

平成 30 年 10 月 4 日

学会誌「大学のメンタルヘルス」統計データ・記載方法などに関するガイドライン

群馬大学健康支援総合センター  
大学のメンタルヘルス統計担当編集委員  
竹内一夫

このガイドラインは、学会誌「大学のメンタルヘルス」投稿原稿の、統計分析の結果の記載に関して、表記や記述のやり方などに関する標準的な情報およびアドバイスを提供するものである。関連するいくつかの学会の投稿規定やガイドラインを参考として、今回原案を提示しましたが、本学会誌は創刊直後であることから、今後投稿されてくる原稿の内容に合わせて、数年後に見直しを行うことも提案したい。

本学会の性質上、所属大学の学生を対象とした調査研究が中心となるため、概して標本数が大きい（数百から数千人）、一般健康集団を対象とした探索的研究やコホート内ケース・コントロール研究が多い（RCT などはまれ）、といった特徴が予想されるため、こうした研究調査の結果を論文にする際に、特に注意が必要な点を章立てに沿って列挙することとした。

なお、こうした統計処理について慣れていない若手の会員のために（註）を付記した。あくまで初学者のためのメモ書きなので、必要でない方は読み飛ばしていただきたい。

0. まず、統計用語や関連する人名について、投稿原稿中にスペルミスが散見されるため、ネットなどで検索して正確に記載することが推奨される。

1. 「はじめに」：通常は、研究目的の最後に、どういう統計学的な分析を主として行ったかについて一言触れておくことが推奨されている。

（註 1. 例えば、仮説を証明する論文であれば、簡単にその仮説を示し、中心となる検証技法を一言記述しておくことよい。）

2. 「対象と方法」：通常は、この章の最後に「統計解析」や「データ解析」といった名称の小項目を立て、論文中で使用したすべての統計手法を記載しておく。仮説とその検証方法については、ここで具体的に記述しておくことが推奨される。すなわち、どういう目的でその分析を行ったのか、なぜこの手法を選んだのかについて、順に簡単に述べておく。

また、この部分で、どの水準で統計学的有意と判定したかを記載しておく。

（註 2：社会医療系では通常は 5%あるいは 1%が用いられる。本来はどれか一つに統一の方が良いが、慣例的にこの 2つが表示されることが多い。なお、当該分野では、それ

以外の、10%、0.25%、0.1%といった有意水準は採用しない方が良い。P 値の表現方法については「結果」のところでも後述。)

さらに、分析に使用した統計パッケージやプログラムの名称とバージョンを、末尾に明記しておくことが推奨される。ネットから入手したプログラムである場合は、入手元のアドレスと入手した日付を記載しておく。

### 3. 「結果」(統計データの記載方法)

・原則：結果の提示にあたっては、分析ごとに標本数、代表値(平均値など)、散布度(標準偏差など)と言った情報を、図表あるいは文章中に明記して、可能な限り、他の研究者が研究結果を後から検証できるような配慮が求められている。「結果」では、算出された数値や検定結果とそれに対する最小限の記述にとどめ、その解釈に関しては「考察」部分で記述することが推奨されている。

・表示する数値の桁数について：通常は、有効数字より1桁多い形で示されることが多い。ただし、同じ変数であっても、有効数字は測定器具や手段によって異なることもあるため、個々に厳密に規定はしない。

(註3：目安として、身長(cm)、体重(kg)、血圧(mmHg)といった測定変数の平均値は、通常は小数点1位までで示すことが多く、t値やカイ二乗値のような検定統計量や相関係数、因子負荷量などに関しては、通常は小数点以下2桁までで示すことが多い。いずれにせよ、このようにして決めた桁数が、図表中と文章中とで異なることのないように統一をする。)

・検定の実施に関して：原則として、多くの測定変数に同じ検定を一律には実施せず、検定回数を必要最小限とするように求められている。多重比較については成書を参照。

(註4：例えば、有意水準5%で判断すると決めた場合、20回に1回は偶然にその出来事が起こることが前提となる。そのため、20回続けて同様な検定を繰り返した場合、その内1回は偶然に有意と判定されることが予測され、結果の解釈に困難が生じる。この影響を最小限にするため、さまざまな多重比較法が公開されているが、適用に当たっては一長一短があるため、成書を参考に選択することが推奨される。)

・原則として検定は両側検定で実施し、片側検定を行う場合はその理由を明記することが推奨されている。

・有意確率を示すP値について：最近では算出されたP値を、そのまま、小数点以下2桁程度で表記することが強く推奨されている。

(註5：例えば、「有意水準5%のもとで、2群の母平均には差がある(P=0.007)。」と言った表現が推奨されている。)

・ノンパラメトリックな検定方法を使用した場合：比較的発現の少ない健康問題を扱うときにノンパラメトリックな検定方法を用いることがあるが、その際、通常は代表値としては平均値ではなく中央値を、散布度としては標準偏差ではなく四分位偏差を示すことが多い。また、これはパラメトリックな手法においても同様であるが、最近では 95%信頼区間を併記することが推奨されている。

#### 4. 「考察」

・原則：考察部分では、統計分析結果に関する「解釈」の文章が中心となり、t 値や F 値と言った検定に関わる数値そのものは記入しないことが推奨される。

・統計学的有意性と実質的な有意性：統計学的に有意であっても実質的には意味のない差である場合がある。本学会の調査研究では、標本数が概して多く（数十から数千）、統計学的検定を実施した際、測定上ほんの僅かな差異であっても結果において「有意である」と判定されることが多いが、解釈をする際には慎重に検討する必要がある。

（註 6．良く見られる例：1,000 人規模の二つの学部の男子学生の平均身長の違いに関して、その差は 1mm 程度であったが、Student の t 検定を実施したところ、統計学的には有意である、という「結果」が得られた。しかし、健康診断会場での身長の測定の仕方の大雑把さを考えると 1mm 程度の差に実質的な意味はないと考えられるので、「考察」では「差がある」とは解釈しなかった。）

（註 7．余談的な補足です。）

そもそも、データ結果の解釈を統計学検定にのみ委ねることへの強い批判がある。折角苦労して情報量豊富なデータを採取したのに、「 $P < 0.05$ 」といったオール・オア・ナッシング的な検定の結果だけを取り上げて、ある集団が別の集団に比べてある測定値が高い、あるいは、ある変数と別の変数の間に強い関連がある、などと即断することへの危惧が一般的になりつつある。

統計学的検定は、一見して判断がつかないような微妙な違いや関連について、確率論的な立場からのひとつの参考意見が得られるのに過ぎない。解釈に当たっては確率論以外のさまざまな専門的な根拠をも十分に踏まえて行われるべき。

参考までに、「統計的有意性と P 値に関する ASA 声明」日本計量生物学会 (<http://biometrics.gr.jp/news/all/ASA.pdf> 2018 年 2 月 5 日閲覧可能)の一部を以下に転載する。

「結語：すぐれた科学の実践に必須の要素であるすぐれた統計学の実践のためには以下の点を強調しておく。すぐれた研究デザインとその実施という原則、多様な数値およびグラフによるデータの要約、研究対象である事象の理解、背景情報に基づく結果の解釈、す

べてを報告すること、そしてデータの要約の意味の適正な論理的かつ定量的理解。ひとつの指標が科学的推論の代わりとはなりえない。」

上記、一見ただけではわかりづらいので、当学会の研究活動に当てはめて少し具体的に言い換えてみます。

すなわち、単に、データの一部に実施した統計学検定の結果だけから、最終的結論を下すべきではない。得られたさまざまな数値の集計やそのグラフといったものをすべて並べて、それらの測定値が算出されるようになるそもそもの原理原則から考えると、そこに示されている数値の高低は、どのように解釈するのが論理的であろうか、と考へ、さまざまな可能性を列挙した上で、最後に、今回は著者はこういう根拠で、この変数の値の違いについて、あるいはこの変数とこの変数との関連について、こういう意味であろうと解釈した、などと記述する、ということかと。

良く見られる例；数百人におよぶ女子学生の健診時の抑うつ度尺度得点と BMI との関連について、Pearson の積率相関係数  $r$  を算出したところ、絶対値は 0.1 台であったが、統計学的検定上は有意となった。しかし、散布図を見るとまとまりがなく、相関係数の絶対値の小ささから考えると、「関連がある」と結論づけるのは無理があり、むしろ「強い関連は見出せなかった」と解釈した。

参考：<http://aoki2.si.gunma-u.ac.jp/Hanasi/StatTalk/jissainoimi.html>