因子分析

1. 因子分析の１因子モデルが導出できる
2. 因子分析の１因子モデルが計算できる
3. 因子分析の２因子モデルが導出できる
4. 因子分析の２因子モデルが計算できる

QC プラネッツ© https://qcplanets.com/

因子分析の１因子モデルが導出できる

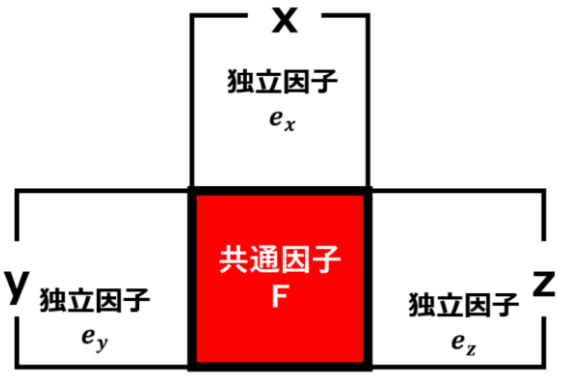
【1】１因子モデルを式で表現

下表のように、例えば、あるデータがあり、変数 x,y,z があるとしましょう。

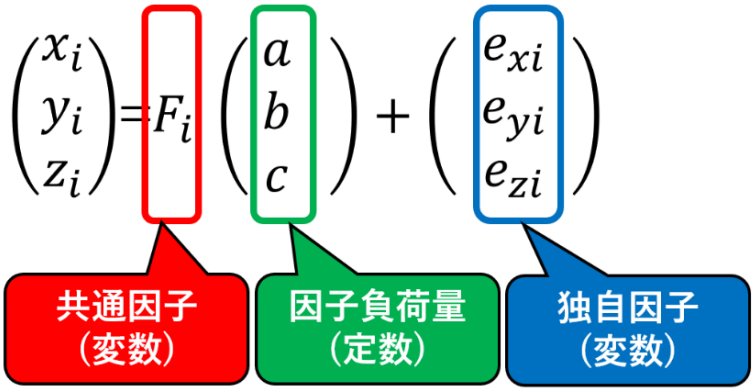
No	x	y	z
1	x_1	y_1	z_1
2	x_2	y_2	z_2
…	…	…	…
n	x_n	y_n	z_n

このデータを、主成分分析のように、ある共通変数 F に集約できなか？を考えます。式と図で表現すると、

$$\begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix} = \begin{pmatrix} aF_i \\ bF_i \\ cF_i \end{pmatrix} + \begin{pmatrix} e_{xi} \\ e_{yi} \\ e_{zi} \end{pmatrix}$$



各値を共通因子と誤差に相当する独自因子に分割します。



式と図から、

●定数である因子負荷量
●変数である共通因子 F
●変数である誤差でもある独自因子 e
にわけて、●変数である共通因子 F に意味を持たせるをやるのが因子分析です。

数式を使って、わかりやすく因子分析を導出していきます。

QC プラネッツ© https://qcplanets.com/

【2】モデル式から分散・共分散を導出
(1) モデル式から分散・共分散を導出

3変数x, y, zで、次の分散
s_x^2, s_y^2, s_z^2
と共分散
s_{xy}, s_{xz}, s_{yz}
を考えます。

① 分散s_x^2

s_x^2, s_y^2, s_z^2のうち、s_x^2を代表として計算します。
$s_x^2 = V(aF + e_x)$
分散の加法性から
$= a^2V(F) + 2aCov(F, e_x) + V(e_x)$
と計算できます。

s_y^2, s_z^2は同様に計算すると、
$s_y^2 = a^2V(F) + 2aCov(F, e_y) + V(e_y)$
$s_z^2 = a^2V(F) + 2aCov(F, e_z) + V(e_z)$
ですね。

② 共分散s_{xy}

s_{xy}, s_{yz}, s_{xz}のうち、s_{xy}を代表として計算します。
$s_{xy} = Cov(aF + e_x, bF + e_y)$
分散の加法性から

$abCov(F, F) + aCov(F, e_y) + bCov(e_x,) + Cov(e_x, e_y)$
$= abV(F) + aCov(F, e_y) + bCov(e_x, F) + Cov(e_x, e_y)$

と計算できます。

s_{yz}, s_{xz}も同様に計算すると、
$s_{yz} = bcV(F) + bCov(F, e_z) + cCov(e_y, F) + Cov(e_y, e_z)$
$s_{xz} =$
$acV(F) + aCov(F, e_z) + cCov(e_x, F) + Cov(e_x, e_z)$
となります。

【3】仮定を入れて式を簡略化
(1) 式を簡略化するため仮定を入れる

●分散の式
$s_x^2 = a^2V(F) + 2aCov(F, e_x) + V(e_x)$
●共分散の式
$s_{xy} =$
$abV(F) + aCov(F, e_y) + bCov(e_x, F) + Cov(e_x, e_y)$ ●分散の式

が長い式なので、仮定を入れて式を簡略化します。

1. 因子間は独立とする 2. 変数を標準化(平均 0,標準偏差 1)とする

つまり、

1. 独立⇒Cov(●,□)=0
2. 標準化⇒S_x^2、S_y^2、S_z^2、$V(F)$はすべて1

では簡略化してみます。

(2) 分散を簡略化

●分散の式

$s_x^2 = a^2V(F) + 2aCov(F, e_x) + V(e_x)$

$1 = a^2 + V(e_x)$

となり、同様に計算して結果をまとめると、

●$1 = a^2 + V(e_x)$
●$1 = b^2 + V(e_y)$
●$1 = c^2 + V(e_z)$

と簡単になりましたね。

(3) 共分散を簡略化

●共分散の式

$s_{xy} = ab$

となり、同様に計算して結果をまとめると、

●$s_{xy} = ab$
●$s_{yz} = bc$
●$s_{xz} = ac$

と簡単になりましたね。

【4】因子負荷量を計算

(1) 因子負荷量を計算

しかも、共分散 s_{xy}, s_{yz}, s_{xz} は実データから計算できるので、a, b, cが計算できますね。

実データから

$s_{xy} \times s_{yz} \times s_{xz} = a^2b^2c^2$ で、この平方根から、

abcとs_{xy}, s_{yz}, s_{xz} の比が a, b, c となり、因子負荷量を計算することができます。

【5】共通因子と誤差を計算

あとは、誤差と共通因子を計算すれば、因子分析の一連の流れが計算できます。

(1) 誤差を計算

誤差$V(e_x), V(e_y), V(e_z)$は関係式から計算できます。

●$V(e_x)$=1-a^2
●$V(e_y)$=1-b^2
●$V(e_z)$=1-c^2

計算はできるのですが、１つ問題があります。

誤差の分散は計算できるが、個別の誤差の値は別の方法で計算が必要。

実は共通因子Ｆの各値を計算してから、個別の誤差の値を計算します。

(2) 共通因子Ｆを計算
実は、推定方法から計算します。この推定方法の導出は QC プラネッツで調査中です。

①共通因子Ｆの計算方法

共通因子$F=ZR^{-1}A$
ただし、
●Z：標準化データ行列
●R：相関係数行列
●A：因子負荷量ベクトル
(1因子モデルはベクトル、2因子以上なら行列)

3 変数、1 因子モデルの場合、具体的に記述すると

$$\begin{pmatrix} F_1 \\ F_2 \\ \vdots \\ F_n \end{pmatrix} = \begin{pmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ \vdots & \vdots & \vdots \\ x_n & y_n & z_n \end{pmatrix} \begin{pmatrix} 1 & r_{xy} & r_{xz} \\ r_{xy} & 1 & r_{yz} \\ r_{xz} & r_{xy} & 1 \end{pmatrix}^{-1} \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

で計算できます。

②誤差 E を計算
誤差 E は残りなので、上の式を使うと、

$$\begin{pmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ \vdots & \vdots & \vdots \\ x_n & y_n & z_n \end{pmatrix} = \begin{pmatrix} F_1 \\ F_2 \\ \vdots \\ F_n \end{pmatrix} \begin{pmatrix} a & b & c \end{pmatrix} + \begin{pmatrix} e_{x1} & e_{y1} & e_{z1} \\ e_{x2} & e_{y2} & e_{z2} \\ \vdots & \vdots & \vdots \\ e_{xn} & e_{yn} & e_{zn} \end{pmatrix}$$

から、誤差行列

$$\begin{pmatrix} e_{x1} & e_{y1} & e_{z1} \\ e_{x2} & e_{y2} & e_{z2} \\ \vdots & \vdots & \vdots \\ e_{xn} & e_{yn} & e_{zn} \end{pmatrix}$$

の個々の値が計算できます。ここまで、各値を計算した上で、共通因子 F の意味を考えて、因子分析ができたといえます。

以上「因子分析の 1 因子モデルが導出できる」を解説しました。

因子分析の１因子モデルが計算できる

【1】データ事例

(1) データ事例

下表のように、例えば、あるデータがあり、変数 x,y,z があるとしましょう。

No	A	B	C
1	30	48	54
2	60	50	60
3	40	42	55
4	80	70	68
5	56	54	71
6	65	70	76
7	22	28	30
8	100	72	96
9	65	70	88
10	32	46	52
合計	550	550	650
平均	55	55	65
標準偏差	23.11	14.21	18.15

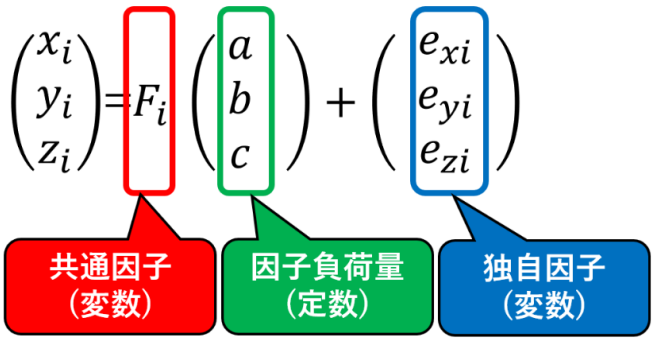
このデータを、主成分分析のように、ある共通変数 F に集約できなか？を考えます。

(2) １因子モデルを式で表現

つまり式と図で表現すると、

$$\begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix} = \begin{pmatrix} aF_i \\ bF_i \\ zF_i \end{pmatrix} + \begin{pmatrix} e_{xi} \\ e_{yi} \\ e_{zi} \end{pmatrix}$$

各値を共通因子と誤差に相当する独自因子に分割します。



式と図から、

●定数である因子負荷量
●変数である共通因子 F
●変数である誤差でもある独自因子 e
にわけて、
●変数である共通因子 F に意味を持たせる
をやるのが因子分析です。

数式を使って、わかりやすく因子分析を導出していきます。

(2) データを標準化
因子分析はモデル式から分散・共分散を導出するときに、データの標準化が必須です。先にデータを標準化しましょう。標準化は $z = \frac{x-\bar{x}}{\sigma}$ です。

No	A	B	C
1	-1.08	-0.49	-0.61
2	0.22	-0.35	-0.28
3	-0.65	-0.92	-0.55
4	1.08	1.06	0.17
5	0.04	-0.07	0.33
6	0.43	1.06	0.61
7	-1.43	-1.9	-1.93
8	1.95	1.2	1.71
9	0.43	1.06	1.27
10	-0.99	-0.63	-0.72
sum	0	0	0
ave	0	0	0
sig	1	1	1

また、分散・共分散も計算すると以下のとおりです。

–	s_x	s_y	s_z
分散	1	1	1
–	s_{xy}	s_{yz}	s_{xz}
共分散	0.878	0.874	0.919

【2】モデル式から分散・共分散を導出
分散と共分散の結果は以下のとおりになります。

- $1=a^2+V(e_x)$
- $1=b^2+V(e_y)$
- $1=c^2+V(e_z)$
- $s_{xy}=ab$
- $s_{yz}=bc$
- $s_{xz}=ac$

【4】因子負荷量を計算
(1) 因子負荷量を計算
共分散 s_{xy}, s_{yz}, s_{xz} は実データから計算できるので、a, b, cが計算できますね。
実データから$s_{xy} \times s_{yz} \times s_{xz} = a^2b^2c^2$ で、この平方根から、
abcとs_{xy}, s_{yz}, s_{xz} の比が a, b, c となり、因子負荷量を計算することができます。

実際に、

●$s_{xy}=ab$=0.878
●$s_{yz}=bc$=0.874
●$s_{xz}=ac$=0.919

全部掛けると
$a^2b^2c^2$=0.878×0.874×0.919=0.840
abc=0.840

●$a=abc/s_{yz}$=0.840/0.919=0.914
●$b=abc/s_{xz}$=0.840/0.875=0.960
●$c=abc/s_{xy}$=0.840/0.878=0.957

となります。

【5】共通因子と誤差を計算
あとは、誤差と共通因子を計算すれば、因子分析の一連の流れが計算できます。
(1) 誤差を計算

誤差$V(e_x), V(e_y), V(e_z)$は関係式から計算できます。

●$V(e_x)$=1-a^2=1-0.914^2=0.165
●$V(e_y)$=1-b^2=1-0.960^2=0.078
●$V(e_z)$=1-c^2=1-0.957^2=0.084

計算はできるのですが、１つ問題があります。

誤差の分散は計算できるが、個別の誤差の値は別の方法で計算が必要。

実は共通因子Ｆの各値を計算してから、個別の誤差の値を計算します。

(2) 共通因子Ｆを計算
実は、推定方法から計算します。この推定方法の導出は QC プラネッツで調査中です。

共通因子$F=ZR^{-1}A$
ただし、
●Z：標準化データ行列
●R：相関係数行列
●A：因子負荷量ベクトル
(1因子モデルはベクトル、2因子以上なら行列)

3 変数、1 因子モデルの場合、具体的に記述すると

$$\begin{pmatrix} F_1 \\ F_2 \\ \vdots \\ F_n \end{pmatrix} = \begin{pmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ \vdots & \vdots & \vdots \\ x_n & y_n & z_n \end{pmatrix} \begin{pmatrix} 1 & r_{xy} & r_{xz} \\ r_{xy} & 1 & r_{yz} \\ r_{xz} & r_{xy} & 1 \end{pmatrix}^{-1} \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

で計算できます。実際に計算してみましょう。

$$\begin{pmatrix} F_1 \\ F_2 \\ F_3 \\ F_4 \\ F_5 \\ F_6 \\ F_7 \\ F_8 \\ F_9 \\ F_{10} \end{pmatrix} = \begin{pmatrix} -1.08 & -0.49 & -0.61 \\ 0.22 & -0.35 & -0.28 \\ -0.65 & -0.92 & -0.55 \\ 1.08 & 1.06 & 0.17 \\ 0.04 & -0.07 & 0.33 \\ 0.43 & 1.06 & 0.61 \\ -1.43 & -1.9 & -1.93 \\ 1.95 & 1.2 & 1.71 \\ 0.43 & 1.06 & 1.27 \\ -0.99 & -0.63 & -0.72 \end{pmatrix} \begin{pmatrix} 1 & 0.3715 & 0.8745 \\ 0.3715 & 1 & 0.7316 \\ 0.8745 & 0.7316 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 0.914 \\ 0.960 \\ 0.957 \end{pmatrix}$$

ここで、相関係数行列の逆行列は、

$$\begin{pmatrix} 12.44 & 7.18 & -16.13 \\ 7.18 & 6.29 & -10.89 \\ -16.13 & -10.89 & 23.07 \end{pmatrix}^{-1}$$

を使います。計算結果は、

$$\begin{pmatrix} F_1 \\ F_2 \\ F_3 \\ F_4 \\ F_5 \\ F_6 \\ F_7 \\ F_8 \\ F_9 \\ F_10 \end{pmatrix} = \begin{pmatrix} -2.248 \\ 0.699 \\ -2.123 \\ 4.857 \\ -1.063 \\ 1.648 \\ -2.189 \\ 2.801 \\ -0.414 \\ -1.968 \end{pmatrix}$$

(3) 誤差 E を計算

誤差 E は残りなので、上の式を使うと、

$$\begin{pmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ \vdots & \vdots & \vdots \\ x_n & y_n & z_n \end{pmatrix} = \begin{pmatrix} F_1 \\ F_2 \\ \vdots \\ F_n \end{pmatrix} \begin{pmatrix} a & b & c \end{pmatrix} + \begin{pmatrix} e_{x1} & e_{y1} & e_{z1} \\ e_{x2} & e_{y2} & e_{z2} \\ \vdots & \vdots & \vdots \\ e_{xn} & e_{yn} & e_{zn} \end{pmatrix}$$

から、誤差行列

$$\begin{pmatrix} e_{x1} & e_{y1} & e_{z1} \\ e_{x2} & e_{y2} & e_{z2} \\ \vdots & \vdots & \vdots \\ e_{xn} & e_{yn} & e_{zn} \end{pmatrix}$$

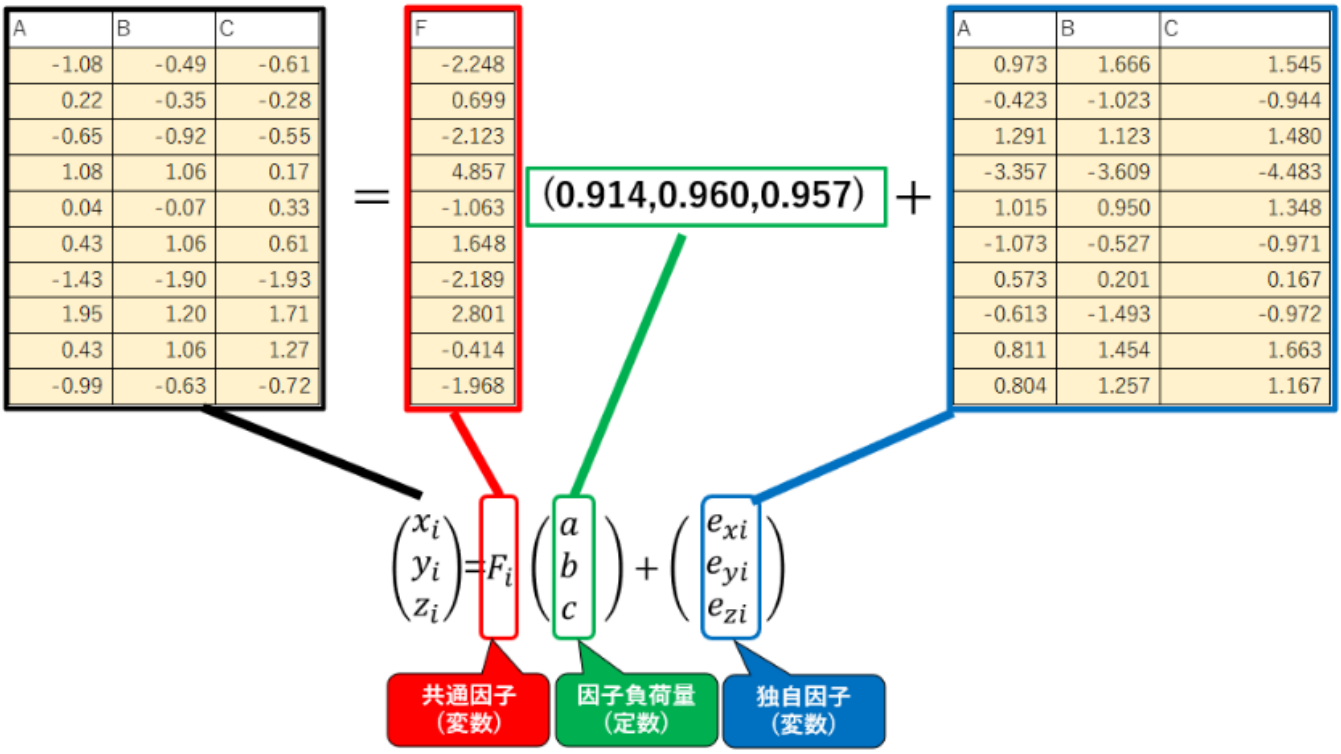
の個々の値が計算できます。

実際に計算すると、下表になります。

QC プラネッツ© https://qcplanets.com/

ei	A	B	C
1	0.973	1.666	1.545
2	-0.423	-1.023	-0.944
3	1.291	1.123	1.48
4	-3.357	-3.609	-4.483
5	1.015	0.95	1.348
6	-1.073	-0.527	-0.971
7	0.573	0.201	0.167
8	-0.613	-1.493	-0.972
9	0.811	1.454	1.663
10	0.804	1.257	1.167

因子分析結果をまとめると、下図になります。



ここまで、各値を計算した上で、共通因子 F の意味を考えて、因子分析ができたといえます。

以上、「因子分析の 1 因子モデルが計算できる」を解説しました。

因子分析の２因子モデルが導出できる

【1】２因子モデルを式で表現

下表のように、例えば、あるデータがあり、変数 x,y,z があるとしましょう。

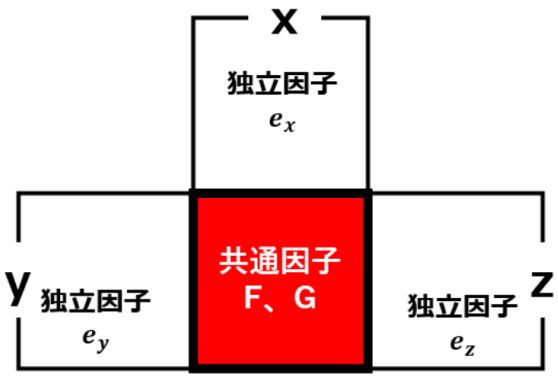
No	x	y	z
1	x_1	y_1	z_1
2	x_2	y_2	z_2
…	…	…	…
n	x_n	y_n	z_n

このデータを、主成分分析のように、２つの共通変数 F、G に集約できなか？を考えます。

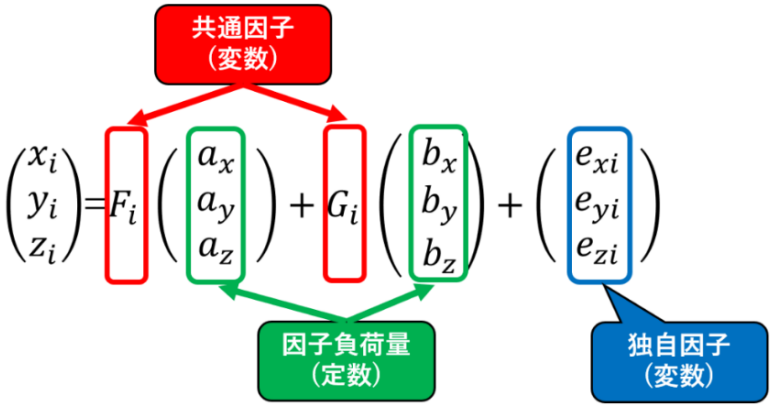
つまり式と図で表現すると、

$$\begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix} = F_i \begin{pmatrix} a_x \\ a_y \\ a_z \end{pmatrix} + G_i \begin{pmatrix} b_x \\ b_y \\ b_z \end{pmatrix} + \begin{pmatrix} e_{xi} \\ e_{yi} \\ e_{zi} \end{pmatrix}$$

各値を共通因子と誤差に相当する独自因子に分割します。



式と図から、



●定数である因子負荷量a, b
●変数である共通因子 F,G
●変数である誤差でもある独自因子 e
にわけて、
●変数である共通因子 F,G に意味を持たせる
をやるのが因子分析です。

数式を使って、わかりやすく因子分析を導出していきます。

QC プラネッツ© https://qcplanets.com/

【2】モデル式から分散・共分散を導出

(1) モデル式から分散・共分散を導出

3変数x, y, zで、次の分散

s_x^2, s_y^2, s_z^2

と共分散

s_{xy}, s_{xz}, s_{yz}

を考えます。

① 分散s_x^2

s_x^2, s_y^2, s_z^2のうち、s_x^2を代表として計算します。

$s_x^2 = V(a_x F + b_x G + e_x)$

分散の加法性から

$= a_x^2 V(F) + b_x^2 V(G) + V(e_x)$

$+ 2a_x b_x Cov(F, G) + 2a_x Cov(F, e_x) + 2b_x Cov(G, e_x)$

と計算できます。

s_y^2, s_z^2は同様に計算すると、

$s_y^2 = a_y^2 V(F) + b_y^2 V(G) + V(e_y)$

$+ 2a_y b_y Cov(F, G) + 2a_y Cov(F, e_y) + 2b_y Cov(G, e_y)$

$s_z^2 = a_z^2 V(F) + b_z^2 V(G) + V(e_z)$

$+ 2a_z b_z Cov(F, G) + 2a_z Cov(F, e_z) + 2b_z Cov(G, e_z)$

ですね。式は１因子モデルより長くなりましたが、機械的に計算できます。

② 共分散s_{xy}

s_{xy}, s_{yz}, s_{xz}のうち、s_{xy}を代表として計算します。

$s_{xy} = Cov(a_x F + b_x G + e_x, a_y F + b_y G + e_y)$

分散の加法性から

$= a_x a_y V(F) + a_x b_y Cov(F, G) + a_x Cov(F, e_y)$

$+ a_y b_x Cov(G, F) + b_x b_y V(G) + b_x Cov(G, e_y)$

$+ a_y Cov(e_x, F) + b_y Cov(e_x, G) + Cov(e_x, e_y)$

と計算できます。

s_{yz}, s_{xz}も同様に計算すると、

$s_{yz} = Cov(a_y F + b_y G + e_y, a_z F + b_z G + e_z)$

$= a_y a_z V(F) + a_y b_z Cov(F, G) + a_y Cov(F, e_z)$

$+ a_z b_y Cov(G, F) + b_y b_z V(G) + b_y Cov(G, e_z)$

$+ a_z Cov(e_y, F) + b_z Cov(e_y, G) + Cov(e_y, e_z)$

$s_{xz} = Cov(a_x F + b_x G + e_x, a_z F + b_z G + e_z)$

$= a_x a_z V(F) + a_x b_z Cov(F, G) + a_x Cov(F, e_z)$

$+ a_z b_x Cov(G, F) + b_x b_z V(G) + b_x Cov(G, e_z)$

$+ a_z Cov(e_x, F) + b_z Cov(e_x, G) + Cov(e_x, e_z)$

と計算できます。

QC プラネッツ© https://qcplanets.com/

【3】仮定を入れて式を簡略化

(1) 式を簡略化するため仮定を入れる

分散の式と共分散の式が長い式なので、仮定を入れて式を簡略化します。

1. 因子間は独立とする ⇒ $Cov(●,□)=0$
2. 変数を標準化(平均 0,標準偏差 1)とする ⇒ S_x^2、S_y^2、S_z^2、$V(F)$, $V(G)$ はすべて 1

では簡略化してみます。

(2) 分散を簡略化

●分散の式

$s_x^2=1$

$=a_x^2V(F)(1)+b_x^2V(G)(1)+V(e_x)$

$+2a_xb_xCov(F,G)+2a_xCov(F,e_x)+2b_xCov(G,e_x)$

$a_x^2+b_x^2+V(e_x)=1$

と簡略化できます。

s_y^2,s_z^2は同様に計算すると、

●s_y^2⇒

$a_y^2+b_y^2+V(e_y)=1$

●s_z^2⇒

$a_z^2+b_z^2+V(e_z)=1$

と簡略化できます。

●$a_x^2+b_x^2+V(e_x)=1$

●$a_y^2+b_y^2+V(e_y)=1$

●$a_z^2+b_z^2+V(e_z)=1$

(3) 共分散を簡略化

●共分散の式

$s_{xy}=\frac{s_{xy}}{1\times1}=$

$\frac{s_{xy}}{s_x^2s_y^2}=r_{xy}$(相関係数)として、

$s_{xy}=r_{xy}$

$=a_xa_yV(F)(1)+a_xb_yCov(F,G)+a_xCov(F,e_y)$

$+a_yb_xCov(G,F)+b_xb_yV(G)(1)+b_xCov(G,e_y)$

$+a_yCov(e_x,F)+b_yCov(e_x,G)+Cov(e_x,e_y)$

$r_{xy}=a_xa_y+b_xb_y$

と計算できます。

s_{yz},s_{xz}も同様に計算すると、

●s_{yz}⇒

$r_{yz}=a_ya_z+b_yb_z$

●s_{xz}⇒

$r_{xz}=a_xa_z+b_xb_z$

と計算できます。

●$r_{xy}=a_xa_y+b_xb_y$

●$r_{yz}=a_ya_z+b_yb_z$

●$r_{xz}=a_xa_z+b_xb_z$

と計算できます。

QC プラネッツ© https://qcplanets.com/

【4】因子負荷量を計算

(1) 因子負荷量を求める方程式

先の結果をまとめると、

●$a_x^2+b_x^2+V(e_x)=1$
●$a_y^2+b_y^2+V(e_y)=1$
●$a_z^2+b_z^2+V(e_z)=1$

と

●$r_{xy}=a_xa_y+b_xb_y$
●$r_{yz}=a_ya_z+b_yb_z$
●$r_{xz}=a_xa_z+b_xb_z$

から、$a_{\bigcirc}$,$b_{\bigcirc}$を計算すればOKですが、

1 因子モデルと違って手計算はかなりキツイ。。。 なので、最小二乗法を活用して Excel で計算します。

(2) SMC 法から因子負荷量の 2 乗和を計算

相関係数行列Rの逆行列のii要素をr^{ii}として、

$h_i^2=1-\frac{1}{r^{ii}}$

$h_i^2=a_i^2+b_i^2$

$V(e_i)=1-h_i^2$

の関係式を使って計算する方法です。この理論の背景は現在調査中です。分かり次第報告します。

(3) 因子負荷量を最小二乗法から計算

SMC法を使って

●$h_x^2=a_x^2+b_x^2$
●$h_y^2=a_y^2+b_y^2$
●$h_z^2=a_z^2+b_z^2$

から、求めます。

●$h_x^2=a_x^2+b_x^2$
●$h_y^2=a_y^2+b_y^2$
●$h_z^2=a_z^2+b_z^2$

と

●$r_{xy}=a_xa_y+b_xb_y$
●$r_{yz}=a_ya_z+b_yb_z$
●$r_{xz}=a_xa_z+b_xb_z$

ここで、Q を

$$Q=((a_x^2+b_x^2)-h_x^2)^2$$
$$+((a_y^2+b_y^2)-h_y^2)^2$$
$$+((a_z^2+b_z^2)-h_z^2)^2$$
$$+2((a_xa_y+b_xb_y)-r_{xy}^2)^2$$
$$+2((a_ya_z+b_yb_z)-r_{yz}^2)^2$$
$$+2((a_xa_z+b_xb_z)-r_{xz}^2)^2$$

と定義して、この Q が最小値（なるべく 0）となる場合を Excel で計算してもらいます。

(4) 最小二乗法の計算を Excel に託す
Q の式を行列表記します。

$$Q=\left(\begin{pmatrix} h_x^2 & r_{xy} & r_{xz} \\ r_{xy} & h_y^2 & r_{xy} \\ r_{xz} & r_{xy} & h_z^2 \end{pmatrix} - \begin{pmatrix} a_x^2+b_x^2 & a_xa_y+b_xb_y & a_xa_z+b_xb_z \\ a_xa_y+b_xb_y & a_y^2+b_y^2 & a_ya_z+b_yb_z \\ a_xa_z+b_xb_z & a_ya_z+b_yb_z & h_z^2 \end{pmatrix}\right)^2$$

上の式を満たすように、
$a_{\bigcirc}, b_{\bigcirc}$ の値を数値計算から求めます。

$$Q=\left\{\left(\begin{pmatrix} a_x^2+b_x^2 & a_xa_y+b_xb_y & a_xa_z+b_xb_z \\ a_xa_y+b_xb_y & a_y^2+b_y^2 & a_ya_z+b_yb_z \\ a_xa_z+b_xb_z & a_ya_z+b_yb_z & a_z^2+b_z^2 \end{pmatrix}\right) - \left(\begin{pmatrix} h_x^2 & h_{xy} & h_{xz} \\ h_{xy} & h_y^2 & h_{yz} \\ h_{xz} & h_{yz} & h_z^2 \end{pmatrix}\right)\right\}^2$$

Qが最小(ほぼ0になるように、
赤行列の各値を数値解析から求める

既知

具体的な計算は、関連記事の計算編で解説します。

因子負荷量 $a_{\bigcirc}, b_{\square}$ が求まります。 ただし、注意してほしいのは、 因子負荷量 $a_{\bigcirc}, b_{\square}$ 求める方程式は自由度が残るので、 因子負荷量は１組ではありませんし、 数値解析の誤差が残ります。

【5】共通因子と誤差を計算
あとは、誤差と共通因子を計算すれば、因子分析の一連の流れが計算できます。

(1) 誤差を計算

誤差 $V(e_x), V(e_y), V(e_z)$ は関係式から計算できます。

誤差はSMC法から求められます。
●$V(e_x)$=1-h_x^2
●$V(e_y)$=1-h_y^2
●$V(e_z)$=1-h_z^2

計算はできるのですが、１つ問題があります。

誤差の分散は計算できるが、個別の誤差の値は別の方法で計算が必要。

実は共通因子Ｆの各値を計算してから、個別の誤差の値を計算します。

(2) 共通因子Ｆを計算
実は、推定方法から計算します。この推定方法の導出は QC プラネッツで調査中です。

★共通因子Ｆの計算方法

共通因子$F=ZR^{-1}A$

ただし、

●Z：標準化データ行列(x_i, y_i, z_iは標準化対応済が前提)

●R：相関係数行列

●A：因子負荷量ベクトル

(1因子モデルはベクトル、2因子以上なら行列)

3 変数、2 因子モデルの場合、具体的に記述すると

$$\begin{pmatrix} F_1 & G_1 \\ F_2 & G_2 \\ \vdots & \\ F_n & G_2 \end{pmatrix} = \begin{pmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ \vdots & \vdots & \vdots \\ x_n & y_n & z_n \end{pmatrix} \begin{pmatrix} h_x^2 & r_{xy} & r_{xz} \\ r_{xy} & h_y^2 & r_{yz} \\ r_{xz} & r_{xy} & h_z^2 \end{pmatrix}^{-1} \begin{pmatrix} a_x & b_x \\ a_y & b_y \\ a_z & b_z \end{pmatrix}$$

で計算できます。

(3) 誤差 E を計算

誤差 E は残りなので、上の式を使うと、

$$\begin{pmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ \vdots & \vdots & \vdots \\ x_n & y_n & z_n \end{pmatrix} = \begin{pmatrix} F_1 \\ F_2 \\ \vdots \\ F_n \end{pmatrix} \begin{pmatrix} a_x & a_y & a_z \end{pmatrix} + \begin{pmatrix} G_1 \\ G_2 \\ \vdots \\ G_n \end{pmatrix} \begin{pmatrix} b_x & b_y & b_z \end{pmatrix} + \begin{pmatrix} e_{x1} & e_{y1} & e_{z1} \\ e_{x2} & e_{y2} & e_{z2} \\ \vdots & \vdots & \vdots \\ e_{xn} & e_{yn} & e_{zn} \end{pmatrix}$$

から、誤差行列

$$\begin{pmatrix} e_{x1} & e_{y1} & e_{z1} \\ e_{x2} & e_{y2} & e_{z2} \\ \vdots & \vdots & \vdots \\ e_{xn} & e_{yn} & e_{zn} \end{pmatrix}$$

の個々の値が計算できます。

ここまで、各値を計算した上で、共通因子 F,G の意味を考えて、因子分析ができたといえます。

【6】因子分析を解析する際の注意点

(1) 注意点

一連の流れを解説してきましたが、どうでしょうか？ 以下の点が因子分析するとき注意が必要です。

1. 因子分析という割に、自分で因子が設定できない(ならば主成分分析でいいじゃん！)
2. 因子負荷量を求める方程式は自由度が残るため、解は１つではない
3. SMC 法などの手法を借りて解くので、精度は低下する
4. 数値解析解から因子Ｆ，Ｇの意味を考えても、結果の精度は低い
5. ２因子モデルで精度が高くないので、多因子になると何を解析しているかがわからなくなる

因子分析は何をやっているのか？よくわからない難しい手法の割に得るものが少ないと QC プラネッツは思います。

因子分析の解析フローを理解して

分析から何を得たいのかはよく考えて使う必要があります。

以上、「因子分析の 2 因子モデルが導出できる」を解説しました。

因子分析の２因子モデルが計算できる

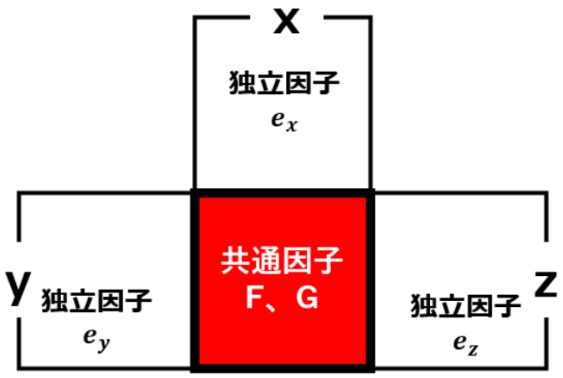
【2】２因子モデルを式で表現
下表のように、例えば、あるデータがあり、変数 x,y,z があるとしましょう。

No	A	B	C
1	30	58	54
2	60	50	60
3	40	60	55
4	80	32	68
5	56	70	71
6	65	54	76
7	22	28	30
8	100	72	96
9	65	80	88
10	32	46	52
和	550	550	650
平均	55	55	65
標準偏差	23.12	15.93	18.15

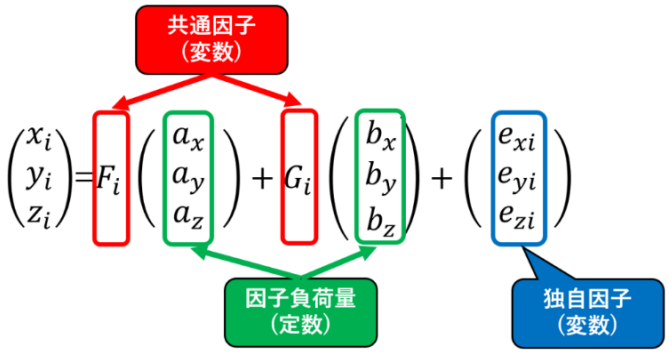
このデータを、主成分分析のように、２つの共通変数 F、G に集約できなか？を考えます。
つまり式と図で表現すると、

$$\begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix} = F_i \begin{pmatrix} a_x \\ a_y \\ a_z \end{pmatrix} + G_i \begin{pmatrix} b_x \\ b_y \\ b_z \end{pmatrix} + \begin{pmatrix} e_{xi} \\ e_{yi} \\ e_{zi} \end{pmatrix}$$

各値を共通因子と誤差に相当する独自因子に分割します。



式と図から、



●定数である因子負荷量a, b
●変数である共通因子 F,G
●変数である誤差でもある独自因子 e
にわけて、
●変数である共通因子 F,G に意味を持たせる
をやるのが因子分析です。

(1) データを標準化
因子分析はモデル式から分散・共分散を導出するときに、データの標準化が必須です。先にデータを標準化しましょう。標準化は $z=\frac{x-\bar{x}}{\sigma}$ です。

No	A	B	C
1	-1.08	0.19	-0.61
2	0.22	-0.31	-0.28
3	-0.65	0.31	-0.55
4	1.08	-1.44	0.17
5	0.04	0.94	0.33
6	0.43	-0.06	0.61
7	-1.43	-1.69	-1.93
8	1.95	1.07	1.71
9	0.43	1.57	1.27
10	-0.99	-0.56	-0.72
和	0	0	0
平均	0	0	0
標準偏差	1	1	1

また、分散・共分散も計算すると以下のとおりです。

–	s_x	s_y	s_z
分散	1	1	1
–	s_{xy}	s_{yz}	s_{xz}
共分散	0.371	0.732	0.874

【2】モデル式から分散・共分散を導出
(1) 詳細な導出
分散・共分散の導出過程は本冊子【因子分析の２因子モデルが導出できる】で解説しています。先にご確認ください。

(2) 結果まとめ
分散と共分散の結果は以下のとおりになります。

●$a_x^2+b_x^2+V(e_x)=1$
●$a_y^2+b_y^2+V(e_y)=1$
●$a_z^2+b_z^2+V(e_z)=1$
と
●$r_{xy}=a_xa_y+b_xb_y$
●$r_{yz}=a_ya_z+b_yb_z$
●$r_{xz}=a_xa_z+b_xb_z$

から、$a_{\bigcirc}$,$b_{\bigcirc}$を計算すればOKですが、

1 因子モデルと違って手計算はかなりキツイ。。。なので、最小二乗法を活用して Excel で計算します。

【3】因子負荷量を計算

(1) SMC 法から因子負荷量の２乗和を計算

相関係数行列Rの逆行列のii要素をr^{ii}として、

h_i^2=1-$\frac{1}{r^{ii}}$

h_i^2=a_i^2+b_i^2

$V(e_i)$=1-h_i^2

の関係式を使って計算する方法です。この理論の背景は現在調査中です。分かり次第報告します。

(2) 因子負荷量を最小二乗法から計算

SMC 法を使って

●$h_x^2=a_x^2+b_x^2$　●$r_{xy}=a_xa_y+b_xb_y$

●$h_y^2=a_y^2+b_y^2$　●$r_{yz}=a_ya_z+b_yb_z$

●$h_z^2=a_z^2+b_z^2$　●$r_{xz}=a_xa_z+b_xb_z$

ここで、Q を

$$
\begin{aligned}
Q=&((a_x^2+b_x^2)-h_x^2)^2\\
&+((a_y^2+b_y^2)-h_y^2)^2\\
&+((a_z^2+b_z^2)-h_z^2)^2\\
&+2((a_xa_y+b_xb_y)-r_{xy}^2)^2\\
&+2((a_ya_z+b_yb_z)-r_{yz}^2)^2\\
&+2((a_xa_z+b_xb_z)-r_{xz}^2)^2
\end{aligned}
$$

と定義して、この Q が最小値（なるべく 0）となる場合を Excel で計算してもらいます。

(3) 最小二乗法の計算を Excel に託す

Q の式を行列表記します。

$$
Q=\left(\begin{pmatrix} h_x^2 & r_{xy} & r_{xz} \\ r_{xy} & h_y^2 & r_{xy} \\ r_{xz} & r_{xy} & h_z^2 \end{pmatrix} - \begin{pmatrix} a_x^2+b_x^2 & a_xa_y+b_xb_y & a_xa_z+b_xb_z \\ a_xa_y+b_xb_y & a_y^2+b_y^2 & a_ya_z+b_yb_z \\ a_xa_z+b_xb_z & a_ya_z+b_yb_z & h_z^2 \end{pmatrix}\right)^2
$$

上の式を満たすように、

$a_{\bigcirc}$,$b_{\bigcirc}$の値を数値計算から求めます。

$$
\boxed{Q}=\left\{\left(\boxed{\begin{pmatrix} a_x^2+b_x^2 & a_xa_y+b_xb_y & a_xa_z+b_xb_z \\ a_xa_y+b_xb_y & a_y^2+b_y^2 & a_ya_z+b_yb_z \\ a_xa_z+b_xb_z & a_ya_z+b_yb_z & a_z^2+b_z^2 \end{pmatrix}}\right) - \boxed{\begin{pmatrix} h_x^2 & h_{xy} & h_{xz} \\ h_{xy} & h_y^2 & h_{yz} \\ h_{xz} & h_{yz} & h_z^2 \end{pmatrix}}\right\}^2
$$

Qが最小(ほぼ0になるように、
赤行列の各値を数値解析から求める

既知

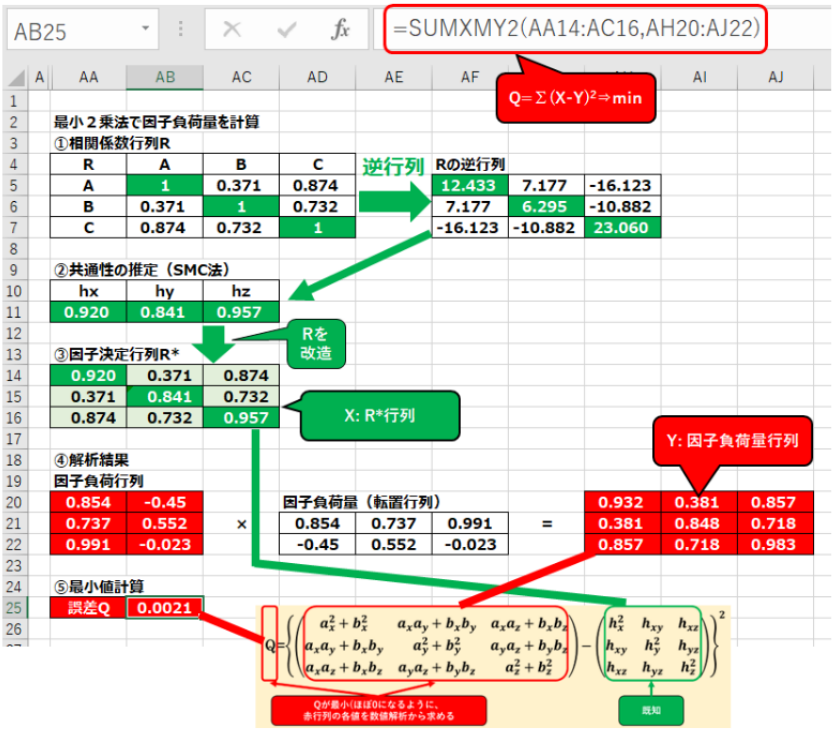
(4) Excel を使った計算方法

上の式で行列表記したものを下図のように代入します。

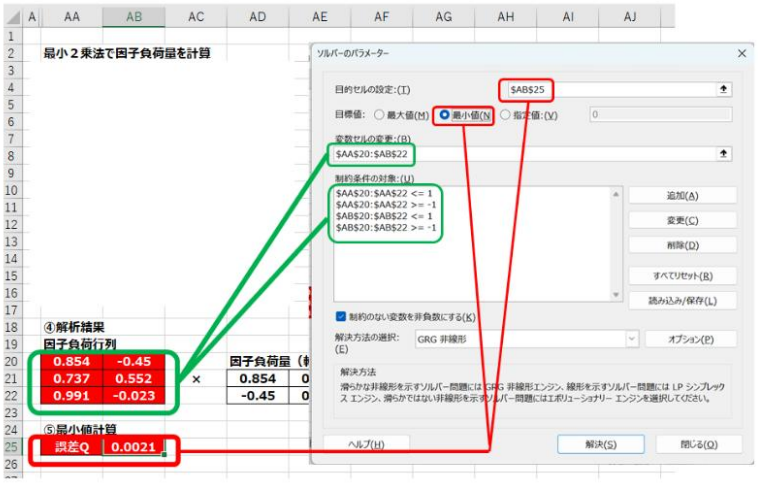
●③因子決定行列 R*

●④因子負荷行列

●⑤最小値計算 Q



Excel の「データ」⇒「ソルバー」で、Q の最小値となる因子負荷量の各値に制約条件を入れて求めます。



注意なのは、

●因子負荷量の値が 1 つに決まらないこと
●数値計算がうまく収束しないこと
●Q の最小値が 0.01 程度まで下がるまで制約条件を変えて解析する必要がある
Excel での計算は一瞬ですが、解の妥当性はよく確認することが必要です。

今回は、Excel がはじき出した値を使って、残りの値を計算します。

●a_x=0.854、a_y=0.737、a_z=0.991
●b_x=-0.45、b_y=0.552、b_z=-0.023

【4】共通因子と誤差を計算
あとは、誤差と共通因子を計算すれば、因子分析の一連の流れが計算できます。

(1) 誤差を計算

誤差$V(e_x), V(e_y), V(e_z)$は関係式から計算できます。

誤差は SMC 法から求められます。 ●$V(e_x)$=1-h_x^2 ●$V(e_y)$=1-h_y^2 ●$V(e_z)$=1-h_z^2

計算はできるのですが、１つ問題があります。

誤差の分散は計算できるが、個別の誤差の値は別の方法で計算が必要。

実は共通因子Ｆの各値を計算してから、個別の誤差の値を計算します。

(2) 共通因子Ｆを計算
実は、推定方法から計算します。この推定方法の導出は QC プラネッツで調査中です。

(i) 共通因子Ｆの計算方法

共通因子$F=ZR^{-1}A$ ただし、 ●Z：標準化データ行列(x_i, y_i, z_iは標準化対応済が前提) ●R：相関係数行列 ●A：因子負荷量ベクトル (1因子モデルはベクトル、2因子以上なら行列)

3 変数、2 因子モデルの場合、具体的に記述すると

$$\begin{pmatrix} F_1 & G_1 \\ F_2 & G_2 \\ \vdots & \\ F_n & G_2 \end{pmatrix} = \begin{pmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ \vdots & \vdots & \vdots \\ x_n & y_n & z_n \end{pmatrix} \begin{pmatrix} h_x^2 & r_{xy} & r_{xz} \\ r_{xy} & h_y^2 & r_{yz} \\ r_{xz} & r_{xy} & h_z^2 \end{pmatrix}^{-1} \begin{pmatrix} a_x & b_x \\ a_y & b_y \\ a_z & b_z \end{pmatrix}$$

で計算できます。実際に計算してみると

$$\begin{pmatrix} F_1 & G_1 \\ F_2 & G_2 \\ \vdots & \\ F_n & G_2 \end{pmatrix} = \begin{pmatrix} -1.08 & 0.19 & -0.61 \\ 0.22 & -0.31 & -0.28 \\ \vdots & \vdots & \vdots \\ -0.99 & -0.56 & -0.72 \end{pmatrix} \begin{pmatrix} 12.433 & 7.177 & -16.123 \\ 7.177 & 6.295 & -10.882 \\ -16.123 & -10.882 & 23.060 \end{pmatrix} \begin{pmatrix} 0.854 & -0.45 \\ 0.737 & 0.552 \\ 0.991 & -0.023 \end{pmatrix}$$

$$= \begin{pmatrix} -0.571 & 1.023 \\ -0.303 & -0.626 \\ -0.545 & 0.579 \\ 0.121 & -1.961 \\ 0.334 & 0.649 \\ 0.615 & -0.142 \\ -1.924 & -0.423 \\ 1.663 & -0.702 \\ 1.294 & 1.141 \\ -0.683 & 0.462 \end{pmatrix}$$

で計算できます。

(3)誤差 E を計算

誤差 E は残りなので、上の式を使うと、

$$\begin{pmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ \vdots & \vdots & \vdots \\ x_n & y_n & z_n \end{pmatrix} = \begin{pmatrix} F_1 \\ F_2 \\ \vdots \\ F_n \end{pmatrix} \begin{pmatrix} a_x & a_y & a_z \end{pmatrix} + \begin{pmatrix} G_1 \\ G_2 \\ \vdots \\ G_n \end{pmatrix} \begin{pmatrix} b_x & b_y & b_z \end{pmatrix} + \begin{pmatrix} e_{x1} & e_{y1} & e_{z1} \\ e_{x2} & e_{y2} & e_{z2} \\ \vdots & \vdots & \vdots \\ e_{xn} & e_{yn} & e_{zn} \end{pmatrix}$$

から、誤差行列

$$\begin{pmatrix} e_{x1} & e_{y1} & e_{z1} \\ e_{x2} & e_{y2} & e_{z2} \\ \vdots & \vdots & \vdots \\ e_{xn} & e_{yn} & e_{zn} \end{pmatrix}$$

の個々の値が計算できます。

実際に誤差行列を計算すると

$$\begin{pmatrix} -0.134 & 0.044 & -0.017 \\ 0.194 & 0.255 & 0.011 \\ 0.077 & 0.396 & 0.002 \\ 0.095 & -0.45 & 0 \\ 0.05 & 0.337 & 0.014 \\ -0.156 & -0.438 & -0.007 \\ 0.026 & -0.043 & -0.03 \\ 0.211 & 0.229 & 0.043 \\ -0.159 & -0.014 & 0.011 \\ -0.204 & -0.316 & -0.029 \end{pmatrix}$$

ここまで、各値を計算した上で、共通因子 F,G の意味を考えて、因子分析ができたといえます。

因子分析の解析フローを理解してデータ事例を使って計算できました。

以上、「因子分析の 2 因子モデルが計算できる」を解説しました。