ホームページ：　分かりやすくまとめた会社実務のための基礎知識と事例
URL　http://www.bisquerhino12.sakura.ne.jp/

主成分分析の手順

2021年　ver. 1.1

倉谷　隆博

９－１．主成分分析とは

主成分分析（PCA: Principal Component Analysis）は、個体に係る複数の変数を対象にそれぞれの変数が持ってる情報をできるだけ失うことなく、変数の数を減らして、新たな変数に組み直すことを可能にする方法である。新たに作った変数を主成分という。また、変数の数を減らすことを「縮約」という。新しい変数は制約条件のもとで元の変数を線形結合して作る。

主成分分析のメリット

①データの持っている特徴を、新たな切り口で理解することができる。

例えば、4教科（変数）の試験のデータがあったとき、総合点を算出して比較することがよく行われる。総合点は試験データを評価する際に新たな情報の切り口を提供する。ここで、総合点を算出するということは、4つある変数を総合点という1つの変数に置き換えることを意味している。4次元のデータを1次元のデータにすることで、次元数を削減している。

しかし、総合点だけで試験結果を評価すると、個々の教科の試験データの情報を無視することになる。総合点は新しい変数ではあるが、元の変数が持っている多くの情報を失っている。

主成分分析は、元の変数が持っている情報をできるだけ失わないで、縮約して新しい変数を作る方法である。

4教科の試験のデータの主成分分析では、分析結果の解釈によるが、例えば理系の力と文系の力というような、縮約した新たな2つの変数でデータを説明することを可能にする。

②データの構造が見えやすくなる。

主成分分析では、元の変数が持っている情報をできるだけ失わないようにして、そのうえ、互いに相関関係がないようにして、新たな変数を作り出す。これによって、データの構造を簡素化して見えやすくしている。例えば、30種類もある変数のデータを前にして、相互関係などのデータ構造を理解するのは極めて難しいと思われる。しかしこれが、例えば5種類の新たな変数で説明できるとなると、データ構造の理解は容易になる。3次元以下にまで縮約できれば、データの構造を図などで可視化して表すことができるようになる。

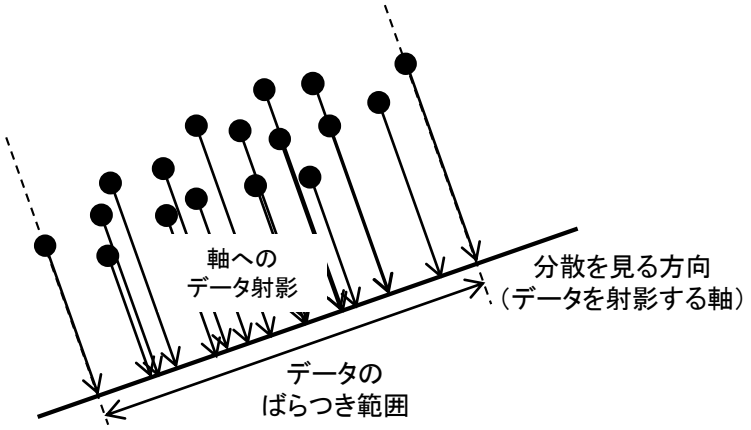
また、新たな変数の間には相関関係がないため、ある変数のデータを見るとき他の変数のデータの影響を全く考慮する必要がない。これはデータの構造を理解する上で大きな力を発揮する。

９－２．主成分分析の進め方

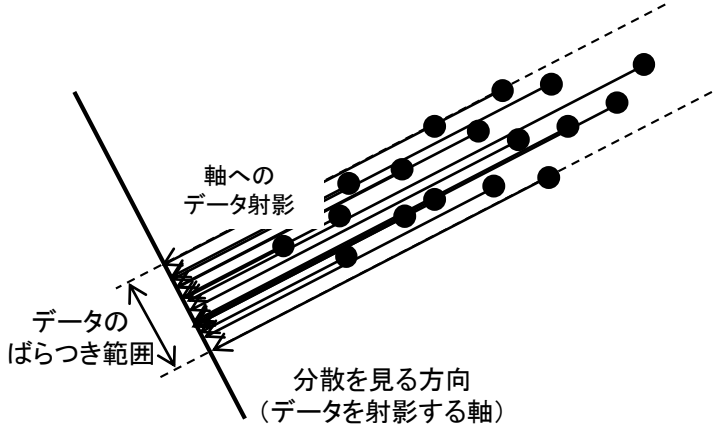
主成分分析は、元の変数の線形結合によって新しい変数を縮約して作る方法である。
新しい変数を作るにあたっては、前提条件がある（下記）。

①新しい変数（主成分）には、元の変数の情報をできるだけ持たせる。
情報量の大きさはデータのばらつき具合（分散）の大きさで測ることができる。
よって、分散が大きくなるように新しい変数を作る。
（注１）：例えば、試験が難し過ぎて試験結果が全員０点であった場合、全員が同じ点数であったということは、点数にばらつきがなかった、点数の分散＝０だったということになるが、このような場合、この試験では、個々人の能力が測定できなかったことになる。何の情報も得られなかったことになる。
（注２）：（図１）と（図２）のデータの分布は同じであるが、データを見る方向（データを射影する軸）を変えるとデータのばらつき具合が異なる。（図１）の軸での分散は大きく、（図２）の軸での分散は小さい。
軸の取り方次第で分散の大きさが変わる。主成分分析ではこのような軸のなかから元の情報をできるだけ失わないように、情報量を表す分散が最大になる軸を探し出す。この軸が主成分になる。最初に作る新しい軸を第１主成分という。

②新しい変数（主成分）が持つ情報には重複がないようにする。
情報に重複がないとは相関がないということを意味する。相関がないように、そして分散が最大になるように新しい変数軸（第２主成分以下）を１つずつ増やしていく。重複のない情報が積み重なっていく。



（図１）データのばらつき範囲が大きい方向で見る場合（イメージ）



（図２）データのばらつき範囲が小さい方向で見る場合（イメージ）

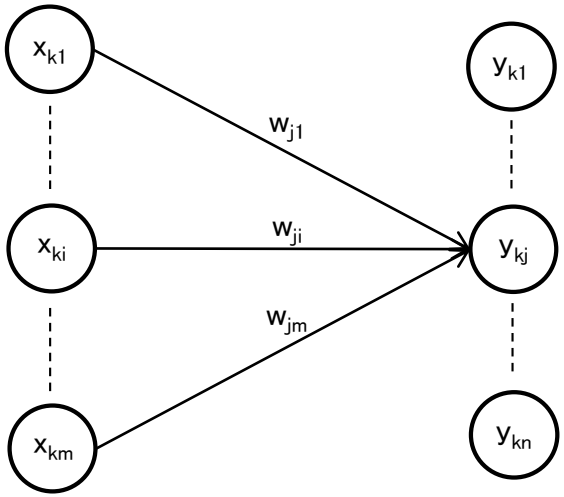
９－３．主成分分析の手順（1/2）

元の変数（軸）を使って、新しい変数（軸）を式（a）の線形結合の式で作る。

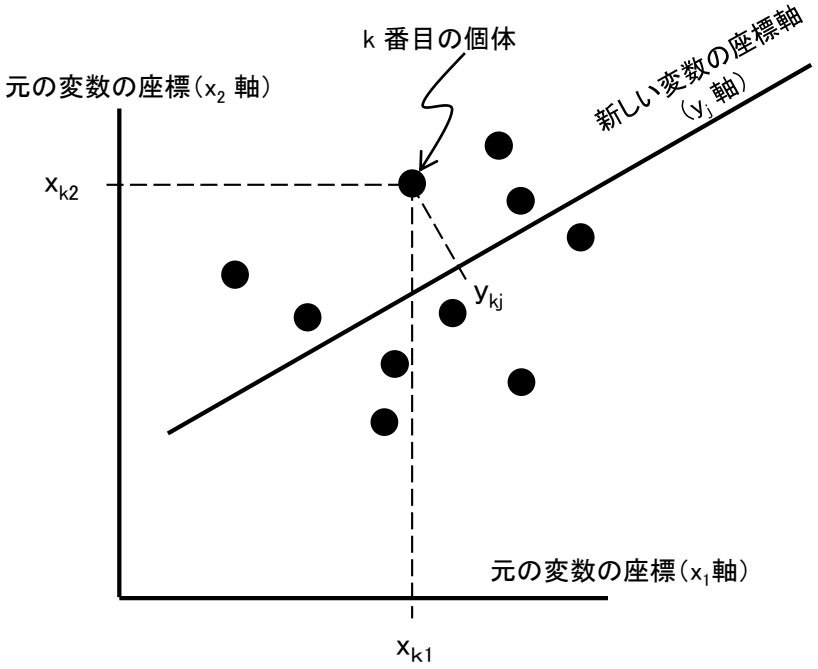
$$y_{kj} = \sum_{i=1}^{m} w_{ji} x_{ki} = w_{j1}x_{k1} + w_{j2}x_{k2} + \cdots + w_{jm}x_{km} \quad ---\ (a)$$

k：個体の数（1, …, p）、 i： 元の変数（軸）の数（1, …, m）、 j： 新しい変数（軸）の数（1, …, n）
x_{ki}：k 番目の個体の i 番目の元の変数（軸）
y_{kj}：k 番目の個体の j 番目の新しい変数（軸）
w_{ji}：k 番目の個体の m 個の元の変数（軸） x_{ki} から j 番目の新しい変数（軸） y_{kj} を作成するとき使う、i 番目の元の変数（軸）の係数

（図3）に式（a）を模式化して示す。m 個の元の変数（軸）から n 個の新しい変数（軸）を作る。新しい変数（変数）の数は元の変数（軸）の数より少なくする。つまり、m > n である。変数（軸）の数（次元）を少なくすることを、次元の削減、あるいは縮約という。
また、（図4）に、式（a）によって元の変数（軸）の座標から新しい変数（軸）の座標に移すときの様子を示す。 図では元の2つの変数（軸） X_{k1} と x_{k2} から、新しい1つの変数（軸） y_{kj} を作っている。なお、式（a）は、本来 x_{ki} の平均値を通る式であるが、ここでは簡略化している。平均値を考慮した場合、（図4）の新しい変数の座標軸はこの平均値を通る座標軸として表される。



（図3）線形結合



（図4）座標変換

９－３．主成分分析の手順（2/2）

ステップ１　第１主成分（軸）の求め方

下記の条件(b)の下で、個体 k の新しい変数（軸）である y_{k1} は、分散 $V(y_{k1})$ が最大となるような係数 w_{1i} と個体 k の元の変数（軸）の i 番目の x_{ki} との線形結合の式（a） で求める。このようにして求めた y_{k1} （軸）は第１主成分（軸）と呼ばれる。
条件(b)がないと、y_{k1} の分散はいくらでも大きくすることができるので、条件(b)を設けておく。軸の大きさが問題なのではなく、データを射影する軸の方向が問題なので、大きさを制約する条件（b）を設けることは問題ない。

$$\sum_{i=1}^{m} w_{ji}^2 = w_{j1}^2 + w_{j2}^2 + \cdots + w_{jn}^2 = 1 \quad ---\ (b)$$

よって、条件（b）を組み込んだ式（c）の最大化問題を解くことになる。

$$Q = V(y_{k1}) \Big/ \sum_{i=1}^{m} w_{1i}^2 \quad ---\ (c)$$

この問題は係数 w_{1i} の連立方程式（固有方程式という）を解く問題になり、解くにあたってはラグランジェの未定乗数法という方法が用いられる。この方程式の解である係数 w_{1i} は固有ベクトルと呼ばれ、また、分散 $V(y_{k1})$はこの方程式の固有値で表される。

ステップ２　第２主成分（軸）の求め方

条件（b）に加えて、

共分散 $Cov(y_{k1}, y_{k2}) = 0$ ---（c）

つまり、第１主成分（軸） y_{k1} と第２主成分（軸） y_{k2} の間には相関がないという条件で、分散 $V(y_{k2})$が最大となるような係数 w_{2i} を求め、この係数を使って式（a）で y_{k2} （軸）を計算する。

ステップ３　第３主成分（軸）の求め方

条件（b）に加えて、

共分散 $Cov(y_{k1}, y_{k3}) = 0$ ---（d1）
共分散 $Cov(y_{k2}, y_{k3}) = 0$ ---（d2）

つまり、第３主成分 y_{k3} （軸）は、第１主成分 y_{k1} （軸）とも第２主成分 y_{k2} （軸）とも相関がないという条件で、分散 $V(y_{k3})$が最大となるような係数を求め、この係数を使って式（a）で y_{k3} （軸）を計算する。

なお、以降も、主成分の数が元の変数の数に到達するまでこの計算を繰り返す。
そのなかで、第１主成分から始まり、主成分の分散（情報量）の累積が全体の分散（情報量）に対して７０～８０％になるまでの主成分を新しい変数として採用する。主成分の数は元の変数の数より少なくなり、縮約される。

９－４．主成分分析の例題(1/4)

ここでは、例題として、１０店舗を、商品品質軸、接客態度軸など6軸で、１０点満点で評価したデータを使い、主成分分析を行う。
主成分分析の目的は、店舗の評価をここに挙げた6軸とは異なる、より少ない数の軸で評価できるようにすることである。そして、少ない数の軸がどのような意味を持つ軸になるのかを確認すことである。
（表1）が元データになる。（表2）は（表1）のデータを平均 = 0、標準偏差 = 1 になるように標準化したものである。（表1）の6軸の評価点は全て同じ単位であり、この場合標準化は必ずしも必要ではないが、一般化して分析を進めるためここではデータの標準化を行った。

（表1）元データ

店舗名	商品品質軸	接客態度軸	接客速度軸	清潔さ軸	保全軸	雰囲気軸
A店舗	9.0	9.5	9.1	8.0	9.0	9.2
B店舗	9.5	8.0	8.5	8.0	7.5	8.2
C店舗	8.7	9.1	8.2	7.5	6.5	7.8
D店舗	9.7	9.4	9.0	8.0	6.0	8.3
E店舗	9.2	6.5	8.2	7.8	5.4	7.6
F店舗	8.8	6.5	7.5	6.4	7.2	6.8
G店舗	7.5	8.0	8.0	7.9	8.0	7.0
H店舗	8.8	9.2	9.1	8.5	8.2	8.3
I店舗	8.0	7.5	6.4	7.2	7.3	5.7
J店舗	8.4	6.5	7.1	6.7	7.0	6.3
平均	8.76	8.02	8.11	7.60	7.21	7.52
標準偏差	0.67	1.24	0.90	0.65	1.06	1.06
分散	0.44	1.54	0.81	0.43	1.13	1.13

（表2）標準化データ

標準化	商品品質軸	接客態度軸	接客速度軸	清潔さ軸	保全軸	雰囲気軸
A店舗	①0.38023	1.25731	②1.16278	0.64550	1.77332	1.66541
B店舗	1.17239	-0.01699	0.45806	0.64550	0.28730	0.67409 ⑦
C店舗	-0.09506	0.91750	0.10571	-0.16137	-0.70338	0.27757
D店舗	1.48925	1.17236	1.04532	0.64550	-1.19873	0.77323
E店舗	0.69710	-1.29129	0.10571	0.32275	-1.79313	0.07931
F店舗	0.06337	-1.29129	-0.71646	-1.93649	-0.00991	-0.71375
G店舗	-1.99623	-0.01699	-0.12920	0.48412	0.78264	-0.51548
H店舗	0.06337	1.00245	1.16278	1.45237	0.98078	0.77323
I店舗	-1.20408	-0.44176	-2.00843	-0.64550	0.08916	-1.80419
J店舗	-0.57035	-1.29129	-1.18627	-1.45237	-0.20804	-1.20940
平均	0.00	0.00	0.00	0.00	0.00	0.00
標準偏差	1.00	1.00	1.00	1.00	1.00	1.00
分散	1.00	1.00	1.00	1.00	1.00	1.00

制約条件付きの最大化問題は、連立方程式（固有方程式）を解く問題になるが、固有方程式には分散共分散行列（標準化されている場合は、相関係数からなる相関行列）が埋め込まれている。（表3）は、（表2）から求めた相関行列である。

例えば、（表3）の商品品質軸と接客態度軸との相関係数は、（表2）の①の商品品質軸のデータと（表2）の②の接客態度軸のデータから求まり、（表3）の③に示すように 0.25007 となる。

（表3）相関行列

相関係数	商品品質軸	接客態度軸	接客速度軸	清潔さ軸	保全軸	雰囲気軸
商品品質軸	1	0.25007	0.59806	0.28123	-0.34467	0.64989
接客態度軸	③0.25007	1	0.71522	0.71014	0.37014	0.72055
接客速度軸	0.59806	0.71522	1	0.77710	0.17907	0.96266
清潔さ軸	0.28123	0.71014	0.77710	1	0.21583	0.71668
保全軸	-0.34467	0.37014	0.17907	0.21583	1	0.23354
雰囲気軸	0.64989	0.72055	0.96266	0.71668	0.23354	1

9－4. 主成分分析の例題(2/4)

（表4）以下は主成分分析の結果である。主成分分析には翔泳社のExcel のビジネス統計ソフト（アドイン）を使用した。

（表4）固有ベクトル（係数）

固有ﾍﾞｸﾄﾙ	第1主成分軸	第2主成分軸	第3主成分軸	第4主成分軸	第5主成分軸	第6主成分軸
商品品質軸	④ 0.30493	-0.61111	⑤ -0.42436	0.08840	0.56881	0.14875
接客態度軸	0.43481	0.24762	0.23956	0.82759	0.07028	0.04882
接客速度軸	0.50326	-0.07526	-0.06486	-0.21554	-0.53238	0.63795
清潔さ軸	0.44013	0.12530	0.60399	-0.47714	0.44036	-0.06474
保全軸	0.13425	0.73377	-0.57032	-0.15577	0.28655	0.10917
雰囲気軸	0.50294	-0.07377	-0.26125	-0.09427	-0.33476	-0.74324

（表5）主成分得点

主成分得点	第1主成分軸	第2主成分軸	第3主成分軸	第4主成分軸	第5主成分軸	第6主成分軸
A店舗	2.60759	1.25069	-0.99214	0.08231	⑨ -0.07951	-0.22625
B店舗	1.24233	-0.51318	⑥ -0.48138	-0.42545	0.56272	-0.04565
C店舗	0.39730	-0.27949	0.48445	0.88853	-0.41140	-0.17455
D店舗	2.00200	-1.55422	0.45260	0.68241	0.05490	0.19828
E店舗	-0.35449	-2.03487	0.58485	-0.91199	-0.14876	-0.16752
F店舗	-2.11532	-0.50181	-1.26728	0.08418	-0.28994	0.14409
G店舗	-0.62223	1.89840	0.83216	-0.46699	-0.45786	0.05703
H店舗	2.20017	0.96659	0.25370	-0.33406	0.14923	0.23852
I店舗	-2.74954	0.89523	0.56603	0.42505	0.69858	-0.08948
J店舗	-2.60781	-0.12734	-0.43298	-0.02399	-0.07795	0.06552

（表6）情報量（分散）

情報量	第1主成分軸	第2主成分軸	第3主成分軸	第4主成分軸	第5主成分軸	第6主成分軸
固有値	3.66975	1.40697	0.48684	⑧ 0.27927	0.13274	0.02443
寄与率	0.61163	0.23450	0.08114	0.04654	0.02212	0.00407
累積寄与率	0.61163	0.84612	0.92726	0.97380	0.99593	1.00000

（表7）主成分負荷量

主成分負荷量	第1主成分軸	第2主成分軸	第3主成分軸	第4主成分軸	第5主成分軸	第6主成分軸
商品品質軸	0.58414	-0.72487	-0.29609	0.04671	0.20724	0.02325
接客態度軸	0.83295	0.29372	0.16715	⑩ 0.43735	0.02561	0.00763
接客速度軸	0.96408	-0.08927	-0.04526	-0.11390	-0.19397	0.09972
清潔さ軸	0.84314	0.14862	0.42143	-0.25215	0.16044	-0.01012
保全軸	0.25717	0.87037	-0.39793	-0.08232	0.10440	0.01706
雰囲気軸	0.96346	-0.08750	-0.18229	-0.04982	-0.12197	-0.11617

（表4）は、主成分を求める際に使用する、元データに対応する係数（固有ベクトル）である。（表5）の主成分得点を計算するのに使用する。
なお、例えば第3主成分軸⑤の係数の2乗の総和は制約条件なので、当然1である。 $(-0.42436)^2 + (0.23956)^2 + \cdots + (-0.26125)^2 = 1$
また。主成分軸同志は直交するように係数を求めているので（無相関）、第1主成分軸④と第3主成分軸⑤を例にあげれば、下表に示すように積和は当然0 になる。$-0.12940 + 0.10416 + \cdots - 0.13139 = 0$

標準化	商品品質軸	接客態度軸	接客速度軸	清潔さ軸	保全軸	雰囲気軸
(表4)の④ 第1主成分軸	0.30493	0.43481	0.50326	0.44013	0.13425	0.50294
(表4)の⑤ 第3主成分軸	-0.42436	0.23956	-0.06486	0.60399	-0.57032	-0.26125
④×⑤	-0.12940	0.10416	-0.03264	0.26583	-0.07657	-0.13139

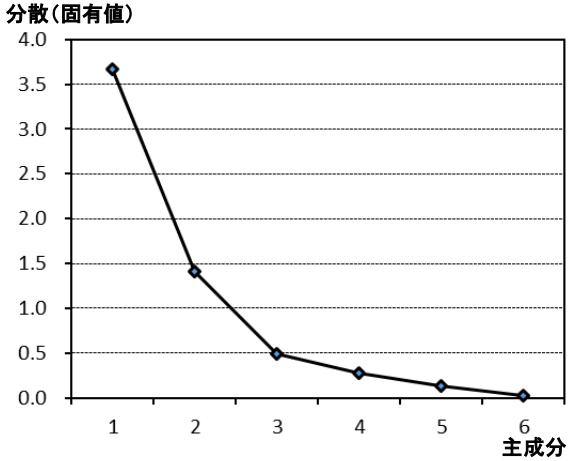
（表5）は、新しい変数（主成分軸）での値である。主成分得点と呼ぶ。例えば、B店舗の第3主成分軸での主成分得点（表5）の⑥は、（表4）の⑤の係数と、対応する（表2）の⑦の元のデータの積和として求める。
$-0.49752 - 0.00407 + \cdots - 0.17611 = -0.48138$
なお、元の変数（表2）の分散の合計 $1 + 1 + \cdots + 1 = 6$ は、新しい変数（表5）の分散の合計つまり、（表6）の固有値の合計と同じになる。
$3.66975 + 1.40197 + \cdots + 0.02443 = 6$ である。

標準化	商品品質軸	接客態度軸	接客速度軸	清潔さ軸	保全軸	雰囲気軸
(表4)の⑤ 第3主成分軸	-0.42436	0.23956	-0.06486	0.60399	-0.57032	-0.26125
(表2)の⑦ B店舗	1.17239	-0.01699	0.45806	0.64550	0.28730	0.67409
⑤×⑦	-0.49752	-0.00407	-0.02971	0.38988	-0.16385	-0.17611

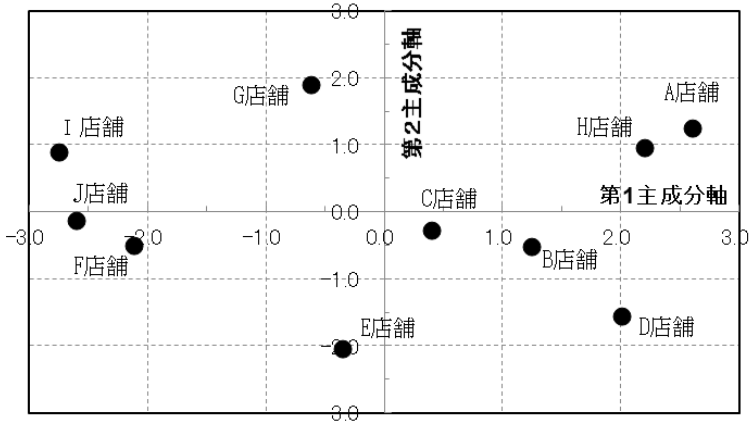
（表6）は主成分ごとの情報量（分散）を示す。分散は固有値で表される。例えば、（表6）の第4主成分軸の固有値⑧は、（表5）の第5主成分軸⑨の分散である。
（表6）の寄与率は、全ての主成分軸の情報量に対する特定の主成分軸の情報量が占める割合である。例えば、第1主成分軸の寄与率は（表6）の第1主成分軸の固有値 3.66975 を全ての主成分軸の固有値の総和 6 で割って求める。なお、例えば、第1主成分軸と第2主成分軸の累積寄与率は、それぞれの寄与率の和である。$0.61163 + 0.23450 = 0.84612$
（表7）は、元の変数と新しい変数との相関係数である。元の変数が新しい変数にどの程度の影響を及ぼしているかを大雑把に把握するために使用する。例えば（表7）の⑩は、（表2）の接客態度軸②と（表5）の第4主成分軸⑨との相関係数である。

９－４．主成分分析の例題(3/4)

（図5）は、固有値を降順で並べたスクリープロット（scree plot，scree とは、山腹の崩れ岩の塊のこと）である。
（表6）から、第１主成分軸の分散の値 3.66975 と第２主成分軸の分散の値 1.40697 は、元の変数の分散の平均値 １ より大きく、平均以上の情報を持っていることが分かっているが、（図5）はそれを視覚的に分かりやすく表したものである。
寄与率は（表6）より、第１主成分軸が６１．１%、第２主成分軸が２３．５%であるので、この２つの軸での累積寄与率は８４．６%になる。つまり、この２つの主成分軸で全体の情報量の８４．６%を説明できていると言える。元の変数軸の数は６つであったが、さほど情報を失うことなく、新しい２つの変数軸に縮約できたことになる。



（図5）スクリ―プロット



（図6）主成分得点の散布図

（図6）は、新しく作成した第１主成分軸と第２主成分軸で店舗をプロットしたものである。元の変数の数は６つであったのでデータを可視化することはできないが、新しい変数は２つであるのでこの散布図のようにデータを可視化できる。この散布図も加味して主成分分析の結果から、新しい２つの変数（軸）は何を表しているのかを考察する。
まず、第１主成分に関しては、（表４）で第１主成分の係数の値を見ると全てプラスであることが分かる。これは元の変数のデータのいずれかが大きくなれば第１主成分の値が大きくなることを示している。このことから、第１主成分はいわば店舗の総合力のようなものを表しているといえる。（図６）から、第１主成分軸の右側にプロットされているA店舗、H店舗、D店舗は総合力が高い店舗、一方、I店舗、J店舗、F店舗は総合力に課題がある店舗だといえる。これは元の変数のデータ（表１）を見てもうなづける。
次いで、第２主成分に関しては、（表４）で第２主成分の係数を見ると保全軸と接客態度軸の係数がプラスの大きな値になっており、一方、商品品質軸の係数はマイナスの大きな値になっている。店舗の保全（修繕）がきちんとしており、かつ接客態度が気持ちいい、いわば印象のいい店舗でありながら商品品質の達成に課題があることを示している軸だと言える。第２主成分軸は商品品質課題度を表していると考えられる。これは、元の変数のデータ（表１）を見ても概ねうなずける。
なお、これらの傾向は、元の変数と新しい変数の相関係数を示す（表７）主成分負荷量を見ても分かる。

9－4. 主成分分析の例題(4/4)

(図7)に、元の変数軸のときのデータ構造と新しい変数(主成分)軸のときのデータ構造を対比しておく。6次元の軸を2次元の軸に縮約している。つまり、もとの変数軸の持つ情報量(分散)をできるだけ維持するために、寄与率の大きな第1主成分軸と第2主成軸の2軸を選んでいる。2軸で、84．7%(61．2＋23．5)の情報量を維持できている。

■ 元の変数軸

変数軸	分散	累積寄与率(%)
商品品質軸	1.0	16.7
接客態度軸	1.0	16.7
接客速度軸	1.0	16.7
清潔さ軸	1.0	16.7
保全軸	1.0	16.7
雰囲気軸	1.0	16.7
合計	6.0	100

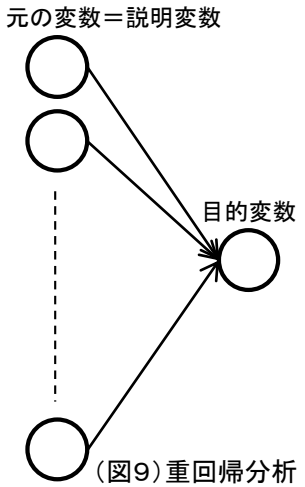
■ 新しい変数軸

変数軸	変数軸を名付け	分散	累積寄与率(%)
第1主成分軸	総合力軸	3.66975	61.2
第2主成分軸	商品品質達成度軸	1.40697	23.5
第3主成分軸	―	0.48684	8.1
第4主成分軸	―	0.27927	4.7
第5主成分軸	―	0.13274	2.2
第6主成分軸	―	0.02443	0.4
合計		6.0	100

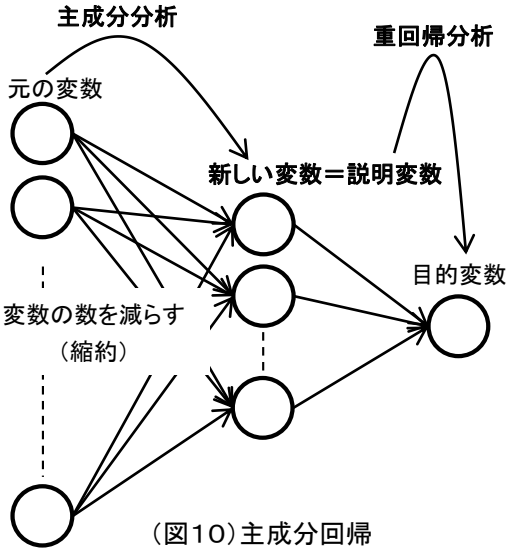
(図7)主成分分析による変数の数の削減(縮約)の様子

9－5. 主成分回帰の概要

このように、主成分分析は、新しい変数を合成することで変数の数を少なく(縮約)する手法であると言える。
この縮約という手法は重回帰分析に使用される。重回帰分析(図9)では説明変数の数が増えると分析が複雑になり、さらには分析結果の解釈が難しくなる。
そこで、まず主成分分析によって説明変数の数を減らす縮約を実施し、その後、重回帰分析を行う、つまり主成分回帰(図10)を行うといった方法である。主成分回帰では新しい変数の間、つまり回帰分析で使用する説明変数の間に相関はないので、重回帰分析を行うとき問題となる多重共線性を全く気にする必要がないという利点がある。



(図9)重回帰分析



(図10)主成分回帰

付録 1. 座標変換の手順　1/3

主成分分析では、(図4)に示すような座標変換を行う。ここでは、例を使って座標変換の手順を示す。

(付表1－1)と(付図1－1)に示す座標軸 x_1 と座標軸 x_2 でプロットされる 10 個のデータを新しい座標軸 y_1 へ移す。この座標変換では、新しい座標軸 y_1 として、10 個のデータの分散を最大にする座標軸を求める。新しい座標軸はその方向をどう設定するかが問題であり、その長さはどうでもいい。そこで座標軸の長さを1に設定して新しい座標軸を求める。ここで、座標軸 y_1 が主成分分析の第1主成分軸になる。

手順1)新しい座標の原点を、元のデータ(付表1－1)の平均値に設定する。(付図1－1)の赤色の点が平均値になる。

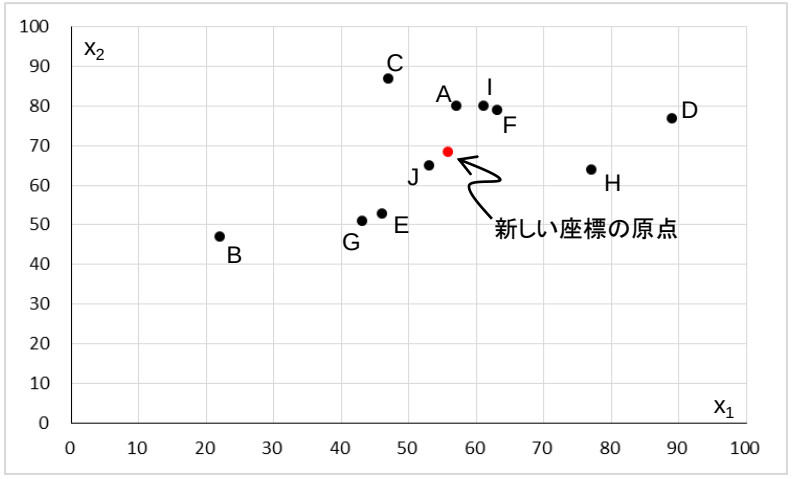
x_1 の平均値 = 55.8,　x_2 の平均値 = 68.3

手順2)新しい座標系の原点を平均値に設定し、平均値をデータの中心に据える。

つまり、データと平均値の差を計算し(付表1－2)、このデータを使って座標変換を行う。なお、(付表1－2)のデータを元の座標軸にプロットしたものを(付図1－2)に示す。

(付表1－1)準備したデータ

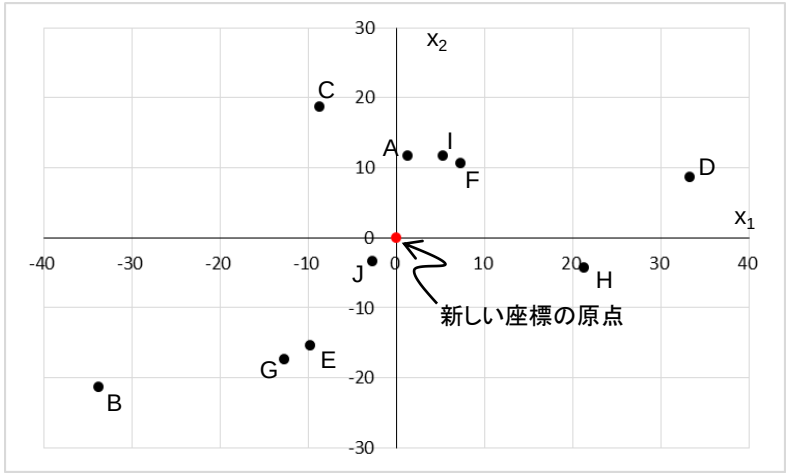
	A	B	C	D	E	F	G	H	I	J
x_1	57.0	22.0	47.0	89.0	46.0	63.0	43.0	77.0	61.0	53.0
x_2	80.0	47.0	87.0	77.0	53.0	79.0	51.0	64.0	80.0	65.0



(付図1－1)　(付表1－1)のデータのプロット図

(付表1－2)新しい座標の原点を中心に置いたデータのプロット

	A	B	C	D	E	F	G	H	I	J
x_1	1.2	-33.8	-8.8	33.2	-9.8	7.2	-12.8	21.2	5.2	-2.8
x_2	11.7	-21.3	18.7	8.7	-15.3	10.7	-17.3	-4.3	11.7	-3.3



(付図1－2)　(付表1－2)のデータのプロット図

付録 1. 座標変換の手順　2/3

手順3）式(a)に示す座標軸変換に使う係数のベクトル w は、下記の固有方程式を解いて求めることができる。

$$\mathbf{S}\mathbf{w} = \lambda \mathbf{w} \quad \cdots \text{（付式1）}$$

この固有方程式は、制約条件付きの最適化問題を解くときに使われるラグランジュ関数の偏微分 = 0 から導かれる式である。
この式に使われている行列 S は、(付表1－2)に示すデータの分散共分散行列である。

$$\mathbf{S} = \begin{pmatrix} S_{x1x1} & S_{x1x2} \\ S_{x2x1} & S_{x2x2} \end{pmatrix} \quad \cdots \text{(付式2)}$$

x_1 の分散　：　$S_{x1x1} = \{ (1.2^2 + (-33.8)^2 + \cdots + (-2.8)^2 \} / 10 = 312.0$　　・・・・（付式３）　10 はデータ数

x_2 の分散　：　$S_{x2x2} = \{ (11.7^2 + (-21.3)^2 + \cdots + (-3.3)^2 \} / 10 = 183.0$　　・・・・　（付式4）

x_1 と x_2 との共分散　：　$S_{x1x2} = S_{x2x1} = \{ (1.2 \times 11.7 + (-33.8) \times (-21.3) + \cdots + (-2.8) \times (-3.3) \} / 10 = 128.6$

・・・・　（付式5）

よって、分散共分散行列は下記のようになる。

$$\mathbf{S} = \begin{pmatrix} 312.0 & 128.6 \\ 128.6 & 183.0 \end{pmatrix} \quad \cdots \text{(付式6)}$$

(付式1)の ベクトル w は 0 ではないので、次の行列式(付式7)が求まり、(付式7)に(付式6)を代入する。

$$|\mathbf{S} - \lambda \mathbf{I}| = 0 \quad \cdots \text{(付式7)} \qquad \text{ただし、} \mathbf{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{、単位行列}$$

$$\begin{vmatrix} 312.0 - \lambda & 128.6 \\ 128.6 & 183.0 - \lambda \end{vmatrix} = 0 \quad \cdots \text{(付式8)}$$

(付式8)から、$(312 - \lambda)(183 - \lambda) - 128.6^2 = \lambda 2 - 495\lambda + 40558 = 0$　・・・(付式9)

(付式9)を解いて、λ (固有値)を求めると、　$\lambda = \{495 \pm (495^2 - 4 \times 40558)^{0.5}\} / 2 \rightarrow \lambda_1 = 391.4 \quad \lambda_2 = 103.6$

まず、(付式1)に、$\lambda_1 = 391.4$ の値を代入して、λ_1 の固有ベクトル $\mathbf{w}_1$ を求める。

$$\begin{pmatrix} 312.0 & 128.6 \\ 128.6 & 183.0 \end{pmatrix} \begin{pmatrix} w_{11} \\ w_{12} \end{pmatrix} = 391.4 \begin{pmatrix} w_{11} \\ w_{12} \end{pmatrix} \quad \cdots \text{(付式10)}$$

(付式10)より、$312\, w_{11} + 128.6\, w_{12} = 391.4\, w_{11}$　・・・(付式11)

$128.6\, w_{11} + 183\, w_{12} = 391.4\, w_{12}$　・・・(付式12)

(付式11)(付式12)より、ベクトルの長さを 1 の単位ベクトル(固有ベクトル)を求めると、$\mathbf{w}_1 = \begin{pmatrix} 0.53 \\ 0.85 \end{pmatrix}$　・・・(付式13)

同様に、λ_2 の固有ベクトルを求めると、$\mathbf{w}_2 = \begin{pmatrix} 0.85 \\ -0.53 \end{pmatrix}$　・・・(付式14)

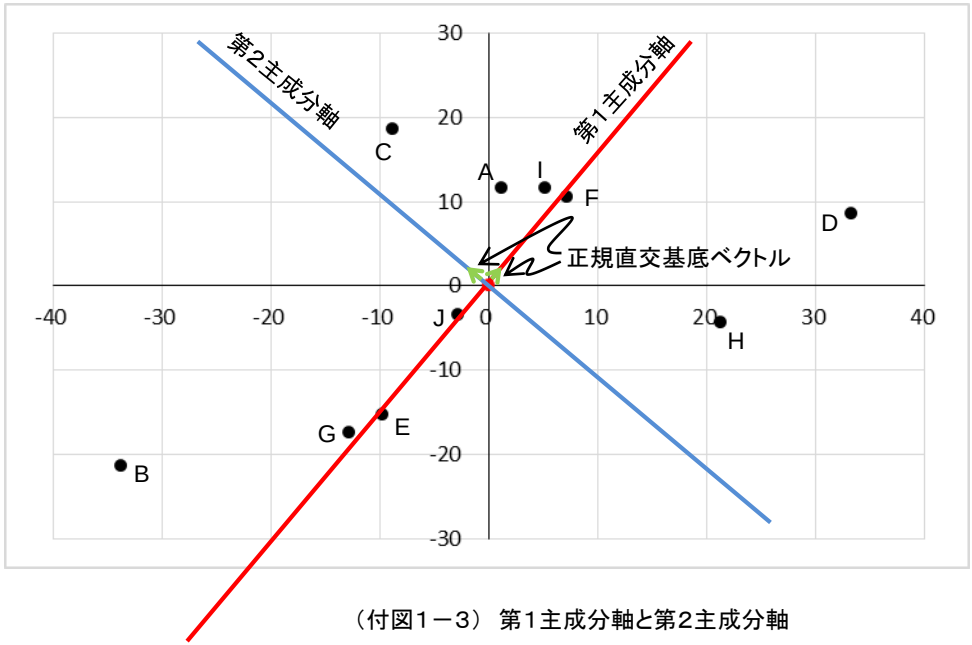
付録 1. 座標変換の手順　3/3

手順4）求めた固有ベクトルを、正規直交基底ベクトルに設定する。
固有値が大きい順に、それに対応する固有ベクトルを第1主成分軸、第2主成分j軸とする。

$\lambda_1 = 391.4 > \lambda_2 = 103.6$ なので、
第1主成分軸は $\mathbf{w}_1$（付式13）、第2主成分軸 は $\mathbf{w}_2$（付式14）になる。

よって、座標変換の式は式(a)を使って、
第1主成分軸に関しては、$y_1 = 0.53\ x_1 + 0.85\ x_2$　・・・（付式15）
第2主成分軸に関しては、$y_2 = 0.85\ x_1 - 0.53\ x_2$　・・・（付式16）

手順5）このようにして求めた新たな座標軸である第1主成分軸と第2主成分軸を（付図1－3）に示す。
第1主成分軸と第2主成分軸は、正規直交基底ベクトルをもとに作成しているので、互いに直交している。



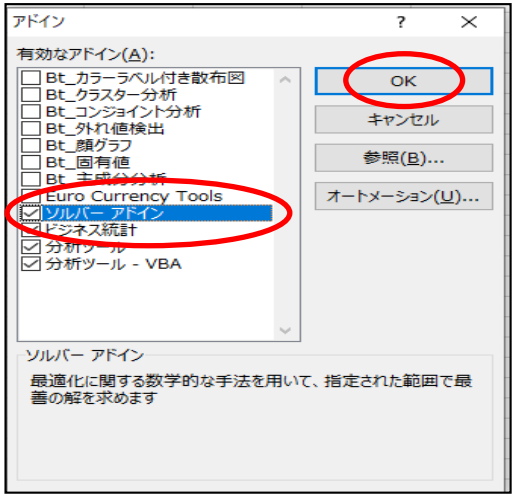
（付図1－3）　第1主成分軸と第2主成分軸

付録 2. Excel を使った主成分分析の方法　1/3

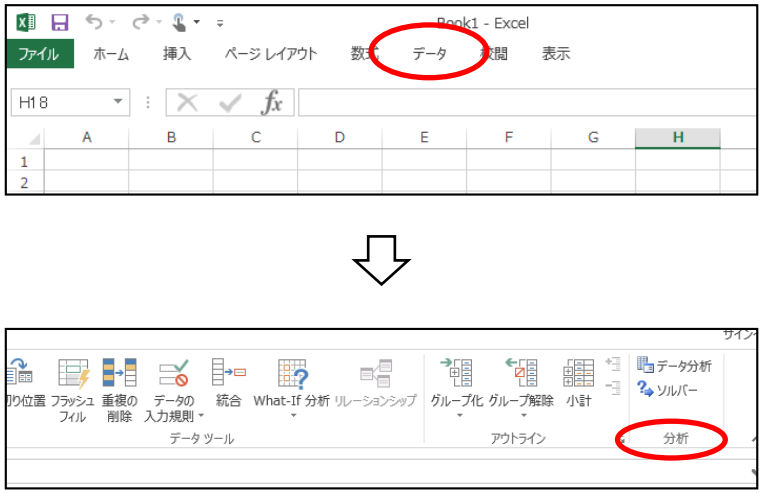
1）統計ソフトのアドイン（機能追加）

①「ファイル」をクリック
②「オプション」をクリック
③「アドイン」をクリック
④「設定」をクリック

⑤「ソルバー」にチェックを入れて、「OK」をクリックする。

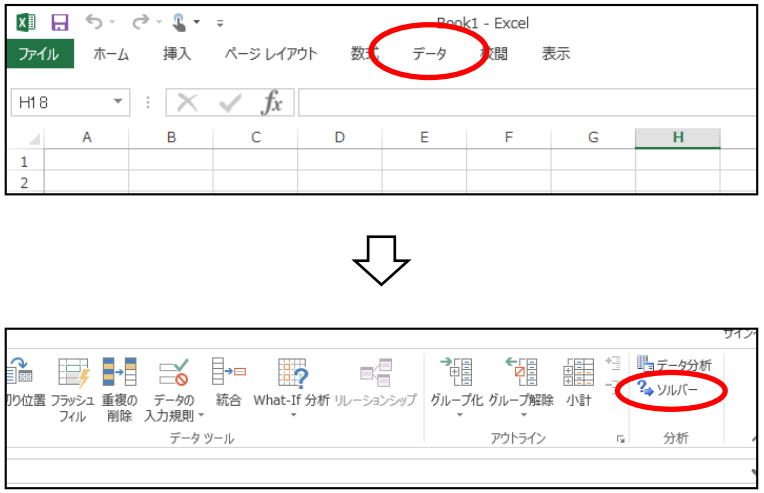


⑥「データ」をクリックして、「分析」が表示されれば、アドインは完了である。

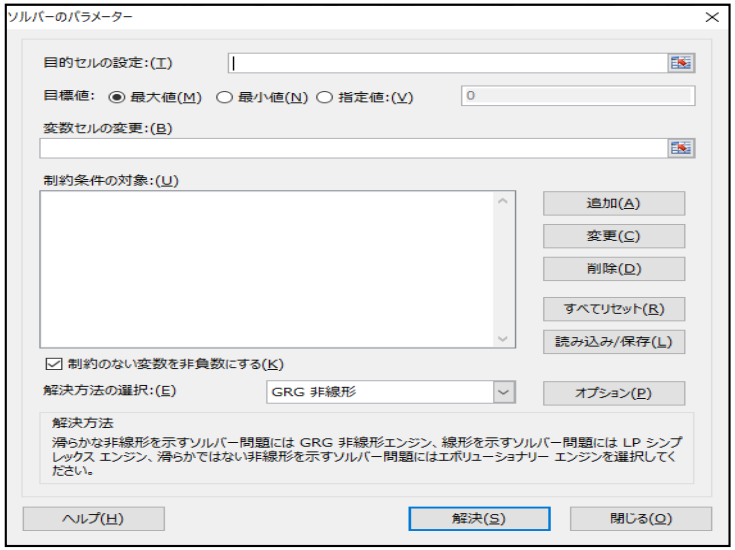


2）主成分分析の準備

①「データ」をクリックして、次いで「分析」にある「ソルバー」をクリックする。



②下記の「ソルバー」が表示される。
この「ソルバー」を使用して主成分分析を行う。



付録 2. Excel を使った主成分分析の方法　2/3

3）主成分分析の実施

ここでは、ソルバーを使用して、第１主成分と第２主成分を求めるところまでを示す。第３主成分以下は、第２主成分の求め方と同じである。

①元のデータを準備する。

標準化	商品品質軸	接客態度軸	接客速度軸	清潔さ軸	保全軸	雰囲気軸
A店舗	0.38023	1.25731	1.16278	0.64550	1.77332	1.66541
B店舗	1.17239	−0.01699	0.45806	0.64550	0.28730	0.67409
C店舗	−0.09506	0.91750	0.10571	−0.16137	−0.70338	0.27757
D店舗	1.48925	1.17236	1.04532	0.64550	−1.19873	0.77323
E店舗	0.69710	−1.29129	0.10571	0.32275	−1.79313	0.07931
F店舗	0.06337	−1.29129	−0.71646	−1.93649	−0.00991	−0.71375
G店舗	−1.99623	−0.01699	−0.12920	0.48412	0.78264	−0.51548
H店舗	0.06337	1.00245	1.16278	1.45237	0.98078	0.77323
I 店舗	−1.20408	−0.44176	−2.00843	−0.64550	0.08916	−1.80419
J店舗	−0.57035	−1.29129	−1.18627	−1.45237	−0.20804	−1.20940
平均	0.00	0.00	0.00	0.00	0.00	0.00
標準偏差	(a) 1.00	(b) 1.00	1.00	1.00	1.00	1.00
分散	1.00	1.00	1.00	1.00	1.00	1.00

例題と同じデータを準備した。
データは標準化している（平均値 = 0　標準偏差 = 1）。 分散は、１である。
　平均値の計算：average
　標準偏差の計算：$varp^{1/2}$
　分散の計算：varp

②ソルバーが算出する、線形変換で使用する係数を入れるカラムを準備する。

(c)	係数1	0	0	0	0	0	0	合計 (d)
	2乗	0	0	0	0	0	0	0

(e)	係数2	0	0	0	0	0	0	合計 (f)
	2乗	0	0	0	0	0	0	0
						積和1－2		0 (g)

第１主成分に関していえば、第１主成分の（c）の係数1は、係数１の２乗和（d）= 1という条件の下で、（a）の分散を最大にする係数1をソルバーを使用して求める。係数１の２乗和を計算するカラム（d）を用意しておく。

第２主成分に関していえば、第２主成分の（e）の係数2は、係数２の２乗和（f）= 1 という条件と、（c）の係数1と（e）の係数2の積和（g）= 0 という条件の下で、（b）の分散を最大にする係数2をソルバーを使用して求める。係数２の２乗和を計算するカラム（f）を用意しておく。また、係数1と係数2の積和を計算するカラム（g）を用意しておく。

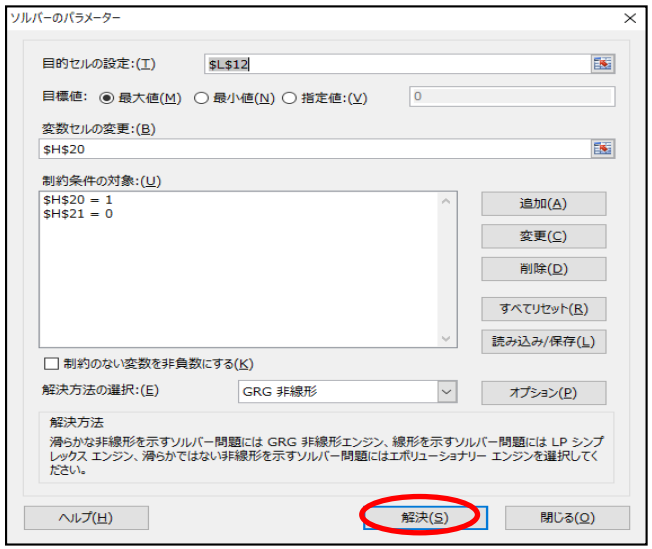
③ソルバーが算出する、線形変換の結果である主成分得点を入れるカラムを準備する。

主成分得点	第1主成分	第2主成分	第3主成分	第4主成分	第5主成分	第6主成分
A店舗	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
B店舗	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
C店舗	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
D店舗	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
E店舗	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
F店舗	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
G店舗	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
H店舗	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
I 店舗	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
J店舗	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
分散	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000

各カラムには、①の元のデータと②の係数の積和の計算式をセットしておく。
また、主成分軸ごとの分散の計算式を収納するカラムを用意しておく。
ソルバーを使用して、最初に第１主成分のカラムのデータを算出し、
ついで第２主成分のカラムのデータを算出する。

付録 2. Excel を使った主成分分析の方法　3/3

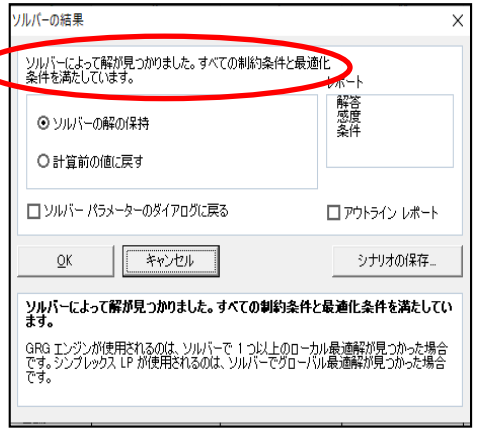
③ソルバーを使って、線形変換の係数と主成分得点を求める。



ここでは、第2主成分の係数2を求めるためのソルバーの使い方を示す。

目的セルの設定には、①の表の分散のセル(b)をセットする。
目標値は、最大値をチェックする、分散(b)を最大化する。
変数セルの変更は、②の表の係数2のセル(e)をセットする。
制約条件の対象は、
　②の表のセル(f)の値を1、
　②の表のセル(g)の値を0 に設定する。
制約のない変数を非負数にする、にはチェックをいれない。
解決方法の選択は、GRG非線形を選ぶ。
以上を設定した上で、解決をクリックする。

問題なく係数2の値が計算できたときは、
右のように「ソルバーによって解が見つかりました。
全ての制約条件と最適化条件を満たしています。」
と表示される。



④ソルバーによって求まった解が表示される。

係数1	0.3050002	0.4346491	0.5032051	0.4402978	0.1348793	0.5027833	合計
2乗	0.0930251	0.1889198	0.2532154	0.1938621	0.0181924	0.2527911	1.000006

係数2	−0.611272	0.247712	−0.075471	0.1253226	0.7335294	−0.074325	合計
2乗	0.3736534	0.0613612	0.0056959	0.0157058	0.5380654	0.0055242	1.0000059
						積和1－2	3.477E-12

主成分得点	第1主成分	第2主成分	第3主成分	第4主成分	第5主成分	第6主成分
A店舗	2.6083125	1.2491667	0.0000000	0.0000000	0.0000000	0.0000000
B店舗	1.2425769	−0.5138911	0.0000000	0.0000000	0.0000000	0.0000000
C店舗	0.3966263	−0.2793985	0.0000000	0.0000000	0.0000000	0.0000000
D店舗	2.0010926	−1.5546991	0.0000000	0.0000000	0.0000000	0.0000000
E店舗	−0.3553229	−2.0347242	0.0000000	0.0000000	0.0000000	0.0000000
F店舗	−2.1152870	−0.5014379	0.0000000	0.0000000	0.0000000	0.0000000
G店舗	−0.6217051	1.8988554	0.0000000	0.0000000	0.0000000	0.0000000
H店舗	2.2006881	0.9658015	0.0000000	0.0000000	0.0000000	0.0000000
I 店舗	−2.7492105	0.8967722	0.0000000	0.0000000	0.0000000	0.0000000
J店舗	−2.6077538	−0.1264292	0.0000000	0.0000000	0.0000000	0.0000000
分散	3.6697753	1.4069806	0.0000000	0.0000000	0.0000000	0.0000000

第1主成分軸の係数1と第2主成分軸の係数2が求まった。
制約条件である、第1主成分軸の係数1の2乗和は1(表示では計算誤差があり、1.000006)、第2主成分軸の係数2の2乗和も1(表示では計算誤差があり、1.0000059)である。
また、係数2のもう一つの制約条件である、係数1と係数2の積和は0(表示では計算誤差があり、3.477×10^{-12})である。

主成分得点は、ここでは、第1主成分軸の値と、第2主成分軸の値を示している。

参考にした書籍

河本薫：「会社を変える分析の力」、 講談社現代新書
小山昇：「数字は人格」、 ダイヤモンド社
永野裕之：「ビジネス×数学＝最強」、 すばる舎
竹内薫：「数学×思考＝ざっくりといかにして問題をとくか」、 丸善出版
中西達夫：「統計データをすぐに分析できる本」、 アニモ出版
中西達夫：「すぐれた判断が統計データ分析から生まれる」、 実務教育出版
豊田裕貴：「マンガでわかる ビジネスを成功に導くデータ分析」、 ナツメ社
向後千春、冨永敦子：「統計学がわかる」、 技術評論社
石井俊全：「意味がわかる統計学」、 ベレ出版
涌井良幸、涌井貞美：「中学数学でわかる統計の授業」、 日本実業出版
涌井良幸、涌井貞美：「統計学の図鑑」、 技術評論社
西内啓：「統計学が最強の学問である」、 ダイヤモンド社
西内啓：「統計学が最強の学問である（実践編）」、 ダイヤモンド社
西内啓：「統計学が最強の学問である（ビジネス編）」、 ダイヤモンド社
森岡毅、今西聖貴：「確率思考の戦略論」、 角川書店
デビッド・マクアダムス：「世界一流企業はゲーム理論で決めている」、 ダイヤモンド社
河村真一ほか：「本物のデータ分析力が身に付く本」、 日経BPムック
末吉正成、末吉美貴：「Excel ビジネス統計分析 この分析できますか？」、 翔泳社
谷岡一郎：「社会調査のウソ リサーチ・リテラシーのすすめ」、 文藝春秋
林知己夫：「調査の科学」、 ちくま学芸文庫
八谷大岳：「データ解析」シリーズ 全15回、（例）データ解析 第1回 ベクトルの復習 – YouTube