

# 心理学論文における数値と統計の書き方

2023年4月18日版

新美亮輔(新潟大学人文学部)

- ※ これは心理学専攻の学生向け資料です。他分野にはあてはまらないことが多数あります。
- ※ この資料は心理学論文の「正しい」書き方を定義するものではありません。大学生・大学院生が心理学論文の「適切な」書式を学ぶこと、卒論等を書く際の参考となることを目的としています。ジャーナルによって書式は異なります。授業レポートでは、担当教員の指示に従って下さい。
- ※ 筆者は統計学の専門家ではありません。もし統計学的な観点から妥当でない記述があれば、ぜひご教示ください。
- ※ 本文中の検定結果や実験計画はすべて架空のものです。

## 目次

### 1 桁数を決めて書く

- [1.1 数値は3桁か4桁にそろえる](#)
- [1.2 「小数点以下〇桁」も意識する](#)
- [1.3 適切な単位を選び、一貫して用いる](#)
- [1.4 略してよいゼロと略してはいけないゼロ](#)
  - [末尾のゼロ](#)
  - [小数点の前のゼロ](#)
- [1.5 計算途中では四捨五入しない](#)
- [1.6 Excelや統計ソフトで数値の最後につく“E+5”とか“e-10”の意味](#)

### 2 統計的検定結果の書式

- [2.1 実例を見るのがいちばん](#)
- [2.2 本文中での統計数値の典型的な書き方](#)
  - [平均値など](#)
  - [z検定](#)
  - [t検定](#)
  - [分散分析](#)
  - [\$\chi^2\$ 検定](#)
  - [マン・ホイットニーのU検定](#)
  - [相関係数](#)
- [2.3 書式の要注意ポイント](#)
  - [何でもかんでもイタリック体にしな](#)

Century フォントは非推奨

半角スペースをたくさん入れる

カイ 2 乗のカイをエックスで代用しない

検定統計量は臨界値（境界値）ではなくデータから実際に算出された値を書く

#### 2.4 $p$ 値の書き方

$p > .05$  とか  $p < .01$  ではなく実際の  $p$  値を書く

いわゆる「有意傾向」について

不等号の向きに注意

大文字 ( $P$ ) か小文字 ( $p$ ) か

有意でないことを表す  $ns$  は本文中では特に必要ない

有意水準を表すときは  $p$  ではなく  $\alpha$

ときどき見かける “ $ps$ ” の意味

#### 2.5 Excel で $p$ 値を求める

$z$  検定（両側検定）

$t$  検定（両側検定）

分散分析（ $F$  値）

$\chi^2$  検定

マン・ホイットニーの  $U$  検定

#### 2.6 有意でなかった結果は書かなくてよいか？

#### 2.7 分散分析表は必要？

### 3 統計的検定結果の説明のしかた

#### 3.1 基本形

1 標本  $t$  検定の報告の例

1 要因 ANOVA の報告の例

#### 3.2 検定結果をわかりやすく書くポイント

検定結果より先に、データの全体像を提示する

検定のデザインを過不足なく記述する

要因名と水準名を使い分ける

統計数値を消しても文が成り立つように書く

有意差のあり・なしだけでなく、差の方向性もわかるように書く

その検定結果は何を意味するのか、言葉でまとめる

検定結果や分析結果が多いときは、図表にまとめる

#### 3.3 複数の検定結果を列挙するとき

カッコに入れ、セミコロンで区切るのが基本

悪い例 (1) 文中での列挙

悪い例 (2) 簡条書き

悪い例 (3) テンプレ文の繰り返し

#### 3.4 検定結果はあくまでも言葉で説明する

#### 3.5 事実と推測を区別して書く

### 3.6 「考察」で何を書けばいいか

どこにも有意差がなかった場合

予想外の有意差があった場合

論文の本題とはあまり関係がないところに差があった場合

### 3.7 発展編：統計的検定のことばかり考えすぎない

## 4 サンプルサイズ決定について報告する

### 4.1 なぜそのサンプルサイズなのか？

#### 4.2 データを取る前にサンプルサイズ設計をした場合

先行研究に基づいて効果量を見積もった場合

中程度の効果量を見積もった場合

研究上の目的から効果量を設定した場合

複数の検定がある場合・多要因分散分析の場合

#### 4.3 他の方法でサンプルサイズを決定した場合

サンプルサイズが予定通りにならなかった場合

#### 4.4 サンプルサイズ不足かも？と思ったら

# 1 桁数を決めて書く

## 1.1 数値は3桁か4桁にそろえる

数値はたいてい3桁か4桁（ときどき2桁）にそろえて書きます。つまり、3桁か4桁になるよう四捨五入します（四捨五入することを英語で **round** というので、「丸める」とも言います）。ただし、**計算途中では四捨五入しません**（[1.5](#)参照）。最終的に論文に書く時にだけ四捨五入します。

- × 123.71064981 → ○ 124 または 123.7
- × -4.198350112 → ○ -4.20 または -4.198

同じ種類の数値は同じ桁数で書くのが原則です。違う種類の数値ならば桁数は違っていてもかまいません。たとえば正答率（%）は3桁、反応時間（ms）は4桁、というのはOKです。

- × 正答率は自転車条件で 68%，飛行機条件で 82.9%だった。
- 正答率は自転車条件で 68.2%，飛行機条件で 82.9%だった。
- 正答率は自転車条件で 68%，飛行機条件で 83%だった。
  
- 新幹線条件の平均正答率は 68.2%，平均反応時間は 1234 ms だった。

## 1.2 「小数点以下○桁」も意識する

データによっては、同じ種類の数値でも3桁の数（862）と4桁の数（1059）が混在することもあります。これはかまいません。無理に揃えることはしません。

- 平均反応時間は通勤電車条件で 862 ms，特急電車条件で 1059 ms だった。
- × 平均反応時間は通勤電車条件で 862.4 ms，特急電車条件で 1059 ms だった。

なぜ 862.4 と 1059 がダメかと言うと、桁数は**データの精度**も考慮して決めるものだからです。通常のPCとキーボードを使った反応時間測定では、1msを下回るような細かい時間は正確に測定できない（誤差の範囲内）と考えられます。それに、たいていの実験では、人間の反応時間の1ms程度の差は小さすぎてあまり意味がありません。ですから、1の位まで書けば十分なわけです。人間の身長を1ミクロン単位で書いても意味がないのと同じです。反応時間でも、たとえば5問の計算問題を解くのにかった時間（平均数10秒）ならば、1秒か0.1秒の精度で書けば十分でしょう（「平均反応時間は32.1秒だった」）。

論文で報告する桁数のひとつの目安は、**測定精度と同じか、測定精度の1桁下**です。たとえば、7件法Likert尺度（1 全く同意しない—7 強く同意する）の評定値は整数（測定精度が小数点以下0桁）なので、その平均値は小数点以下1桁程度で報告します。測定精度や計算精度、実質的に意味があるかを考慮し、小数点以下何桁まで書くか決めましょう。あわせて、桁数が3、4桁程度におさまるような単位を選びましょう。

分散分析の  $F$  や  $t$  検定の  $t$  などの検定統計量は、小数点以上が 1 桁か 2 桁になるのが普通なので、小数点以下 1, 2 桁程度まで書きます。 $p$  値は 1 の位には意味がない（略してしまうことも多い）ので、小数点以下 3 桁程度まで書きます。相関係数は小数点以下 2 桁まで書くことが多いようです。[APA Publication Manual](#)（以下『APA マニュアル』）では、相関係数、 $F$ ,  $t$ ,  $\chi^2$ , Cohen の  $d$  などは小数点以下 2 桁、 $p$  値は小数点以下 2 桁または 3 桁と指定されています。なお、 $r = .40$  のような末尾のゼロを略してはいけません（[1.4](#) 参照）。

### 1.3 適切な単位を選び、一貫して用いる

0.000000538 km/h とか 12501506904 秒 のような数値は桁が多すぎて不便です。桁数が 3, 4 桁になるような単位を選びましょう。日本心理学会の『[執筆・投稿の手びき](#)』（以下『手びき』）は、数値が 0.1 から 1000 の範囲になるように単位を選ぶとしています。

- × 0.000288 km → ○ 288 mm, 28.8 cm, 0.288 m, 0.29 m など
- × 445091 Hz → ○ 445.1 kHz (キロヘルツ), 445 kHz など

実験に用いたモニタの大きさや観察距離など、1 つの実測値を報告する場合は、測定精度そのままで報告します。たいてい、1 mm（ミリメートル）単位で測りますよね（メジャーの目盛は 1 mm 単位）。ですから、288 mm とか 28.8 cm と書くのが適切。一方、従属変数では複数の測定値の集計結果（平均値など）を報告することが多いです。たとえば参加者が描いた 10 本の線の長さを測り、平均値を報告するとします。1 mm（0.1 cm）単位でしか測定していないのに、平均値を 4.568 cm と書くのは過剰です。4.6 cm とか 45.7 mm がよいでしょう。測定精度と同じか測定精度の 1 桁下で書きます。百分率の平均値（平均正答率など）なら、小数点以下 1 桁程度で書くことが多いです（たとえば「62.5%」）。

同じ種類の数値はできるだけ単位を統一しましょう。刺激の大きさがある場所では cm 単位で、別の場所では mm 単位で書いていたりすると、とても理解しづらくなります。

金額や人口など、どうしても大きい桁数で書く必要がある場合は、3 桁ごとに桁区切りのコンマを入れて 14,630,229 と書くと見やすいです（ヨーロッパではコンマでなくピリオドで桁を区切る習慣もありますが、論文ではコンマが標準的）。「億円」とか「万人」を単位にして桁数を減らすのもよいです。自分の研究の目的に照らして最適な表記を選んで下さい。

### 1.4 略してよいゼロと略してはいけないゼロ

#### 末尾のゼロ

論文では、安易に略してはいけません。学校の算数や数学では 12 と 12.0 は同じですが、実験・調査論文では意味が違います。なぜならば、

- 12 cm → 実際の測定値は 11.500... cm から 12.499... cm の間
- 12.0 cm → 実際の測定値は 11.9500... cm から 12.0499... cm の間

という意味だからです。精度が違うのです。もし 1 mm 単位で測定したならば、12.0 cm と書くことには意味があります。同様に、たとえば正答率（正答数 ÷ 全試行数）ならば、

- 0.67 → 実際に算出された正答率の値は 0.66500...から 0.67499...の間  
 0.670 → 実際に算出された正答率の値は 0.669500...から 0.670499...の間

という意味になります。現実のデータを扱うときには、誤差や精度の問題が必ずつきまとうということを覚えておきましょう。

Excel が勝手にゼロを略してしまう場合には、右クリックして「セルの書式設定」→「表示形式」タブで「数値」を選び、「小数点以下の桁数」を指定して下さい。

#### 小数点の前のゼロ

0.12 のような小数点の前のゼロは、割合・確率・相関係数・信頼性係数など定義上絶対に 1 を超えない数値では略してもよいです。1 を超える可能性がある数値では、略さない方がよいでしょう。略してしまうと、脱字と区別がつかないからです。

- $p = .345$     ○  $r = -.67$   
 △  $t(50) = .89, p = .378, d = .08$  → ○  $t(50) = 0.89, p = .378, d = 0.08$   
 (  $t$  や Cohen の  $d$  は 1 を超える可能性がある )

とはいえ、特に古い文献では、 $t$  などでも小数点前のゼロが略されていることは割とあります。また、論理的には 1 を超えることができても実際のデータでは 1 を超えることが少ない数値 (Cohen の  $d$  もそうです) では、小数点前のゼロを略してしまうこともあるようです。

小数点前のゼロは略さなくても OK ですが、心理学論文では略することが多く、『APA マニュアル』は  $p$  や  $r$  では必ず略すとしています。いずれにせよ、略すなら略す、略さないなら略さないで統一すること。ただ、本文中では略さないが図表では省スペースのために略す、というスタイルもあります。

### 1.5 計算途中では四捨五入しない

四捨五入して桁数をそろえるのは、最終的に論文に書くときだけです。Excel で集計しているときなど、計算の途中ではいちいち四捨五入しません。何度も四捨五入をすると、誤差 (いわゆる「丸めの誤差」) が積み重なって最終的な計算結果がかなり変わってしまうことがあるからです。たとえば、 $1400 \div 3 \times 12.5$  という計算を考えます。

#### × 途中で四捨五入する場合

$$1400 \div 3 = 466.6666... \rightarrow 467 \text{ (ここで 1 回四捨五入)}$$

$$467 \times 12.5 = 5837.5 \rightarrow \mathbf{5838}$$

#### ○ 途中で四捨五入しない場合

$$1400 \div 3 = 466.6666... \text{ (四捨五入しない)}$$

$$466.6666... \times 12.5 = 5833.3333... \rightarrow \mathbf{5833}$$

最終的に報告する値が 5 も違っています。実際の集計や分析ではもっと何回も計算が行われますから、このような誤差はさらに大きくなる可能性があります。

厳密に言えば、Excel などのソフトが行っている計算も、こういう誤差を含んでいます。が、コンピュータは大きい有効桁数で数値を扱っている (たいてい 15 桁くらい) 上に、誤

差を減らすいろいろな工夫がされているので、通常の用途では誤差が小さすぎて気にならないだけです。現実のデータの計算には、多かれ少なかれ、常に誤差がつきまとうのです。

## 1.6 Excel や統計ソフトで数値の最後につく“E+5”とか“e-10”の意味

ソフトの出力結果で、ときどき数値の後に E+5 とか+e8 とかついていることがあります。これは、桁数がとても多い数値を表すときの慣習です。意味は下のとおり。

$$1.203E+5 = 1.203 \times 10^5 = 1.203 \times 100000 = 120300$$

$$1.203E-5 = 1.203 \times 10^{-5} = 1.203 \times 0.00001 = 0.00001203$$

つまり指数で表しているわけです (E は指数 **exponential** の意味)。これはあくまでもソフトが表示しやすいだけの便宜的な書き方なので、ソフトからのコピーで論文に 1.203E+5 と書かないように。こんな桁数の大きい値は、単位を変えるなどして小さい桁数に改めます。統計ソフトでは非常に小さい  $p$  値がよくこの指数標記で出てきますが、論文では 0.001 より小さい  $p$  値は単に  $p < .001$  と書けば OK です ([2.4](#) 参照)。人口や金額など、どうしても大きい桁数で書く必要がある場合は、 $9.807 \times 10^8$  のように「 $\times 10^n$ 」式の指数標記にするか、980,700,000 のように桁区切りのコンマ (3 桁ごと) を使うと読みやすくなります。

## 2 統計的検定結果の書式

### 2.1 実例を見るのがいちばん

統計的検定はとて多くの種類があるので (*100 statistical tests* という本があるくらいです), あらゆる検定法の結果の書き方を規定したマニュアルは存在しません。ですから、『手びき』や『APA マニュアル』に加えて, 実際の論文でどう書かれているかを見てみましょう。

[『心理学研究』の最新号の論文を見て下さい](#)。卒論なら, 似た実験・分析をしている先行研究の論文をいくつか読んでいるでしょうから, それらも参考にしましょう。検索では学会抄録などの「論文っぽいもの」もよく出てきますが, プロが校正していないので, たとえ有名な先生が書いたものでも書式の参考にはしない方がよいです。

### 2.2 本文中での統計数値の典型的な書き方

本文中での統計数値類の典型的な書き方を挙げます。あくまでも典型的な例なので, こう書かなければ絶対ダメというわけではありません。また, 図表での書き方は別です。

#### 平均値など

$M = 409.2 \text{ ms}, SD = 55.0$	(平均と標準偏差)
$M = 409.2 \text{ ms}, SD = 55.0, N = 43$	(サンプルサイズを加えた場合)
$M = 3.2, SD = 1.09$	(評定値など, 単位がない場合もある)
$M = 12.3 \text{ cm}, 95\% \text{ CI } [11.9, 12.7]$	(平均とその 95%信頼区間)
$Mdn = 0.49$	(中央値)

平均値を報告する際, 標準偏差をカッコで書き添える書き方もよく使われます。

参加者の年齢は平均 20.9 歳 ( $SD = 1.78$ ) だった。

データ全体の総サンプルサイズとデータの一部のサンプルサイズ (たとえば, 参加者をいくつかの群に分けたときの 1 つの群の参加者数) を区別したいときは, 前者を大文字の  $N$  で, 後者を小文字の  $n$  で表すことがあります (そうでないこともあります)。

#### z 検定

$$z = 2.10, p = .036$$

自由度はありません。両側検定では,  $z$  値はマイナスを略し絶対値を書くことが多いです。

#### t 検定

$$t(42) = 2.56, p = .014$$

$$t(42) = 2.56, p = .014, d = 0.77 \quad (\text{効果量として Cohen の } d \text{ を書く場合})$$

$$t(42) = 2.56, p = .014, d = 0.77, 95\% \text{ CI } [0.14, 1.40] \quad (d \text{ の } 95\% \text{ 信頼区間を書く場合})$$



カッコの中に自由度を 1 つ書きます。ジャーナルによっては自由度を下のように表すこともありますが、心理学分野では少数派です。

$$t_{(42)} = 2.56, p = .014 \quad (\text{心理学ではあまり多くない書き方})$$

$$t = 2.56, df = 42, p = .014 \quad (\text{心理学ではあまり多くない書き方})$$

自由度は基本的に整数ですが、ウェルチの検定（等分散を仮定しない 2 標本の  $t$  検定）では、自由度を小数で書くことがあります。

$z$  検定と同じく、両側検定では  $t$  値のマイナスを略し絶対値を書くことが多いです。効果量  $d$  も負の値がありえますが、両側検定ではマイナスを略して報告しても問題ありません（ $d$  は絶対値が大きいほど効果が大きいことを意味します）。ただし、実験間で効果の方向性が異なるか議論したい場合などは、どちらの条件の平均値が高い（低い）場合を正の（負の） $d$  で表すかを明記しておき、一貫してその表し方を使うとよいでしょう。

対応のない  $t$  検定の効果量としては Cohen の  $d$  が代表的ですが、Hedges の  $g$  もあります（大久保・岡田, 2012, 3.2; 南風原, 2014, p. 62）。1 標本  $t$  検定では  $d_z$  です（Lakens, 2013）。対応のある  $t$  検定では対応のない場合と同じ  $d$  か、または差得点による  $d_z$  が報告されるようです（大久保・岡田, 2012, p. 68; 南風原, 2014, pp. 73–74）。

$t$  検定には種類があるので、どの検定なのかわかるような書き方をしましょう。特に、**対応のある検定をしたら、そのことを明記**しましょう。単に「 $t$  検定を行った」とだけ書いてあると、読者は対応のない検定だと思ってしまうかも知れません。

## 分散分析

$$F(3, 69) = 1.53, p = .214$$

自由度は 2 つあります。自由度の順番には意味があるので、逆にしないように。1 つ目が  $F$  の分子の自由度、2 つ目が  $F$  の分母（誤差項）の自由度です。

$$F(3, 69) = 1.53, p = .214, \eta_p^2 = .06 \quad (\text{効果量として偏イータ 2 乗を加えた場合})$$

$$F(3, 69) = 1.53, p = .214, MSE = 8.78 \quad (MSE \text{ を加えた場合})$$

分散分析の効果量は  $\eta^2$ （イータじじょう）、 $\eta_p^2$ （偏イータじじょう）、 $\eta_G^2$ （一般化イータじじょう）、 $\omega^2$ （オメガじじょう）などです（大久保・岡田, 2012, Lakens, 2013）。心理学論文では  $\eta_p^2$  がよく報告されるようです（Fritz et al., 2012）。どれでなければいけない、という決まりはありませんが、同じ論文の中では統一した方がよいでしょう。このほか、分散分析では  $MSE$ （分散分析表の誤差項の  $MS$  のこと、 $MS_e$  とも）を報告することもあります。 $MSE$  は効果量ではありません。効果量と  $MSE$  の両方を書いてもかまいません。

効果量がきわめて小さい場合（ $\eta_p^2$  が 0.01 を下回るなど）には、

$$F(3, 69) = 0.93, p = .431, \eta_p^2 < .01 \quad (\text{効果量がとても小さかった場合})$$

のように不等号で表現することもあります（ $\eta_p^2$  は負の値にはなりません）。また、 $F$  値が 1 未満の場合には、その効果がないことがほぼ確実なので、

$$F < 1$$

とだけ書いて、詳しい数値を省略することがあります。ただ、このような書き方は、その効果が研究目的上まったく重要ではなく、かつ、検定結果が多すぎて略さないと煩雑な場合だけにしましょう。その効果があるかどうか研究目的上重要なポイントであれば、たとえ  $F$  値が 1 未満でも、 $p$  値や効果量などをきちんと報告すべきです。

なお、有意でも有意でなくても、原則としてすべての主効果・交互作用の結果を書きましょう（交互作用が多いときには一部を省略することもあります）。多重比較をした場合は、多重比較法の名前（Bonferroni 法、Holm 法、Tukey の HSD 法など）も書きましょう。

### $\chi^2$ 検定

$$\begin{array}{ll} \chi^2(3) = 9.01, p = .029 & \text{(自由度をカッコ内に書く)} \\ \chi^2(3, N = 62) = 9.01, p = .029 & \text{(サンプルサイズも書く場合)} \\ \chi^2(1) = 6.35, p = .012, \phi = .25 & \text{(効果量として } \phi \text{ [ファイ] を書く場合)} \end{array}$$

適合度の検定なのか独立性の検定なのかを本文中で明記しましょう。連続性の修正を行った場合は、そのことを書いておきましょう（連続性の修正をしなかったことは特に書かなくても OK）。サンプルサイズは、「方法」で書いてあるなどして明確であれば、書かなくても大丈夫です。逆に、サンプルサイズが異なる検定が立て続けにいくつも出てくるような場合には、その都度サンプルサイズを明記した方が親切かも知れません。

$\chi^2$  検定の効果量としては Cohen の  $w$  があります（Cohen, 1988, Chapter 7; Cohen, 1992）。 $\omega$ （オメガ）と書かれることもあります。独立性の検定（クロス集計表）では、 $\phi$ （ファイ、クロス集計表が  $2 \times 2$  の場合）や Cramér（クラメール）の連関係数  $V$ （ $2 \times 2$  の場合に限らない）が報告されます（大久保・岡田, 2012, 3.5.2; 南風原, 2002, p.187）。 $V$  は  $\phi$  を拡張したものなので、記号  $\phi$  で表すこともあるようです。

### マン・ホイットニーの $U$ 検定

$$\begin{array}{ll} U = 150, p = .533 & \\ U = 151.5, p = .563 & \text{(} U \text{ が } 0.5 \text{ 刻みになる場合)} \\ U = 151.5, p = .563, r = .10 & \text{(効果量として } r \text{ を加える場合)} \end{array}$$

自由度はありません。検定統計量  $U$  は基本的に整数ですが、結び（2 群で等しい値）がある場合、0.5 がつくことがあります。なお、1 つの検定において  $U$  は 2 つの値があります。一般的には小さい方の  $U$  だけを報告します（この場合、 $U$  が小さいほど  $p$  値も小さくなるので注意）。2 つの  $U$  の合計は各群のサンプルサイズの積になります（つまり、 $U_1 + U_2 = n_1 \times n_2$ ）。たいてい、群の順序を入れ替えて統計ソフトに与えると、もう一方の  $U$  が出てきますので、試してみてください。

効果量としては、 $r$  や Cliff の  $d$  が用いられます。 $r$  は、順位（rank）の点双列相関係数（point-biserial correlation,  $r_{pb}$ ）です（データを順位にしたものと、群を 0, 1 でコーディングしたものの間の相関係数）。サンプルサイズが十分大きく、 $U$  検定を  $z$  検定で近似する場合は、検定統計量  $z$  を  $\sqrt{N}$  で割れば  $r_{pb}$  が近似的に求まります。 $r$  として順位双列相関係数（rank-biserial correlation coefficient,  $r_{rb}$ ）を報告することもあるようです。Cliff の  $d$  は Cliff の  $\delta$ （デルタ）と書かれることもあります。論文では、単に  $d$  とだけ書くと Cohen の

$d$ と紛らわしいので、(少なくとも初出時は) Cliff's  $d$ と明記しておくといよいでしょう。

### 相関係数

$r = -.23, p = .041$  (マイナス - はハイフン - と違うことに注意)

$r = -.23, N = 80, p = .041$  (サンプルサイズも書く場合)

$r = -.23, 95\% \text{ CI} [-.43, -.10]$  ( $r$ の95%信頼区間を書く場合)

有意な相関かどうかという検定(無相関検定)の  $p$  値を加えることがありますが、サンプルサイズが大きくなればすぐ有意になるので、この  $p$  値は参考程度に思っておきましょう。効果量を加える必要はありません( $r$ はそれ自体が効果量と言えます)。

報告する相関係数がいくつもある場合は、区別するために、 $r_{HL}$ (High条件とLow条件の間の相関)、 $r_{WM}$ (ある従属変数と作業記憶 working memory 課題の成績との相関)のように記号を作ってもかまいませんが、初出時に必ず記号の意味を説明しましょう。

なお、「相関係数」と名前をつくものは、ここで例を挙げている  $r$ (ピアソンの積率相関係数、もっとも代表的な「相関係数」)以外にもありますので、注意しましょう。

## 2.3 書式の要注意ポイント

### 何でもかんでもイタリック体にしない

数値、カッコ、等号や不等号、ギリシャ文字はイタリック体にしません。イタリック体にするのは、平均の  $M$  や検定統計量 ( $z, t, F$ ) などの統計量の記号、サンプルサイズの  $N$  や  $n$ 、確率の  $p$ 、数式で変数名として使う  $a, b$  などです。ただし、統計量の記号であってもギリシャ文字はイタリック体にしません ( $\chi^2, \eta^2, \alpha, \beta$  など)。

統計量ではない統計用語の略語(分析手法名など)は、イタリック体にしません。たとえばANOVA(分散分析)、GLM(一般化線形モデル)、log(対数)など。統計量である平均の標準誤差(standard error of mean)はSEMとイタリック体、分析手法名である構造方程式モデリング(structural equation modeling)の略SEMは立体(イタリック体でない)です。

このように、何らかの数値を表しうる統計量・スコア・変数等の名前はイタリック体にすることが多いですが、例外もあります。たとえば、『APA マニュアル』ではCI(信頼区間)、AIC(赤池情報量規準)などは立体です。迷ったら論文の実例や『APA マニュアル』(第7版) Table 6.5を参考にしましょう。基本的には、イタリック体にする確信があればイタリック体にし、よくわからなければ立体にしておく、という方針でよいと思います。

### Century フォントは非推奨

Word のデフォルトである Century フォントはイタリック体にならず、ただの斜体にしかならないので、論文には不向きです。イタリック体は、斜体(ax)とは違い、筆記体に近い字形(ax)になります。論文ではイタリック体になる Times New Roman がよく使われます。

△	$r = .70, p < .001$	$df = n - 1$	(Century)
△	$r = .70, p < .001$	$df = n - 1$	(MS 明朝)
○	$r = .70, p < .001$	$df = n - 1$	(Times New Roman)
○	$r = .70, p < .001$	$df = n - 1$	(Century Schoolbook)

### 半角スペースをたくさん入れる

英文や数式は半角スペースがないと非常におかしいです。また、読みづらいです。

×  $F(3,69)=0.93, p=.431, \eta^2<.01$       × 平均速度は  $1.03 \text{ deg/s} (SD=0.04)$  だった。

半角コンマの後、等号や不等号の前後、数値と半角の単位の間などには半角スペースが必要です。スペースを使うのは英語の基本中の基本ですから、半角で書いている時は必ず半角スペースを使って下さい（これは英語文献の表記にも共通の注意点です）。

○  $F(3, 69) = 0.93, p = .431, \eta^2 < .01$       ○ 平均速度は  $1.03 \text{ deg/s} (SD = 0.04)$  だった。

半角スペースが必要な場所を色で示すと、

$F(3, 69) = 0.93, p = .431, \eta^2 < .01$       平均速度は  $1.03 \text{ deg/s} (SD = 0.04)$  だった。

となります。 $F$  とカッコの間は、間隔がおかしくなりやすいので、適宜スペースを入れたり、カーニングを調整したりしましょう（Wordならフォントの「詳細設定」の「文字間隔」）。日本語論文では、統計数値の報告でもカッコだけは全角を使うこともあります。

次のケースでは、スペースを入れません。

- ◆ 数の桁区切りの半角コンマの後      3,000 ms      (× 3, 000 ms)
- ◆ 小数点として使う半角ピリオドの後       $SD = 12.3$       (×  $SD = 12.3$ )
- ◆ 負の数値を表すマイナス符号の後       $r = -.78$       (×  $r = - .78$ )
- ◆ 数値と%または角度の単位 (° ' ") の間      12.3% 45.6°      (× 12.3 % 45.6°)

### カイ2乗のカイをエックスで代用しない

$\chi^2$  を  $x^2$  とか  $x^2$  としてはいけません。これは決して細かいことではなく、「ミ」を「三」で代用するくらいダメな書き方です。必ずギリシャ文字（半角）のカイを使うこと。統計ソフトの出力など、ギリシャ文字が出力できない環境では  $x$  で代用されることがありますが、論文ではダメです。「かい」と打って漢字変換（スペースキー押し）すると全角の「 $\chi$ 」が出てきますから、これを選択してフォントを Times New Roman にすると、半角の  $\chi$  になります。また、半角の  $\chi$  は Word の「挿入」→「記号と特殊文字」から直接入力できます。

### 検定統計量は臨界値（境界値）ではなくデータから実際に算出された値を書く

$t$  や  $F$  などの検定統計量は、実際にデータから算出された値を書きます。臨界値（ $p$  値がちょうど .05 とか .01 になる検定統計量、つまり統計の教科書の後ろの数表に載ってるやつ）は書きません。たとえば、 $t$  検定をして  $t$  値が 3.45（自由度 20）と算出されたとします。有意水準 5% で両側検定の場合、自由度 20 で  $t$  の臨界値は 2.09 で、実際の  $t$  値（3.45）はこれを上回っているので有意と判断することになります。この場合の書き方は、下のとおり。

- $t(20) = 3.45, p < .05$       (データから算出された  $t$  値を書く)
- ×  $t(20) = 2.09, p < .05$       (臨界値を書くのは誤り)
- ×  $t(20) = 3.45 > 2.09, p < .05$       (不等号で臨界値を書くことはしない)
- ◎  $t(20) = 3.45, p = .003$       ( $p$  値もデータから算出した値を書くのが◎)

$t$  の臨界値は調べればすぐわかるので、書く必要はないのです。 $z, F, \chi^2$  などでも同じです。

## 2.4 $p$ 値の書き方

### $p > .05$ とか $p < .01$ ではなく実際の $p$ 値を書く

$p$  値を不等号で書くのは、古い書き方です。統計ソフトが一般的ではなかった時代には、実際の  $p$  値の計算が大変で、検定表で臨界値を調べて検定する方が簡便だったからです。現在では Excel や統計ソフトで実際の  $p$  値が簡単にわかる (2.5 参照) ことが多いですから、データから計算された実際の  $p$  値を書くことが推奨されています。

△  $t(20) = 2.22, p < .05$        $F(2, 34) = 1.23, p > .05$       (有意水準と比べる書き方)

△  $t(20) = 2.22, p < .04$        $F(2, 34) = 1.23, p > .3$       (有意水準ではない)

○  $t(20) = 2.22, p = .038$        $F(2, 34) = 1.23, p = .305$       (具体的な  $p$  値を書く)

×  $t(20) = 2.22, p = .038 < .05$        $F(2, 34) = 1.23, p = .305 > .05$       (くどい)

ただし、実際の  $p$  値が .001 より小さい場合は、不等号を使って  $p < .001$  と書きます。

×  $t(20) = 5.67, p = .0000015$  (0 が多くて読みづらい)

×  $t(20) = 5.67, p = .000$

○  $t(20) = 5.67, p < .001$

×  $t(20) = 5.67, p < .00001$  (心理学論文で用いられる有意水準は最小で .001)

△  $t(20) = 5.67, p < 10^{-3}$  (間違いではないが、心理学では一般的でない表記)

$p$  値は小数点以下 2 桁または 3 桁に四捨五入して論文に書きます。 $p$  値が非常に小さく、Excel や統計ソフトで  $p = 0.00000$  となっていたり、 $5.48E-05$  のような指数表記 (1.6 を参照) になっていたりする場合は、そのまま書いてはいけません。 $p < .001$  と書きます。

なお、複数の  $p$  値をまとめて簡単に報告する際には、そのうち最小あるいは最大の  $p$  値のみを不等号により示して「 $ps > .27$ 」のように書くこともあります。

### いわゆる「有意傾向」について

$p$  値が .05 から .10 の間の場合、「有意傾向」と表現する習慣があります。この表現はダメというわけではありませんが、あたかも「本当は有意である」かのような誤解を生むという 批判もあります。あくまでも、有意でない結果として考察しましょう (4.4 も参照のこと)。

### 不等号の向きに注意

実際の  $p$  値がわからず、不等号で  $p$  値を書く場合は、不等号の向きに注意。逆になっているレポートをけっこう見かけます。 $t$  検定や  $z$  検定の両側検定、分散分析や  $\chi^2$  検定の場合、

$t, z, F, \chi^2$  (の絶対値) が 5% 臨界値より大きいなら、5% 水準で有意なので、 $p < .05$

$t, z, F, \chi^2$  (の絶対値) が 5% 臨界値より小さいなら、5% 水準で有意でないので、 $p > .05$

です。検定統計量 ( $t, z, F, \chi^2$ ) が臨界値より 大きいときは、 $p$  値は有意水準より 小さいという関係です (マン・ホイットニーの  $U$  検定ではこの限りでない。2.2 の説明を参照)。

### 大文字 (P) か小文字 (p) か

心理学分野では小文字の  $p$  の方が一般的です。『手びき』や『APA マニュアル』でも小文字です。大文字で  $t(12) = 3.68, P = .003$  のように書くジャーナルもありますが、レポートや卒論では特段の理由がない限り小文字にしておけばよいでしょう。

### 有意でないことを表す ns は本文中では特に必要ない

統計的に有意でないことを表す略語  $ns$  は、論文の本文中ではあまり使うべき理由がありません。 $ns$  だけだと、具体的な  $p$  値も有意水準もわからず、情報量が少ないからです。図表などでは、省スペースのために  $ns$  を使うことがあります。

- $t(20) = 0.45, p = .658$
- △  $t(20) = 0.45, p > .10$  (具体的な  $p$  値を書いた方がよい)
- △  $t(20) = 0.45, ns$  (具体的な  $p$  値を書いた方がよい)
- △  $t(20) = 0.45, p = .658, ns$  (くどい)

### 有意水準を表すときは $p$ ではなく $\alpha$

5%水準で検定したことを表すために  $p = .05$  と書くのは誤りです。 $p$  は実際にデータから得られた  $p$  値を表す記号です。有意水準を明記したい場合は、

- ……について  $z$  検定 ( $\alpha = .05$ ) を行ったところ、……
- ……か有意水準 5%で  $z$  検定を行ったところ、……

のように、 $\alpha$  (アルファ) で表すか、言葉で説明します。次のような書き方はしません。

- × ……について  $z$  検定 ( $p = .05$ ) を行ったところ、……
- × ……か有意水準  $p = .05$  で  $z$  検定を行ったところ、……
- × ……に統計的な有意差は見られなかった ( $z = 1.23, p = .05$ )。

両側  $z$  検定で  $z = 1.23$  なら、 $p$  値は  $.22$  になりますから、上の例では ( $z = 1.23, p = .22$ ) と書くのが正解。もちろん、 $\alpha = .05$  で有意差はありません。なお、全角  $\alpha$  ではなく半角  $\alpha$  を使いましょう (全角  $\alpha$  を Times New Roman フォントに変更、もしくは「挿入」→「記号と特殊文字」)。

### ときどき見かける “ps” の意味

「ピーエス」ではなく、 $p$  の複数形 (「 $p$  ず」) です。複数の検定結果をざっくり簡単に報告するときに使われます。つまり、 $ps > .05$  と書いてあれば、複数の  $p$  値があったけれど、どれも  $.05$  より大きかったよ、という意味です。たとえば、ある課題の正答率が、参加者の特性を表す 8 つの変数 (年齢, VIQ, PIQ, 作業記憶容量……) のいずれとも有意な相関を示さなかったとき、それが論文の中であまり重要な結果ではなく、8 個の相関係数とその検定結果をいちいち全部書くほどでなければ、「これら 8 つの変数のいずれも正答率との間に有意な相関は示さなかった ( $ps > .05$ )。」とだけ書いてすませることがあります。

同じように、複数の相関係数を  $rs$  と表したり、複数の信頼区間を  $CI$ s と表したりします。ちなみに、 $ps$  でも  $rs$  でも  $s$  はイタリック体にしません ( $p$  と  $r$  はイタリック体)。

## 2.5 Excel で $p$ 値を求める

たいていの  $p$  値は Excel 関数で計算できます (統計数値表に載っている 5% 臨界値を入力して、約 0.05 になるか試してみてください)。赤字斜体の所には、数値を手打ち入力するのではなく、できるだけセルを直接指定しましょう。たとえば  $t$  値が D5 のセルに、自由度が D6 のセルに算出してあるなら、=T.DIST.2T (ABS (D5), D6) です。丸めの誤差を減らすため (1.5 参照)、手動で四捨五入した値は使わないようにしましょう。

### $z$ 検定 (両側検定)

= (1 - NORM.S.DIST (ABS ( $z$  値), TRUE)) \* 2

(ABS () は、絶対値を求める Excel 関数)

### $t$ 検定 (両側検定)

=T.DIST.2T (ABS ( $t$  値), 自由度)

(Excel 関数 T.DIST.2T () は、 $t$  値として負の値を指定するとエラーになるので注意)

### 分散分析 ( $F$ 値)

=F.DIST.RT ( $F$  値, 分子の自由度, 分母つまり誤差項の自由度)

( $F$  値は負になりません。2 つの自由度の順番には意味があるので、逆にしないように。論文で書く順番と同じです。 $F(2, 38)$  なら、=F.DIST.RT ( $F$  値, 2, 38) です。)

### $\chi^2$ 検定

=CHISQ.DIST.RT ( $\chi^2$  値, 自由度)

### マン・ホイットニーの $U$ 検定

サンプルサイズが十分な場合 (森・吉田, 1990, p. 206 によれば各群とも 16 以上) には、正規分布近似により  $z$  検定が使えますから、NORM.S.DIST () 関数で (近似的な)  $p$  値が求められます。詳しくは各種テキストを参照してください (森・吉田, 1990; 芝田, 2021, p. 197 など)。なお、 $U$  検定は JASP で実行でき、 $p$  値も出力されます。

上で紹介している Excel 関数は古いバージョンの Excel では使えないことがあります。古いバージョンの Excel では、NORMSDIST, TDIST など別の名前の (古い) 関数を使います。数値の書き方も違いますから、注意して下さい。

## 2.6 有意でなかった結果は書かなくてよいか?

そんなことはありません。有意でない結果は、有意な結果と同じくらい大切な情報です。「効果がなさそうだ/弱そうだ」という結論も、実験・調査をして初めてわかることです。研究目的上の必要があつて行った検定ならば、きちんと検定結果を報告しましょう。

検定結果が多いときには、一部を省略することもあります。有意かどうかにかかわらず、研究の目的上重要性が低く、読者の関心も薄いと思われる検定結果は省略してかまいません。ただ、有意な結果はいちおう省略せずに書いておく、という傾向はあるようです (これは「有意でないなら省略する」と同じではないことに注意)。また、数値は省略するが有意だった

かどうかだけ書いておく、という手もあります。たとえば「統制群では、長距離条件 ( $p = .037$ ) を除き、すべてチャンスレベルとの間に有意差はなかった ( $ps > .1$ )」といったように。

もし、どの結果は略してどの結果は詳しく書くべきかがわからないなら、自分のデータをきちんと解釈できていないということです。「結果」を書く前に、データを整理して、アウトラインを作りましょう。アウトラインを作ってから、本文を書き進めて下さい。

## 2.7 分散分析表は必要？

通常、論文に分散分析表は載せません。ただ、授業のレポートでは勉強のために分散分析表を載せると指示されることがあります。その場合は、レポートの本文中に表として挿入するか、レポート末尾に付録として載せることが多いですね。なお、論文に分散分析表を載せるときは、Excel や統計ソフトの出力そのままコピペはダメです。体裁を整えて下さい。

### × コピペそのままの分散分析表

Table of Analysis of Variance

source	SS	df	MS	F	p
subject	15606.1333333	9	1734.0148148		
A:Factor A	22210.0666667	2	11105.0333333	3.725	0.0443 *
error[AS]	53657.2666667	18	2980.9592593		
Total	91473.4666667	29			

+  $p < .10$ , \*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .005$ , \*\*\*\*  $p < .001$

### ○ 適切な書式の分散分析表

表 1

実験 1 反応時間の結果についての分散分析表

変動因	SS	自由度	MS	F
参加者	15606	9	1734	
乗り物	22210	2	11105	3.73 *
誤差	53657	18	2981	
全体	91473	29		

\*  $p < .05$



### 3 統計的検定結果の説明のしかた

#### 3.1 基本形

おおむね、下のような順序が基本です(もちろん、実験デザインや前後の文脈との関係で、一部を略したり、順序を変えて書いたりもします)。

- 1) 平均値など検定対象となるデータを報告 (本文中に書くか、図表を参照する)
- 2) 検定の目的をひとことで説明
- 3) 検定のスペックやデザインを記述 (なに検定なのか、要因は何か、など)
- 4) 検定結果の統計数値を記述
- 5) 検定結果を言葉で要約・解釈

##### 1 標本 $t$ 検定の報告の例

(赤字の番号は実際には書かない。なお、検定結果は架空のものです。以下同じ)

- 1) 楽しさ評定の平均値は 3.9 ( $SD=1.1$ ) だった。
- 2) この平均評定値が中立点 (3「どちらとも言えない」と異なっているか確かめるため)
- 3) 1 標本  $t$  検定を行ったところ、
- 4) 有意差があった ( $t(27) = 4.17, p < .001$ )。
- 5) 平均評定値は中立点より有意に高く、サドガエルの声を聴くことは楽しいと評定されたと言える。

##### 1 要因 ANOVA の報告の例

- 1) 映画の内容理解度テストの平均正答率は図 2 のようになった。
- 2) 映画鑑賞時に食べたおやつ (柿の種・ばかうけ・サラダホープ) が平均正答率に影響したか、
- 3) 対応のある 1 要因分散分析により検討した。
- 4) 予想に反しておよつの有意な主効果は認められず ( $F(2, 46) = 1.12, p = .335, \eta_p^2 = .03$ )、
- 5) 内容理解度テストの成績がおやつによって異なるとは言えなかった。

検定結果はまず文で説明し、その後に検定統計量、自由度、 $p$  値、効果量、信頼区間などを付記します。検定結果は上の例のように ( ) 内に書くか、または

1 標本  $t$  検定を行ったところ有意差が認められた、 $t(27) = 4.17, p < .001$ 。

のように、コンマで区切ってから書きます (英語論文では割とよくある書き方)。どちらでもよいですが、同じ論文の中では統一した方がよいでしょう。

#### 3.2 検定結果をわかりやすく書くポイント

##### 検定結果より先に、データの全体像を提示する

論文の「結果」を、検定結果で書き始めてはいけません。まずデータの紹介をします。特に、検定対象である統計量 (平均値など) を、きちんと報告しましょう。図表で示すことが多いですが、量が少なければ本文中に書くこともあります。図表か本文のどちらか一方で報告すればよく、図表に示した平均値を本文中にも書く必要はありません。

### 検定のデザインを過不足なく記述する

どんな検定を行ったのか、読者に追試ができるような情報を書きましょう。対応のある検定であれば、そのことを書きましょう（何も書かれていなければ、読者は対応のない検定だと思ってしまう）。 $\chi^2$  検定では、適合度の検定なのか独立性の検定なのかを明記し、連続性の修正を行ったならばそのことを明記しましょう。 $t$  検定など通常は両側検定を使う検定で、あえて片側検定を使った場合には、そのことを理由つきで説明しましょう。

分散分析では、要因計画をきちんと書きましょう。「手続き」などで詳しく要因計画を説明してあって、それと同じ計画であることが文脈上明らかな場合には、「結果」では簡単に書いておけばよいです（「事前知識と標的物体による 2 要因分散分析を行ったところ…」）。しかし、異なる計画の分散分析を複数報告する場合などは、その都度、ていねいに記述した方が親切です（「事前知識（あり・なし）を参加者間要因、標的物体（ツチノコ・ビッグフット・チュパカブラ）を参加者内要因とする、2 要因混合計画分散分析を行ったところ…」）。

### 要因名と水準名を使い分ける

要因 (factor) や独立変数 (independent variable) は、水準 (level) や条件 (condition) とは違います。要因名 (独立変数名) と水準名 (条件名) には、違う名前をつけましょう。たとえばストループ課題なら、文字色と文字の意味の「一致性」(congruency) が要因名・独立変数名で、「一致」(congruent) と「不一致」(incongruent) が水準名・条件名です。

要因を水準名・条件名の列挙で書くとわかりづらいので、できるだけやめましょう。

- × 一致不一致と教示のありなしの交互作用                   (「一致」や「あり」は水準名)
- 一致性と教示の交互作用                                       (要因名は「一致性」と「教示」)

要因名には、簡潔でわかりやすい数文字の名前を決めて下さい。要因名のかわりに「条件」とか「要因」のような広すぎる言葉を使うと文意があいまいになりやすいので、要注意です。

- △ 正再生数に条件の影響があるか  $t$  検定で検討したところ、…
- × 正再生数に Hz の影響があるか  $t$  検定で検討したところ、…
- × 正再生数に妨害音の周波数が高いか低いかが影響したか  $t$  検定で検討したところ、…
- 正再生数に妨害音周波数の影響があるか  $t$  検定で検討したところ、…

要因名や水準名は「序論」や「方法」で定義し、一字一句変えずにその表記を使います。不用意に表記を変えると読者を混乱させます。従属変数名など他のキーワードでも同様です。用語の表記の一覧表を原稿ファイル冒頭にでも作るとよいです（提出時には削除する）。

### 統計数値を消しても文が成り立つように書く

論文は「文」であり、統計の数値はオマケにすぎません。統計の数値は（ ）かコンマで本文とは区切って書き、統計を読み飛ばしても日本語の文が成立するようにしましょう。

- × …対応のある 1 要因分散分析を行ったが、芋煮会の主効果は  $F(2, 46) = 1.12, p = .335, \eta_p^2 = .03$  となって有意でなく、実験 1 の結果は再現されなかった。
- …対応のある 1 要因分散分析を行ったが、芋煮会の主効果は有意でなく ( $F(2, 46) = 1.12, p = .335, \eta_p^2 = .03$ )、実験 1 の結果は再現されなかった。

### 有意差のあり・なしだけでなく、差の方向性もわかるように書く

単に有意差があった、とだけ書いてあっても、差の方向性（どちらの条件が高かった／低かったのか）がわからなければ、読者は結果をうまく飲み込めません。差の方向性によって、結論はまったく違ってしまうからです（予想通りだったのか、予想と逆だったのか？）。有意差があった場合は、どちら方向の差だったのかわかるよう、言葉で補足しましょう。

△ …*t* 検定を行ったところ、有意差があった ( $t(20) = 2.22, p = .038, d = 0.45$ )。次に、…

○ …*t* 検定を行ったところ、谷川岳条件の反応時間は苗場山条件に比べ有意に長くなっていった ( $t(20) = 2.22, p = .038, d = 0.45$ )。次に、…

### その検定結果は何を意味するのか、言葉でまとめる

検定は研究上の目的があるから行います。検定結果は、単に報告するだけでなく、それが研究目的上どんな意味を持っているのか、言葉で説明しましょう（たとえば、予想どおりの差だったのか、予想に反する差だったのか？ 実験 1 と同じ結果だったのか、異なる結果だったのか？）。

### 検定結果や分析結果が多いときは、図表にまとめる

研究計画によっては、多数の検定を行うこともあります。しかし、検定統計量や *p* 値をただ羅列したり、箇条書きにしたりしてはいけません。本文中で説明すると煩雑な場合は、図表にしましょう。たとえば、相関行列のすべての相関係数の有意性検定の結果や、20 個の質問項目の結果すべてに同じデザインの分散分析を適用した結果（20 個の分散分析の結果）は、表にまとめるとよいでしょう。

## 3.3 複数の検定結果を列挙するとき

### カッコに入れ、セミコロンで区切るのが基本

○ その結果、乗り物と移動時間のどちらも主効果が有意だった（順に、 $F(2, 34) = 18.94, p < .001, \eta_p^2 = .53$ ;  $F(1, 34) = 5.71, p = .023, \eta_p^2 = .14$ ）。

○ …3 要因分散分析の結果、いずれの主効果も有意ではなかった（乗り物、 $F(2, 204) = 1.89, p = .15$ ; 性別、 $F(1, 102) = 1.94, p = .17$ ; 年齢、 $F(2, 102) = 1.04, p = .36$ ）。

セミコロンは必ず半角を使い、セミコロンの後ろに半角スペースを入れます。

×（実験 1,  $r = .04, p = .551$ ; 実験 2,  $r = .12, p = .290$ ）

○（実験 1,  $r = .04, p = .551$ ; 実験 2,  $r = .12, p = .290$ ）

### 悪い例 (1) 文中での列挙

× …3 要因分散分析の結果、乗り物の主効果は  $F(2, 204) = 5.67, p = .004$ 、性別の主効果は  $F(1, 102) = 19.42, p < .001$ 、年齢の主効果は  $F(2, 102) = 4.12, p = .019$  だった。

カッコに入れるかコンマで区切り、できるだけ文末にまとめて列挙しましょう。

- ……3 要因分散分析の結果、いずれの主効果も有意だった（乗り物、 $F(2, 204) = 5.67, p = .004$ ; 性別、 $F(1, 102) = 19.42, p < .001$ ; 年齢、 $F(2, 102) = 4.12, p = .019$ ）。

列挙しないならば、文中にその都度カッコを入れるかコンマで区切り、統計の数値を書きま  
す。このとき、**統計の部分を読み飛ばしても文が成立している必要があります**。「文」が先  
にあって、統計の数値はあとからオマケとして挿入する、と考えましょう。

- ……3 要因分散分析の結果、乗り物の主効果は有意で（ $F(2, 204) = 5.67, p = .004$ ）、性  
別の主効果も有意となり（ $F(1, 102) = 19.42, p < .001$ ）、さらに年齢の主効果も有意だ  
った（ $F(2, 102) = 4.12, p = .019$ ）。
- ……3 要因分散分析の結果、乗り物の主効果は有意で、 $F(2, 204) = 5.67, p = .004$ 、性別  
の主効果も有意となり、 $F(1, 102) = 19.42, p < .001$ 、さらに年齢の主効果も有意だっ  
た、 $F(2, 102) = 4.12, p = .019$ 。

#### 悪い例 (2) 簡条書き

- × ……3 要因分散分析を行った。その結果、  
乗り物の主効果： $F(2, 204) = 5.67, p = .004$   
性別の主効果： $F(1, 102) = 19.42, p < .001$   
年齢の主効果： $F(2, 102) = 4.12, p = .019$   
となった。

統計に限らず、簡条書きは原則としてやめましょう。論文は「文」です。必ず作文して下さい。そして、文は段落になっていなければなりません。

あまりにも検定結果が多い場合には、図表にまとめることも検討しましょう。また、そもそも本当にそんなに大量の検定結果を書く必要があるのか、考えてください。論文の本筋とあまり関連がなく重要でない結果ならば、割愛したり、付録にまとめたりするのも手です。教員は、どれだけたくさん検定をしたかでレポートを評価するわけではありません。

#### 悪い例 (3) テンプレ文の繰り返し

- × ……3 要因分散分析を行った。その結果、乗り物の主効果は有意だった（ $F(2, 204) = 16.31, p < .001$ ）。性別の主効果は有意だった（ $F(1, 102) = 62.62, p < .001$ ）。年齢の主効果は有意だった（ $F(2, 102) = 6.68, p = .002$ ）。乗り物と性別の交互作用は有意だった（ $F(2, 204) = 11.18, p < .001$ ）。乗り物と年齢の交互作用は有意でなかった（ $F(4, 204) = 1.11, p = .352$ ）。性別と年齢の交互作用は有意でなかった（ $F(2, 102) = 0.54, p = .586$ ）。3 者間交互作用は有意でなかった（ $F(4, 204) = 1.58, p = .182$ ）。

これを読んでも、結局どういう結果なのかまるでわかりません。読者に何を理解してもらいたいのかを考えず、ただ検定結果を書き写しているだけです。検定をしたら、いきなり本文を書くのではなく、まず結果を言葉に置き換え、簡条書きなどの形に整理し、本文のアウトラインのメモを作ります。アウトラインに肉付けする形で、本文を書きましょう。

- ……3 要因分散分析を行った。まず、すべての主効果が有意だった（乗り物,  $F(2, 204) = 16.31, p < .001$ ; 性別,  $F(1, 102) = 62.62, p < .001$ ; 年齢,  $F(2, 102) = 6.68, p = .002$ ）。**図2からもわかるように、平均得点は男性より女性の方が、また年齢が高いほど高くなっていました。ただし、乗り物と性別の交互作用が有意で ( $F(2, 204) = 11.18, p < .001$ )、乗り物の効果は性別によって異なることが示唆された。**実際に、乗り物の単純主効果は女性では有意だったが ( $F(2, 204) = 25.67, p < .001$ )、男性では有意ではなかった ( $F(2, 204) = 1.81, p = .166$ )。**つまり、女性では乗り物の種類による好みの違いが出ていたと言える。**これ以外の交互作用は、3 者間交互作用を含めていずれも有意ではなかった ( $ps > .10$ )。**まとめると、性別にかかわらず年齢が高いほど得点は高くなり、女性では乗り物の種類による違いが見られた。**

そこで、女性での乗り物の効果を詳しく検討するため乗り物の単純主効果について下位検定を行うと、……

赤字の文のように、**要所要所で検定結果を言葉で説明しなおす**だけで、段違いにわかりやすくなります。読者の頭の中にグラフを描いてあげるイメージです。また、グラフや表を見ながら検定結果を読むとわかりやすいので、**図表を適宜参照**しましょう。多要因の分散分析で交互作用が出ているなど、検定結果が複雑な場合には、かなり工夫して書かないとわかりにくいからです、読者の身になっていねいに説明しましょう。

あまり重要なポイントでなければ有意でない結果はまとめてしまうなど、検定結果が多いときにはメリハリをつけましょう（ただし、授業レポートでは勉強のため検定結果は省略せずすべて書くように指示されることもありますので、先生の指示に従ってください）。

### 3.4 検定結果はあくまでも言葉で説明する

検定結果の書き方というと、どうしても数値や記号の書き方ばかり気にしがちですが、実は一番大切なのは文です。論「文」ですから、あくまでも文が先にあって、検定統計量や  $p$  値、グラフや表はあとからつけ加えてある、という体裁です。グラフと検定結果だけ書いておしまい、というわけにはいきません。ストーリーをはっきりさせて、筋道立てた説明をすることが重要です。もし結果を言葉でうまく説明できないならば、それは書き手が自分のデータを解釈しきれていないということです。よく頭をひねってください。

それでも、どうしてもやっぱりデータを言葉に置き換えにくいこともあります。中途半端な結果だったり、互いに矛盾するような結果がいくつもあつたり、3 者間交互作用が複雑すぎてどう考えても理解できない場合など。これはもう、仕方がない。どうしてもわからなかったら、**素直にわからないと書けばいい**のです。

- この点については解釈が難しいが、総合考察で議論する。
- 予想に反して3 者間交互作用が有意だったが、実験の目的とは関連が薄く、効果量も小さいので、解釈しない。

こういうのはOKです。すべてを100%鮮やかに説明できる研究なんてめったにありません！ただ、敗北宣言には違いないですね。本当にどうしてもわからない時だけにしましょう。

### 3.5 事実と推測を区別して書く

統計的検定結果の説明をはじめ、論文では、事実と推測をかなり厳密に区別して書き分けて下さい。具体的には、**文末で確からしさのレベルを表現する**ようにします。

- 実験2では、統制キャラクター条件に比べてかわいいキャラクター条件の平均正答率が有意に高かった。したがって、教材に用いられるキャラクターのかわいさが学習効果を高めたと考えられる。キャラクターのかわいさは、学習者の動機づけや教材の誘目性を高めることによって学習を促進するのかも知れない。

下線部分が、確からしさのレベルを表現しています。「高かった。」は断言していて、推測の余地のない事実を表しています。統制条件より正答率が有意に高いという検定結果だったこと自体は間違いのない事実ですから、断言が許されます。一方、「のかも知れない。」は、かなり推測を含んだ表現です。キャラクターのかわいさが動機づけや誘目性を媒介して学習を促進するというメカニズムは筆者の推測にすぎないからです。この説が本当に正しいかは、動機づけや誘目性を直接に操作・測定する別の研究によって検証する必要があります。

では、「と考えられる。」はどうでしょうか。これは、事実だろうと思われるが100%事実とまでは言い切れない（少し推測が混ざっている）ことを表現しています。かわいいキャラクター条件で正答率が高いという実験結果はキャラクターのかわいさそのものが原因ではなく交絡など他の原因によって起こっていた可能性が完全には排除できないからです。

論文では、このような「確からしさのレベル」の微妙な書き分けが求められます。逆に言えば、この種の表現を誤って使うと、論文の科学的妥当性が損なわれます。

- × 実験2では、統制キャラクター条件に比べてかわいいキャラクター条件の平均正答率が有意に高かったと考えられる。したがって、教材に用いられるキャラクターのかわいさが学習効果を高めたと考えられる。キャラクターのかわいさは、学習者の動機づけや教材の誘目性を高めることによって学習を促進する。

専門家がこれを読むと、「検定が正しいか自信がないのか…？」とか、「いや、促進するなんて断言しちゃまずいでしょう、動機づけや誘目性なんて調べてもいないのに…」と感じて、困ってしまいます。

このような表現に明確な決まりはありませんが、少なくとも以下の3つのレベルを意識して書き分けるとよいでしょう（個人的には、レベル3をさらに2つに分ける感覚です）。

#### レベル1（疑いようのない事実）

…だった。…である。…した。

#### レベル2（推測を含むが、事実の可能性がかなり高い）

…と言える。…と考えられる。…と言ってよい。

#### レベル3（推測を含み、事実かは議論の余地がある）

…と推測される。…と思われる。…示唆される。…可能性がある。

…のかも知れない。…ではないだろうか。…でもおかしくはない。

### 3.6 「考察」で何を書けばいいか

実験・調査論文では、統計的分析結果に基づいて議論するのが原則です。と言うと、検定結果を「結果」で書いてしまったら、「考察」で他に何を書いたらいいのかわからない、と言われることがあります。「考察」は「考察」ですから、推測が入ってもかまわないのです。一生懸命頭をひねって、解釈や推測を繰り広げてください。

ただし、ここまでは検定結果から直接言えること、ここから先は推測、とはっきり区別をつけて書くことが大切です (3.5 参照)。また、たとえ推測であっても、できるだけ状況証拠や部分的な証拠 (参加者の内観報告とか、同じような解釈をしている先行研究とか) を挙げるようにしましょう。

#### どこにも有意差がなかった場合

「本当に差がない」可能性と、「本当は差があるのに、何らかの原因で差が出なかった」可能性の 2 つがあります。それぞれ、具体的にはどんな可能性が考えられるか、(できれば間接的な証拠も挙げながら) 考察してください。

後者の可能性はいくらでも考えられます。天井効果や床効果、交絡変数の統制が不十分だった、サンプルサイズが小さすぎた、独立変数の操作が不十分だった、などです。平均値や標準偏差のような集団の値にとらわれすぎずに、個人データや分布の傾向を全体的にながめてみると、何かわかるかもしれません。交絡の可能性については、データを何らかの基準で分けて分析するとヒントが得られることがあります (たとえば、実験の前半と後半に分けてみる、刺激のカテゴリーで分けてみる、ある方略を使っていたことを実験後に報告していた参加者とそうでない参加者に分けてみる、など)。ただし、こういった追加分析は、**有意でなかった差を有意にするために行うべきではありません** (藤島・樋口, 2016)。追加分析の結果は、論文の主要な結論としてではなく、今後考慮すべき問題や将来の研究のヒントとして書くのがよいでしょう (「実際に〇〇の効果があるか検討する実験を行う必要があるかも知れない」等)。また、論文に書く際には、事前の計画にはなかった分析 (ad hoc analysis) であることを明言しましょう (「事後的に分析を試みたところ、…」 「事後の探索的な検討として、…」)。あたかも最初からその分析をする計画だったかのように書くことは避けるべきです。

サンプルサイズは、本来はデータを取る前に先行研究などを参考にして計画するべきなのですが (4.2), 卒論研究ではそれをしていなかったり、計画した数の参加者が締切までに集まらなかったりということもあるでしょう。また、何らかの原因で先行研究より分散が大きく、計画したサンプルサイズでは不十分だったということもありえます。検定力 ( $1-\beta$ ) や効果量を考慮に入れて議論するとよいでしょう (4.4)。統計学的な (標準化された) 効果量だけでなく、従属変数の単位での効果量 (たとえば、条件間の平均値差) を考察することも役立つかもしれません (平均値差が先行研究と同程度なのに標準化された効果量が小さいなら、未知のノイズ要因や交絡変数のせいで分散が大きくなったのかも?)。

いわゆる「有意傾向」 ( $p$  値がギリギリ 5%以下にならず、5%–10%だった場合) の結果を、あたかも効果があったかのようにゴリ押しして書くのはやめましょう。検定では、 $p$  値が有意水準を上回ったらあくまでも「有意でない」という結論です。

### 予想外の有意差があった場合

仮説検証型の研究が多い心理学のレポートや卒論では、すべて予想通りでないといけない気がしてくるものですが、予想外の結果は何も悪くありません。むしろ、**予想外だからこそ実際に実験をした意義がある**というものです。予想外の結果を「実はこういうことだったんだ！」とビシッと解釈している論文は、これぞ科学的発見！という感じがして、カッコイイものです。ぜひ、予想外の結果をどうやったら説明できるか、考えてみてください。

とは言え、この場合もやはり「本当に予想が間違っていた」（実験結果が正しい）可能性と、「本当は予想が正しかったのに、何らかの原因で別の結果になってしまった」（実験結果が間違っている）可能性があります。後者の可能性（未知の交絡変数など）についても、考察をめぐらせてみましょう。

### 論文の本題とはあまり関係がないところに差があった場合

無視せず、ぜひ考察してみてください。たとえば、独立変数を2通りの方法で操作していて、2つの先行研究にならって念のためそうしただけなのだけれども、結果を見るとこの2通りの方法で割と違う結果になっていた、というようなケースです。

論文の本筋と関係が薄くても、何か面白いことを示していないか、未知の現象が背後にないかなど、いろいろと推測してみましょう。新しい研究につながるヒントが見つかるかもしれません。「実は〇〇だった可能性も考えられる。したがって、原因を明らかにするためには、〇〇を統制した上で〇〇を操作するなど、別の検討が必要だろう。」といった調子です。

## 3.7 発展編：統計的検定のことばかり考えすぎない

この資料では触れませんが、結果の説明で最も威力を発揮するのが図表です。わかりやすい図表を作り、効果的に使いましょう。よい論文は、図表だけを見れば、研究の目的・方法・結果がだいたいわかります。統計的検定結果も、必ず本文中に書かなければならないわけではありません。図表に検定結果の情報を盛り込むこともできます。特にデータが多い場合には効果的です。

また、検定だけが統計分析ではありません。心理学分野でも、帰無仮説検定至上主義は長いこと批判されています。検定は便利な一方、さまざまな限界や欠点、落とし穴もあることが意識されるようになりました。検定結果を過度に神聖視しないことも大切です。これは、検定結果を無視して考察してよいという意味ではありません。検定には限界や欠点もあるから、それを考慮せずに検定結果を盲信してはいけませんよ、という意味です。

具体的には、 $p$  値だけではなく効果量や検定力 ( $1-\beta$ ) を考慮したり、検定以外の分析手法（一般化線形モデルやベイズ推定など）を使ったりすることの重要性が指摘されています。たとえば、有意差がないことは厳密には「差があるとは言えない」だけであって、「差がない」ことの積極的な証拠とは言えません。積極的に「差がない」と結論できるかどうかは、ベイズファクター（BF）によって議論できます（[Aczel et al., 2018](#); [岡田, 2018](#); BF の算出は [JASP](#) や R の [BayesFactor](#) パッケージでできます）。大学院進学を考えている方や、研究者を目指している方は、こういった点についても専門書などで学ぶことをおすすめします。



## 4 サンプルサイズ決定について報告する

### 4.1 なぜそのサンプルサイズなのか？

サンプルサイズ（参加者数など）は、論文で必ず書きますよね。一方、どうしてそのサンプルサイズに決めたのかは、これまで心理学論文ではあまり書かれませんでした。しかし近年では、サンプルサイズが妥当だという根拠を報告すること（*sample size justification*）が求められるようになってきました。データを取る前にサンプルサイズ設計（その研究に必要なサンプルサイズの推計）をしていれば、それを報告して下さい。実際には事前に行わなかったのに、あたかも事前にサンプルサイズ設計を行ったかのように装って書くのは厳禁です。

サンプルサイズ設計そのものについては、村井・橋本（2017, 2018）、豊田（2009）、水本・竹内（2011）などを参考にしてください。この資料では、効果量に基づいた帰無仮説検定のためのサンプルサイズ設計の報告について主に説明しますが、実際にはサンプルサイズはさまざまな（場合によっては複数の）理由で決定されるものです（Lakens, 2022）。

### 4.2 データを取る前にサンプルサイズ設計をした場合

最高ですね。胸を張ってサンプルサイズ設計を報告しましょう。「方法」の冒頭あたりに書くことが多いです。参加者の属性や代表性を厳密に考慮する必要がある研究では「サンプルサイズの決定」「サンプリング方法」といった独立の項目を作ることありますが、たいへいは「参加者」等の項目の中でさらっと書くだけでも十分です。

帰無仮説検定のための事前のサンプルサイズ設計では、使う予定の**検定法**、その検定の**有意水準** ( $\alpha$ )、達成したい**検定力** ( $1 - \beta$ )、見積もった**効果量**の情報が重要です。これらを報告しましょう。検定法によっては、自由度などその他の想定されるパラメータを書くこともあります。有意水準は慣例に従い.05とします。検定力は.80がよく使われますが、研究目的によっては.90や.95を使うこともあります。効果量は、使う予定の検定法に合ったものを使います。2標本 *t* 検定なら Cohen の *d*、分散分析なら *f* などです。

このほか、計算に用いたツール（G\*Power, PANGEA, R の [pwr パッケージ](#)等）、停止規則（*stopping rule*）なども説明することがあります。検定結果の報告ほど定式化された書き方のきまりはありません。必要な情報が含まれてさえいれば、基本的には OK です。

#### 先行研究に基づいて効果量を見積もった場合

サンプルサイズ設計をややていねいに報告すると、下のようになります。

（効果量や先行研究は架空のものです。以下同じ）

メタ分析（Osushi, 2019）によれば、寿司おごられ経験による長期的な自尊感情向上の効果量の中央値は  $d = 0.45$  である。この効果を *t* 検定 ( $\alpha = .05$ ) による介入群と対照群の比較で検出するには、検定力.80 でサンプルサイズ 158（各群 79）が必要と推定された。10%程度のドロップアウトを想定し、176（各群 88）を目標とした。

研究計画が他の場所できちんと書かれており、文脈上明らかであれば、

メタ分析 (Osushi, 2019) による効果量  $d=0.45$  を用い、 $\alpha=.05$ 、 $1-\beta=.80$  の下、10% のドロップアウトを考慮してサンプルサイズは 176 (各群 88) とした。

のように簡潔に書くこともできます (ジャーナル論文ではこれくらい短いことも普通です)。最低限、効果量・有意水準・検定力の 3 つの値は書いておきましょう。

メタ分析はないが、いくつかの関連研究の効果量を参考にする、という場合もあります。

こしあん選択行動に対する広告文脈の効果量は、あんパンを用いた越野 (2008) では Cohen's  $w=0.25$ 、おはぎを用いた粒川 (2012, 実験 2) では  $w=0.34$  と算出できた。予備実験 (付録 1 を参照) では  $w=0.29$  であった。これらを参考に  $w=0.30$  を仮定すると、2 (こし・つぶ)  $\times$  2 (文脈あり・なし) の独立性検定 ( $df=1, \alpha=.05, 1-\beta=.80$ ) にはサンプルサイズ 88 が必要なので、外れ値による除外を考慮し 95 名を募集した。

このほか、研究対象の現象そのものズバリの効果量はわからないので、ある分野の研究を広く集計すると効果量はだいたいこれくらい、というメタ・メタ分析を参考にする場合があります。たとえば、心理学研究で報告される現象の典型的な効果量は  $d=0.40$  程度とされま (Open Science Collaboration, 2015)。他にも、社会心理学領域では  $r=.21$  (Richard et al., 2003)、心理療法やアセスメントの研究では  $r=.20-.30$  程度 (Hemphill, 2003) など (ただし、こういった値は出版バイアスを含み、実際はもっと小さい可能性があることに注意)。このような研究を根拠として引用しながらサンプルサイズ設計を説明すればよいでしょう。

#### 中程度の効果量を見積もった場合

参考になりそうな既存データもなく、効果量を見積もることが難しい場合は、Cohen(1988) が提唱する「中程度」(medium) の効果量 (Cohen, 1992, Table 1 にも載っています) を仮定してサンプルサイズ設計をすることがよく行われています。たとえばこんな調子です。

G\*Power 3.1.9.2 を用いてサンプルサイズ設計を行った。牛丼サイズの主効果に Cohen (1988) による中程度の効果量 ( $f=0.25$ ) を仮定し、有意水準 5%、検定力 80% で必要なサンプルサイズとして各群 53 (総計 159) を計画した。

#### 研究上の目的から効果量を設定した場合

研究目的によっては、目標とする効果量を個別に決めてサンプルサイズ設計をすることがあります。たとえば新しい臨床的な介入法の開発が目的の研究では、介入効果の有無を調べるとき、本人も自覚できないようなわずかな症状の改善や、すでに広く使われている介入法より小さな効果を見つけても、あまり意味がないかもしれません。そこで、最低限これくらいの効果があるとわかれば研究目的にとって意味がある、と言えるような目標効果量を決め、それを検出できるようなサンプルサイズを設計します。こういう場合は、その目標効果量の根拠を含めて、サンプルサイズ設計を説明しましょう。

#### 複数の検定がある場合・多要因分散分析の場合

1 つのサンプル (データ) に対して、複数の異なる目的の分析を行う研究計画もあります。その場合は、それぞれの分析について必要なサンプルサイズを推計し、大きい方のサンプル

サイズを採用すればよいので（サンプルサイズ不足を避けるため）、そのことを簡単に説明しておけば大丈夫です。主たる解析と副次的解析がある、というように、分析の優先順位が決まっている場合には、主たる解析に基づいてサンプルサイズを求めることも可能です。その場合は、「主たる解析である 2 要因分散分析に対して、効果量  $d=0.45$  を用い、 $\alpha=.05$ ,  $1-\beta=.80$  としてサンプルサイズを算出したところ…」のように説明しておきましょう。

多要因の分散分析を行う研究も似たような状況になります。2 要因計画なら、要因 A の主効果、要因 B の主効果、A×B 交互作用の 3 つの効果が検定されますので、それぞれについて必要なサンプルサイズを推定することが可能です。この場合、3 つの必要サンプルサイズのうち最も大きいものを採用するという考え方もありますが、研究目的上重要な効果の検定に必要なサンプルサイズを採用するのが素直な方法です。交互作用には研究上関心がなく、2 つの主効果を調べるのが目的なら、それぞれの主効果に必要なサンプルサイズ（のうち大きい方）を採用すればよく、逆に交互作用を検討するのが研究目的であれば、交互作用に必要なサンプルサイズを採用すればよいでしょう。「本研究の目的は交互作用の検討であり、それに必要なサンプルサイズを推定すると…」のように書いておけば OK です。

#### 4.3 他の方法でサンプルサイズを決定した場合

先行研究のサンプルサイズをまねた、現実の事情によってサンプルサイズが限られていた（研究期間が限られている、参加の条件を満たす人がそもそも少ない…）、といった場合です。既存データによる後ろ向き研究でも、サンプルサイズはあらかじめ決まってしまうことが多いでしょう。そういった経緯や理由を素直に書きましょう。

越野（2008）や粒川（2012）にならい、50 名を募集した。

第 67 回ワールドあんこフェスティバル期間中に会場で参加者を募集し、結果としてあん職人 34 名が調査に参加した。

茶道経験者については、著者の所属機関および近隣大学の茶道部に依頼し、縁故法により可能な限り募集した。最終的に 19 名の参加を得た。

実際には事前のサンプルサイズ設計をしていないのに、あたかも事前にそのサンプルサイズを設計したかのように書くのは、研究倫理に反しますからやめましょう（たとえ結果的に十分なサンプルサイズだったとしても。事後的なサンプルサイズの検討については [4.4](#) を参照）。効果量に基づく事前のサンプルサイズ設計をしなかったからと言って、その研究がただちにダメだということにはなりません。たとえば、少数のエキスパートを調べる研究には、サンプルサイズの小ささというデメリットを越える科学的メリットがあるかも知れません。

##### サンプルサイズが予定通りにならなかった場合

予定との多少のズレは問題ありません。設計したサンプルサイズは 80 だが 77 になった、という程度なら、単に「結果として 77 名が参加した」とだけ書けば OK です。

一方で、事前に設計したサンプルサイズと実際のサンプルサイズが大きく異なってしまうこともあるかも知れません。これも、正直に書きましょう。

…80名を目標としたが、結果として、研究期間中に40名が参加した。

…80名を募集したが、想定より多い160名が回答した。

まあ、参加者というのは、思うようには集まらないものです…（じっと手を見る）。ただし、サンプルサイズが予定より少なかった場合は、それを考慮に入れて慎重に結果を考察しましょう。有意な結果が出ても、それは小さなサンプルサイズのせいで偶然生じた大きな効果量のせいかもしれません（サンプルサイズが小さいと、極端な結果になる可能性が高まる）。

サンプルサイズが大きすぎることは、参加者に不要な負担を与えることになり、研究倫理上好ましくありません。また、有意になるまでサンプルサイズを追加する、というのは典型的な  $p$  値ハッキングであり（[池田・平石, 2016](#)）、厳に慎むべきです。過大なサンプルサイズは、サンプリングの停止規則（stopping rule）を事前に適切に作っておけば、基本的には防げます。しかし、オンライン実験で予想を超えて参加が集中し、募集を締め切るのが遅れた、といったことはあるかも知れません。そういう場合は、「方法」や注で正直に事情を書けばよいでしょう。なお、大きすぎるサンプルサイズは、高すぎる検定力を生み、ささいな効果でも検出してしまいます（数万人のサンプルサイズなら、ほとんどゼロのような効果量なのに有意な結果になることも…）。もし有意な結果になったら、実際に得られた効果量（標本効果量）について「結果」や「考察」で議論するとよいかも知れません。

#### 4.4 サンプルサイズ不足かも？と思ったら

データを取った後にサンプルサイズ不足かも？と不安になるのは、実際の研究では往々にしてあることです。具体的には、

- ◆ 事前のサンプルサイズ設計をしなかった場合
- ◆ 予定よりサンプルサイズがかなり小さくなってしまった場合
- ◆ 当然あるはずの効果（先行研究で何度も再現されている効果など）がなかった場合
- ◆ グラフを見ると結構な差がありそうなのに、有意差がなかった場合
- ◆ いわゆる「有意傾向」の結果となった場合

などです。このような場合にどんな分析を報告すべきかは、必ずしも定まっていません。いわゆる事後の検定力分析（post hoc power analysis）、つまり得られたデータから算出した検定力（標本検定力）の報告が勧められることもありますが、現在では、あまり意味がないともされています（[大久保・岡田, 2012](#), pp. 161–162）。

考えられる一つ目の方法は、事後の検討であることを明記した上で、サンプルサイズが十分だったか検討することです。事前のサンプルサイズ設計と同じ方法で、メタ分析などを参考にもともと期待された効果量を見積り（[4.2](#) 参照）、その検出に必要なサンプルサイズを求めます。それよりも実際のサンプルサイズが大きければ（見積もった効果量に対する実際のサンプルサイズでの検出力が.80より大きければ）、結果論ですが、サンプルサイズの正当化になるでしょう。この検討の結果、明らかなサンプルサイズ不足が判明した場合には、サンプルサイズ不足を踏まえた慎重な考察をしましょう（たとえ有意な結果でも、確実な結論が得られたかのようにゴリ押しして書くのは望ましくありません）。

二つ目の方法は、**実際のサンプルサイズで検出できるもっと小さい効果量** ( $\alpha = .05, 1 - \beta = .80$  の下で) を計算し、報告しておくことです。この値を報告すること自体が必ずしもサンプルサイズの正当化にはなりません。読者が実験結果を評価したり、他の研究の結果と比較したりする際の目安にはなるでしょう。