広島大学大学院教育学研究科紀要　第三部　第51号　2002　171－180

心理学における初歩的統計使用の要注意事項集

松田文子・三宅幹子[1]・橋本優花里
山崎理央・森田愛子・小嶋佳子[2]

(2002年9月30日受理)

Notices about using elementary statistics in psychology

Fumiko Matsuda, Motoko Miyake, Yukari Hashimoto
Rio Yamasaki, Aiko Morita, and Yoshiko Kojima

Improper uses of elementary statistics that were often observed in beginners' manuscripts and papers were collected and better ways were suggested. This paper consists of three parts: About descriptive statistics, multivariate analyses, and statistical tests.

Key words: Descriptive statistics, Multivariate analyses, Statistical tests, Improper uses
キーワード：記述統計，多変量解析，統計的検定，不適切な使用

卒業論文，修士論文，博士論文，投稿論文をチェックする立場にあるものは，くりかえし同じような注意を学部生，院生，投稿者に与えることになる。その中には統計にまつわることも多い。筆者らはいずれも統計学の専門家ではないので，高度の誤用について論ずる能力はないが，ごく一般的な初歩的誤用や配慮不足については，チェックした経験とチェックされた経験の両方を豊富に有している。そこで論文を作成する学生の自己チェック用に役立つような，初歩的統計使用の要注意事項集を作成することにした。もちろん以下に述べることは原則論であるので，一律にあてはまらないことは言うまでもない。その点についても紙面の許すかぎり書きこみたいと思っている。

記述統計上の注意

素データ(測定値)集団の特徴を，より効率的に記述して理解しやすくするために行う数的要約を，記述統計と呼ぶ。これは統計的処理の基本である。理解しやすい効率的記述という目的に添わない記述統計は避けなければならない。

代表値と散布度　平均値や中央値といった代表値を文中，表，図に示すときには，標準偏差や四分領域といった対応する散布度を出来るだけつけるようにし，結果の読みとりに際しても，一度は散布度を考慮することを習慣にすると良い。ただし，表や図に散布度をつけるときは，表や図が見づらくならないように気をつける必要がある。

測定値の尺度水準　名義尺度，順序尺度，間隔尺度，比尺度の定義的区別は，心理学における統計学の初歩で学び，それぞれの尺度水準で使用可能な統計量や検定法についても教わる。それにもかかわらず，評定尺度法(たとえば，1．大変○○である～5．まったく○○でない)で得られた数値に対し，ためらいもなく間隔尺度上の測定値に対してしか求められない統計量(たとえば平均値)を求めていることが多い。これは，間隔尺度の場合，検定法等がもっとも進んでいて，多様でスマートな分析が可能になるためと思われるが，一言「便宜上，間隔尺度上の数値と見なし」と断り，自分がラフで大胆なことをしているという自覚を持ちたいものである。高度の分析を必要としないのであれば(たとえば，代表値と散布度，および一要因の代表値の差の検定の程度)，その尺度にふさわしい統計量を求めるほうが望ましい。

有効数字について　有効数字への配慮の足りない論文原稿は多い。多くの場合，コンピュータが算出した意味のない小さな位の数値まで書く，という不注意で

[1]萩国際大学
[2]旭川大学女子短期大学部

ある。数値の有効性は，素データと素データに対して統計的処理を行った結果の両方について，考える必要がある。

まず，素データ(測定値)であるが，これは測定器具の精度と心理学的な有意味性とによって決まる。たとえばある刺激をある時間(標準時間)提示した後，その提示時間と同じ長さと思う間だけ参加者にキーを押させることにより，標準時間に対する再生時間を測定したとしよう。そのタイマーの精度がmsの単位まで信頼できるものであれば，測定器具精度の点からは1msまで有効であるが，個人内変動や個人間変動の観点からあるいは標準時間の精度から，100ms以下の変動は心理学的に意味がないということであれば，測定値は1/10sまで有効とすればよい。逆に，調査における評定尺度法の場合，ある態度について0から100まで1きざみのデータを心理学的には得たいと思っても，測定器具としてのヒトは多くの場合そのような精度を持っていないので，せいぜい5段階や7段階の評定値を有効とせざるをえない(実際には5段階や7段階に識別することも，調査内容や評定者によっては難しい)。

次にそうして得られた測定値を処理した時の有効数字についても気をつける必要がある。このことについて松田(1991b)は，次のように要領よく述べている。

> 「39人の児童の平均発言時間」を求める場合，39人の全発言時間を39で割る。割り切れないからと小数点以下たくさんの数字を並べることは，全く不要，無意味なことで，せいぜい生の測定値の有効な位(今の例では1/10秒)の一つ下の位の単位(1/100秒)まで書けばよい。たまたま39で割ると，ぴったり割り切れて14秒であった場合も14.00秒と記して，他と有効数字の位をそろえなければいけない。一般に平均値($\Sigma X/N$)は，元の測定値の有効数字の位より，一つ下の位まで求める。それは，総和を求める(ΣX)ことにより，しばしば有効数字の桁数がふえ，以下に述べるように枚挙数(N)(一つ，二つと数えられる数。上例では児童数39)で割っても有効数字の桁数はかわらないからである。たとえば，小数点第一位まで有効な測定値1.1，4.2，5.1，3.4，2.8，5.4の平均値を求める場合，まずΣX＝22.0を算出する。この有効数字は3桁である。したがって$\Sigma X/N$＝22.0/6＝3.666…の有効数字は3桁であるから，平均値は3.76となり，元の測定値より一つ下の位まで有効となっている。
>
> 平均値を求める場合にみたように，元の測定値がaの位まで有効である場合，このような測定値のいくつかを加えたり引いたりしても，aの位まで有効であり，枚挙数を乗除しても有効数字の桁数は変わらない。
>
> aの位まで有効である数値とbの位まで有効である数値を加えたり引いたりした場合は，位の大きい方までしか有効でない。たとえば15.48＋112.4＝127.88であるが，有効なのは，小数点第一位までであるから，小数点第二位は四捨五入して127.9とする。
>
> n桁有効な数字とm桁有効な数字の積や商では，有効数字の桁数の少ない方と答の有効数字の桁数を一致させる。12.5×3.1＝38.75であるが，有効数字2桁にして39とする。n桁有効な数字の平方根もn桁とすればよい。
>
> ただし，最終的な統計量を求める前にいくつかの演算をする場合は，演算の途中では，有効数字より3桁または4桁余分の数字をとり，最終の結果の表示の所で，有効な位にまで四捨五入するのが安全である。(p.172-173)

加減乗除による有効数字の位や桁数の変化は，測定値は真値の近似値であり，たとえば測定値が15.0(小数点第一位まで有効)のとき，真値は14.95～15.05の間にあると推定されることによる。上例の15.48＋112.4の真値は一番小さいときは15.475＋112.35＝127.825であり，一番大きいときは15.485＋112.45＝127.935でありうるから，小数点第一位は8か9であるという意味を持っているが，小数点第二位は0から9まで何でもあり得て意味がない。乗除についても自分で試してみれば，上記のことが納得できるだろう。

さて，このように平均値の場合，有効数字の桁数が一般に増える(素データが整数であれば，平均値は小数点第一位まで示すことが多い)。その誤った般化ではないかと思うのだが，100未満の枚挙数のパーセンテージを出すときも，小数点第一位まで示している例は非常に多い。平均値の場合と異なり，パーセンテージの場合は有効数字の桁数は素データと変わらないことに注意しよう。Figure 1 は，人数をパーセンテージに直したとき，実際にはふえない有効数字の桁数を1桁増やしてしまっている例である。

表　記述統計量を表にまとめることも多いが，わかりやすい表を書くことは結構難しい。表には縦線はまず使用せず，横線を基本とする(これは主に印刷費の節約から出てきた制約なので，著者の書いた表が図と同様そのまま写真製版される場合は，このことをあまり気にする必要はない)。表の中の見出しは，カテゴ

Table 4

発生原因別に見た対人葛藤の発生回数と第三者の行動の有無

	物の取り合い	不快な働きかけ	ルール違反	イメージ	仲間入り	偶発	トラブル	計
発生回数	38	29	14	52	18	10	2	163
第三者あり	15	12	8	18	13	3	2	71
割合（%）	39.5	41.4	57.1	34.6	72.2	30.0	100.0	43.6

※「割合」＝「第三者あり」／「発生回数」

Figure 1. パーセンテージで表したとき，有効数字の桁数を増やしてしまった例（広島大学心理学研究第1号，2001より）。

A

Table 1

記憶セットの種類と大きさによる反応時間と誤答率の違い

記憶セット条件	文字		色パッチ		
	L1	L4	C1	C2	C3
反応時間（ms）	579.8	671.7	709.7	731.8	781.6
（*SD*）	105.6	128.4	193.0	182.3	172.4
ミス（%）	0.0	0.6	39.2	45.4	46.8
（*SD*）	0.0	0.8	19.0	20.7	18.2
フォールス・アラーム（%）	0.3	1.3	12.1	19.4	24.2
（*SD*）	1.1	2.5	6.1	13.5	11.1

B

	文字		色パッチ		
測度	L1	L4	C1	C2	C3

C

Table 1

記憶セットの種類と大きさによる平均反応時間と平均誤答率の違い（（　）内は *SD*）

記憶セット	反応時間(ms)	ミス(%)	ﾌｫｰﾙｽ･ｱﾗｰﾑ(%)
文字			
L1	579.8(105.6)	0.0(0.0)	0.3(1.1)
L4	671.7(128.4)	0.6(0.8)	1.3(2.5)
色パッチ			
C1	709.7(193.0)	39.2(19.0)	12.1(6.1)
C2	731.8(182.3)	45.4(20.7)	19.4(13.5)
C3	781.6(172.4)	46.8(18.2)	24.2(11.1)

Figure 2. 表の見出しの付け方。
Aは，表の見出しの好ましくない例（広島大学心理学研究第1号，2001より）。
Bは，見出し部分を修正したもの。Cは，もっとも比較したい数値を縦に並べたもの。

リーごとにまとめ，同一カテゴリー内の下位の見出しをカバーする範囲に横線を引く。たとえばFigure 2のAの場合は，Bのほうが望ましいだろう。このほうが，文字，色パッチという見出しの下に，それぞれL1とL4，C1とC2とC3の見出しがあることがはっきりする。また，Aの表の記憶セット条件のように，横向きに内容を示す見出しは，通常使わない。表題，見出し，表の注などで，必要があれば明らかにする。

表の数値は横よりも縦に比較するほうが一般に易しい。したがって一番比較したいものが縦に来るようにする。Figure 2 のAの表の場合，Cの表の方が見やすくないだろうか。なお，表中の数値の小数点の位置

がそろうように気をつける。

折れ線グラフ　折れ線グラフは，横軸の変数に対する縦軸の値の連続的変化を示すものであるから，通常横軸が間隔尺度か比尺度の場合に用いるべきである(ただし知能検査や性格検査の因子ごとの得点についての全体的パターン，すなわちプロフィールをみる場合は，名義尺度に対しても折れ線グラフを使用する場合も多い)。したがって横軸が名義尺度の場合は棒グラフ等の方が望ましい。Figure 3 の左は横軸が名義尺度であるとき折れ線グラフを用いた例である。この場合，横軸上の順番が一義的に決まらないので，折れ線の上昇，下降の意味も一義的に決まらない。横軸が順序尺度の場合は APA マニュアル(APA, 2001)にも折れ線グラフが見本に載っているように，かなりよく使用されるが，折れ線の上昇，下降の勾配の意味が一義的に定まらないことに注意する必要がある。

イタリック指定　英文字であらわされる統計用語は一般にイタリックで書かれる。このことは初心者にも比較的良く知られているので，平均値としての M や分散分析の F はまずイタリックになっているが，総数を示す N や n がイタリックになっていないことは多い。総数が統計用語であると感じられにくいからであろう(Figure 4 は，確率 p はイタリックにしているが，人数の n をイタリックにしていない例)。なお N と n を使いわけるときは，N は全体の総数，n は全体の中にいくつかの群がある場合の各群の数を示すのに用いる。

小数点を含む数値　小数点を含む数値で，その値が理論上±1の範囲を超えないものである場合(たとえば，確率 p，相関係数 r，因子負荷量)，小数点の前の0を省略することが多い。もちろん付いていても誤り

(a)読み取りの深さ　読み取りの深さの結果は FIG. 5に示されている。分散分析の結果，挿入学習の効果のみが有意であった ($F=4.39, df=1/54, p<0.05$)。t 検定の結果，黙読条件では関連有条件の方が直後条件よりも有意に成績がよく ($t=2.25, df=18, p<0.05$)，音読条件においても，関連有条件の方が直後条件よりも有意に成績がよかった ($t=2.22, df=18, p<0.05$)。

(b)読み取りの正しさ　読み取りの正しさの結果は FIG. 6に示されている。分散分析の結果，挿入学習の主効果のみが有意であった ($F=4.56, df=1/54, p<0.05$)。t 検定の結果，黙読条件では関連有条件の方が直後条件よりも有意に成績がよく ($t=2.25, df=18, p<0.05$)，音読条件では，関連有条件及び関連無条件の方が，直後条件よりも有意に成績がよかった ($t=2.33, df=18, p<0.05$; $t=2.59, df=18, p<0.05$)。

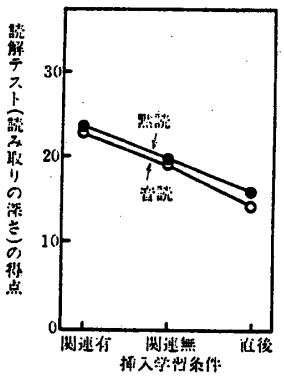

FIG. 5　各条件における読解テスト（読み取りの深さ）の成績

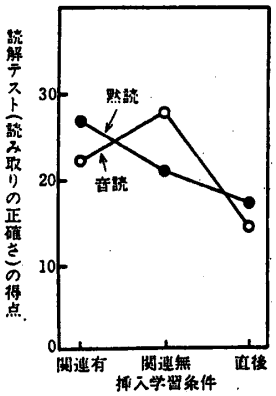

FIG. 6　各条件における読解テスト（読み取りの正確さ）の成績

ライアン法による多重比較を行った。さらに，本研究では各カテゴリーにおける各性に特徴的な出現傾向を明らかにするため，学年の主効果のみが有意な場合でも性別における学年の単純主効果の検定を行った。

However, it had been predicted that the effect of the encoding status (i.e., the priming effect) would be different in the study conditions, and indeed the interaction between study condition and encoding status approached statistical significance, $F(2, 90) = 2.32$, $p = .10$. Thus, a planned 2 (test task) × 2 (encoding status) ANOVA was conducted for each study condition.

Figure 3. 左：横軸が名義尺度であるのに折れ線グラフを用いている例であり，かつ交互作用が有意でないのに，断りなく下位検定を行っている例(教育心理学研究第28巻，1980より)。
右：交互作用が有意でないときも単純主効果の検定を行う理由を述べている例(教育心理学研究第47巻，1999，および *Japanese Psychological Research,* Vol.43, 2001 より)。

ではないが，ない方が表の場合であればすっきりするし，それで失われる情報もない。

統計ソフトの設定によっては，浮動小数点表示(仮数部と指数部の2組の数字を用いて数を表示する方式)によるデータの入力や結果の出力が可能である。3.814E3 = 3.814×10^3, 3.814E−3 = 3.814×10^{-3} の意味なので，出力時にEをerrorの略号と勘違いして，うろたえないように。

多変量解析における注意

複数の変数について，それらの相互関係を同時に考慮して，より効率的に結果を記述し，理解しやすくするために行う数値要約を，特に多変量解析とよぶ。

相関関係と因果関係　独立変数を人為的に操作し，しかも剰余変数を完全に統制していない限り，独立変数と従属変数の間の因果関係を明らかに出来ない。したがって調査研究は基本的に相関関係しか明らかに出来ない。しかし，私達は多くの場合因果関係を明らかにしたい(科学するとはものごとを説明することであり，説明するとは事象間の因果関係を明らかにすることである)。したがって，過去100年間の固定電話の一般家庭の普及率と虫歯の罹病率の相関係数の高さから「電話線を使って虫歯菌が移動する」というような因果関係を推測しがちである。しかし，2つの事象(XとY)間の相関関係でさえ，「XがYの原因」，「YがXの原因」，「ZがXとYの両方の原因でXとYの間には直接的には関係がない」，あるいはこれらの3つのうちの2つ以上の複合が考えられる。ましてや多くの変数の相互関係にもとづく多変量解析の場合，その解析法がモデルとして因果関係を想定しているか否か，想定している場合，自分が当てはめた変数の因果関係は論理的・文献的・常識的説得性を有しているかを十分吟味する必要がある。想定していない場合は，結果の解釈に因果関係を持ち込むときに，それを妥当とする他の論拠を必要とする。

結果の解釈　多変量解析は一般に多人数から多くの変数についてデータをとり，さまざまな推定を行って，かなり多数の未知数の解を得る。したがってその計算過程は一般に複雑であり手計算で行うことはまずなく，解析プログラムを自作することもまずなく，多くの場合，既製のソフトを使用することになる。逆に言えば，解析の基礎となる理論も計算手順も何も知らなくても，素データをソフトにほうりこめば結果は出てくる。これは便利なことであるが，かなり危険なことでもある。たとえば，因子分析で因子負荷量が負であることを考慮しない因子の命名を行い(「期待通りの授業であった」「触発されて本や資料を調べた」「将来の教師の仕事に役立つ」の3項目に高く負に負荷している因子を「有効性」と命名)，それ故結果の解釈がまったく逆になっている例がある(「学生から見た教職科目の授業の評価と改善に関する研究」プロジェクト，1997)。また千石(2001)は，因子分析とクラスター分析の結果の誤った解釈の興味深い例となっている(著者の知名度の高さと新書版という出版形態からみて，影響力の大きさを危惧する)。あたりまえのことだが，解析の基礎となる理論や仮定はよく理解しておかねばならない。

先述のように多変量解析は一般に求める未知数が多い。そこで求めた解の一部のみを示すほうが表がすっきりする場合も少なくないが，他方，それは結果の解釈の妥当性についての読者による判断の手がかりを少なくすることでもある。たとえば，Figure 5のような因子分析結果の書き方では(因子負荷量絶対値が.50以上の項目のみを提示)，因子の命名(因子の解釈)の妥当性を他因子への負荷量や捨てられた項目内容(30項目中10項目が捨てられていると思われる)から吟味することは出来ない。

表2　操作確認の平均値（標準偏差）

	必要サポート		
必要サポート操作確認	高条件（n=13）		低条件（n=12）
予想得点	7.30 (.75)	<	8.50 (1.62) †
自信	1.54 (.52)	<	2.12 (.58)***
必要サポート[1)]	2.18 (.77)		2.19 (.73)
安静時脈拍数	78.19 (9.65)		74.0 (8.93)

***p<.01，†p<.10

1）得点が高いほど必要サポートが高いことを示す.

Figure 4. 英文字の統計用語をイタリックで示す際に，nをイタリックにし忘れた例（広島大学心理学研究第1号，2001より）。

Table 1　中学生の自己実現の因子分析結果

質問項目	抽出因子 I	II	III	IV	V	VI	VII	h^2	M	SD
I 自分自身の肯定										
20 理想像を求めることが悪い結果をもたらすこともある。	-.69							.48	3.51	1.23
22 私はいつも相手の気持ちが気になる。(R)	.67							.56	1.93	1.15
1 私は過去を後悔することが多い。(R)	.59							.58	2.26	1.33
12 私は少しぐらい批判されても，あまり自信を失わない。	.51							.56	2.93	1.39
II 自分の弱点の受容										
28 私は他の人からの批判を素直に聞ける。		.71						.57	3.22	1.13
13 私は自分の弱さを素直に認めることができない。(R)		.63						.56	3.22	1.32
14 私は，自分の中にある矛盾を受け入れることができる。		.59						.40	3.28	1.09
17 私は完璧でなくても満足できる。		.56						.48	3.33	1.50
VII 他者の両極的評価										
16 私にとって，面白味のない人もいる。							.74	.43	4.33	1.01
3 私は友人に対して友好的な感情だけでなく，イヤな感情も出している。							.51	.46	3.15	1.19
寄与率(%)	12.05	8.33	8.04	7.22	5.81	5.47	5.09			
累積寄与率(%)	12.05	20.38	28.42	35.64	41.45	46.92	52.01			

注 (R)は逆転項目である。

Figure 5. 主要因子以外の因子負荷量や捨てられた項目の結果を示していない因子分析の例（広島大学教育学部紀要第49号，2000より）。

統計的検定における注意

統計的仮説検定の問題点　記述統計の結果の解釈のための1つの客観的指標として，統計的検定の結果がしばしば利用される。この目的に沿わない統計的検定の結果の利用をしないように注意する必要がある。もっともよく用いられかつ方法論的に発展している統計的検定は，帰無仮説の棄却を目指す統計的仮説検定である。しかしこの方法には重大な問題点が3つある。1つ目は，母集団からのランダムな標本抽出を前提にしている点である。しかし実験にしろ調査にしろ，標本のランダム抽出が行われていることはまずない。したがって，用いた標本から母集団を逆に良心的に(!)想定するという方法を，多くの場合採らざるを得ない(松田，1991a)。近くにいる大学生10人に実験に参加してもらった結果を，全人類に当てはめていないだろうか。近隣の協力してくれた5学級の担任と児童の結果を，日本の全学校の全学級に当てはめていないだろうか。母集団の想定，すなわち結果の一般化の範囲には，十分注意する必要がある。2つ目は，総数が大きくなれば，心理学的な有意味性とは無関係に，帰無仮説は必ず棄却されて統計的に有意味になる，ということである(この例は後に示すが，Cohen, 1994，とそれに対するHubbard, et al., 1995，のcommentはこのことの理解に大変役立つ)。3つ目は，統計的に有意でなかったときは結論保留であり，統計的には何もいえない，ということである。

以上のような事実は，統計的仮説検定の結果を過大に評価すべきでないことを示している。統計的仮説検定は，偶然でもかなり生じ得るようなわずかの差や傾向から過大な結論を導かないための歯止め，という程度の役割であると考えればよい。初心者はとかく検定結果を絶対視しがちである。もちろん上記のような問題点を克服する試みもなされているが(たとえば，南風原・芝，1987；橘，1997)，主流は依然として統計的仮説検定であるので，以下それについての要注意点を述べる。

検定法の選択　どのような検定法を用いるかを決める際には，検定の目的，測定値の尺度水準，標本が互いに独立かそれとも連関があるか，検定の対象となる記述統計量の種類，要因数，要因内の条件数等を考慮しなければならない。そして，ある検定目的に対して複数の検定法が適用可能なことも多い。明らかに検定力の低いものは除くとして，とりあえずいろいろな方法を試みるのがよいだろう。たとえば，ある性格特性(実験変数)がある作業量にどのような影響を与えるか，を100人の大学生について調べたとする。特性の強さを示す得点と作業量の相関係数を求め，有意に0より大きいかどうか検定することもできる。特性の強さが極端な場合だけ影響がありそうであれば，特性値の両極の人を25%ずつ抽出して(もちろん，別な抽出法も考えられる)，両群の平均作業量を求め，t検定することもできる。ただし，論文にするときは，冗長性を避けるために，検定結果を精選する必要がある。

TABLE 1 渡航前のCMIに基づくI・II・III群の間における各自覚症指標の平均値に関する差の検定結果 (†: p<.10, •: p<.05, ••: p<.01, •••: p<.001)

CMI実施時期

	CMI区分	渡航前	渡航直後	1か月後	2か月後	4か月後	6か月後
I・II群の比較結果	A				†		
	B	***	**	*	†		
	C	*	*	†			
	D	***	**	*			
	E						
	F	*			†	*	*
	G	**	†				
	H						
	I	***	*	*	**	†	†
	J	**		†			
	K						
	L	**	*				
	CIJ	***	**	*	*		†
	A-L	***	***	*	*	†	*
	M	***	**	*	**	**	*
	N	**	*		†	†	
	O	*	*			*	
	P	***	*		*	*	
	Q	*	†		†	*	†
	R	***	*	**	*	***	*
	M-R	***	***	*	**	***	*
I・III群の比較結果	A						
	B	***	**	†			
	C	†	**			*	*
	D	*	*		*	**	**
	E	**	***	**	***		
	F	*	***	†		*	**
	G	*	***	*	***	***	***
	H	†	**	*	*		*
	I	***	*	**	**	***	***
	J	***			**		*
	K						
	L	*				**	
	CIJ	***	**	***	†	***	*
	A-L	***	***	***	*	*	***
	M	***	***	***	***	***	***

Figure 6. 記述統計量も検定統計量もなく，帰無仮説が棄却できる有意水準のみを示した例(平均値は各群別，実施時期別に，CIJ, A-L, M-Rのみ，別に図示されている)(教育心理学研究第34巻，1986より)。

記述統計との関係　統計的検定は記述統計量の解釈に際して，1つの客観的手がかりをあたえるものであるから，検定統計量のみ書いて記述統計量を示さない，ということは出来るだけ避けたい。極端な場合には，Figure 6のように，記述統計量がないだけでなく，F値やχ^2値といった検定統計量もなく，ただただ何%の有意水準で帰無仮説が棄却できるのかのみ書いてある場合がある。このような書き方は，平均値の差の検定であれば，有意水準の高さが差の大きさと一対一対応するという誤解を読者に与えやすいし，多分書き手がそのように誤解している場合も多いのだろう。何%の有意水準で棄却できるかは，差の大きさだけでなく，分散の大きさやデータ数が影響する。したがって，検定結果を書くときには，検定の対象となる記述統計量はもちろん，検定方法と検定統計量もできるだけきちんと書く必要がある。なぜなら，統計的検定には検定法それぞれに前提条件があり，また推定値を用いており，絶対的なものではないからである。はたしてその統計的検定が妥当なものであるかどうか，という判断をするための最小限の情報として，検定方法と検定統計量は出来るだけ読者に提供すべきである。

検定結果の書き方　検定結果の書き方には絶対的なものはないが，執筆規定で指示されている場合はそれに従い，それがない場合は，一般に流布している様式に従うのが，読者には分かりよい。F検定であれば，$F(1, 117) = 4.71, p<.05$のように書き，Fの( )内の数値は，検定対象の要因の自由度(df_1)と誤差項の自由度(df_2)である。$F = 4.71, df = 1/117, p<.05,$ のように書くこともある。t検定では，$t(60) = 1.99, p<.05,$ のように書き，( )内が自由度である。χ^2検定では，$\chi^2(4, N = 90) = 10.51, p<.05,$ のように書き，χ^2の( )内の最初の数値が自由度である。確率は.10, .05, .01, .001がよく用いられるが，不等号の記号を用いないで$p = .03$のように書いてもよい。帰無仮説を棄却する有意水準をあらかじめ定めて文中に記しておけば，いちいちp値を書く必要はなく，有意であったか否かのみ記せばよい。たとえば，「以下，検定はすべて5%の有意水準で行う。•••••。○○の主効果が有意であった($F(1, 117) = 4.71$)。」

なお，2つの統計量の比較をする場合，通常は両側検定を行うが，片側検定を行う場合は，そのことを必ず記す必要がある。片側検定の場合，両側検定に比べて帰無仮説を棄却しやすい(差が有意になりやすい)が，片側検定をするには，それを妥当とする理由が必要であるし，それを記す必要がある。

交互作用と単純主効果の検定　実験変数が複数ある時は，1要因の分散分析をいくつも繰り返すのではなく，分散の違いなどの問題がない限り，交互作用が明らかになるように，多要因の分散分析を1つ行う。そして，原則的には交互作用が有意であったときに，その意味するところを明らかにする客観的手がかりを得るために，単純主効果等の下位検定を行う。したがってFigure 3の左のように，交互作用が有意でない時，なんのことわりもなく，さらに条件別の検定を行うのはまずいだろう。ただし，それが研究の主目的の条件であり，しかも別な群では交互作用が有意で単純主効果の検定をしているとか，10%水準ならば交互作用が有意である，というような場合，特にことわりを述べて行うことは，それが結果理解のための有効な情報を提供するなら，許されるのではないかと思う(Figure 3

の右。ただしこれを許さない人もいる)。

多重比較 3条件以上の代表値を含む要因の検定を行って，その要因の主効果が有意であった場合は，どの代表値とどの代表値の間に有意差があるかという多重比較を行う。多重比較には様々な方法があるので，研究目的と研究方法に適したものを用い，文中にはその方法名を必ず記す必要がある。主効果が有意であったのに，同じ有意水準で多重比較を行ったとき，どの対にも有意差が出てこない，ということは理論的にも実際にもあり得る。なお，要因中の条件が2つの時は，主効果が有意であるということは2条件間の代表値の間に有意差があるということであるから，多重比較を行う必要はない。多重比較については，森・吉田(1990)の説明が分かりよい。

分割表とχ^2検定 分割表の独立性の検定には一般にχ^2検定が用いられる。各セル内の度数は相互に独立でなければならない。「分割表＝χ^2検定」の結びつきが強いために，分割表になっているとなんでもχ^2検定をしてしまいがちである。たとえば，事前テストの正答・誤答と事後テストの正答・誤答を組み合わせた2×2の分割表に，テストを受けた40人の児童をふりわけてχ^2検定した場合は，独立性の検定であるから，事前テストの正・誤と事後テストの正・誤に連関があるか否かを検定しているのであって(事前テストで正答の人は事後テストでもやはり正答のことが多いということがあるかどうか)，事前テストから事後テストへ正答者がふえたかどうかの検定ではない。しかし，正答者数(率)の変化の検定のつもりでχ^2検定をするという誤りをしがちである(このような場合はこの分割表の中の，事前テストが誤答で事後テストは正答の人数と事前テストは正答で事後テストは誤答の人数を比較するマクニマーの検定などの方法がある)。

各セル内の度数が独立でない分割表にもχ^2検定をしてしまうことがある。Figure 7は，5人の参加者が，理論モデルの下で発言すると予測される条件下(Predicted Mention)で実際に発言した回数(Actually Mention)と発言しなかった回数(Actually Skip)，および理論モデルの下で発言しないと予測される条件下(Predicted Skip)での両者の回数を示している。これを2×2の分割表(セル内の数値はTotalの13，4，8，34)にして$\chi^2(1, N = 59) = 22.0$を求め，有意であったのでモデルの正しさを検証した(Predicted Mention条件下では Actually Mention が多く，Predicted Skip条件下では Actually Skip が多い)と論文中にある。しかし各セルの値(発言回数)は5人の参加者の発言回数を合わせたもので，参加者内と参加者間の要因が交絡し，独立なデータになっていない(このような参加者内データと参加者間データをいっしょにしてしまう誤りは，χ^2検定にかぎらず，他の検定でも初心者に時々みられる誤りである)。χ^2検定を行うためには，これが59人のデータでなければならない。(この5人のデータであれば，各人の2つの条件下での言及率を求め，角変換して対応のある場合のt検定が可能であろう。)

統計的検定の結果と心理学的結果 何度も述べるようだが，統計的検定は，記述統計量の解釈に際しての，1つの客観的な手がかりである。統計的に有意(significant)であることが即，心理学的に有意味

TABLE 4
Model-Data Fit for Subject R Solving Eight Problems

	Predicted Mention		Predicted Skip	
Prob. No.	Actually Mention	Actually Skip	Actually Mention	Actually Skip
1	1	2	1	1
2	1	0	2	3
3	2	0	0	4
4	1	0	0	5
5	2	0	0	8
6	3	0	1	2
7	3	0	3	2
8	0	2	1	9
Total	13	4	8	34

Figure 7. 5人の被験者の総発言回数59回にもとづいて誤ってχ^2検定を行った例(*Cognitive Science,* Vol.14, 1990より)。

表4　授業方法と児童生徒の授業態度

		小	中	
授業方法	「よくある」「わりとある」の割合の合計			
	1時間中、先生の話を聞いている授業	31.8	37.5	***
	先生が質問をして、自分たちが答える授業	67.8	73.6	***
	グループで話し合う授業	59.9	46.8	***
	クラス全員で話し合う授業	49.6	20.7	***
	自分の課題を考えながら進めていく授業	41.0	26.5	***
	先生に言われたドリルやプリントをする授業	68.0	69.5	***
授業態度	「とてもそう」「わりとそう」の割合の合計			
	わからないときに、先生に質問しやすい	53.0	53.5	***
	先生に教えてもらう時間をとりやすい	39.4	36.1	***
	学習に集中できない	31.6	34.7	***
	発表がしやすい	48.3	46.3	***
	みんなが授業でかつやくできる	49.8	40.0	***
	むだ話が多い	62.0	70.8	***
	先生との会話が多い	53.7	56.8	***
	学習が楽しい	65.5	53.2	***
	授業中に手紙などを書いても先生に気づかれない	29.4	43.7	***

統計上の検定の結果、＊は5％水準で、＊＊は1％水準で、
＊＊＊は0.1%水準で有意であることを示す。以下の表も同様。

Figure 8. 被調査者の人数が多いため，検定結果が心理学的に有意味な情報をもたらしていない例(小学生4026人，中学生3405人)(広島大学大学院教育学研究科紀要第50号，2001より)。

(significant)であるわけではない。帰無仮説の棄却をめざす統計的仮説検定においては，参加者の人数が多くなれば，どんなわずかな差でも有意となる。たとえば，Figure 8 は小学生4,026人，中学生3,405人が 4 段階評定をしたとき，上位 2 つの評定を行った人数の%の小学生と中学生の差の検定(多分 χ^2検定)の結果を示しているが，わずか0.5%の差でも0.1%水準で有意になっている。すなわち，この検定は心理学的に意味のある情報をほとんどもたらしていない。

また逆に，統計的に有意でないことが，かならずしも心理学的に有意味でないことを意味しない。要因A(条件 a_1 と条件 a_2)と要因B(条件 b_1 と条件 b_2)の主効果と交互作用を 2 要因の分散分析で調べたところ，交互作用が有意であり，単純主効果の検定の結果，条件 a_1 では要因Bの効果が有意であり，条件 a_2 では要因Bの効果が有意でない，という結果が出た場合，「条件 a_1 と条件 a_2 で要因Bの効果が異なる」という解釈をする前に，条件 a_1 と条件 a_2 の下での 2 つの平均値とその差をじっくり眺めてみよう。どちらもよく似た傾向だが，条件 a_1 の方が少々その傾向が強いだけなのかもしれない。逆に，要因Aの主効果のみ有意であった場合，「条件 b_1 と条件 b_2 で条件 a_1 と条件 a_2 効果はかわらない」という解釈も慎重にしなければいけない。「有意でない」ということは統計的には結論保留である。しっかりと記述統計量を見てもう少し参加者を増やせば，別の明瞭な結果が得られるのではないか，と考えてみる必要もある。あるいは，少し視点を変えて，別なデータの扱い方や検定法を工夫してみることも大切である。

【引用文献】

APA 2001 *Publication manual of the American Psychological Association.* Washington, DC: APA

Cohen, J. 1994 The earth is round ($p < .05$). *American Psychologist,* **49**, 997-1003.

「学生から見た教職科目の授業の評価と改善に関する研究」プロジェクト　1997　学生から見た教職科目の授業の評価と改善に関する研究　平成 8 年度教育学部教育研究特別経費報告書　広島大学教育学部

南風原　朝和・芝　祐順　1987　相関係数および平均値の差の解釈のための確率的指標　教育心理学研究 **35**, 259-265.

Hubbard, R., et al. 1995 Comment. *American Psychologist,* **50**, 1098-1103.

松田文子　1991a　簡単な統計　松田伯彦・松田文子(編著)　新版　教育心理学研究法ハンドブック－教師教育のために－　北大路書房

松田文子　1991b　論文・レポートの書き方　松田伯彦・松田文子(編著)　新版　教育心理学研究法ハンドブック－教師教育のために－　北大路書房

森　敏昭・吉田寿夫　1990　心理学のためのデータ解

析テクニカルブック　北大路書房
千石　保　2001　新エゴイズムの若者達－自己決定主義という価値観－　PHP 研究所
橘　敏明　1997　確率化テストの方法－誤用しない統計的検定－　日本文化科学社