

Excel を用いた情報可視化技術をめぐる一考察

— データと図表の関連付け—

An Essay on Information Visualization Skills via Excel : Associating Data with Charts

持元江津子

Yoshimoto Etsuko

要 約

本論文では、広く普及している表計算ソフトの Excel が装備しているグラフ描画機能に注目し、統計データの尺度水準を軸にデータと図表を関連づけ、Excel を用いた情報可視化技術がより有効に活用されるための具体的な作業方法について整理した。そして、初歩的な図表例を多く挙げ、尺度水準を考慮したデータの扱い、つまりデータの Excel 適応化がデータ可視化を容易にし、それ自体がデータ分析上有用な技術となることを明らかにした。

Key Words : Excel, 情報可視化技術, データ, データの Excel 適応化, 尺度水準, 表, グラフ

1. はじめに

情報可視化手段としての表計算ソフト Excel の活用方法はさまざまである。最も身近に感じられる活用方法は、セルに入力した、言葉及び数値で表されたデータから決められた手順を踏んでグラフを描くことだろう。Excel では、関数や VBA を用いたグラフ以外の図示的表現も可能である。データ表や言葉のみで構成された各種様式等を含む多様な表形式の文書作成（表計算機能を活用することもある）や、データの表計算ツール作成（資金管理など会計処理の簡易化をめざすものなど）も、情報可視化技術の範疇にあると捉えられる。

潜在的ないし顕在的に情報を内包するデータを処理するという観点から、情報可視化のために Excel を有効活用するには、データの特徴を見極めつつデータ及びデータをまとめた表を使い回して、Excel のもつグラフ描画のバリエーション機能を活用してゆく技術が必要である。そのような技術を使いこなすことができれば、データが潜在的に持つ情報を目に見える形にし、顕在的に持っている情報さえより鮮明に浮かび上がらせ、より豊かなデータ分析が行えるであろう。

そこで、本論文では、Excel のグラフ描画機能を有効活用する前提条件となる技術として、統計データの特徴の見極めに統計データの尺度水準を考慮することを提案する。それは、初歩的でありながら、Excel のグラフ描画機能を中心とする情報可視化技術を支える重要な要素となる。よって、本論文では、統計データの種類と Excel で描画可能な基本的なグラフとの関係を、尺度水準を軸として整理することを目的とする。この問題はこれまで十分には検討されてこなかったものであり、筆者にとっても積み残していた課題であった（拙稿^[1]参照）¹⁾

尚、本論文では Excel 2007 または Excel 2008 for Mac の使用を前提とする。これらは本論文

の執筆時点における最新版ではないが、その使用はその時点で最新版よりも十分に普及しているという理由により妥当しているであろう。

2. 尺度水準とデータの計測

統計データを分類するための尺度水準（Level of Measurement）についての考察は、1946年に S. S. Stevens が著した論文「On the Theory Scales of Measurement（計測の尺度の理論）」^[2]が有名である。現代において尺度水準として理解されている概念は、概ねこれに基づいている。これによると、尺度は、①名義尺度（Nominal Scale）、②順序尺度（Ordinal Scale）、③間隔尺度（Interval Scale）、④比例尺度（Ratio Scale）の4種に類別される。それぞれの性質は次の通りである。

①名義尺度（Nominal Scale）においては、何らかの現象において場合分けを行いたいとき、それぞれの場合に名義（カテゴリーの名称）を与え、その名義に当てはまる事象の個数や人数などを計測し、データとする。各カテゴリーの間で重要さなどの程度の別はなく、それぞれの名義は対等である。

②順序尺度（Ordinal Scale）では、好感度や満足度などの程度について順序付けを行い、それぞれの順序付きカテゴリーに当てはまる事象の個数や人数などを計測し、データとする。データは言葉で表され、順序はあるが、数量的な差異は不明である。

③間隔尺度（Interval Scale）には、順序を有し、事象を数値で計測できる故に得られたデータの差異を数量的に示しうるものが当てはまる。「1と2」、「2と3」といった数値間の間隔は一定である。摂氏や華氏で測る温度がこれに該当する。「0（ゼロ）」は任意に与えられ、相対的な意味しか持たない。

④比例尺度（Ratio Scale）は、順序など③間隔尺度に対応する性質を全て備え、且つ、「0（ゼロ）」に絶対的な意味のあるデータが当てはまる。人口や重さなど主に物理的に観察される量を表す数値データがこれに該当する。

4種の尺度水準とデータの計測については上述の通りである。統計データとしての扱いやすさと精度の高さという観点に基づくと、①名義尺度から④比例尺度まで順に水準が上がっていくものと解釈できる。尚、①と②は質的データ、③と④は量的データとして計測される。

次章では、データを Excel のワークシート上でのグラフ描画に適応させるために、どのようなことを行えるのかについて尺度水準を軸にして考察する。

3. 尺度水準とデータの Excel 適応化

以下、4種の尺度水準を軸に、主にデータの Excel 適応化について考察する。

①名義尺度に該当するデータの場合：

名義そのものに意味があるため、集計表を作成する際などに、名義の代わりに数字を割り振るなどの加工はあまり行われな。しかし、名義が2種類のみである場合には、ダミー変数「1,0」を与え、数量として扱えるようにできる。この場合、データは③間隔尺度に昇格する。

また、各名義カテゴリーに当てはまる事象の個数などのデータを百分率(%)で表すことも多い。どの名義カテゴリーの個数が多かったか、どの内訳比率(%)が大きかったかなどが重要になる。

②順序尺度に該当するデータの場合：

この場合、①名義尺度の場合に似て、個数などのデータを百分率(%)で表し、順序カテゴリー一別の内訳比率を明らかにすることが少なくない。

また、データ分析上、順序という観点で並べて中央に位置するデータが、より程度の高い方に振れているのか、あるいはより低い方に振れているのかなどに注目することもある。しかしながら、数量的な差異が不明であるため、原データの分布は明らかでない。

また、Stevens は既述の論文^[1]において「Intelligence (知性)」を例にとり、これは本来②順序尺度に該当するが、知性が「0 (ゼロ)」の間はおらず、等間隔の数値を割り振れば、③間隔尺度に昇格すると述べている。

この Stevens の言及と同様に、好感度や満足度にも高い方から「5, 4, 3, 2, 1」というように数字を割り振ると、データの尺度水準が③間隔尺度に昇格する。

しかし、好感度には「好悪」という相反する心理、満足度にも「満足または不満」という相反する心理が含まれ、「どちらでもない」という「0 (ゼロ)」相当の立場が順序の中心にある。とはいえ、数値化の際に「0 (ゼロ)」という数字に与えることはまずない。マイナスの順序というのが実感しづらいということもある。よって、順序カテゴリーに「0 (ゼロ)」以下の数値を与えることがあまりなく、これらは④比例尺度に昇格し得ない。

③間隔尺度に該当するデータの場合：

Excel でグラフ描画を行うために特別な準備をすることはあまりない。より高度な分析を行うために、例えば温度の移動平均などを Excel 上で計算しておくことはありうる。また、平均値やデータのばらつき具合を示す標準偏差などが重要視されやすく、それらは Excel 関数を利用すれば容易に算出できる。また、重回帰分析等に備えてデータの標準化を行うこともありうる。

④比例尺度に該当するデータの場合：

上述の③間隔尺度の場合と同様、Excel でグラフ描画を行うために特別な準備をすること

はあまりない。より高度な分析のために、指数や百分率(%)、千分率(‰)で表された変化率ないし変動率等々をExcel上で計算することはある。

以上より、データのExcel適応化に関して、質的データである①名義尺度と②順序尺度に該当するデータの数値化ないし数量化は可能かどうかの判断が、まず鍵となるであろう。

4. 尺度水準と表、グラフの関係

本章では、尺度水準に応じてデータ集計表などをどのように整理し、どのようなグラフを描いていくのかについて、具体例を挙げながら検討していく。

①名義尺度に該当するデータの場合：

まず表1のようなデータが得られたとする。²⁾「好き」というカテゴリに入る人数と「嫌い」の人数を集計し、適切な集計結果表に書き換えたものが表2である。表2から人数に注目した棒グラフ(図1)と内訳比率を明示する円グラフ(図2)を描くことができる。

グラフ化するにあたり表2のような集計結果表を書くことは、Excelを活用する情報可視化技術として重要である。もっともExcelを使用する場合は、内訳比率の計算をしなくても円グラフを描く際に自動的にパーセンテージを表示させることができるので、このような作業の手間を省くことは可能である。

また、データをダミー変数化することにより、③間隔尺度のデータと同様に、カテゴリ別の人数集計が容易になって

表1 グラフの読み書きについて

個人整理番号	回答：ダミー変数 (好き 1, 嫌い 0)	
1	好き	1
2	嫌い	0
3	嫌い	0
4	嫌い	0
5	嫌い	0
6	嫌い	0
7	嫌い	0
8	嫌い	0
9	嫌い	0
10	嫌い	0
11	好き	1
12	嫌い	0
13	嫌い	0
14	嫌い	0
15	好き	1
16	好き	1
17	嫌い	0
18	嫌い	0
19	嫌い	0
20	嫌い	0
21	嫌い	0
22	嫌い	0
23	嫌い	0
24	嫌い	0
25	嫌い	0
26	嫌い	0
27	嫌い	0
28	嫌い	0
29	嫌い	0
30	嫌い	0
31	嫌い	0
32	嫌い	0
計		4



図1 グラフの読み書きについて (計32人)

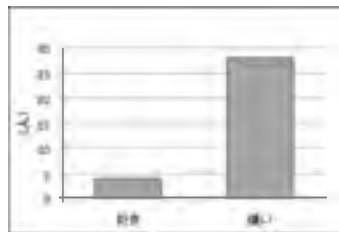


図2 グラフの読み書きについて

表2 グラフの読み書きについて

回答	(人)	(%)
好き	4	12.5
嫌い	28	87.5
計	32	100

いることにも注目すべきであり、これも有用な技術の一つである。³⁾

②順序尺度に該当するデータの場合：

表3は、小中学生が携帯電話を持つことに対する意見についてのデータである。表1と表2の場合と同様に、表4を作成する。同様に図3の円グラフと図4の棒グラフも描画する。図3の円グラフでは、賛成及び反対の順序に応じて円グラフのパイの取り方を決めている。多くの円グラフは内訳比率の大きい順に12時の位置を基線としてパイを時計回りに取るが、グラフを読む際の論理的な思考を助けるべく図3のように臨機応変に対応するのが望ましい。こうしておけば、賛成側と反対側のいずれの回答者が過半を占めるのかも伝わりやすくなる。

③間隔尺度に該当するデータの場合：

ある決まった地点で観測した1日の温度変化などを折線グラフで表すことが想定できるが、ここでは表3のデータをどう扱うかに光を当てることとする。

回答「賛成～反対」に「賛成」から順に「4, 3, 2, 1」という数値を与え、表5を作成する。無回答は有効回答でないため、例えば「0」などの数値をこれに割り振ることはない。無回答の3人を除いた算術平均が「約2.97」で、中間の「2.5」より大きくこの集団の回答の傾向はやや賛成寄りであることが分かる。そして、図5の折線グラフを新たに描く。このようなグラフを描くことにより、やや賛成寄りの回答の多かったという情報を可視化できるのである。

尚、図5では無回答が含まれるため、ところどころで折

表3 小中学生の携帯電話を持つことについて

個人整理番号	回答
1	どちらかと言えば賛成
2	どちらかと言えば反対
3	どちらかと言えば賛成
4	どちらかと言えば賛成
5	賛成
6	どちらかと言えば賛成
7	どちらかと言えば賛成
8	賛成
9	無回答
10	どちらかと言えば反対
11	どちらかと言えば反対
12	どちらかと言えば賛成
13	どちらかと言えば賛成
14	どちらかと言えば反対
15	賛成
16	無回答
17	どちらかと言えば賛成
18	どちらかと言えば賛成
19	どちらかと言えば反対
20	どちらかと言えば賛成
21	賛成
22	賛成
23	無回答
24	どちらかと言えば賛成
25	どちらかと言えば反対
26	どちらかと言えば反対
27	反対
28	賛成
29	賛成
30	賛成
31	どちらかと言えば賛成
32	どちらかと言えば賛成

表4 小中学生の携帯電話を持つことについて

回答	(人)	(%)
賛成	8	25.0
どちらかと言えば賛成	13	40.6
どちらかと言えば反対	7	21.9
反対	1	3.1
無回答	3	9.4
計	32	100.0



図3 小中学生の携帯電話を持つことについて (計32人)

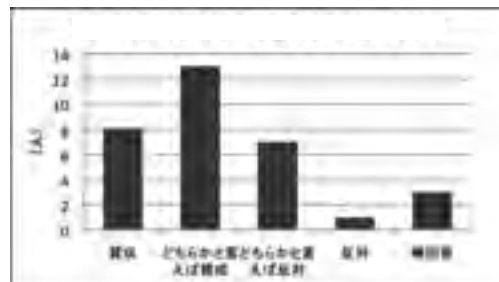


図4 小中学生の携帯電話を持つことについて

線グラフが途切れている。無回答は有効回答でないため、必要に応じ無回答を排除して作表・作図することもある。

④比例尺度に該当するデータの場合：

人口や貿易統計の長期的推移などのデータのグラフ化を思い浮かべるとよい。本論文では我が国の輸出入額の推移を例にグラフ描画を試みる^[3]。表6は輸出入額と前年比増減率を示している。ここでは前年比増減率をExcel上で計算している。この表をもとに図6から図10までのグラフの描画が可能である。特に図6を作成した後は、Excelのグラフの種類変更機能を使用することで、容易に図7と図8を作成することが可能である。

図9には若干の手間をかけている。輸入額の全てのデータの値をExcel上で負の値に置き換えて入力する。図8をコピーしてデータの選択を変更する。具体的には輸入のデータ範囲を負に置き換えた値の範囲に差し替える。さらに輸入額の目盛数値に「-（マイナス）」が表示されないよう縦軸の書式設定を変更する。このようにすれば図9が描ける。

図10のような前年比増減率の折線グラフを描くと、正負に揺れる変動の実態を、年別の数量変化を表す棒グラフなどよりも詳細且つ正確に伝えられることが分かるであろう。

こういったグラフのバリエーションを試すことも、Excelを活用する情報可視化技術の一つであり、データ分析に有用である。

5. まとめ

以上より、統計データの尺度水準から最適なグラフを1対1対応で決めることは難しいが、尺度水準を判別すれば、Excel上でデータにどのような手を加えるとよりよいグラフが描けるのかについて、ヒントを得られることが明らかになった。データの尺度水準を考慮しながらグラフを描画しやすくする表を整え、必要に応じてデータを加工し、データの尺度水準を引き上げるなどしてExcelのグラフ描画機能を有効活用すること、それがExcelを活用する情報可視化技術の重要な一部を占めている。

表5 小中学生の携帯電話を持つことについて

個人整理番号	回答 対立～反対:	
1	どちらかと書えば賛成	3
2	どちらかと書えば反対	2
3	どちらかと書えば賛成	3
4	どちらかと書えば賛成	3
5	賛成	4
6	どちらかと書えば賛成	3
7	どちらかと書えば賛成	3
8	賛成	4
9	無回答	
10	どちらかと書えば反対	2
11	どちらかと書えば反対	2
12	どちらかと書えば賛成	3
13	どちらかと書えば賛成	3
14	どちらかと書えば反対	2
15	賛成	4
16	無回答	
17	どちらかと書えば賛成	3
18	どちらかと書えば賛成	3
19	どちらかと書えば反対	2
20	どちらかと書えば賛成	3
21	賛成	4
22	賛成	4
23	無回答	
24	どちらかと書えば賛成	3
25	どちらかと書えば反対	2
26	どちらかと書えば反対	2
27	反対	1
28	賛成	4
29	賛成	4
30	賛成	4
31	どちらかと書えば賛成	3
32	どちらかと書えば賛成	3

図5 小中学生の携帯電話を持つことについて

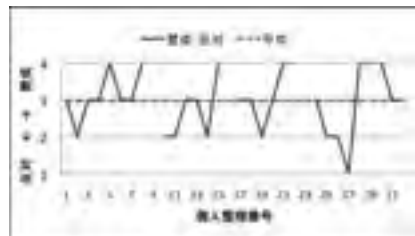


表6 年別輸出入額（確定値）と前年比増減率の推移

年	輸出		輸入		'95	2935247958	304379502202	310
	年	前年比%	年	前年比%				
1990	298301057		342795551		'95	33453964502	139	37464145741
1991	458776775	54.0	737247958	117.0	'96	34432500947	75	37556307374
1992	458243179	-0.2	730307652	-0.9	'97	34809206598	14	38074754056
1993	458340428	0.2	857469443	16.8	'98	40320293751	15.9	38321126549
1994	585535037	27.8	851385437	-0.4	'99	35789713887	-15.2	3857117070
1995	773115948	31.4	859749911	0.1	'00	33375797321	-5.6	389112573
1996	907229071	24.4	1162704360	39.7	'01	33939782158	1.9	24706313539
1997	1326985536	44.3	1542950900	32.6	'02	37822524626	11.4	2327677281
1998	1505567826	0.6	1097574295	-29.7	'03	41456925674	9.6	31835297238
1999	1244387203	-20.2	1205876732	8.7	'04	42537892974	2.2	3790151527
1990	1439627161	11.3	1678627363	24.6	'05	430226444	1.1	27527413260
1991	1524314578	4.9	2097755201	29.4	'06	43202448725	6.5	28826357239
1992	1789317267	18.1	2029748890	-3.0	'07	43487557887	0.7	28704207343
1993	1567767144	-10.5	2427387177	19.3	'08	41530595171	-4.5	248751331
1994	1467348567	-7.4	2557510493	7.8	'09	4473131706	7.7	37954706
1995	3042627304	23.7	2940846741	2.9	'10	50937997859	13.9	40336162573
1996	359550700	-13.7	142877255	-16.6	'11	5064500938	-0.6	385364723
1997	3759765207	8.8	479777497	33.3	'12	47347558341	6.1	33786100663
1998	4569735348	24.2	4675427477	1.4	2000	5165797720	9.1	40336422768
1999	573645762	-21.3	5408472791	-5.7	2001	48979244371	-5.2	42475533002
1990	495426759	-20.4	6797290595	25.7	2002	5710395735	17.4	4727503245
1991	3347763283	-20.1	650906130	-7.7	2003	54948300137	-4.7	4476707137
1992	3306072348	-4.9	7225978235	4.6	2004	61159972054	12.1	43276632346
1993	7001426255	13.5	7424335047	43.9	2005	65656344157	7.3	5224935121
1994	7620752927	8.8	7606321923	-0.1	2006	75246772322	14.6	61344292077
1995	7634370717	0.1	7710026976	-0.1	2007	85931437872	11.5	71305296427
1996	9534678464	20.5	79227468610	-2.0	2008	81070807877	-5.7	72294747226
1997	27648970427	8.6	7913779700	0.3	2009	54170614283	-33.1	57459777779
1998	70355840363	-3.0	7677024005	-17.6	町原省貿易統計			
1999	72331578857	9.6	24245730997	44.9	http://www.itsm.go.jp/2009/01/11/01/01.html			

図6 年別輸出入額（確定値）の推移

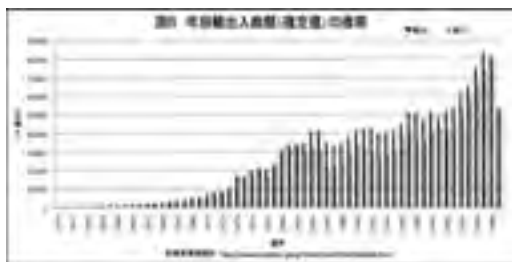


図7 年別輸出入額（確定値）の推移

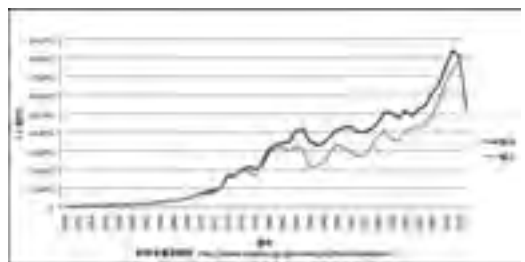


図8 年別輸出入額（確定値）の推移

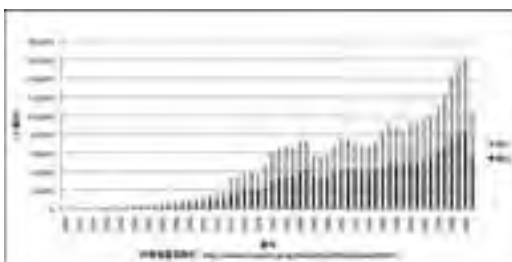
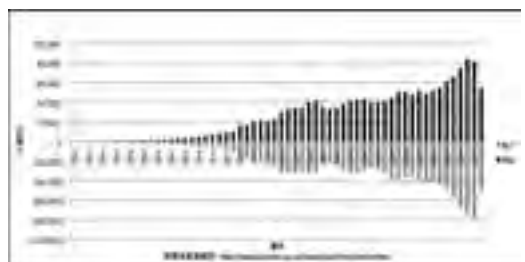
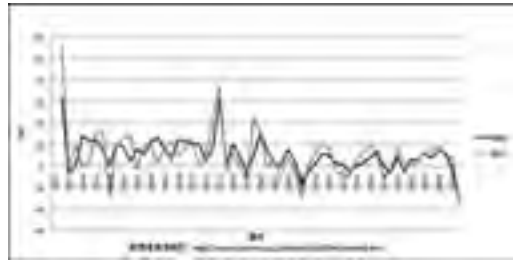


図9 年別輸出入額（確定値）の推移



しかしながら、よりよいグラフ描画という観点からは、データの尺度水準と最適なグラフ選びとの関連性について、Excelから一旦離れた上でのさらなる検討が必要となろう。これを筆者の今後の課題としたい。

図10 年別輸出入額（確定値）前年比増減率



【注】

- 1) 拙稿^[1]より「データにふさわしいグラフを判断するコツとして、尺度水準とグラフとの関係を可能な限り示した。とはいえ、この部分はグラフ化スキルを向上させるのに肝要であるにも関わらず、実際には判断基準の開発段階にある。今後、指導する側が考察・検討をもっと重ねて、なるべく早期に理論化すべき部分と考えられる。」(p.121)
- 2) 表1は聖泉大学における平成22年度統計学受講生対象に行ったアンケート結果を活用している。アンケートは2010年4月20日の統計学講義時に実施した。以下、本論文で示す図表は断りが無い限り上記アンケート結果を活用したものである。
- 3) ダミー変数化等の数値化ないし数量化により、基礎統計量を算出することができ、③間隔尺度や④比例尺度のデータと同等の回帰分析や因子分析など高度な分析が可能になる。

【引用・参考文献】

- [1] 持元江津子, 「リテラシー教育を考慮した統計学講義への取り組み」, 『聖泉論叢 (17)』, 2010年, pp.117-126。
- [2] S. S. Stevens: “On the Theory of Scales of Measurement”, Science, New Series, Vol. 103, No. 2684, 1946, pp.677-680, American Association for the Advancement of Science.
- [3] 財務省貿易統計: <http://www.customs.go.jp/toukei/info/index.htm> 情報最終確認日 2011年1月5日。