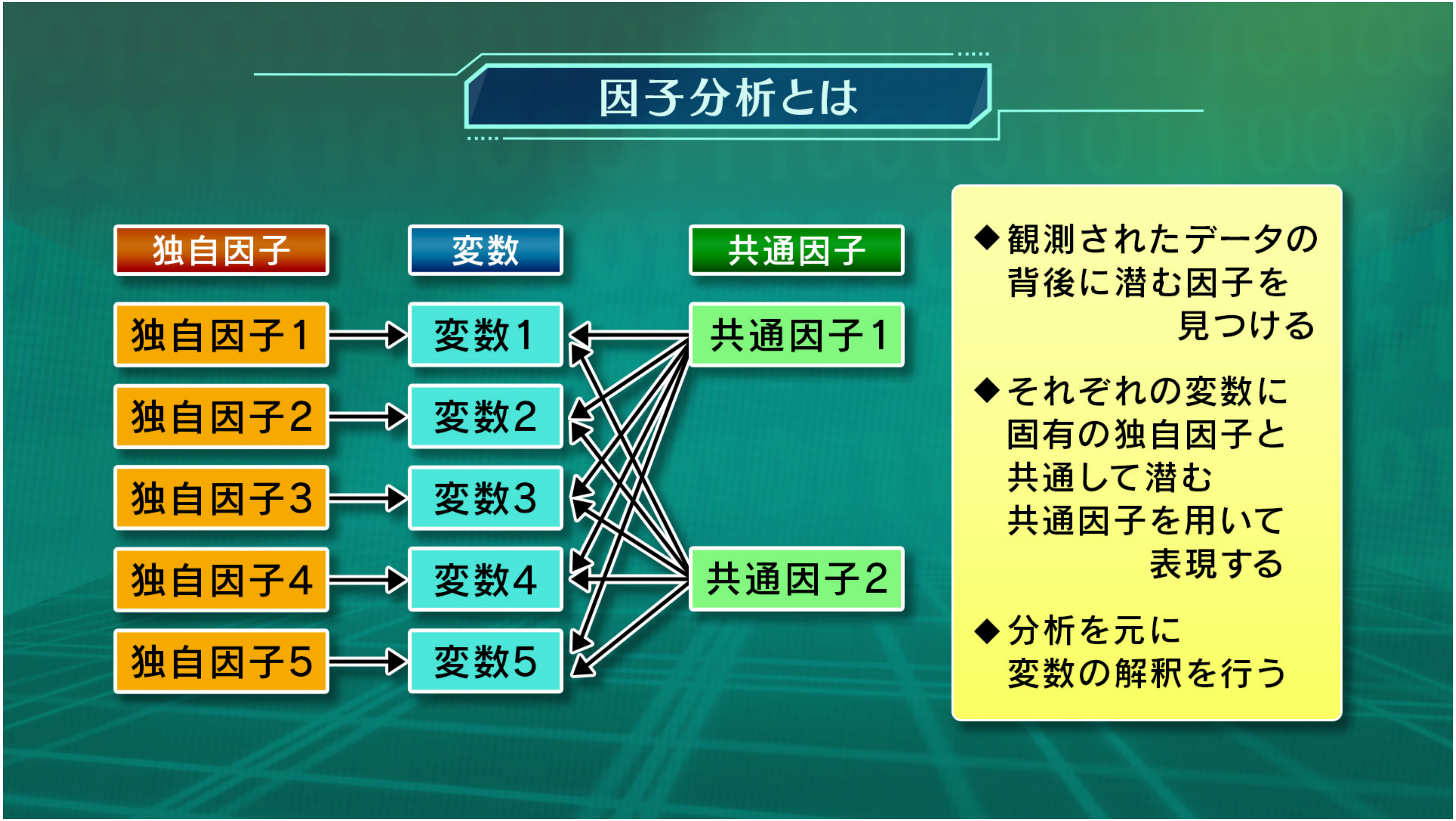
データの分析と知識発見



因子分析とは
独自因子
変数
共通因子
独自因子1
独自因子2
独自因子3
独自因子4
独自因子5
変数1
変数2
変数3
変数4
変数5
共通因子1
共通因子2
◆観測されたデータの背後に潜む因子を見つける
◆それぞれの変数に固有の独自因子と共通して潜む共通因子を用いて表現する
◆分析を元に変数の解釈を行う

主成分分析と因子分析

カットですか？

主成分分析

$$
\begin{aligned}
z_1 &= a_{11} y_1 + a_{12} y_2 + \cdots + a_{1r} y_r \\
z_2 &= a_{21} y_1 + a_{22} y_2 + \cdots + a_{2r} y_r \\
&\vdots \\
z_m &= a_{r1} y_1 + a_{r2} y_2 + \cdots + a_{mr} y_n
\end{aligned}
$$

因子分析

$$
\begin{aligned}
Y_1 &= c_{11} F_1 + c_{12} F_2 + \cdots + c_{1m} F_m + d_1 E_1 \\
Y_2 &= c_{21} F_1 + c_{22} F_2 + \cdots + c_{2m} F_m + d_2 E_2 \\
&\vdots \\
Y_n &= c_{n1} F_1 + c_{n2} F_2 + \cdots + c_{nm} F_m + d_n E_n
\end{aligned}
$$

$$Y = CF + DE$$

因子分析の定式化と用語

$$
\begin{aligned}
Y_1 &= c_{11}F_1 + c_{12}F_2 + \cdots + c_{1m}F_m + d_1E_1 \\
Y_2 &= c_{21}F_1 + c_{22}F_2 + \cdots + c_{2m}F_m + d_2E_2 \\
&\vdots \qquad\qquad\qquad\qquad \vdots \\
Y_n &= c_{n1}F_1 + c_{n2}F_2 + \cdots + c_{nm}F_m + d_nE_n
\end{aligned}
$$

$$Y = CF + DE$$

因子分析の定式化と用語

$$
\begin{aligned}
Y_1 &= c_{11}F_1 + c_{12}F_2 + \cdots + c_{1m}F_m + d_1E_1 \\
Y_2 &= c_{21}F_1 + c_{22}F_2 + \cdots + c_{2m}F_m + d_2E_2 \\
&\;\;\vdots \qquad\qquad\qquad \vdots \\
Y_n &= c_{n1}F_1 + c_{n2}F_2 + \cdots + c_{nm}F_m + d_nE_n
\end{aligned}
$$

共通因子

$$
Y = CF + DE
$$

因子分析の定式化と用語

$$
\begin{aligned}
Y_1 &= c_{11}F_1 + c_{12}F_2 + \cdots + c_{1m}F_m + d_1 E_1 \\
Y_2 &= c_{21}F_1 + c_{22}F_2 + \cdots + c_{2m}F_m + d_2 E_2 \\
&\ \ \vdots \qquad\qquad\qquad\qquad \vdots \\
Y_n &= c_{n1}F_1 + c_{n2}F_2 + \cdots + c_{nm}F_m + d_n E_n
\end{aligned}
$$

独自因子

$$
Y = CF + DE
$$

因子分析の定式化と用語

$$
\begin{aligned}
Y_1 &= c_{11} F_1 + c_{12} F_2 + \cdots + c_{1m} F_m + d_1 E_1 \\
Y_2 &= c_{21} F_1 + c_{22} F_2 + \cdots + c_{2m} F_m + d_2 E_2 \\
&\;\;\vdots \qquad\qquad\qquad\qquad \vdots \\
Y_n &= c_{n1} F_1 + c_{n2} F_2 + \cdots + c_{nm} F_m + d_n E_n
\end{aligned}
$$

因子負荷量

$$
Y = CF + DE
$$

共通因子 と 独自因子

Fのことを**共通因子**　Eを**独自因子**という

① 共通因子 独自因子は平均0 分散を1とする

② 共通因子はどの独自因子とも相関がない

③ 独自因子はそれぞれ独立で相関がない

④ 共通因子間には相関がある場合とない場合の両方を考える

相関がない場合を**直交解**　ある場合を**斜交解**という

因子分析の手順

共通因子の個数を決める

- カイザーガットマン基準（相関行列の固有値が1以上の個数）
- スクリープロット（相関行列の固有値のグラフから判断）

因子負荷量を求める

- 最尤法
- 最小二乗法
- 主因子法

因子軸の回転を行う

結果を解釈する

因子分析の定式化と用語

$$Y_1 = c_{11}F_1 + c_{12}F_2 + \cdots + c_{1m}F_m + d_1E_1 \qquad h_1^2 = 1 - d_1^2$$

$$Y_2 = c_{21}F_1 + c_{22}F_2 + \cdots + c_{2m}F_m + d_2E_2 \qquad h_2^2 = 1 - d_2^2$$

$$\vdots \qquad\qquad\qquad \vdots \qquad\qquad\qquad \vdots$$

$$Y_n = c_{n1}F_1 + c_{n2}F_2 + \cdots + c_{nm}F_m + d_nE_n \qquad h_n^2 = 1 - d_n^2$$

独自性 d_i^2

共通性 h_i^2

因子の回転

因子負荷量は一意に決まるわけではない

回転によって 特徴が分かりやすく より解釈しやすい軸を探す

直交回転

- ◇ バリマックス法
- ◇ オーソマックス規準を最大化

斜交回転

- ◇ プロマックス法
- ◇ バリマックス解を元にプロクラステス回転を行い特徴を際立たせる

因子分析の定式化と用語

$$
\begin{aligned}
Y_1 &= c_{11}F_1 + c_{12}F_2 + \cdots + c_{1m}F_m + d_1E_1 \\
Y_2 &= c_{21}F_1 + c_{22}F_2 + \cdots + c_{2m}F_m + d_2E_2 \\
&\ \ \vdots \qquad\qquad\qquad\quad \vdots \\
Y_n &= c_{n1}F_1 + c_{n2}F_2 + \cdots + c_{nm}F_m + d_nE_n
\end{aligned}
$$

因子負荷量2乗和 $\displaystyle\sum_{i=1}^{n} c_{ij}^2$

Rによるシミュレーション

```
> w1 <- read_csv("fac01.csv")
> head(w1)
> w2 <- cor(w1[,-1] )
> y <- eigen(w2)$values
> x <- 1:length(y)
>  ggplot() + geom_point(aes(x=x,y=y) )+
+ geom_line(aes(x=x,y=y))+  geom_hline(yintercept = 1)
```

架空の試験の結果
因子数の決定

```
> w3 <- factanal(w1[,-1],   2,   rotation="promax")
> w4 <- as_tibble(w2$loadings[1:5,1:2])
> w4 <- w4 %>% mutate(subject=rownames(w2$loadings))
> ggplot(w4,aes(x=Factor1,y=Factor2) ) + geom_point()+
 + geom_text(aes(label=subject),vjust=-1 )+
 + ylim(-0.2,0.9)
```

factanal:因子分析

rotation
何もしない: “none”
バリマックス法: “varimax”
プロマックス法: “promax”

as_tibble データフレームに変換
#
ylim : y軸の範囲を指定