統計データの可視化の試み
―図形による表現とその描画プログラム―

高　橋　昌　也（短期大学部，情報メディア学科）

An Attempt for visualizing some statistical data
―Representation by figures and their drawing programs―

Masaya TAKAHASHI (Department of Information and Multimedia Technology)

Abstract

We offer a class on fundamental statistical processing at Department of Business and Information Technology, Fukuoka Institute of Technology, Junior College. In the recent years, the intelligibility and satisfaction of students' evaluation is decreasing every year slightly. In this paper, we discuss the factors and the visualization of educational materials and statistical data as the measure. Especially, we discuss the visualizations of mean, median, mode, variance, standard deviation and correlation coefficient, and discuss the programming for the visualizations. Furthermore, we also consider the relation between the strength of the correlation and the correlation coefficient in this paper.

Key words: *Statistical processing, statistical data, students' evaluation, visualization, programming*

1．はじめに

最初に，本稿では統計処理により作成された図表や算出された数値のことを「**統計データ**」と定義する。例えば，ヒストグラムや度数分布表，回帰直線のようなグラフや表，平均値，標準偏差や相関係数のような数値などである。

社会に出て仕事をする上で統計処理は重要なツールであることは，様々な媒体の巻頭ページ等で指摘されている。[2),4−6),8),10−11),13),17−18),20−21)]そこで，本学ビジネス情報学科では「ビジネス統計学」という科目を設置し，基礎的な統計処理の方法と，その方法を用いて Excel で処理する方法を習得できるように，初歩的なテーマを選び，例題を用いて解説している。2 年生後期の選択の専門科目であるので，受講者は統計処理そのものにある程度興味がある学生，若しくは，専門科目の卒業要件単位をある程度残してしまっている学生である。

授業はパソコンによる演習形式であり，内容は以下の(01)～(15)のとおりである。

(01) 授業の進め方，情報の要約
（度数分布表とヒストグラム）
(02) 基礎統計量の計算－Ⅰ（平均値，中央値，最頻値），演習問題
(03) 基礎統計量の計算－II（分散，標準偏差），演習問題
(04) データの相関－Ⅰ（散布図，相関係数），演習問題
(05) データの相関－II（順位相関），演習問題
(06) データの相関－III（無相関の検定），演習問題
(07) 演習問題の解説－Ⅰ
(08) これまでの振り返りと理解度の確認－Ⅰ
(09) データの相関－IV（相関表），演習問題
(10) データの相関－Ⅴ（相関表から相関係数を求める），演習問題
(11) データの相関－VI（独立性の検定），演習問題
(12) 回帰直線－Ⅰ（回帰分析），演習問題
(13) 回帰直線－II（分散分析表と検定），演習問題
(14) 演習問題の解説－II
(15) これまでの振り返りと理解度の確認－II

なお，教科書として『すぐわかる統計解析』[4)]（以降では簡潔に「現在の教科書」と表記する）を使用し，また補助教材として適宜プリントを配付している。

毎回の授業は原則として**「基礎」→「実用」→「実践」**の順に段階的に進めていく。ここで，「基礎」「実用」「実践」はそれぞれ以下のとおりである。

平成30年 5 月29日受付

基礎：　現在の教科書[4]を使って**各統計データの意義**や計算方法を説明する。

実用：　現在の教科書[4]等の例題に対して，実際に Excel を使用して統計データを作成する手順を，**作成手順書**としてプリントにて配付し説明する。

実践：　演習問題を配付し統計データを作成させる。

各段階を補足すると以下のようになる。

「基礎」の段階：　数学的な根拠や証明はまったく行わない。また，各統計データの意義として「それぞれの統計データが何を意味するものなのか，何を表現しているものなのか」ということを説明する。（付録の図11に統計データの１つである相関係数の計算方法を説明しているページのコピーを掲載しておく。）

「実用」の段階：　統計データを作成する手順を説明する「作成手順書」とその手順書通りに作成した Excel シートを配付し，それらを使ってモニター画面を通して作成手順を説明する。このとき，現在の教科書の計算方法を代行する Excel の関数があれば，それらを積極的に使用し，作成手順を簡略化する。（付録の図12に Excel による相関係数の作成手順を記述しているプリントを掲載しておく。）

「実践」の段階：　演習問題は，作成手順書と例題を参考にすれば比較的容易に統計データを作成できるレベルのものである。

筆者はこの数年，上記の要領で「ビジネス統計学」を担当してきた。それに対する学生の授業評価の概要は以下の表1.1のとおりである。

表1.1　学生授業評価の概要

	受講人数	理解度	満足度
2012年度後期	17	3.6	3.7
2013年度後期	18	3.3	3.2
2014年度後期	25	3.3	3.3
2015年度後期	11	3.2	3.3
2016年度後期	10	3.1	3.3

表1.1の結果は「2016年度の満足度を例に採ると，大いに満足（５点）１人，まあまあ満足（４点）３人，どちらともいえない（３点）４人，少し不満（２点）２人，大いに不満（１点）０人で，それらを平均すると3.3点となる。」という方法で算出されたものである。

この結果から，理解度・満足度とも僅かではあるが年々低下傾向にあり，この傾向は理解度において顕著である。このままでは近い将来それらがいずれも3.0を下回ってしまうことが十分予測できる。

そこで，本稿ではその原因を考察とその対策を述べる。特に，大きな対策である「統計データの可視化」，具体的には統計データの図形として表現について述べる。さらに，Excel の機能で表現できない図形を作成するために作成したプログラムの仕様全般についても述べる。また，相関係数と相関の強さの関係について考察する。

2．学生授業評価の低下傾向の原因の考察と教材の可視化

学生授業評価の低下傾向の理由を2.1節で考察し，包括的な２つの対策を提案する。またそれらの具体的案について，１つは2.2節で述べる。もう１つは本稿の主要テーマであるが，記述すべき事項が多いため，改めて次章で述べることにする。

2.1　原因の考察

学生の状況を把握し評価するための最も客観的な方法は試験である。そこで「ビジネス統計学」で2016年度に実施した中間試験と期末試験の結果から授業評価の低下傾向の原因を分析することとする。なお，学生の授業評価は中間試験の前後の時期に実施されるので，中間試験の結果を重視する。

中間試験は**「模擬テスト－１」→「解説－１」→「本試験－１」**の順に段階的に実施した。「模擬テスト－１」「解説－１」「本試験－１」はそれぞれ以下のとおりである。

模擬テスト－１：　「本試験－１」の１週間前の授業の冒頭で配付する。（付録の図13参照。）授業では最初の40 分程度でこの問題を解答させた。

解説－１：　「模擬テスト－１」に引き続き，その正答を Excel ファイルにて配付し，解説を行った。正答はそれまでの授業の「基礎」「実用」の段階で説明した計算方法や作業手順書に則って，事前に作成しておいたものである。さらに，「本試験－１は模擬テスト－１と『同じような』問題であり，『同じような』というのは，データの個数と値のみ変えているだけで，それら以外はすべて模擬テストと同じであるという意味である。」と告知した。

本試験－１：　告知どおり，模擬テスト－１と「同じような」問題で模擬テスト－１の１週間後に実施した。

期末試験の実施も同様である。このように**「模擬テスト」→「解説」→「本試験」**という段階を踏んで試験を実施した理由は，受講する学生の多様性を鑑み，「何を勉強すればよいのか」を学生に対して明確に指示することであった。つまり，毎回の授業の「基礎」「実用」の枚葉を理解し，「実践」をきちんとしておくように，最悪でも模擬テストの解

法だけでも理解し実践できるようにしておくようにというメッセージを出すことであった。

ところが，2016年度の中間試験と期末試験の得点分布は以下の表2.1のとおりとなった。表の結果は「中間試験の０～９を例に採ると，中間試験の素点が０点～９点の学生は３人である。」ということを表している。さらに，特に成績の良くなかった中間試験の設問別の正答率は以下の表2.2のとおりである。（正答率の高い順にソートしている。）

表2.1　中間・期末試験の得点分布

得点範囲	中間人数	期末人数
０～９	3	0
10～19	2	0
20～29	1	1
30～39	3	0
40～49	2	2
50～59	1	0
60～69	0	1
70～79	0	3
80～89	0	5
90～100	4	4
平均点	43.81	75.75

これらの表から中間試験の時点では，受験した16人中12人もの学生が毎回の授業の「基礎」「実用」を理解しておらず，「実践」でやったことが身についていない，そのために殆どの設問に正答を出すことができないということが分かる。

また，設問（24）の相関係数については「値は－１～１の範囲になる」と何度も説明したにもかかわらず，20や82.1という答のまま提出した学生もいた。

期末試験では，学生自身が「これではまずい」と考えたのか，また筆者が模擬テスト時に「同じような」というのは，「データの個数と値のみ変えているだけで，それら以外はすべて模擬テストと同じである」ということをしつこく説明した甲斐があったのか，成績が全体的に向上したが，16人中３人が改善できなかった。

しかし，中間試験の結果と期末試験の結果の差から以下の考察が導かれる。

考察−１：「基礎」「実用」の内容は（必要に迫られて）きちんと取り組めばかなり理解できるが，下記の（A）～(C)の理由で，統計処理という作業に対してあまり親しみを感じず，学期の始めからきちんと取り組めなかった。

表2.2　中間試験の設問別正答率

設問	正答率	設問	正答率
(22)	81.25%	(21)	43.75%
(25)	75.00%	(37)	37.50%
(21)	62.50%	(11)	31.25%
(04)	56.25%	(12)	31.25%
(23)	50.00%	(13)	31.25%
(39)	50.00%	(14)	31.25%
(40)	50.00%	(15)	31.25%
(01)	43.75%	(20)	31.25%
(02)	43.75%	(07)	25.00%
(03)	43.75%	(08)	25.00%
(05)	43.75%	(10)	25.00%
(24)	43.75%	(19)	25.00%
(26)	43.75%	(34)	25.00%
(27)	43.75%	(36)	25.00%
(28)	43.75%	(06)	18.75%
(29)	43.75%	(09)	18.75%
(30)	43.75%	(16)	18.75%
(31)	43.75%	(17)	18.75%
(32)	43.75%	(18)	18.75%
(33)	43.75%	(38)	18.75%

（A）「基礎」の段階で使用している現在の教科書[4]は統計データの計算方法が見開きの左側のページ，その例題が同右側のページに記述されており，分かり易く構成されているが，文章と式が中心で，数学が本当に苦手な学生にとって視覚的に理解できない。

＜可視化されていない。＞

（B）「実用」の段階で作成手順書をプリントで配付しているが，これも文章中心である。その作成手順どおりに予め作成しておいたExcelシートも同時に配付しているが，作成手順全体の流れが見えにくい。

＜可視化されていない。＞

（C） 上記（A）（B）の作業で作成した統計データはヒストグラム，散布図，回帰直線を除けば，殆どのものが数値データまたはそれらが羅列された表であるので，おしなべて統計データのイメージが掴みにくい。

＜可視化されていない。＞

考察−2： 学生の授業評価は中間試験の前後の時期に実施されるので，学期の始めから学生が少しでも統計処理という作業に対して親しみを感じ，「基礎」「実用」の内容を理解することにきちんと取り組めていれば，学生の授業評価ももう少し高くなる。

上記2つの考察を整理し，1つにまとめると以下のようになる。

考察−3： 学生授業評価の低下傾向の大きな原因は，統計処理という作業にあまり親しみを感じず，「基礎」「実用」の内容を理解することが学期の始めからきちんと取り組めなかったことにある。そしてその原因は「教材が可視化されていない」「扱う統計データが可視化されていない」ことである。

以上の議論より，学生授業評価の低下傾向への対策として，下記の対策を提案する。

対策： 学生授業評価の低下傾向への対策は「教材の可視化」「統計データの可視化」である。そしてこれらを通して，学期の始めから学生が少しでも統計処理という作業に対して親しみを感じ，「基礎」「実用」の内容を理解することにきちんと取り組めるようにする。

以下では，「教材の可視化」の具体策について考察する。

2.2 教材の可視化

まず教材の可視化として以下の方策を立てる。

方策−1： 2016年度後期現在で使用している教材，つまり文章と式が中心の現在の教科書[4]と，同じく文章中心のExcel での処理手順のプリント配付を止める。

方策−2： その代わりに，図解入りの説明が多い『できるビジネスパーソンのための Excel 統計解析入門』[20]を新規の教科書として採用する。（付録の図14に新規の教科書[20]の Excel による相関係数の作成手順を掲載しておく。）

理由は以下の通りである。

理由−1： 現在の教科書[4]の計算方法を理解できなくても Excel の関数機能を使えば簡単に統計データを算出できるケースは結構多い。実際，付録の図11及び図12で示した相関係数もそうである。よって，「それぞれの統計データが何を意味するものなのか，何を表現しているものなのか」を言葉や文章で説明し理解させることができれば，本来の計算方法を敢えて説明する必要はない。新規の教科書[20]には「統計データが意味するもの，表現するもの」の説明が記述されている。

理由−2： 新規の教科書[20]では Excel での作成手順も図解入りで説明されており，可視化されていて分かり易い。

理由−3： 新規の教科書[20]の筋立てが本稿第1章で述べられている筆者の授業内容に比較的近い。

以上より，「教材の可視化」について下記の具体策を提案する。

具体策（教材の可視化）： 新規の教科書[20]を教材として採用することにより，教材の可視化を図ることとする。また，必要に応じて新規に補足資料を作成し，配付する。

2.3 教材の可視化による効果

前節で提案した具体策を2017年度後期の授業で実施した。その結果，早速以下のような効果が表れた。

効果−1： 学生の授業評価が，2016年度以前は1章の表1.1の状況であったのが，理解度4.0，満足度3.9となり，大幅に改善された[15]。

効果−2： また，2016年度の中間・期末試験の平均点が2.1節の表2.1の状況であったのが，2017年度は中間が77.16点，期末が83.00点となり[15]，理解度のより大幅な改善を裏付ける結果となった。

（詳細は拙著『統計データとその処理過程の可視化の試み』[15]を参照されたし。）

3．統計データの可視化

前章では，「教材の可視化」の具体策を提案し，2.3節のような結果が得られた。しかし，統計データの大部分は数値であり，それらは当然数字で表現されている。そして，その数字を見ただけで，その意味すること，表現していることを理解できる学生は殆どいない。

そこで本章では，数値として作成される「統計データの可視化」について下記の具体策を提案する。

具体策（統計データの可視化）： 以下の（A）〜（C）の要領で「統計データの可視化」を行う。

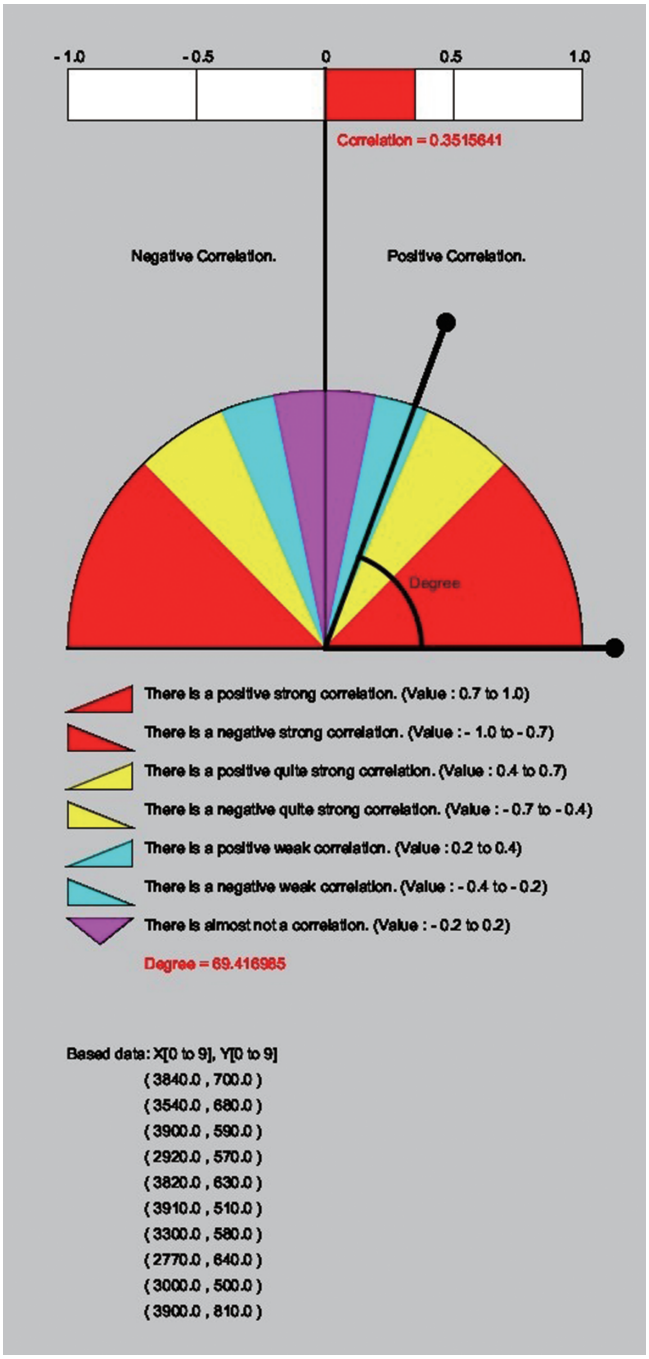


図１　相関係数を表現する図形

（A）　数値として作成される統計データとその意味すること，表現していることを図形で表現して可視化する。

（B）　現在の教科書[4]で説明していた数値の計算方法も，できるだけこれらの図形を使って説明する。

（C）　Excel で作成できない図形は processing というプログラミング言語（以降 processing と省略）により作成する。それらの描画プログラムについては起動方法やデータ入力方法等も解説する。

なお，作成すべき統計データはたくさんあるので，すべて図形で表現するには膨大な時間と紙面を必要とする。本稿では相関係数の図形表現についてのみ述べる。

3.1　相関係数

相関係数を表現する図形の例を上記の図１に示す。この図形は，現在の教科書[4]で新生児体重とその母親の胎盤重量のペアとして示されている表3.1の10件のデータに対して，以下の(1)(2)を表現している。

(1)　新生児体重とその母親の胎盤重量という２種類のデータの相関係数が約0.3516である。

(2)　その２種類のデータには「やや正の相関がある。」

表3.1　図１の基になる数値データのペア

No	新生児体重	母親の胎盤重量	No	新生児体重	母親の胎盤重量
0	3840	700	5	3910	510
1	3540	680	6	3300	580
2	3900	590	7	2770	640
3	2920	570	8	3000	500
4	3820	630	9	3900	810

長方形の全体は相関関係の値として取りうる範囲である－１～１を表現し白色で塗り潰す。実際に相関係数として計算された値を赤色で示し，長方形の中に上書きする。また，黒い長針が半円の左側にあるか右側にあるか，どの色の部分を通っているかにより，２種類のデータの相関が「どのような状態なのか」を示す。

このような図形を描画する Excel の機能はないので，描画プログラムを processing により作成した。本節では，以下の(3)(4)について記述する。

(3)　表現する図形を学生に説明するための**「図形の説明」**。

(4)　学生がプログラムを正しく使用するために必要な**「ユーザインタフェース」**。

3.2　相関係数の図形の説明

まず，受講すると予想される学生に対する図形と併記される数値の説明について述べる。彼らの特性を考慮し，数式や記号等をできるだけ排除する。入力データは１個以上存在するとする。図形の構成要素は下記の表3.2の「大項目」の３つであり，それぞれの要素はさらに「小項目」に細分化される。

まず，大項目の３要素を下記のように定義する。

表3.2　図形の概要

大項目	小項目	備考
相関係数	固定長方形	白く塗り潰す
	可変長方形の長さ	赤く塗り潰す
	可変長方形の位置	
相関の程度	半円	7段階に分けて色分けする
	長針と短針	計算された値が属す段階を示す
説明図	説明図	相関の程度を説明する

相関係数：　図1の例の新生児体重とその母親の胎盤重量のような，2種類のデータの関連性を示す値であり，最初に縦横の長さはともに一定の白色の長方形を描き，その中に1つだけ赤い長方形を描く。白色の長方形を**「固定長方形」**，赤色の長方形を**「可変長方形」**と呼ぶ。求められた相関係数の値によって描かれる長さと位置が変化するため，「可変」という表現を用いる。

相関の程度：　相関係数は2種類のデータの「なす角」を基に三角関数の cos を用いて－1～1の間の数値で表すことを意味している[4]ので，相関の程度を「半円」，**「長針と短針」**により表現する。

説明図：　これについては大項目と小項目が同じであるので，詳細は後述する。

次に，小項目の6要素を以下のように定義する。

固定長方形：　相関係数一般について考えられる最大幅を描く。左辺は相関係数としてあり得る最小値の－1を表し，右辺は同最大値の1を表す。中央の上から下への左辺と右辺にともに平行な黒い直線は，相関係数が正の値か負の値かを分ける0を示すラインである。

可変長方形の長さ：　**「可変長方形」**は計算された相関係数を描画する長方形である。従って，その値が0でなければ出現する。縦の長さは白い長方形と同じであるが，横の長さは「固定長方形の横の長さ×相関係数の絶対値÷2」となる。

可変長方形の位置：　相関係数が負の値の場合，右辺が中央線となる。相関係数が正の値の場合，左辺が中央線となる。（図8の例では左辺が中央線である。）

半円：　相関の程度は以下の表3.3と表3.4の組合せにより決定する。

表3.3は「相関があるかないか」を，表3.4は相関がある場合，その程度を示す。ここで，表3.3の「相関の正負」と「採りうる値」は2種類のデータから得られる散布図との関連で定義されており，現在の教科書[4]の記述を基にしたものである。また，「絶対値の値」とは相関係数として採りうる値の絶対値の範囲のことをいう。なお，「相関の強さ」と「絶対値の値」の関係も現在の教科書[4]を基に導き出したものである。（「強い相関がある」と「かなり相関がある」等の，相関の強さの境界値は『文献によってさまざまで，あいまいなところがあるが』[3),16),20)]本稿では，半円で表現したときの「強い相関がある」の部分の面積が最も小さく抑えられている現在の教科書[4]の境界値である－0.7，－0.4，－0.2，0.2，0.4，0.7を採用することにした。これらの境界値については，別のところで考察する。）

表3.3　半円の領域と相関の正負の関係

半円の領域	相関の正負	採りうる値
右側（紫色の右側）	正の相関がある	0.2～1
中程（紫色の部分）	ほとんど相関がない	－0.2～0.2
左側（紫色の左側）	負の相関がある	－1～－0.2

表3.4　半円の色と相関の強さの関係

色	相関の強さ	絶対値の値
赤	強い相関がある	0.7～1
黄	かなり相関がある	0.4～0.7
青	やや相関がある	0.2～0.4

長針と短針：　短針は半円の中心から x 軸上に固定する。長針と短針の「なす角」は，2種類のデータの「なす角」に等しいものになるように定める。従って，長針が表3.3の半円のどの領域を，表3.4のどの色の部分を通るかにより，相関の程度を表すこととなる。図1の例では，長針は紫色の右側を通っているので「正の相関がある」ということになり，さらに青い部分を通っているので「やや相関がある」ということになり，従って2種類のデータには「やや正の相関がある」ということになる。

説明図：　上記「相関の程度」のところで説明した半円の下に描く。表3.3と表3.4を1つにまとめて説明した図である。説明内容は下記の表3.5の通りである。この図がなければ，このプログラムのユーザはいちいちこの3.2節を読まないと相関の程度を判断できないであろう。

また，3.1節と同様の目的で計算結果領域と数値領域を以下のように定義する。

表3.5　説明図の表記内容

半円の領域	色	相関の強さ	絶対値の値
右側	赤	強い正の相関がある	0.7〜1
	黄	かなり正の相関がある	0.4〜0.7
	青	やや正の相関がある	0.2〜0.4
中程	紫	ほとんど相関がない	−0.2〜0.2
左側	青	やや負の相関がある	−0.4〜−0.2
	黄	かなり負の相関がある	−0.7〜−0.4
	赤	強い負の相関がある	−1〜−0.7

計算結果領域：　図形の上側に相関係数の値と２種類のデータの「なす角」を角度で書き込む。その角度は相関係数の値に基づいて算出される。図１の例では，相関係数の値は約0.3516，「なす角」は約69.4度である。

数値領域：　図形の下側に，相関係数の値の基になった２種類のデータのペアを，入力された順番に上から下へ書き込む。図８の例では図形の下側に（3840，700），（3540，680），（3900，590），（2920，570），（3820，630），（3910，510），（3300，580），（2770，640），（3000，500），（3900，810）が入力された順番に，上から下へ実数表現で書き込まれている。

3.3　ユーザインタフェース

本項では，学生がプログラムを正しく実行させるための操作方法について述べる。

まず，その操作方法の概略を以下の**「操作手順」**として記述する。

操作手順：

OP-1：　プログラムファイルの入ったフォルダを開く。

OP-2：　ファイル名「Book5」，ファイル形式「Excel の csv 形式」のデータファイルに入力するデータを記述する。記述方法の詳細は後述する。

OP-3：　必要に応じて，プログラムの一部を変更する。必要な変更の詳細は後述する。

OP-4：　実行ボタン（プログラムファイルの左上の三角形のボタン）をクリックしてプログラムを実行させる。

OP-5：　ファイル名「Correlation-001」，ファイル形式「jpg」の出力ファイルに図形が出力される。

OP-6：　必要に応じて，「ペイント」等のソフトを使って出力ファイルを加工する。

OP-7：　出力ファイル Correlation-001.jpg のサイズは，プログラムの２行目

size（500，1500）；....（3.2.1）

で定めているが，入力データの個数が多くなると，それに伴い下側の数値の記述部分が縦長になり，上記で定めたサイズでは描き切れなくなる。そのような場合，この２行目の縦の長さを定めている値1500をもっと大きな値に変更する。また，相関係数が１に極めて近い値になると，3.2節で定義した「長針」が x 軸に極めて近づき，出力ファイルの右側からはみ出す恐れがある。そのような場合，この２行目の横の長さを定めている値500をもっと大きな値に変更する。

上記操作手順の**「OP-2」**のデータファイルへの記述方法は以下のとおりである。

DF-1：　A1セルにダミーの文字列，例えば「Data」と書き込み，A2セルからA列の n＋1行目のセルまで縦に，B2セルからB列の n＋1行目のセルまで縦にそれぞれ n 個の入力データを記述する。

DF-2：　実数値が含まれる場合，すべてのデータが整数になるまで『すべてのデータを10倍にし，倍率を1/10にする』を繰り返す。具体例を下記の**例題3.1**に示す。

例題3.1：　元のデータを

110，15.25，20.125，100.7，11

（５件）とすると，データファイルには

110000，15250，20125，100700，11000

と記述し，倍率を0.001とする。

次に，上記操作手順の**「OP-3」**について述べる。上記操作手順の**「OP-2」**で，元のデータに実数値が含まれている場合，上記操作「DF-2」で入力データの変更とその倍率調整を行ったが，それに伴い，プログラムの一部も次のような変更を行う必要がある。

プログラムの10行目が

float SCL＝1.0；....（3.2.2）

となっているが，これは「現時点での倍率は1.0倍である」ことを意味しているので，この10行目の倍率1.0を上記操作手順の**「DF-2」**で変更した値に書き直す。上記**例題3.1.1**については，

float SCL＝0.001；....（3.2.3）

と書き直す。

3.4　相関係数と相関の強さの関係に関する考察

今回「統計データの可視化」の一環として，相関係数と相関の強さを現在の教科書[4]を基にして図１のように図形化した。特に，相関の強さは教科書[4]を基に表3.3及び表3.4のように定義し，それを図１の下部の半円とその７つの分割で表現している。ここで筆者はその７つの分割について，赤い部分，つまり「強い正または負の相関がある」となるエリアが黄色（かなり正または負の相関がある），青色（やや正または負の相関がある），紫色（ほとんど相関がない）

のエリアに比べて広すぎるように見える。そこで筆者は，3.2節で相関の強さの境界値は『文献によってさまざまで，あいまいなところがあるが』[3),16),20)]と記されている状況をもう少し調べ，その上で境界値について新たな提案を試みることにした。

まず，筆者は現在の教科書[4)]以外にいくつかの書籍[2),6),10−11),13),17−18),20−21)]とWebサイト[1),3),7),12),14−16),19)]の相関係数に関する部分を調査し，表3.3及び表3.4の相関の強さの段階の区分やその境界値の設定がどのように記述されているのか調査した。調査結果の概要は以下の⑴〜⑶のとおりである。（調査の詳細は付録に記述している。）

⑴　書籍に関しては，相関の強さの段階の区分，その境界値の設定の両方ともさまざまで，あいまいなところがある。
⑵　Webに関しては，特に近年は現在の教科書[4)]と同様の区分と境界値を用いているところが多いが，境界値に関しては異なる値を用いているところもある。
⑶　1956年にGuilfordが提唱して以来，現在の教科書[4)]と同様の区分と境界値を用いることが多いが，ほとんどの場合，その根拠や原典が示されていない。[3),16)]つまり，慣例的に用いられているだけである。

調査の結果，区分については現在の教科書[4)]の7段階に固まりつつあるが，それらの境界値については依然として『文献によってさまざまで，あいまいなところがある』という状況と考えてよいであろう。

そこで筆者は相関の強さの区分とその境界値として以下の提案を行うことにする。提案の基本概念は，現在の教科書[4)]を基にした表3.3及び表3.4を「強い正または負の相関がある」，「かなり正または負の相関がある」，「やや正または負の相関がある」，「ほとんど相関がない」の各エリアの広さを原則として等しくするように変更しようというものである。

提案−1：　「強い相関がある」，「かなり相関がある」，「やや相関がある」，「ほとんど相関がない」のそれぞれに正・負に分けた8段階に区分し，それぞれを原則22.5度ずつに半円を分割する。具体的には以下の表3.6のように設定する。『相関の正負はあまり問題ではなく，その強さを重視する。』という考え方を採る。

提案−2：　「強い相関がある」，「かなり相関がある」，「やや相関がある」のそれぞれに正・負に分けた6段階に「ほとんど相関がない」を加えた合計7段階に区分し，それぞれを180/7＝約25.7度ずつに半円を分割する。具体的には以下の表3.7のように設定する。『ほとんど相関がない場合は相関の正負は問題にならないが，それ以外の場合は正負も重視する。』という考え方を採る。

表3.6　相関の強さの区分とその範囲（提案−1）

相関の強さ	範　囲
強い正の相関がある	$-1 \leqq r < -0.92$
かなり正の相関がある	$-0.92 \leqq r < -0.71$
やや正の相関がある	$-0.71 \leqq r < -0.38$
ほとんど相関がない	$-0.38 \leqq r \leqq 0.38$
やや負の相関がある	$0.38 < r \leqq 0.71$
かなり負の相関がある	$0.71 < r \leqq 0.92$
強い負の相関がある	$0.92 < r \leqq 1$

境界値はそれぞれ±0.38，±0.71，±0.92となる。

表3.7　相関の強さの区分とその範囲（提案−2）

相関の強さ	範　囲
強い正の相関がある	$-1 \leqq r < -0.9$
かなり正の相関がある	$-0.9 \leqq r < -0.65$
やや正の相関がある	$-0.65 \leqq r < -0.25$
ほとんど相関がない	$-0.25 \leqq r \leqq 0.25$
やや負の相関がある	$0.25 < r \leqq 0.65$
かなり負の相関がある	$0.65 < r \leqq 0.9$
強い負の相関がある	$0.9 < r \leqq 1$

境界値はそれぞれ±0.25，±0.65，±0.9となる。

上記2つの提案に基づいて，表3.2のデータで相関係数を表現した図形がそれぞれ以下の図2及び図3である。それぞれの「考え方」において各色の面積が原則として平等であることがわかる。ただし，同じデータでも「提案−1」では「ほとんど相関がない」となり，「提案−2」では従来どおり「やや正の相関がある」となり，強さの評価が分かれてしまうところが興味深い。

最後に，本項では新たに2つの境界値を提案したが，境界値はあくまでも「目安」であり，取り扱う分野によって異なる値が設定されても構わないであろうと考える。

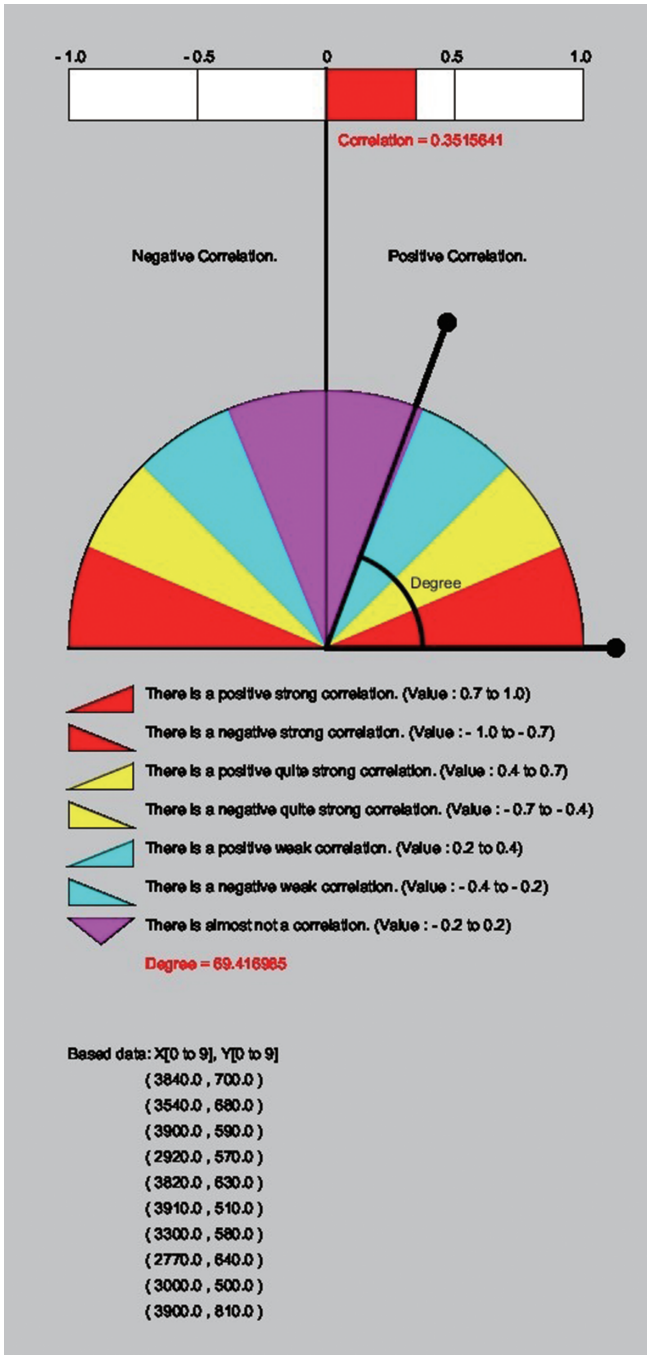


図２　「提案－１」に基づいた相関係数を表現する図形

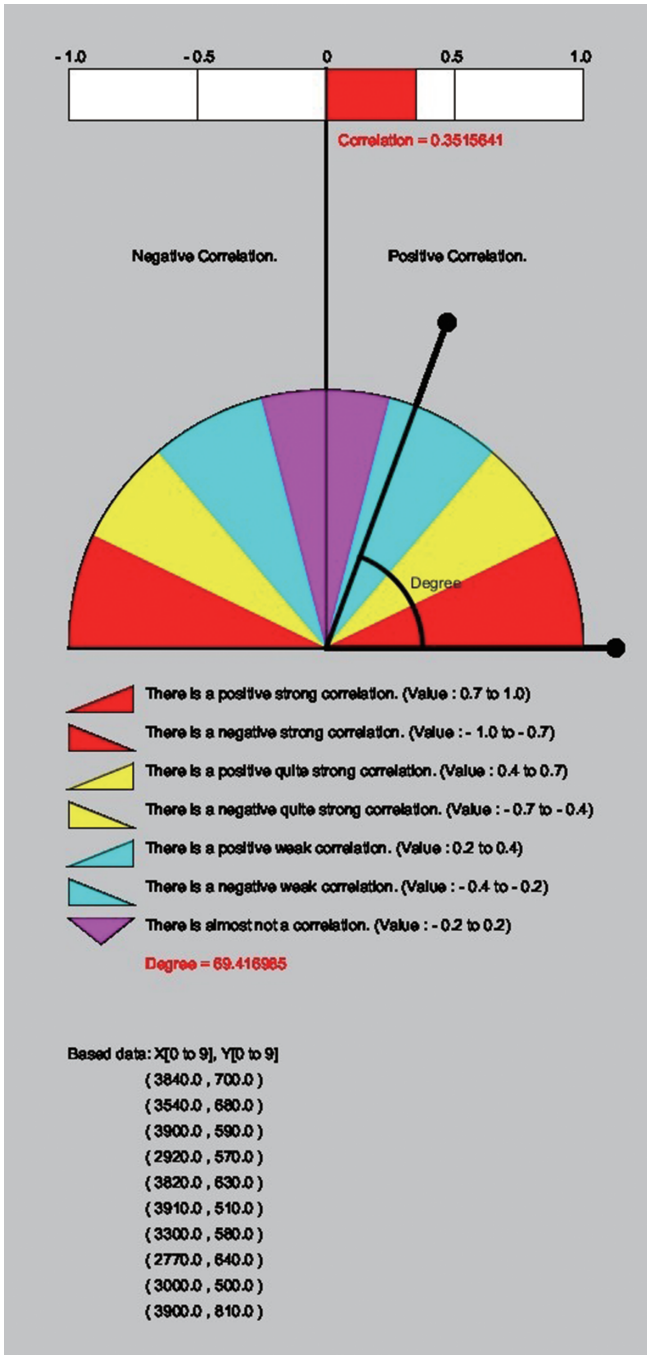


図３　「提案－２」に基づいた相関係数を表現する図形

4．結論

本稿では，筆者が福岡工業大学短期大学部ビジネス情報学科で担当している「ビジネス統計学」という科目の学生の授業評価が僅かずつではあるが低下傾向を示している原因を，2.1節で述べたように，以下の２つの対策を立てた。

教材の可視化：　2.2節で述べたように，新規の教科書[20]を教材として採用することにより，教材の可視化を図ることとする。また，必要に応じて新規に補足資料を作成し，配付する。

統計データの可視化：　第３章で述べたように，数値で表現されていることの多い統計データ（統計処理により作成された図表や算出された数値のこと）を図形により可視化する。本稿では差し当たり，相関係数の図形表現を試みた。これらについてはExcelによる作図機能がないため，processingというプログラミング言語により作図描画プログラムを作成し，そのプログラムを操作するための使用マニュアルも作成した。

教材の可視化については2017年度から早速実施し，学生による授業評価を大幅に改善することができた。統計データの可視化については，2018年度後期の授業で図形描画プログラムを試行予定である。

また，相関係数と相関の正負や強さ（強い相関，弱い相関，無相関など）の関係を現在の教科書[4]を基に図形で表現したとき，下記の（(1)(2)の区分を表すエリアが(3)〜(7)の区分を表すエリアより広く広く見えた。

(1)　強い正の相関がある。
(2)　強い負の相関がある。
(3)　かなり正の相関がある。
(4)　かなり負の相関がある。
(5)　やや正の相関がある。
(6)　やや負の相関がある。
(7)　ほとんど相関がない。

そこで3.3節で述べたように，相関係数と相関の正負の強さについて記述されている文献を調査した結果，区分については上記(1)〜(7)の７段階に固まりつつあるが，それらの境界値については依然として『文献によってさまざまで，あいまいなところがある』という状況と考えてよいということになった。

そこで筆者は相関の強さの区分とその境界値として3.3節の表3.6及び3.7の提案を行った。提案の基本概念は，上記(1)〜(7)の各エリアの広さを原則として等しくするように変更しようというものである。

今後の課題としては，補助教材を充実させ，もっとたくさんの統計データの可視化を図り，数学や数字に強くない人々に統計処理・統計学に対する理解を深め，親しみを持ってもらえるようにすることである。

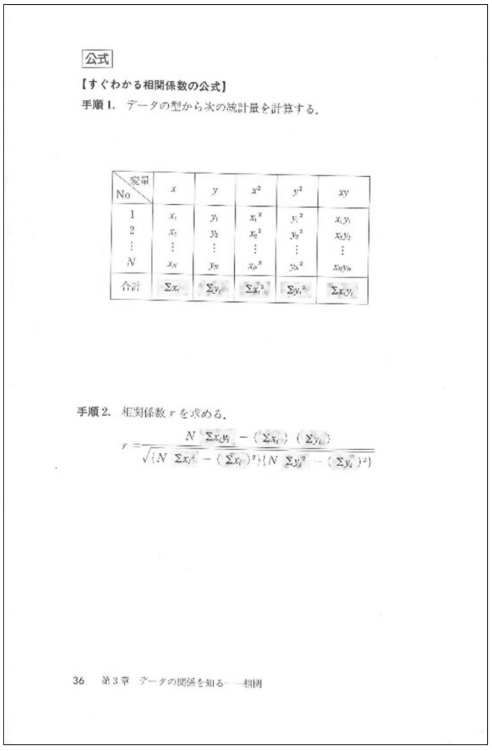
参考文献

1）有馬昌宏：『第13章　相関』, https://www.ai.u-hyogo.ac.jp/~arima/lectures/JT-13.pdf（2017年８月29日現在）
2）浅野晃：社会人１年生のための統計学教科書，SB Creative，2014年（初版）
3）井口豊：『統計学の基準値の由来』，https://note.chiebukuro.yahoo.co.jp/detail/n190275（2017年８月29日現在）
4）石村貞夫：すぐわかる統計解析，東京図書，2010年（第34刷）
5）泉恵理子他（編）：日経ビジネスアソシエ「仕事の数字に強くなる！」，日経BPムック，2014年
6）片谷教孝，松藤敏彦：環境統計学入門，オーム社，2003年（初版第１刷）
7）金久保正明：『相関係数』，http://www.sist.ac.jp/~kanakubo/research/statistic/soukankeisuu.html（2017年８月29日現在）
8）木村捨雄：『今，なぜ統計教育が必要なのか？』，http://www.naruto-u.ac.jp/kyozai/toukei/ts/main_4_k.html（2017年６月29日現在）
9）J.P.Guilford: Fundamental statistics in psychology and education, McGraw Hill, 1956 (New York)
10）向後千春，富永敦子：統計学がわかる―回帰分析・因子分析編―，技術評論社，2008年（初版）
11）小寺平治：新統計学入門，裳華房，2004年（第14版）
12）志堂寺和則：『自動車感性評価学』，http://cog.inf.kyushu-u.ac.jp/~shidoji/japanese/statistics/07Correlation.pdf（2017年８月29日現在）
13）関口やす夫，篠原靖忠，小森尚志：基礎数理統計，共立出版，1980年（初版第11刷）
14）高木方隆：『回帰分析と相関係数』，http://www.infra.kochi-tech.ac.jp/takagi/Survey2/9Regression.pdf（2017年８月29日現在）
15）高橋昌也：『統計データとその処理過程の可視化の試み』，福岡工業大学 FD Annual Report Vol.8, pp.63-71, 2017.
16）タナカタロウ（匿名）：『相関係数の大きさに対する目安の歴史的変遷』，https://drive.google.com/file/d/0B3B5DERtTZI2V1ZMa1lPVGhXOTQ/view（2017年８月29日現在）
17）所一夫：数理統計概要，槙書店，1978年（第14刷）
18）縄田和満：Excelによる統計入門，朝倉書店，2007年（初版第１刷）
19）林田智弘：『相関分析』，http://hil.hiroshima-u.ac.jp/sys2/c/sobun.pdf（2017年８月29日現在）
20）日花弘子：できるビジネスパーソンのためのExcel統

計解析入門，SB Creative，2016年（初版第１刷）
21）牧野都治：統計の知識，森北出版，1994年（初版第16刷）

付録

見開き左側のページ



公式

【すぐわかる相関係数の公式】

手順 1. データの型から次の統計量を計算する。

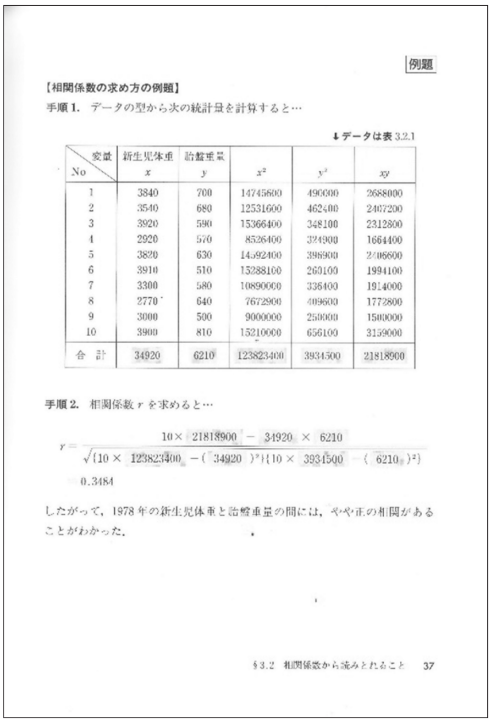
No＼変量	x	y	x^2	y^2	xy
1	x_1	y_1	x_1^2	y_1^2	x_1y_1
2	x_2	y_2	x_2^2	y_2^2	x_2y_2
⋮	⋮	⋮	⋮	⋮	⋮
N	x_N	y_N	x_N^2	y_N^2	x_Ny_N
合計	Σx_i	Σy_i	Σx_i^2	Σy_i^2	Σx_iy_i

手順 2. 相関係数 r を求める。

$$r=\frac{N\Sigma x_iy_i-(\Sigma x_i)(\Sigma y_i)}{\sqrt{\{N\Sigma x_i^2-(\Sigma x_i)^2\}\{N\Sigma y_i^2-(\Sigma y_i)^2\}}}$$

36　第3章　データの関係を知る——相関

見開き右側のページ



例題

【相関係数の求め方の例題】

手順 1. データの型から次の統計量を計算すると…

↓データは表 3.2.1

No＼変量	新生児体重 x	胎盤重量 y	x^2	y^2	xy
1	3840	700	14745600	490000	2688000
2	3540	680	12531600	462400	2407200
3	3920	590	15366400	348100	2312800
4	2920	570	8526400	324900	1664400
5	3820	630	14592400	396900	2406600
6	3910	510	15288100	260100	1994100
7	3300	580	10890000	336400	1914000
8	2770	640	7672900	409600	1772800
9	3000	500	9000000	250000	1500000
10	3900	810	15210000	656100	3159000
合　計	34920	6210	123823400	3934500	21818900

手順 2. 相関係数 r を求めると…

$$r=\frac{10\times 21818900-34920\times 6210}{\sqrt{\{10\times 123823400-(34920)^2\}\{10\times 3934500-(6210)^2\}}}$$

$$=0.3484$$

したがって，1978年の新生児体重と胎盤重量の間には，やや正の相関があることがわかった。

§3.2　相関係数から読みとれること　37

図11　教科書[4]の相関係数の作成手順

Excel による相関係数の求め方

以下のデータ(A 型データ)はある村の農家 25 世帯の田畑の耕作面積(単位はアール)である。このデータから相関係数を Excel により求める手順を示す。

1. まず、B2 セル～D27 セルに以下のような表を作成する。

No	田	畑
1	21	40
2	25	39
3	24	41
4	23	45
5	24	44
6	31	39
7	27	44
8	22	50
9	32	41
10	29	46
11	36	41
12	24	57
13	30	52
14	35	48
15	37	46
16	41	43
17	27	37
18	58	50
19	42	48
20	31	59
21	26	67
22	38	56
23	41	60
24	36	68
25	48	71

2. C29 セルに「 相関係数 =」と入力する。
3. D29 セルに式「 =correl(C3:C27,D3:D27) 」を入力し Enter キーを押す。

〈相関係数の計算〉

図12　配付プリントの相関係数の作成手順

2016年11月22日3限目実施

ビジネス統計学　模擬テスト[1]

<持込：すべて可>

＜ 問 題 用 紙 ＞

以下の表や計算式の空欄を埋めよ。解答は解答用紙の対応する番号欄に記入すること。また、小数点以下が3桁以上になる場合は、3桁目を四捨五入して、小数点以下を2桁にすること。(例：2.61512… → 2.62)

表の作成や計算は **Excel** を使って行ってください。

x、y の2種類のデータをとったところ、以下のような結果を得た。

No	1	2	3	4	5	6	7	8	9	10
x	1	2	3	4	6	7	9	10	11	12
y	8	13	13	4	1	14	10	6	9	5
No	11	12	13	14	15	16	17	18	19	20
x	15	18	20	22	23	26	28	29	31	35
y	11	2	6	4	13	7	1	14	11	12

手順1：　y についての度数分布表を作ると以下のようになった。

階級	階級値	度数	相対度数 (%)	累積度数	累積相対度数 (%)
0 ～ 3	1.5	(01)	(06)	(11)	(16)
3 ～ 6	4.5	(02)	(07)	(12)	(17)
6 ～ 9	7.5	(03)	(08)	(13)	(18)
9 ～12	10.5	(04)	(09)	(14)	(19)
12～15	13.5	(05)	(10)	(15)	(20)

手順2：　x の平均値は (21) 、x の中央値は (22) であり、
y の平均値は (23) 、y の中央値は (24) であり、
y の最頻値は (25) である。

1

2016年11月22日3限目実施

ビジネス統計学　模擬テスト[1]

<持込：すべて可>

手順3：　x の分散は (26) 、標準偏差は (27) 、
標本分散は (28) 、標本標準偏差は (29) であり、
y の分散は (30) 、標準偏差は (31) 、
標本分散は (32) 、標本標準偏差は (33) である。

手順4：　相関係数は (34) である。

手順5：　最後に、有意水準を0.05として無相関の検定を行う。
相関係数は (34) であり、x と y のペアの個数は (35) であるので、
検定統計量は式

(34) × ((35) − 2)^1/2 / (1 − (34)^2)^1/2

で計算すると (36) となる。また、自由度は式

(35) − 2

で計算すると (37) となり、この値と有意水準0.05から t 分布の値は (38) となる。検定統計量の値と t 分布の値を比較すると
(36) (39) (38) （ (39) には等号または不等号が入る。）
となるので、

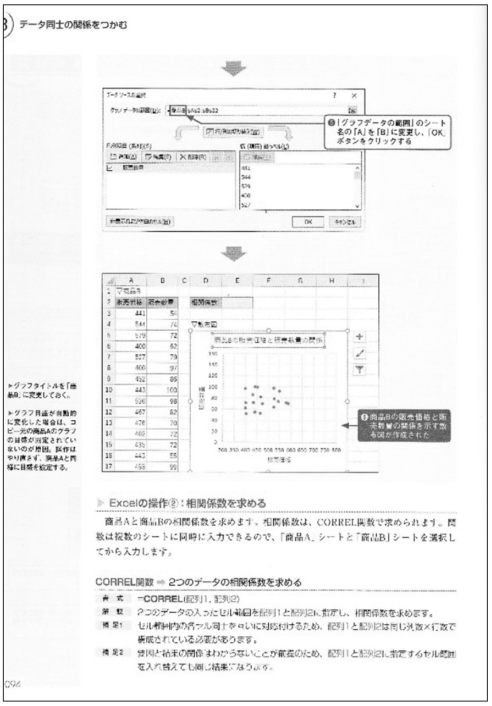
「相関が (40) 。」

という結論が得られる。

学籍番号＿＿＿＿＿＿　**氏名**＿＿＿＿＿＿＿＿＿＿

2

図13　2016年度模擬テスト問題



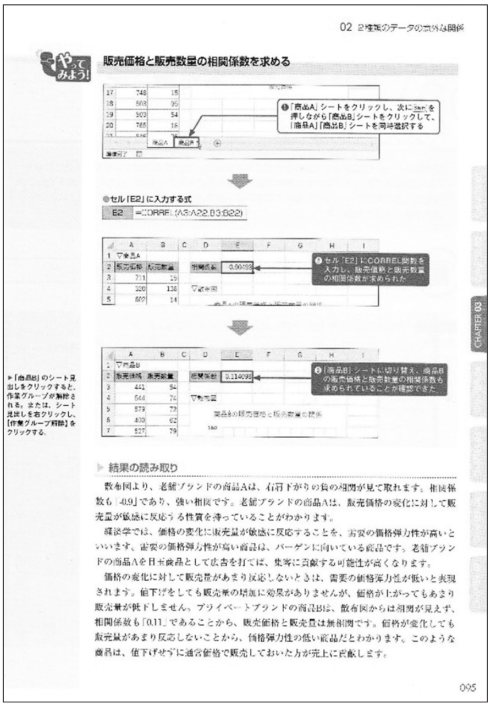


図14　新規採用教科書[20]の相関係数計算の解説

相関の強さの区分とその範囲に関する調査結果

まず，書籍についての調査結果は以下の１～５のとおりである。

１．記述がないものが２冊。[6),13)]ただし，２冊とも相関係数の計算方法については記述あり。

２．表3.5.2及び表3.5.3と概ね同じものが１冊。[10)]

３．３段階に区分しているものが１冊。[21)]詳細は下記の表A.1のとおりである。

表 A.1　相関の強さの区分とその範囲(1)

相関の強さ	範囲
正の相関	$0<r\leqq1$
無相関	$r=0$
負の相関	$-1\leqq r<0$

４．５段階に区分しているものが３冊。[11),17),18)]詳細は下記の表A.2～A.4のとおりである。

表 A.2　相関の強さの区分とその範囲(2)[11)]

相関の強さ	範囲
強い正の相関	$r\fallingdotseq1$
正の相関	$0<r<1$
無相関	$r\fallingdotseq0$
負の相関	$-1<r<0$
強い負の相関	$r\fallingdotseq-1$

表 A.3　相関の強さの区分とその範囲(3)[17)]

相関の強さ	範囲
完全相関	$r=1$
順相関	$0<r<1$
無相関	$r=0$
逆相関	$-1<r<0$
完全逆相関	$r=-1$

表 A.4　相関の強さの区分とその範囲(4)[18)]

相関の強さ	範囲
完全相関	$r=1$
正の相関	$0<r<1$
無相関	$r=0$
負の相関	$-1<r<0$
負の完全相関	$r=-1$

５．７段階に区分しているが範囲が表3.5.2及び表3,5,3と異なるものが２冊。[2),20)]詳細は下記の表A.5及び表A.6のとおりである。

表 A.5　相関の強さの区分とその範囲(5)[2)]

相関の強さ	範　囲	
	正の相関	負の相関
強い相関	$r\fallingdotseq1$	$r\fallingdotseq-1$
中くらいの強さの相関	r：0.7付近	r：−0.7付近
ほとんど相関がない	r：0～0.5付近	r：−0.7付近～0
無相関	$r\fallingdotseq0$	

※実際には，「0.5は中くらいの強さの相関ではなく，ほとんど相関がないことを示し，0.7くらいのとき，中くらいの強さの相関になります。」という表現である。

表 A.6　相関の強さの区分とその範囲(6)[20)]

相関の強さ	範　囲	
	正の相関	負の相関
完全相関	$r=1$	$r=-1$
強い相関	$0.5\leqq r<1$	$-1<r\leqq0.5$
弱い相関	$0.2\leqq r<0.5$	$-0.5<r\leqq-0.2$
無相関	$-0.2<r<0.2$	

次にWebサイトについての調査結果は以下の６及び７のとおりである。

６．表3.5.2及び表3.5.3と概ね同じものが４件。[1),7),12),14)]

７．７段階に区分しているが範囲が表3.5.2及び表3,5,3と異なるものが１件。[19)]詳細は下記の表A.7のとおりである。ただし，「これらはあくまでも目安であって，データ数や扱っている対象によって変化する。」という注釈が付いている。

表 A.7　相関の強さの区分とその範囲(7)[19)]

相関の強さ	範　囲	
	正の相関	負の相関
強い相関あり	$0.8\leqq r\leqq1$	$-1\leqq r\leqq-0.8$
相関あり	$0.6\leqq r<0.8$	$-0.8<r\leqq-0.5$
弱い相関あり	$0.4\leqq r<0.6$	$-0.6<r\leqq-0.4$
ほとんど相関なし	$-0.4<r<0.4$	