の8つの処理組合せは自由度7を持つ: 自由度3は, 3つの主効果が持ち, 自由度4は, 交互作用が持つ (3つの2因子交互作用と1つの3因子交互作用).

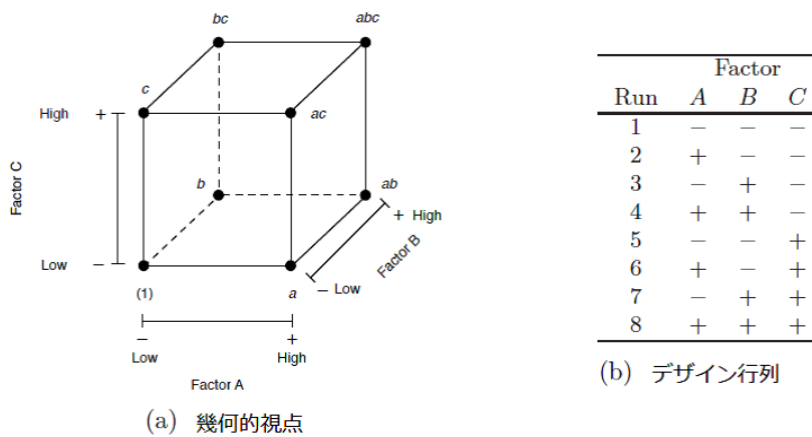


Figure 6.3: 2^3 計画における処理の組合せ .

B と C が低水準のとき, A の効果は $[a - (1)]/n$ である. B が高水準で C が低水準のとき, A の効果は $[ab - b]/n$ である. C が高水準で B が低水準のとき, A の効果は $[ac - c]/n$

である．最後に， B と C が高水準のとき， A の効果は $[abc - bc]/n$ である．だから， A の平均効果は次のようになる．

$$A = \frac{1}{4n} [a - (1) + ab - b + ac - c + abc - bc] \quad (6.6)$$

これは，図 fig:fig.6.4 の右側面における4つの処理の組合せ間の対比である． A の効果は， A が高水準にある4回のランの平均から A が低水準(\bar{y}_{A-})4回のランの平均を引いたものである．式で表すと次のようになる．

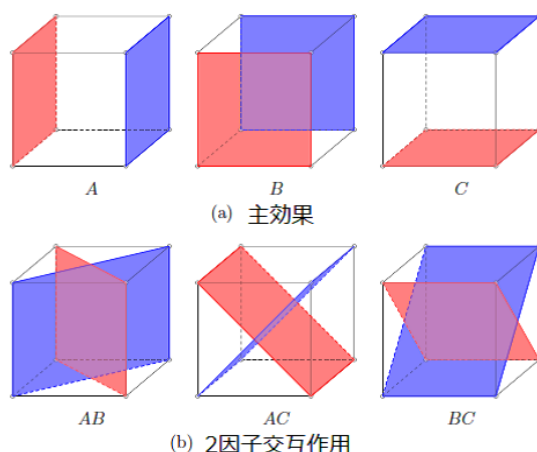


Figure 6.4: 対比の幾何的表現．どの図も高水準は青で，低水準は赤で表示している．

$$\begin{aligned} A &= \bar{y}_{A+} - \bar{y}_{A-} \\ &= \frac{a + ab + ac + abc}{4n} - \frac{(1) + b + c + bc}{4n} \end{aligned} \quad (6.7)$$

2因子交互作用効果も同様に求めることができる．なぜなら交互作用は，1つの因子の，もう1つの因子の2つの水準に対する平均の差だからである．便宜的に，この差の半分を交互作用 $A \times B$ の効果という．交互作用 $A \times B$ の効果 AB は，

$$AB = \frac{abc - bc + ab - b - ac + c - a + (1)}{4n} \quad (6.8)$$

あるいは

$$AB = \frac{abc + ab + c + (1)}{4n} - \frac{abc + ab + c + (1)}{4n} \quad (6.9)$$

となる。

最後に、交互作用 $A \times B \times C$ の効果 ABC は、 C の異なる 2 つの水準に対する交互作用 $A \times B$ の差、つまり次式のようになる。

$$\begin{aligned}
 ABC &= \frac{1}{4n} \{ [abc - bc] - [ac - c] - [ab - b] + [a - (1)] \} & (6.10) \\
 &= \frac{1}{4n} \{ abc - bc - ac + c - ab + b + a - (1) \}.
 \end{aligned}$$

これまでの全ての方程式は、エッチング効果の推定に関連した対比を示している。これらの対比から、表 6.2 に示すように、“+”（高水準）と“-”（低水準）の表を展開することができる。主効果の符号が決定すると、残りの効果は、適切な列の行単位での積により得ることができる。表 6.2 は、興味深い特徴をいくつか持つ。(1) I 列を除いた全ての列で、プラス記号とマイナス記号の数は等しい。(2) 任意の列の行単位での積の和はゼロになる（直交性による）。(3) I は単位元なので、これに他の列を掛けても他の列は変化しない。(4) 任意の列の積は、表の中にある。例えば、 $A \times B = AB$ であり、 $AB \times B = AB^2 = A$ である。

最後に、効果の平方和は、次のように定義される。

$$SS = \frac{(\text{Contrast})^2}{8n}. \quad (6.11)$$

Table 6.2: 2^3 計画における効果計算の符号。

処理の組合せ	I	A	B	AB	C	AC	BC	ABC
(1)	+	-	-	+	-	+	+	-
a	+	+	-	-	-	-	+	+
b	+	-	+	-	-	+	-	+
ab	+	+	+	+	-	-	-	-
c	+	-	-	+	+	-	-	+
ac	+	+	-	-	+	+	-	-
bc	+	-	+	-	+	-	+	-
abc	+	+	+	+	+	+	+	+

プラズマエッチング実験 (plasma.txt) は、 2^3 計画であり、枚様式プラズマエッチング装置における窒化エッチング工程を開発するために実施された。因子は電極ギャップ、ガスフロー（反応ガスとして C_2F_6 を使用）、光電陰極に与える RF 電力である。各因子の水準は 2 で、実験を 2 回繰り返した。データファイルの読み込みは、通常の形で行うことができる。

```
plasma <- read.table("plasma.txt", header=TRUE)
plasma
```

データを R のデータフレームに変換する前に，データを次に示す．

```
Run  A  B  C  R1  R2
1  1  -1 -1 -1  550  604
2  2   1 -1 -1  669  650
3  3  -1  1 -1  633  601
4  4   1  1 -1  642  635
5  5  -1 -1  1 1037 1052
6  6   1 -1  1  749  868
7  7  -1  1  1 1075 1063
8  8   1  1  1  729  860
```

ANOVA モデルを適用する前に，次のようにして R の “long” フォーマットに変換する必要がある．

```
plasma.df <- data.frame(etch=c(plasma$R1, plasma$R2),
  rbind(plasma[,2:4], plasma[,2:4]))
plasma.df[,2:4] <- lapply(plasma.df[,2:4], factor)
```

次に，`aov()` 関数を用いて，全ての主効果と交互作用を求める（表 6.3）．

```
plasma.df.aov <- aov(etch ~ A * B * C, data=plasma.df)
summary(plasma.df.aov)
```

表 6.3 より，主効果 A, C ，交互作用 $A \times C$ が有意である．通常の有意水準 5% では，他のどの効果も有意ではない．各効果の寄与率は，対応する平方和を計算することにより簡単に求めることができる．

Table 6.3: プラズマエッチング実験に対する分散分析表．

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
A	1	41310.56	41310.56	18.34	0.0027
B	1	217.56	217.56	0.10	0.7639
C	1	374850.06	374850.06	166.41	0.0000
A:B	1	2475.06	2475.06	1.10	0.3252
A:C	1	94402.56	94402.56	41.91	0.0002
B:C	1	18.06	18.06	0.01	0.9308
A:B:C	1	126.56	126.56	0.06	0.8186
Residuals	8	18020.50	2252.56		

第7章 2^k 要因実験におけるブロック化と交絡化

第8章 2水準系直交表

第9章 3水準系，混合系直交表

第10章 回帰モデルへのあてはめ

第11章 応答曲面法と設計

第12章 ロバストパラメータ設計とロバスト プロセス研究

第13章 変量因子の実験

第14章 枝分かれ法と分割法

第15章 計画と分析に関する他のトピック

標準正規分布とスチューデントの t 分布

Table 1: 標準正規分布 $\mathcal{N}(0, 1)$ の累積分布関数 .

Table 2: 標準正規分布 $\mathcal{N}(0, 1)$ の 100β パーセント点 $u(\beta)$.

Table 3: スチューデントの t 分布 . Quantiles $u(\beta)$ of the order β from $\mathcal{T}(\nu)$.

目次

2.1	引っ張り強度の点グラフ(上図)と200個の正規乱数のヒストグラム(下図)	4
2.2	ヒストグラムで用いたのと同じ200個の正規乱数の密度推定	4
2.3	ポルトランドセメント引っ張り強度実験データの箱ひげ図	6
2.4	正規分布の密度関数	7
2.5	第1種の誤りの確率 α と第2種の誤りの確率 β	8
2.6	強度の検証実験	11
3.1	エッチング速度データ	14
3.2	ANOVAモデルのチェック	18
3.3	分散分析モデル(2)のチェック	19
3.4	エッチング速度データに対する1元配置分散分析のHDS検定の信頼区間	21
3.5	1元配置分散分析に対するOC曲線の例(エッチング速度の例1)	24
4.1	代用血管実験の結果	27
4.2	代用血管実験の結果(続き)	28
4.3	代用血管データのモデルチェック	29
4.4	効果プロット	31
4.5	触媒実験. 触媒の種類別(1, 2, 3, 4)の関数として各ブロックで測定された 応答	35
4.6	Tukeyの95%同時信頼区間	35
4.7	10ブロック×10因子のBIBD	37
5.1	材料の種類-バッテリー寿命試験の温度プロット	41
5.2	(a) 効果図. (b) 診断プロット	42
5.3	\bar{y}_{ij} 対 \hat{y}_{ijk} のプロット	43
5.4	不純物データに対する高次項の検証	45
5.5	ボトル詰めデータセット	47
5.6	非線形な温度効果を入れた場合のモデルのあてはめ	49
5.7	切削ツール研究の等高線図	51
5.8	intensity データ	52
6.1	2^3 計画における処理の組合せ	54

6.2	(a) 応答曲面 . (b) 等高線図	58
6.3	2^3 計画における処理の組合せ	58
6.4	対比の幾何的表現 . どの図も高水準は青で , 低水準は赤で表示している . .	59

目 次

3.1	エッチング速度実験に対する分散分析モデルの結果	15
3.2	Bonferroni の方法	20
3.3	Hochberg の方法	20
3.4	Tukey の HSD 法	20
3.5	エッチング速度データに対するノンパラメトリックな多重比較の適用 LB/UP-CI: 95%信頼区間の下限と上限; p 値: 片側, 両側 p 値	25
4.1	モデル $y = \mu + \text{PSI}_i + \text{block}_j$ に対する結果	28
4.2	ロケットデータの分析	31
4.3	BIB 分析の要約	34
5.1	3^2 バッテリー実験の ANOVA 表	40
5.2	ボトル詰めデータに対する飽和モデルの結果	48
5.3	ボトル詰めデータの縮小モデルの結果. 有意な有効化のみを示している	48
5.4	バッテリー寿命の気温の 2 次項を追加した当てはめ	48
5.5	intensity データに適用した ANOVA モデルの結果	53
6.1	収量実験に対する分散分析表	57
6.2	2^3 計画における効果計算の符号	60
6.3	プラズマエッチング実験に対する分散分析表	61
1	標準正規分布 $\mathcal{N}(0, 1)$ の累積分布関数	71
2	標準正規分布 $\mathcal{N}(0, 1)$ の 100β パーセント点 $u(\beta)$	72
3	スチューデントの t 分布. Quantiles $u(\beta)$ of the order β from $\mathcal{T}(\nu)$	73

関連図書

1. J-M Azaïs and J-M Bardet. *Le modèle linéaire par l'exemple*. Dunod, 2005.
2. R. A. Becker, W. S. Cleveland, M. J. Shyu, and S. Kaluzny. A tour of trellis display. Technical report, Bell Labs, 1995. Also available at <http://cm.bell-labs.com/cm/ms/departments/sia/wsc/webpapers.html>.
3. D. Benoist. *Plans d'Expériences : Construction et Analyse*. Lavoisier, 1994.
4. P. D. Berger and R. E. Maurer. *Experimental Design with Applications in Management, Engineering and the Sciences*. Duxbury Press, 2002.
5. J. M. Bland and D. G. Altman. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, pages 307-310, 1986.
6. R. Christensen. *Plane Answers to Complex Questions. The Theory of Linear Models*. New York: Springer, 3rd edition, 2002.
7. P. Dagnelie. *Principes d'expérimentation. Planification des expériences et analyses de leurs résultats*. Gembloux, Presses agronomiques, 2003.
8. S. Dudoit and M. J. van der Laan. *Multiple Testing Procedures with Applications to Genomics*. New York: Springer, 2008.
9. J. J. Faraway. *Linear Models with R*. Chapman & Hall/CRC, 2005.
10. J. J. Faraway. *Extending the linear Model with R. Generalized Linear, Mixed Effects and Nonparametric Regression Models*. Chapman & Hall/CRC, 2006.
11. R. A. Fisher. *Statistical Methods, Experimental Design, and Scientific Inferences*. Oxford: Oxford University Press, 1990.
12. J. E. Gentle. *Random Number Generation and Monte Carlo Methods*. New York: Springer, 2003.
13. J. Goupy. *Pratiquer les plans d'expérience*. Genève, Ellipses, 2004.

14. K. Hinkelmann and O. Kempthorne. *Design and Analysis of Experiments. Volume 2: Advanced Experimental Design*. John Wiley & Sons, Inc., 2005.
15. Julian Faraway July. *Practical regression and anova using R*, 2002.
16. B. Lecoutre, M-P. Lecoutre, and J. Poitevineau. Uses, abuses and misuses of significance tests in the scientific community: won't the bayesian choice be unavoidable? *International Statistical Review*, 69:399-418, 2001.
17. M-P. Lecoutre, J. Poitevineau, and B Lecoutre. Even statisticians are not immune to misinterpretations of null hypothesis significance tests. *International Journal of Psychology*, 38:37-45, 2003.
18. U. Munzel and L. A. Hothorn. A unified approach to simultaneous rank test procedures in the unbalanced one-way layout. *Biometrical Journal*, 43(5):553-569, 2001.
19. K. R. Murphy and B. Myers. *Statistical Power Analysis: A Simple and General Model for Traditional and Modern Hypothesis Tests*. Mahwah, NJ: Lawrence Erlbaum Associates, 2nd edition, 2004.
20. J. C. Pinheiro and D. M. Bates. *Mixed-Effects Models in S and S-PLUS*. New York: Springer, 2000.
21. F. W. Satterthwaite. An approximate distribution of estimates of variance components. *Biometrics Bulletin*, 2:110-114, 1946.
22. L. E. Toothaker. *Multiple Comparison Procedures*. SAGE Publications, 1993.
23. J. W. Tukey. Comparing individual means in the analysis of variance. *Biometrics*, 5:99-114, 1949.
24. J. W. Tukey. One degree of freedom for non-additivity. *Biometrics*, 5:232-242, 1949.
25. W. J. van der Linden, B. P. Veldkamp, and J E Carlson. Optimizing balanced incomplete block designs for educational assessments. *Applied Psychological Measurement*, 28:317-331, 2004.
26. W. N. Venables and B. D. Ripley. *Modern Applied Statistics with S*. New York: Springer, 4th edition, 2002.
27. B. L. Welch. The generalization of "student's" problem when several different population variances are involved. *Biometrika*, 34:28-35, 1947.

28. B. L. Welch. On the comparison of several mean values: an alternative approach.
Biometrika, 38:330-336, 1951.