



# グラフィカルモデリング

*RICOH*  
*Quality Management Division*  
*Motohisa Hirono*

<ファイル:ES-C.DAT>

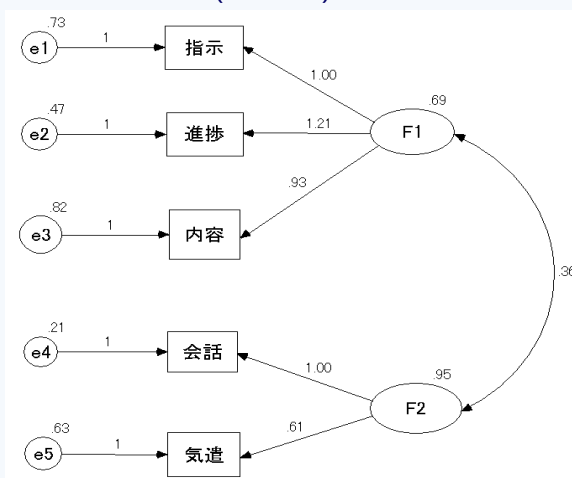
## 相関係数行列

解析対象個体数 = 101

No.	指示	進捗	内容	会話	気遣
指示	1.000	0.563	0.421	0.425	0.270
進捗	0.563	1.000	0.571	0.293	0.180
内容	0.421	0.571	1.000	0.170	0.158
会話	0.425	0.293	0.170	1.000	0.539
気遣	0.270	0.180	0.158	0.539	1.000

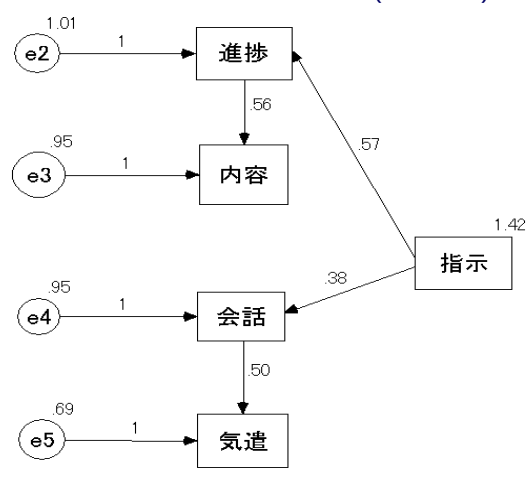
変数間の相関から傾向は見てくるが内部構造は分からない  
主成分分析や因子分析は表線から構造探索  
グラフィカルモデリングは裏線から構造探索  
により探索的に因果モデルまで作成できる

因子分析(表線)モデル



$\chi^2 = 8.42, df=4, p=0.077$   
 GFI=0.966, AGFI=0.874, NFI=0.938  
 RMR=0.068

グラフィカルモデル(裏線)



$\chi^2 = 2.19, df=6, p=0.709$   
 GFI=0.985, AGFI=0.963, NFI=0.973  
 RMR=0.040

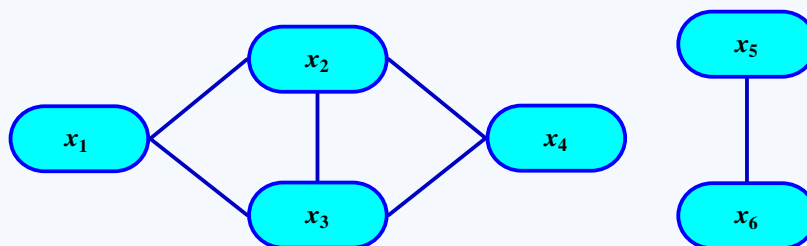
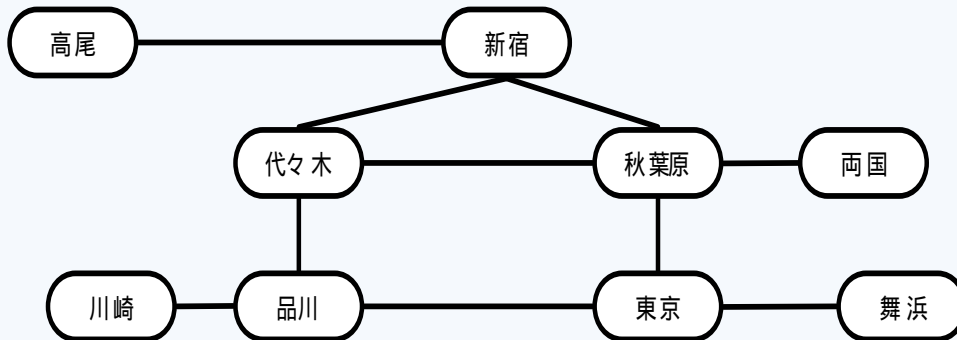
本日のプログラム

お話し内容	タイムスケジュール(あくまで予定)
量的変数のグラフィカルモデリングとは	30分の超特急簡易版
G-GMでできること 事前分析 無向独立グラフ 有向独立グラフ 連鎖独立グラフ 古典的多変量解析(主成分分析&クラスタ分析)	60分のハンズオン 10分のウォーミングアップ 15分のハンズオン 15分のハンズオン 10分のハンズオン ハンズオン(時間次第)
質的変数のグラフィカルモデリングとは	30分の超特急簡易版
L-GMでできること 事前分析 無向独立グラフ 有向独立グラフ 連鎖独立グラフ 古典的多変量解析(対応分析&対数線形モデル)	60分のハンズオン 10分のウォーミングアップ 15分のハンズオン 15分のハンズオン 10分のハンズオン ハンズオン(時間次第)

- **本編:書き下ろしテキスト(4章構成)**
- **応用編:行動計量学会での発表要旨(2編)**
- **簡易マニュアル**
- **ソフトウェア(CD 或はUSBメモリ)**

- **グラフィカルモデリングにおけるグラフ**
  - **グラフとその意味**
  - **多変量解析におけるグラフの利用**
  - **因果関係を表すグラフ**
  - **解析結果の解釈におけるグラフの役割**

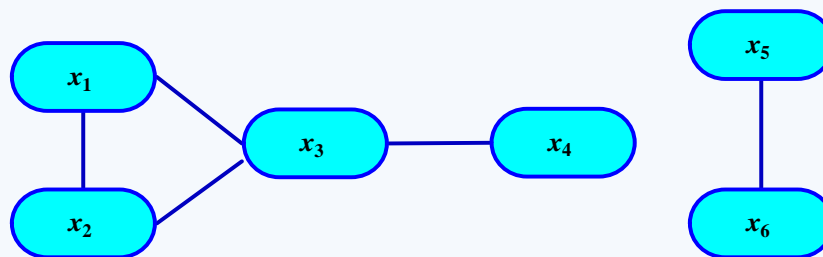
- データを視覚的に要約する統計グラフ
  - 円グラフ, 棒グラフ, 折れ線グラフ
- グラフィカルモデリングでのグラフ
  - グラフ理論に基づいて統計と融合したもの
  - 例: 鉄道の路線網



- 頂点(x1)と頂点(x2)のように1つの線で直接結ばれている関係: **直接効果**
- 頂点(x1)と頂点(x4)のように1つの線では結ばれていないが, 間に頂点(x2), (x3)を介することで道が通じている関係: **間接効果**
- 頂点(x1)と頂点(x5)のように複数の線をたどっても道が通じていない関係: **独立性**

➡ グラフィカルモデリングでは, このような関係を統計理論に結び付けている

- 多変量データから, 変数間の関係を探索する
    - 対称的關係
    - 因果的關係
  - 変数間の関係をグラフと統計量で表す
  - グラフから読み取れる因果・相関構造の解釈が重要
    - 直接効果
    - 間接効果
- 偏相関の概念 条件付独立性の概念



直接効果・間接効果を分類せよ

- 線と矢線の区別
  - 線に矢のないグラフ: 無向グラフ (Undirected graph)  
対称的關係  
2つの因子に交互作用がある: 対称的關係
  - 矢のあるグラフ: 有向グラフ (Directed graph)  
因果的關係, 時間的順序關係
- 間接効果: 連関図による説明
  - 1次原因は問題点に向い矢線がある
  - 2次原因は1次原因に向い矢線があるが, 問題点への矢線がない
  - これは, 階層的に間接効果と直接効果を使い分けている

## ● グラフィカルモデリング

- 直接効果, 間接効果は確率論で厳密に区分される
- グラフィカルモデリングのグラフは定量的(客観)関係
- 矢線の関係
  - 頂点:変数:有向グラフの場合は因果関係の記述
  - 有向グラフは, データを分析しながら構築する(探索的データ解析)
- 統計的因果関係
  - ある原因変数と結果変数において, 他の原因変数の値を固定した状態で, 原因変数の値を変化させると結果変数の分布の位置が変化する
- パス解析(path analysis)
  - 多変量の間関係を有向グラフ(統計的因果関係)のみでグラフにモデル化する方法:パスダイアグラム
  - パスダイアグラムの作成は事前に先見情報からモデルを作成し, 実際のデータを使い検証する(確証的データ解析)

## 第3の変量の影響を除外する方法

<ファイル:県データ-C.CSV>

### ● 関数関係が不明な場合

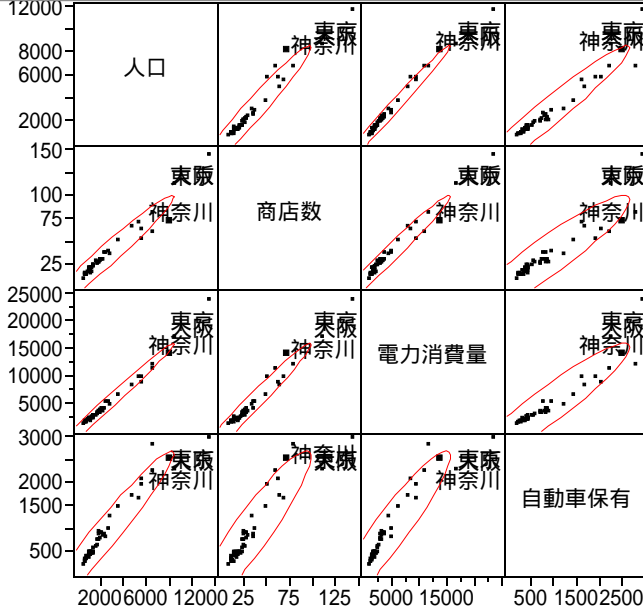
- 第3の変量との比を取る  
(経済データ等に有効な初等的処置)
  - 県の経済データについて、人口、電力消費量、自動車保有台数、商店数の人口(=県の規模)によらない実質的相関を求めたい場合には、各変量を人口で割っておく

### ● 関数関係が認められる場合

- 回帰分析の残差を利用する
  - 直線的傾向が認められない場合は2次項、変数変換などを試みる
- 偏相関係数の利用

- 「数量データ分析入門」のSFCオリジナルデータ  
( <http://www.dsci.sfc.keio.ac.jp/> )
- 47都道府県の経済データが入っている

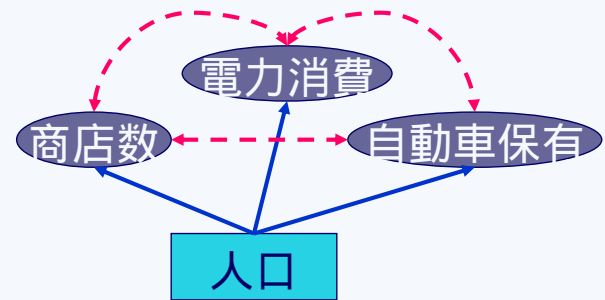
散布図行列



相関

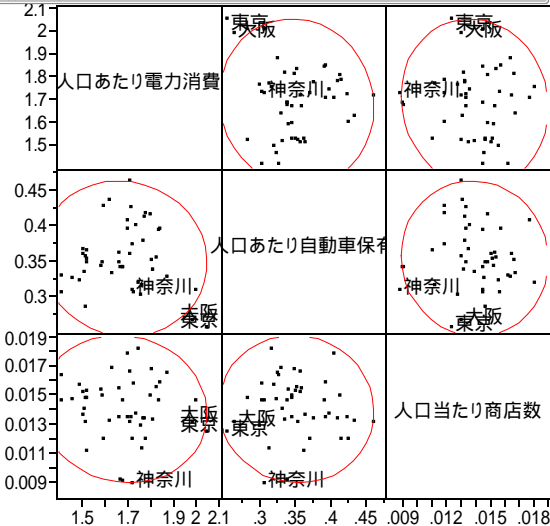
	人口	商店数	電力消費量	自動車保有
人口	1.0000	0.9756	0.9919	0.9690
商店数	0.9756	1.0000	0.9841	0.9265
電力消費量	0.9919	0.9841	1.0000	0.9407
自動車保有	0.9690	0.9265	0.9407	1.0000

それぞれの変量の相関係数が0.90以上であるが、因果関係があると思われるか？



- 人口と商店数、電力消費、自動車保有の関数が分からないなら、各変量を人口で割っておこう。
- 人口で割った新たな変量について散布状況や相関係数を調べてみよう。

散布図行列



相関

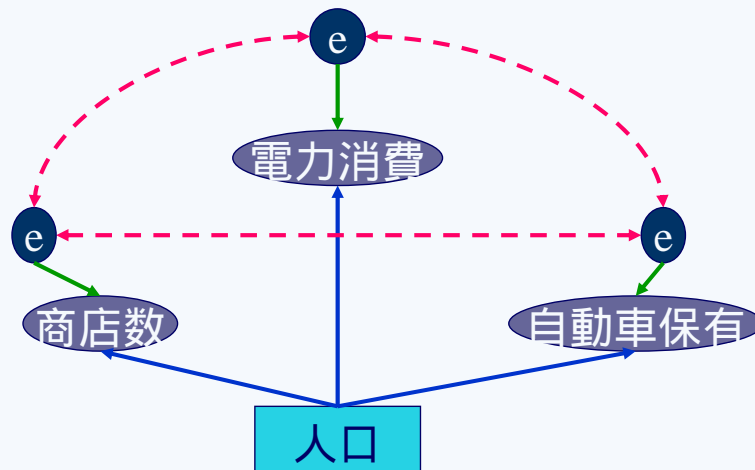
	人口あたり電力消費	人口あたり自動車保有	人口あたり商店数
人口あたり電力消費	1.0000	-0.0614	-0.0703
人口あたり自動車保有	-0.0614	1.0000	-0.0321
人口あたり商店数	-0.0703	-0.0321	1.0000

強い正相関の関係がなくなりました。

人口の影響を取り除いてしまえば、3変量間には実質的な関係はないことが分かり、これは実感に合う。

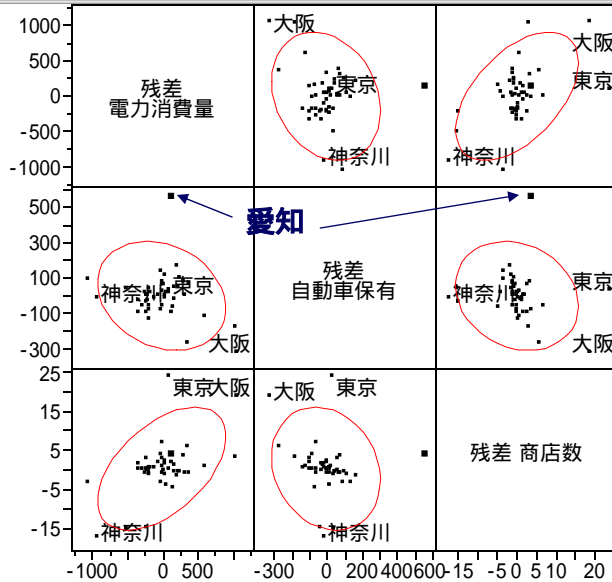
● 回帰残差の利用

- 人口の影響を取り除くためのスマートな方法
- 回帰分析の残差を利用する
- 回帰残差は、目的変数について説明変数で説明がつかない部分である
- 残差間の相関を調べると実質的な相関を調べることができる



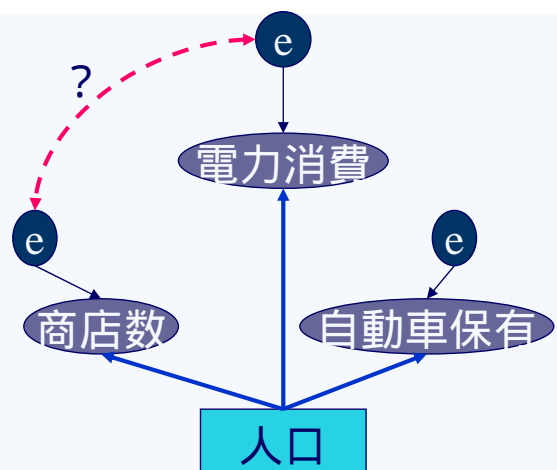
- 人口と各変数との関係は、やや曲線的にも見えるので2次式を当てはめる
- その残差をプロットすると以下のようなになる

散布図行列



相関

	残差 電力消費量	残差 自動車保有	残差 商店数
残差 電力消費量	1.0000	-0.2263	0.5235
残差 自動車保有	-0.2263	1.0000	-0.1913
残差 商店数	0.5235	-0.1913	1.0000





イントロの終わり  
これからが始まりです。



- はじめに
- IC工程の例
- 相関・回帰関係、層別と擬似相関
- グラフィカルモデリング(GM)とは

&lt;ファイル:IC工程-C.CSV&gt;

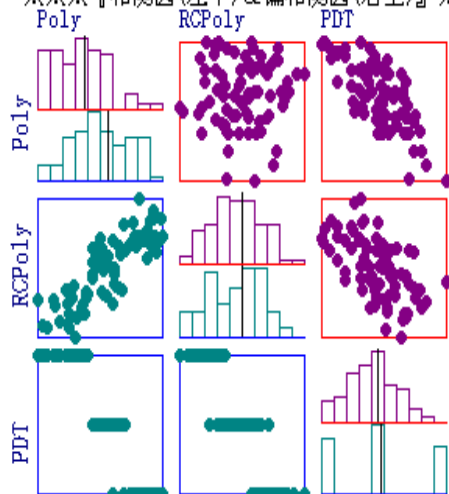
- 半導体 (IC) 製造工程では、IC製造のために、温度条件など数多く (250以上) の工程設定条件がある。
- ICの性能を評価する電気特性も100以上ある。
- 工程の品質管理では、平均位置や工程ばらつきなどには大いに気を配る必要がある
- いま、2つの電気特性が測られていて、それはICの酸化温度 (PDT) に大きく影響する

## 1.3.1 IC工程の相関関係

20/136

## ● 散布状態を調べた

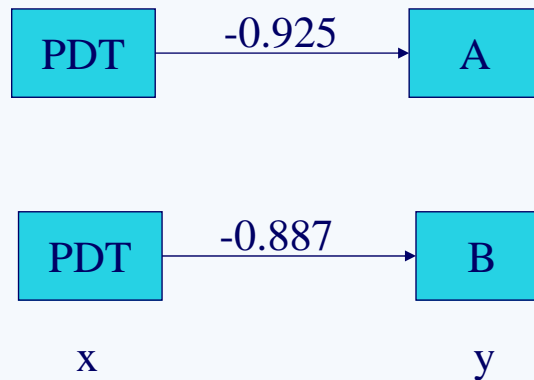
※※※『相関図(左下)&amp;偏相関図(右上)』※※※ 変数の数= 3



	<i>Poly</i>	<i>RCpoly</i>	<i>PDT</i>
<i>poly</i>	1.000	0.826	-0.925
<i>RCpoly</i>	0.826	1.000	-0.887
<i>PDT</i>	-0.925	-0.887	1.000

- ・PDTは3つの値しか取らない
- ・管理された状態でのデータである

### ● PDTと電気特性A&Bの関係を調べてみよう

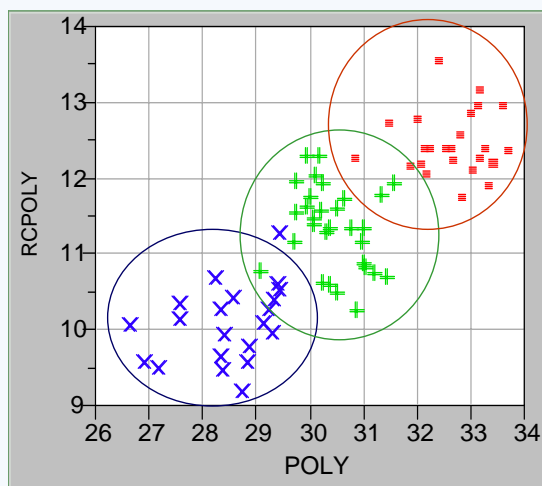


PDTによって、電気特性A,Bは制御 (= 管理可能)

- 工業データはこのように原因と結果とがはっきりしている  
つまり、因果関係が知見により簡単に判断できる
- 社会調査のデータでは因果関係があいまいな場合が多々ある  
つまり、分析者によって都合よく用いられるかも？

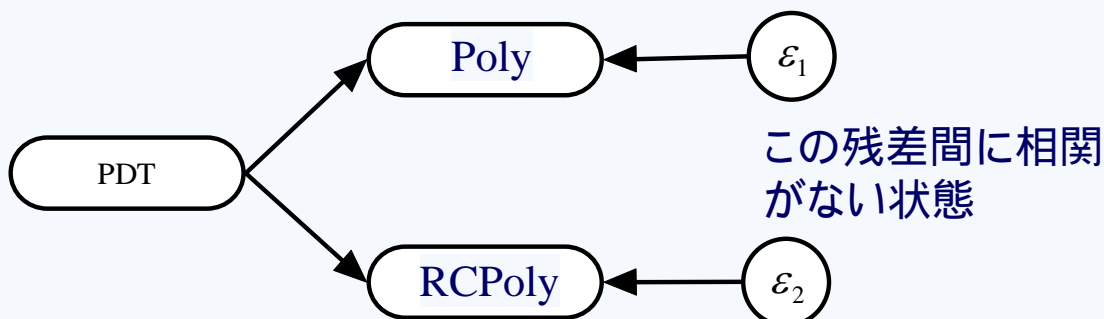
### 1.3.3 IC工程の層別分析

- 世の中の多くの事柄は、原因で層別 (= 分類:層別には強い意図がある) により解決可能
- 工業データは比較的層別が簡単



それぞれの相関の状態は、非常に弱いことが見てとれる

- 原因となる第3の変量で2つの結果変量について、層別散布図を描いたとき相関が認められない状態を擬似相関という



A,Bの相関の原因は、PDTによるものであることが判明した状態である

- 偏相関情報について

2つの残差を加えた相関係数行列

	<i>Poly</i>	<i>RCPoly</i>	<i>PDT</i>	<i>R_P</i>	<i>R_RC</i>
<i>P</i>	1.000				
<i>RC</i>	0.826	1.000			
<i>PDT</i>	-0.925	-0.887	1.000		
<i>R_P</i>	0.380	0.013	0.000	1.000	
<i>R_RC</i>	0.011	0.462	0.000	0.028	1.000

この相関がゼロに近い値であることに注目!!

ここは計算上ゼロではないが、  
実質的な意味がないため、はじめからゼロとして分析する

<ファイル: 3変数-R.DAT>

$$R = \begin{pmatrix} 1.00 & & \\ 0.80 & 1.00 & \\ 0.70 & 0.56 & 1.00 \end{pmatrix} \Rightarrow P = \begin{pmatrix} - & & \\ 0.69 & - & \\ 0.51 & 0 & - \end{pmatrix}$$

$$r_{12 \cdot 3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1-r_{13}^2)(1-r_{23}^2)}} = \frac{0.80 - 0.70 \times 0.56}{\sqrt{(1-0.70^2)(1-0.56^2)}} = 0.69$$

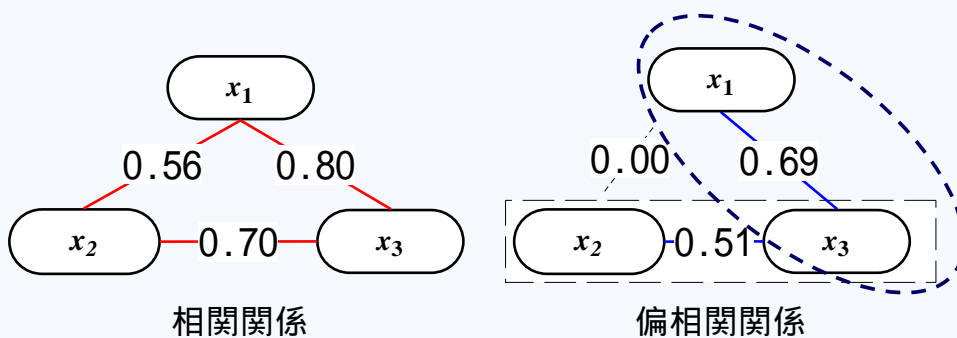
$$r_{13 \cdot 2} = \frac{r_{13} - r_{12}r_{23}}{\sqrt{(1-r_{12}^2)(1-r_{23}^2)}} = \frac{0.70 - 0.80 \times 0.56}{\sqrt{(1-0.80^2)(1-0.56^2)}} = 0.51$$

$$r_{23 \cdot 1} = \frac{r_{23} - r_{12}r_{13}}{\sqrt{(1-r_{12}^2)(1-r_{13}^2)}} = \frac{0.56 - 0.80 \times 0.70}{\sqrt{(1-0.80^2)(1-0.70^2)}} = 0$$

層別と条件付独立性 (偏相関がない)

グラフィカルモデリングの基本的概念は、「条件付独立性」である。  
 3つの変数  $x_1, x_2, x_3$  があるとき、  
 $x_3$  を制御できる下で  $x_1, x_2$  が無関係(独立)であるかどうか  
 $x_3$  で層別して  $x_1, x_2$  が独立であるかどうか  
 を検討することである。

変数間の関係に条件付独立性が発見できれば、  
問題の構造を分解して考えることができる。



- 偏相関係数を理解するのに(重)回帰モデルからの残差を考えることは大切
- 実際の計算はもっと簡単にできる

4変数以上の偏相関係数はRの逆行列よりPを求めることで計算できる.

相関係数行列  $R = (r_{ij})$  とする

その逆行列を  $R^{-1} = (r^{ij})$  (逆行列では添え字を上付きにする) とする

$x_i$  と  $x_j$  以外の全ての変数を与えたとき

$$x_i \text{ と } x_j \text{ の偏相関係数は, } r_{ij \bullet \text{rest}} = -\frac{r^{ij}}{\sqrt{r^{ii} r^{jj}}} \text{ となる}$$

(rest は残りの変数という意味) .

逆行列の対応する要素を2つの対角要素の平方根で割って基準化符号を反転させる .

## 4変数の偏相関係数の計算-1

<ファイル:4変数-R.DAT>

$$R = \begin{pmatrix} 1.00 & & & \\ r_{12} & 1.00 & & \\ r_{13} & r_{23} & 1.00 & \\ r_{14} & r_{24} & r_{34} & 1.00 \end{pmatrix} = \begin{pmatrix} 1.00 & & & \\ 0.80 & 1.00 & & \\ 0.80 & 0.64 & 1.00 & \\ 0.80 & 0.64 & 0.64 & 1.00 \end{pmatrix}$$

逆行列を求める

$$R^{-1} = \begin{pmatrix} 1.00 & & & \\ r^{12} & 1.00 & & \\ r^{13} & r^{23} & 1.00 & \\ r^{14} & r^{24} & r^{34} & 1.00 \end{pmatrix} = \begin{pmatrix} 6.333 & & & \\ -2.222 & 2.778 & & \\ -2.222 & 0 & 2.778 & \\ -2.222 & 0 & 0 & 2.778 \end{pmatrix}$$

基準化する

$$r_{12 \bullet 34} = r_{12 \bullet \text{rest}} = -\frac{r^{12}}{\sqrt{r^{11} r^{22}}} = \frac{-2.222}{\sqrt{6.333 \times 2.778}} = 0.53$$

$$P = \begin{pmatrix} - & & & \\ r_{12 \bullet rest} & - & & \\ r_{13 \bullet rest} & r_{23 \bullet rest} & - & \\ r_{14 \bullet rest} & r_{24 \bullet rest} & r_{34 \bullet rest} & - \end{pmatrix} = \begin{pmatrix} - & & & \\ 0.53 & - & & \\ 0.53 & 0 & - & \\ 0.53 & 0 & 0 & - \end{pmatrix}$$

### ● 偏相関係数が求まらない

- 変数間に何かしらの関係式があるとき相関係数行列の逆行列が求まらない
- すべての変数が線形独立でないならば相関係数の逆行列が求まらない
- 多重共線性がある

### ● 偏相関係数行列から相関係数行列を求める方法

- 先の手順の逆を行うとよい

$$P = \begin{pmatrix} - & & & \\ r_{12 \bullet rest} & - & & \\ r_{13 \bullet rest} & r_{23 \bullet rest} & - & \\ r_{14 \bullet rest} & r_{24 \bullet rest} & r_{34 \bullet rest} & - \end{pmatrix} = \begin{pmatrix} - & & & \\ 0.53 & - & & \\ 0.53 & 0 & - & \\ 0.53 & 0 & 0 & - \end{pmatrix}$$

対角要素(-)は-1であることに注意して $-P$ の逆行列を計算する

$$(-P)^{-1} = \begin{pmatrix} p^{11} & & & \\ p^{12} & p^{22} & & \\ p^{13} & p^{23} & p^{33} & \\ p^{14} & p^{24} & p^{34} & p^{44} \end{pmatrix} = \begin{pmatrix} 6.357 & & & \\ 3.369 & 2.786 & & \\ 3.369 & 1.786 & 2.786 & \\ 3.369 & 1.786 & 1.786 & 2.786 \end{pmatrix}$$

基準化する

$$r_{12} = \frac{p^{12}}{\sqrt{p^{11} p^{22}}} = \frac{3.369}{\sqrt{6.357 \times 2.786}} = 0.80$$

$$R = \begin{pmatrix} 1.00 & & & \\ r_{12} & 1.00 & & \\ r_{13} & r_{23} & 1.00 & \\ r_{14} & r_{24} & r_{34} & 1.00 \end{pmatrix} = \begin{pmatrix} 1.00 & & & \\ 0.80 & 1.00 & & \\ 0.80 & 0.64 & 1.00 & \\ 0.80 & 0.64 & 0.64 & 1.00 \end{pmatrix}$$

- 相関係数行列と偏相関係数行列には上記のような関係があり,これは母集団,標本共に成立する関係である
- この関係を使えば,変数ごとに回帰モデルの残差を計算しなくてもよい

- **擬似相関: Spurious correlation**

本来 $x_1$ と $x_2$ の間に相関関係は存在しないにも関わらず,別の原因変数 $x_3$ のために, $x_1$ と $x_2$ も変化して,その結果,相関関係があるようにみえること

- **偏相関: partial correlation**

擬似相関が存在するかも知れない状況で, $x_1$ と $x_2$ の実質的な相関を考えること.これは $x_3$ の影響を取り除いた後の $x_1$ と $x_2$ の相関

$x_3$ の値を固定したときの $x_1$ と $x_2$ の相関

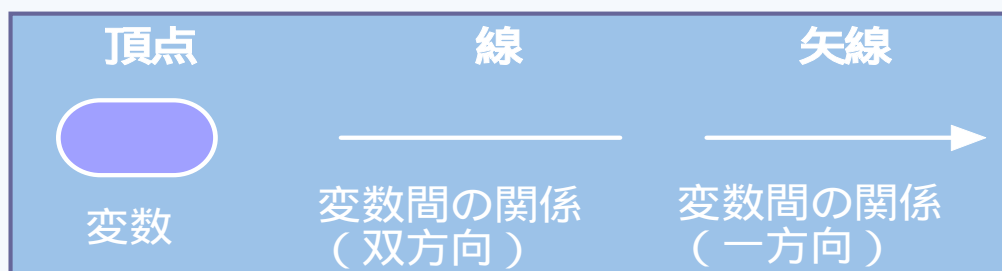
- **偏相関係数: partial correlation coefficient**

偏相関の大きさを定量的に表す特性



グラフィカルモデル  
 調査研究データを解析するための強力なモデリング・ツールの1つで、  
 統計量と知見からモデルを探索的に構築できる

解析のアウトプット  
 統計量と独立グラフ (点と線) で表現



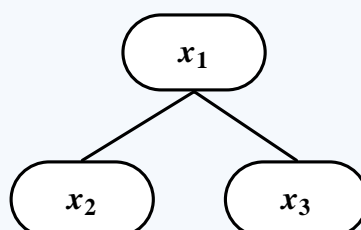
## 1.4.1 独立グラフの作り方-1

<ファイル: 3変数-R.DAT>

$$\Pi = \begin{pmatrix} 1.00 & & \\ \rho_{12} & 1.00 & \\ \rho_{13} & \rho_{23} & 1.00 \end{pmatrix} = \begin{pmatrix} 1.00 & & \\ 0.80 & 1.00 & \\ 0.70 & 0.56 & 1.00 \end{pmatrix}$$

$$\Lambda = \begin{pmatrix} - & & \\ \rho_{12 \cdot 3} & - & \\ \rho_{13 \cdot 2} & \rho_{23 \cdot 1} & - \end{pmatrix} = \begin{pmatrix} - & & \\ 0.69 & - & \\ 0.51 & 0 & - \end{pmatrix}$$

● 母偏相関係数行列の様子をグラフで表すと



- 2次元正規分布では「無相関」 = 「独立」
- 多次元正規分布でも「無相関」 = 「独立」
- 「偏相関係数 = 0」 = 「条件付独立」

記号による表現

$$\rho_{23} = 0 \quad x_2 \perp x_3$$

$$\rho_{23 \cdot 1} = 0 \quad x_2 \perp x_3 \mid x_1$$




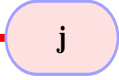
相関の立場

独立性の立場

- 量的変数のグラフィカルモデリングでは、
- 多変量正規分布を仮定した下で、
- 条件付独立の関係をグラフを用いて表現する

2つの変数( $x_i, x_j$ ) ( $i \neq j$ )に対して残りの変数を与えたとき

$\rho_{ij \cdot rest} = 0$     条件付独立: 線が切れる

$\rho_{ij \cdot rest} \neq 0$      条件付独立: 線が切れて  
でない いない

統計量

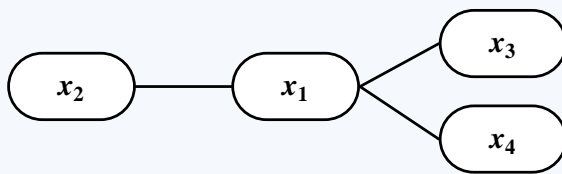
独立グラフ(Independence graph)

## ● 4変数の場合の例

&lt;ファイル:4変数-R.DAT&gt;

$$\Pi = \begin{pmatrix} 1.00 & & & \\ \rho_{12} & 1.00 & & \\ \rho_{13} & \rho_{23} & 1.00 & \\ \rho_{14} & \rho_{24} & \rho_{34} & 1.00 \end{pmatrix} = \begin{pmatrix} 1.00 & & & \\ 0.80 & 1.00 & & \\ 0.80 & 0.64 & 1.00 & \\ 0.80 & 0.64 & 0.64 & 1.00 \end{pmatrix}$$

$$\Lambda = \begin{pmatrix} - & & & \\ \rho_{12 \bullet rest} & - & & \\ \rho_{13 \bullet rest} & \rho_{23 \bullet rest} & - & \\ \rho_{14 \bullet rest} & \rho_{24 \bullet rest} & \rho_{34 \bullet rest} & - \end{pmatrix} = \begin{pmatrix} - & & & \\ 0.53 & - & & \\ 0.53 & 0 & - & \\ 0.53 & 0 & 0 & - \end{pmatrix}$$



$$x_2 \perp x_3 \mid (x_1, x_4) \Leftrightarrow \rho_{23 \bullet rest} = 0$$

$$x_2 \perp x_4 \mid (x_1, x_3) \Leftrightarrow \rho_{24 \bullet rest} = 0$$

$$x_3 \perp x_4 \mid (x_1, x_2) \Leftrightarrow \rho_{34 \bullet rest} = 0$$

## 独立グラフの見方-1

- 線で直接つながっていない変数対は, 残りの変数を与えた下で条件付独立
- 線で直接つながっていない変数対について, その変数対を間接的につなぐ経路に含まれる変数のみを与えた下で条件付独立
- 変数間の関係
  1. 直接関係
  2. 変数を与えると関係が断ち切れる (条件付独立)

線を鉄道の路線, 変数を駅と考える

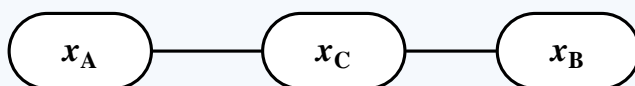
・AからBへ行くとき, どの駅もおることなく線路で結ばれているならば, AとBとは直接的な関連がある



### 独立グラフ: 直接関係

線を鉄道の路線, 変数を駅と考える

・AからBへ行くとき, 途中にCだけをとおらなければならない  
Cを与えた時にAとBとは条件付独立な関連がある



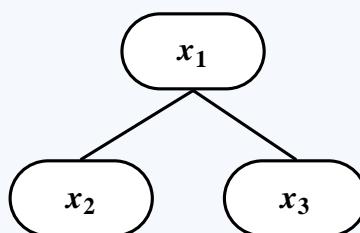
### 独立グラフ

$$x_A \perp x_B \mid x_C$$

$$\rho_{AB \cdot C} = 0 \Rightarrow$$

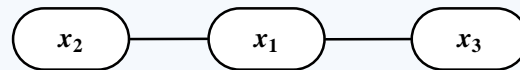
$$\rho_{AB} = \rho_{AC} \times \rho_{CB}$$

相関の乗法則

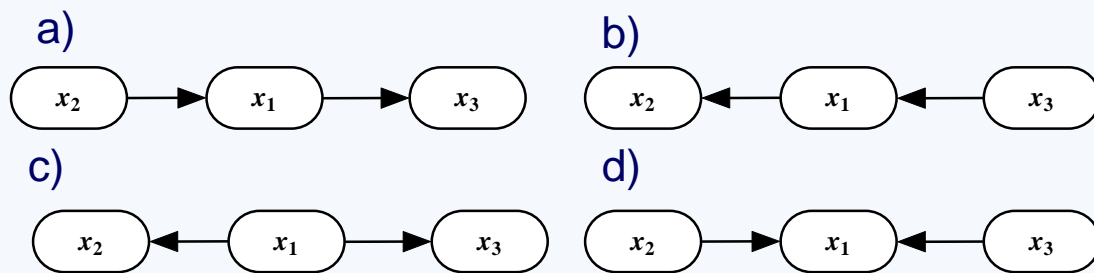


- $x_1$ を与えた下で,  $x_2, x_3$ は条件付独立
- $x_1$ の値を固定すると $x_2, x_3$ は独立
- $x_3$ の値をコントロールする目的で $x_2$ を変化させても,  $x_1$ が固定されていると目的は達成できない
- $x_3$ の制御には $x_1$ を操作するのがよい

- 無向グラフ(undirected graph)  
独立グラフで矢印のない(線のみで構成されている)グラフ
- 有向グラフ(directed graph) 因果グラフ(causality graph)  
変数間の因果関係を考慮して矢印をつけたもの

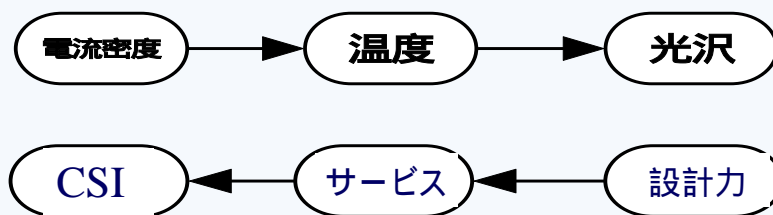


適当に矢印をつける



変数の具体的意味を考えなければ, a), b)は同じ

### パターンa), b)



- 電流密度を増やすと発熱のためにメッキ液の温度が上がる
- メッキ液の温度が上がるためにメッキ面の光沢が悪くなる
- この工程では, メッキ液を冷却装置に循環して温度を一定に保つことが出来れば, 電流密度と光沢とは無関係になる(条件付独立)

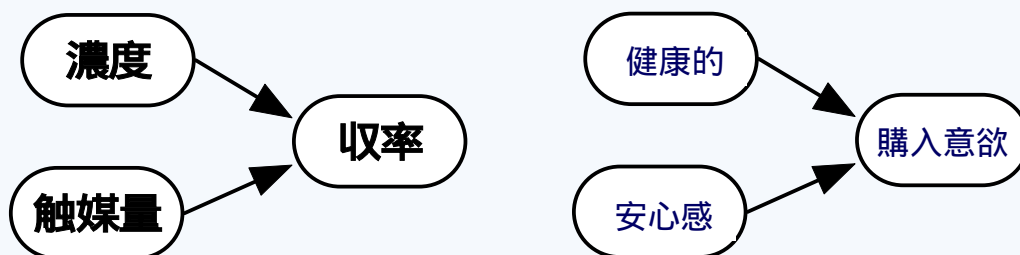
(電流密度  $\perp$  光沢 | 温度)



- 処理温度が製品硬度と製品強度の両方に影響を与えている
- 処理温度を一定に制御できれば、硬度と強度は無関係になる(条件付独立)

(硬度 ⊥ 強度 | 温度)

V字合流

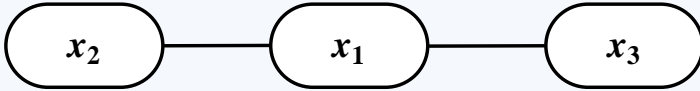


- 実験計画法により原料濃度と触媒量は独立になるようにデータが取られた
- 原料濃度が高くなると収率は高くなる
- 触媒量が高くなると収率は高くなる

- 偏相関係数行列のみの情報からでは, 方向が分からないから, a) ~ c) のどのタイプが分からない
- d)のタイプは得られない
- 因果グラフにおける条件付独立性
  - a) ~ c)では, 独立グラフの場合と同様に $x_1$ を与えたときに $x_2, x_3$ が条件付独立であると考えることができる
  - d)はそうはいかない
    - 結果を与えた下で2つの原因が条件付独立とは本末転倒

## 1.4.3 共分散選択-1

&lt;ファイル:3変数-R.DAT&gt;



$$\Pi = \begin{pmatrix} 1.00 & & \\ 0.80 & 1.00 & \\ 0.70 & 0.56 & 1.00 \end{pmatrix} \quad \Lambda = \begin{pmatrix} - & & \\ 0.69 & - & \\ 0.51 & 0 & - \end{pmatrix}$$

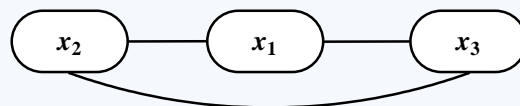
- d)は, 2つの原因は独立である.
- 2つの変数の相関係数は0.56でゼロではない
- 相関行列から偏相関行列を求めて作成した独立グラフに基づくとd)の因果グラフは対象外

● d)の結果のとき母相関係数行列を計算すると

$$\Pi = \begin{pmatrix} 1.00 & & \\ \rho_{12} & 1.00 & \\ \rho_{13} & \rho_{23} & 1.00 \end{pmatrix} = \begin{pmatrix} 1.00 & & \\ a & 1.00 & \\ b & 0 & 1.00 \end{pmatrix}$$

$$\rightarrow \Lambda = \begin{pmatrix} - & & \\ \rho_{12 \cdot 3} & - & \\ \rho_{13 \cdot 2} & \rho_{23 \cdot 1} & - \end{pmatrix} = \begin{pmatrix} - & & \\ a/\sqrt{1-b^2} & - & \\ b/\sqrt{1-a^2} & -ab/\sqrt{(1-a^2)(1-b^2)} & - \end{pmatrix}$$

$a \neq 0, b \neq 0$ より偏相関係数は全てゼロでない



## Dempster の定理-1

<ファイル:3変数2-R.DAT>

● モデリング

データ解析の立場から、データより計算した標本相関係数行列に基づいてモデルを構築する  
つまり、データに基づいてモデルを推測・探索し、その妥当性を検証すること(アナリストの責任)

$$R = \begin{pmatrix} 1.00 & & \\ r_{12} & 1.00 & \\ r_{13} & r_{23} & 1.00 \end{pmatrix} = \begin{pmatrix} 1.00 & & \\ 0.80 & 1.00 & \\ 0.70 & 0.60 & 1.00 \end{pmatrix}$$

$$r_{23} = 0.60 \square 0.80 \times 0.70 = r_{12} \times r_{13}$$

$$\Rightarrow \rho_{23} = \rho_{12} \times \rho_{13} \quad , \quad \rho_{23 \cdot 1} = 0 \quad \rightarrow \text{を仮定する}$$



- 共分散選択 covariance selection  
偏相関係数のいくつかを0と置いた相関構造モデルを採用するアプローチ

$$R = \begin{pmatrix} 1.00 & & & \\ r_{12} & 1.00 & & \\ r_{13} & r_{23} & 1.00 & \\ & & & \ddots \end{pmatrix} = \begin{pmatrix} 1.00 & & & \\ 0.80 & 1.00 & & \\ 0.70 & 0.60 & 1.00 & \\ & & & \ddots \end{pmatrix}$$

$$P = \begin{pmatrix} - & & & \\ r_{12 \cdot 3} & - & & \\ r_{13 \cdot 2} & r_{23 \cdot 1} & - & \\ & & & \ddots \end{pmatrix} = \begin{pmatrix} - & & & \\ 0.67 & - & & \\ 0.46 & 0.09 & - & \\ & & & \ddots \end{pmatrix}$$

$r_{23 \cdot 1} = 0.09$ はゼロに近いので母偏相関係数 $\rho_{23 \cdot 1} = 0$ を相関構造モデルとして採用する

- 問題は, ある母偏相関係数をゼロとした時に他の母数の推定値がどのような値となるか
- Dempsterの定理
  - 1)  $x_2, x_3$   $x_3$ の母偏相関係数は制約によってゼロとする
  - 2) 制約のない $(x_1, x_2), (x_1, x_3)$ の母相関係数の推定値は, それぞれ元の標本相関係数の値のままとする

$\rho_{23\cdot1} = 0$ を採用すると

$$P_0 = \begin{pmatrix} - & & \\ r_{12\cdot3} & - & \\ r_{13\cdot2} & r_{23\cdot1} & - \end{pmatrix} = \begin{pmatrix} - & & \\ a & - & \\ b & 0 & - \end{pmatrix} \quad \text{とし定理2)より}$$

$$R_0 = \begin{pmatrix} 1.00 & & \\ r_{12} & 1.00 & \\ r_{13} & r_{23} & 1.00 \end{pmatrix} = \begin{pmatrix} 1.00 & & \\ 0.80 & 1.00 & \\ 0.70 & c & 1.00 \end{pmatrix} \quad \text{逆行列を計算する}$$

$$R_0^{-1} = \begin{pmatrix} r^{11} & & \\ r^{12} & r^{22} & \\ r^{13} & r^{23} & r^{33} \end{pmatrix} = \frac{1}{|R_0|} \begin{pmatrix} 1.00 - c^2 & & \\ 0.70c - 0.80 & 0.51 & \\ 0.80c - 0.70 & 0.56 - c & 0.36 \end{pmatrix} \quad \begin{array}{l} 0.56 - c = 0 \text{より} \\ c = 0.56 \end{array}$$

$$|R_0| = -0.13 + 1.12c - c^2$$

$$R_0 = \begin{pmatrix} 1.00 & & \\ r_{12} & 1.00 & \\ r_{13} & r_{23} & 1.00 \end{pmatrix} = \begin{pmatrix} 1.00 & & \\ 0.80 & 1.00 & \\ 0.70 & 0.56 & 1.00 \end{pmatrix}$$

● 0.60 0.56でわずかに0.04しか変化していない

$$P_0 = \begin{pmatrix} - & & \\ r_{12\cdot3} & - & \\ r_{13\cdot2} & r_{23\cdot1} & - \end{pmatrix} = \begin{pmatrix} - & & \\ 0.69 & - & \\ 0.51 & 0 & - \end{pmatrix}$$

偏相関係数の1つをゼロとおくのは簡単に計算できるが、複数の偏相関係数をゼロとおき、Dempsterの定理を満たすには複雑な反復計算が必要である。



Wermuth Scheidtのアルゴリズムで解く

- 共分散選択: 幾つかの偏相関係数をゼロにおいた相関構造モデルにより, 観測された相関係数行列を近似するアプローチ
- どの偏相関係数をゼロとするかはデータからの判断による
- ちょうど回帰分析の変数選択問題と同じ  
回帰の変数選択:  $2^p$ とおり  
共分散選択:  $2^{p(p-1)/2}$ とおりもある!!
- 効率よく, よいモデルに到達する方法として, 変数減少法を使い, 相関係数の残差をみながら変数選択(追加と削除)を行うアプローチをとる

<ファイル:IC工程-C.CSV> デモ:表1.8 & 図1.13

#### 1.4.4 モデルの適合度-1

- 縮約モデル Reduced Model(RM)  
共分散選択により, 偏相関係数のいくつかをゼロとおいたモデルを縮約モデルと呼ぶ
- RMが適切であるかどうかを定量的に評価する必要がある
- 逸脱度の導入 Deviance  
フルモデルFMの下での相関係数行列をRと表し, 共分散選択で採択された縮約モデルRMの母相関係数の推定値  $\hat{\Pi}$  との適合具合を逸脱度,

$$dev(RM) = n \log \frac{|\hat{\Pi}|}{|R|} \quad |\cdot| \text{は行列式}$$

帰無仮説の下で

$$dev(RM) = n \log \frac{|\hat{\Pi}|}{|R|} \square \chi^2(df)$$

自由度 $df$ はゼロとした偏相関係数の数

$$|\hat{\Pi}| \geq |R|$$

モデルが標本相関行列によく適合していれば  
 $|\hat{\Pi}|, |R|$ がほぼ等しくなるであろう



多次元構造

モデリング

偏相関の縮約

偏相関 有りや無しや

独立グラフ

線 有りや無しや

母相関の推定

適合度評価



### ● 逸脱度は2つある

- 縮約モデル $RM$ とフルモデル $FM$ との違いを測っているもの  
制約をつけることでどれだけデータから乖離しているか

$$dev(RM) = n \log \frac{|\hat{\Pi}|}{|R|} \square \chi^2(df)$$

- 共分散選択の各ステップ毎の縮約モデルの違いを測る逸脱度も考えることができる。

$$dev(RM_{i-1}, RM_i)$$

$$= dev(RM_i) - dev(RM_{i-1}) \square \chi^2(df = 1)$$

考え方は、検定の帰無仮説を採択する立場にたつので基準となる値をどこに置くか明快ではない。2つの平均値の差の検定Stepでの等分散の検定の目的と同じ立場にある

$$dev\{RM(i)\} = n \log \frac{|\hat{\Pi}_{(i)}|}{|\mathbf{R}|}$$

$$dev\{RM(i), RM(i-1)\} = dev\{RM(i)\} - dev\{RM(i-1)\}$$

*FM* に対しては  $p$  値 = 0.5 を ,  
*RM*<sub>(i-1)</sub> に対しては  $p$  値 = 0.2 ~ 0.3 を判断基準に



全体モデル: フルモデルからの乖離の量  
 縮約モデル: 1つ前の縮約モデルからの乖離の量  
 G-GMは変数減少法によりモデル探索  
 統計的に得られた相関情報から考えて推測論的に  
 許容可能な縮約の段階を客観的に評価する

#### 1.4.4 モデルの適合度-4

##### その他の適合度指標: 標本数が多い場合

$$GFI = 1 - \frac{tr \left[ \left\{ \hat{\Pi}_{(i)}^{-1} (\mathbf{R} - \hat{\Pi}_{(i)}) \right\}^2 \right]}{tr \left[ \left\{ \hat{\Pi}_{(i)}^{-1} \mathbf{R} \right\}^2 \right]}$$

0.95を目安に

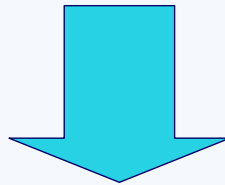
$$AGFI = 1 - \frac{p(p+1)}{2df} (1 - GFI)$$

$$NFI = 1 - dev\{RM(i)\} / dev(NM)$$

$$SRMR = \sqrt{\frac{2}{p(p+1)} \sum_{i \leq j} (r_{ij} - \hat{\rho}_{ij})^2}$$

0.05を目安に

相関係数行列の当てはまりを問題にしているので、  
いくら着目した要素の偏相関係数が小さくても、  
それを0とおくと逸脱度などの指標が悪くなることもある。



元の相関とモデルでの相関の差  
すなわち残差の最大値に着目して  
共分散選択(線の切断と接続)を対話的に  
行うとよい

第1部終了  
お疲れ様でした

次からはハンズオンです。



&lt;ファイル:ES-C.DAT&gt;

## .従業員満足度に関する調査データの分析

→ 上司に対する評価と満足感

- 1.データの概要
- 2.分析の戦略とステップ概要
- 3.事前の分析
- 4.無向独立グラフによる変数間構造の探索
- 5.無向グラフからの因果想定
- 6.有向グラフによる確認
- 7.他の古典的手法との関連について

## 1.データの概要

62/136

## 事例の概要:

ある企業では、無作為に抽出した101人のテマリーダーに対して、ES(従業員満足度)調査を行った

本解析では以下の上司との関わりかたに関する5項目を使う

変数名	質問内容
指示	あなたの上司は、あなたの仕事ぶりについて指示を与えていると思いますか
進捗	あなたの上司は、あなたの仕事の進捗度を知っていますか
内容	あなたの上司は、あなたの仕事内容を知っていますか
会話	あなたの上司は、あなたと気楽に話ができると思いますか
気遣	あなたの上司は、あなたの職場の雰囲気気に気を使っていると思いますか



基本統計量

解析対象個体数 = 101

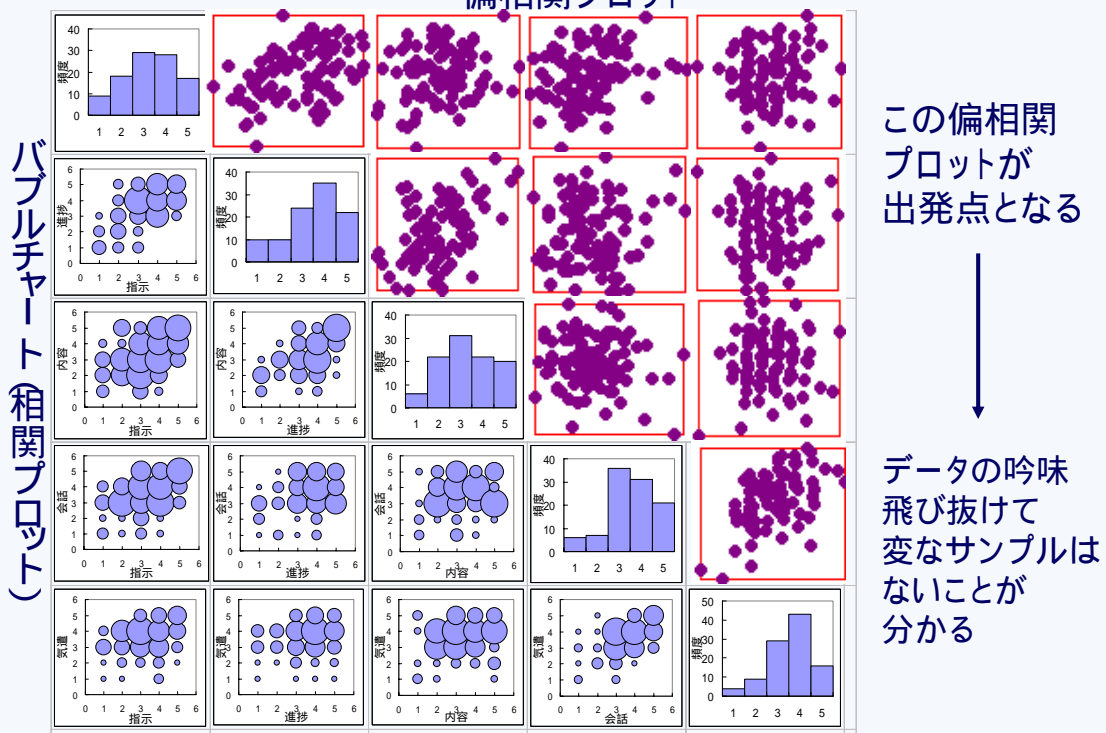
変数名	平均値	標準偏差	最大値	最小値	変動係数	歪度	尖度
1指示	3.2574	1.1971	5.0000	1.0000	0.3675	-0.2283	-0.7959
2進捗	3.4851	1.2216	5.0000	1.0000	0.3505	-0.6029	-0.4710
3内容	3.2970	1.1962	5.0000	1.0000	0.3628	-0.0605	-0.9579
4会話	3.5347	1.0823	5.0000	1.0000	0.3062	-0.4773	-0.0915
5気遣	3.5743	0.9934	5.0000	1.0000	0.2779	-0.6155	0.1959

相関係数行列

解析対象個体数 = 101

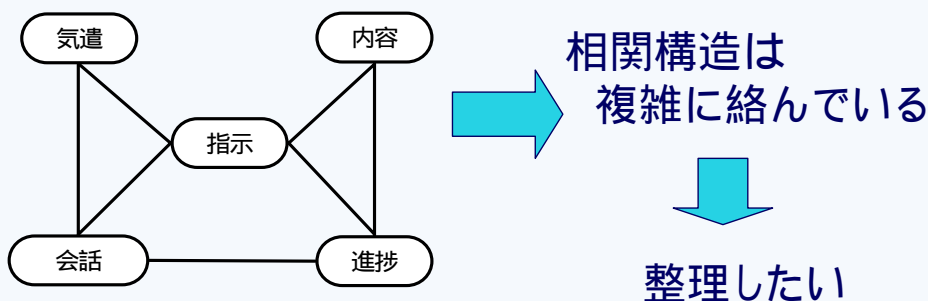
No.	指示	進捗	内容	会話	気遣
指示	1.000	0.563	0.421	0.425	0.270
進捗	0.563	1.000	0.571	0.293	0.180
内容	0.421	0.571	1.000	0.170	0.158
会話	0.425	0.293	0.170	1.000	0.539
気遣	0.270	0.180	0.158	0.539	1.000

偏相関プロット





1.相関プロットと偏相関プロットは見え方が違う  
 相関係数の検定では,  $|r| \geq 0.2$  で5%有意



相関には, 他の変数の影響も含まれているから  
 直接効果と間接効果を分離したい

**鍵を握るのは, どうやら指示の変数であることが分かる**

3.無向独立グラフによる

変数間構造の探索 -

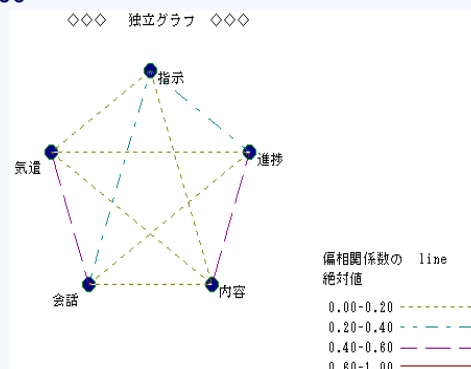
ここでは, すべての変数の役割は同じであると考えて統計量に沿って無向独立グラフを作成する.

左下三角が偏相関係数の推定値, 右上三角が相関係数の差

GF1= 1.000 AGFI= 1.000 NFI= 1.000 SRMR= 0.000

偏相関係数の推定

	指示	進捗	内容	会話	気遣
1指示	***				
2進捗	0.380	***			
3内容	0.151	0.448	***		
4会話	0.273	0.091	-0.078	***	
5気遣	0.034	-0.032	0.072	0.489	***



**太字の相関係数が0とおく候補となる  
 順次0とおきモデリングする**

変数間構造の探索 -

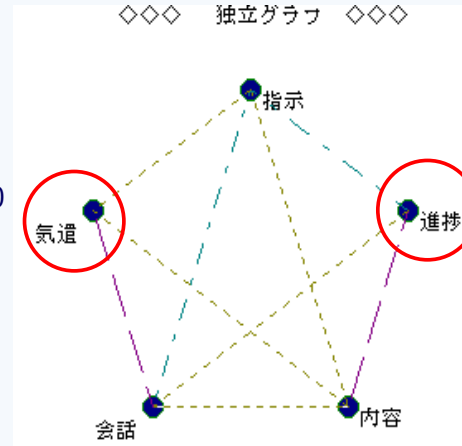
左下三角が偏相関係数の推定値，右上三角が相関係数の差

n= 101 逸脱度=0.103(df=1) p=0.7478 逸脱度=0.103(df=1) p=0.7478

GFI= 1.000 AGFI= 0.994 NFI= 0.999 SRMR= 0.005

偏相関係数の推定

	指示	進捗	内容	会話	気遣
1指示	***				
2進捗	0.379	***			-0.020
3内容	<b>0.152</b>	0.447	***		残差
4会話	0.280	<b>0.075</b>	-0.071	***	
5気遣	0.022	<u>0.000</u>	0.058	0.487	***



進捗 - 気遣の偏相関係数が0  
進捗 - 気遣の線が切断



グラフと共分散選択が連動

変数間構造の探索 -

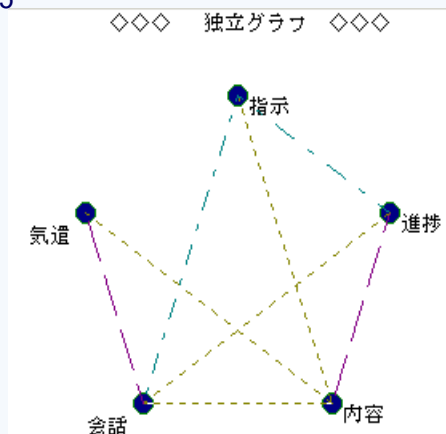
左下三角が偏相関係数の推定値，右上三角が相関係数の差

n= 101 逸脱度=0.159(df=2) p=0.9236 逸脱度=0.056(df=1) p=0.8136

GFI= 0.999 AGFI= 0.995 NFI= 0.999 SRMR= 0.005

偏相関係数の推定

	指示	進捗	内容	会話	気遣
1指示	***				0.016
2進捗	0.379	***			-0.014
3内容	<b>0.154</b>	0.447	***		
4会話	0.289	<b>0.075</b>	-0.075	***	
5気遣	<u>0.000</u>	<u>0.000</u>	0.066	0.493	***



今度は，指示と気遣の線が切断された  
このようにして逐次的に線を切断していく

変数間構造の探索 -

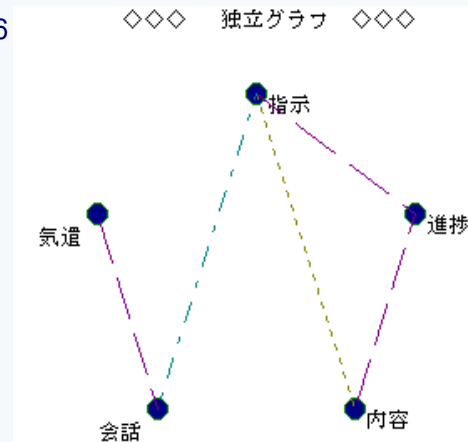
左下三角が偏相関係数の推定値，右上三角が相関係数の差

n=101 逸脱度=1.575(df=5) p=0.9042 逸脱度=0.526(df=1) p=0.4684

GFI= 0.994 AGFI= 0.982 NFI= 0.989 SRMR= 0.026

偏相関係数の推定

	指示	進捗	内容	会話	気遣
1指示	***				0.039
2進捗	0.405	***		0.053	0.049
3内容	0.136	0.446	***	-0.011	0.060
4会話	0.309	-0.000	0.001	***	
5気遣	0.001	0.001	0.000	0.501	***



偏相関の構造が次第にシンプルになってきた  
もう一つ線を切断してみよう

変数間構造の探索 -

左下三角が偏相関係数の推定値，右上三角が相関係数の差

n=101 逸脱度=3.760(df=6) p=0.7091 逸脱度=2.185(df=1) p=0.1394

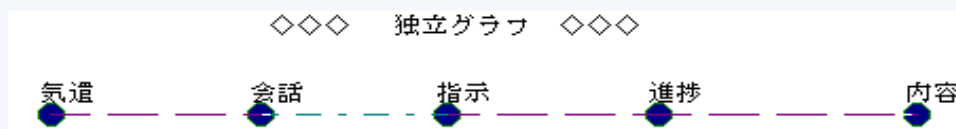
GFI= 0.985 AGFI= 0.963 NFI= 0.973 SRMR= 0.040

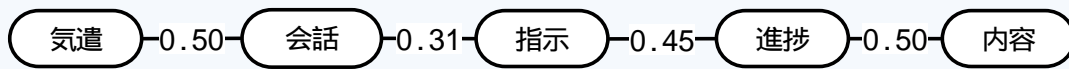
偏相関係数の推定

	指示	進捗	内容	会話	気遣
1指示	***		0.099		0.039
2進捗	0.455	***		0.053	0.049
3内容	0.001	0.498	***	0.032	0.082
4会話	0.312	0.000	0.000	***	
5気遣	0.001	0.000	0.001	0.501	***

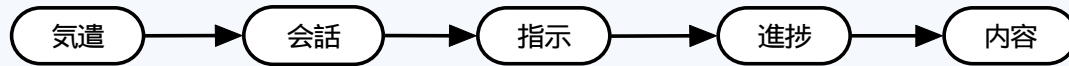
全体としてのあてはまりは  
良好であるが

1つ前の縮約モデルからの  
適合はやや悪い

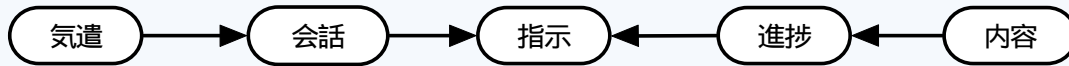




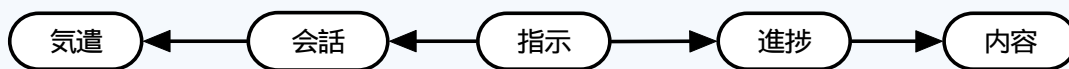
案1. 集団維持能力 ~ 実行計画能力へ



案2. 集団維持能力と実行計画能力が的確な指示を生む

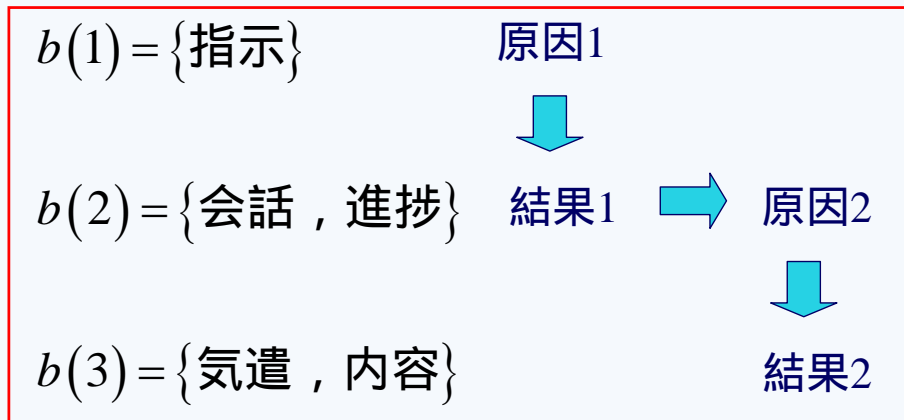


案3. 指示を根として集団維持能力と実行計画能力が必要とされる



V字合流と意味的に案3を採択する  
案2.の場合にはV字合流が存在する

連鎖独立グラフを求める  
そのために無向独立グラフから想定した因果から  
変数を3階層に分類する



$b(1) = \{\text{指示}\}$  ,  $b(2) = \{\text{会話, 進捗}\}$  による無向独立グラフの作成

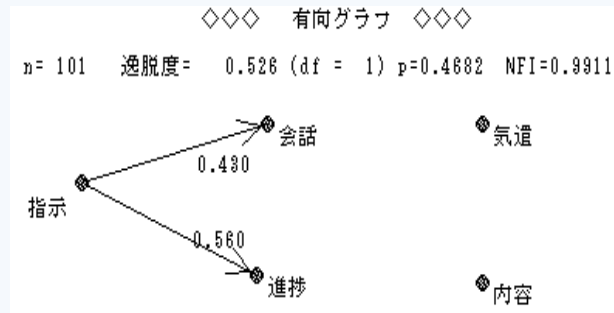
左下三角が偏相関係数の推定値, 右上三角が相関係数の差

n= 101 逸脱度=0.526(df=1)p=0.4682 逸脱度=0.526(df=1)p=0.4682

GFI= 0.997 AGFI= 0.979 NFI= 0.991 SRMR= 0.022

偏相関係数の推定

	指示	会話	進捗
1指示	***		
2会話	0.362	***	0.054
3進捗	0.525	0.000	***



グラフの係数はパス係数

$b(1) = \{\text{指示}\} + b(2) = \{\text{会話, 進捗}\} \& b(3) = \{\text{気遣, 内容}\}$   
による無向独立グラフの作成

**制約: 指示, 会話, 進捗間の線は凍結する(3つの役割は原因)**

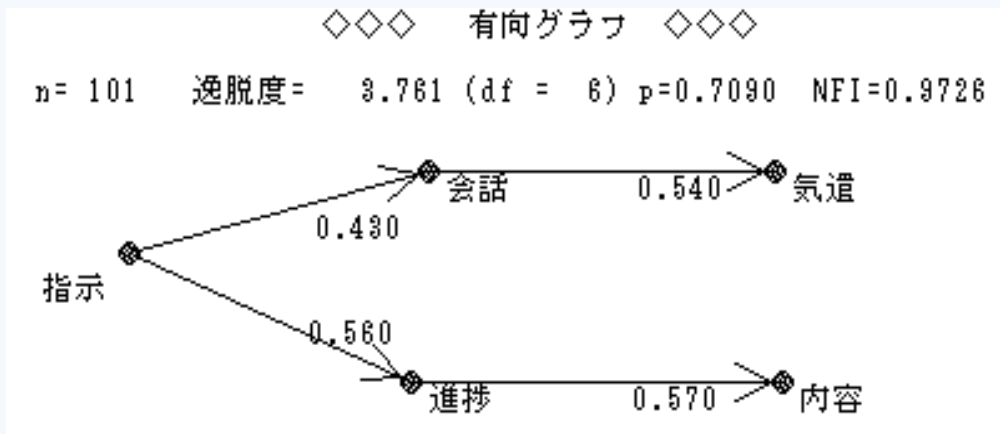
左下三角が偏相関係数の推定値, 右上三角が相関係数の差

n= 101 逸脱度=3.234(df=5) p=0.6639 逸脱度= 2.185(df=1) p=0.1394

GFI= 0.987 AGFI= 0.962 NFI= 0.976 SRMR= 0.033

偏相関係数の推定

	指示	進捗	会話	内容	気遣
1指示	***		0.099	0.039	
2進捗	0.440	***		0.021	
3会話	0.285	0.054	***	0.002	
4内容	0.000	0.497	0.000	***	0.066
5気遣	0.001	-0.000	0.500	0.001	***

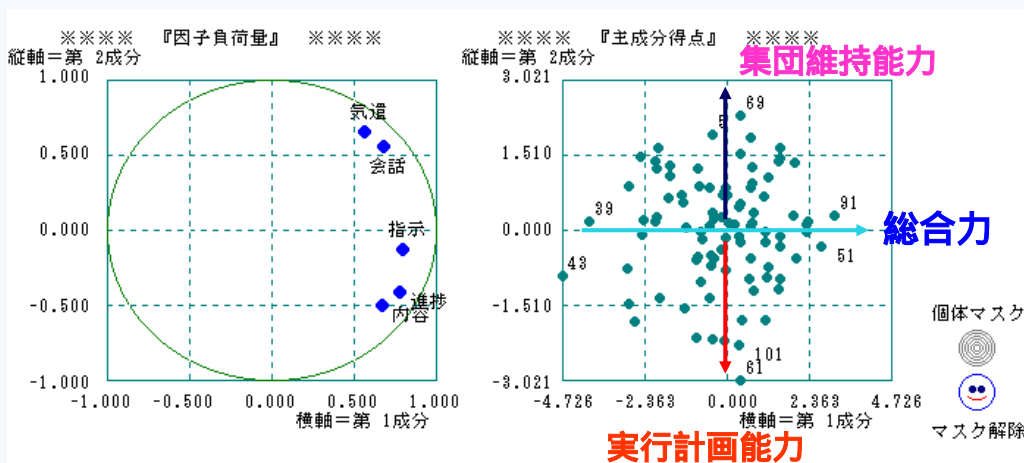


逸脱度や自由度, p値は, 無向グラフの場合と同じ  
無向グラフからの仮説が検証できた

6.他の古典的手法との関連についてー

1.主成分分析

主成分分析は回転による合成変数により軸の解釈を行う. 基本は共通性をみている



分類・マップ化は得意, しかし その先どうするか鍵がない

## 2.因子分析

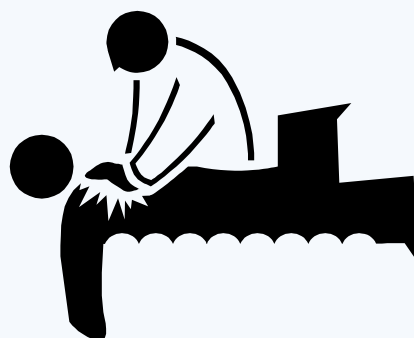
因子分析は，因子の単純構造化と潜在因子の同定が特徴，軸の解釈は因子分析の方が分かりよい  
因子間構造の研究のためにSEMへ発展

バリマックス回転後の因子負荷量				プロマックス回転後因子パターン		
変数名	1 因子	2 因子	共通性	変数名	1 因子	2 因子
指示	0.7134	0.3720	0.6474	指示	0.6798	0.2686
進捗	0.8654	0.1357	0.7673	進捗	0.8766	-0.0020
内容	0.8352	0.0034	0.6976	内容	0.8670	-0.1344
会話	0.2078	0.8514	0.7680	会話	0.0757	0.8499
気遣	0.0598	0.8664	0.7542	気遣	-0.0805	0.8900

この場合，プロマックス回転はうまくいっていない

直交解では2因子の存在が示唆される

いよいよG-GMを使います  
リラックスしましょう。



● G-GM適用上の注意

- データ形式を覚えよう
- 操作のくせを理解してください
- 因果グラフの逐次性を理解しよう

● 事例の注意

- 今回は操作を理解するためのもの
- 変数は誰もが理解できるもの
- 決して課題が解決した事例ではありません

G-GMのコマンド一覧

ウィンドウメニュー	下位メニュー	下位メニュー	下位メニュー	下位メニュー
ファイル	データ行列(CDA)形式 相関係数行列形式 CSV(データ行列)形式	(隠ボタン)	(隠ボタン)	
モニタリング	基本統計量 相関係数行列	多変量連関図	単回帰分析 偏相関分析	(隠ボタン)
多変量解析	主成分分析	主成分分析	因子真荷量 真荷量グラフ 固有値グラフ 主成分分析の終了	バリマックス回転 プロマックス回転
		ヘルプ	バージョン情報 ヒント	
	主変数選択	主変数選択	寄与率グラフ サブセットの保存 主変数選択の終了	
		ヘルプ	バージョン情報 ヒント	
	グラフィカルモデリング	無向グラフ	GM	GMの終了
			共分散選択	手動選択 自動切断 自動接続
			推定モデル	独立グラフ 相関係数の表示
			詳細設定	変数の並替え 収束判定値
			ヘルプ	線の凍結 自動切断規準 バージョン情報 ヒント
		有向グラフ	作図オプション	線の選択 矢線の選択 パス係数 初期位置 矢の大きさ 矢の角度
	クラスター分析			
ヘルプ	バージョン情報 ヒント			



- データ行列(CDA)形式
  - サンプルデータは \*\*\* - C.dat
- 相関係数行列形式
  - サンプルデータは \*\*\* - R.dat
- CSV(データ行列)形式
  - サンプルデータは \*\*\* - C.csv

### 例) ES-C.DAT ファイルの中身

101,5	n(個体数),p(変数の個数)
指示,進捗,内容,会話,気遣	変数名
1,1,1,1,1	変数タイプ(1:量的 2:質的)
5,5,5,5,5	データ
4,3,3,5,3	データ
(以降略)	



<ファイル:市販乳の外観イメージ-R.DAT>

<ファイル:IC工程-R.DAT>

<ファイル:従業員満足度-C.CSV>

**前半が終了**  
**これから質的変数のGMが始まりです。**



# Applied Categorical Data Analysis (質的変数のグラフィカルモデリング)

89

## 質的変数のイメージ(層別効果)

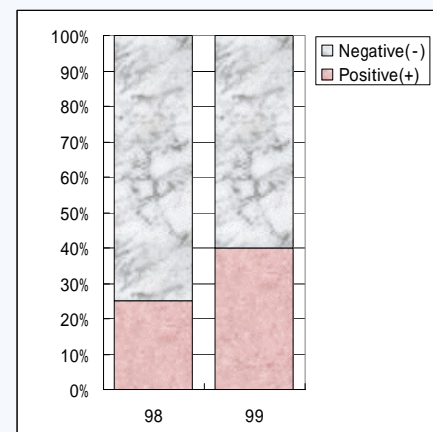
90/136

### ● Classified as belonging to one of two possible categories

- Presence and Absence
- Male and Female
- Good and Bad
- Positive(+) and Negative(-)

$$p = n_1/n \quad (0 \leq p \leq 1)$$

$$n_1 + n_2 = n$$



ある分類項目で分けてみたら差が認められるか？

違いを探す(QCの見方) 原因系の層別因子の水準で切り分け

## 質的変数の分析で使うかなと思われる古典的な手法

- KI法、なぜ3分析など定性的な分解
  - クロス表の解析, 帯グラフ  
(何故かビジネスマンは円グラフが多い)
  - 数量化理論第II類
  - AID(CART、パーティション、etc.)
  - (累積)ロジスティック回帰分析
  - 対応分析(数量化理論第III類)
  - 多重対応分析
- など

- 2元表の解析
  - 独立性, Chi-square, Deviance, etc
- 3元表の解析
  - シンプソン・パラドックス, etc
- 対数線形グラフィカルモデリング
  - 条件付独立, 対数線形モデル etc
- L-GMのデモと例題

2元分割表(a×b分割表)

	B <sub>1</sub>	B <sub>2</sub>	...	B <sub>j</sub>	...	B <sub>b</sub>	計
A <sub>1</sub>	n <sub>11</sub>	n <sub>12</sub>	...	n <sub>1j</sub>	...	n <sub>1b</sub>	n <sub>1+</sub>
A <sub>2</sub>	n <sub>21</sub>	n <sub>22</sub>	...	n <sub>2j</sub>	...	n <sub>2b</sub>	n <sub>2+</sub>
⋮	⋮	⋮	...	⋮	...	⋮	⋮
A <sub>i</sub>	n <sub>i1</sub>	n <sub>i2</sub>	...	n <sub>ij</sub>	...	n <sub>ib</sub>	n <sub>i+</sub>
⋮	⋮	⋮	...	⋮	...	⋮	⋮
A <sub>a</sub>	n <sub>a1</sub>	n <sub>a2</sub>	...	n <sub>aj</sub>	...	n <sub>ab</sub>	n <sub>a+</sub>
計	n <sub>+1</sub>	n <sub>+2</sub>	...	n <sub>+j</sub>	...	n <sub>+b</sub>	n

A, B Item Category

n<sub>ij</sub> 度数 (計数值)

$$n_{i+} = n_{i1} + n_{i2} + \dots + n_{ij} + \dots + n_{ib} = \sum_{j=1}^b n_{ij} \quad n_{+j} = n_{1j} + n_{2j} + \dots + n_{ij} + \dots + n_{aj} = \sum_{i=1}^a n_{ij}$$

$$n = n_{11} + n_{12} + \dots + n_{ij} + \dots + n_{ab} = \sum_{i=1}^a \sum_{j=1}^b n_{ij}$$

QCの例題から 3つの組立ラインの不良

表 不良の分類データ

No.	ライン			不良の種類			
	A <sub>1</sub>	A <sub>2</sub>	A <sub>3</sub>	強度不良	精度不良	外観不良	その他
1	1	0	0	0	1	0	0
2	0	1	0	1	0	0	0
3	1	0	0	1	0	0	0
4	0	0	1	0	0	1	0
5	0	1	0	0	0	0	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
298	1	0	0	1	0	0	0
299	0	0	1	1	0	0	0
300	0	1	0	0	0	1	0

<ファイル:QCの例題-T.DAT>

ラインによって  
不良の種類  
の割合は同じか

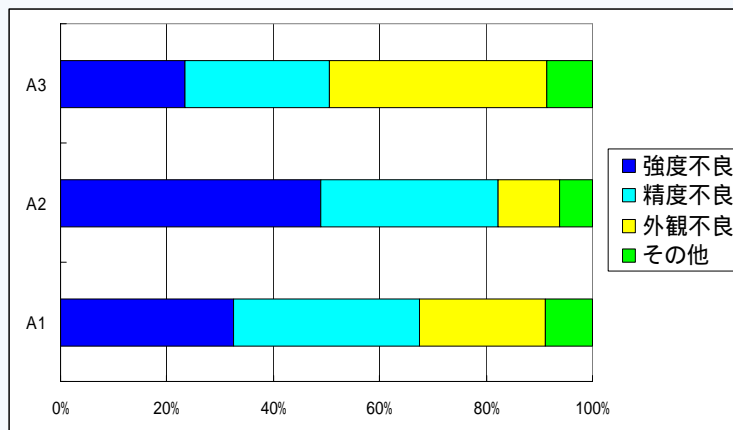
表 3×4分割表(不良の分類調査)

ライン	強度不良	精度不良	外観不良	その他	計
A <sub>1</sub>	n <sub>11</sub> = 40	n <sub>12</sub> = 43	n <sub>13</sub> = 29	n <sub>14</sub> = 11	n <sub>1+</sub> = 123
A <sub>2</sub>	n <sub>21</sub> = 47	n <sub>22</sub> = 32	n <sub>23</sub> = 11	n <sub>24</sub> = 6	n <sub>2+</sub> = 96
A <sub>3</sub>	n <sub>31</sub> = 19	n <sub>32</sub> = 22	n <sub>33</sub> = 33	n <sub>34</sub> = 7	n <sub>3+</sub> = 81
計	n <sub>+1</sub> = 106	n <sub>+2</sub> = 97	n <sub>+3</sub> = 73	n <sub>+4</sub> = 24	n = 300

集計する

層別因子

反応



● 発生確率の比較

- 帰無仮説  $H_0$ : すべての  $i$  と  $j$  に対して  $p_{ij} = p_{i+} \times p_{+j}$
- 対立仮説  $H_1$ : ある  $i$  と  $j$  に対して  $p_{ij} \neq p_{i+} \times p_{+j}$

3.1.1 独立性の検定 (交互作用がない = 独立性)

2元分割表の確率

	B <sub>1</sub>	B <sub>2</sub>	...	B <sub>j</sub>	...	B <sub>b</sub>	計
A <sub>1</sub>	p <sub>11</sub>	p <sub>12</sub>	...	p <sub>1j</sub>	...	p <sub>1b</sub>	p <sub>1+</sub>
A <sub>2</sub>	p <sub>21</sub>	p <sub>22</sub>	...	p <sub>2j</sub>	...	p <sub>2b</sub>	p <sub>2+</sub>
⋮	⋮	⋮	...	⋮	...	⋮	⋮
A <sub>i</sub>	p <sub>i1</sub>	p <sub>i2</sub>	...	p <sub>ij</sub>	...	p <sub>ib</sub>	p <sub>i+</sub>
⋮	⋮	⋮	...	⋮	...	⋮	⋮
A <sub>a</sub>	p <sub>a1</sub>	p <sub>a2</sub>	...	p <sub>aj</sub>	...	p <sub>ab</sub>	p <sub>a+</sub>
計	p <sub>+1</sub>	p <sub>+2</sub>	...	p <sub>+j</sub>	...	p <sub>+b</sub>	1

$$p_{i+} = p_{i1} + p_{i2} + \dots + p_{ij} + \dots + p_{ib} = \sum_{j=1}^b p_{ij}$$

$$p_{+j} = p_{1j} + p_{2j} + \dots + p_{ij} + \dots + p_{aj} = \sum_{i=1}^a p_{ij}$$

$$1 = p_{11} + p_{12} + \dots + p_{ij} + \dots + p_{ab} = \sum_{i=1}^a \sum_{j=1}^b p_{ij}$$

$$p_{ij} = \frac{n_{ij}}{n}$$



● 帰無仮説の下での期待度数  $m_{ij}$  は

- 得られたサンプル総数は  $n$  とする

- 周辺確率  $P_{i+}, P_{+j}$  の推定値  $\tilde{p}_{i+} = \frac{n_{i+}}{n}, \tilde{p}_{+j} = \frac{n_{+j}}{n}$

- 各セルの確率の推定値  $\tilde{p}_{ij} = \tilde{p}_{i+} \times \tilde{p}_{+j} = \frac{n_{i+} \times n_{+j}}{n^2}$

- 各セルの期待度数は, 結局  $m_{ij} = n \times \tilde{p}_{ij} = \frac{n_{i+} \times n_{+j}}{n}$

期待度数の計算

期待度数  $m_{ij}$  (不良の分類調査)

ライン	強度不良	精度不良	外観不良	その他	計
A <sub>1</sub>	$m_{11} = 43.46$	$m_{12} = 39.77$	$m_{13} = 29.93$	$m_{14} = 9.84$	$m_{1+} = 123$
A <sub>2</sub>	$m_{21} = 33.92$	$m_{22} = 31.04$	$m_{23} = 23.36$	$m_{24} = 7.68$	$m_{2+} = 96$
A <sub>3</sub>	$m_{31} = 28.62$	$m_{32} = 26.19$	$m_{33} = 19.71$	$m_{34} = 6.48$	$m_{3+} = 81$
計	$m_{+1} = 106$	$m_{+2} = 97$	$m_{+3} = 73$	$m_{+4} = 24$	$m = 300$



● 期待度数の計算例

$$m_{11} = \frac{n_{1+} \times n_{+1}}{n} = \frac{123 \times 106}{300} = 43.46$$

$$m_{12} = \frac{n_{1+} \times n_{+2}}{n} = \frac{123 \times 97}{300} = 39.77$$

$$m_{13} = \frac{n_{1+} \times n_{+3}}{n} = \frac{123 \times 73}{300} = 29.93$$

- 独立性の検定の帰無仮説の内容

$$p_{ij} = p_{i+} \times p_{+j}$$

- 両辺に対数をとると,

$$\ln p_{ij} = \ln p_{i+} + \ln p_{+j} = \mu + \alpha_i + \beta_j$$

- 2元配置の分散分析と同様に, 各アイテム内の効果の和を0とおく制約をつける

$$\sum_{i=1}^a \alpha_i = \alpha_1 + \alpha_2 + \dots + \alpha_a = 0 \quad \sum_{j=1}^b \beta_j = \beta_1 + \beta_2 + \dots + \beta_b = 0$$

- 対立仮説の内容は  $p_{ij} \neq p_{i+} \times p_{+j}$

- 両辺対数をとると  $\ln p_{ij} \neq \ln p_{i+} + \ln p_{+j} = \mu + \alpha_i + \beta_j$

- 行効果と列効果の単純な和では表現できない  
( $A_i, B_j$ ) の組合せ効果, 交互作用が存在する

$$\ln p_{ij} = \mu + \alpha_i + \beta_j + \underline{(\alpha\beta)_{ij}}$$

$$\sum_{i=1}^a (\alpha\beta)_{ij} = \sum_{j=1}^b (\alpha\beta)_{ij} = 0$$

- 交互作用の有無のチェック
  - 帰無仮説  $H_0$ : すべての  $i$  と  $j$  に対して
$$(\alpha\beta)_{ij} = 0$$
  - 対立仮説  $H_1$ : ある  $i$  と  $j$  に対して
$$(\alpha\beta)_{ij} \neq 0$$
- 交互作用の自由度は  $df=(a-1)(b-1)$

## 3.1.2 対数線形モデル-3

- $\ln p_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$  をフル(飽和)モデル  
Full Model: *FM*
  - 確率  $\hat{p}_{ij} = \frac{n_{ij}}{n}$
- $\ln p_{ij} = \mu + \alpha_i + \beta_j$  を縮約モデル1  
Reduced Model: *RM1*
  - 確率  $\tilde{p}_{ij} = \frac{n_{i+} \times n_{+j}}{n^2}$

無意味なモデル

- $\ln p_{ij} = \mu + \alpha_i$

- $\ln p_{ij} = \mu + \beta_j$

を縮約モデル2,3

Reduced Model: RM2, RM3

- 確率  $\tilde{p}' = \frac{n_{i+}}{nb}$      $\tilde{p}'' = \frac{n_{+j}}{na}$

- $\ln p_{ij} = \mu$  をナルモデル

Null Model: NM

- 確率  $p^0 = \frac{1}{ab}$

- Pは確率    乗法モデル

- モデルの差は, 比で見る  $(\hat{p} / \tilde{p})^2$      ~~$(\hat{p} - \tilde{p})^2$~~

- 対数をとると  $2(\ln \hat{p}_{ij} - \ln \tilde{p}_{ij})$

●

- セルの度数(重み)を考慮  $2n_{ij}(\ln \hat{p}_{ij} - \ln \tilde{p}_{ij})$

- 2元表全体では  $2 \sum_{i=1}^a \sum_{j=1}^b n_{ij} (\ln \hat{p}_{ij} - \ln \tilde{p}_{ij})$

- 度数で書くと  $2 \sum_{i=1}^a \sum_{j=1}^b n_{ij} \left( \ln \frac{n_{ij}}{n} - \ln \frac{m_{ij}}{n} \right) = 2 \sum_{i=1}^a \sum_{j=1}^b n_{ij} \left( \ln \frac{n_{ij}}{m_{ij}} \right)$

- A ItemのCategory数が  $a (\geq 2)$  なので, 2項分布を拡張する

$$f(n_1, n_2, \dots, n_a) = \frac{n!}{n_1! n_2! \dots n_a!} p_1^{x_1} p_2^{x_2} \dots p_a^{x_a} = \frac{n!}{\prod_{i=1}^a n_i!} \prod_{i=1}^a p_i^{x_i}$$

- 平均, 分散, 共分散  $\{n_i\}_{i=1}^a, \sum_{i=1}^a n_i = n$

$$E(N_i) = np_i \quad \text{Var}(N_i) = np_i(1 - p_i)$$

$$\text{Cov}(N_i, N_k) = -np_i p_k \quad (i \neq k)$$

- A, B Item,  $a \times b$  分割表の生起確率に拡張

$$f(n_{11}, n_{12}, \dots, n_{ab}) = \prod_{i=1}^a \left( \frac{n_{i+}!}{\prod_{j=1}^b n_{ij}!} \right) \prod_{j=1}^b (p_{ij})^{n_{ij}}$$

$$\ln L(\hat{p}) = \sum_{i=1}^a \left[ \ln(n_{i+}!) - \sum_{j=1}^b \ln(n_{ij}!) + \sum_{j=1}^b n_{ij} \ln(p_{ij}) \right]$$

● 確率分布に多項分布を考える

- FM, RM各モデルの対数尤度は

$$\ln L(\hat{p}) = \sum_{i=1}^a \left[ \ln(n_{i+}!) - \sum_{j=1}^b \ln(n_{ij}!) + \sum_{j=1}^b n_{ij} \ln(\hat{p}_{ij}) \right]$$

$$\ln L(\tilde{p}) = \sum_{i=1}^a \left[ \ln(n_{i+}!) - \sum_{j=1}^b \ln(n_{ij}!) + \sum_{j=1}^b n_{ij} \ln(\tilde{p}_{ij}) \right]$$

- モデルの差は  $\Delta = \sum_{i=1}^a \sum_{j=1}^b n_{ij} (\ln \hat{p}_{ij}) - \sum_{i=1}^a \sum_{j=1}^b n_{ij} (\ln \tilde{p}_{ij})$
- 対数尤度比  $dev(RM) = 2 \sum_{i=1}^a \sum_{j=1}^b n_{ij} (\ln \hat{p}_{ij} - \ln \tilde{p}_{ij})$

● FMからRMの差を逸脱度  
(モデルとデータとの乖離度)

$$dev(RM) = 2 \sum_{i=1}^a \sum_{j=1}^b n_{ij} (\ln \hat{p}_{ij} - \ln \tilde{p}_{ij}) = 2 \sum_{i=1}^a \sum_{j=1}^b n_{ij} \ln \frac{n_{ij}}{m_{ij}} = 2 \sum_{i=1}^a \sum_{j=1}^b n_{ij} \ln \frac{nn_{ij}}{n_{i+}n_{+j}}$$

- $dev(RM)$ は近似的に  $df=(a-1)(b-1)$ の  $\chi^2$  分布にしたがう
- $dev(RM) \geq \chi_0^2(df, \alpha)$ ならば帰無仮説を棄却
- 交互作用  $(\alpha\beta)_{ij}$  が存在する --- > FM採択

$$\begin{aligned}
 dev(RM) &= 2 \sum_{i=1}^3 \sum_{j=1}^4 n_{ij} \ln \frac{n_{ij}}{m_{ij}} \\
 &= 2 \times \left\{ 40 \ln \frac{40}{43.46} + 43 \ln \frac{43}{39.77} + 29 \ln \frac{29}{29.33} + 11 \ln \frac{11}{9.84} \right. \\
 &\quad + 47 \ln \frac{47}{33.92} + 32 \ln \frac{32}{31.04} + 11 \ln \frac{11}{23.36} + 6 \ln \frac{6}{7.68} \\
 &\quad \left. + 19 \ln \frac{19}{28.62} + 22 \ln \frac{22}{26.19} + 33 \ln \frac{33}{19.71} + 7 \ln \frac{7}{6.48} \right\} \\
 &= 25.63 > \chi^2 \{ (3-1)(4-1) = 6, 0.05 \} = 12.59
 \end{aligned}$$

- 有意水準5%有意
- 不良品の作られたラインと不良品の種類に関連がある

- 仮説検定; 背反な2つの仮説  $H_0, H_1$  について真偽を確かめる
  - 逸脱度が棄却限界値より大きいならば帰無仮説を棄却する

モデリング; 観測データから, どちらのモデル  $FM, RM$  を支持するのが安全かを確かめる

- 逸脱度の大きさを確率的に評価して, 妥当性を判定する

$$p = \Pr \{ \chi^2 \geq dev(RM) \}$$

3元分割表

		C <sub>1</sub>	C <sub>2</sub>	...	C <sub>c</sub>
A <sub>1</sub>	B <sub>1</sub>	n <sub>111</sub>	n <sub>112</sub>	...	n <sub>11c</sub>
	B <sub>2</sub>	n <sub>121</sub>	n <sub>122</sub>	...	n <sub>12c</sub>
	⋮	⋮	⋮	...	⋮
	B <sub>b</sub>	n <sub>1b1</sub>	n <sub>1b2</sub>	...	n <sub>1bc</sub>
A <sub>2</sub>	B <sub>1</sub>	n <sub>211</sub>	n <sub>212</sub>	...	n <sub>21c</sub>
	B <sub>2</sub>	n <sub>221</sub>	n <sub>222</sub>	...	n <sub>22c</sub>
	⋮	⋮	⋮	...	⋮
	B <sub>b</sub>	n <sub>2b1</sub>	n <sub>2b2</sub>	...	n <sub>2bc</sub>
⋮	⋮	⋮	...	⋮	
A <sub>a</sub>	B <sub>1</sub>	n <sub>a11</sub>	n <sub>a12</sub>	...	n <sub>a1c</sub>
	B <sub>2</sub>	n <sub>a21</sub>	n <sub>a22</sub>	...	n <sub>a2c</sub>
	⋮	⋮	⋮	...	⋮
	B <sub>b</sub>	n <sub>ab1</sub>	n <sub>ab2</sub>	...	n <sub>abc</sub>

$$n_{ij+} = n_{ij1} + n_{ij2} + \dots + n_{ijk} + \dots + n_{ijc} = \sum_{k=1}^c n_{ijk}$$

$$n_{i+k} = n_{i1k} + n_{i2k} + \dots + n_{ijk} + \dots + n_{ibk} = \sum_{j=1}^b n_{ijk}$$

$$n_{+jk} = n_{1jk} + n_{2jk} + \dots + n_{ijk} + \dots + n_{ajk} = \sum_{i=1}^a n_{ijk}$$

$$n_{i++} = n_{i11} + n_{i12} + \dots + n_{ijk} + \dots + n_{ibc} = \sum_{j=1}^b \sum_{k=1}^c n_{ijk}$$

$$n_{+j+} = n_{1j1} + n_{1j2} + \dots + n_{ijk} + \dots + n_{ajc} = \sum_{i=1}^a \sum_{k=1}^c n_{ijk}$$

$$n_{++k} = n_{11k} + n_{12k} + \dots + n_{ijk} + \dots + n_{abk} = \sum_{i=1}^a \sum_{j=1}^b n_{ijk}$$

$$n = n_{111} + n_{112} + \dots + n_{ijk} + \dots + n_{abc} = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c n_{ijk}$$

バート表

	A <sub>1</sub>	A <sub>2</sub>	...	A <sub>a</sub>	B <sub>1</sub>	B <sub>2</sub>	...	B <sub>b</sub>	C <sub>1</sub>	C <sub>2</sub>	...	C <sub>c</sub>
A <sub>1</sub>	n <sub>1++</sub>											
A <sub>2</sub>	0	n <sub>2++</sub>										
⋮	⋮	⋮	...									
A <sub>a</sub>	0	0	...	n <sub>a++</sub>								
B <sub>1</sub>	n <sub>11+</sub>	n <sub>21+</sub>	...	n <sub>a1+</sub>	n <sub>+1+</sub>							
B <sub>2</sub>	n <sub>12+</sub>	n <sub>22+</sub>	...	n <sub>a2+</sub>	0	n <sub>+2+</sub>						
⋮	⋮	⋮	...	⋮	⋮	⋮	...					
B <sub>b</sub>	n <sub>1b+</sub>	n <sub>2b+</sub>	...	n <sub>ab+</sub>	0	0	...	n <sub>+b+</sub>				
C <sub>1</sub>	n <sub>1+1</sub>	n <sub>2+1</sub>	...	n <sub>a+1</sub>	n <sub>+11</sub>	n <sub>+21</sub>	...	n <sub>+b1</sub>	n <sub>++1</sub>			
C <sub>2</sub>	n <sub>1+2</sub>	n <sub>2+2</sub>	...	n <sub>a+2</sub>	n <sub>+12</sub>	n <sub>+22</sub>	...	n <sub>+b2</sub>	0	n <sub>++2</sub>		
⋮	⋮	⋮	...	⋮	⋮	⋮	...	⋮	⋮	⋮	...	
C <sub>c</sub>	n <sub>1+c</sub>	n <sub>2+c</sub>	...	n <sub>a+c</sub>	n <sub>+1c</sub>	n <sub>+2c</sub>	...	n <sub>+bc</sub>	0	0	...	n <sub>++c</sub>



<ファイル:3元分割表-T.DAT>

表 多肢選択のデータ形式

No.	項目A	項目B	項目C
1	1	2	1
2	2	1	1
3	1	1	2
4	a	2	c
5	2	b	2
⋮	⋮	⋮	⋮
n-2	2	1	c
n-1	1	1	2
n	2	b	2

同じデータ

3つの質的変数のデータの形式

No.	項目A				項目B				項目C			
	A <sub>1</sub>	A <sub>2</sub>	...	A <sub>a</sub>	B <sub>1</sub>	B <sub>2</sub>	...	B <sub>b</sub>	C <sub>1</sub>	C <sub>2</sub>	...	C <sub>c</sub>
1	1	0	...	0	0	1	...	0	1	0	...	0
2	0	1	...	0	1	0	...	0	1	0	...	0
3	1	0	...	0	1	0	...	0	0	1	...	0
4	0	0	...	1	0	1	...	0	0	0	...	1
5	0	1	...	0	0	0	...	1	0	1	...	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
n-2	0	1	...	0	1	0	...	0	0	0	...	1
n-1	1	0	...	0	1	0	...	0	0	1	...	0
n	0	1	...	0	0	0	...	1	0	1	...	0

3.2.2 層別とシンプソンパラドックス

例 n=100の3元表データ

表3.5 3元分割表

		C <sub>1</sub>	C <sub>2</sub>
A <sub>1</sub>	B <sub>1</sub>	20	5
	B <sub>2</sub>	15	7
A <sub>2</sub>	B <sub>1</sub>	4	11
	B <sub>2</sub>	6	32

表3.6 バート表

	A <sub>1</sub>	A <sub>2</sub>	B <sub>1</sub>	B <sub>2</sub>	C <sub>1</sub>	C <sub>2</sub>
A <sub>1</sub>	47					
A <sub>2</sub>	0	53				
B <sub>1</sub>	25	15	40			
B <sub>2</sub>	22	38	0	60		
C <sub>1</sub>	35	10	24	21	45	
C <sub>2</sub>	12	43	16	39	0	55

- **バート表は量的データの多変量連関図に対応**
  - 対角ブロックはヒストグラムに対応
  - 非対角ブロックは散布図に対応
- **各変数組の逸脱度は相関係数に対応**

● 項目 B , C 間を調べる

- 項目 A を考慮しないとき 有意

$$dev(RM)=6.09 > \chi^2(1,0.05)=3.84$$

- 項目 A を考慮(A の category で層別) するとき

- A1 有意でない  $dev(RM)=0.86$

- A2 有意でない  $dev(RM)=0.79$

● 併合した場合の検定結果と層別した場合の検定結果が食い違う現象を

シンプソン・パラドックス(Simpson's Paradox)

3元分割表で使われる表記

$$p_{ijk} = \frac{n_{ijk}}{n}$$

$$p_{ij+} = p_{ij1} + p_{ij2} + \dots + p_{ijk} + \dots + p_{ijc} = \sum_{k=1}^c p_{ijk}$$

$$p_{i+k} = p_{i1k} + p_{i2k} + \dots + p_{ijk} + \dots + p_{ibk} = \sum_{j=1}^b p_{ijk}$$

$$p_{+jk} = p_{1jk} + p_{2jk} + \dots + p_{ijk} + \dots + p_{ajk} = \sum_{i=1}^a p_{ijk}$$

$$p_{i++} = p_{i11} + p_{i12} + \dots + p_{ijk} + \dots + p_{ibc} = \sum_{j=1}^b \sum_{k=1}^c p_{ijk}$$

$$p_{+j+} = p_{1j1} + p_{1j2} + \dots + p_{ijk} + \dots + p_{ajc} = \sum_{i=1}^a \sum_{k=1}^c p_{ijk}$$

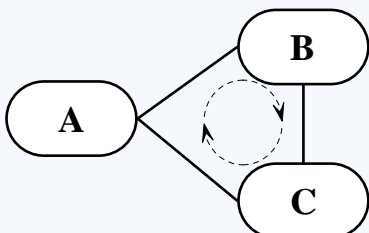
$$p_{++k} = p_{11k} + p_{12k} + \dots + p_{ijk} + \dots + p_{abk} = \sum_{i=1}^a \sum_{j=1}^b p_{ijk}$$

$$1 = p_{111} + p_{112} + \dots + p_{ijk} + \dots + p_{abc} = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c p_{ijk}$$

●  $P_{ijk}$  に対する FM (Full Model) は

$$\ln p_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk}$$

制約条件



$$\sum_{i=1}^a \alpha_i = \sum_{j=1}^b \beta_j = \sum_{k=1}^c \gamma_k = 0$$

$$\sum_{i=1}^a (\alpha\beta\gamma)_{ijk} = \sum_{j=1}^b (\alpha\beta\gamma)_{ijk} = \sum_{k=1}^c (\alpha\beta\gamma)_{ijk} = 0$$

$$\sum_{i=1}^a (\alpha\beta)_{ij} = \sum_{j=1}^b (\alpha\beta)_{ij} = \sum_{i=1}^a (\alpha\gamma)_{ik} = \sum_{k=1}^c (\alpha\gamma)_{ik} = \sum_{j=1}^b (\beta\gamma)_{jk} = \sum_{k=1}^c (\alpha\beta)_{jk} = 0$$

● 両者の考え方の違い

- 対数線形モデルでは3因子交互作用のみの削除は考慮できる

$$\ln p_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} \quad \cancel{(\alpha\beta\gamma)_{ijk}}$$

- 対数線形グラフカルモデルでは3因子交互作用のみの削除は不可

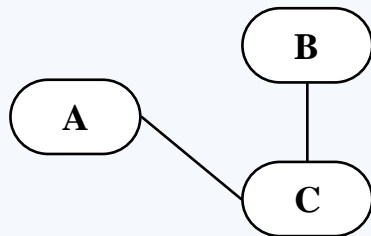
ある2因子交互作用を削除すると、  
それを含む高次の交互作用も削除される

$$\ln p_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} \quad \cancel{(\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk}}$$

$$\ln p_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\beta\gamma)_{jk} \quad \cancel{(\alpha\gamma)_{ik} + (\alpha\beta\gamma)_{ijk}}$$

$$\ln p_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} \quad \cancel{(\alpha\beta)_{ij} + (\alpha\beta\gamma)_{ijk}}$$

- A × B を削除する場合は  
A × B × C も削除しなければならない



$$\ln p_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk}$$

- C を与えた下で A と B は条件付独立  
 $A \perp B \mid C$

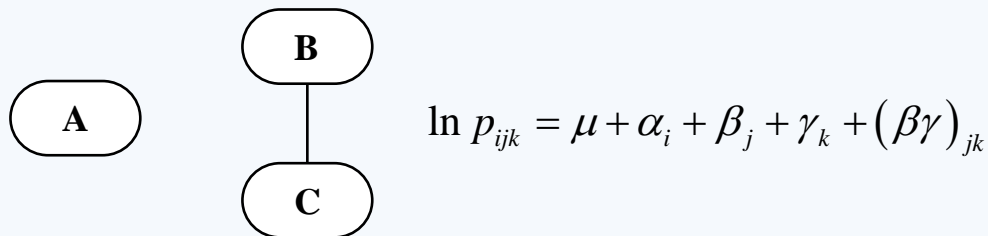
- $A \perp B \mid C$  での確率計算は確率の乗法定理より、

$$\Pr(A \cap B \mid C) = \Pr(A \mid C) \times \Pr(B \mid C)$$



$$\frac{p_{ijk}}{p_{+++k}} = \frac{p_{i+k}}{p_{+++k}} \times \frac{p_{+jk}}{p_{+++k}}$$

- すでに  $A \times B \times C$  は削除されているので  
A × C のみを削除する

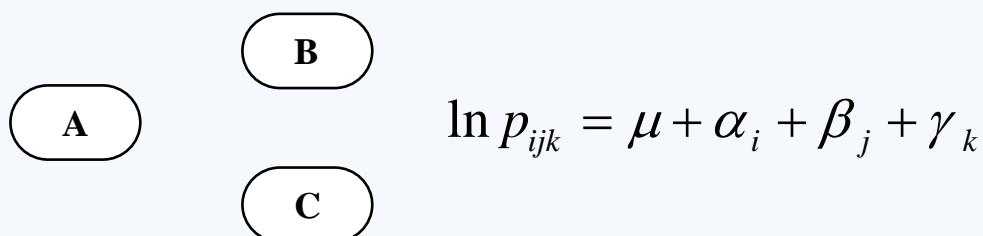


$A \perp (B, C)$  : A は (B と C) と独立である

$$\Pr(A \cap B \cap C) = \Pr(A) \times \Pr(B \cap C)$$

$$p_{ijk} = p_{i++} \times p_{+jk}$$

- すでに  $A \times B \times C$  は削除されているので  
B × C のみを削除する



A, B, C は互いに独立

$$\Pr(A \cap B \cap C) = \Pr(A) \times \Pr(B) \times \Pr(C)$$

$$p_{ijk} = p_{i++} \times p_{+j+} \times p_{++k}$$

- 2因子交互作用の有無を独立グラフの辺の有無と対応させる
- ある2因子交互作用を削除すると, それを含む高次の交互作用も削除される
- グラフィカルモデリングでは表現できないモデルがある例)

$$\ln p_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\alpha\beta\gamma)_{ijk}$$

$$\ln p_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk}$$

- モデルの評価は逸脱度とp値

- FM(フルモデル)の下での $p_{ijk}$ の推定値  $\hat{p}_{ijk} = \frac{n_{ijk}}{n}$

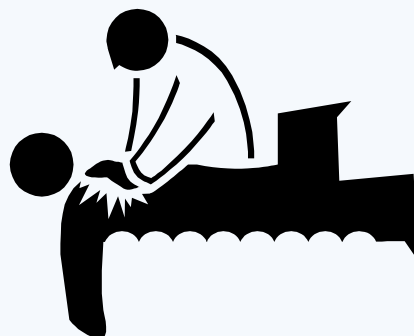
- RM(縮約モデル)の下での $p_{ijk}$ の推定値  $\tilde{p}_{ijk} = \frac{m_{ijk}}{n}$

- 逸脱度  $dev(RM) = 2 \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c n_{ijk} (\ln \hat{p}_{ijk} - \ln \tilde{p}_{ijk}) = 2 \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c n_{ijk} \ln \frac{n_{ijk}}{m_{ijk}}$

- p値  $p = \Pr(\chi^2 \geq dev(RM))$

- FMからの乖離は,  $p$ 値が $0.50 = 1/2$ を目安
  - $RM$ は $FM$ とほぼ同等
  - $RM$ はデータに適合している
- 辺の除去( $RM1-RM2$ )の指標
  - $dev(RM1, RM2) = dev(RM2) - dev(RM1)$ が近似的に  $\chi^2$  分布に従うことを使う
  - $p$ 値が $0.30$ 程度より大ならば,  $RM1$ と $RM2$ はほぼ同等

いよいよL-GMを使います  
リラックスしましょう。



● L-GM適用上の注意

- データ形式を覚えよう
- 操作のくせを理解してください
- 因果グラフの逐次性を理解しよう

● 事例の注意

- 今回は操作を理解するためのもの
- 変数は誰もが理解できるもの
- 決して課題が解決した事例ではありません

L-GMのコマンド一覧

ウィンドウメニュー	下位メニュー	下位メニュー	下位メニュー	下位メニュー
データ	データ行列(CDA)形式 多元分割表形式 CSV(データ行列)形式	(隠ボタン)		
モニタリング	パート表 2変数間の逸脱度 多変量連関図	度数グラフ 分割表グラフ		
多変量解析	多重対応分析	多重対応分析	数量化得点 布置の連関図 固有値グラフ 多重対応分析の終了	(隠ボタン) 布置図
		ヒント	バージョン情報 ヒント	
	グラフィカルモデリング	無向グラフ	GM	層別2元分割表 GMの終了
			変数(交互作用)選択	手動選択 自動切断 自動接続
			推定モデル	独立グラフ 対数線形モデル
			詳細設定	変数の並替え 線の凍結 自動切断規準
			ヘルプ	バージョン情報 ヒント
		有向グラフ	作図オプション	線の選択 矢線の選択 初期位置 矢の大きさ 矢の大きさ角度
ヘルプ	バージョン情報 ヒント			



- データ行列(CDA)形式
  - サンプルデータは \*\*\* - N.dat
- 分割表形式
  - サンプルデータは \*\*\* - T.dat
- CSV(データ行列)形式
  - サンプルデータは \*\*\* - N.csv

### 例) ソフト-N.datファイルの中身

307,5	n(個体数),p(変数の個数)
始読,取説,検索,混乱,操作	変数名
2,2,2,2,2	変数タイプ(1:量的 2:質的)
1,2,2,2,2	データ
2,2,2,2,2	データ

(以降略)

例) 信頼性-T.datファイルの中身

3	p(変数の個数)
“会社”,3,“機種”,4,“故障”,2	変数名,水準数
32,11,56,20,21, 3, 8, 1	頻度
39,16,22, 9,86,11,25, 3	
8, 3, 5, 3,10, 3,87, 5	

例) エッセル-N.CSVファイルの中身

The screenshot shows an Excel spreadsheet with the following data:

	A	B	C	D	E	F	G	H
1	216	6						
2	菓子替	親しみ	甘補給	定番	甘物欲	おやつ		
3	2	2	2	2	2	2		
4	1	1	1	1	1	2		
5	2	2	2	1	2	2		
6	1	1	1	2	2	2		
7	2	1	2	2	2	1		
8	1	1	2	1	1	2		
9	2	1	1	1	1	1		
10	2	2	1	2	1	2		
11	1	2	1	2	1	1		
12	2	1	1	1	1	1		
13	1	1	1	1	1	2		

Annotations in the image:

- n,p**: Points to the values 216 and 6 in the first row.
- 変数名**: Points to the variable names in the second row.
- 変数タイプ**: Points to the numerical values in the subsequent rows.
- データ**: Points to the numerical values in the subsequent rows.
- セル内をクリアにしておく**: A text box on the right side of the spreadsheet.

<ファイル:画質-N.CSV>

<ファイル:メロディ-N.CSV>

<ファイル:選好-T.DAT>

後半も終了  
お疲れ様でした。楽しいランチを!!

