

sample



Minitab 多変量解析

株式会社構造計画研究所
Minitabチーム

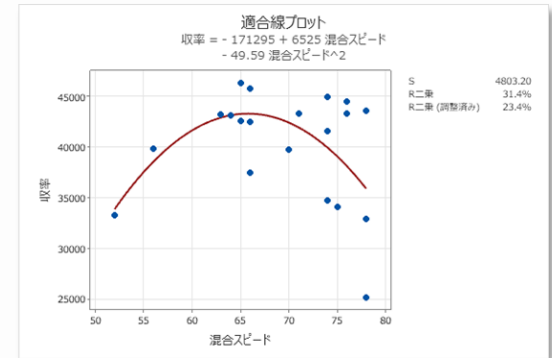
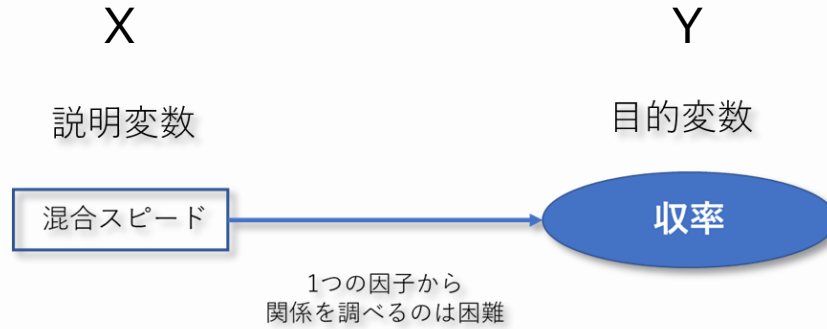
- 多変量解析
 - 単変量と多変量
 - 多変量解析の種類
 - 散布図の例と使い方
 - 例：燃費データ
- 回帰分析
 - 単回帰分析
 - 重回帰分析
 - 例：発泡酒の売上
- 主成分分析
 - 主成分分析とは
 - 主成分分析がやっていること
 - 主成分分析の幾何学的なイメージ
 - 例：環境変数
- PLS（部分的最小二乗回帰）
 - PLSとは
 - 例：運転条件の調査

- 因子分析
 - 主成分分析と因子分析の違い
 - 因子分析の手法
 - 例：求職者の評価
- クラスタリング
 - クラスタリングとは
 - 階層クラスタリングと非階層クラスタリング
 - 階層クラスタリング：Ward法
 - 例：運転条件の調査 Ward法
 - 非階層クラスタリング：K-means法
 - 例：運転条件の調査 K-means法
- 決定木（CART）
 - 機械学習とは
 - 教師あり学習、教師なし学習
 - 機械学習の様々な手法
 - 決定木とは
 - 例：パルプの漂白、ガラスの製造工程

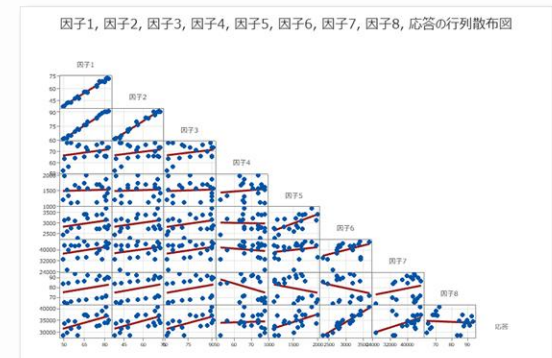
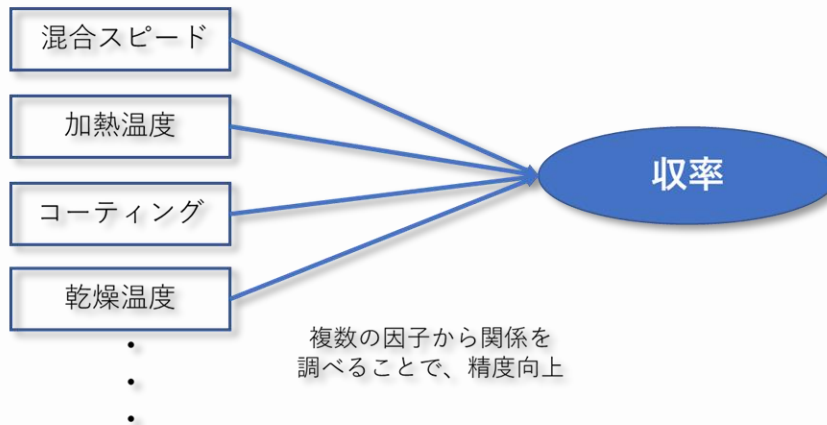
多変数を可視化する

多変量解析の概要と可視化

単変量



多変量



一方、分析が複雑になるため
パラメータ推定のテクニックが必要

多変量解析

要約

大量の入力データを数学的に効率的に処理し、回帰や分類の予測精度向上に役立てる。

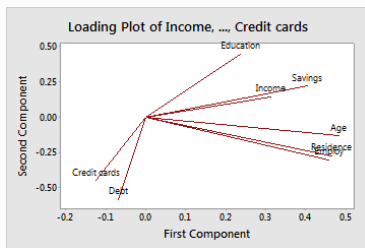
分類

入力（あるいは、出力でも可）の類似するグループを作り、管理を効率化する。

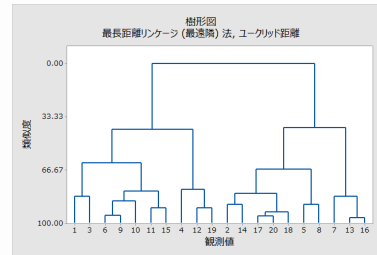
予測

入力に対する出力の関係をモデル化し、影響度の評価や出力の予測を行う。

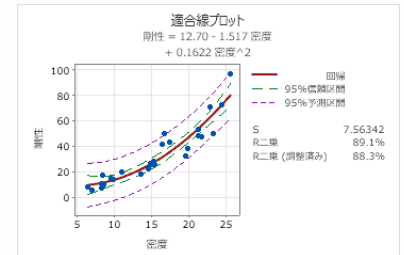
主成分分析



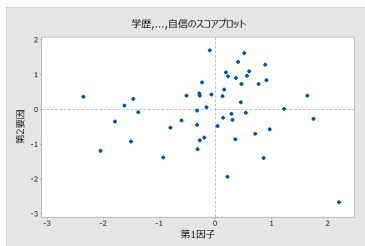
クラスター分析



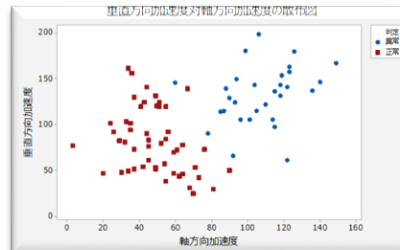
回帰分析



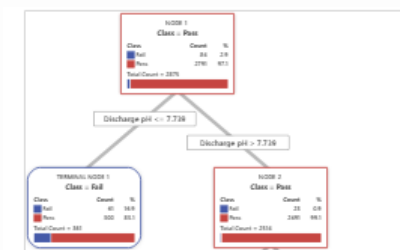
因子分析



判別分析



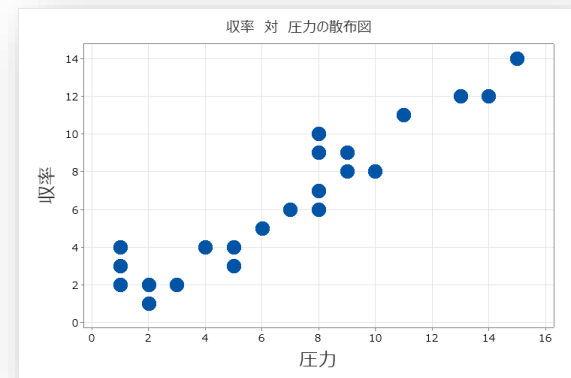
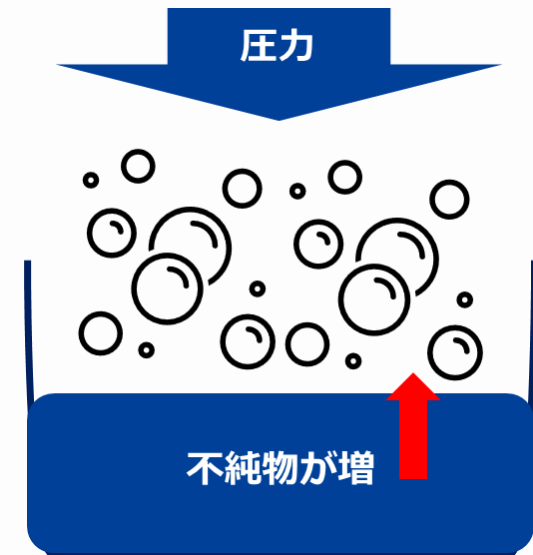
決定木(CART)



回帰分析は予測や多変量の相関関係を調べることで使われることが多いです。今回のように要因を探索するときの回帰分析には注意が必要です。

まず、大前提として「相関関係は必ずしも因果関係を意味しない」ことは基本となります。

- 化学工程の作業標準のひとつに「泡が出た時には、圧力を増やせ」とあります。※1
- 不純物が増すと泡が発生するメカニズムになっており、圧力をかけると泡は現象します。
- 不純物は生成物の収率低下に繋がります。
- 圧力と収率の相関は高く、回帰分析を行うと関係性があるように思えますが、**圧力が収率の因果関係があるわけではない**点に注意です。



※1 参考図書 朝倉書店『統計的因果推論 -回帰分析の新しい枠組み-』 著 宮川 雅巳

次元を縮約し、抽象的な表現を得る

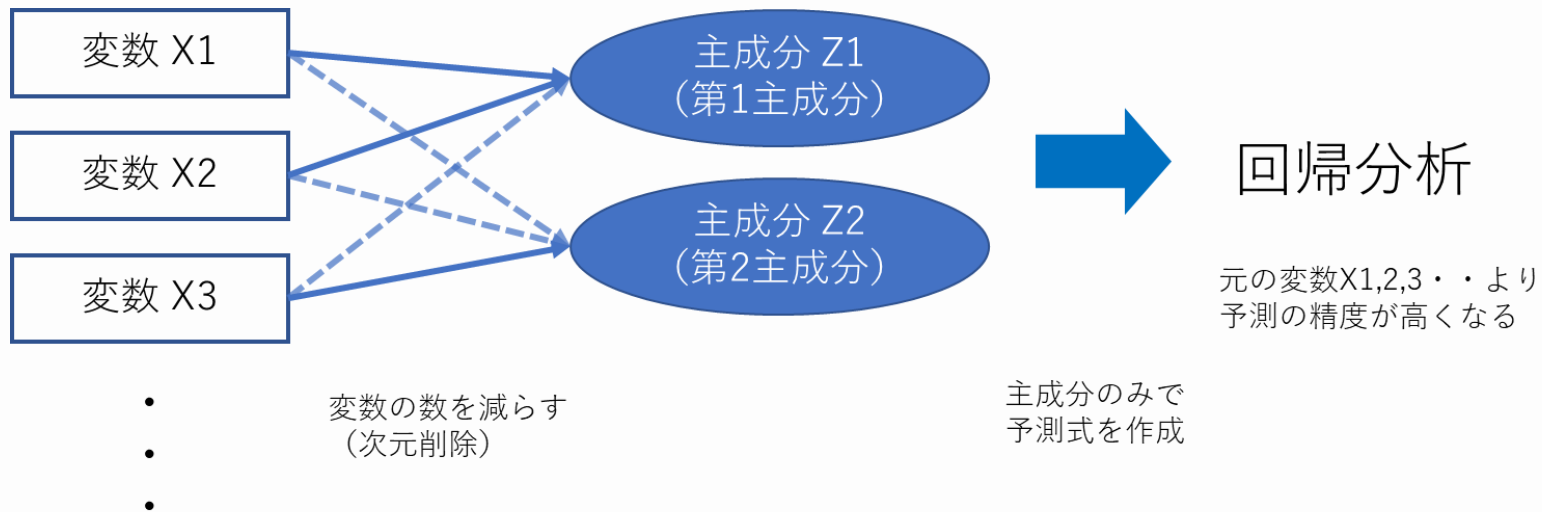
主成分分析

■ 主成分分析とは

主成分分析は、解析しようとしている多次元データを、そこに含まれる情報の損失をできるだけ少なくして低次元のデータに縮約する手法です。

主に次のような目的で使われます。

- 多数の変数を統合して、新たに総合的な変数を作成する。
- 観測対象のグループ分けをする。
- 回帰分析のような他の解析手法を併用して、データを別の観点から吟味する。



学籍番号	数学	英語	物理	化学	国語
1	77	71	73	70	76
2	56	63	54	59	72
3	61	60	62	62	59
4	68	78	66	60	77
5	71	75	74	72	66
6	87	81	88	86	68
7	62	64	62	61	63
8	80	92	77	71	61
9	80	61	83	93	63
10	65	56	66	70	61
11	67	62	71	74	60
12	56	68	57	60	95
13	62	66	63	66	72
14	80	76	79	68	80

元データ：
14人の生徒の5科目のテストの点数

生徒の学力に関して共通する特徴は？

学籍番号	数学	英語	物理	化学	国語
1	77	71	73	70	76
2	56	63	54	59	72
3	61	60	62	62	59
4	68	78	66	60	77
5	71	75	74	72	66
6	87	81	88	86	68
7	62	64	62	61	63
8	80	92	77	71	61
9	80	61	83	93	63
10	65	56	66	70	61
11	67	62	71	74	60
12	56	68	57	60	95
13	62	66	63	66	72
14	80	76	79	68	80

元データ
(14人の生徒の5科目のテストの点数)



主成分分析を行うと...

学籍番号		
1	0.57531	0.63369
2	-2.41045	-0.12734
3	-1.38068	-1.23551
4	-0.62745	1.32902
5	0.69717	0.06062
6	3.23801	0.39126
7	-1.32278	-0.65725
8	1.94123	0.96350
9	2.37711	-1.55472
10	-0.69169	-1.53909
11	0.10693	-1.30993
12	-2.45236	1.76060
13	-1.12048	-0.05529
14	1.07013	1.34044

主成分スコア(主成分得点)

学籍番号	数学	英語	物理	化学	国語
1	77	71	73	70	76
2	56	63	54	59	72
3	61	60	62	62	59
4	68	78	66	60	77
5	71	75	74	72	66
6	87	81	88	86	68
7	62	64	62	61	63
8	80	92	77	71	61
9	80	61	83	93	63
10	65	56	66	70	61
11	67	62	71	74	60
12	56	68	57	60	95
13	62	66	63	66	72
14	80	76	79	68	80

元データ
(14人の生徒の5科目のテストの点数)

$$\text{理系能力} = 5.6\text{数学} + 0.1\text{英語} \\ + 4.5\text{物理} + 2.1\text{化学} + 0.2\text{国語}$$



学籍番号	理系能力	文系能力
1	0.57531	0.63369
2	-2.41045	-0.12734
3	-1.38068	-1.23551
4	-0.62745	1.32902
5	0.69717	0.06062
6	3.23801	0.39126
7	-1.32278	-0.65725
8	1.94123	0.96350
9	2.37711	-1.55472
10	-0.69169	-1.53909
11	0.10693	-1.30993
12	-2.45236	1.76060
13	-1.12048	-0.05529
14	1.07013	1.34044

主成分スコア(主成分得点)

学籍番号	数学	英語	物理	化学	国語
1	77	71	73	70	76
2	56	63	54	59	72
3	61	60	62	62	59
4	68	78	66	60	77
5	71	75	74	72	66
6	87	81	88	86	68
7	62	64	62	61	63
8	80	92	77	71	61
9	80	61	83	93	63
10	65	56	66	70	61
11	67	62	71	74	60
12	56	68	57	60	95
13	62	66	63	66	72
14	80	76	79	68	80

元データ
(14人の生徒の5科目のテストの点数)

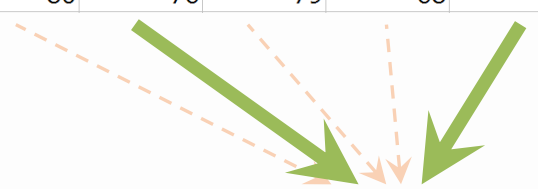
$$\text{文系能力} = 5.6\text{数学} + 5.6\text{英語} \\ + 4.5\text{物理} + 2.1\text{化学} + 6.4\text{国語}$$

学籍番号	理系能力	文系能力
1	0.57531	0.63369
2	-2.41045	-0.12734
3	-1.38068	-1.23551
4	-0.62745	1.32902
5	0.69717	0.06062
6	3.23801	0.39126
7	-1.32278	-0.65725
8	1.94123	0.96350
9	2.37711	-1.55472
10	-0.69169	-1.53909
11	0.10693	-1.30993
12	-2.45236	1.76060
13	-1.12048	-0.05529
14	1.07013	1.34044

主成分スコア(主成分得点)

学籍番号	数学	英語	物理	化学	国語
1	77	71	73	70	76
2	56	63	54	59	72
3	61	60	62	62	59
4	68	78	66	60	77
5	71	75	74	72	66
6	87	81	88	86	68
7	62	64	62	61	63
8	80	92	77	71	61
9	80	61	83	93	63
10	65	56	66	70	61
11	67	62	71	74	60
12	56	68	57	60	95
13	62	66	63	66	72
14	80	76	79	68	80

5つの詳細なデータをすべて見なくても、たった2つの抽象的なデータを見るだけでも各生徒の特性(得意・不得意)を簡単に理解することができます。



学籍番号	理系能力	文系能力
1	0.57531	0.63369
2	-2.41045	-0.12734
3	-1.38068	-1.23551
4	-0.62745	1.32902
5	0.69717	0.06062
6	3.23801	0.39126
7	-1.32278	-0.65725
8	1.94123	0.96350
9	2.37711	-1.55472
10	-0.69169	-1.53909
11	0.10693	-1.30993
12	-2.45236	1.76060
13	-1.12048	-0.05529
14	1.07013	1.34044

機械学習の基礎を知ろう

決定木 (CART)

機械学習とは、与えられた問題や課題に応じてコンピューター自身が学習し、解決を行う仕組み全体のことをいいます。

機械学習と共に使われる言葉として**人工知能 (AI)**がありますが、これは幅広い意味で使われ、大まかに「人間と同じような処理を行える技術」と考えることができます。機械学習は人工知能の分類の1つです。

人工知能

機械学習

- ・教師あり学習
- ・教師なし学習
- ・強化学習

機械学習には様々な手法がありますが、データに対して、次のいずれかに分類されます。

- **教師あり学習**

- 問題の答えが含まれたデータを使って機械学習のモデルを学習させる手法

- **教師なし学習**

- 与えられたデータから法則や構造を自動的に抽出する手法

- **強化学習**

- ある環境下で報酬を最大化するように学習していく手法

データの扱い、意図で変わってきます。

統計は、結果の**根拠を示す場合に効果を発揮し**、

機械学習は、**予測する時に効果を発揮**します。

ただ、両者は明確に区別できるものではなく、技術的にも使われるシーンも重なる部分が多いです。製造業においては、なぜこのような結果になったのかという原因を説明するときには統計手法(SQC手法)を使い、従来よりもより精度の高い予測を行いたい(従来の手法で思ったような結果がでない) 時には機械学習を使っていくシーンが多いようです。

機械学習が優れており、SQCが不要ということはありませんので、問題解決のツールとして両方をうまく使い分けていく必要があります。

教師あり

手法名	分類	回帰
回帰	-	○
SVM(サポートベクターマシン)	○	○
決定木	○	○
ニューラルネットワーク	○	○
kNN	○	○

教師なし

手法名	クラスタリング	次元削除
主成分分析(PCA)	-	○
LSA	-	○
NMF	-	○
K-means法	○	-

- **決定木**とは、特徴量の値に対して条件を「Yes or No」の条件によって予測を行う機械学習です。
- 右図のような枝分かれで結果が表示されるので、学習結果の理解・解釈がしやすいのが特徴です。
- 数値データとカテゴリデータの両方を使うことができます。

Alternative Tree Diagram

