

(教育講座)

スポーツ健康学の科学論文で使えるパラメトリック統計： js-STARを使って簡単に実践してみよう！

Useful method of parametric statistics for a scientific article in sports and health sciences:
practice easily using js-STAR!

加藤 雄一郎*

Yuichiro KATO*

1. はじめに

スポーツや身体活動で起きる様々な現象を評価するとき、そのパフォーマンスの差異やトレーニング効果などをみるには、客観的なデータを計測する必要がある。例えば、立ち幅とびで跳んだ距離、脚の膝関節伸展筋力、最大酸素摂取量、ボールのあての正確性などである。トレーニングや練習の効果を評価するためには、その客観的なデータの結果に基づいて判断しなければならない。しかし、実際の指導現場においては、コーチ自身の経験則に基づいて主観的に善し悪しを判断しがちである。例えば野球の打撃パフォーマンスを向上させるためには、ゴロを打つ練習ばかりするのではなく、ライナーやフライを織り交ぜながらバリエーションに変化をもたせた練習をする方が効果が高いことがわかっている。このことを多様練習の効果という (Catalano & Kleiner, 1984)。専門家による主観的な経験則は、正しい場合もあれば、誤っている場合もある。その危険性をできる限り最小限にして、確率論的に正しい判断ができるように科学的根拠 (エビデンス) を用いる必要がある。

科学的な思考では事象の判断基準に客観性が求められるが、事象を客観的に測定したとしても、そこには必ずばらつき (誤差) が存在することになる。つまり計測機器を用いて測定さえすれば、その違いを平均値だけで評価したら良いと考えるのは誤りとなる。そして、その誤差を考慮するために「統計」の使用が必然となるのである (市原, 1994)。科学論文を執筆する際には、測定した客観的なデータをもとに、生起する現象が何と関連性を持つのか、どの条件間で差異が生じているのかを導き出さなければならない。しかし、その客観的な数値にも誤差 (偶然誤差) があるということを理解しておく必要がある。「統計」を実施する目的は、このような誤差が生じ

る状況のなかで、抽出した標本から母集団を推測し、より一般的な結論を導き出すことにある。

抽出した標本からのデータから2つのグループの平均値に差がありそうだと考えるとき、母集団の平均値も同じように差があるであろうということを推定する検定方法を統計的仮説検定 (statistical hypothesis testing) と言う。このとき科学論文の中で頻繁に出てくる用語がp値であり、このp値がどれくらい小さければ平均値に差があると判断できるかを設定しているのが有意水準5%である。例えば、2種類の持久的トレーニングの効果をみるためにAというトレーニングを10名 (A群) に、Bというトレーニングを10名 (B群) の集団に実施したとしよう。トレーニング開始前の最大酸素摂取量が両群とも同じであったとして、トレーニング終了後にA群の平均値は55 mL/kg/minであり、B群の平均値は53 mL/kg/minであったとする。この場合、トレーニングAの方が有効と言えるのであろうか？平均値で比較すれば確かにA群の方が大きな値となっているが、トレーニングした10名全員が同じような変化を示すことはほとんどない。そこには必ずといっていいほど個人差がある。この個人間の変動を考慮して、確率論的に両群の間に差異があるのかどうかを検定しなければならないということになる。検定の結果が $p < 0.05$ であれば「差がある」と判断することになる。これは帰無仮説である「A群とB群の最大酸素摂取量が同じである」ことが5%未満の確率でしか起こらないことを意味している。このように得られたデータ間に意味のある差があるかないかを検定するのが統計分析ということになる。

そこで本稿では、著者の研究グループによる測定データを参照しながら、基本的な統計手法を紹介する。使用する統計ソフトはインターネット上にフリーで公開され

*平成国際大学スポーツ健康学部

E-mail:y.kato@hiu.ac.jp

ているjs-STAR（注1）である。js-STARは、ウェブブラウザを用いて数値データをコピー&ペーストすれば簡単に統計値を計算してくれるため、パソコンのOSの違いによる動作要件がなく、価格の高い統計ソフトを購入しなくてもよい点でとても優れている。

	A	B	C
1	Trial	Light	Sound
2	1	305	304
3	2	316	310
4	3	325	316
5	4	311	316
6	5	299	307
7	6	310	298
8	M	①	
9	SD	②	

図1. 個人データの表計算ソフト・ワークシートへの入力方法

2. 平均値と標準偏差

最大酸素摂取量、脚筋力、ボール速度、反応時間といった計測データは、連続する変数であり、間隔尺度ということになる。ここで個人や集団を代表する値が必要となる。代表値として最も一般的に用いられるのが平均値（mean, M）である。個々のデータを x_i 、平均値を \bar{x} とすると次の式で表される。

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} (x_1 + x_2 + x_3 + \dots + x_n)$$

データが平均値からどの程度散らばっているかを示す数値が、標準偏差（standard deviation, SD, σ ）である。標準偏差は次の式で表される。

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} = \sqrt{\frac{1}{n} [(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2]}$$

これらは表計算ソフトを用いることで簡単に計算することができる。例えば、被験者Aの光刺激と音刺激に対する単純反応時間のデータ（6試行）を入力する際には、ワークシートの列に項目、行にデータを入力する。図1に示したM（平均値）とSD（標準偏差）のセル（①、②）には、スプレッドシート関数（M=AVERAGE, SD=STDEV.P）を入力する。そうすると平均値311、標準偏差8となる。この平均値である311 msが個人の代表値となる。対象となった被験者の代表値を集めたものが集団のデータということになる。このデータから母集団のばらつきを推定することになるが、このときには標準偏差の計算式の分母にあるn（サンプル数）をn-1とす

ることになる。分母をn-1で求めた標準偏差のことを標本標準偏差という。このときのスプレッドシート関数はSD=STDEVとなる。

3. データの事前準備

個人の代表値から、集団のデータを作成する際には、個人データの整理と同じようにワークシートの列に項目、行にデータ（個人の代表値）を入力する（図2参照）。図2には、被験者24名（男性、14名；女性、10名）による光刺激と音刺激に対する単純反応時間を示してある。図1の個人データと異なるのは、B列に「性別」の項目が追加されている。図2に示した数値の並びは、js-STARに限らず、他の統計ソフトでも同様なので、統計をかける事前準備として、このようなワークシートを作成しておくことが重要となる。統計分析用のファイルをインポートする場合には、平均値や標準偏差などのデータは入れないようにする。js-STARでは、一括代入エリアに表計算ソフトのシートから、必要な箇所の数値をコピー&ペーストすることになる。本稿では便宜上、図2に示すワークシートに平均値と標準偏差を入れておく。

	A	B	C	D
1	Sub.	性別	Light	Sound
2	1	Female	297	277
3	2	Female	297	291
4	3	Female	311	298
5	4	Female	336	329
6	5	Female	352	346
7	6	Female	307	311
8	7	Female	304	278
9	8	Female	298	317
10	9	Female	311	296
11	10	Female	264	267
12	11	Male	317	323
13	12	Male	273	280
14	13	Male	273	291
15	14	Male	281	267
16	15	Male	264	292
17	16	Male	292	291
18	17	Male	280	271
19	18	Male	293	284
20	19	Male	277	279
21	20	Male	304	289
22	21	Male	274	262
23	22	Male	285	281
24	23	Male	282	282
25	24	Male	258	265
26	M		294	289
27	SD		22	23

図2. 代表値を入力した集団データのワークシートへの入力方法

4. t-検定（2標本の差の検定）

独立している2群、関連している2群の平均値の差異を検定したい場合には、t-検定を用いる。t-検定は、比較する群が2つの場合に限られる。3つ以上の群を比較したいときには第6章で説明する分散分析を用いる。

独立している2群とは、被験者が異なる2つの集団データを扱う場合のことを言う。図2のサンプルを例にす

ると、光刺激に対する反応時間について、男性と女性で差異があるのかどうかを検定したい場合のことである。このときのt検定は「対応のないt検定」を用いることになる。js-STARでは、分析ツールにある「t検定（参加者間）／ノンパラ」を使用する。「t検定（参加者間）／ノンパラ」をクリックすると図3に示す画面が表示される。まずは参加者数の設定を行う。図2のサンプルデータは女性、男性の順番で並んでいるので、群1には「10」、群2には「14」を入力する（図3①）。これで「②データ入力セル」の形がワークシートと同じになるはずである。次に表計算ソフトのワークシートにあるC列の2行目から25行目までの数値データを「③一括代入エリア」へコピー＆ペーストすれば良い。あとは「④計算！」ボタンを押せばt検定を実行してくれる。そうしたならば、「結果エリア」にt検定の結果が表示される（図4⑤参照）。t値は2.8393、自由度は13、有意水準は $p < 0.05$ であることが分かる。記述する場合は「光刺激に対する反応時間は、男性（284 ms）の方が女性（308 ms）よりも有意に速かった（ $t(13) = 2.8393, p < 0.05$ ）」となる。これをグラフにする場合には、平均値を○△で表し、標準偏差を上下のヒゲ（エラーバー）で表すことになる（図5参照）。そして、両群間にアスタリスクを表示しておくが良い。

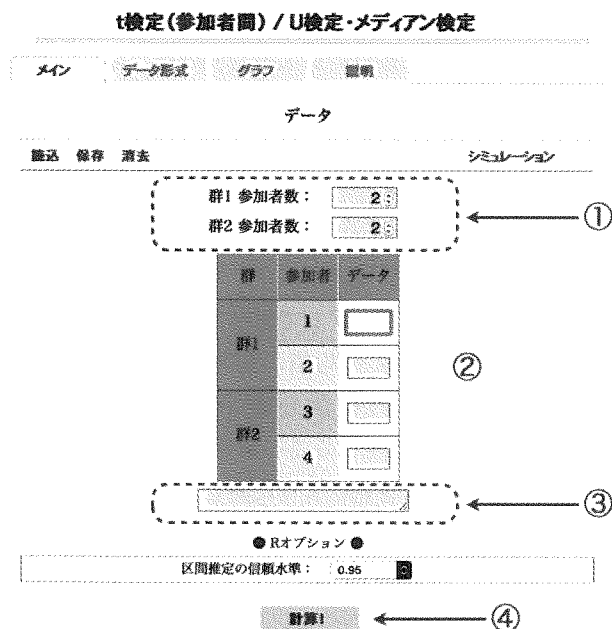


図3. 対応のないt検定における入力画面（js-STAR）

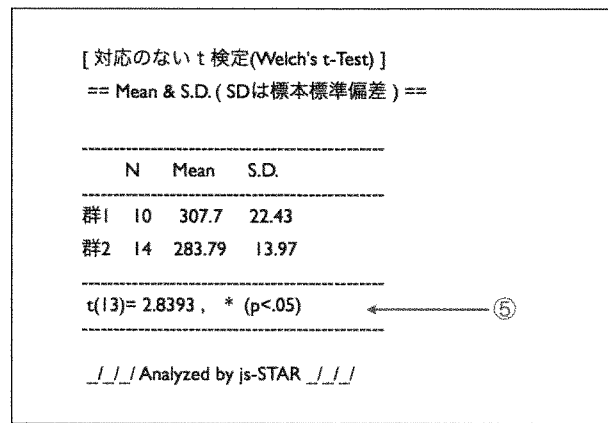


図4. 対応のないt検定の分析結果

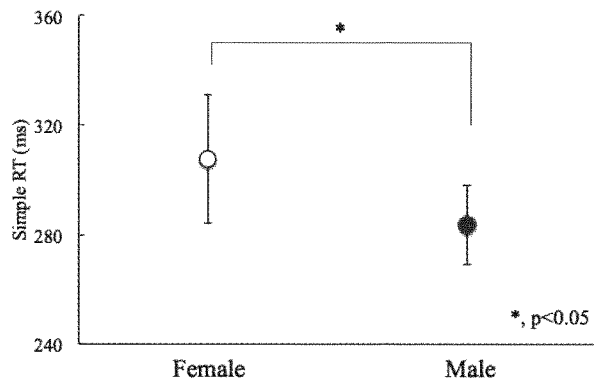


図5. 光刺激に対する全身単純反応時間における男女差

関連している2群とは、同じ被験者が2つの異なる条件で実施したデータを扱う場合のことを言う。図2のサンプルを例にすると、光刺激に対する反応時間と、音刺激に対する反応時間で差異があるのかどうかを検定したい場合のことである。平均値をみると光刺激が294 ms、音刺激が289 msとなり、音刺激に対する反応時間の方が速い。これに有意な差異があるかどうかを検定したいということになる。このときのt検定は「対応のあるt検定」を用いることになる。js-STARでは、分析ツールにある「t検定（参加者内）／ノンパラ」を使用する。図2のサンプルデータを用いる場合、被験者数は24名となるため、参加数は24に設定する。次に表計算ソフトのワークシートにあるC列とD列の2行目から25行目までの数値データを「一括代入エリア」へコピー＆ペーストし、計算をすれば図6のような結果が表示される。t値は1.9013、自由度は23、 $p > 0.05$ であることが分かる（ \dagger は $0.05 < p < 0.1$ という意味）。記述する場合は「光刺激と音刺激に対する反応時間の間に有意差はなかった。」となる。このようにt検定には2種類があり、「対応なし」が「参加者間」、「対応あり」が「参加者内」と覚えておこう。

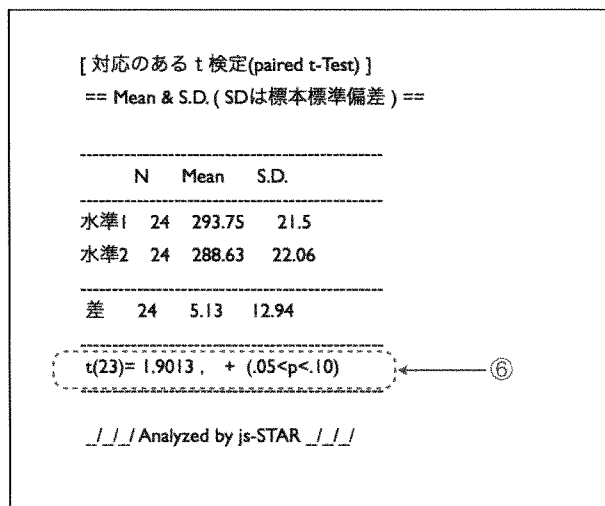


図6. 対応のあるt-検定の分析結果

5. 相関係数 (2 標本の関連性)

相関係数は、2 標本の連続変数の間にどの程度の直線関係があるのかどうかをみる指標である。一方の変数が大きくなると、もう一方の変数も大きくなる関係は「正の相関関係」と言い、小さくなる関係は「負の相関関係」と言う。図7は、サッカーのインサイドキックにおけるボール初速度と足部のスイング速度の関係を表した散布図である。散布図の左側は標的までの距離が12 m、右側は18 mと遠くなっている。どちらの距離においてもボール初速度とスイング速度に正の相関がありそうに見える。そこで相関係数を算出することが大きな意味を持つ。なぜなら相関係数の大きさは、2つの標本の相関関係の強さを表しているからである。js-STARでは、分析ツールにある「相関係数の計算と検定」を使用する。ここでは、12 m条件と18 m条件におけるそれぞれの数値データを表計算ソフトから「一括代入エリア」へコピー&ペーストし、計算ボタンをクリックする。そうすると図8のような結果が表示される。12 mの距離からのキックでは、相関係数は $r=0.602$ 、 $p>0.05$ であり、相関係数が有意でないことが分かる。一方、18 mの距離からのキックでは、相関係数は $r=0.908$ 、 $p<0.01$ であり、有意な相関が

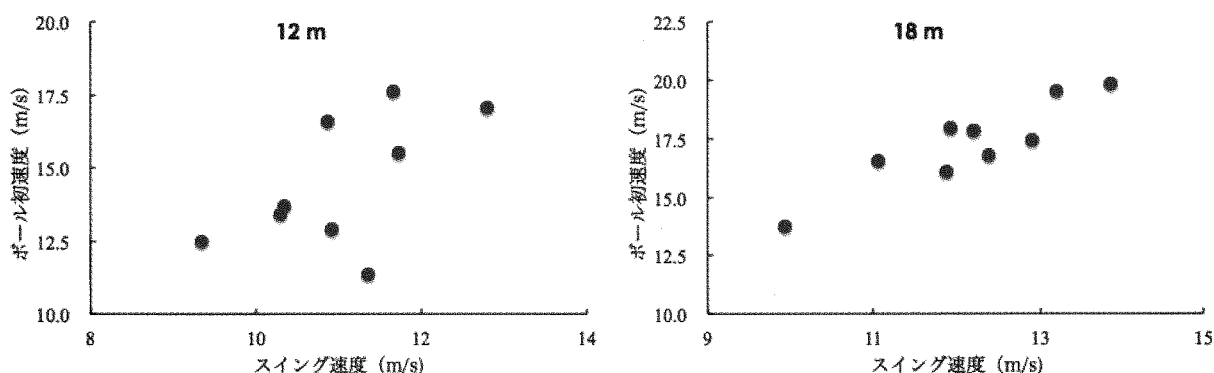


図7. インサイドキックにおけるボール初速度と足部スイング速度の関係

見られたことが分かる。記述する場合は「ボール初速度とスイング速度の間の相関関係は、12 mの距離からのキックではなかったが、18 mの距離からのキックには非常に強い相関関係が認められた ($r=0.908, p<0.01$)。」となる。ボール初速度とスイング速度の相関関係は、距離の遠い18 mからのインサイドキックの方が、距離の近い12 mからよりも強いということが言える。

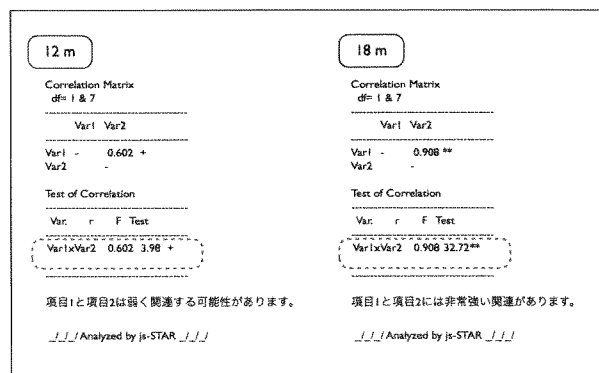


図8. 標的までの距離が12 mと18 mにおける相関分析の結果

6. 分散分析 (3 群以上の差の検定)

3つ以上ある群の平均値の差を検定したい場合には、分散分析 (Analysis of variance, ANOVA) を用いる。例えば、野球の走塁において3種類の走法の塁間時間を比較したい場合である (1 要因分散分析)。また、A、Bという2種類のトレーニング方法による全身持久力への効果を、トレーニング前後の最大酸素摂取量でみる場合などである (2 要因分散分析)。したがって、多くの実験研究においては、分散分析を用いることになるだろう。独立多群の平均値の差異を検定したい場合は、繰り返しなしの分散分析 (Factorial ANOVA) を用いることになる。独立多群とは、野球選手、サッカー選手、水泳選手のように3つ以上のグループでデータを分ける場合のことを言う。関連多群の平均値の差異を検定したい場合は、繰り返しありの分散分析 (Repeated measures ANOVA) を用いることになる。関連多群とは、低強度、中等強度、高強度の運動のように同じ被験者が3つ以上

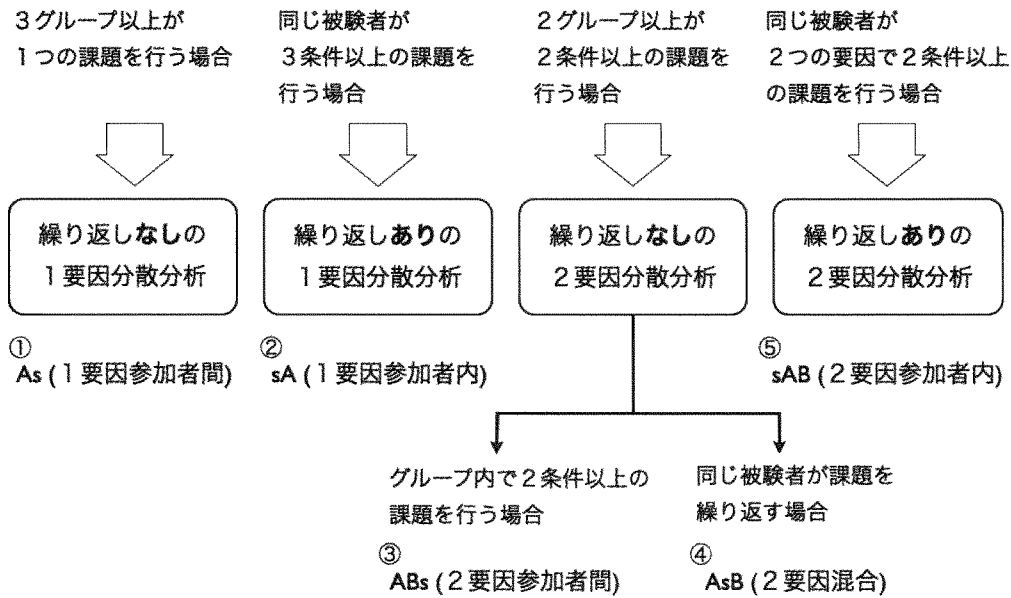


図9. 実験デザインに対応した分散分析の手法のフローチャート

の条件でデータを分ける場合のことを言う。図9には、どのような実験デザインで研究を実施した際に、どの統計手法を使えば良いのかフローチャートに示している。数字で記してある方法が、js-STARで用いる分析ツールの名称となる。実験計画を立てるとき、どの実験デザインを用いることになるのか事前に検討しておくことが大切である。3要因以上の実験デザインは、有意差検定が厳しくなること、結果の解釈が難しくなることがあるため、できるだけ2要因までで計画することが推奨される。図10は、それぞれの分散分析を行う際のjs-STARにおける分析ツールの「データ入力セル」を示してある。この図を見れば、表計算ソフトのワークシートにどのように数値を入力すれば良いかが分かるであろう。

離を示してある。3つの視覚条件は、標的が見えるコントロール条件、標的が遮蔽されている標的遮断条件、アイマスクをした視覚遮断条件である。被験者数は16名であった。js-STARでは、分析ツールにある「sA（1要因参加者内）」を使用する。図12にある表示がなされるので、参加者数に被験者の人数16（①）、水準数に視覚条件数の3（②）を入力し、多重比較の項目を「Bonferroni（③）」にチェックを入れておく。3条件の数値データを表計算ソフトから「一括代入エリア」へコピー＆ペーストし、計算ボタンをクリックする。そうすると図13のような結果が表示される。①をみるとF値が5.97、 $p < 0.01$ であることがわかる。②をみると条件についての自由度（群間変動）が $df = 2$ 、ばらつきについての自由度（群内変動）が $df = 30$ となる。記述する場合は「分散分析を行った結果、3つの条件間に有意なF値が認められた（ $F(2,30) = 5.97, p < 0.01$ ）」となる。つまり、3つの条件間のどこかに有意な差が生じていることを意味する。F値が有意である場合は、3つの条件間の差異を検定するために、多重比較を行うことになる。2標本t検定を繰り返すと第1種の過誤が生じ、誤って判定する確率が高くなってしまっているので、多重比較を用いることが重要である。③をみるとBonferroni法による多重比較検定の結果が示されている。有意差がある場合には、不等号で示されるので、どちらが大きいのか小さいのか分かるようになっている。「ns.」はnot significantで「有意差なし」の意味となる。記述する場合は「多重比較の結果、誤差距離は、コントロール条件よりも標的遮断、視覚遮断条件の方が有意に大きく（85.8 vs. 125.3, 122.1 cm）、標的遮断と視覚遮断条件の間には有意差はみられなかった。」となる。

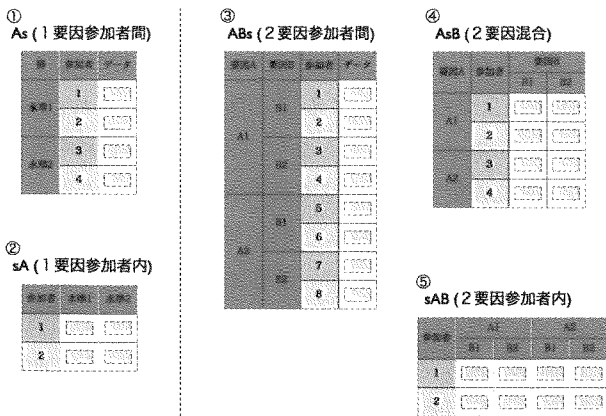


図10. 各分散分析に対応したjs-STARのデータ入力セル

まずは繰り返しありの1要因分散分析を行ってみる。図11には、3つの視覚条件において、10 mの距離から標的に向けて下手投げでボールを投げる投射課題の誤差距

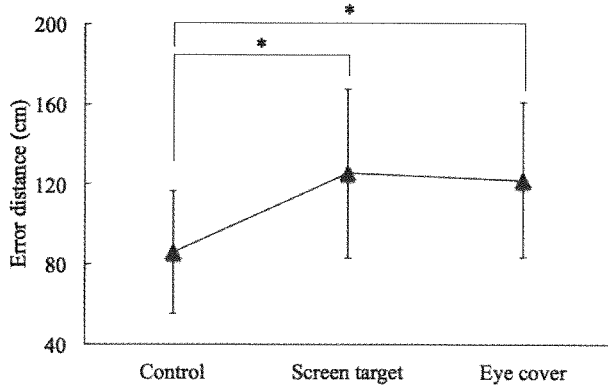


図11. コントロール、標的遮断、視覚遮断条件における標的への投射の誤差距離

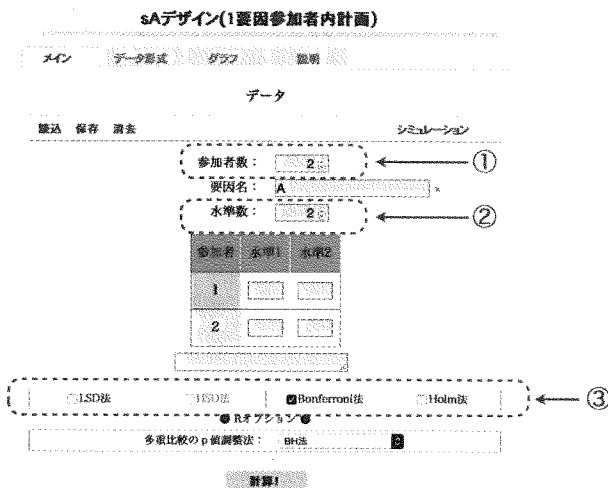


図12. 繰り返しありの1要因分散分析における入力画面(js-STAR)

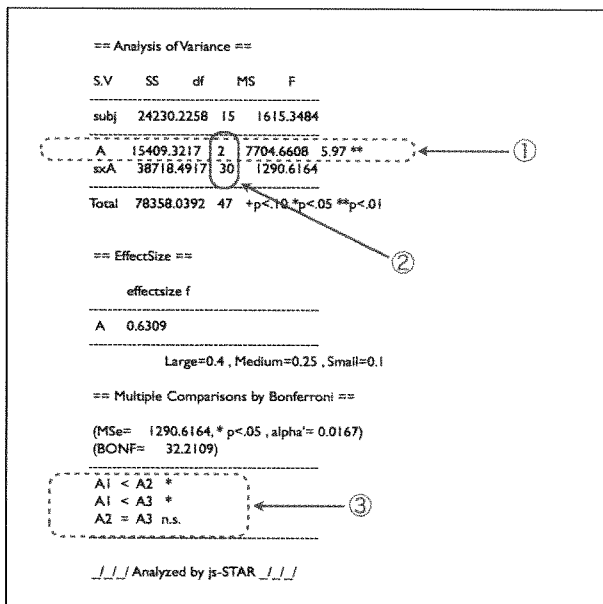


図13. 1要因分散分析の結果と多重比較検定

次に繰り返しありの2要因分散分析を行ってみる。図14は、24名の被験者を対象に光刺激と音刺激に対する単純反応時間、Go/NoGo反応時間の平均値を示したものである。単純反応課題は、光刺激については赤色の刺激が

提示されたらジャンプする、音刺激については500 Hzの音が提示されたらジャンプするというものであった。Go/NoGo反応課題は、光刺激では赤色または青色のどちらか一方の刺激が提示され、赤色であったらジャンプし、青色であったらジャンプしないというものであった。音刺激では500 Hzまたは1 kHzのどちらか一方の刺激が提示され、500 Hzであったらジャンプし、1 kHzであったらジャンプしないというものであった。この実験デザインでは、刺激の違いによる要因と、反応課題の違いによる要因の2つの要因があることになる。また、それぞれの要因に2つの水準があることになる。この場合、js-STARでは分析ツールにある「sAB (2要因参加者内)」を使用する。図14にある表示がなされるので、参加者数に被験者の人数24 (①)、要因Aの水準数に2 (②)、要因Bの水準数に2 (③)を入力し、多重比較の項目を「Bonferroni (④)」にチェックを入れておく。要因Aが刺激の違いであり、要因Bが反応課題の違いとなる。データの並びは横1列で、要因A1の要因B1、B2、要因A2の要因B1、B2の順番にしておく。4条件の数値データを表計算ソフトから「一括代入エリア」へコピー&ペーストし、計算ボタンをクリックする。そうすると図16のような結果が表示される。2要因分散分析では、1つの要因の条件間の平均値に有意差がある場合を「主効果」と言う。2つの要因が作用しあって、条件間の平均値に有意差がある場合を「交互作用」と言う。

図16の①②が主効果の結果であり、刺激の違いである要因AはF値が5.28, $p<0.05$ となり、反応課題の違いである要因BはF値が131.75, $p<0.01$ となる。③が交互作用の結果であり、F値が13.80, $p<0.01$ であることが分かる。記述する場合は「2要因分散分析の結果、刺激の違いによる要因と反応課題の違いによる要因に有意な主効果が認められ (F (1,23) =5.28, $p<0.05$; F (1,23) =131.75, $p<0.01$), 2つの要因による交互作用が認められた (F (1,23) =13.80, $p<0.01$ 。)」となる。交互作用が認められたため、下位検定である単純主効果の検定をすることになる。結果はすでに算出済みである (図17参照)。要因内の水準が3つ以上ある場合は、多重比較検定を行うことになる。A at B1 (①)は、単純反応時間において光刺激と音刺激の間に有意差がないことを示している。A at B2 (②)は、Go/NoGo反応時間において光刺激と音刺激の間に有意差があることを示している (354 vs. 377 ms)。この実験課題においては、①②が重要な指標となる。つまり、光刺激と音刺激に対する全身反応時間は、単純反応課題では差異がなかったが、Go/NoGo反応課題になると音刺激の方が遅くなることを示唆し

ている。B at A1 (③) と B at A2 (④) は、光・音刺激の両条件において、単純反応時間の方がGo/NoGo反応時間よりも有意に速いことを示している。

図14のデータでは交互作用があったが、もし、交互作用がなく、主効果のみがあった場合はどうなるであろうか。例えば、分散分析の結果、反応課題の違いである要因Bにのみ主効果があったとする。この場合、刺激の違いである光刺激に対する反応、または音刺激に対する反応は、反応時間に影響する要因としては除かれることになる。そして、両条件の単純反応時間がGo/NoGo反応時間よりも速かったという解釈になる (292 vs. 362 ms)。すなわち図14における白○の2条件のデータをひとまとまり、黒●の2条件のデータをひとまとまりとして、その平均値間に有意な差があることを意味することになる。このように2要因分散分析では、交互作用があるのか、ないのかでデータの取り扱いに違いがあるので、気をつけておく必要がある。

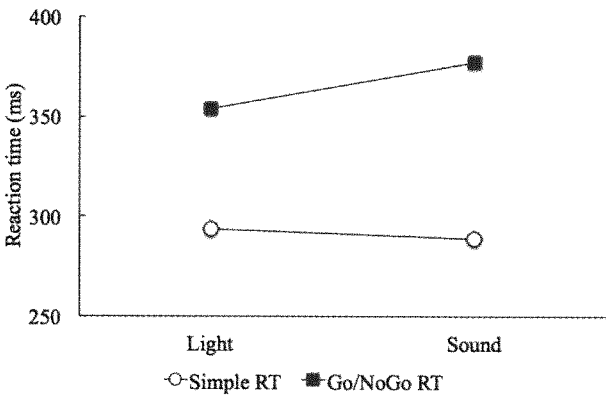


図14. 光刺激と音刺激に対する単純反応時間とGo/NoGo反応時間

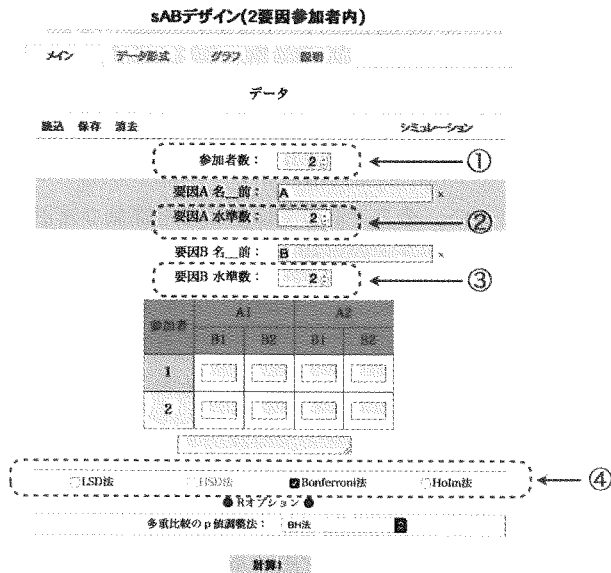


図15. 繰り返しありの2要因分散分析における入力画面 (js-STAR)

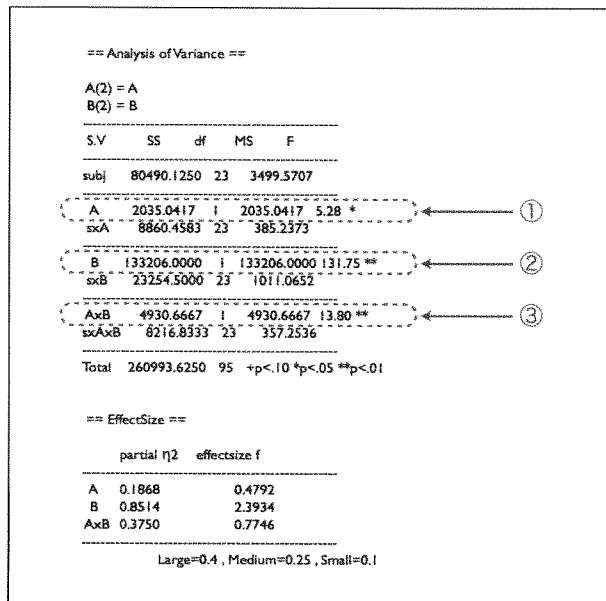


図16. 2要因分散分析による主効果と交互作用の結果

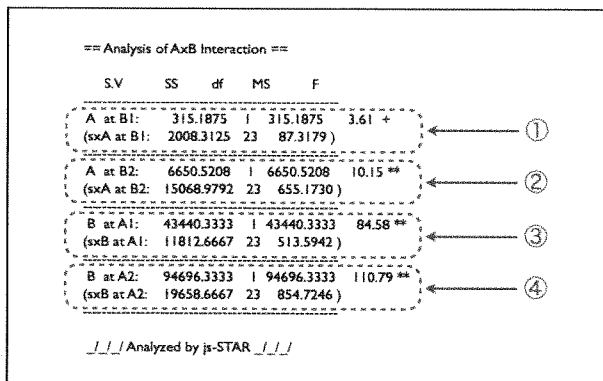


図17. 2つの要因による交互作用が有意な場合の下位検定 (単純主効果)

7. まとめ

科学論文を執筆する際には、抽出された標本のデータから母集団のデータを推計し、ある集団間におけるデータの平均値に差異があるかどうかを検定することになる。確率論的に最も確からしい根拠を示すには、「統計」は避けては通れない。そこで本稿では、とても高価な統計ソフトを使用しなくても統計値を算出してくれるjs-STARによる統計手法を解説した。数値データを表計算ソフトのワークシート内に適切な行列に並べてさえいれば、js-STARの「一括代入エリア」へコピー＆ペーストすれば良いことになる。統計値をすぐに算出してくれるので、とても簡便に利用することができる。本稿で紹介したのは、利用頻度の高いt-検定、相関係数、分散分析の使用方法である。特に分散分析を用いる場合は、実験デザインとの関連性があるため、事前の研究計画の段階でどの分散分析を用いることになるのか考慮する必要がある。実験デザインと分散分析の関係を2要因まで示したので、どの統計手法を用いれば良いのか分かるであろう。何れにしても統計学の基本的知識を持っていた方が理解のすすむことは間違いない。js-STARは、OSを選ばずインターネット上で手軽に利用することができる。これを活用することで、学生諸君には、より質の高い科学論文を執筆して欲しいと思う。

文献

- Catalano, J. F. & Kleiner, B. M. (1984) Distant transfer and practice variability. *Perceptual and Motor Skills*, 58, 851-856.
- 市原清志. (1994) バイオサイエンスの統計学：正しく活用するための実践理論. 南江堂.
- 神田善伸. (2014) EZRでやさしく学ぶ統計学：EBMの実践から臨床研究まで. 中外医学社.
- 対馬栄輝. (2020) 医療統計解析使いこなし実践ガイド. 羊土社.

注釈

(注1) js-STARは田中 敏, Nappaが開発したウェブブラウザ上で使用できる統計分析プログラムである。
<http://www.kisnet.or.jp/nappa/software/star/index.htm>

(2021年2月17日受理)