多変量解析の基礎として線形代数を学ぶというのが、この講義の趣旨でしたから、細かいことよりも、多変量解析がどのように線形代数とつながっているのか、大きな枠組みがわからば良いと思っていました。その趣旨からすると、主成分分析まで行きたかったのですが、さすがに時間的に無理だったようです。話しておきたかったのは、疑似逆行列を使った重回帰分析と、特異値分解、主成分分析です。多変量解析が、線形代数と関係をざっくりまとめてしまえば、「説明変数が多くて、関係性が把握しにくいデータを、不必要な説明変数を切り捨て、いくつかの説明変数（主成分）に集約し、関係性を把握する。」のが、多変量解析です。線形代数では、関係性を行列で表現しますが、その行列を直交的な関係に投影して、素それぞれの。直交軸の説明力の大きさを、固有値として表現します。今まで説明してきたのは、正則な行列（ざっくり言えば正方行列ですが、形だけが正方行列であっても正則でない行列はいくらでもあります。もう少し数学的に言えば、それに具多的な被説明変数を与えて、それを連立方程式と考えた時に、唯一解が得られる行列が正則です。）でした。また、正方行列でない行列については、その転置行列と掛け合わせて得られる、対称行列を扱ってきました。しかし、実際のデータは、サンプルサイズが測定項目の数よりはるかに多いのが普通でしょう。この場合に、最適な係数行列を計算するのが重回帰分析ですね。ですから、重回帰分析も、データを近似的な最適解に圧縮しているのです。行列では行方向にできることは列方向でもできるのが普通ですから、列方向に圧縮することも可能です。そのための数学的なテクニックが、特異値分解です。また、実施に列方向の項目を集約化しているのが主成分分析です。どちらも線形代数の本、多変量解析の本やネットの解説に転がっています。私のブログにもありますので、そのあたりは自習でお願いします。ということで、その入り口として、線形代数的に疑似逆行列を使って重回帰分析をします。安瀬そのようなことが出来るのかという解説は、私がブログに書いた解説を読んでください。その後、特異値分解、主成分分析を読んでください。ここでは、線形代数と多変量解析の関係が何となく感覚的にわかれば十分だと思います。
いつもは、計算しやすく解説しやすいように、サンプルデータを作るのですが、今回は、それだけの時間的余裕がないので、ネットで拾ったデータを使います。

表１　店の売り上げと店の場所・モーニング・限定メニュー

最寄り駅までの時間	モーニング	店限定商品	年間売上(万円)
6	0	2	7800
3	1	4	8718
1.5	1	5	9401
4	1	1	8596
7	0	0	7235
1.5	1	6	9396
9	0	2	7749
2	1	6	9288
7	0	1	7581
8	1	4	8434

モーニング:モーニングサービス有；１、モーニングサービス無：０

これを定数項も含めて行列形式で書くと以下のようになり、定数項も含めた、係数行列Ｂの最適化問題になります。

$$\begin{pmatrix} 1 & 6 & 0 & 2 \\ 1 & 3 & 1 & 4 \\ 1 & 1.5 & 1 & 5 \\ 1 & 4 & 1 & 1 \\ 1 & 7 & 0 & 0 \\ 1 & 1.5 & 1 & 6 \\ 1 & 9 & 0 & 2 \\ 1 & 2 & 1 & 6 \\ 1 & 7 & 0 & 1 \\ 1 & 8 & 1 & 4 \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} 7800 \\ 8718 \\ 9401 \\ 8596 \\ 7235 \\ 9396 \\ 7749 \\ 9288 \\ 7581 \\ 8434 \end{pmatrix}$$

以下のように行列に名前を付けて

$$A = \begin{pmatrix} 1 & 6 & 0 & 2 \\ 1 & 3 & 1 & 4 \\ 1 & 1.5 & 1 & 5 \\ 1 & 4 & 1 & 1 \\ 1 & 7 & 0 & 0 \\ 1 & 1.5 & 1 & 6 \\ 1 & 9 & 0 & 2 \\ 1 & 2 & 1 & 6 \\ 1 & 7 & 0 & 1 \\ 1 & 8 & 1 & 4 \end{pmatrix}、B = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{pmatrix}、Y = \begin{pmatrix} 7800 \\ 8718 \\ 9401 \\ 8596 \\ 7235 \\ 9396 \\ 7749 \\ 9288 \\ 7581 \\ 8434 \end{pmatrix}$$

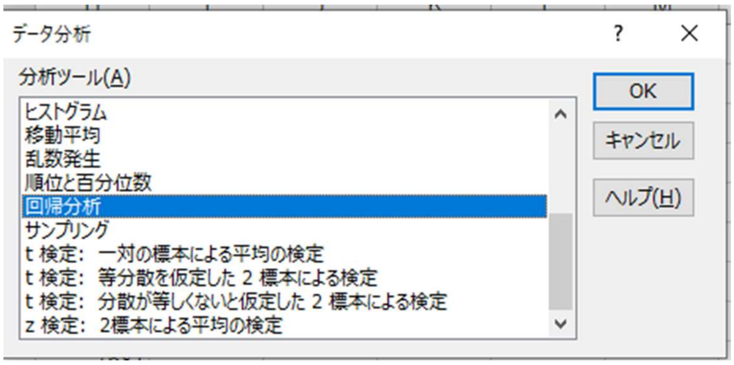
$\|AB - Y\|$のノルムを最小化するわけですが、普通は２次のユークリッドノルムを最小化するので、これを最小二乗法と言います。

これを、解くのは、どのように計算しても構いませんが、ここでは、結果を急ぐので、エクセルのデータ分析を使います。エクセルのデータ分析には重回帰が組み込まれています。

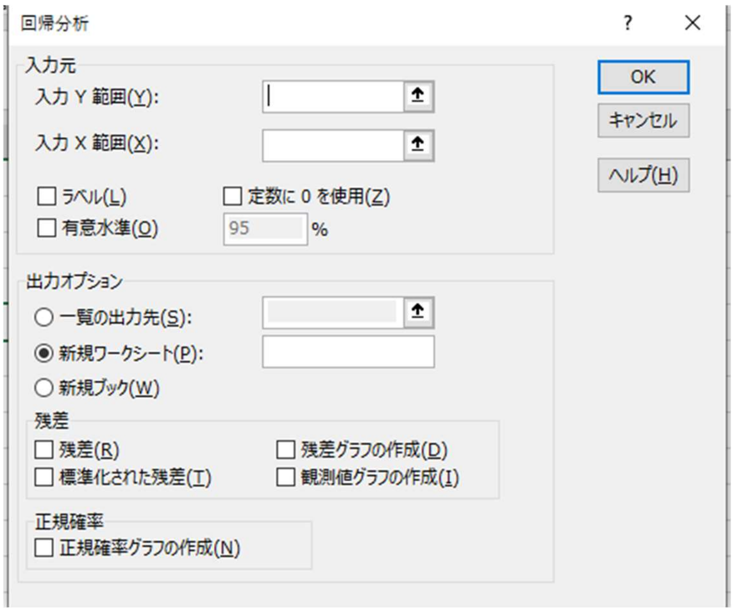
まず、エクセル上に説明変数と非雪面変数の Y の行列を作ります。

	最寄駅までの時間	モーニング	店限定商品数	年間売り上げ（万円）
3	6	0	2	7800
4	3	1	4	8718
5	1.5	1	5	9401
6	4	1	1	8596
7	7	0	0	7235
8	1.5	1	6	9396
9	9	0	2	7749
10	2	1	6	9288
11	7	0	1	7581
12	8	1	4	8434

次に、上部の機能の選択でデータ→データ分析とクリックすると下の画面が出てきます。ここで回帰分析を選択して OK をクリック。



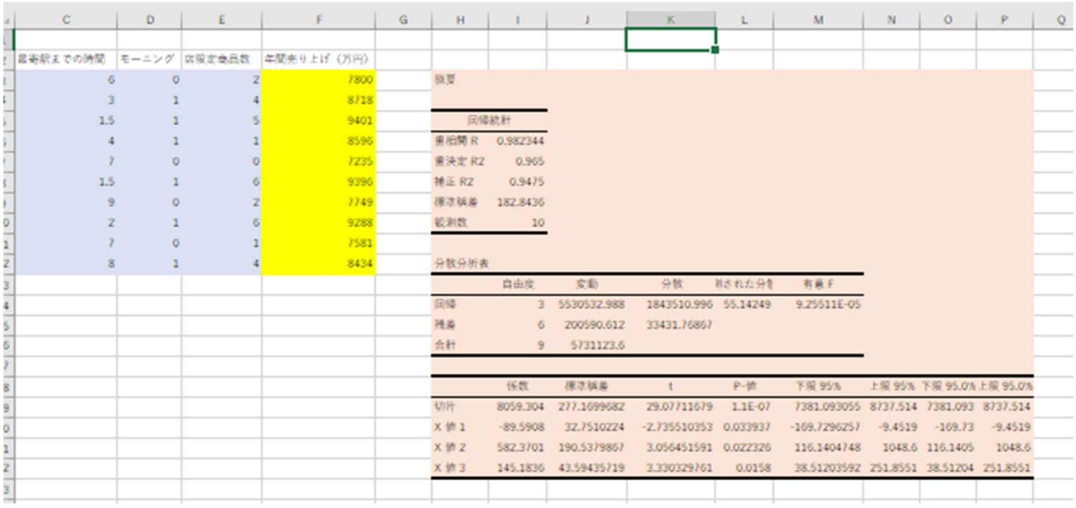
以下の画面になります。



この画面で、入力 Y 範囲(Y)は下図の黄色で示した Y の行列を、入力 X 範囲(X)は青で示し

た A の行列を選択します。有意水準は必要だと思う有意水準に設定してください。その上で、一覧の出力先を指定し、OK をクリックします。

すると、回帰分析の結果が示されます。



最寄駅までの時間	モーニング	店限定食品数	年間売り上げ（万円）
6	0	2	7800
3	1	4	8718
1.5	1	5	9401
4	1	1	8596
7	0	0	7235
1.5	1	6	9396
9	0	2	7749
2	1	6	9288
7	0	1	7581
8	1	4	8434

結果の部分を拡大すると、次のようになっています。

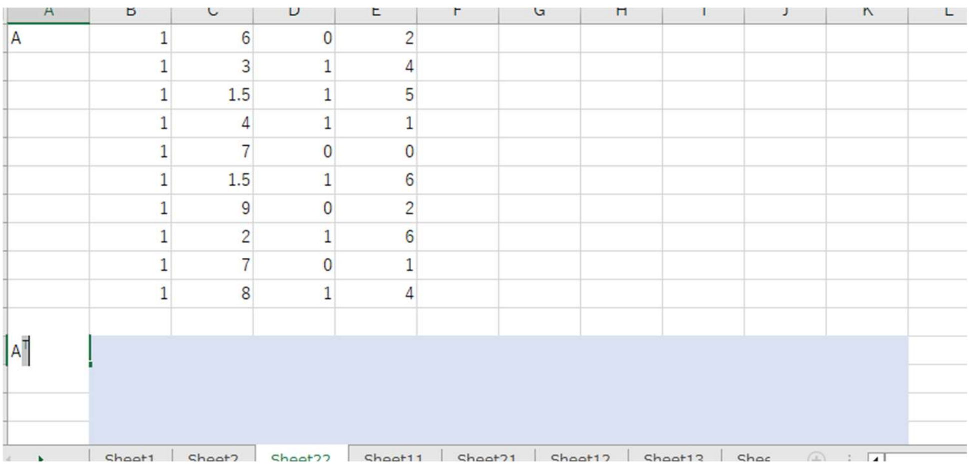
概要								
回帰統計								
重相関 R	0.982344							
重決定 R2	0.965							
補正 R2	0.9475							
標準誤差	182.8436							
観測数	10							
分散分析表								
	自由度	変動	分散	された分散	有意 F			
回帰	3	5530532.988	1843510.996	55.14249	9.25511E-05			
残差	6	200590.612	33431.76867					
合計	9	5731123.6						
	係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%
切片	8059.304	277.1699682	29.07711679	1.1E-07	7381.093055	8737.514	7381.093	8737.514
X 値 1	-89.5908	32.7510224	-2.735510353	0.033937	-169.7296257	-9.4519	-169.73	-9.4519
X 値 2	582.3701	190.5379867	3.056451591	0.022326	116.1404748	1048.6	116.1405	1048.6
X 値 3	145.1836	43.59435719	3.330329761	0.0158	38.51203592	251.8551	38.51204	251.8551

最後の表に定数項（切片）も含めて、最小二乗法で求めた、最適な係数とその有意性が島されます。
これは、知っている人は知っているが、知らなくてもどうでも良いことなのですが、とりあえず、結果がどのようなるか知らないと次の話につながりませんので、どうのようになるかだけを示しておきます。やりたかったことはこの計算ではありません。

エクセル上に定数項も含めた説明変数側の行列を作ります。これを A とします。

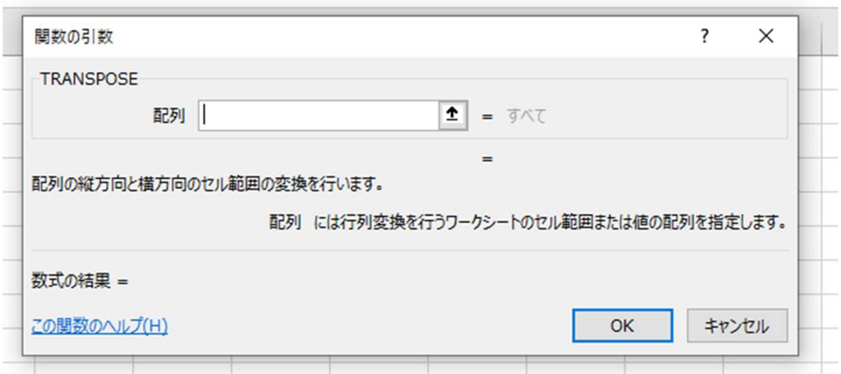
A	1	6	0	2
	1	3	1	4
	1	1.5	1	5
	1	4	1	1
	1	7	0	0
	1	1.5	1	6
	1	9	0	2
	1	2	1	6
	1	7	0	1
	1	8	1	4

次にこの転置行列を作ります。これをA^Tとします。エクセルで転置行列を作るときは、数式の検索・行列のところにあるＴRANSPOSE というコマンドです。このコマンドを動かすにはコツがあって、それを知らないと絶対に動きません。まず、このコマンドでは実施の結果が、一つのセルに出てくるのではなくて、行列で出てきますから、その行列を確保して、そこでこのコマンドを指定します。下図の領域です。その上で、TRANSPOSE を選択します。



A	B	C	D	E
A	1	6	0	2
	1	3	1	4
	1	1.5	1	5
	1	4	1	1
	1	7	0	0
	1	1.5	1	6
	1	9	0	2
	1	2	1	6
	1	7	0	1
	1	8	1	4
A T				

ＴＲANSPOSE を選択すると、下図のような元の行列をしていするウインドウが出て来ます。



この画面で、行列Ａの範囲を指定して、Enter を押せばよいのですが、続いてＯＫを押し

てしまうと、行列は出てきません。ＯＫではなくて、Ctrl と Shift を押しながら、Enter を押します。すると、以下のように転置行列が表示されます。

A^T	1	1	1	1	1	1	1	1	1	1
	6	3	1.5	4	7	1.5	9	2	7	8
	0	1	1	1	0	1	0	1	0	1
	2	4	5	1	0	6	2	6	1	4

A^Tに A を掛けてA^TAを作ります。これは分散共分散行列ですね。エクセルでの行列の掛け算は数式の数学/三角の MMULT というコマンドです。これも行列が出力されるから、先に領域を確保して、その上で、ＭＭＵＬＴを選択します。掛けられる行列、書ける行列を指定した後に、やはりＯＫではなくて、Ctrl と Shift を押しながら、Enter を押します。すると、以下のように行列の積が表示されます。

A^TA	10	49	6	31
	49	312.5	20	113.5
	6	20	6	26
	31	113.5	26	139

次にこの分散共分散行列の逆行列を求めます。エクセルではこの計算は、数式の数学/三角のところにある MINVERSE というコマンドです。これも出力が行列ですから、出力範囲の行列を推定してから、INVERSE を実行し、Ctrl と Shift を押しながら、Enter を押して、出力します。次のような結果になります。

$(A^TA)^{-1}$	2.29791	-0.25141	-0.67933	-0.18012
	-0.25141	0.032084	0.079278	0.015044
	-0.67933	0.079278	1.085935	-0.11635
	-0.18012	0.015044	-0.11635	0.056846

計算の確かさを確認するために、$(A^TA)((A^TA)^{-1})$を計算して、単位行列になることを確認しておいた方が良いでしょう。結果は次のようになりました。

$(A^TA)((A^TA)^{-1}$	1	7.22E-16	8.88E-16	2.22E-16
	0	1	1.78E-15	8.88E-16
	2.7E-15	2.22E-16	1	2.22E-16
	-4E-15	1.78E-15	3.55E-15	1

計算の誤差を考えると、十分な結果でしょう。$(A^TA)^{-1}$に右からさらに、A^Tを掛けます。

$(A^TA)^{-1}A^T$	0.429175	0.143846	0.340844	0.432803	0.538008387	0.16072	-0.32507	0.035013	0.357885	-1.11323
	-0.02882	-0.01571	-0.04879	-0.02876	-0.02682553	-0.0337	0.06743	-0.01771	-0.01178	0.144711
	-0.43636	0.179025	-0.05625	0.607369	-0.12437596	-0.1726	-0.19853	-0.13296	-0.24073	0.575418
	0.02383	-0.02396	0.010318	-0.17946	-0.07481861	0.06716	0.068961	0.074685	-0.01797	0.051255

これを。左から Y の行列にかけます。

$$\begin{bmatrix} 0.429175 & 0.143846 & 0.340844 & 0.432803 & 0.538008387 & 0.16072 & -0.32507 & 0.035013 & 0.357885 & -1.11323 \\ -0.02882 & -0.01571 & -0.04879 & -0.02876 & -0.02682553 & -0.0337 & 0.06743 & -0.01771 & -0.01178 & 0.144711 \\ -0.43636 & 0.179025 & -0.05625 & 0.607369 & -0.12437596 & -0.1726 & -0.19853 & -0.13296 & -0.24073 & 0.575418 \\ 0.02383 & -0.02396 & 0.010318 & -0.17946 & -0.07481861 & 0.06716 & 0.068961 & 0.074685 & -0.01797 & 0.051255 \end{bmatrix} \begin{bmatrix} 7800 \\ 8718 \\ 9401 \\ 8596 \\ 7235 \\ 9396 \\ 7749 \\ 9288 \\ 7581 \\ 8434 \end{bmatrix} = \begin{bmatrix} 8059.304 \\ -89.5908 \\ 582.3701 \\ 145.1836 \end{bmatrix}$$

結果、部分だけを拡大すると、

8059.304
-89.5908
582.3701
145.1836

となります。これは、エクセルのデータ分析で行った回帰分析の結果と完全に一致しています。

つまり、$(A^TA)^{-1}A^T$には、A が正則の行列の場合に、逆行列A^{-1}が、

$$AB = Y$$

という方程式の両辺に、左からA^{-1}という行列を掛けることによって。

$$A^{-1}AB = A^{-1}Y$$

$$IB = A^{-1}Y$$

$$B = A^{-1}Y$$

という、計算によって、唯一解 B を与えるのと同様に、正則でない A について

$$AB = Y$$

という式に対して、左から、$(A^TA)^{-1}A^T$を掛けることによって

$$AB = Y$$

$$(A^TA)^{-1}A^TAB = (A^TA)^{-1}A^TY$$

$$IB = (A^TA)^{-1}A^TY$$

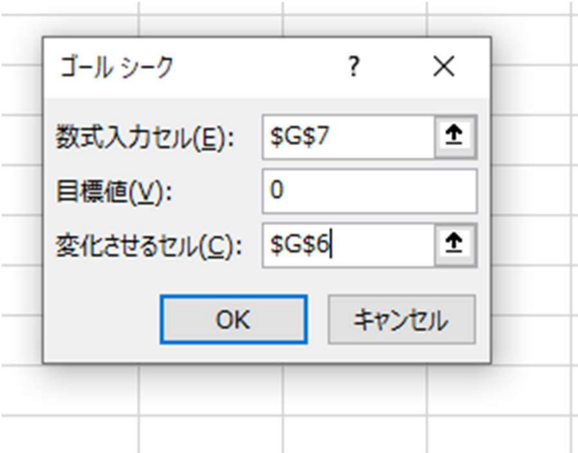
$$B = (A^TA)^{-1}A^TY$$

という形で、最適解、近似解を与えるということです。そういうことから、$(A^TA)^{-1}A^T$のことを、疑似逆行列 Pseudo inverse matrix と言います。何か手品のような感じですが、どうしてそうなるのかということは、私のブログを読んでください。大切なことは、こういう線形代数学的な変形によって、データを集約できるという感覚を持つことです。ここで、こんなことをした目的は、線形代数と多変量解析の関係の感覚的な理解なのですが、エクセルで行列の計算をしたり、重回帰したできると、少し便利なことがあります。エータを入力している途中で、何かと何かに関係がありそうだと思いつくことがあります。そんな時に、エクセル上ですぐ儒回帰などが出来ると便利でしょう。分散共分散行列や相関行列が見たくなることもあります。そんな時に、直ちに、行列の転置や掛け算、逆行列の計算が、簡単なコマンドで計算できると楽でしょう（ちなみに、漁悪豪列のコマンドは、数学/三角のｔところにある MINVERSE というコマンドです。)。ただし、何でもエクセルでしない方が良いということも確かです。例えば、固有値の計算も出来ますがあまりお勧めではありません。いろいろなやり方が考えられますが、私は、データーのところにある、What if 1 分析のゴール・シークという機能で、多次方程式を解くというやり方を試してみました。

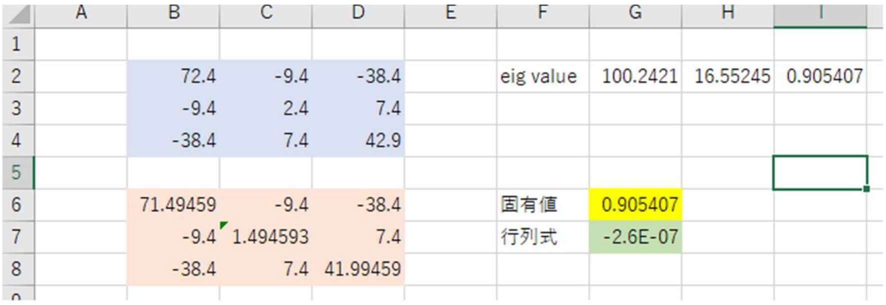
まず、エクセル上に、固有値を求める行列を作ります（青の部分）。次に、暫定的に入れる固有の欄を作って、そこに適当な数値を入れます（黄色のセル）。ここでは、まず、１を入れました。次に、この固有値の値を、先ほど作った行列の対角成分から差し引いた行列を作ります（赤の部分）。この行列の行列式を作ります（緑のセル）（ここの関数は=MDETERM(B6:D8))。

	A	B	C	D	E	F	G	H	I
1									
2		72.4	-9.4	-38.4		eig value	100.2421	16.55245	0.905407
3		-9.4	2.4	7.4					
4		-38.4	7.4	42.9					
5									
6		72.4	-9.4	-38.4		固有値	1		
7		-9.4	2.4	7.4		行列式	1502.3		
8		-38.4	7.4	42.9					

これで準備完了です。ここでゴール・シークを開けて、ゴールのセル、ゴールとする値、変化させるセルを指定して、OK を押します。ここでは、ゴールのセル（数式入力セル）はＧ７（緑色)、ゴールとする値（目標値）は当然、行列式＝０で、変化させるセルは G6(黄色)です。



これらを入力して OK を押せば、しばらくして答えが出ます。



	A	B	C	D	E	F	G	H	I
1									
2		72.4	-9.4	-38.4		eig value	100.2421	16.55245	0.905407
3		-9.4	2.4	7.4					
4		-38.4	7.4	42.9					
5									
6		71.49459	-9.4	-38.4		固有値	0.905407		
7		-9.4	1.494593	7.4		行列式	-2.6E-07		
8		-38.4	7.4	41.99459					

得られた固有値は 0.905407 ですが、３次の正方行列ですから、固有値は最大３つあります。そこで、固有値（G6）に 10 をいれて、もう一度、ゴールシークを実行すると、16.55245 が得られます。さらに、G6 に 100 を入れて、ゴールシークを実行すると 100.2421 が得られます。

固有ベクトルを求めるには、solver を使って、固有値を与えて、基底ベクトルを変化させて、最適化するという計算方法が考えられます。とりあえず、やってみます。以下の行列計算を$\boldsymbol{e}$について解けば良いのですから

$$\left(\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right)\begin{pmatrix} e1 \\ e2 \\ e3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\boldsymbol{e} = \begin{pmatrix} e1 \\ e2 \\ e3 \end{pmatrix}$$

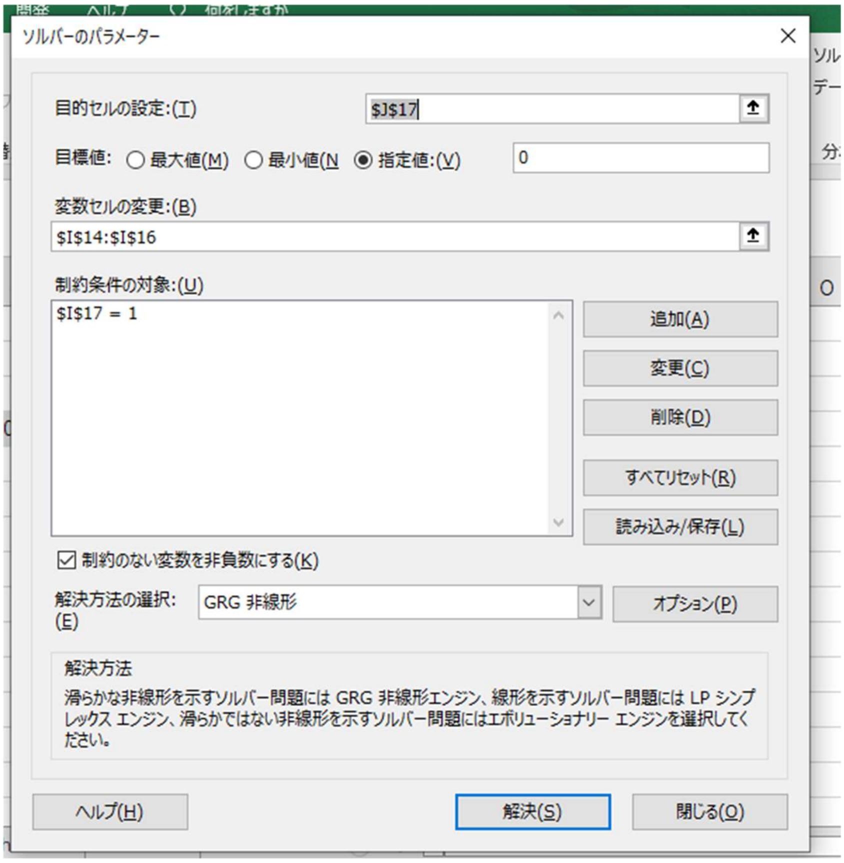
, $\boldsymbol{e}$は基底となってる単位行列で$e1^2 + e2^2 + e3^2 = 1$

エクセル上に固有ベクトルを求める行列を作り（青の部分）、求める固有ベクトルが属する固有値のセルをつくり（灰色の部分）、対処となる行列の対角成分から固有値を差し引いた行列を作り（赤の部分）、これに、暫定的に作った固有ベクトル、ここでは

$\boldsymbol{e} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$とした。緑の部分、単位ベクトルという制約があるので、ベクトルの成分の平方和(SS)を作り、これを1とした。その上で、赤の行列に緑の行列を掛けて、黄色のベクトルを得た。このベクトルを$\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$とすべく、ソルバーを実行するのだが、その目的値として、求めたベクトルの平方和のセルを作った（橙色の部分 J17）

A	B	C	D	E	F	G	H	I	J	K
	72.4	-9.4	-38.4							
	-9.4	2.4	7.4							
	-38.4	7.4	42.9							
					固有値	100.2421				
	-27.8421	-9.4	-38.4				e1	1	-27.8421	
	-9.4	-97.8421	7.4				e2	0	-9.4	
	-38.4	7.4	-57.3421				e3	0	-38.4	
							SS	1	2338.105	

その上で、ソルバーを実行した。ソルバーのパラメータは以下の通り。パラメータを入力し、解決を押す



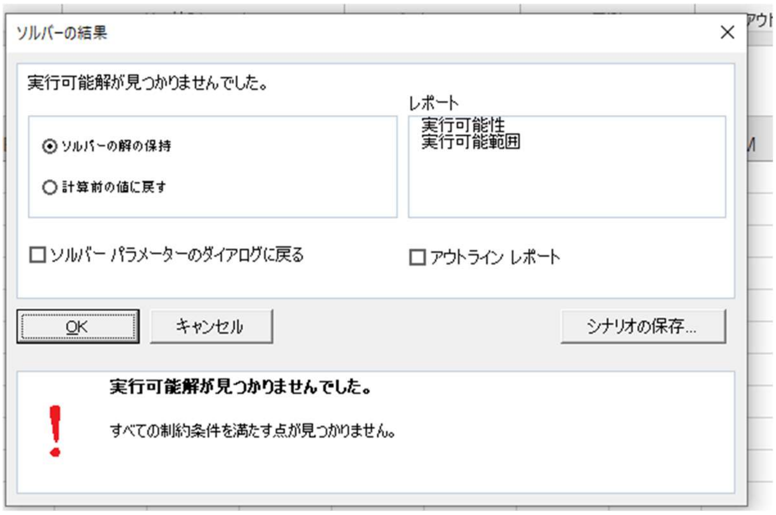
下図がその結果である。

	A	B	C	D	E	F	G	H	I	J
10		72.4	-9.4	-38.4						
11		-9.4	2.4	7.4						
12		-38.4	7.4	42.9						
13						固有値	16.55245			
14		55.84755	-9.4	-38.4				e1	0.571479	0.001073
15		-9.4	-14.1525	7.4				e2	0.048804	-0.00078
16		-38.4	7.4	26.34755				e3	0.819164	-0.0007
17								SS	1	2.26E-06
18										

この例では、固有値 16.55245 を入れた。この場合には、

$$\boldsymbol{e} = \begin{pmatrix} 0.571479 \\ 0.048804 \\ 0.819164 \end{pmatrix}$$

と、おそらく正しい計算結果が出る。しかし、固有値 100.2421 や 0.905407 を入れると、下図のように、計算できないという返答が返ってくる。



おそらく、私のノートパソコンの計算能力が低く、固有値が大きすぎたり小さすぎたりすると、計算が途中でオーバーフローしてしまうのだろう。数値を適当に大きくしたり、小さくしたりすれば、答えが出るかもしれないが、そんなことまでしたくはない。そもそも、固有値を求めるときにも、何回も試行的にゴール・シークを繰り返さなくてはならない。そんなことをしたい暇人はそんなにいないだろう。そういう意味で、エクセル上で、固有値、固有ベクトルを計算するのはお勧めでない。R ならば、極めて簡単にあっという間に答えが返ってくる。ただし、エクセル上で線形代数的な計算をしてみることは、自分の理解を確かめたり、理解を深めるのに役に立つ、ゲーム感覚で解法を考えるのは面白いかもしれない。ネット上にそのような試みの紹介もされている。

実は、この後、特異値分解について、何かやってみるつもりでいました。特異値分解は、線形代数を理解するキーのようなところがありますし、工学分野も含めていろいろなとこ

ろで使われているテクニックだからです。特異値分解を逆方向にした演算で、疑似逆行列が作れるというのをやりたかったのですが、ここ投げた計算事例を使って、両者がピタリと一致するというのを見せたかったのですが、今のところうまくいきません。近いところまでは行くのですが、細かい数値が一致しません。ということで投げ出してしまいました。エクセルで固有値や固有ベクトルを計算する方法を考えるという、つまらないゲームに夢中になって、時間が足りなくなってしまいました。ということで、どなたか上手なやり方を知っていたら教えてください。