因子分析

探索的因子分析： 共通因子の意味を仮定しないで分析→ 意味を考える

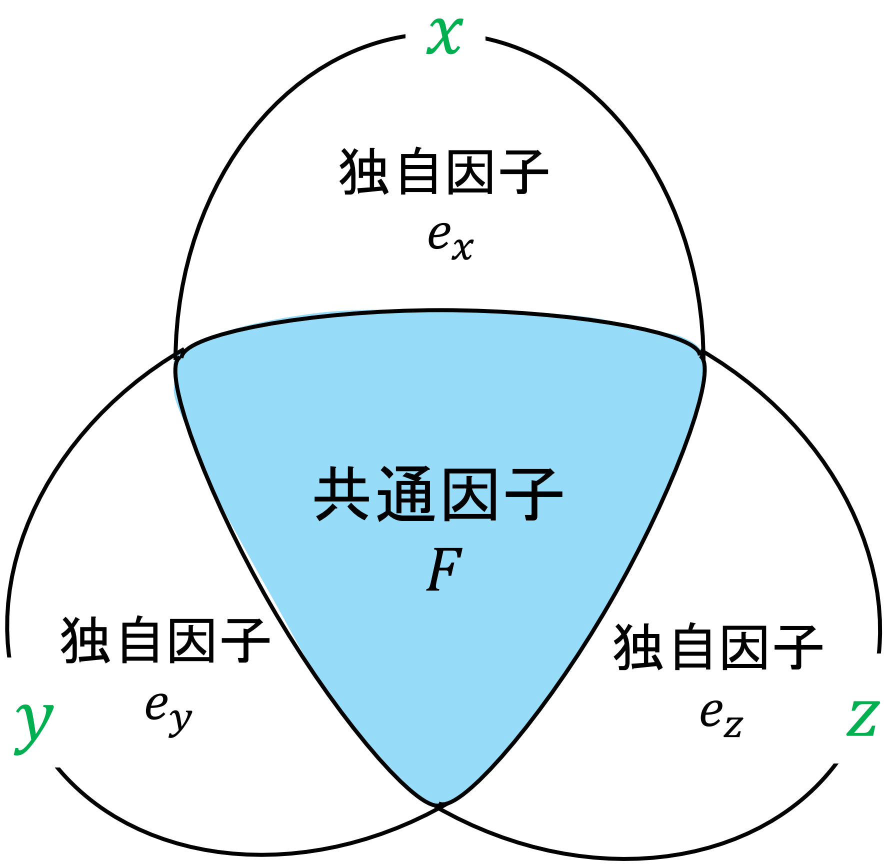
確認的因子分析： 意味を仮定→ 分析

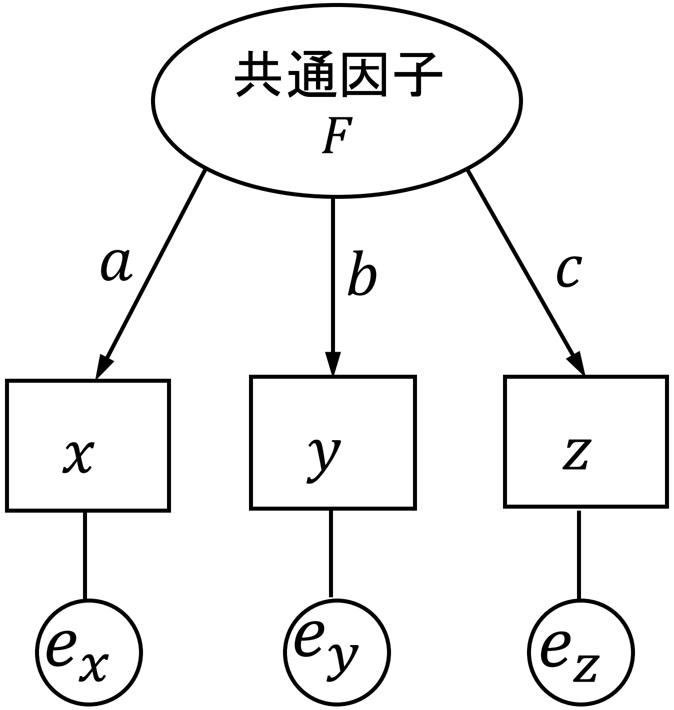
$$x = aF + e_x$$
$$y = bF + e_y$$
$$z = cF + e_z$$

a、b及びcは因子負荷量

	x	y	z
	万人	件/千人	件/千人
	人口	旅券発行	婚姻率
北海道	554	17.0	5.17
青森	139	11.6	4.55
岩手	135	12.9	4.66
宮城	234	19.4	5.46
秋田	111	13.9	4.00
山形	119	16.8	4.56
福島	205	18.3	4.92
茨城	296	26.8	5.25
栃木	201	24.4	5.52
群馬	201	23.5	5.14

〜





xの分散$S_x{}^2$（＝V_x）と共分散S_{xy}

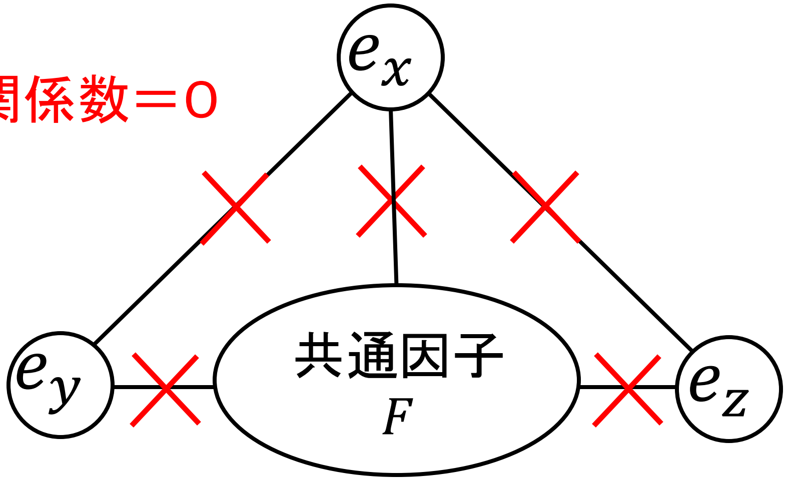
$$S_x{}^2 = V(aF + e_x) = a^2V(F) + aCov(F, e_x) + V(e_x)$$

$$S_{xy} = Cov\left(aF + e_x, bF + e_y\right)$$

$$= abCov(F, F) + aCov\left(F, e_y\right) + bCov(e_x, F) + Cov\left(e_x, e_y\right)$$

$$= abV(F) + aCov(F, e_y) + bCov(e_x, F) \;+\; Cov\left(e_x, e_y\right)$$

$$\left.\begin{aligned} &Cov(F, e_x) = Cov\left(F, e_y\right) = Cov(F, e_z) = 0 \\ &Cov\left(e_x, e_y\right) = Cov\left(e_y, e_z\right) = Cov(e_z, e_x) = 0 \end{aligned}\right\}$$ 相互に相関係数＝0



$$S_x{}^2 = V(aF + e_x) = a^2V(F) + V(e_x)$$

$$S_{xy} = abV(F)$$

$$S_x{}^2 = 1 \quad V(F) = 1$$ ←標準化

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

$$S_{xy} = ab \qquad S_{yz} = bc \qquad S_{zx} = ca$$

	人口	旅券発行	婚姻率
北海道	554	17.0	5.17
青森	139	11.6	4.55
岩手	135	12.9	4.66
宮城	234	19.4	5.46
秋田	111	13.9	4.00
山形	119	16.8	4.56
福島	205	18.3	4.92
茨城	296	26.8	5.25
栃木	201	24.4	5.52
群馬	201	23.5	5.14
≀			
熊本	182	19.8	5.17
大分	120	19.8	5.25
宮崎	114	15.5	5.47
鹿児島	172	15.4	5.05
沖縄	138	19.8	6.28
平均値	272	24	5
σ	261.0	7.8	0.6

標準化

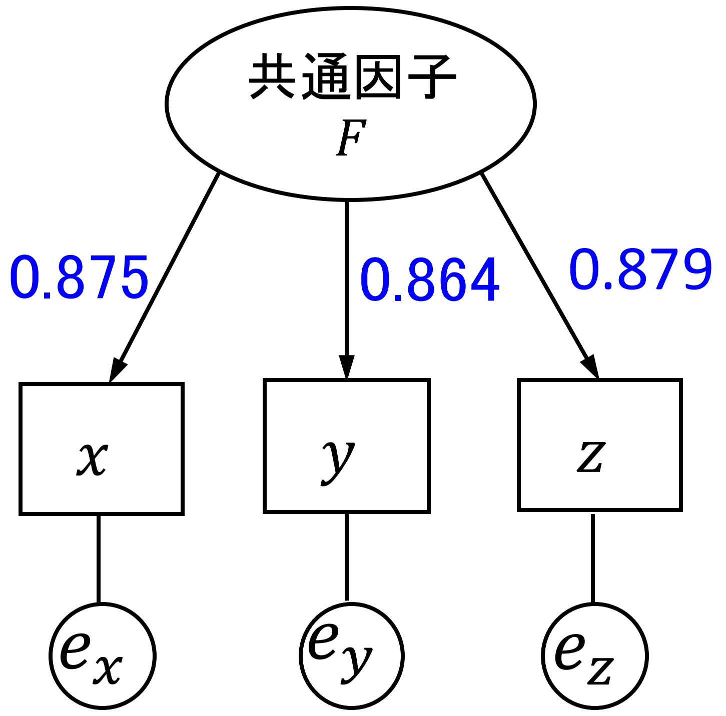
⇨

人口	旅券発行	婚姻率
1.08	−0.90	−0.09
−0.51	−1.59	−1.19
−0.52	−1.43	−1.00
−0.14	−0.60	0.43
−0.62	−1.30	−2.17
−0.59	−0.93	−1.17
−0.26	−0.74	−0.53
0.09	0.35	0.05
−0.27	0.04	0.53
−0.27	−0.07	−0.14
	≀	
−0.34	−0.55	−0.09
−0.58	−0.55	0.05
−0.60	−1.10	0.44
−0.38	−1.11	−0.30
−0.51	−0.55	1.88

Sxy	0.756	ab
Syz	0.760	bc
Szx	0.770	ca
abc	0.665	
a	0.875	
b	0.864	
c	0.879	

共分散

=COVAR(G3:G49,H3:H49)



標準化

$$\frac{\text{データ} - \text{平均値}}{\text{標準偏差}\sigma}$$

前ページより

$S_{xy} = ab$

$S_{yz} = bc$

$S_{zx} = ca$

$a^2\ b^2\ c^2 = ab \times bc \times ca = 0.756 \times 0.760 \times 0.770$

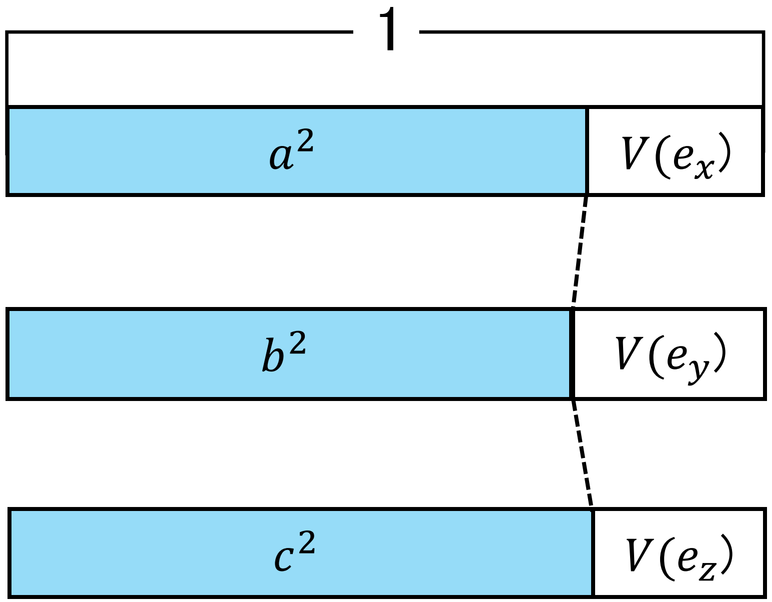
$abc = 0.665$

↓

$a = 0.875$

$b = 0.864$

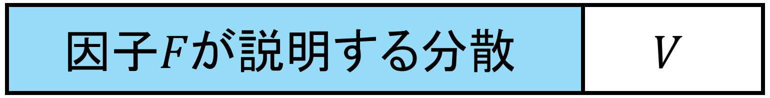
$c = 0.879$



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

a =0.875 $\rightarrow$ $a^2 = 0.766$
b =0.864 $\rightarrow$ $b^2 = 0.746$
c =0.879 $\rightarrow$ $c^2 = 0.773$

全分散量$= {S_x}^2 + {S_y}^2 + {S_z}^2 = 1 + 1 + 1 = 3$
因子Fが説明する分散の総和$h^2 = a^2 + b^2 + c^2 = 0.766 + 0.746 + 0.773 = 2.285$
寄与率$= \dfrac{\text{因子}F\text{が説明する分散}}{\text{全分散量}V} = \dfrac{2.285}{3} = 0.762$



寄与率　0.762

式の導出

$f(x) = ax + b$

	x	y	z
1	x_1	x_1	x_1
2	x_2	x_2	x_2
…	…	…	…
n	x_n	x_n	x_n

$$aCov(x, x) = \frac{1}{n}\{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2\} = V(x)$$

$$V(ax + b) = \frac{1}{n}[\{(ax_1 + b) - (a\bar{x} + b)\}^2 + \cdots + \{(ax_n + b) - (a\bar{x} + b)\}^2]$$
$$= a^2 \frac{1}{n}\{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2\} = a^2V(x)$$

$$V(x + y) = \frac{1}{n}[\{(x_1 + y_1) - (\bar{x} + \bar{y})\}^2 + \cdots + \{(x_n + y_n) - (\bar{x} + \bar{y})\}^2]$$
$$= \frac{1}{n}[\{(x_1 - \bar{x}) + (y_1 - \bar{y})\}^2 + \cdots + \{(x_n - \bar{x}) + (y_n - \bar{y})\}^2]$$
$$= \frac{1}{n}\{(x_1 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2\} + 2\frac{1}{n}\{(x_1 - \bar{x})(y_1 - \bar{y}) + \cdots + (x_n - \bar{x})(y_n - \bar{y})\}$$
$$+ \frac{1}{n}\{(y_1 - \bar{y})^2 + \cdots + (y_n - \bar{y})^2\} = V(x) + 2Cov(x, y) + V(y)$$

$$Cov(ax, by) = \frac{1}{n}\{(ax_1 - a\bar{x})(by_1 - b\bar{y}) + \cdots + (ax_n - a\bar{x})(by_n - b\bar{y})\}$$
$$= ab\frac{1}{n}\{(x_1 - \bar{x})(y_1 - \bar{y}) + \cdots + (x_n - \bar{x})(y_n - \bar{y})\} = abCov(x, y)$$

$$Cov(x + y, z) = \frac{1}{n}\{(x_1 + y_1 - \bar{x} - \bar{y})(z_1 - \bar{z}) + \cdots + (x_n + y_n - \bar{x} - \bar{y})(z_n - \bar{z})\}$$
$$= \frac{1}{n}\{(x_1 - \bar{x})(z_1 - \bar{z}) + \cdots + (x_n - \bar{y})(z_n - \bar{z})\} + \frac{1}{n}\{(y_1 - \bar{y})(z_1 - \bar{z}) + \cdots + (y_n - \bar{y})(z_n - \bar{z})\}$$
$$= Cov(x, z) + Cov(y, z)$$