©NSCA JAPAN
Volume 21, Number 5, pages 20-25

特集 feature

論文の書き方8

相関と回帰（その2）

星川佳広　CSCS, NSCAジャパン編集委員, 東海学園大学准教授

本稿は、「論文の書き方1：投稿論文を書こう」（2013年10月号）の続編パート7です。NSCAジャパンウェブサイトに掲載中の「投稿論文（事例報告）の書き方」を参照しながら読み進めてください。

【投稿論文（事例報告）の書き方】
NSCAジャパンウェブサイトTOP　→　［指導者の育成］　→　［事例報告・研究論文］、「投稿要領」部分

前々号までにすでに紹介済みですが、Excelには代表的な統計や検定を実施するための「分析ツール」が用意されています（メニュー→データ→データ分析）[1]。「相関」と「回帰」分析についても分析ツールを利用することができます。前号ではグラフ作成や関数によって「相関」、「回帰」の諸指標を得る方法を紹介しましたが、今回は「分析ツール」を利用する方法を紹介します。「分析ツール」ではより詳細に分析に付随する数値を知ることができます。

さて前号の続きで、あなたはある大学野球チームでS&Cコーチをしているとします。そして選手に対してレジスタンストレーニングの指導を始めました。一定期間の後、1RMの向上と50m走タイムの改善が関係しているかを確認することにしました。

その結果が**表1**のA1：E17の領域であったとします。**表1**のワークシートは、目的変数（y）としてトレーニング前後の50m走タイムの変化、説明変数（x_1, x_2, x_3）としてトレーニング前後のスクワット、ベンチプレス、レッグカールの1RMの変化としています。列Fの重回帰式については後述します。

1．相関

分析ツールにおいて「相関」を選択すると**図1**のボックスが出現します。このボックスの「入力範囲」には、**表1**の例ではB2：E17とします。「先頭行をラベルとして使用」をチェックすると、B2：E17の先頭行つまりB2：E2の行はデータでなくラベルとして扱われます。「出力オプション」で分析結果を出力する場所を指定しますが、任意の場所を選びOKをクリックすると**図2**が出力されます。**図2**は「相関クロス」と呼ばれるものですが、50m走タイムおよび各1RMの変化の2つの変数間の相関係数が一覧にて示されています。相関係数は、前号紹介したcorrel関数によっても計算することができますが、分析ツールを使うとすべての変数間の相関係数を一気に計算してくれるので、特に変数が多い場合は作業を短時間に済ますことができて便利です。

ただし、残念ながら分析ツールも相関係数の有意性までは計算してくれません。前号第4節で紹介した脚注の式[2]に従ってご自身で各相関係数が有意かどうか計算してみてください。50m走

1　「データ分析」がリボン内にない場合は、ファイル→オプション→アドイン→データ分析

2　$t=\frac{r\times\sqrt{n-2}}{\sqrt{1-r^2}}$　　r 相関係数、n 被検者数

表1　ワークシート例						
	A	B	C	D	E	F
1			1 RMの変化（kg）			
2	選手	50m走タイムの変化（秒）	スクワット	ベンチプレス	レッグカール	重回帰式
3	a	0.21	5	0	2.5	−0.0478399
4	b	0.1	10	0	7.5	−0.1523328
5	c	0.05	0	20	5	0.0934765
6	d	0.01	25	20	5	−0.3835222
7	e	−0.05	10	15	5	−0.1099388
8	f	−0.2	20	10	7.5	−0.3179007
9	g	−0.21	5	12	0	−0.0130153
10	h	−0.33	10	10	5	−0.1225546
11	i	−0.38	15	2	5	−0.238140
12	j	−0.44	20	12	5	−0.3083078
13	k	−0.44	35	25	10	−0.5707991
14	l	−0.46	25	0	12.5	−0.4476253
15	m	−0.65	25	15	10	−0.4052313
16	n	−0.7	40	5	5	−0.7075688
17	o	−0.75	30	10	7.5	−0.5087002

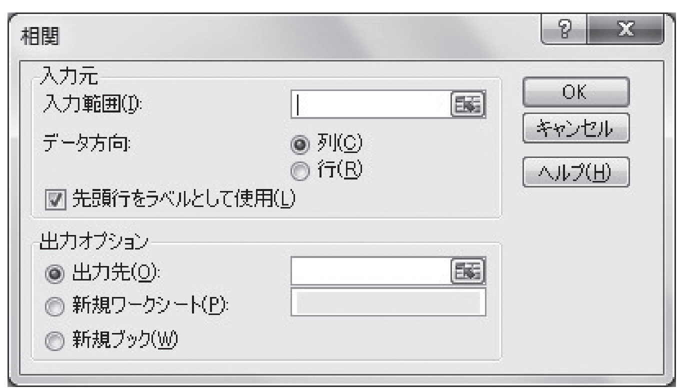

図1　データ分析（相関）

タイムの変化とスクワット間の$r=-0.7457$は、$p=0.00071\cdots$、スクワットとレッグカール間の$r=0.544762$は、$p=0.0178\cdots$になるはずです。ともに$p<0.05$ですから５％水準で有意な相関関係にあると判断できます。一方、50m走タイムの変化とレッグカール間の$r=-0.4225$は$p=0.0583\cdots$と計算されるはずです。したがって５％水準では有意と判断できません。

2. 回帰分析

前号のベースランニングタイム（目的変数）と50m走タイム（説明変数）の関係では、目的変数はひとつの説明変数によって説明されていました。このような回帰を「単回帰」といい、回帰式は$y=ax+b$と表現されます。その一方で今回の**表１**では、50m走タイムの変化（目的変数）がスクワット、ベンチプレス、レッグカールの1RMの変化という複数の説明変数によって説明されます。このような回帰を「重回帰」といいます。重回帰での回帰式は、$y=ax_1+bx_2+cx_3+\cdots+z$のように表現されます。**表１**の例では、50m走タイムの変化＝a×スクワットの変化＋b×ベンチプレスの変化＋c×レッグカールの変化＋dとなります。ここでa、b、cの値が負（マイナス）であれば、スクワット、ベンチプレス、レッグカールの変化は目的変数を減らす方向（50m走タイムタイムを短縮する方向）に作用するはずです。このように重回帰分析を利用すると、目的変数に対してそれぞれの説明変数がどの程度どのように貢献しているか、つまり50m走タイムの変化に対してどのレジスタンストレーニングが最も貢献しているかを数字として知ることができます[3]。これは私たちS&C専門職にとっては非常にありがたい分析方法ではないかと思います。

さて、分析ツールにおいて「回帰分析」を選択すると**図３**のボックスが出現します。分析ツールの「回帰分析」は、もともと重回帰分析での利用を想定したつくりになっています。そこで本稿でもそれに沿って分析ツールの使い方を重回帰分析によって説明しますが、

3　重回帰式　$y=ax_1+bx_2+cx_3+\cdots+z$のa, b, cを「偏回帰係数」といいます。偏回帰係数の値が、目的変数に対してx_1、x_2、x_3それぞれの説明変数がどの程度貢献しているかの情報を与えますが、a, b, c にはx_1、x_2、x_3の値の大きさそのものも影響するので、偏回帰係数を標準化して「標準偏回帰係数」を計算する必要があります。この計算方法はここでは割愛しますが、標準偏回帰係数は非常に重要な数値です。「重回帰分析」を理解するには、ここで記述する以上の様々なことを知る必要があります。他書で「重回帰分析」の理解を深めつつ勉強してください。

	50m走タイムの変化（秒）	スクワット	ベンチプレス	レッグカール
50m走タイムの変化（秒）	1			
スクワット	−0.74547	1		
ベンチプレス	−0.03199	0.13111	1	
レッグカール	−0.4225	0.544762	0.023277	1

図２　データ分析（相関）の結果

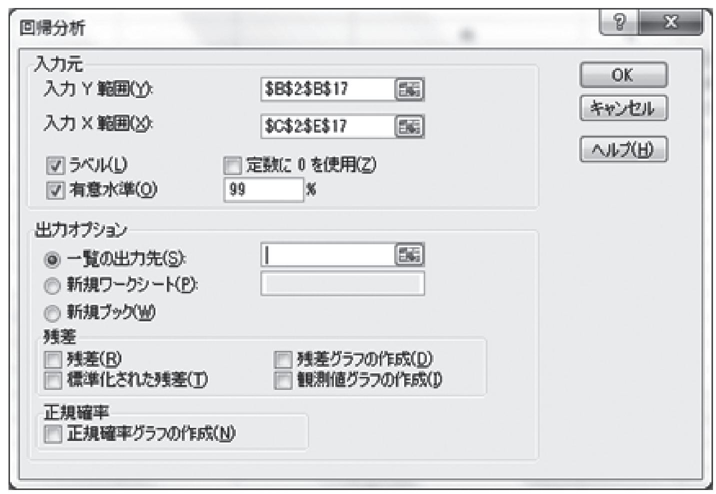

図３　データ分析（回帰分析）[4]

理解の順番としては単回帰→重回帰のほうがわかりやすいと思います。したがって、次節の単回帰の説明も見ながら読み進めてもらえればと思います。

分析ツールによって重回帰分析を行なう場合、**図３**ボックスの「入力Y範囲」には目的変数が入力されているセル（列）を指定します。例えば**表１**では50ｍ走タイムの変化が目的変数ですから$B$2：$B$17です。「入力X範囲」には説明変数が入力されている複数の列を指定するので、**表１**の例では$C$2：$E$17です。ここでもし「入力X範囲」に$C$2：$C$17のようにひとつの列を指定すれば、それはひとつの説明変数による単回帰分析を行なうことになります（次節参照）。「ラベル」をチェックすると先頭行はデータでなくラベルとして扱われます。「定数に０を使用」にチェックすると出力される回帰式は強制的に切片が０になります。「有意水準」は99％としておきましょう。各係数の99％信頼区間を算出してくれます（95％信頼区間は、この指定にかかわらず自動的に算出してくれます）。出力オプションで分析結果を出力する任意の場所を指定してOKをクリックすると**図４**が出現します。

図４の重要な点のみをピックアップして説明します。まず先に**図４**の下方に示された“切片”、“スクワット”、“ベンチプレス”、“レッグカール”と書かれている表を見てください。そのすぐ右隣の係数と書かれている列にある数字が切片および回帰係数になります。すなわち、

50ｍ走タイムの変化＝−0.0191×スクワットの変化＋0.0025×ベンチプレスの変化−0.0018×レッグカールの変化＋0.0521…（＊）
（小数点第5位を四捨五入）

の重回帰式が得られました。＊式を使えば、スクワット、ベンチプレス、レッグカールの1RMの変化から50ｍ走タイムの変化を予測することができるわけです。**表１**のワークシートF列には、＊式を各被検者のスクワット、ベンチプレス、レッグカールの変化データに当てはめて求めた50ｍ走タイム変化の予測値が入力されています。仮にスクワット、ベンチプレス、レッグカールの変化という３つの説明変数のみで完璧に50ｍ走タイムの変化を予測できたならば、**表１**のF列の数値と実際の50ｍ走タイムの変化であるB列の数値は完全に一致するはずです。

図４の上部、回帰統計と書かれた部分の“重相関Ｒ”は、重回帰式（＊）による予測値（F列）と実測値（B列）との相関係数で、「重相関係数」といいます。重相関係数は一般的に大文字のRを使って表現します。すなわち重相関係数は、重回帰式（＊）が実際の値をどれくらい適切に予測しているかのあてはまりの良さを示しています。当然ながらr（前号参照）と同じように−1〜1の間を取ります。“重決定Ｒ２”とは「決定係数（R^2）」を表します。決定係数（R^2）

4　ボックス下部の「残差」や「正規確率」のオプションは、回帰分析に慣れてきたらチェックを入れて出力される内容を確かめてください。重回帰による予測値と実測値の差異の分析などを行なうことができます。

概要

回帰統計	
重相関 R	0.74858186
重決定 R2	0.5603748
補正 R2	0.44047701
標準誤差	0.22546041
観測数	15

分散分析表

	自由度	変動	分散	観測された分散比	有意 F
回帰	3	0.71273697	0.23757899	4.67377112	0.02432569
残差	11	0.55915637	0.0508324		
合計	14	1.27189333			

	係数	標準誤差	t	P-値	下限95%	上限95%	下限99.0%	上限99.0%
切片	0.05210651	0.15520235	0.33573272	0.74338872	−0.2894916	0.39370458	−0.429922	0.53413498
スクワット	−0.01907995	0.00617652	−3.08910846	0.01030245	−0.0326744	−0.00548551	−0.038263	0.00010314
ベンチプレス	0.00252316	0.00772512	0.32661760	0.75008743	−0.0144797	0.01952604	−0.0214696	0.02651589
レッグカール	−0.00181865	0.02310717	−0.07870508	0.93868071	−0.0526772	0.04903988	−0.073585	0.06994774

図４　データ分析（回帰分析）の結果１（重回帰）

は、結果的に重相関係数(R)の二乗値と等しくなりますが、その定義は、重回帰式による予測値の分散（**表１**F列の分散）／実測値の分散（**表１**B列の分散）です。この式は、実測値の分散のうち重回帰式の予測によって説明できる分散の割合を意味しており、決定係数（**図４**の場合0.56…）を使って「重回帰式が実測値を56％説明する」といった表現をしたりします。“補正R２”は、決定係数を目的変数と被検者の数で補正したものです。なぜなら、説明変数を増やすとその説明変数が目的変数と無関係であっても、決定係数は１に近づく（表面上、良い予測ができるようになる）性質があるからです。そこで“補正R２”のほうで重回帰式の当てはまりの良さを評価することもあります。

図４中段の分散分析表では、“有意F”と記された列に0.0243…と出力されています。これは＊式が５％水準で有意であることを表します。つまり、＊式は意味ある予測ができていることになります。

再度**図４**下の“切片”、“スクワット”、“ベンチプレス”、“レッグカール”と書かれている表に戻ります。表の６～７列目には“下限95％”および“上限95％”と書かれていますが、これは＊式の各係数の95％信頼区間を表します。同様に表の８～９列目には、各係数の99％信頼区間が出力されています[5]。例えば“スクワット”については、回帰係数は95％の確率で－0.0326744～－0.00548551の範囲内に、99％の確率で－0.038263～0.00010314の範囲内に存在することを表します。逆にいえばスクワットの回帰係数がこれらの範囲外になる確率はそれぞれ５％、１％未満ということです。

スクワットの回帰係数が上限95％信頼区間の上限においてもマイナス（－0.0054855）ということは、スクワット１RMの回帰係数（－0.0190799）は5％水準で有意に0より小さいことを意味します。つまり、スクワット１RMの増加は目的変数（50m走タイムの増加）に対して95%の確率でマイナス（50m走タイムを短縮する方向）に作用する、あるいは逆に、スクワット１RMの増加が目的変数を増やす方向に作用する確率は5%未満しかないわけです。その一方でベンチプレス、レッグカールの回帰係数は95%信頼区間の上限が０を超えてしまってい

5　95％信頼区間は自動的に出力され、99％信頼区間は図３の有意水準で99％としたため出力される

ます。つまり係数が0となることもありうるわけで、この場合、ベンチプレス、レッグカールの1RMが増えようが減ろうが目的変数(50m走タイムの変化)には全く影響しません。したがって、ベンチプレス、レッグカールについては回帰係数(それぞれ0.00252…、−0.00181…)が出力されていますが、この係数は5%水準で0と差がなく有意ではありません。すなわち、ベンチプレス、レッグカールの1RMの増加は50m走タイムの変化に貢献しないと判断できます。

図4の出力結果をまとめれば以下のようになります。「目的変数(50m走タイムの変化)に対してスクワット、ベンチプレス、レッグカール1RMの変化の3つを強制的に説明変数として組み込み重回帰分析を行なった結果、＊式を得ることができた。しかし、実際に50m走タイムの変化に有意に貢献していたのはスクワット1RMの増加のみであった。」[6]

3. 回帰分析(単回帰)

前号で紹介したとおり、単回帰の場合(y=ax+b)、回帰係数aと切片bはそれぞれslope関数、intercept関数によっても知ることができます。しかし、分析ツールの回帰分析を利用すれば上述どおりそれらの95%信頼区間をも知ることができます。

図5は、**図3**の「入力X範囲」に**表1**のスクワットのデータ($C$2:$C$17)のみを指定した場合の出力です。**図5**をみておわかりになるとおり、分析ツールの回帰分析では単回帰分析を行なったとしても、重回帰分析と同じフォーマット(“重相関R”、“重決定R2”など)で出力されます。Excelは単回帰分析を重回帰分析の説明変数がひとつの場合として、重回帰と単回帰を区別なく扱っているようです。しかし、皆さんが単回帰分析を行なって論文を書く場合は、相関係数はRではなくrを使って表現しましょう。結局は、**図5**の“重相関R”によって(±の違いがありますが)実質的にxとyの相関係数rを知ることができます。**図5**の“重相関R”の数値0.7454…と**図2**相関クロス内の50m走タイムの変化とスクワット1RMの変化の相関係数−0.7454…は±の違いがありますが同じであることを確認してください。また、**図5**の補正R2は、**図4**の補正R2よりも高い値になっています。これは3つ(スクワット、ベンチプレス、レッグカールの変化)の説明変数による回帰式よりも、スクワットというひとつの説明変数による回帰式のほうが予測が

概要								
回帰統計								
重相関 R	0.74547241							
重決定 R2	0.55572912							
補正 R2	0.52155444							
標準誤差	0.20848634							
観測数	15							
分散分析表								
	自由度	変動	分散	観測された分散比	有意 F			
回帰	1	0.70682816	0.70682816	16.2614271	0.001422266			
残差	13	0.56506517	0.04346655					
合計	14	1.27189333						
	係数	標準誤差	t	P-値	下限95%	上限95%	下限99.0%	上限99.0%
切片	0.06787931	0.10224703	0.66387562	0.51837119	−0.153012	0.28877059	−0.2401169	0.37587557
スクワット	−0.0191207	0.00474159	−4.032546	0.00142227	−0.029364	−0.0088771	−0.0334037	−0.0048377

図5　データ分析(回帰分析)の結果2(単回帰)

6　図2の相関クロスでは、50m走タイムの変化に対して、スクワットだけでなくレッグカールの変化も相関係数は有意であったのに、重回帰分析ではスクワットのみしか有意にならなかった！　→これはスクワットとレッグカールの変化が相関関係にあるために起こる。このような現象を「**共線性**」といい、本来、重回帰分析での説明変数は互いに無関係な(共線性のない)変数を選ぶことが前提となる。

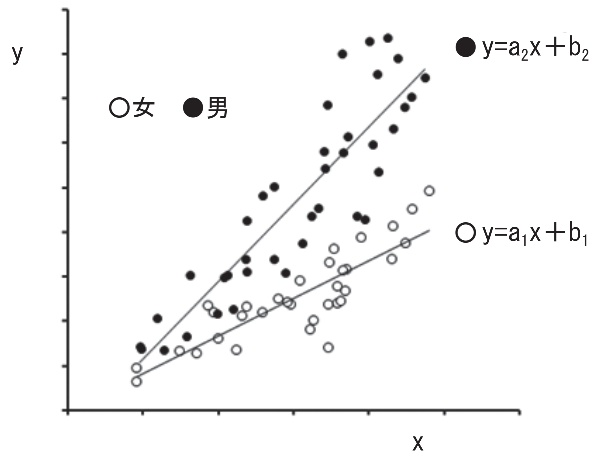

図６　ふたつの回帰式の比較例

良い（当てはまりが良い）ことを意味します。

図５下方の“切片”、“スクワット”の係数からは、50m走タイムの変化(y)をスクワットの変化(x)のみで回帰した場合、y=－0.019x＋0.068の回帰式（少数点第４位を四捨五入）になることがわかります。そして各係数の95％信頼区間は、切片については－0.153012～0.28877059、回帰係数については－0.029364～－0.0088771であることが読み取れます。

この回帰係数の95%信頼区間は、ふたつの回帰式の違いを検定するときなどに有効です。例を**図６**に挙げてみましょう。**図６**の○は女性、●は男性のデータとします。男性も女性もxとyは相関関係にありますが、xが大きくなるとyが大きくなる程度は男性のほうが女性よりも大きいように見受けられます。今、このデータよりxとyの関係性に男女差があることを示したい場合はどうすればよいでしょうか？　そのためには女性○、男性●の回帰式が有意に異なることを示せばよいわけで、各係数の95％信頼区間を利用します。

まず男女それぞれで分析ツールを利用して回帰式を求めましょう。今、女性：$y=a_1x+b_1$、男性：$y=a_2x+b_2$の回帰式が得られたとします。**図６**ではa_2がa_1より大きい（回帰直線の傾きが大きい）ようにみえますが、a_2の95％信頼区間の下限がa_1の95％信頼区間の上限よりも大きければ、a_2はa_1より５％水準で有意に大きいということができます。同様にして切片についてもb_1、b_2に有意差があるかどうか調べることができます。これらの手順を踏めば、逆に回帰式には有意差がないということを主張する場合にも応用可能です。

4. おわりに

「論文の書き方」として８回にわたり連載をしてきました。ここで一区切りとさせていただきます。「論文とは何か」からはじまり、手持ちのデータを整理する手段として、グラフの書き方やExcelを使った統計・検定方法について解説してきました。

想定した読者は、「論文を作る意思があり、見よう見まねで試みてはいるものの、なかなかうまく書き出せない人」でした。自分が行なっている指導や取り組みを成果として残したい、というNSCAジャパン会員の要望は高まっているように感じます。実際に何人かの読者から「論文を書いてみようと思います」という嬉しい反響もいただきました。ぜひとも投稿が増えることを願っています。

しかし一方で、想定が合致せずに十分に噛み砕けていなかったり、むしろ簡単すぎることもあったと思います。ご容赦ください。特に統計については短い文章の中で十分に説明できていません。ここまで読まれたことを期に、次のステップの本にぜひともあたってほしいと思います。

皆さんに「論文を書こう」と諭しながら、私自身もこのように書くという作業をすると何回も止まってなかなか進めませんでした。書くのは口で説明するのとは比較にならないプレッシャーがあります。しかし終わるとひとつまっとうした感とともに自分が成長したことも感じます。ありがとうございました。◆