『経営学論集』，第18巻　第4号，113-122，2008
KYUSHU SANGYO UNIVERSITY, *KEIEIGAKU RONSHU*（*BUSINESS REVIEW*）Vol. 18, NO. 4 , 113-122, 2008

〔論　説〕

Excelのチャートオプション“近似曲線の追加”機能の評価

孫　　宏　傑

［要　旨］

MS Excelのチャートツールは教育やビジネスなどの分野で多用されている。しかし，チャートツールの正確性についてはあまり検証されていない。チャートツールの出力は図であるから，目視によるチェック以外の検証法は少ないが，検証が不十分なツールの利用は非常に危ういと考えられる。本稿はチャートツールの一部，“近似曲線の追加”オプションに限定してその振るまいの正確さを検討する。“近似曲線の追加”機能の異常な振るまいがMSのサポートKB211967［4］，KB829249［5］，Aoki［1］で公開されたが，稀なケースかあるいは一般的な問題かについては言及していない。本稿ではこれについて，処理アルゴリズムと描画エンジンの2方面から検討する。結論として，“近似曲線の追加”機能はアルゴリズムに問題があり，少し大きいデータ（あるいは数の多いデータ）を適用すると異常な結果を得ることが多い。また，描画エンジンは正確に数式のデータを近似曲線に変換している場合もあれば，まったくできない場合も存在する。加えて，“近似曲線の追加”を利用するとき，Excelの設計上の原因でまったく意味のない数式とR^2値が出力されることがあるので，注意が必要である。なお，本稿でWindows版のExcel 2000，Excel2003とExcel2007を対象とする[1]。

1　はじめに

Excelのチャートツールの“近似曲線の追加”機能には線形近似，多項式近似，累乗近似，対数近似，指数近似，移動平均がある。ここで，移動平均はチャート上に移動平均曲線を描くだけで，数値や数式の出力がないので検討の対象外とする。残りの近似曲線はすべて線形回帰を利用しているので，まず線形回帰が正確に行われているかを検討する。

Excelには，基本的に3種の線形回帰法がある。すなわち，関数LINEST()，解析ツールパック（Add-in）の中の回帰分析，そしてソルバー（Add-in）である。解析ツールパックの回帰分析はLINEST()を使っているとMS社では公表している［4］。LINEST()に関しては，Excel2003より前のバージョンではアルゴリズムが原因で計算結果が不安定になり易いという問題があった［2，7，8，9，10］。Excel2003ではQRアルゴリズムなどの手法が導入され［6］，計

[1] Excel2000（9.0.2812），Excel2003（11.5612.5605），Excel2007（12.0.6024.5000）を利用している。アップデートで最新の状態に維持している。

算結果が基本的に信頼できるように変更された［3，9］。唯一発見されたバグはアルゴリズムの問題ではなく結果の表示処理によるものだとも報告されている［3］（このバグはExcel 2007ではなくなったが，MSからの説明はないままである）。本稿ではExcel2003のLINEST()を比較の基準とする。ソルバーに関してはやはり計算の結果が不安定になるという問題があり［7，8，9］，ここでは使わないものとする。

本稿では，まず2節でチャートツールの回帰処理が信頼できるかどうかを検討する。MS社によれば，チャートツールの回帰処理はLINEST()を使わず，独自のルーチンで処理しているとある［5］。これまでのLINEST()(Excel2003より前のバージョン）の問題が多く報告されたことを考えると，検討する必要があろう。3節で描画エンジンが正確に数式のデータを曲線に変換できるかを検討する。MS社は正確に描画できると言っているが［4］，これも検証する必要があろう。また，“近似曲線の追加”機能を利用するとき，時々異常な現象が発生することが報告されている［1，4，5］。4節で“近似曲線の追加”機能の設計上の問題について検討する。最後に，本稿のまとめを行う。

2．各バージョンのチャートツールの回帰処理

MS社のKB829249［5］はチャートツールの回帰処理が関数LINEST()を使わず，独自のルーチンで処理していることと，線形回帰の定数項をゼロに指定して使う場合R^2がマイナスになることと解説している。しかし，定数項をゼロに指定していない場合の処理結果については言及していない。当然，利用者は処理結果には問題がないと解釈するであろう。ここでは，果たしてこのような解釈が妥当かどうかを検討する。

テストに利用するデータセットは下記のとおりである。

データセット1

x	y
10…001	10…00.000
10…002	10…00.000
10…003	10…00.900
10…004	10…01.100
10…005	10…01.010
10…006	10…00.990
10…007	10…01.100
10…008	10…00.999
10…009	10…00.000
10…010	10…00.001

n個の0

データセット2

x	y
10	3.2
15	3.8
30	5.2
45	5.9

データセット3

x	y
1	3.2
2	3.8
3	5.2
4	5.9

データセット１は Simonoff［10］のデータセットを x と y の値を変更できるようにアレンジしたものである。データのサイズ（組数）は10で，データ自身は何の意味も持たない。n（０～８）は０の個数を表わし，n＝０のとき，１行目のデータが x＝101，y＝10.000となる。n＝８のとき，Simonoff のデータセットになる。n を０から８まで変更しても，Excel の表示できるデータ範囲（およそ15桁の精度，Excel のヘルプで計算の仕様と制限を検索すれば表示される）を超えないことに注意されたい。データセット２と３は意味を持たないデタラメなデータである。

まず，Excel2000において，チャートの回帰処理と関数 LINEST()が同じ問題点を持っているかどうかを検討する。データセット１において，n＝０および n＝３～８のときのデータを作成して，散布図に変換する。作成した散布図に“近似曲線の追加”機能を利用して，線形近似直線を追加する。チャート上に出力された数式を表１と表２にまとめる。

表１　Excel2000のチャートの回帰出力と LINEST()の出力（y=ax+b）

		a	b	R^2
n＝０	＊１	2.93333333334886E－03	1.03005333333327E＋01	2.83131022819118E－04
（x が３桁）	＊２	2.93333333336549E－03	1.03005333333299E＋01	2.83131022816958E－04
n＝３	＊１	2.93331261837121E－03	9.70726545454546E＋03	2.83127028710063E－04
（x が６桁）	＊２	2.93329839270277E－03	9.70726402758856E＋03	2.83131021563253E－04
n＝４	＊１	2.93323863636364E－03	9.70677527272720E＋04	2.83111347270682E－04
（x が７桁）	＊２	2.93231130477975E－03	9.70682825675081E＋04	2.83130988699042E－04
n＝５	＊１	3.02663438256659E－03	9.67330401937046E＋05	3.01720561501965E－04
（x が８桁）	＊２	2.93321299638989E－03	9.70668463903430E＋05	2.83130991240578E－04
n＝６	＊１	0	1.11848106666667E＋07	0
（x が９桁）	＊２	1.78571428571429E－02	8.21428622607143E＋06	－7.04551917762973E－03
n＝７	＊１	（出力なし）	（出力なし）	（出力なし）
（x が10桁）	＊２	0	1.00000000610000E＋08	0
n＝８	＊１	－1.25000000000000E－01	1.07374182400000E＋09	－6.66666666666667E－01
（x が11桁）	＊２	－1.25000000000000E－01	2.25000000129750E＋09	－5.38274368539939E－01

＊１：Excel2000のチャートの回帰出力　＊２：Excel2000の LINEST()の出力

表２　Excel2000のチャートの回帰出力と LINEST()の出力（y=ax）

		a	R^2
n＝０	＊１	1.00496404363245E－01	－3.13158975602113E－01
（x が３桁）	＊２	1.00496404363245E－01	－3.13158975602077E－01
n＝３	＊１	1.00000599886930E－01	－3.09751669822629E－01
（x が６桁）	＊２	1.00000599886930E－01	－3.09751511229785E－01
n＝４	＊１	1.00000059998869E－01	－3.09731195622009E－01
（x が７桁）	＊２	1.00000059998869E－01	－3.09748062158085E－01
n＝５	＊１	1.00000005999989E－01	－3.09657320872274E－01
（x が８桁）	＊２	1.00000005999989E－01	－3.09747717317754E－01

n＝6 (x が9桁)	＊1	1.00000000600000E−01	−1.53846153846154E−01
	＊2	1.00000000600000E−01	−3.09747682747596E−01
n＝7 (x が10桁)	＊1	1.00000000060000E−01	0.0
	＊2	1.00000000060000E−01	−3.09747678528894E−01
n＝8 (x が11桁)	＊1	1.00000000006000E−01	3.33333333333333E−01
	＊2	1.00000000006000E−01	−3.09747652748181E−01

＊1：Excel2000のチャートの回帰出力　＊2：Excel2000のLINEST()の出力

表1と表2から次のことがわかる。

1．チャートツールの回帰出力とLINEST()は似たような動きをしている。小さいデータに対して一致する桁数が多く，大きなデータになればなるほど一致する桁数が少なくなる。表1ではxが7桁のときから，両方の結果（bの値）が著しく異なってきている。これはMS社が公表した別々のルーチンで処理していることと一致する。そして表2のxが11桁のとき，チャート上のR^2の出力が0.333333333333333であることを注意されたい。理論上R^2の値は0と1の間にあるので，この値だけを見て処理がおかしいと判断できない。

2．チャートツールの回帰処理もLINEST()と同じように，定数項bがゼロでない場合でも，マイナスのR^2値が出力される（表1でn＝8のケース）。したがって，チャートツールの回帰処理も信頼性が低いと結論できる。おそらく，LINEST()と同じ回帰処理のアルゴリズムが使われているものと推測できる。

次に検討するのはExcel2003とExcel2007がExcel2000のチャートツールの回帰処理と異なるかどうかである。MS社から変更したとの発表はないが，過去に発表していない変更もあったので，チェックしてみる。

ここで，定数項bがゼロでないケースだけを計算する（bがゼロの場合，KB829249［5］で結果が不正確になると言明している）。結果を表3にまとめた。表3も表1と表2と同じ方法でデータを収集した。そして，比較のため，Excel2003のLINEST()の結果も加えた。

表3　Excel2000，2003と2007のチャートの回帰出力(y=ax+b)

		a	b	R^2
n＝0 (x が3桁)	＃1	2.93333333334886E−03	1.03005333333327E+01	2.83131022819118E−04
	＃2	2.93333333334886E−03	1.03005333333327E+01	2.83131022819118E−04
	＃3	2.93333333333408E−03	1.03005333333333E+01	2.83131022816245E−04
	＃4	2.93333333333335E−03	1.03005333333333E+01	2.83131022815896E−04
n＝3 (x が6桁)	＃1	2.93331261837121E−03	9.70726545454546E+03	2.83127028710063E−04
	＃2	2.93331261837121E−03	9.70726545454546E+03	2.83127028710063E−04
	＃3	2.93330094127944E−03	9.70726377249053E+03	2.83124782616335E−04
	＃4	2.93333333334664E−03	9.70726053333200E+03	2.83131022818718E−04

n＝4	＃1	2.93323863636364E−03	9.70677527272720E+04	2.83111347270682E−04
（x が7桁）	＃2	2.93323863636364E−03	9.70677527272727E+04	2.83111347270682E−04
	＃3	2.93352069276752E−03	9.70670739393939E+04	2.83167520468981E−04
	＃4	2.93333333366634E−03	9.70672605330003E+04	2.83131022880175E−04
n＝5	＃1	3.02663438256659E−03	9.67330401937046E+05	3.01720561501965E−04
（x が8桁）	＃2	3.02663438256659E−03	9.67330401937046E+05	3.01720561501965E−04
	＃3	3.03755326704545E−03	9.69623738181818E+05	3.03486270728742E−04
	＃4	2.93333333366635E−03	9.70667260530003E+05	2.83131022895016E−04
n＝6	＃1	0	1.11848106666667E+07	0
（x が9桁）	＃2	0	1.11848106666667E+07	0
	＃3	1.28355061349693E−02	8.71669006134969E+06	5.06684631919927E−03
	＃4	2.93333332872751E−03	9.70666726099391E+06	2.83131022022914E−04
n＝7	＃1	（出力なし）	（出力なし）	（出力なし）
（x が10桁）	＃2	（出力なし）	（出力なし）	（出力なし）
	＃3	9.78381374722838E−02	2.08320198669623E+06	2.15855390798226E+00
	＃4	2.93333313681861E−03	9.70666674570481E+07	2.83130986241493E−04
n＝8	＃1	−1.25000000000000E−01	1.07374182400000E+09	−6.66666666666667E−01
（x が11桁）	＃2	−1.25000000000000E−01	1.07374182400000E+09	−6.66666666666667E−01
	＃3	7.56985294117647E+00	−7.46882180517647E+10	2.02946968826593E+01
	＃4	2.83130986241493E−04	9.70666647551264E+08	2.83131410198266E−04

＃1：Excel2000のチャートの回帰出力　＃2：Excel2003のチャートの回帰出力
＃3：Excel2007のチャートの回帰出力　＃4：Excel2003の LINEST()

表3から次のことが読み取れる。

1．Excel2003と Excel2000のチャートツールの回帰処理はまったく同じである。

2．Excel2007のチャートツールの回帰処理は変更されている。しかし，依然不正確な結果が出力されている。たとえば，n＝7（x が10桁）の場合，R^2が2.16で，n＝8（x が11桁）の場合，R^2が20.29であった。これは R の理論値よりかなり離れている。

3．Excel2003の LINEST()の結果と比較すれば，データが小さい場合（x が7桁以下），四つの処理結果がほぼ同じである。データが大きい場合，Excel2000，Ex2003と Excel2007のチャートの回帰処理結果はともに不正確になる。

以上から，MS 社は Excel2007のチャートツールの回帰処理を変更したが，問題がクリアできていないことがわかる。なぜ Excel2003の LINEST()を使わないのかは不明である。

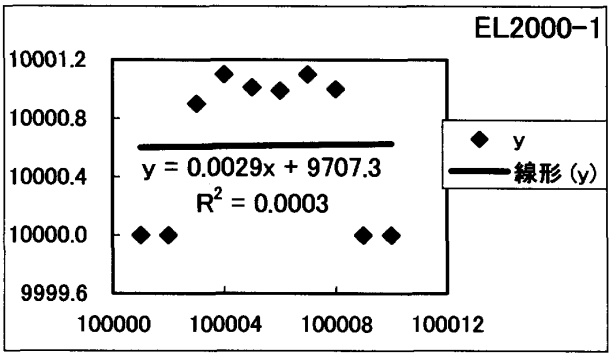
3．描画エンジンの振るまい

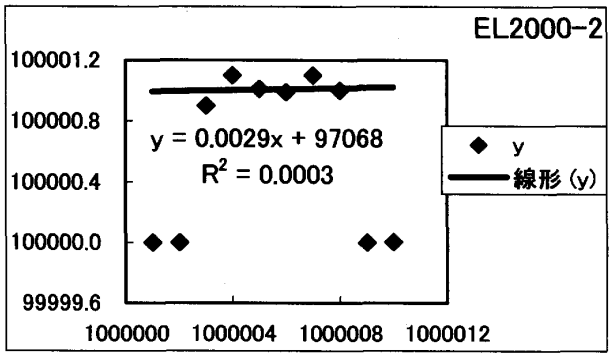
作成されたチャートの正確さはアルゴリズム以外に描画エンジンの動きもひとつの要因となる。ここで，チャートの作成過程を追って描画エンジンの動きを検討する。なお，表3から Ex-

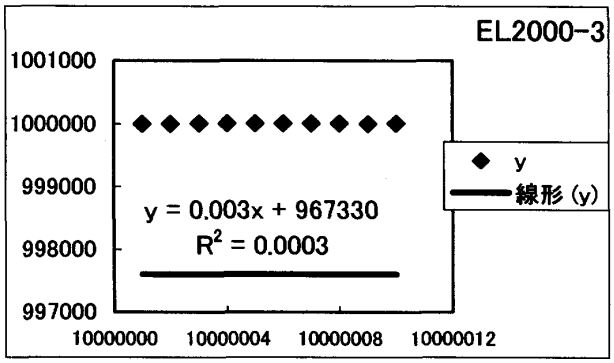
cel2003とExcel2000のチャートツールの回帰処理は同じなので，Excel2000とExcel2007のチャートについて検討する。

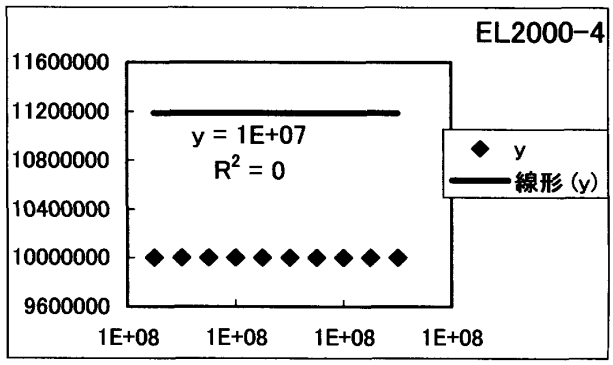
チャートはデータセット1のn＝3～8（xが6桁から11桁まで）のケースに対して作成した。各データに対して，正確なチャートのイメージは図EL2000-1のようで，直線がほぼ真ん中を通る。しかし，作成されたチャートには上へ偏ったり(EL2000-2，EL2000-4)，下へ偏ったり（EL2000-3，EL2000-6），そして，凡例にマークが表示されただけで，直線は作成されていない（EL2000-5）など多様である。Excel2007のチャートも，同様なパターンが読み取れる。次に，これらの現象が描画エンジンによるものかを検討する。

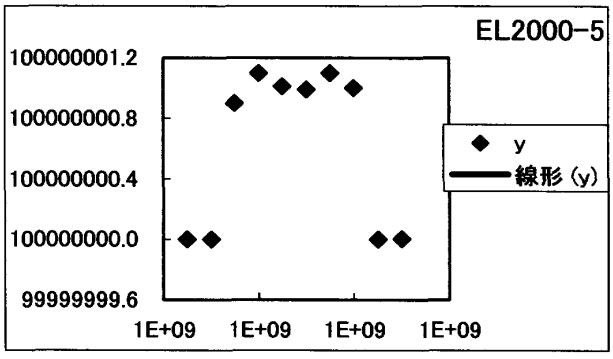
試しに，EL2000-3の回帰式にxの値を代入してyの予測値を計算した。その結果は表4にまとめた。

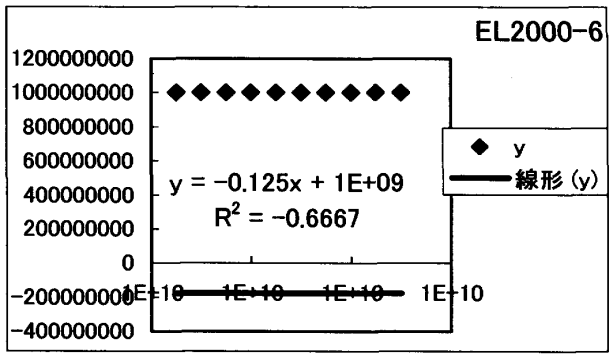


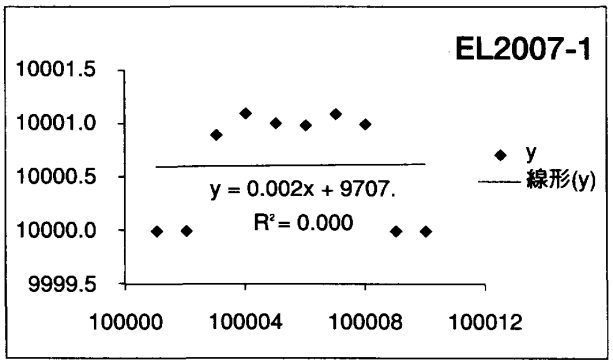


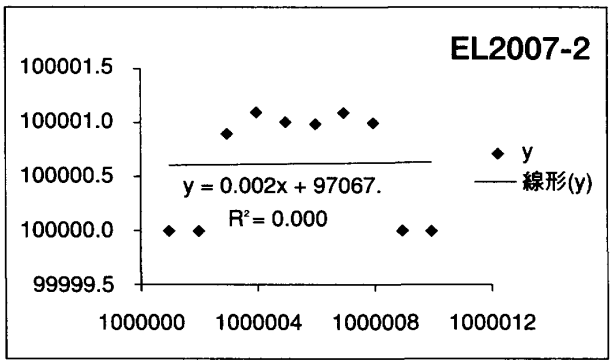


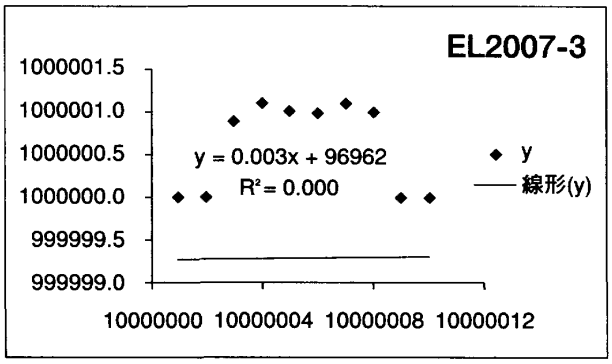


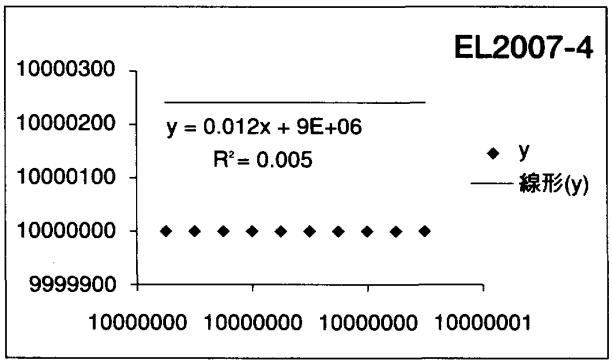


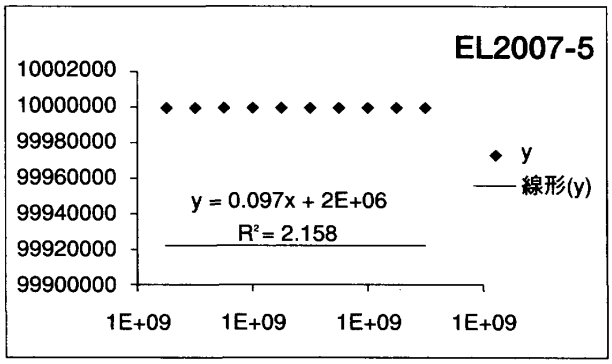












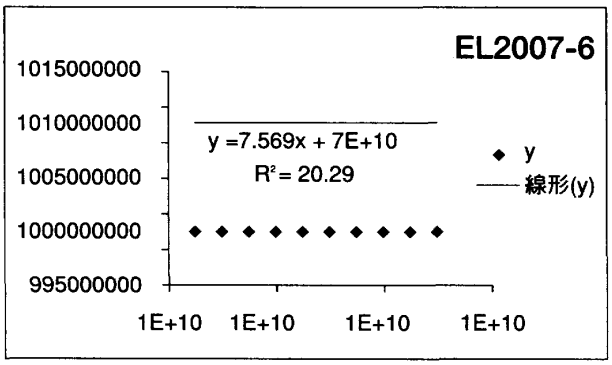


表 4　EL2000- 3 のデータ

x	y	y の予測値
10000001	1000000.000	997596.749
10000002	1000000.000	997596.752
10000003	1000000.900	997596.755
10000004	1000001.100	997596.758
10000005	1000001.010	997596.761
10000006	1000000.990	997596.764
10000007	1000001.100	997596.767
10000008	1000000.999	997596.770
10000009	1000000.000	997596.773
10000010	1000000.001	997596.776

表４からわかるようにEL2000-３の回帰直線が下方へ偏っているのは回帰式の結果であり，描画エンジンは忠実にデータを描いたことがわかる。同じ理由で，ほかの上下へ偏っていた直線も回帰式によるものであって，描画エンジンの不正ではない。しかし，EL2000-５はどういう原因か不明である。凡例に線形のマークが作成されたことは描画エンジンが作動していたが，何らかの原因で線の描画ができなかったと考えられる。これは描画エンジンに渡された回帰式に異常発生したか（たとえば，NaN），あるいは描画エンジンの異常であるかは同定できないが，直線が作成されていないので，描画エンジンの欠陥と結論付けても問題にはなるまい。

Excel2007の各チャートと対応のExcel2000のチャートと比較すると，相対的に偏りが小さいことがわかる。しかし，xの値が大きくなると偏りも大きくなることがわかる。

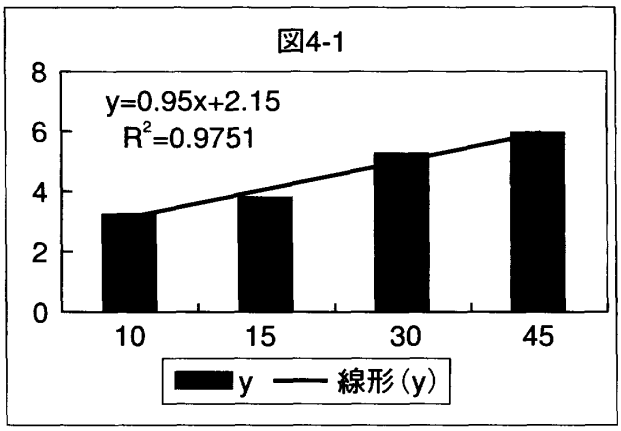
なお，Excel2007のチャートにはバグ２つがあった。１つはEL2007-３のbの出力である。bの桁数が６桁だが，デフォルトのまま５桁しか出力されない。もう１つはEL2007-４に現れた。水平軸の数値が本来９桁なのに，デフォルトのままでは８桁しか表示されない。また，デフォルトのままで出力された数値の端数処理がゼロに近い方に丸めるという方法をとっていることに注意されたい。たとえば，EL2007-１にa＝0.002と表示されたが，通常の丸め方法ではa＝0.003となる。ゼロに近い方に丸めるという方法はIEEE754規格のひとつであるが，同規格の中にあり，普通使われている近い値に丸めるという方法を採用していない理由は不明である。

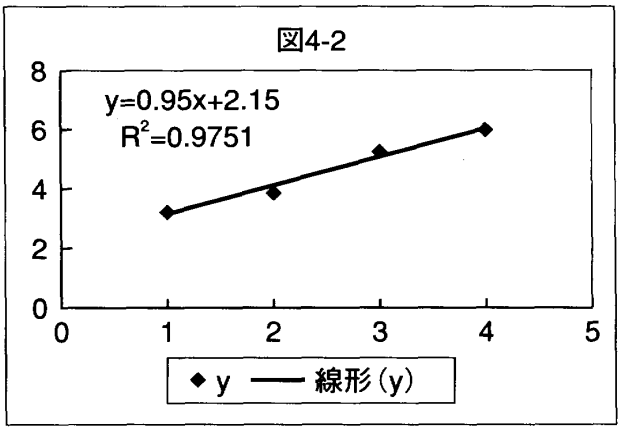
４．“近似曲線の追加”機能の設計上の問題

Excelのチャートの中で，“近似曲線の追加”機能を適用できるのは散布図，面グラフ，縦棒グラフ，横棒グラフ，折れ線グラフ，株価チャートとバブルチャートのようである。これらのチャートに近似曲線の数式を表示させるとき，まったくデータに合わない数式が出力されることがある。右図がその例である。

図４-１と図４-２がデータセット２と３を縦棒グラフと散布図に変換したものである。両データセットはまったく異なるが，出力された数式とR^2値はまったく同じである。

実は，Excelのチャートでは横軸に数値を表示するものと文字列を表示するものの２種類がある［11］。設計上，縦棒グラフの横軸が等間隔に文字列として表示され，散布図の横軸が数値軸として表示される。図４-１を見ればスケールが合っていないことは一目瞭然である。横軸が文字列の場合でも，回帰処理が可能な理由はExcelが自動的に１から順に整数を割り当てているからである［４］。当然，出力された数式とR^2値はナンセンスな値である。これらの理由で，図４-１と図４-２の数式とR^2値がまったく同じとなる。MS社がこの事実を認識してい





たにもかかわらず［4，5］，数式とR^2の出力を無効にしていないのは誤解を招く一因となろう。

なお，縦棒グラフと散布図のケースがAoki［1］サイトに比較があった。本稿で付け加えたいのは縦棒グラフだけの異常現象ではなく，両方の軸が同時に数値軸でない場合，必ず縦棒グラフのようなナンセンスな数式とR^2が計算される。これは，Excelの設計ミスの一例と考えてよいと思う。

5．おわりに

本稿では，Excelのチャートツールのオプション，“近似曲線の追加”機能について検討してきた。移動平均をのぞいて，残りの近似曲線がすべてアルゴリズムに起因する結果の不安定性という問題がある。適用できるデータの大きさ，データセットの大きさがかなり制限されている。加えて描画エンジンにも欠陥が見られる。また，利用するときチャートの軸が両方とも数値軸であるかどうかを注意する必要がある。片方が数値軸ではない場合，数式やR^2の出力はまったく無意味（ナンセンス）である。

今は計算の安い時代になっている。したがって，チャートが正確かどうかわからない場合，別のソフトを利用してチェックしたほうがより安全である。フリーソフトのR（www.r-project.org）は強力な作図機能を持ち，よく検証されていて信頼性が非常に高い。ひとつの解決法と考えてよい。

謝辞　ご精読を頂きました経営学部の福田　馨氏に心より感謝の意を表します。

参考文献

1．Aoki,S. Sep. 2007. aoki2. si.gunma-u.ac.jp/Hanasi/excel/index.html

2．Heiser, D.A. 2007. TECHNICAL ARTICLES AND REPORTS, www.daheiser.info

3．Foxes team, Dec. 2007. digilander.libero.it/foxes/StRD_Benchmarks/X_NIST_StRD_Excel 2003_Results.htm

4．KB211967. support.microsoft.com/kb/211967/en-us (December 2007)

5．KB829249. support.microsoft.com/kb/829249/en-us (December 2007)

6．KB828533. support.microsoft.com/kb/828533/ja (December 2007)

7．McCullough,B.D., Wilson,B., 1999. On the accuracy of statistical procedures in Microsoft Excel 97, Comput. Statist. Data Anal. 31, 27-37.

8．McCullough,B.D., Wilson,B., 2002. On the accuracy of statistical procedures in Microsoft Excel 2000 and XP. Comput. Statist. Data Anal. 40, 713-721.

9．McCullough,B.D., Wilson,B., 2005. On the accuracy of statistical procedures in Microsoft Excel 2003. Comput. Statist. Data Anal. 49, 1244-1252.

10. Simomoff, J.S. 2006. Statistical analysis using Microsoft Excel. Available on www.stern.nyu.edu/~jsimonof/classes/1305/pdf/excelreg.pdf

11. Walkenbach, J. 2003. Excel Charts, Wiley Publishing, Inc. pp 147.