1. 平均を比較する (一元配置 ANOVA)

この章では心理統計の中で最も広く使われているツールの一つ，“分散分析” として知られている手法を紹介します。普通は ANOVA と呼ばれています。基本的な技術はロナルド・フィッシャー卿によって 20 世紀初頭に開発されたもので，不便な用語を使っているのも彼によるところです。ANOVA という用語は，二つの意味で若干ミスリーディングです。第一に，この技術名は分散といっていますが，ANOVA は平均の違いを検証することに興味があるのです。第二に，いくつかの異なる手法が全て ANOVA として引用されていますが，中には関係が薄いものもあるのです。この本の後の方では，全く違う状況に適用される異なる手法の ANOVA に出会すことになりますが，この章の目的は最も単純な形式の ANOVA についてだけ考えることにします。ここでは異なる群について観測しており，これらの群の違いが関心のある結果変数において差があるかどうかに興味があります。これは**一元配置 ANOVA** と呼ばれる問いの立て方になります。

この章の構造は次のようになっています。：セクション **??** ではこの章を通じて例として利用することになる，架空データについて紹介します。データの導入が終われば，一元配置 ANOVA が実際どのように働くのか，そのメカニクスを記述し (セクション **??**) ，それから JASP でどのように実行するかに注目していきます (セクション **??**)。この二つのセクションが，この章の中核となります。この章の残りの箇所では，ANOVA を実行するときには避けられない重要なトピックスの範囲について議論します。例えば効果量をどう計算するのかとか (セクション **??**)，事後検定や多重比較の補正 (セクション **??**)，ANOVA が依存している仮定 (セクション **??**) などです．私たちはまた，これらの仮定をチェックする方法や，もし仮定が満たされなかったら何ができるのかについても論じます (セクション **??** から **??**)。それから，反復測定 ANOVA についてセクションもカバーしていきます (セクション **??** から **??**)．この章の終わりには，ANOVA とほかの統計的ツールとの関係についても少し紹介します (Section **??**)。

1.1

データセットについて

あなたが *Joyzepam* と呼ばれる新しい抗うつ剤をテストする臨床試験に参加するとしましょう。薬の性能を公平に検証するために，この試験では 3 つの異なる薬を含めて検証することにします。一つはプラセボ，もう一つは既に抗うつ/抗不安剤として知られている *Anxifree* です。最初のテストでは，抑うつを抑えるのに変化があるのかを検証するために 18 人の参加者が集められました。薬は心理学的セラピーと一緒に用いられることがありますから，あなたの研究では 9 人が認知行動療法 (CBT) を受けていて，残りの 9 人は受けていないという状況にあります。参加者の処遇はランダムに割り当てられているので (もちろん二重盲検法で，です)，3 種類の薬それぞれについて，3 人が CBT をうけ，3 人がセラピーを受けていない人ということになります。心理学者は各薬の処方後 3 ヶ月たったあとで，個々人の気分を査定し，個々人の気分が全体的に改善されたかどうかを −5 点から +5 点の尺度で評定してもらいました。この研究デザインのデータファイルを `clinicaltrial.csv` にアップロードしてあります。データセットには三つの変数，`drug`,`therapy`,`mood.gain` が含まれています。

この章の目的として，私たちが本当に関心を持っているのは，`mood.gain` に対して `drug` の効果があるかどうかです。最初にやるべきことは，記述統計を計算していくつかのグラフを描くことです。第 **??**章で，これを JASP でどのようにするか，‘Descriptives’ – ‘Descriptive Statistics’ の中の ‘Split’ ボックス’ を使う方法をお見せしました。その結果を図 **??**に示してあります。

プロットがはっきり示しているように，Joyzepam 群は Anxifree やプラセボ群よりも大きな改善が見られています。Anxifree 群は統制群よりも大きな気分の向上がみられますが，その差は大きくありません。私たちが知りたい答えは，これらの差が “本当に” あるものかどうか，あるいはそれがただの偶然なのか?ということです。

1.2 ANOVA のしくみ

私たちの臨床試験データに与えられた問いに答えるために，一元配置分散分析をすることになります。まずは一から統計的ツールを組み立てる難しい方法を見せるところからはじめ，JASP に組み込まれているかっこいい ANOVA 関数にアクセスできなくても計算できることを示します。注意深く読み解いて欲しいと思います。そして ANOVA がどういう仕組みなのかを本当に理解するために，1,2 回はこの長い道のりに挑戦してみてください。あなたがこのやり方を掴み取ったら，なにがあっても二度と同じやり方でやらなくても構いませんから。

前のセクションで私が示した実験デザインは，三つの異なる薬による気分の変化の平均を比較することに興味があるということを，強く示していました。つまり，私たちがやろうとしている分析は t

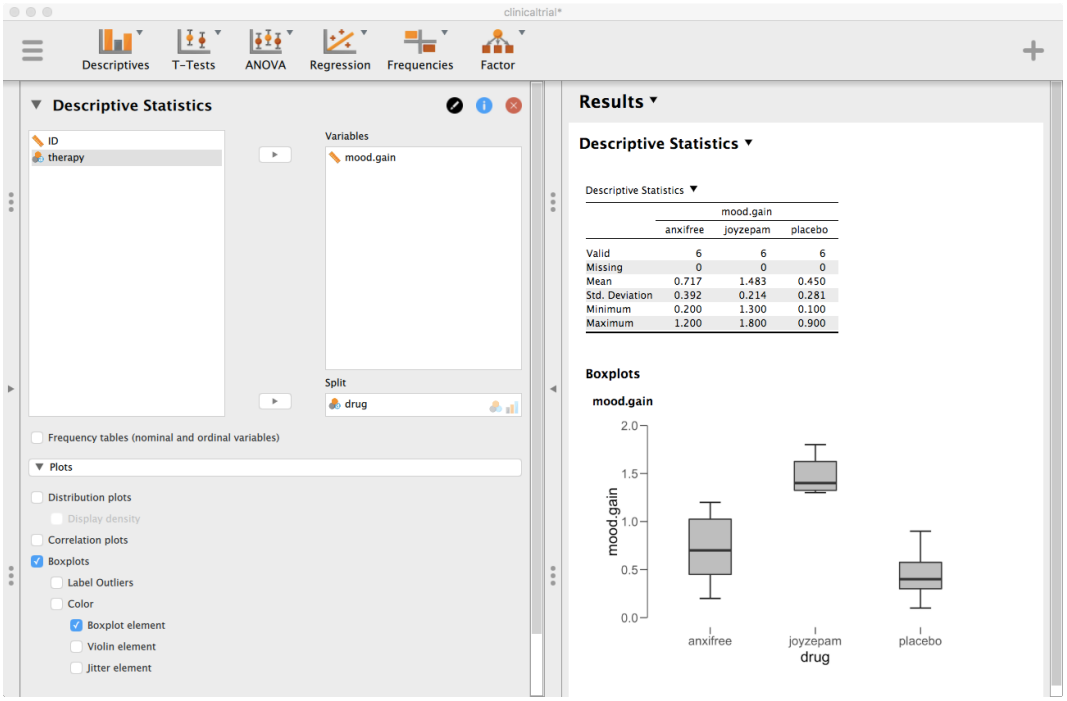


Figure1.1　記述統計と mood gain について薬の設定ごとに分けて描いたボックスプロットの JASP のスクリーンショット。

検定 (第 **??**章) に似ていますが，二つ以上のグループが含まれていることになります。ここで μ_P をプラセボによって作られた気分変容の母平均を示しているとしましょう。そして μ_A と μ_J がそれぞれ Anxifree と Joyzepm，この二つの薬による平均だとします。それから (すこし悲観主義的ですが) 私たちが検証しようとしている帰無仮説は，これら三つの母平均が同じだというものです。つまり二つの薬のどちらも，プラセボに比べてもなんの効果もない，というものです。この帰無仮説は次のように書くことができます。

$$H_0 \quad : \quad \text{次の式が正しい；} \mu_P = \mu_A = \mu_J$$

結果的に，私たちの対立仮説は，三つの異なる処遇の少なくとも一つは，他とは異なるということになります。これを数学的に書くと少しトリッキーに見えます。というのも (この後議論するように)，帰無仮説が間違っている時でもほんのちょっとした違いしかないからです。対立仮説は次のように書くことができます。

$$H_1 \quad : \quad \text{次の式が正しく } \underline{\text{ない}} \text{ ;} \mu_P = \mu_A = \mu_J$$

帰無仮説は私たちがこれまで見てきた検定のどれと比べても，トリッキーな感じ満載です。どうしたらよいでしょうか? この章のタイトルからして，“分散分析をする” というのが妥当な推測になりますが，“分散を分析する” のが平均についてなんらかの有用な知識を得る助けになるのが何故か，と

いうのがあまりはっきりしません。実際のところ，これが，人が初めて ANOVA に出会うときに感じる最大の概念的な困難なのです。ANOVA がどのうように働くかをみるためには，分散について話を始めるのが最も良いことを私は見つけました。実際，分散を記述する数式を使った，ある種の数学ゲームをプレイしながら話をしたいと思います。すなわち，分散の周りで遊ぶところから始めると，これが興味の対象である平均にとって便利なツールになるということがわかります。

1.2.1 Y の分散についての二つの式

まず，いくつかの表記法を導入するところから始めましょう。ここではグループの総数を表すのに，G を使います。三つの薬に関するデータの場合は，$G = 3$ 群あることになります。次に，全体のサンプルサイズを N とします；つまり，私たちのデータセットには全部で $N = 18$ 人いることになります。同様に，第 k 番目の群にいる人は，N_k と書くことにします。私たちの仮想的臨床検査の場合，サンプルサイズは三つの群全てにおいて $N_k = 6$ です。[*1]最後に，結果変数を Y と書くことにします。私たちの例では，Y は気分の変化です。特に，第 k 群の第 i 番目のメンバーに生じた気分変化は，Y_{ik} と書きます。同様に，この実験における 18 人全員について，気分変化の平均を撮ったものを $\bar{Y}$ とし，第 k 群における 6 名の気分変化の平均は，$\bar{Y}_k$ とします。

さて，これで表記法が出揃ったので，式を書き始めることができます。始めるにあたって，セクション **??**で使われていた分散の式を思い出しましょう。私たちが記述統計を行っていた，あの懐かしい日々に戻って。Y の標本分散は次のように書くのでした。

$$\text{Var}(Y) = \frac{1}{N} \sum_{k=1}^{G} \sum_{i=1}^{N_k} \left(Y_{ik} - \bar{Y} \right)^2$$

この数式はセクション **??**の分散の式と，ほとんど同じように見えます。違いは今回二つの添字があることだけです：私は群ごと (すなわち，k の値) に足し合わせて，それから群内の人 (すなわち i の値) について足し合わせました。これは本当に表面的な違いです。もしここで，サンプルにおけるある人 p の結果変数の値として Y_p を使っていたら，添字は一つだけになったでしょう。二つの添字を使っている唯一の理由は，私が人を群に分けたからであり，群の中の人の番号を割り当てたからです。

ここで具体的な例を見るとわかりやすいでしょう。$N = 5$ 人の人がそれぞれ $G = 2$ 群に分類された，次の表を見てみましょう。適当に“イケてる”グループを群 1，“イケてない”グループを群 2 としましょう。ここには 3 人のイケてる人 ($N_1 = 3$) と，2 人のイケてない人 ($N_2 = 2$) がいます。

[*1] 全ての群で同じ数の観測がなされている場合，その実験デザインは“バランス (のとれている)”デザインだといいます。バランスはこの章のトピックである，一元配置 ANOVA の時にはそれほど問題になりません。もっと複雑な ANOVA をやろうとすると，より重要な問題になってきます。

名前	人	群	群番号	群の中の番号	ダサさ
	p		k	i	Y_{ik} or Y_p
Ann	1	cool	1	1	20
Ben	2	cool	1	2	55
Cat	3	cool	1	3	21
Dan	4	uncool	2	1	91
Egg	5	uncool	2	2	22

ここでは，2 種類の異なるラベリング技術を使われていることに注意してください。“人” 変数は p で表ますので，サンプルの中の p 番目の人のダサさを Y_p として表現することができます。例えば，この表では Dan は 4 番目なので，$p = 4$ とすることができます。さて，この “Dan” なる人物のダサさ Y について話すとき，彼がどんな人であったとしても，彼のダサさを $Y_p = 95$ で $p = 4$ である，という参照をすることができます。しかし，Dan を参照する方法はこれだけではありません。もう一つの方法として，Dan が “イケてない” グループ ($k = 2$) に所属しており，イケてない群 ($i = 1$) のリストの最初の人だということもできます。ですから同じように Dan のダサさを参照するのに，$Y_{ik} = 91$ で $k = 2$ かつ $i = 1$ ということもできるのです。

言い換えると，各対象者 p が一つの組み合わせ k に対応しているので，上で挙げた式は実質的に元の分散の式と同じ，つまり次のようになるのです。

$$\mathrm{Var}(Y) = \frac{1}{N}\sum_{p=1}^{N}\left(Y_p - \bar{Y}\right)^2$$

どちらの式でも，サンプルにおける全ての観測例を足し上げることになります。ほとんどの場合，より単純な表記である Y_p という書き方をします。Y_p という式は二つの中で明らかに単純な方の書き方ですね。しかし，ANOVA をするときは各対象者がどちらの群に所属しているかを保持した書き方であることが重要になるため，Y_{ik} をつかってこれを書き表すこともあるのです。

1.2.2 分散から平方和へ

オウケイ，分散がどのように計算されるかを大体掴んだところで，**平方の総和**と呼ばれるものを定義しましょう。表記は SS_{tot} です。これはとても単純です。分散を計算する時には平均偏差の二乗を平均するわけですが，その代わりにそれを単に足し合わせます。

ですので平方の総和の数式は，分散の式とほとんど同じです。

$$\mathrm{SS}_{tot} = \sum_{k=1}^{G} \sum_{i=1}^{N_k} \left(Y_{ik} - \bar{Y}\right)^2$$

ANOVA の文脈で分散を分析することについて話をするときは，実際の分散ではなく平方の総和をつかって実行していることになります。平方の総和を使う利点の一つは，それを異なる二種類の変動に分解することができる点です。

まず，**群内平方和**について話しましょう。そこでは群平均から個々人がどれほどずれているかを見ることができます。

$$\mathrm{SS}_w = \sum_{k=1}^{G} \sum_{i=1}^{N_k} \left(Y_{ik} - \bar{Y}_k\right)^2$$

ここで $\bar{Y}_k$ は群平均です。例えば，$\bar{Y}_k$ は k 番目の薬を与えられた人による，気分変容の平均値です。ですから，実験に参加した全員の平均と個々人を比較するのではなく，おなじ群にいる人の平均と比較していることになります。結果的に，SS_w の値は平方和の総和よりも小さくなります。というのも，そこには群の違い，すなわち薬が人の気分に与える影響の違いがあるかどうかを完全に除外しているからです。

次に，群の違いだけを捉えた変動を記述する第 3 の表記を定義しましょう。これは全体平均 $\bar{Y}$ と群平均 $\bar{Y}_k$ の間のずれを見ることになります。

この変動の大きさを評価するために，**群間平方和**を計算することになります。

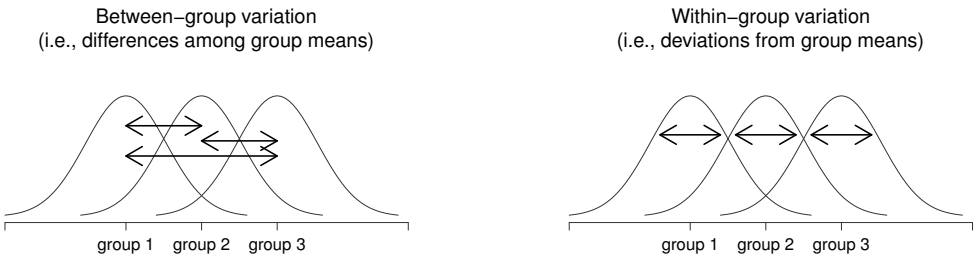
$$\begin{aligned}\mathrm{SS}_b &= \sum_{k=1}^{G} \sum_{i=1}^{N_k} \left(\bar{Y}_k - \bar{Y}\right)^2 \\ &= \sum_{k=1}^{G} N_k \left(\bar{Y}_k - \bar{Y}\right)^2\end{aligned}$$

実験におけるすべての人の中にある全変動 SS_{tot} を示すことはそれほど難しくありません。実際には，群間の変動 SS_b と群内の変動 SS_w を足しあわせるのですすなわち，

$$\mathrm{SS}_w + \mathrm{SS}_b = \mathrm{SS}_{tot}$$

イェイ。

さて，何がわかったのでしょう？ 結果変数に伴う変動全体 (SS_{tot}) は“異なる群の標本平均の違いに伴う変動”(SS_b) と，“その残りの変動”(SS_w) を足し合わせたものに切り分けられるということで



(a) (b)

Figure1.2 “群間” 変動 (パネル a) と “群内” 変動 (パネル b) を図示したもの。左の図では，矢印が群間平均の差を示しています。右図では矢印が群内の変動を強調しています。

. . . . . .

す。[*2]では母平均が群ごとに異なっているかどうかを検証するには，どうすれば良いのでしょう? フムゥ。待てよ。ちょっとまってください。今思うに，これこそまさに私たちが探していたものです。もし帰無仮説が真であるとすれば，全ての平均はそれぞれほとんど同じものになると思えませんか? そしてそれが意味するのは，SS_b が非常に小さい，少なくとも “それ以外の全ての変動” である SS_w よりは小さいことが期待できるでしょう。ムムッ。仮説検定が始まる予感がします。

1.2.3 平方和から *F*-検定へ

最後のセクションでみたように，ANOVA の背後にある本質的なアイデアは SS_b と SS_w という二つの平方和の値を比較するところにあります。群間変動の SS_b が群内変動 SS_w に比べて大きい時，違う群の母平均は互いに等しいとは言えない，という推論をする根拠をもつことになります。これを実際の帰無仮説検定に変換するために，“ちょっといじくり回す” ことが必要です。最初にお見せするのは，検定統計量として何を計算するのかであり，それは ***F* 比**です。それからなぜ私たちがそんな風にするのかを理解してもらおうと思います。

平方和の値を *F* 比に変換するために最初にしなければならないのは，SS_b と SS_w に関する**自由度**を計算することです。普通は，自由度はある計算に関わるユニークな “データ点” の数から，満たす必要のある “制約” の数を引いたものに対応しています。私たちが計算している群内変動は，群平均 (*G* 個の制約) の周りにある個々人の観測による変動 (*N* データ点) です。対して私たちが興味のある群間変動は，全体平均 (1 個の制約) の周りにある群平均 (*G* データ点) の変動です。つまり，こ

[*2] SS_w は AVOVA では誤差，すなわち SS_{error} と表されることもあります。

こでの自由度は次のようになります。

$$\begin{aligned} \mathrm{df}_b &= G - 1 \\ \mathrm{df}_w &= N - G \end{aligned}$$

オーケー，とても単純ですね。次にすることは，平方和を “平均平方” に変換することで，これは自由度で割ることで計算できます。

$$\begin{aligned} \mathrm{MS}_b &= \frac{\mathrm{SS}_b}{\mathrm{df}_b} \\ \mathrm{MS}_w &= \frac{\mathrm{SS}_w}{\mathrm{df}_w} \end{aligned}$$

最後に F 比を計算するために，群間平均平方を群内平均平方で割ります。

$$F = \frac{\mathrm{MS}_b}{\mathrm{MS}_w}$$

非常に一般的な意味で，F 統計量の背後にあるものは直感的にわかります。F の値がより大きいことは，群間変動が群内変動よりも相対的に大きいことを意味します。結果的に，F の値が大きいことは，帰無仮説に対立するより大きな証拠を得たことになります。しかし，F がどれぐらい大きければ，実際に H_0 を棄却できるのでしょう？ これを理解するためには，ANOVA が何であるか，平均平方とは何なのかを，もう少し深く理解しなければなりません。次のセクションではこの詳細について説明していきますが，実際に検定が何をしているかに興味のない読者のために，ここではそれを省略しましょう。帰無仮説検定を完成させるために，帰無仮説が真である時の F の標本分布について知らなければなりません。驚くなかれ，帰無仮説のもとでの F統計量の標本分布は F 分布です。第 **??** 章での F 分布についての議論を思い出してもらいたいのですが，F 分布は二つのパラメータを持っていて，それが二つの自由度に対応しています。最初の自由度 df_1 は群間の自由度 df_b，第二の自由度 df_2 は群内の自由度 df_w です。

一元配置分散分析に含まれる，全ての重要な数字の要約を，計算に使った数式とともに，表 **??**に示しました。

1.2.4 データのためのモデルと F の意味

ANOVA の根本的なレベルでは，二つの異なる統計モデルである H_0 と H_1 が競合します。帰無仮説と対立仮説をこのセクションの冒頭で論じた時，これらのモデルが実際に何を表すのかについては少し不完全なまま紹介しました。今からそれを撤回して詳述していきますが，たぶんあなたはそんなことをする私のことが嫌いになるでしょう。思い出してもらいたいのですが，私たちの帰無仮説は全ての群平均は相互に等しいというものでした。もしそうであるなら，結果変数 Y_{ik} について考える自然な方法は，一つの母平均 μ に，そこの変動を加えたものとして，個々のスコアを記述

Table1.1　分散分析の中に含まれる重要な全ての数字は，“標準的な” 分散分析表の中に組み込まれます。この形式では全ての数字 (p 値と呼ばれる奇妙な数式とコンピュータがないと計算できないものを除いて) が示されています。

	自由度 (df)	平方和 (SS)	平均平方 (MS)	F-統計量	p-値
between groups	$\text{df}_b = G - 1$	$\text{SS}_b = \sum_{k=1}^{G} N_k(\bar{Y}_k - \bar{Y})^2$	$\text{MS}_b = \frac{\text{SS}_b}{\text{df}_b}$	$F = \frac{\text{MS}_b}{\text{MS}_w}$	[complicated]
within groups	$\text{df}_w = N - G$	$\text{SS}_w = \sum_{k=1}^{G}\sum_{i=1}^{N_k}(Y_{ik} - \bar{Y}_k)^2$	$\text{MS}_w = \frac{\text{SS}_w}{\text{df}_w}$	-	-

することです。この変動というのは普通 ε_{ik} で表され，伝統的にそれは観測に伴う誤差とか**残差**と呼ばれています。でも気を付けてください。“有意差” という言葉を見た時と同じように，“エラー (誤差)” には統計における専門的な意味を持っていて，日常的な定義とは異なります。日常用語で言う “エラー” は，ちょっとした失敗と言う意味ですが，統計的にはそうではありません (少なくとも，必ずしもそうではありません)。それを考慮すると，“残差” の方が “エラー (誤差)” よりも良い言葉かもしれません。統計学ではどちらの言葉も，“残っている変動”，すなわちこのモデルが説明できない “あまりもの” を意味しています。

いずれにせよ，帰無仮説を統計モデルとして書き表すと，こんな感じになります。

$$Y_{ik} = \mu + \varepsilon_{ik}$$

ここで (後ほど議論する) ある仮説を立てます。それは残差の値 ε_{ik} は正規分布に従うというもので，その平均は 0, 標準偏差 σ はすべての群を通じて同じであるというものです。第 **??**章で導入した表記法を使うと，この仮定は次のように書くことができます。

$$\varepsilon_{ik} \sim \text{Normal}(0, \sigma^2)$$

対立仮説 H_1 についてはどうでしょうか? 帰無仮説と対立仮説の唯一の違いは，各群は異なる母平均を持ちうるということです。ですから，実験における第 k 番目の群の母平均を μ_k とするなら，H_1 に対応する統計モデルは次のようになります。

$$Y_{ik} = \mu_k + \varepsilon_{ik}$$

ここでもう一度，誤差項が平均 0，標準偏差 σ の正規分布に従うと仮定します。すなわち，対立仮説もまた次のように書けるわけです。

オウケイ，さあ H_0 と H_1 を支える統計モデルについて記述してきましたが，そろそろ平均平方の値が何を測っているのか，そして F の解釈にそれがどう関係するのかを，もう少しはっきりいうべき段階に来ました。照明を初めてあなたを退屈がらせることはしませんが，群内平均平方である MS_w が，誤差分散 σ^2 の推定量 (専門的な意味は，第 **??**章を参照) とみなすことができることがわかります。群間平均平方 MS_b も推定量ですが，それが推定しているのは誤差分散に加えて，群平均間の真の差分についての量も入っています。もしこの量を Q と書くなら，F 統計量は基本的に次のようになっています[a]。

$$F = \frac{\hat{Q} + \hat{\sigma}^2}{\hat{\sigma}^2}$$

ここで，もし帰無仮説が正しければ，真の値は $Q = 0$ ですし，もし対立仮説が正しければ $Q > 0$ でしょう (**Hays1994**). つまり，少なくとも F の値は *1* よりも大きくならなければ，帰無仮説を棄却する可能性がないのです。これは，F 値が 1 よりも小さくなることがないことを意味するのではないことに注意してください。これが意味するのは，帰無仮説が真であれば，F 比の標本分布の平均は 1[b]であり，安全に帰無仮説を棄却するためには F の値は 1 よ理大きくなければならないということです。標本分布についてもう少し正確にいうと，帰無仮説が正しい時，MS_b も MS_w も残差 ε_{ik} の分散推定量であることに注意してください。もしこれらの残渣が正規分布に従っていたとすると，ε_{ik} の分散の推定値がカイ二乗分布に従うのではないか，と考えるかもしれません。というのも (セクション **??**で論じたように)，それこそカイ二乗分布そのものだからです。すなわち，正規分布するものを二乗して足し合わせたものだからです。そして，F 分布は (もう一度，その定義に戻りますが) 二つの χ^2 分布の比を取ったものであり，それが私たちの標本分布なわけです。明らかに，私は多少これをいうときに大袈裟に表現していますが，大まかに言って，本当に私たちの標本分布からきたものになります。

[a] 第 **??**章まで読めば，要因の水準 k における “操作の影響” がどのように α_k の値で定義されるかを目にするでしょう (セクション **??**)。そこでは，Q は重みづけられた操作の影響の平方，つまり $Q = (\sum_{k=1}^{G} N_k \alpha_k^2)/(G-1)$ となっていることがわかります。

[b] あるいは，もし正確さにこだわるのであれば，$1 + \frac{2}{df_2 - 2}$.

1.2.5 実例

ここまでの議論はかなり抽象的で，少し理論的な話でしたので，今度は実際の例を見ることでもう少し有用な話をみていこうと思います。そのために，私がこの章の始めに導入した臨床試験データに戻りましょう。最初に計算した記述統計量からわかるのは，群平均でした。つまり，気分の向上はプラセボで 0.45 高，Anxifree で 0.72，Joyzepam で 1.48 でした。それを念頭に置いて，1899 年のよ

うなパーティーをしましょう[*3]，そして紙と鉛筆で計算をするのです。最初の五人についてだけ計算することにします。だって地獄の 1899 年じゃないし，私はとっても怠け者ですから。群内平方和，SS_w を計算するところから始めるとしましょう。まず，計算しやすくするために便利な表を作ることにします。

群 k	アウトカム Y_{ik}
プラセボ	0.5
プラセボ	0.3
プラセボ	0.1
anxifree	0.6
anxifree	0.4

この段階で表に含めることができるのは，ロウデータそのものでしかありません。つまり群化変数 (すなわち `drug`) と結果変数 (すなわち `mood.gain`) を各人についてのものだけ，です。結果変数が私たちの式表現によると，Y_{ik} に対応していることに注意してください。次の計算ステップはこの研究の一人におけるひとりについて，対応する群平均 $\bar{Y}_k$ に書き下すことです。これは少し繰り返しになりますが、記述統計を行う際にグループの平均値を計算しているので、特に難しいことではありません。

群 k	アウトカム Y_{ik}	**群平均** $\bar{Y}_k$
プラセボ	0.5	**0.45**
プラセボ	0.3	**0.45**
プラセボ	0.1	**0.45**
anxifree	0.6	**0.72**
anxifree	0.4	**0.72**

さて，これらを書き出したので，再び一人一人について，対応する群平均からのズレを計算する必要があります。つまり，$Y_{ik} - \bar{Y}_k$ の引き算をしたいのです。そうした上で，それらを二乗します。そうすると，ここにあるような数字を得ることができます。：

群 k	結果 Y_{ik}	群平均 $\bar{Y}_k$	群平均からの**偏差** $Y_{ik} - \bar{Y}_k$	**偏差の平方** $(Y_{ik} - \bar{Y}_k)^2$
プラセボ	0.5	0.45	**0.05**	**0.0025**
プラセボ	0.3	0.45	**-0.15**	**0.0225**
プラセボ	0.1	0.45	**-0.35**	**0.1225**
anxifree	0.6	0.72	**-0.12**	**0.0136**
anxifree	0.4	0.72	**-0.32**	**0.1003**

最後のステップは結構ストレートです。群内平方和を計算するために，全ての観測に対して偏差平

[*3] あるいは，もう少し正確に，“1899 年の，友達もいないし 1899 年にはなんの意味もなかった計算をするしかないことをしましょう。だって ANOVA は 1920 年代になるまで出てこないんですから”

方を足し合わせて行きます。

$$\begin{aligned} SS_w &= 0.0025 + 0.0225 + 0.1225 + 0.0136 + 0.1003 \\ &= 0.2614 \end{aligned}$$

もちろん，実際にはデータセットの中にある全 18 人分についてこの計算をした，正しい答えが欲しいわけです。最初の 5 人分だけじゃなくてね。そうしたければ，神とペンを使って計算し続けてもらってもいいんですけど，それはちょっと面倒ですよね。その代わりとして，専用のスプレッドシート・アプリである OpenOffice や Excel を使うと随分と楽に計算できます。是非ご自身でやってみてください。そうすると、群内平方和の値が 1.39 になるはずです。

オーケイ。群内分散 SS_w の計算が終わったので，群間平方和，SS_b に向き合う時が来ました。この計算は非常によく似ています。大きな違いは，全ての観測に対して群平均 $\bar{Y}_k$ と観測値 Y_{ik} の差分を計算する代わりに，全体平均 $\bar{Y}$(この回は 0.88 ですが) と群平均 $\bar{Y}_k$ の差を全ての群に対して計算するところです。

群 k	群平均 $\bar{Y}_k$	全体平均 $\bar{Y}$	偏差 $\bar{Y}_k - \bar{Y}$	偏差平方 $(\bar{Y}_k - \bar{Y})^2$
プラセボ	0.45	0.88	-0.43	0.19
anxifree	0.72	0.88	-0.16	0.03
joyzepam	1.48	0.88	0.60	0.36

ただし，群間の計算をするときは，この偏差平方にその群に含まれる観測度数 N_k をかけなければなりません。というのも，ある群における観測はどれも (すべての N_k について)，群間の差を作るのに貢献しているからです。ですから，プラセボ群に 6 人いて，プラセボ群平均の全体平均からの差が 0.19 であれば，全体としてこの 6 人が関わった群間分散は $6 \times 0.19 = 1.14$ になります。ということで，計算のための表を少し拡張しなければなりません。

群 k	… …	偏差平方 $(\bar{Y}_k - \bar{Y})^2$	サンプルサイズ N_k	重み付き偏差平方 $N_k(\bar{Y}_k - \bar{Y})^2$
placebo	…	0.19	6	1.14
anxifree	…	0.03	6	0.18
joyzepam	…	0.36	6	2.16

そして，今やこの研究における全ての群に関する群間平方和が，“重み付き偏差平方” を足し合わせることによって得られるわけです。

$$\begin{aligned} SS_b &= 1.14 + 0.18 + 2.16 \\ &= 3.48 \end{aligned}$$

みてきたように，群間の計算は短かったですね。さて，平方話の計算は SS_b and SS_w, になりましたので，ANOVA の残りはたいしたことありません。次のステップは自由度の計算です。私たちのデータは $G = 3$ の群，そして $N = 18$ 人の観測度数をもっていますから，全体の自由度は簡単な引き算で算出できます。

$$
\begin{aligned}
\mathrm{df}_b &= G - 1 &= 2 \\
\mathrm{df}_w &= N - G &= 15
\end{aligned}
$$

次に，群内分散と群間分散それぞれについて，平方和と自由度を計算したのですから，前者を後者で割ることで平均平方が計算できますね。

$$
\begin{aligned}
\mathrm{MS}_b &= \frac{\mathrm{SS}_b}{\mathrm{df}_b} &= \frac{3.48}{2} &= 1.74 \\
\mathrm{MS}_w &= \frac{\mathrm{SS}_w}{\mathrm{df}_w} &= \frac{1.39}{15} &= 0.09
\end{aligned}
$$

ほとんど終わったようなもんです。この平均平方は，我々の興味ある検定統計量である F 値を計算するのに使われます。群間 MS の値を群内 MS の値で割ることによってその数字が得られます。

$$
F = \frac{\mathrm{MS}_b}{\mathrm{MS}_w} = \frac{1.74}{0.09} = 19.3
$$

イヤッホーイ！ 興奮してきましたね?さてこの検定統計量を手に入れたわけですから，最後のステップはこの検定が有意な結果を出してくれたかどうかを見極めることです。第 **??**章で議論したように，“昔” に戻って統計のテキストを開いたり，後ろの方のセクションに載ってる大きな表のをペラペラめくると，特定のアルファの値 (帰無仮説が棄却される範囲)，たとえば 0.05,0.01,0.001 など，に対応した，自由度 2 と 15 の F 値の閾値を見つけることができます。

そうすると，アルファが 0.001 のときの F の閾値は 11.34 であることがわかりました。これは私たちが計算した F よりも小さいので，$p < 0.001$ と言うことができます。しかしこれは昔ながらのやり方であって，今では賢い統計ソフトウェアがあなたのために正確な p 値を計算してくれますよ。実際，正確な p 値は 0.000071 になります。さて，タイプ I のエラー率について極めて保守的であったとしても，我々は帰無仮説を棄却するのに十分な保証が得られたと言えるでしょう。この時点で，基本的に終了です。計算を終えたらば，これら全ての数字を表 **??**のようにして，分散分析表にまとめるのが伝統的なやり方です。我々の臨床試験データについての，分散分析表は次のようになります。

	自由度	平方和	平均平方	F-統計量	p-値
群間	2	3.48	1.74	19.3	0.000071
群内	15	1.39	0.09	-	-

今日では，こうした表を自分で作る理由がないように思うかもしれませんが，ほとんどの統計ソフトウェア (JASP もそうです) は ANOVA の結果をこうした表にまとめる傾向があることに気づくと思います。なので，読み方に慣れていた方がいいでしょう。とはいえ，ソフトウェアが完全な分散分析表を出力してくれますが，あなたが書くときに全部を含める理由はあんまりありません。この結果をレポートする標準的な方法は，次のように書くことです。

> 一言配置の分散分析では，投薬による気分の向上の有意な効果が示された ($F(2, 15) = 19.3, p < .001$)。

ふー。短い一文を書くために，たいした苦労をするもんだ。

1.3 ANOVA を JASP で実行する

あなたがこのセクションの最後を読んだときに，特にあなたが私のアドバイスに従って紙とペンで (あるいはスプレッドシートで) 計算したのであれば，どんなふうに感じるかは手に取るようにわかります。ANOVA の計算を自分でやることは最悪です。手順通りに計算する量がとてもおおいので，ANOVA をしようと思うたびに何度も何度も計算するのは面倒なのです。

1.3.1 あなたの ANOVA のための JASP

あなたの人生を楽にするために，JASP は ANOVA を実行してくれます.... 万歳！ 'ANOVA'-'ANOVA' 分析，とすすみ，`mood.gain` 変数を '従属変数' ボックスに移動させ，それから `drug` 変数を '固定効果' ボックスに動かします。こうすると図**??**のような結果が示されます。[*4]'追加オプション' の '効果量の推定' の下にある，η^2 チェックボックスにもチェックを入れたので，結果の表にもそれが反映されています。効果量についてはあとで触れることにします。

ANOVA

ANOVA - mood.gain

Cases	Sum of Squares	df	Mean Square	F	p	η^2
drug	3.453	2.000	1.727	18.611	< .001	0.713
Residual	1.392	15.000	0.093			

Note. Type III Sum of Squares

Figure1.3　気分の向上と投薬の関係についての JASP による分散分析表

JASP の結果の表には，平方和，自由度，そのほか今は興味のないいくつかの統計量が示されます。しかし，JASP には "群間" とか "群内" といった表示はしていません。その代わりに，もっと意味のある名前がついています。今回の例では，群間分散は**投薬**がアウトカム変数に及ぼした影響に対

[*4] JASP の結果は，丸め誤差のあった上の文章で述べたものより正確です。

応しており，群内分散は残差誤差とも呼ばれる “残った” 変動分に対応しています。これらの数字を，セクション **??**で手計算した数字と比較すると，四捨五入によるズレを除いてほぼ同じであることがわかると思います。群間平方和は $\text{SS}_b = 3.453$ で，群内平方和は $\text{SS}_w = 1.392$，そしてそれぞれの自由度は 2 と 15 です。F-値と p-値も計算されていて，それも四捨五入によるズレを除いて，先程の長く面倒な方法で計算したものと同じになっていることがわかります。

JASP の結果の表には，平方和，自由度，そのほか今は興味のないいくつかの統計量が示されます。しかし，JASP には “群間” とか “群内” といった表示はしていません。その代わりに，もっと意味のある名前がついています。今回の例では，群間分散は**投薬**がアウトカム変数に及ぼした影響に対応しており，群内分散は残差誤差とも呼ばれる “残った” 変動分に対応しています。これらの数字を，セクション **??**で手計算した数字と比較すると，四捨五入によるズレを除いてほぼ同じであることがわかると思います。群間平方和は $\text{SS}_b = 3.453$ で，群内平方和は $\text{SS}_w = 1.392$，そしてそれぞれの自由度は 2 と 15 です。F-値と p-値も計算されていて，それも四捨五入によるズレを除いて，先程の長く面倒な方法で計算したものと同じになっていることがわかります。

1.4 効果量

ANOVA において効果量を測定するには，二つの異なる方法がありますが，最も一般的に使われているのは η^2 (**eta squared**) と偏 η^2 です。一要因の分散分析では，どちらも同じになりますので，今はとりあえず η^2 について説明します。η^2 は実際のサンプルについて，次のように定義されます。

$$\eta^2 = \frac{\text{SS}_b}{\text{SS}_{tot}}$$

これだけです。図 **??**の分散分析表についてみてみると，$\text{SS}_b = 3.45$ で $\text{SS}_{tot} = 3.45 + 1.39 = 4.84$ です。ここから，η^2 の値は次のように計算できます。

$$\eta^2 = \frac{3.45}{4.84} = 0.71$$

η^2 の解釈も，実に直接的です。アウトカム変数 (`mood.gain`) において，予測変数 (`drug`) が説明する分散の比率をあらわしているわけです。値が $\eta^2 = 0$ であれば，両者になんの関係もないことを表しますし，$\eta^2 = 1$ であれば完璧な関係にあることになります。さらに良いことに，η^2 の値は，セクション **??**)で説明したR^2 にかなり近いもので，それと同じように解釈することができます。

多くの統計的教科書では η^2 を ANOVA における基本的な効果量の測度だと説明していますが，Daniel Lakens がブログで興味深いことを言っています。それによると，η^2 はデータ分析業界における最適な効果量測度ではない，なぜなら推定量のバイアスがあるからだ，ということです (`http://daniellakens.blogspot.com.au/2015/06/why-you-should-use-omega-squared.html`)。

ありがたいことに，JASP にはオメガの二乗 (ω^2) という選択肢もついていて，これはよりバイアスが少なく，イータの二乗と並んで使われているものです。

1.5 多重比較と事後の (Post hoc) 検定

2 群以上で ANOVA をして有意な影響をみたときは，実際はどの群がどの群と差があったのか知りたくなるでしょう。投薬の例では，帰無仮説は三つの薬 (プラセボと Anxifree と Joyzepam) が気分に与える影響は同じというものでした。しかし考えてみると，帰無仮説は実際には三つの異なることを同時に主張しているのです。つまり，主張は次の通りです。

- あなたの競争相手である薬 (Anxifree) はプラセボと変わりない (つまり，$\mu_A = \mu_P$)。
- あなたの薬 (Joyzepam) はプラセボと変わりない (つまり, $\mu_J = \mu_P$)。
- Anxifree と Joyzepam の効果は同じぐらい (i.e., $\mu_J = \mu_A$)。

この三つの主張のどれかが偽であれば，帰無仮説も偽になります。ですから，我々が帰無仮説を棄却するとき，このなかの少なくとも一つは真であることになります。でもどれでしょう？ 三つの命題全てに興味がありますよね。あなたが本当に知りたいのは，あなたの新薬 Joyzepam がプラセボより良いはずだ，というものですから，既存の一般的な代替品 (つまり Anxifree) と比べてどの程度の効果があるのかを知っておくのは良いことでしょう。Anxifree がプラセボにくらべてどれぐらい効果があるのかをチェックするのもまた，いいことのはずです。Anxifree は既に他の研究者によって，プラセボに対する効果の検証がしっかり行われているはずですが，あなたの研究でも先行研究と同じ結果が出ることを示すかどうかのチェックをするのも，いいことのはずです。

この三つの異なる命題を使って帰無仮説を特徴付けるとしたら，8 つのあり得た “世界の状態” を区別する必要があります。

可能性:	$\mu_P = \mu_A$?	$\mu_P = \mu_J$?	$\mu_A = \mu_J$?	どの仮説か?
1	✓	✓	✓	帰無仮説
2	✓	✓		対立仮説
3	✓		✓	対立仮説
4	✓			対立仮説
5		✓	✓	対立仮説
6		✓		対立仮説
7			✓	対立仮説
8				対立仮説

帰無仮説を棄却することで，我々は#1 が真である世界を信じることはできないと決めたわけです。次の質問は，残る 7 つの可能性のうち我々が正しいと考えることができるのはどれか？ ということです。この状況に置かれた時にも，データを見るのが助けになります。例えば，図 **??**をみると，Joyzepam はプラセボや Anxifree より良いようですが，Anxifree とプラセボに実質的な違いはないように思えます。しかし，これにはっきりと答えるのは難しくいので，いくつかの検定をして助けてもらうことにします。

1.5.1 “ペアごとの pairwise” *t*-検定

さて問題解決のために，何をすべきでしょう？ それぞれの平均のペア (プラセボ vs.Anxifree, プラセボ vs.Joyzepam,Anxifree vs. Joyzepam) は既にあるのですから，それぞれに t 検定をしてどうなるかみてみるというのはどうでしょう？ それを JASP でするのは簡単です。ANOVA の ‘事後の検定’ オプションへ行き，‘drug’ 変数を右側のアクティブボックスに動かします。するとすぐに，`drug` の三つのレベルに対して，ペアごとの t-検定全てが表示されます。図**??**のように。

Post Hoc Tests

Post Hoc Comparisons - drug

		Mean Difference	SE	t	p_{bonf}	p_{holm}
anxifree	joyzepam	−0.767	0.176	−4.360	0.002	0.001
	placebo	0.267	0.176	1.516	0.451	0.150
joyzepam	placebo	1.033	0.176	5.876	< .001	< .001

Figure1.4　JASP による事後のペアごとの t-検定結果

1.5.2 多重検定のための補正

前のセクションでは，ここでの問題についてたくさんの t 検定で対応するというヒントを与えました。これらの分析を実行するときに懸念されるのは，“釣り探検” に出かけてしまったのではないかということです。何か有意になるんじゃないかと期待しながら，理論的なガイダンスなしにたくさんのたくさんの検定を行ってしまいました。このような理論無視の群間比較は，**事後検定 post hoc**

analysis と言われます (“post hoc” はラテン語で “after this” という意味です).[*5]

事後の検定をするのはいいのですが，注意が必要なのです。例えば，前のセクションでやったような分析は，それぞれ個別に行われた t-検定ですが，これは 5% のタイプ I エラー (つまり $\alpha = 0.5$) で三つの検定を行ったことになります。もし私の ANOVA が 10 群について行われていたのだとしたら，わたしは 45 回の “事後の”t 検定をして，ある一つの水準が他の一つと有意に異なるかどうか，それぞれ検証することになります。あなたはそのうちの 2 つか 3 つぐらいが，偶然有意になってしまうかも，と思うかもしれません。第 **??**章でみたように，帰無仮説検定の背後にある原則は，タイプ I エラーをコントロールしたいというものであったのですが，今や私はたくさんの t 検定を一度に，ANOVA の結果に基づいて判断する目的で実施しており，この時，一連の検定を通じた実際のタイプ I エラーは，完全に制御不能になっています。

この問題についての一般的な解決策は，p 値を調整することです。これは一連の検定全体を通じたエラー発生率をコントロールすることが目的です。(**Shaffer1995**). 事後の検定をする時は，普通はこのやり方で補正され (いつもではありません)，この手続きは**多重比較の補正**と呼ばれます。時には “同時推論の補正” と言われることもあります。いずれにせよ，この補正のやり方にはいくつかの異なる方法があります。このセクションとセクション **??**では，これらのいくつかについて取り上げますが，多くの他のやり方があるんだということは知っておいてください。(**Hsu1996**)

1.5.3 Bonferroni の補正

最も単純な補正は **Bonferroni の補正**と呼ばれるもので，それは実にとーーーーーってもシンプルなものです。(**Dunn1961**) m 個それぞれの検定をするような事後分析を想像してみましょう。どれかがタイプ I エラーを引き起こす確率の合計が，最大でも α になることを保証するようにしたいと思っています。[*6]このとき，Bonferroni の補正は単に “あなたのもとの p 値を m で割れ” というだけです。もし元の p 値を表すのに p と表記するなら，p'_j を補正した値の書き方として，Bonferroni の補正は次のようになります。:

$$p' = m \times p$$

ですから，もしあなたが Bonferroni の補正を使おうとするのなら，$p' < \alpha$ の時に帰無仮説を棄却するようにしてください。この補正の裏にあるロジックはとても直接的ですよね。m 個の異なる検定をするなら，各検定のタイプ I エラーは大きくても α/m になるはずで，全体的なタイプ I エラーは α 以上になり得ないのですから。これはとても単純なことなので，元の論文で筆者は次のように書いています。

[*5]もしあなたが，ある箇所を比較して他のところはしない，という理論的な基盤を持っていれば, 話は違ってきます。そういうときは，あなたは “post hoc な” 分析をしようとしているのではなくて，“計画された比較” をしようとしているわけです。こういう状況については本書の後半 (セクション **??**) で行いますが，今は話を単純化しておきましょう。

[*6]すべての調整法がそうしようとしているわけではない，ということに注意しておきましょう。ここで私が述べているのは，“ファミリーワイズのタイプ I エラー ’ ’ をコントロールするアプローチというやつです。しかし，他の事後分析では，“偽検出率” のコントロールを目指しているものもあり，ちょっと違うものです

> ここで述べた手法はとてもシンプルで一般的なものですから，以前にも誰かが使っていたでしょう。しかし，先行例を見つけることはできませんでしたので，おそらくこの超単純さのせいで，統計学者はいい方法だと気づけなかったのだと思います。(**Dunn1961**)

Bonferroni の補正を JASP で使うためには，‘補正’ オプションの ‘Bonferroni’ のチェックボックスをクリックします。そうすると，ANOVA の結果の表の中に，Bonferroni の方法で補正された p 値の列を見つけることができるでしょう (**??**)。

1.5.4 Holm の補正

Bonferroni の補正がとてもシンプルなものだったわけですが，それが常にベストなものというわけではありません。代わりによく使われるのが，**Holm の補正**というものです (**Holm1979**)。Holm の補正のアイデアは，あなたが検定を順番にやっていく時に，最も小さい (元の)p 値から初めて，最大のものに進んでいくというものです。第 j 番目の大きさを持つ p 値は，次のいずれかになります。

$$p'_j = j \times p_j$$

(つまり，最大の p 値は変化させないままにしておいて，二番目に大きな p 値は 2 倍に，三番目に大きな p 値は 3 倍に... というふうにしていきます)。あるいはまた，

$$p'_j = p'_{j+1}$$

どれか一つでも大きくなったときにこうします。これはちょっと混乱させるような書き方ですので，もう少しゆっくり説明しましょう。Holm の補正が何をするか，というのは次の通りです。まず，p 値を小さいものから大きいものへと，並べ替えてください。一番小さな p 値について，m 倍して終わりです。しかし，他のどれも二段階プロセスを経ていませんね。たとえば，もしあなたが二番目に小さな p 値を動かしたら，それを $m - 1$ 倍しなければなりません。このかけられた数が，あなたが最後に手に入れた補正された p 値よりも大きければ，取っておきましょう。しかし最後のそれよりも小さければ，最後の p 値をコピーします。これがどういう働きをするかを見るために，次の表を見てください。ここには 5 つの p 値についての Holm の補正計算が示されています。

元の p	順序 j	$p \times j$	Holm p
.001	5	.005	.005
.005	4	.020	.020
.019	3	.057	.057
.022	2	.044	.057
.103	1	.103	.103

これでわかったでしょうか？

少し計算が面倒ですが，Holm の補正はいくつかの良い特性を持っています。Bonferroni よりも良く (つまり，タイプ II エラーがより低く)，直感に反してタイプ I エラーについては同じなのです。結果として，実践ではよりシンプルな Bonferroni の補正を使う理由がなくなります。常に，より洗練された Holm の補正が効率的だからです。ですから，あなたが多重比較の補正をする時は Holm 法でいくべきでしょう。図**??**には Bonferroni と Holm の補正された p 値が示されています。

1.5.5 事後検定の記載

最後に，どの群が他と比べて有意に異なっていたかを決める事後分析を行ったら，あなたが書くだろう結果の文章はこんなふうになります。:

> 事後検定 (Holm の補正された p を使った) では，Joyzepam が Anxifree($p = 0.001$) とプラセボ　（$ p < 0.001 $）よりも有意に大きな気分変容をもたらすことが示された。Anxifree はプラセボに比べて良いという証拠は見つからなかった ($p = .15$)。

あるいはもし，あなたが $p < 0.001$ とだけ書くのは嫌だというのであれば，'設定'-'結果' にいき，'正確な p 値を表示する' を選択しておけば，正確な p 値を計算することができます。どちらにせよ，あなたが使った Holm の補正による調整済み p 値を使ったことを書いておくことが肝要です。そしてもちろん，既に関係する記述統計量 (すなわち，群平均や標準偏差) をどこかに書いてあることを想定しています。だって p 値だけではほとんど情報がないですからね。

1.6 一要因 ANOVA の仮定

あらゆる統計的検定と同じように，分散分析もデータについて，特にその残差についての仮定の上に成り立っています。知っておくべき仮定は次のとおりです。: 正規性，分散の均一性，独立性

セクション **??**のことを思い出して欲しいのですが，全体を読んでいなくても，せめて斜め読みぐらいはして欲しいのですが，私は ANOVA を支える統計モデルについて次のように説明したのでした。

$$
\begin{array}{lrcl}
H_0: & Y_{ik} & = & \mu + \varepsilon_{ik} \\
H_1: & Y_{ik} & = & \mu_k + \varepsilon_{ik}
\end{array}
$$

これらの式で，μ は一つの全体平均を表していてすべての群を通じて同じものです。また μ_k は第

k 番目の群の母平均を表しています。ここで注目しなければならないのは，我々のデータが一つの全体平均で表現できる (帰無仮説) のか，異なる群特有の平均値があるのか (対立仮説) ということです。これはもちろん，実際の研究仮説にとって重要なことです！ しかし，検定の手続きはいずれも，暗に，残差についてもある仮定を置いていて，そこでは ϵ_{ik} が次のようになっているのです。

$$\epsilon_{ik} \sim \text{Normal}(0, \sigma^2)$$

このちょっとした仕掛けがないと，数式がうまく働かないのです。つまり，正確にいうなら，計算して最後に F 統計量を出すことはできますが，F 統計量が実際にあなたが測ろうとしたものをちゃんと測っていたかどうか保証できず，F 検定に基づいて引き出したものが間違っていることになるのです。

さて，では残差についての仮定が正しいかどうかをどうやってチェックしたらいいでしょう？ そうですね，上で述べたように，一つの文章には三つの要素がふめ込まれているので，個別に対応することを考えましょう。

- **分散の均一性**。私たちは母標準偏差 (すなわち σ) について一つの値しか用意してないことに注意してください。各群に個別の値 (つまり σ_k) を考えることもできるのです。これは分散の均一性の仮定として知られています (等分散性ということもあります)。ANOVA は母標準偏差について，すべての群で同じであるという仮定をしているのです。これについてはセクション **??**で大々的に論じます。
- **正規性**。残差は正規分布することが仮定されています。セクション **??**で見たように，これは QQ プロットをみることで (あるいは Shapiro-Wilk 検定をすることで) 検査できます。ANOVA の文脈におけるこれの話は，セクション **??**で論じます。
- **独立性**。独立性の仮定は少しトリッキーです。これが意味することは基本的に，ある残差について知っていても，それは他のどんな残差についてなにも語らないというものです。すべての ϵ_{ik} 値は他のどの残差について，どんな “配慮” もしないし，“関係” も持たないことが仮定されています。これを検証する単純明快な方法というのはないのですが，この仮定を明らかに満たさない状況というのはあるのです。例えば，もしあなたが反復測定デザインをしているとすると，各被験者は二つ以上の状況に晒されるわけですが，この時独立性は保持されていません。いくつかの観測値関係に何らかの関係があるのは，同じ人に対応しているのですから明らかです！ この時は，反復測定 ANOVA(セクション **??**を参照) のような手法を使わなければなりません。

1.6.1 分散の均一性についての仮定をチェックする

分散について予備的な検定を行うことは，波が船の定期便が出向するのに適した状態

かどうかを確認するために，手漕ぎボートで海に出かけるようなものだ。

– George Box (**Box1953**)

諺にあるように，猫の皮を剥ぐ方法は一つではありませんし，分散の均一性の過程を献呈するのもいくつかの方法があります (何らかの理由で，誰もそれを口にしませんが)。私がこの話で見たことのある，最も一般的に使われる検定は，**Levene の検定** (**Levene1960**) で，これは **Brown-Forsythe 検定** (**BrownForsythe1974**) に関わりのある方法です。

標準的な Levene 検定や Brown-Forsythe 検定をやると，どちらであれ検定統計量として F あるいは W で表されるものが出てきますが，これは Y_{ik} の代わりに Z_{ik} を使うだけで，あとは普通の ANOVA で計算される F 統計量と同じやり方で計算されます。これを念頭において，JASP でこの検定をどうやるかを見ていきましょう。

Levene 検定は本当にシンプルなんです。アウトカム変数として Y_{ik} があるとしましょう。新しい変数として，Z_{ik} を定義します。これは群平均からの偏差の絶対値で，次のように定義されます。

$$Z_{ik} = \left|Y_{ik} - \bar{Y}_k\right|$$

オーケイ，こうすることで何がいいんでしょう？ では Z_{ik} が実際に何を意味していて，我々は何を検定しようとしているのかを考えていきましょう。Z_{ik} の値は第 k 群における i 番目の観測がその群平均からどの程度離れているかの測度です。そしてここでの帰無仮説は，群が同じ分散を持っている，すなわち群平均からの全体的な偏差が同じであるというものでした。ですから，Levene 検定における帰無仮説は，Z の母平均が全群で同じであるというものです。ふむぅ。ところで私たちが今知りたいのは，全部の群平均が同じであるという帰無仮説の統計的な検定でした。どこかでみたことありましたっけ？ そうです，これこそ ANOVA で，Levene 検定は新しい変数 Z_{ik} について ANOVA をすることそのものなのです。

Brown-Forsythe 検定のほうはどうでしょうか？ 何が違うんでしょう？ 何にもです。Levene 検定との唯一の違いは，変換された変数 Z の作られ方で，これが少し違います。群の平均ではなく群の中央値からの偏差を使うだけです。つまり，Brown-Forsythe 検定は

$$Z_{ik} = |Y_{ik} - \text{median}_k(Y)|$$

ここで $\text{median}_k(Y)$ は群 k の中央値です。

1.6.2 Levene 検定を JASP で行う

オーケイ，ではどうやって Levene 検定をすれば良いでしょうか。本当に簡単なんですが - ANOVA

の下にある ‘仮定のチェック’ オプション，これの ‘均一性の検定’ チェックボックスをクリックするだけです。そうして結果を見てみると，図**??**に示しましたが，検定結果が非有意 ($F_{2,15} = 1.45, p = .266$) であることがわかります。ですから，分散の均一性の過程は満たされている，と言えそうです。しかし，外見だけでは騙されます！ もしサンプルサイズが大きかったら，ANOVA の頑健性に問題を与えるような分散の均一性の仮定が破られていない時でも，Levene 検定は有意な効果を示しうるのです。これは上の引用にある George Box が指摘したことです。同様に，もしサンプルサイズがとても小さければ，分散の均一性の過程は満たされず，Levene 検定も非有意 (i.e.$p > .05$) になることがあるかもしれません。これが意味するのはつまり，仮定についてのあらゆる統計的検定は，群/カテゴリーごとの平均周りにある標準偏差をプロットしてみないとわからない，ということです... それが似通っている (つまり分散が均一である) かどうかを見ないとね。

Assumption Checks

Test for Equality of Variances (Levene's)

F	df1	df2	p
1.450	2.000	15.000	0.266

Figure1.5　JASP における一要因 ANOVA の Levene 検定結果出力

1.6.3　分散の均一性についての仮定を取り除く

今回の例では，分散の均一性の仮定は大丈夫だったことがわかりました。Levene 検定は非有意 (標準偏差のプロットも見ながら) で，心配することはないようです。しかし，実際にはこんな幸運ばかりではありませんよね。分散の均一性の仮定が破られた時，どうやって ANOVA を救えば良いのでしょう？ ここで t 検定についての議論を思い出せば，この問題に以前出会っていたことに気づきます。Student の t 検定は等分散を仮定していますが，仮定が成り立たない時の解決策は Welch の t 検定を使う，というものでした。実際，**Welch1951** は ANOVA についてのこの問題をどうやって解決するかを示してくれています (the **Welch one-way test**)。JASP にも `One-Way ANOVA` 分析が組み込まれています。Welch の補正を組み込むためには，‘仮定のチェック’–‘均一性の補正’ の下にある ` Welch’ オプションを選択するだけでいいのです。この結果は図 **??**に示しています。

ここで何が起こっているかを理解するために，前のセクション **??**で最初にやった ANOVA で得ら

ANOVA - mood.gain

Cases	Homogeneity Correction	Sum of Squares	df	Mean Square	F	p	η^2
drug	None	3.453	2.000	1.727	18.611	< .001	0.713
drug	Welch	3.453	2.000	1.727	26.322	< .001	0.713
Residual	None	1.392	15.000	0.093			
Residual	Welch	1.392	9.493	0.147			

Note. Type III Sum of Squares

Figure1.6　JASP の一要因 ANOVA の一部としての，Welch's 均一性の補正

れた数字と比較してみましょう。後戻りする面倒を避けるために，前回の最後に得られた数字を書いておきます。：$F(2, 15) = 18.611$, $p < 0.001$ ですね。これは図**??**に示された One-Way ANOVA の Homogeneity Correction の下，None に示されています。

オーケイ，もとの ANOVA では結果として $F(2, 15) = 18.6$ が得られてましたが，Welch の補正をしたら $F(2, 9.49) = 26.32$ になってますね。言い換えると，Welch の補正は群内自由度が 15 から 9.49 に減っていて，結果の F 値が 18.6 から 26.32 に増えています。

1.6.4　正規性の仮定をチェックする

正規性の検定はもう少し直接的です。知っておくべきことはセクション **??**にほとんど書いてあります。やるべきことは，QQ プロットを描く，これだけです[*7]JASP で QQ プロットをするには，'仮定のチェック' に行き，'残差 Q-Q プロット' をチェックします。その結果は図 **??**に示したとおりで，私にはちゃんと正規分布しているように見えます。

1.7　正規性の仮定を取り除く

さて正規性のチェックの仕方を見たところで，正規性が破られた時にどうしたらいいか，と疑問を抱くのは当然ですよね。一要因 ANOVA の文脈では，最も簡単な解決策はノンパラメトリックな検定 (すなわち，確率分布のもつ特別な仮定を一切含まないものに立脚したもの) に切り替えることです。ノンパラメトリック検定については以前，第 **??**章でやりました。二つの群があるときは，

[*7]Shapiro-Wilk 検定も行うべきですが，これは今の JASP に実装されてません。Shapiro-Wilk 検定が有意でなければ (つまり $p < .05$ なら)，この正規性の仮定は破られていないことを示しています。しかし，Levene 検定と同様に，もしサンプルサイズが大きければ Shapiro-Wilk の検定で有意になっても，偽陽性である可能性があります。分析にあたって実質的な問題がないような，正規性の仮定が破られていない状況であっても。そして同様に，とても小さいサンプルでは，偽陰性になってしまうかもしれません。だから目で見てわかる WQQ プロットが重要なのです。

Mann-Whitney か Wilcoxon 検定が，あなたの必要としたノンパラメトリックな代替品を提供してくれます。三つ以上の群があるばあいは，**Kruskal-Wallis 順位和検定** (**KruskalWallis1952**) を使うことができます。そう，これこそ次に説明しようとしているものです。

1.7.1 Kruskal-Wallis 検定の背後にあるロジック

Kruskal-Wallis 検定はある意味 ANOVA によく似ています。ANOVA では Y_{ik} から話を始めました。ここでアウトカム変数は第 k 群の i 番目の人を意味しています。Kruskal-Wallis 検定では，これら Y_{ik} の値を全て順位付けして，順序データの分析をするのです。

R_{ik} を第 k 群の i 番目のメンバーに与えられた順位だとしましょう。ここで $\bar{R}_k$ を計算し，第 k 群における観測値の平均順位を考えます。

$$\bar{R}_k = \frac{1}{N_K} \sum_i R_{ik}$$

そして全体平均 $\bar{R}$ も計算します。

$$\bar{R} = \frac{1}{N} \sum_i \sum_k R_{ik}$$

これで，全体平均 $\bar{R}$ からの偏差平方を計算することができるようになりました。個々のスコアについてこれを計算する，つまり $(R_{ik} - \bar{R})^2$ を計算することで，ik 番目の観測値が全体平均の順位からどれだけずれているかについての，“ノンパラメトリックな” 測度を手に入れたことになります。次に全体平均から群平均の偏差平方を計算する，つまり $(\bar{R}_k - \bar{R})^2$ を算出すると，その群が全体平均の順位からどれぐらいずれているかのノンパラメトリックな測度を手に入れたことになります。覚えておいて欲しいのは，いまから ANOVA と同じロジックを辿っていき，以前やったようにここでの順序の平方和を定義するということです。まず “全体の順序平方和” を計算します。

$$\text{RSS}_{tot} = \sum_k \sum_i (R_{ik} - \bar{R})^2$$

それから “群間順序平方和” を次のように算出します。

$$\begin{aligned} \text{RSS}_b &= \sum_k \sum_i (\bar{R}_k - \bar{R})^2 \\ &= \sum_k N_k (\bar{R}_k - \bar{R})^2 \end{aligned}$$

さて，もし帰無仮説が真で群間にどんな差も認められないなら，群間平方和 RSS_b はとても小さくなり，全体順序和 RSS_{tot} よりもグッと小さくなると思われるでしょう。質的には，こ

れは ANOVA で F 統計量を出そうとした時と同じようなものなのですが，技術的な理由から Kruskal-Wallis 検定統計量は普通 K で表され，少し違った方法で計算します。すなわち，

$$K = (N-1) \times \frac{\text{RSS}_b}{\text{RSS}_{tot}}$$

として，もし帰無仮説が真なら K の標本分散は近似的に $G-1$ の自由度 (ここで G は群の数です) を持ったカイ二乗分布に従います。より大きな K の値が出れば，帰無仮説とより一貫性がないということになります。これは一方向検定です。K が十分大きければ，H_0 を棄却することになります。

1.7.2 補足

前のセクションで書いたのは，Kruskal-Wallis 検定の背後にあるロジックです。概念的なレベルでは，この検定がどういう働きをするのかを考えた方がいいでしょう。しかし，純粋に数理的な側面を考えるのは不必要に複雑です。その導出をして見せようとは思いませんが，ちょっとした代数的ごまかし[a]を使って，K の式が次のように書けることを示しておきましょう。

$$K = \frac{12}{N(N-1)} \sum_k N_k \bar{R}_k^2 - 3(N+1)$$

これは，あなたが K を算出する時にみる数式の最後の形です。この形式は，先ほどのセクションで示したものよりも簡単ですが，全体的に意味をなさないように見えますね。前に示した K の考え方の方が，順位に基づいた ANOVA のアナロジーとして良いように思えます。しかし最後に得られた検定統計量は，元の ANOVA で使われるものから見ると随分違うものに見える，ということは知っておいて欲しいのです。

いやまて，もっとあります！ なぜいつももっとあるんでしょうね？ 今までの話は，ローデータに紐づけられていない時は，常に真なのです。すなわち，同じ値を持つ変数が二つとない場合は，です。もし同順位があれば，この計算に補正項を入れなければなりません。こうなると，もっとも勤勉な読者でさえも気にしなくなったと思います (あるいは，同順位の項は今すぐ注目しなければ，という意見にはならないと思います)。ですから，さっさとどうやって計算するかを示して，なぜこれがこんな風になるのかというつまらない証明はパスしちゃいましょう。ローデータの度数分布表を作ったとして，f_j を j 番目の一つしかない数字だとします。ちょっと抽象的ですから，度数分布表の具体例，`mood.gain` をデータセット `clinicaltrials.csv` から取り出しましょう。

```
0.1 0.2 0.3 0.4 0.5 0.6 0.8 0.9 1.1 1.2 1.3 1.4 1.7 1.8
  1   1   2   1   1   2   1   1   1   1   2   2   1   1
```

この表を見ると，三番目の要素は 2 という数字を持っていることがわかります。これは `mood.gain` が 0.3 というのに対応しているので，二人の気分が 0.3 ポイント上昇したことがわかります。も

うひとつ。先ほど導入した数式的に表現するなら，$f_3 = 2$ ということですね。イェイ。さて，こうなってくると補正項は次のようになります。

$$\text{TCF} = 1 - \frac{\sum_j {f_j}^3 - f_j}{N^3 - N}$$

Kruskal-Wallis 統計量の同順位の値は，K をこの量で割った時に得られます。JASP が計算するのは，この同順位補正版です。やっとこさ，Kruskal-Wallis 検定の理論についての話を終えることができます。Kruskal-Wallis 検定の同順位補正項の計算の仕方を知らない時に感じる不安を取り除けたので，一息つけたのではないかと思います。違います？

[a] jiggery-pokery. 専門用語です。

1.7.3 Kruskal-Wallis 検定を JASP で実行する

Kruskal-Wallis 検定が実際にどうなるのかを理解しようとして大変怖い思いをしたわけですが，この検定を実行するのはごく簡単です。というのも，JASP は ANOVA に ‘ノンパラメトリック’ パートを持っているからです。あなたがすべきことは，グループ変数 `drug` をアクティブボックスに動かすことだけです。そうすると Kruskal-Wallis は図**??**にあるように，χ^2 = 12.076, df = 2, p-値 = 0.002 であることを示します。

1.8 反復測定の一要因 ANOVA

一要因反復測定 ANOVA 検定は，三群以上の間の有意な差異を検定するもので，そこでは各群において同じ実験参加者が使われます (あるいは各実験参加者が他の実験群における参加者と密接に関係がある場合です)。このため，各実験群には常に同じ数のスコア (データ点) があることになります。このタイプの実験デザインと分析は，‘対応のある ANOVA’ とか ‘Within 計画の ANOVA’ とも呼ばれます。

反復測定 ANOVA の背後にあるロジックは，独立した ANOVA(‘Between 計画の ANOVA’ と呼ばれることもあります) と非常に似ています。前のことを思い出して欲しいのですが，Between 計画の ANOVA では全分散が二つの要素，群間分散 (SS_b) と群内分散 (SS_b) に切り分けられ，それぞれを対応する自由度で割ることで，MS_b と MS_w にして (Table **??**参照)，F-ratio を次のようにして計算するのでした。

$$F = \frac{\text{MS}_b}{\text{MS}_w}$$

反復測定 ANOVA では，F-ratio は同じように計算されますが，独立した ANOVA では分母にくる MS_w のもとになった群内分散 (SS) ですが，反復測定では SS_w が二つのパートに分離します。各群で同じ被験者を使いますから，個々人の間にある個人差に伴う分散 ($SS_{subjects}$ と表されるもの) を，群内分散から取り除くことができるのです。これがどうやって計算されるかと言う，技術的な細部についてはこれ以上分け入ることはしませんが，要するに各被験者が被験者要因の各水準になるということです。この被験者内要因の分散はほかの被験者間要因と同じように計算されます。SS_w から $SS_{subjects}$ を引き算することで，SS_{error} の項はより小さいものになります。

$$\text{Independent ANOVA: } SS_{error} = SS_w$$

$$\text{Repeated Measures ANOVA: } SS_{error} = SS_w - SS_{subjects}$$

この SS_{error} の項についての変化は，より強い統計的検定を引き出してくれますが，これは SS_{error} の減少分と誤差項の自由度の減少が相殺しあった上での話 (自由度は (n - k)[*8]から (n - 1)(k - 1) になります。(独立 ANOVA 計画にはより多くの被験者がいることを思い出して))。

1.8.1 JASP による反復測定 ANOVA

まずデータが必要ですね。**Geschwind1972** が言ったように，脳卒中の後に生じる言語障害を正確に把握するためには，ダメージを受けた脳の特定の領域を診断する必要があります。ある研究者は，ブロッカの失語症 (脳卒中のあとに一般的に経験される言語障害) を患っている 6 人の患者が経験したある言語障害を特定することに興味があるとします。

Table1.2　3 つの実験課題において成功した試行の数

患者	スピーチ	概念	文法
1	8	7	6
2	7	8	6
3	9	5	3
4	5	4	5
5	6	6	2
6	8	7	4

患者は三つの言語再任課題をやるように言われます。最初の課題 (スピーチ生成) では，患者は一

[*8](n - k)：(被験者の数 - 群の数)

つの単語が実験者によって大声で読み上げられた後，それを反復するように求められました。第二の課題 (概念) では，言葉の理解のテストで，患者はたくさんの写真とその名前を対応させるよう求められました。第三の課題 (文法) では，正しい言葉の順序についての知識を検証するようなテストで，患者は文法的に正しくない文章を並べ替えるよう求められました。各被験者は全ての課題をやりました。患者が課題をする順番は，患者間でカウンターバランスが取られました。それぞれの課題は 10 回試行されます。各患者が課題に成功した数が，表**??**に示されています。これらのデータは `broca.csv` ファイルにあり，JASP に読み込むことができます。

一元配置反復測定 ANOVA を JASP で実行するには，'ANOVA' から '反復測定 ANOVA' をクリックし，次のように進めてください (図 **??**を参照。),

- 反復測定要因名を入力してください。これは全ての被験者に反復された条件を記述するラベルを，選べるようにするためのものです。例えばスピーチ，概念，文法の課題が全ての被験者に課されたのですから，適切なラベルは '課題' でしょうか。この新しい用意名は，分析における独立変数を意味します。JASP でこれを実行するには，単に 'RM Factor 1' をクリックして，名前を入れるだけです; そこは強調表示されて，新しい名前の入力を待っていると思います–打ち込むだけ！
- 次に '課題' 要因のそれぞれの **水準**名を入れたいと思います。反復測定要因テキストボックスに，三つの水準を追加する必要があることに注意してください。この三つの水準は，三つの課題を意味しています。: `speech`, `conceptual`, and `syntax` です。水準に対応するラベルに変えてください。それぞれをクリックして，新しい名前を入力するだけです。
- それから変数を左のボックスに移動させ，'反復測定のセル' テキストボックスに入れます。変数名がさっき入力した水準名と合致しているか確認してください。
- 最後に，仮定チェックのオプションの下，'球面性のチェック' テキストボックスをクリックします (今はひとまず私を信じて!)

反復測定 ANOVA の JASP の出力は図**??**のようになります。結果を見る前に，Mauchly の球面性の検定をしなければなりません。これは条件間の分散が等しいという仮定を検定するものです (実験条件間の異なるスコアの広がりが，ほとんど同じだという意味です)。図**??**にあるように，Mauchly の検定は有意水準が $p = .720$ です。Mauchly の検定は有意ではなかった (つまり今回の研究例では $p > .05$ だった) ので，分散の間に有意な違いがない (つまり大体等しく，球面性を仮定できる) と結論づけるのが妥当だということになります。

ところで l，もし Mauchly の検定が有意 ($p < .05$) であれば，分散間は有意に異なっていることになるので，球面性の仮定が満たされていないことになります。この場合，一要因 ANOVA で得られた F 値の補正をする必要があります。すなわち，

- "球面性の検定" 表の Greenhouse-Geisser の値が $> .75$ であれば，Huynh-Feldt の補正を使う

べきです。

- Greenhouse-Geisser の値が $< .75$ であれば，Greenhouse-Geisser の補正を使うべきです。

補正された F 値はどちらも，仮定のチェックオプションの下にある球面性の補正チェックボックスにチェックを入れることで得られます。補正された F 値は図**??**の結果の表に示されています。

この分析では，Mauchly の球面性検定の p 値は $p = .720$ (つまり $p > 0.05$) でした。ですから，球面性の仮定が守られていると考えて，F 値の補正は必要ないことになります。そこで，球面性補正の出力にある 'なし' の反復測定の '課題' の値を使います。:$F = 6.93, df = 2, p = .013$ で，スピーチ，理解，文法という各言語課題の成功回数は有意に異なっている，ということができます ($F(2, 10) = 6.93, p = .013$)。

普通，結果を解釈するために記述統計量をレビューするべきです。'追加オプション' のメニューに行き，'課題' を '周辺平均' の下にあるアクションボックスに入れることで，JASP 上でこれらの数値を算出できます。この結果が図**??**にありますが，各条件の平均だけでなく 95%CI も示されています。被験者が達成した課題数の平均を比較すると，ブロッカの失語症はスピーチ (平均 = 7.17) と言語理解 (平均 = 6.17) の課題ではそれなりの成績を上げています。しかし，文法課題ではかなりパフォーマンスが悪く (平均 = 4.33), 事後検定でもスピーチと文法のパフォーマンスには有意な差があります。

1.9 ANOVA と Student の t-検定との関係

さて，ANOVA の話を終わらせる前に一つ指摘しておきたいことがあります。多くの人が驚くと思うんですが，知っておいて損はありません。2 群の ANOVA は Student の t 検定と同じものなのです。いや，ほんと。似ているというのではなくて，実際あらゆる意味で等価なのです。これが常に真であることを証明しようとは思いませんが，一つ具体的な例を示しましょう。我々の `mood.gain ~ drug` というモデルを，ANOVA で実行する代わりに，`therapy` を予測子として使うことを考えてみましょう。ANOVA を走らせると，F 統計量は $F(1, 16) = 1.71$ であり，p 値は 0.21 になります。ここには 2 群しかありませんから，実際には ANOVA に頼る必要はなく，Student の t 検定をすることにします。こうすることで何が起こるかみてくださいよ。：t-統計量は $t(16) = -1.3068$ で p 値は 0.21 になります。不思議なことに，p 値は一致します。もう一度 $p = .21$ になったのです。しかし，検定統計量はどうでしょう？ ANOVA の代わりに t 検定をしたときには，ちょっと違う答えが出ていて，$t(16) = -1.3068$ でした。しかしこれには，かなり直接的な関係があるのです。t 統計量を二乗すると，先ほどの F 統計量を得ます。ほら，$-1.3068^2 = 1.7077$ でしょ。

1.10 要約

この章には多くのことが含まれていますが，まだいくつも取りこぼしたものがあります。わかり切ったことですが，一つ以上のグループ化変数がある状況に興味がある場合の ANOVA をどうやるかについては説明していません。これは第 **??**章で論じることになります。議論してきた内容について，キートピックスと言えば次のようになるでしょう。

- ANOVA がどのように働くかについての基本的なロジック (セクション **??**) と，それを JASP でどう実行するか (セクション **??**)。
- ANOVA での効果量をどうやって計算するか (セクション **??**)。
- 事後の検定と多重比較の時の補正 (セクション **??**).
- ANOVA の仮定 (セクション **??**)
- 分散の均一性の過程をどうやってチェックするか (セクション **??**) と，その仮定が破られたときにどうするか (Section **??**).
- 正規性の仮定をどうやってチェックするか (セクション **??**) と，その仮定が破られたときにどうするか (Section **??**).
- 反復測定 ANOVA (セクション **??**) とそれに等価なノンパラメトリックな Friedman 検定 (セクション **??**)。

この本の全てのチャプターについてそうですが，私が依拠しているいくつかの異なるソースがあります。が，最も影響を受けている一冊を挙げるとすれば，**Sahai2000** です。この本は初心者向けではありませんが，ANOVA の背後にある数学的な側面を理解しようとする，少し進んだ読者にとっては素晴らしい本だと言えるでしょう。

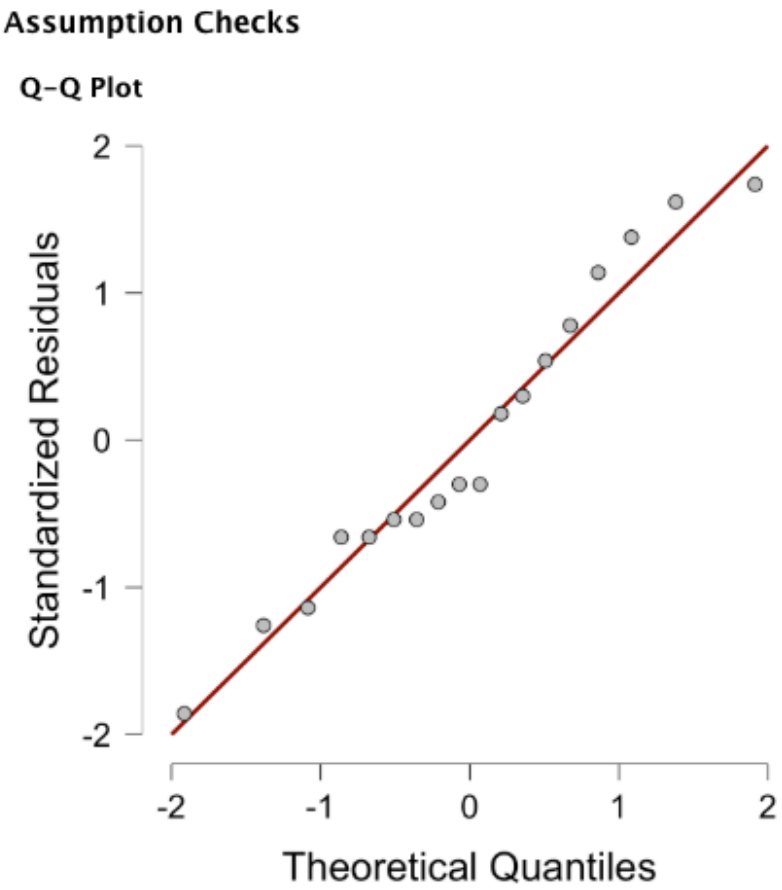


Figure1.7　JASP で作った QQ プロット

Kruskal-Wallis Test

Factor	Statistic	df	p
drug	12.076	2	0.002

Figure1.8　JASP における一要因ノンパラメトリック ANOVA である，Kruskal-Wallis 検定

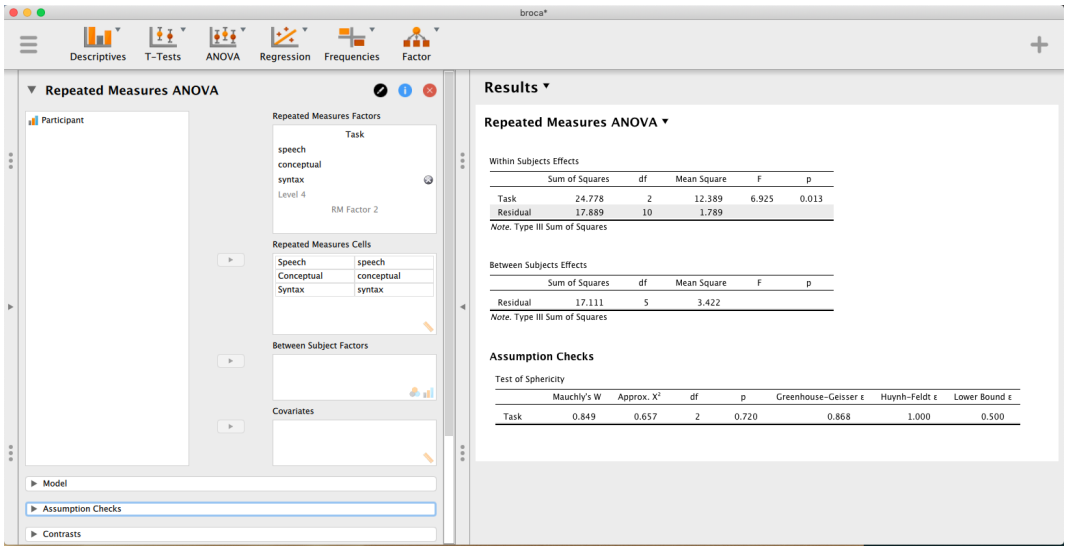


Figure1.9　JASP における反復測定 ANOVA

Assumption Checks

Test of Sphericity

	Mauchly's W	Approx. X^2	df	p	Greenhouse-Geisser ε	Huynh-Feldt ε	Lower Bound ε
Task	0.849	0.657	2	0.720	0.868	1.000	0.500

Figure1.10　一要因反復測定 ANOVA の出力: Mauchly の球面性テスト

Within Subjects Effects

	Sphericity Correction	Sum of Squares	df	Mean Square	F	p
Task	None	24.778	2.000	12.389	6.925	0.013
	Greenhouse-Geisser	24.778	1.737	14.265	6.925	0.018
	Huynh-Feldt	24.778	2.000	12.389	6.925	0.013
Residual	None	17.889	10.000	1.789		
	Greenhouse-Geisser	17.889	8.685	2.060		
	Huynh-Feldt	17.889	10.000	1.789		

Note. Type III Sum of Squares

Figure1.11　一要因反復測定 ANOVA の出力: 被験者内効果の検定

Marginal Means

Marginal Means – Task

			95% CI	
Task	Marginal Mean	SE	Lower	Upper
speech	7.167	0.624	5.825	8.509
conceptual	6.167	0.624	4.825	7.509
syntax	4.333	0.624	2.991	5.675

Figure1.12　反復測定 ANOVA の出力: '周辺平均' ダイアログより，記述統計量が示されます。