

於：2018.12.05

代々木

JUSE活用事例シンポ

STATWORKSとの出会いから、 SEM因果分析に出会うまで

(株)リコー

事業開発本部 事業統括室

廣野 元久

本日、ご紹介したいこと

- 1. 自己紹介とSTAT Worksとの出会い
- 2. GMとSEMとの出会い
 - GM : グラフィカルモデリング
 - SEM : 構造方程式モデル
- 3. SEM因果分析を使ったケーススタディ
 - ・主要都市の降水日の相関
 - ・回帰分析を超えて
 - ・システム思考の調査分析

1.1 自己紹介

- 1984(株)年リコー入社品質管理本部配属(当時)
 - ・努力の怪アリ？で統計語を「カタコト」話せる(石の上に35年)
- JUSEのMA研で統計解析の面白さにはまる（1990年代）
 - ・日科技連のMAセミナーで奥野先生・吉澤先生・芳賀先生の大家に出会う
 - ・MA研の合宿は椿先生・今泉先生と合宿運営（毎回徹夜）
- PCを活用したSQC教育の構築(階層別品質技術研修)
 - ・社内で初のPC研修
 - （多変量解析・実験計画(含：パラメータ設計)・データ解析基礎）
- QM・CS・標準・安全の実施責任者（QM推進室長）
- SF(スマートファシリティ)事業所長
 - ・50歳の手洗いで新規事業開発を拝命

1.2 STAT Worksとの出会い

年代	STAT Worksに関連する出来事
1984～ 1990	・JSQC/JSAの各研究会で飯塚・広津・椿・宮川・永田各先生に教えを頂く機会を得る ・奥野・吉澤・芳賀先生との出会い⇒MA研とSTAT Works（当時はQCAS-MA）に出会う 「累積法、精密累積法、そして累積カイ二乗法」（品質：宮川先生と共著）⇒ロジスティック回帰に興味を持つ
1991～ 2000	・MA研運営委員（日科技連）/東京理科大（神楽坂）非常勤講師（実験計画法） 「寿命データにおける芳賀のロジスティック回帰近似の適用」（R&Mシンポジウム/JUSE-MAシンポジウム） 「設計開発におけるSQC」（品質）⇒PCAの診断グラフがSTAT Worksの主成分得点グラフに採用？ 「ロジット回帰七変化」（日科技連MA専門）⇒寄与率の考え方がSTAT Worksの予測判定グラフに採用？ 「分散分析,回帰分析における残差」・「ロジスティック回帰分析、機能窓への応用」（標準化と品質管理） 「グラフィカルモデリングの実際」（日科技連：共著）★日経品質管理文献賞
2001～ 2010	・MA研運営委員（日科技連）/慶應義塾（湘南藤沢）非常勤講師（多変量解析・統計解析） ・STAT WorksのSEM因果分析編に貢献 「グラフィカルモデリングのためのG-GM&L-GMデータ解析システム」（計算機統計） 「データ解析の鉄人に学ぶ探索的データ解析」（行動計量学会）でGMを紹介
2011～ 2018	・MA研運営委員（日科技連） 「SEM因果分析入門」（日科技連：山口先生と共著） 「アンスコムの数値例で学ぶ統計的方法23講」（日科技連：永田先生と共著）★日経品質管理文献賞 「目からウロコの統計学」（日科技連）

2.1 GM&SEMとの出会い

	GM&SEMに関連する出来事
1996	・JSQC(日本品質管理学会)テクノメトリクス研究会に参画 ・初代主査の宮川先生のリーダーシップでGM(グラフィカルモデリング)の研究がスタート ・芳賀(敏)先生によるDOS版のGMのソフトウェアの開発 ⇒積極的に社内のCS/ES調査や工程データに適用してみた ⇒これは使える。グループワークに向いている
1999	・2代主査の永田先生を中心として「グラフィカルモデリングの実際」が出版 ⇒芳賀(敏)先生のDOS版ソフトを先生として廣野がWin版のソフトを開発
2000	・好みの計量研に参画し、芳賀(麻)先生の紹介でSEMを知る ⇒なんでパス図に1月も2月もかけるのか不思議（研究者が羨ましくも工業への適用に？）
2001	・行動計量学会「牛乳の買いたさの構造を探る3」（芳賀(麻)先生と共著） ⇒探索的因果分析GMから確証的因果分析SEMの展開を提案（後にSTAT Worksが反映してくれた）
2002	G-GMとL-GMのシステムを計算機統計に投稿 （ガウシアングラフィカルモデリングと対数線形グラフィカルモデリングが可能なWinソフト）



STAT WorksのSEM因果分析編にGM+SEMが搭載された

2.2 GM・SEMとは

変数間の相関情報から統計的な理論により内部構造を単純化しグラフで表現する

■ 標本統計量

標本相関係数行列

$$R = \begin{pmatrix} 1.00 & & & & \\ 0.56 & 1.00 & & & \\ 0.68 & 0.68 & 1.00 & & \\ 0.27 & 0.18 & 0.32 & 1.00 & \\ 0.42 & 0.29 & 0.31 & 0.54 & 1.00 \end{pmatrix}$$

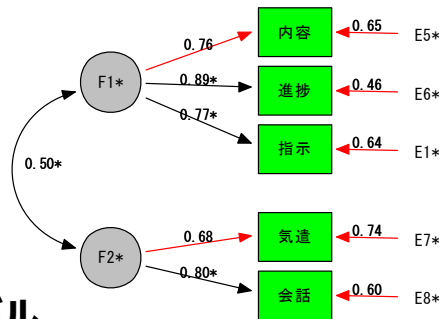


標本偏相関係数行列

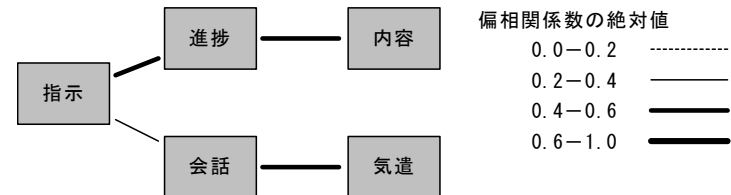
$$P = \begin{pmatrix} - & & & & \\ 0.16 & - & & & \\ 0.46 & 0.51 & - & & \\ -0.06 & -0.13 & 0.24 & - & \\ 0.28 & 0.11 & -0.12 & 0.50 & - \end{pmatrix}$$

■ 出力されるグラフ

パス図



(無向)独立グラフ



■ 推定されたモデル

モデルの相関係数行列

$$\hat{\Pi} = \begin{pmatrix} 1.00 & & & & \\ 0.58 & 1.00 & & & \\ 0.68 & 0.67 & 1.00 & & \\ 0.26 & 0.25 & 0.30 & 1.00 & \\ 0.31 & 0.30 & 0.35 & 0.54 & 1.00 \end{pmatrix}$$

モデルの偏相関係数行列

$$\hat{\Lambda} = \begin{pmatrix} - & & & & \\ 0.00 & - & & & \\ 0.53 & 0.57 & - & & \\ 0.00 & 0.00 & 0.00 & - & \\ 0.28 & 0.00 & 0.00 & 0.50 & - \end{pmatrix}$$

モデルの相関係数行列

$$\hat{\Pi} = \begin{pmatrix} 1.00 & & & & \\ 0.46 & 1.00 & & & \\ 0.68 & 0.68 & 1.00 & & \\ 0.23 & 0.11 & 0.16 & 1.00 & \\ 0.42 & 0.20 & 0.29 & 0.54 & 1.00 \end{pmatrix}$$

2.3 因果分析の勧め

- ・問題解決では層別の考え方が大切

例) 品質管理活動では「なぜなぜ分析」・新QC七つ道具の「連関図法」

- ・「連関図」や「なぜなぜ分析」は定性的な因果分析で人間心理が影響する

⇒経験による誤解や偏り、初めに結論ありきで誤った仮説を導く恐れがある

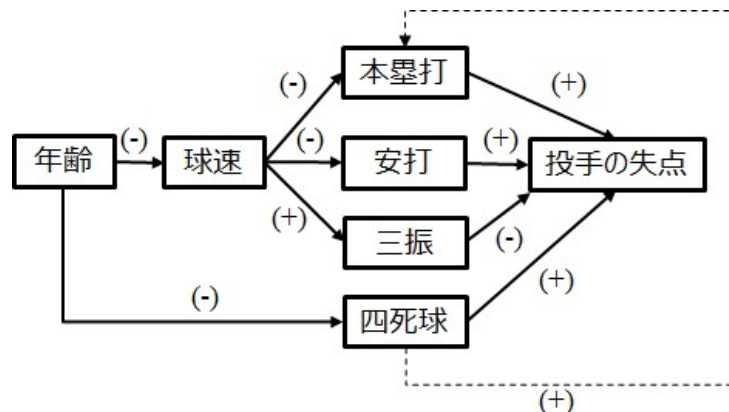
- ・GMやSEMは数値データを使った定量的な因果分析

⇒数値データを使った「なぜなぜ分析」(因果分析)はシステム思考に向いている

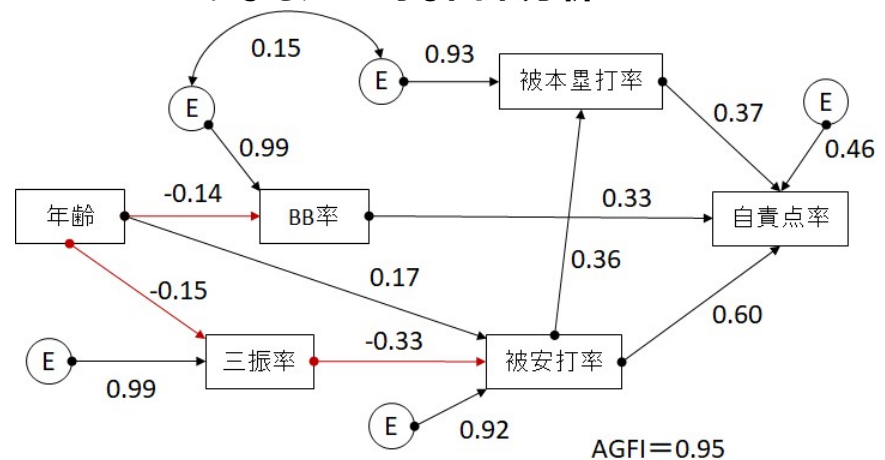
- ・定量的「なぜなぜ分析」をお勧めしたい

⇒以下はプロ野球の先発投手の失点の因果分析

■投手が点をとられる理由は(定性)



■SEMによる定量的な因果分析



2.4 GMとSEMの比較

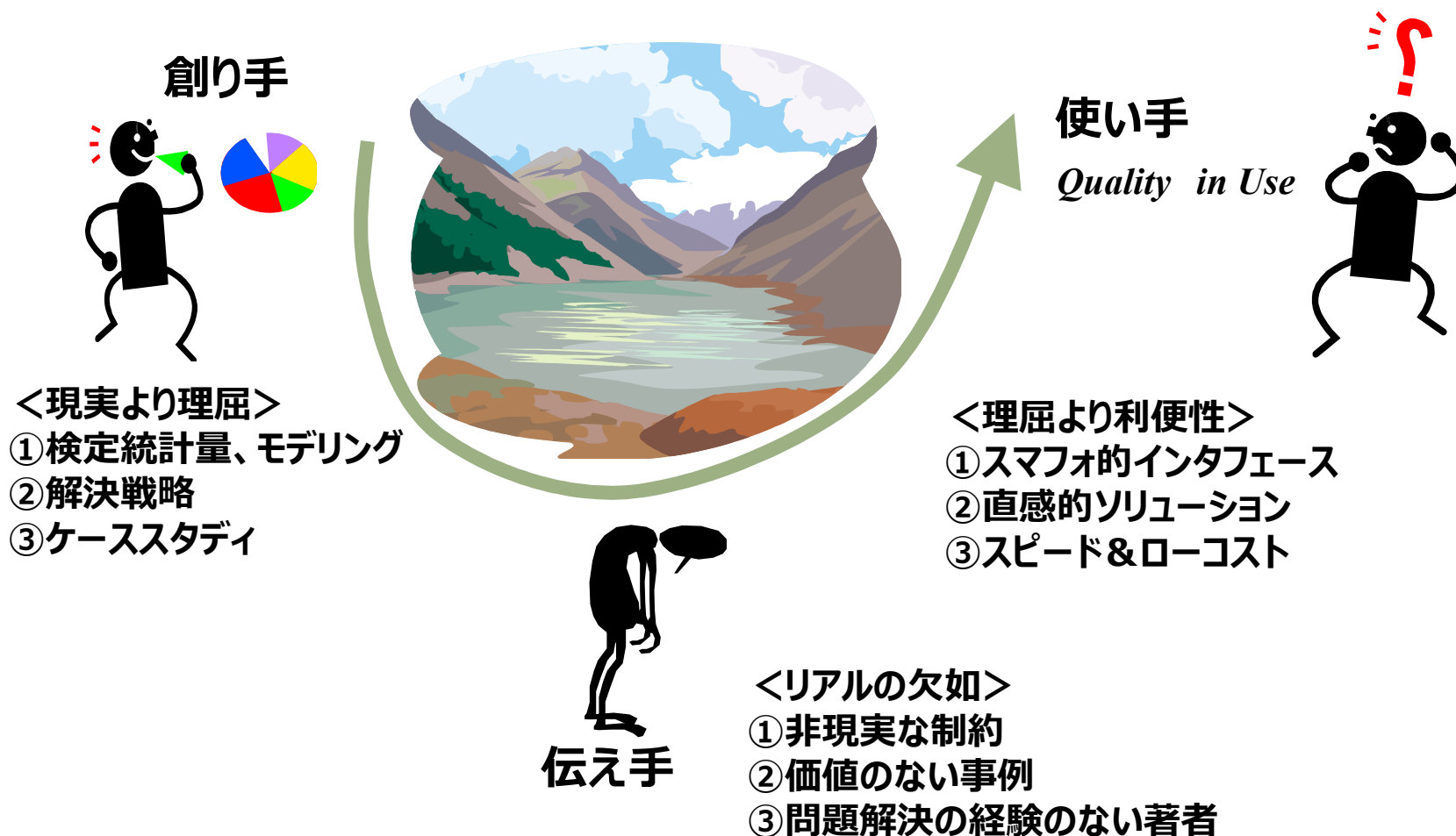
比較する項目	GM（裏線形）	SEM（表線形）
アプローチ	探索的	確証的
出発行列	偏相関係数行列	相関係数行列
ターゲット	母偏相関係数行列	母相関係数行列
特徴	条件付き独立性 (相関係数がゼロか否かは興味なし)	構造(連立)方程式 (相関係数がゼロか否かに興味あり)
モデルの選定	曖昧な仮説で十分	明解な仮説が必要
推定するもの	母偏相関係数	パス係数
モデル選択の方法	偏相関係数の絶対値 (変数減少法)	相関係数のカイ2乗 (ワルド検定・LM検定)
潜在変数 (含：誤差変数)	考慮しない	観測値から効果と誤差を分解
因果分析の方法	階層構造を与え段階的に推定	確証的なモデルで一度に推定

因果仮説の検証ステップ例（山口・廣野(2011)「SEM因果分析入門」、日科技連出版）

- ❶ GMの無向グラフでクリークに着目して因子発見⇒SEMで仮説検証
- ❷ 求めた因子を使いGMの無向グラフ⇒GMによる有向グラフで因子間構造を推定
- ❸ SEMによる統合モデルでパス係数を推定して因果仮説を同定する

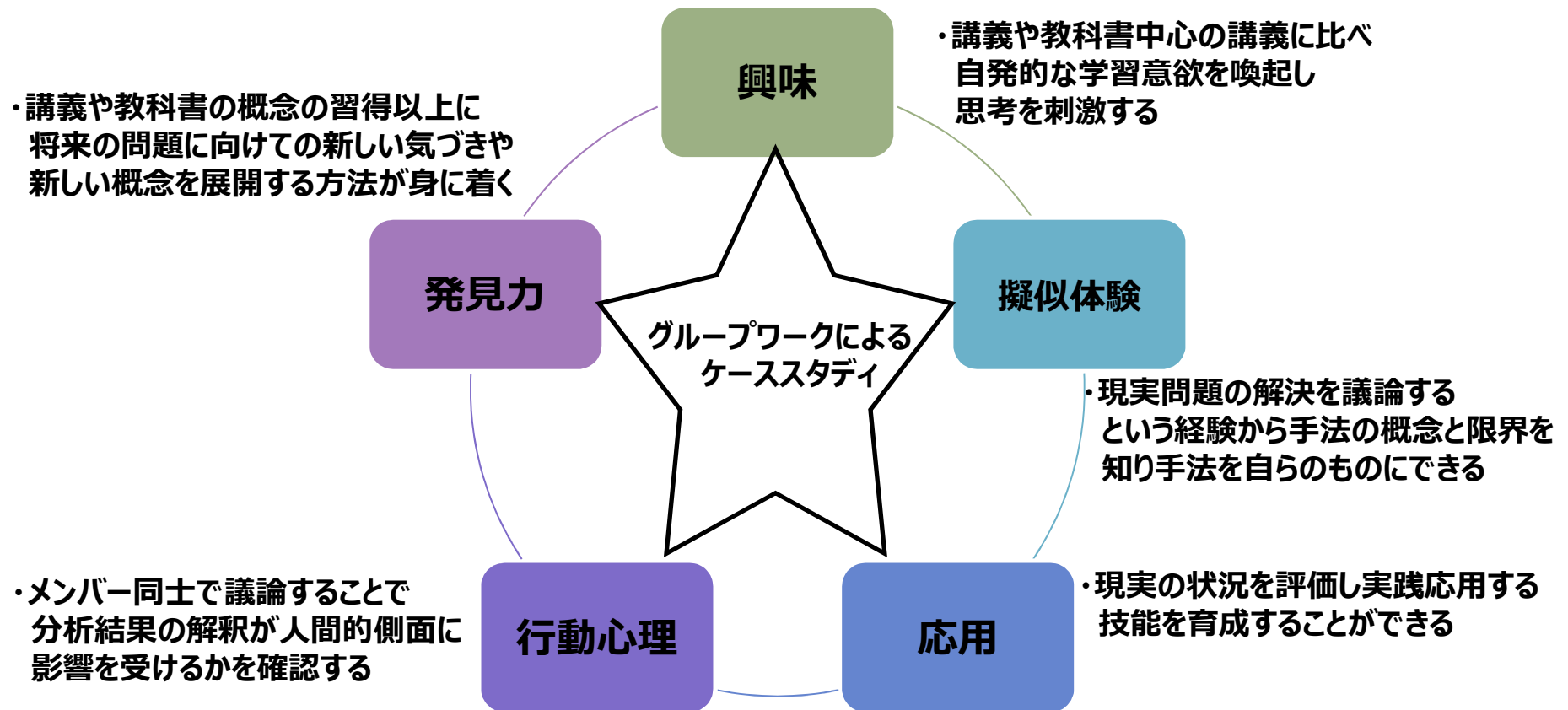
3.1 手法開発と実利用の間の深い谷がある

現実とモデリングとの間にある死の谷



3.2 統計学習にケーススタディを誘う

- ・今の教育は手順を習得させることに拘泥しすぎ？(座学や独学に偏ると一元的な解釈に陥る)
- ・企業活動は常にグループワークだからチームで問題解決に取り組むのが本来の姿
- ・統計教育にもグループワークを取り入れることができる（ケースメソッドの活用）
- ・因果分析の仮説成長アプローチはグループワークによるケーススタディに有効

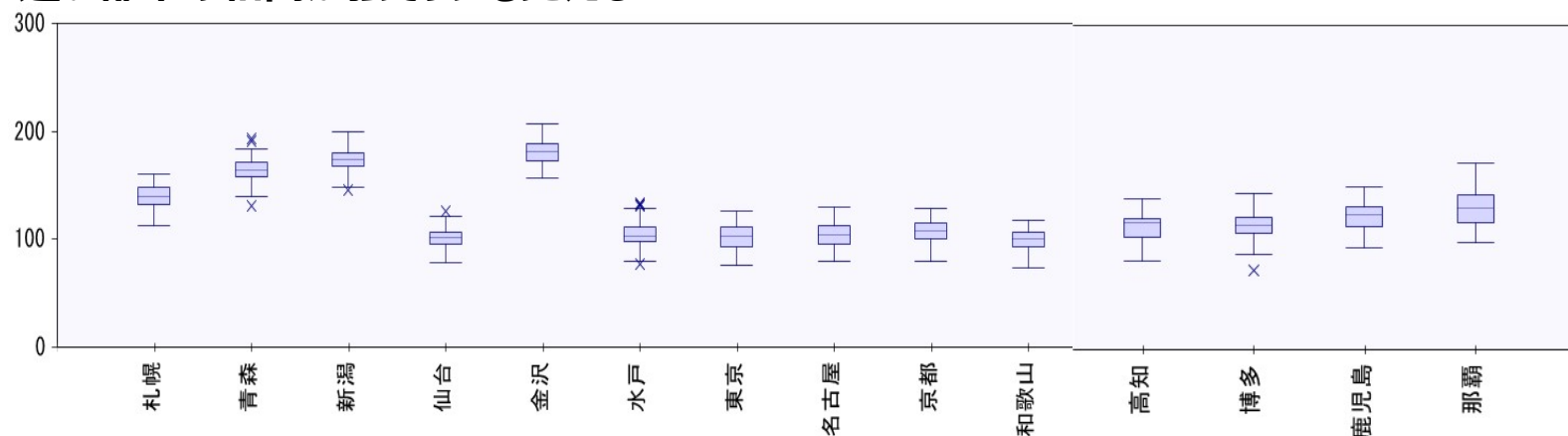




3.3 主要都市の降水日の相関

1927年以降の主要都市の降水日の相関係数などを調べてみると

距離が近い都市の相関が強そうにも見える……



		サンプル数: 84 +: 0.6 以上 ++: 0.8 以上														
No	変数名	札幌	青森	新潟	仙台	金沢	水戸	東京	名古屋	京都	和歌山	広島	高知	博多	鹿児島	那覇
2	札幌	1.000	0.318	0.148	-0.060	0.126	-0.021	-0.099	0.027	-0.141	-0.056	0.023	-0.097	0.103	0.072	-0.032
3	青森	0.318	1.000	0.417	-0.024	0.375	-0.098	-0.242	-0.103	-0.275	-0.132	-0.121	-0.186	-0.244	-0.066	0.200
4	新潟	0.148	0.417	1.000	0.437	0.691+	0.278	0.156	0.269	0.113	0.173	0.233	-0.037	0.097	0.227	0.104
5	仙台	-0.060	-0.024	0.437	1.000	0.409	0.562	0.537	0.405	0.395	0.342	0.355	0.203	0.373	0.325	0.116
6	金沢	0.126	0.375	0.691+	0.409	1.000	0.360	0.214	0.346	0.277	0.256	0.226	0.049	0.230	0.332	0.198
7	水戸	-0.021	-0.098	0.278	0.562	0.360	1.000	0.775+	0.523	0.488	0.454	0.409	0.276	0.306	0.251	0.157
8	東京	-0.099	-0.242	0.156	0.537	0.214	0.775+	1.000	0.627+	0.586	0.595	0.499	0.422	0.408	0.271	0.104
9	名古屋	0.027	-0.103	0.269	0.405	0.346	0.523	0.627+	1.000	0.756+	0.668+	0.646+	0.521	0.514	0.326	0.085
10	京都	-0.141	-0.275	0.113	0.395	0.277	0.488	0.586	0.756+	1.000	0.757+	0.702+	0.662+	0.634+	0.512	0.047
11	和歌山	-0.056	-0.132	0.173	0.342	0.256	0.454	0.595	0.668+	0.757+	1.000	0.685+	0.667+	0.585	0.436	0.033
12	広島	0.023	-0.121	0.233	0.355	0.226	0.409	0.499	0.646+	0.702+	0.685+	1.000	0.645+	0.708+	0.492	-0.043
13	高知	-0.097	-0.186	-0.037	0.203	0.049	0.276	0.422	0.521	0.662+	0.667+	0.645+	1.000	0.565	0.486	0.102
14	博多	0.103	-0.244	0.097	0.373	0.230	0.306	0.408	0.514	0.634+	0.585	0.708+	0.565	1.000	0.712+	-0.025
15	鹿児島	0.072	-0.066	0.227	0.325	0.332	0.251	0.271	0.326	0.512	0.436	0.492	0.486	0.712+	1.000	0.184
16	那覇	-0.032	0.200	0.104	0.116	0.198	0.157	0.104	0.085	0.047	0.033	-0.043	0.102	-0.025	0.184	1.000

3.4 GMで無向グラフを作る

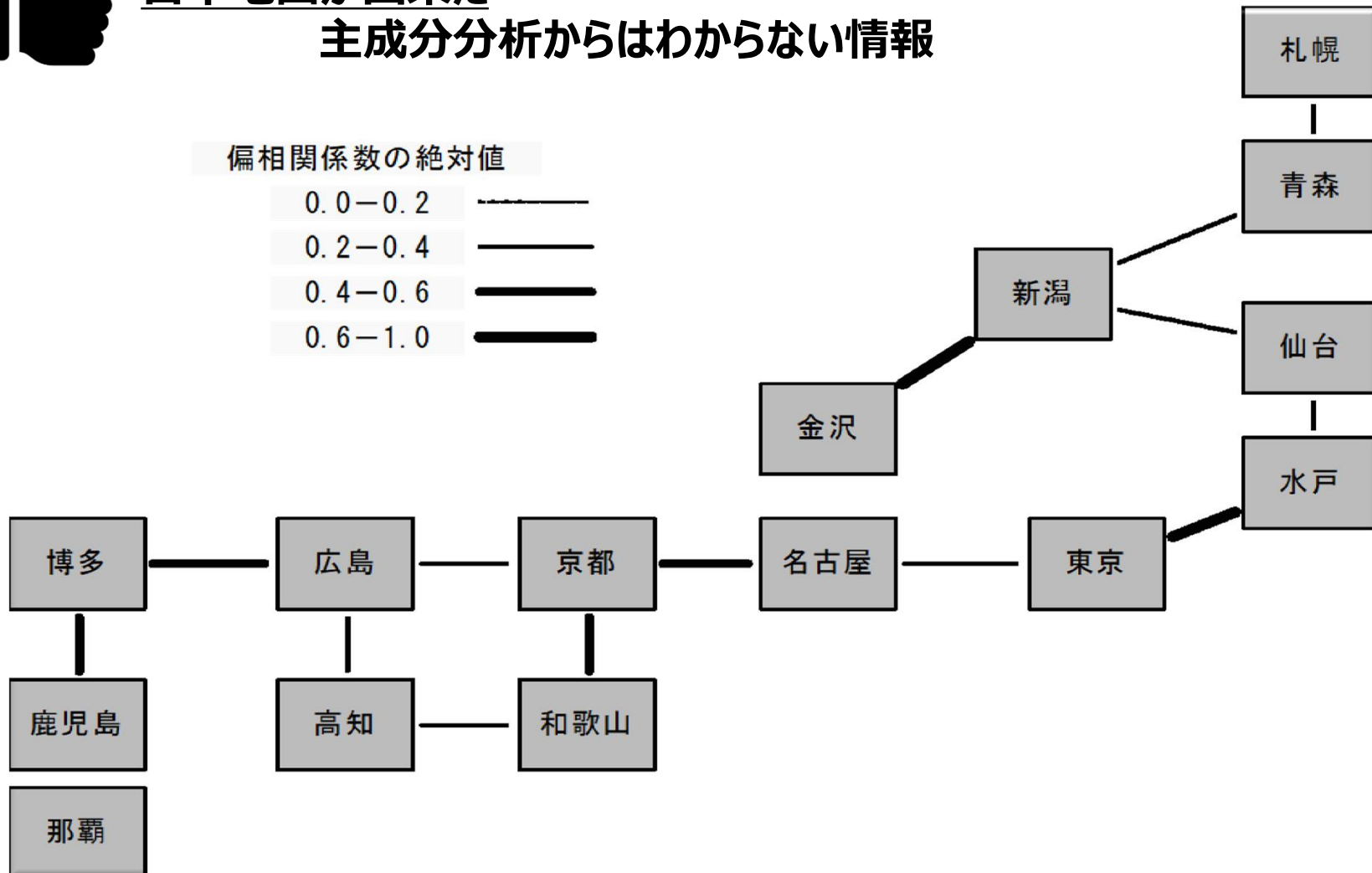


日本地図が出来た

主成分分析からはわからない情報

偏相関係数の絶対値

0.0-0.2	-----
0.2-0.4	-----
0.4-0.6	=====
0.6-1.0	=====





3.5 回帰分析を超えて

- 下の表は6組のデータセットそれぞれから相関係数行列を求めたもの
6組それぞれの重回帰式を求めてください（フルモデル）

ケース①	x_1	x_2	x_3	y	ケース②	x_1	x_2	x_3	y
x_1	1.000				x_1	1.000			
x_2	0.000	1.000			x_2	0.686	1.000		
x_3	0.000	0.000	1.000		x_3	0.000	-0.001	1.000	
y	0.544	0.544	0.544	1.000	y	0.751	0.698	0.513	1.000
ケース③	x_1	x_2	x_3	y	ケース④	x_1	x_2	x_3	y
x_1	1.000				x_1	1.000			
x_2	0.858	1.000			x_2	0.934	1.000		
x_3	0.225	0.256	1.000		x_3	0.913	0.823	1.000	
y	0.837	0.830	0.515	1.000	y	0.970	0.940	0.926	1.000
ケース⑤	x_1	x_2	x_3	y	ケース⑥	x_1	x_2	x_3	y
x_1	1.000				x_1	1.000			
x_2	-0.685	1.000			x_2	-0.715	1.000		
x_3	0.000	0.000	1.000		x_3	0.034	-0.285	1.000	
y	-0.008	0.399	0.742	1.000	y	0.438	-0.186	0.679	1.000

同じ重回帰式が
得られる



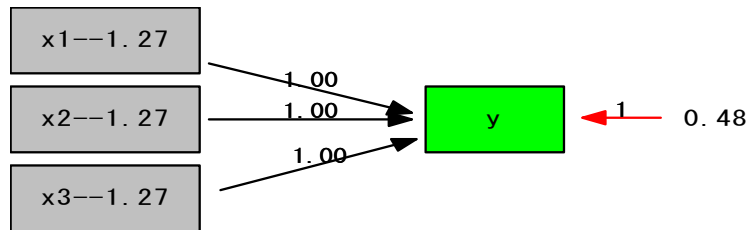
$$\hat{y} = 5.00 + 1.00x_1 + 1.00x_2 + 1.00x_3$$

3.6 SEMで説明変数の内部構造を理解

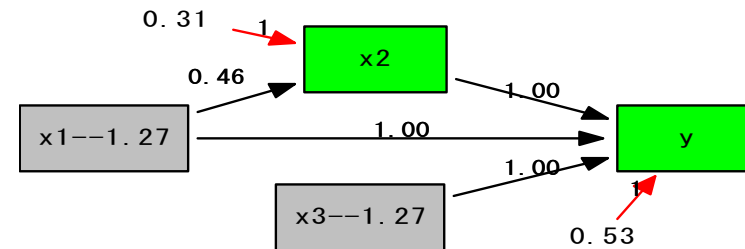


SEMにより説明変数の内部構造の違いが確認できた
因果分析の極意は要因で特性を層別すること

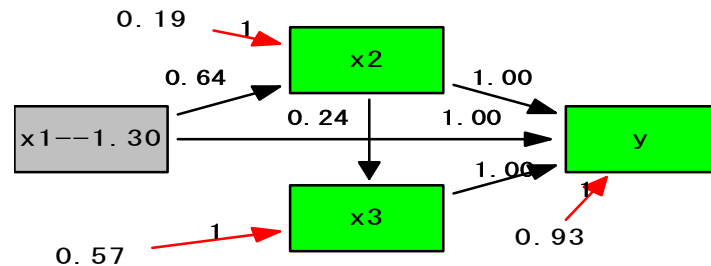
① x_1 の総合効果 : 1



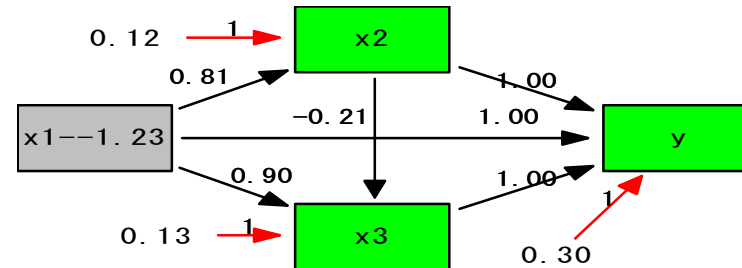
② x_1 の総合効果 : 1.46



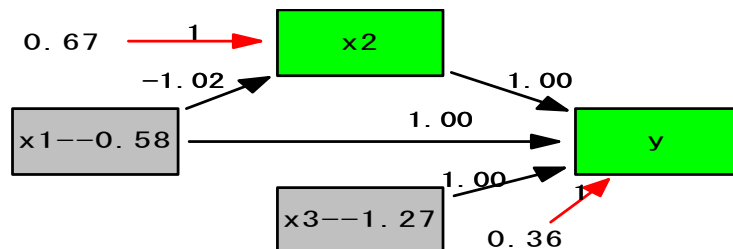
③ x_1 の総合効果 : 1.79



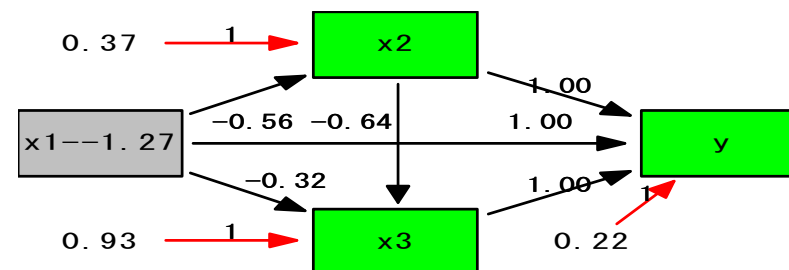
④ x_1 の総合効果 : 2.54



⑤ x_1 の総合効果 : -0.02



⑥ x_1 の総合効果 : 0.45





3.7 システム思考の調査分析

従業員満足度から改善ネタ探しをしてください

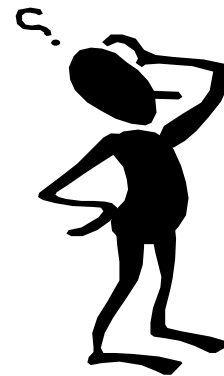
職場における部下から見た上司の関わりについてのアンケート結果

以下は10の変数についての説明

- ・指示：業務への適切な指示与え方
- ・活用：業務結果や資料の有効活用
- ・内容：業務内容の把握
- ・評価：部下の仕事の適切な評価
- ・受入：部下の仕事結果の受入
- ・進捗：業務の進捗管理
- ・会話：部下との会話
- ・気遣：部下への気遣い
- ・叱咤：部下の仕事ぶりへの励まし
- ・非難：部下の仕事ぶりのまずさへの叱責

回転後の因子パターン(4因子)

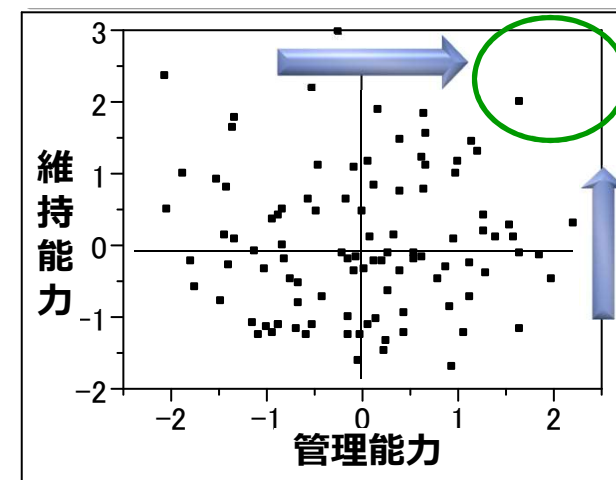
指示	0.788	0.358	0.131	-0.080
叱咤	0.795	-0.084	-0.026	0.348
活用	0.710	0.137	0.368	0.113
非難	0.189	0.012	0.076	0.951
内容	0.561	0.104	0.686	0.020
進捗	0.720	0.251	0.441	0.140
気遣	0.053	0.870	0.081	0.063
会話	0.098	0.779	0.179	-0.026
評価	0.135	0.381	0.823	0.071
受入	0.226	0.701	0.210	-0.028



何ができるか？

＜因子分析＞

因子分析で重要因子を探索する



3.8 GMを使い因果仮説の探索



<構造の探索>

無向グラフで構造探索し因果グラフに渡す



<因果の探索>

因果グラフでモデル探索する

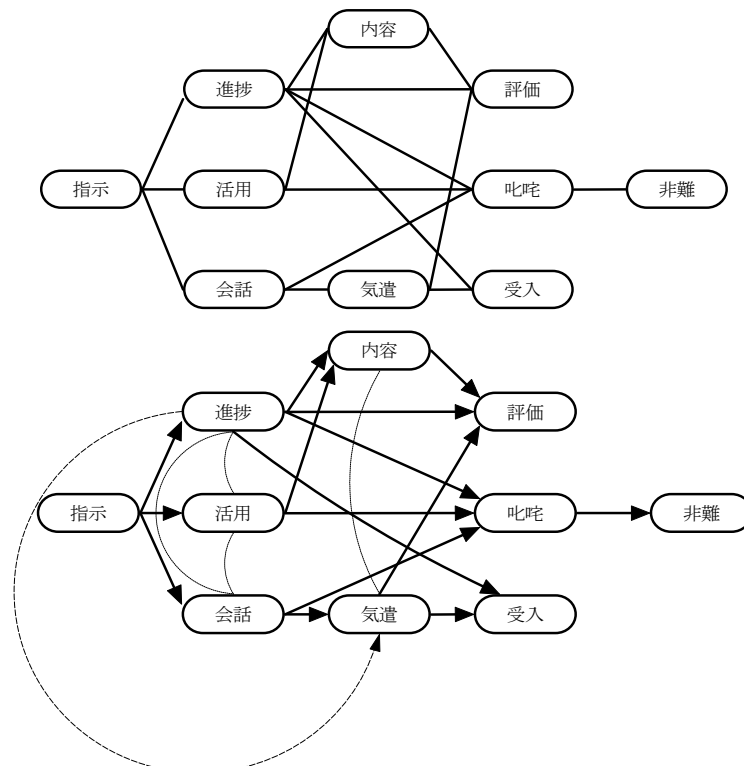


図9.17 独立グラフ（上）からの因果モデル（下）の想定

$$b(1) = \{\text{指示}\}$$

$$b(2) = \{\text{進捗, 活用, 会話}\}$$

$$b(3) = \{\text{内容, 気遣}\}$$

$$b(4) = \{\text{評価, 叱咤, 受入}\}$$

$$b(5) = \{\text{非難}\}$$

フルモデルとの比較：逸脱度=23.743 自由度=29 p値=0.7415

適合度指標：NFI=0.946

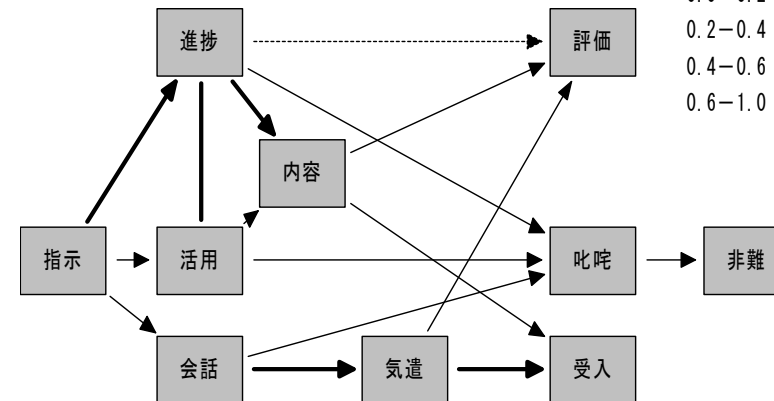
偏相関係数の絶対値

0.0-0.2

0.2-0.4

0.4-0.6

0.6-1.0

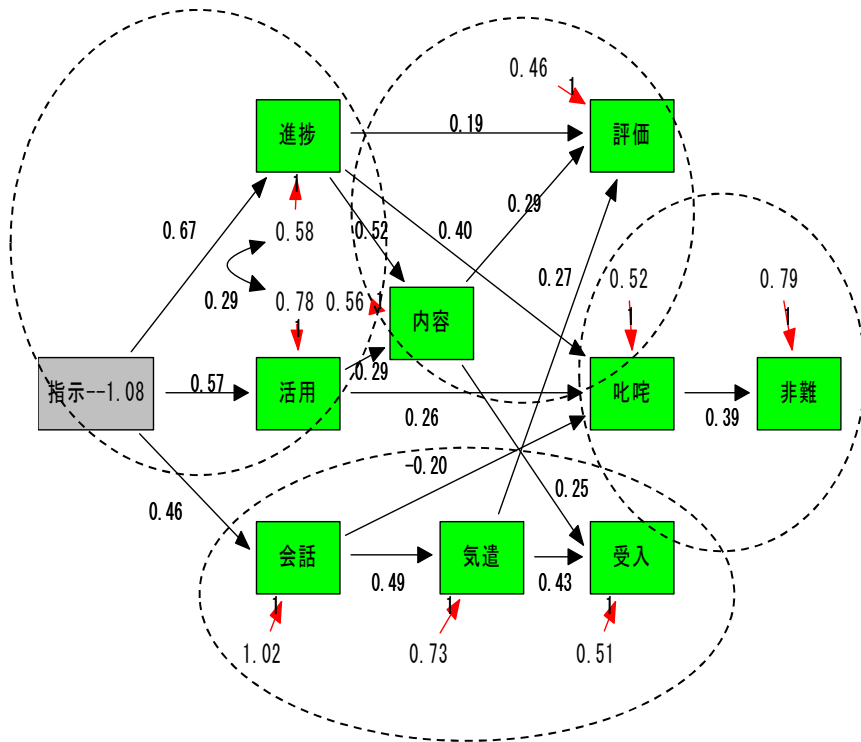




3.9 SEMを使い因果仮説の同定

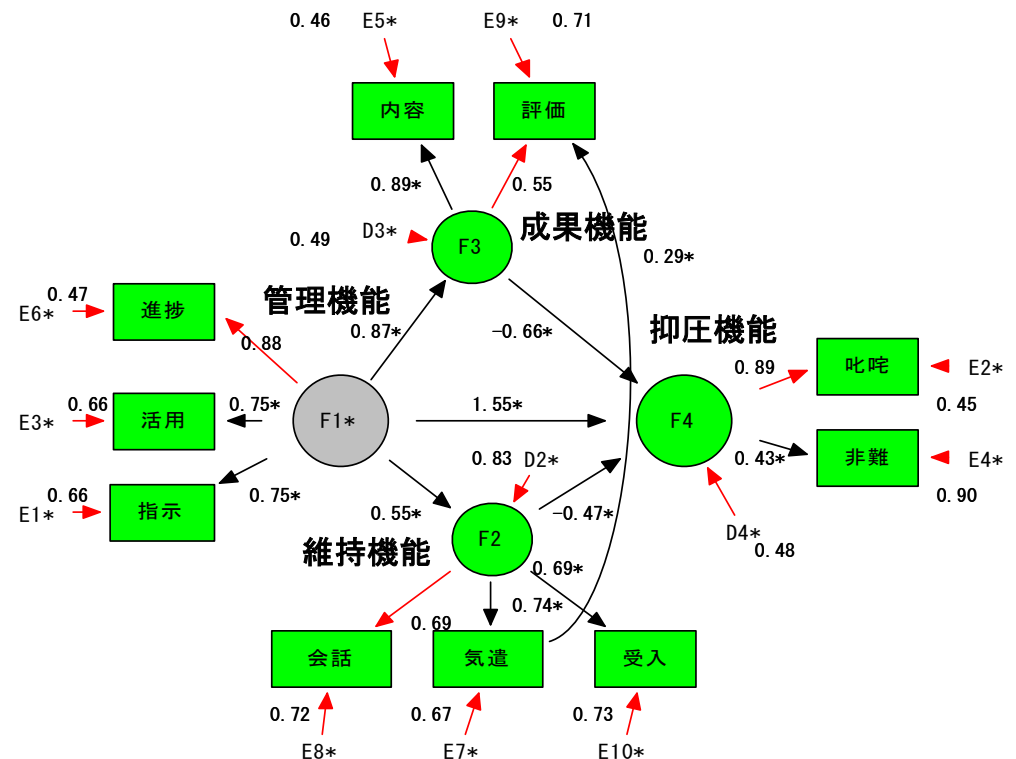
<因子の探索>

GMの因果モデルをSEMに渡す
(STAT Worksでは有向独立グラフが
SEMに引き継がれる)



<因子の解釈>

クリークに着目しながら因子を追加



管理機能→抑圧機能: 直接効果 1.55
管理→成果→抑圧: 間接効果 -0.57
管理→維持→抑圧: 間接効果 -0.26

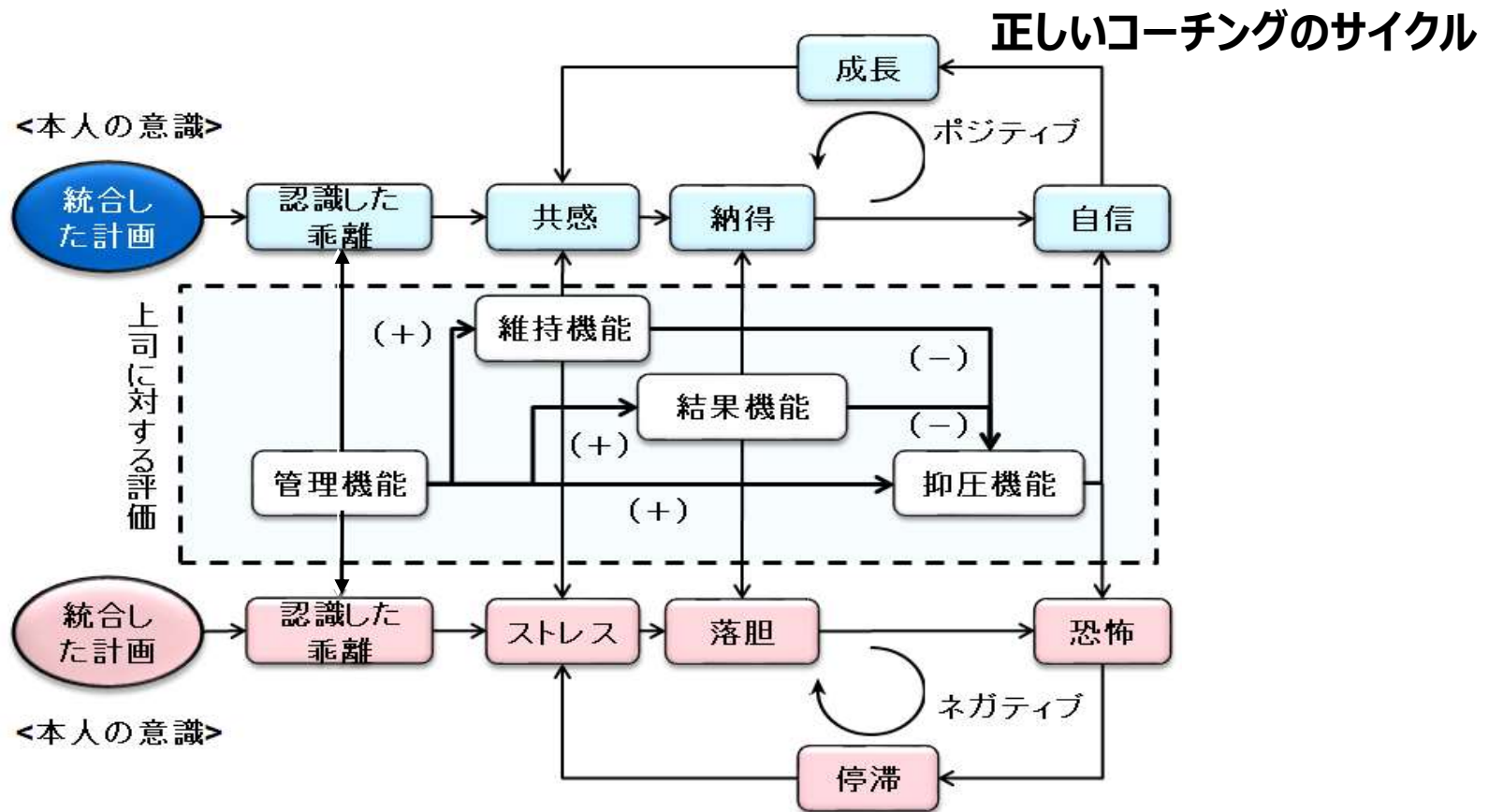
総合効果

0.72
(直接効果の約半分)

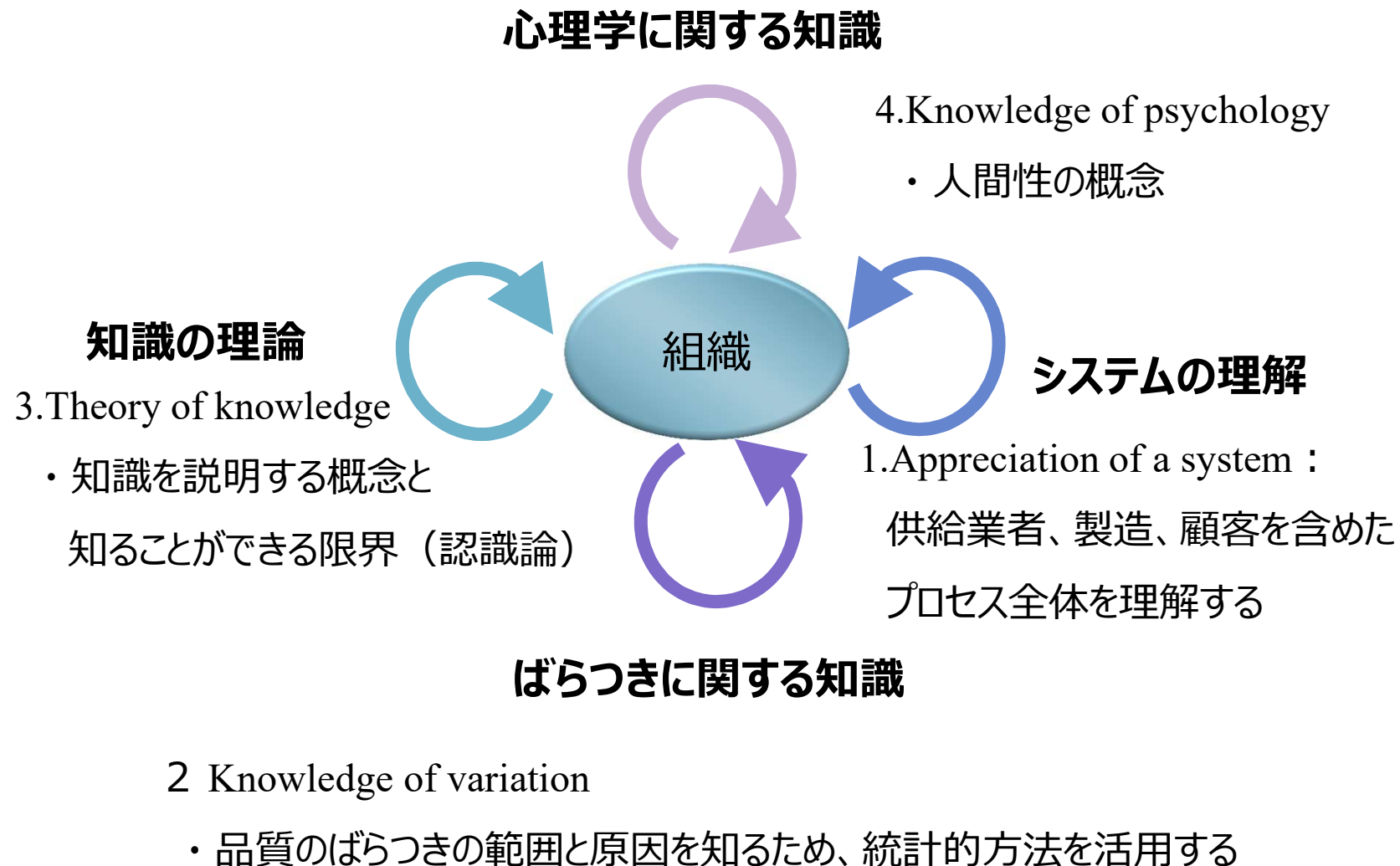
3.10 システム思考で因果仮説を説明



デミングは、「上司は部下の評価をつけることが仕事ではなく、誤りを犯す環境を取り除き部下によい仕事をしてもらうために存在する」と述べている



3.11 デミング博士の深遠なる知識の体系



3.12 統計ソフトウェアが持つべき能力

1. 利便性---誰もが自由に使いこなせる道具
2. 客観性---科学的な仮説検証
3. 作業性---問題解決の思考を邪魔しない
4. 伝達性---直感的なデータの可視化
5. 正確性---単なる集計では見えない事や錯覚を暴く
6. 学習性---モデリング志向により組織知の成長を促す
7. 蓄積性---報告資料の標準化をサポート

3.13 分析者が気を配るべきポイント

1. 無作為化 --- データの素性を明らかにする
強く意識しない限り無作為標本は得られない
2. 限界 --- 扱う手法の前提条件を確認する
現実のデータは手法の前提を満たすものなど得られない
3. 擬似相関 --- 背後に本質的な要因が潜んでいないか考える
調査研究では要因系の変数が手に入ることは困難である
4. 層別 --- 要因系で結果を層別する
調査研究ではワルドが指摘したように選択バイアスを受ける変数が多い
5. 仮説検証 --- 有意差の意味と実質的な違いを認識する
標本数を増やせば無意味な僅かな差も有意になる
6. 予測精度 --- 過去の情報で将来を予測する限界を理解する
統計モデルに完成はなく、データが得られるたびに成長する宿命にある
7. 散布図 --- 相関関係と回帰関係を区別する
調査研究のデータから描かれる散布図の各点は様々な素性を持つ

本著作物は原著作者の許可を得て、株式会社日本科学技術研修所（以下弊社）が掲載しています。本著作物の著作権については、制作した原著作者に帰属します。

原著作者および弊社の許可なく営利・非営利・イントラネットを問わず、本著作物の複製・転用・販売等を禁止します。

所属および役職等は、公開当時のものです。

■公開資料ページ

弊社ウェブページで各種資料をご覧ください <http://www.i-juse.co.jp/statistics/jirei/>

■お問い合わせ先

(株)日科技研 数理事業部 パッケージサポート係 <http://www.i-juse.co.jp/statistics/support/contact.html>