第8回メモ

**ハッシュタグ：#ybkstat**

因子分析

# 分析手法の分類

重回帰分析

|  |  |  |  |
| --- | --- | --- | --- |
| ラベル | 説明変数1 | 説明変数2, ... | 目的変数 |
|  | 量的データ | | 量的データ |

Rではlm()を使って分析する。

数量化理論I類

|  |  |  |  |
| --- | --- | --- | --- |
| ラベル | 説明変数1 | 説明変数2, ... | 目的変数 |
|  | 質的データ | | 量的データ |

Rではlm()を使って分析する。

判別分析

|  |  |  |  |
| --- | --- | --- | --- |
| ラベル | 説明変数1 | 説明変数2, ... | 目的変数 |
|  | 量的データ | | 質的データ |

Rでは目的変数を数値に変換してからlm()を使って分析する。

数量化理論II類

|  |  |  |  |
| --- | --- | --- | --- |
| ラベル | 説明変数1 | 説明変数2, ... | 目的変数 |
|  | 質的データ | | 質的データ |

Rでは目的変数を数値に変換してからlm()を使って分析する。

ロジスティック回帰分析

|  |  |  |  |
| --- | --- | --- | --- |
| ラベル | 説明変数1 | 説明変数2, ... | 目的変数 |
|  | 質的データ | | 量的データ（範囲限定） |

Rでは目的変数をロジットに変換してからlm()を使って分析し、ロジスティック関数で目的変数に戻す。

主成分分析

|  |  |  |
| --- | --- | --- |
| ラベル | 変数1 | 変数2, ... |
|  | 量的データ | |

Rではprcomp()を使って分析する。

因子分析

|  |  |  |
| --- | --- | --- |
| ラベル | 変数1 | 変数2, ... |
|  | 量的データ | |

Rではfactanal()を使って分析する。

# 因子分析

## 主成分分析と因子分析

{太郎, 花子}の{国語, 英語, 数学}の試験の得点データがあったとしよう。

|  |  |  |  |
| --- | --- | --- | --- |
|  | 国語 | 英語 | 数学 |
| 太郎 | 国語太郎 | 英語太郎 | 数学太郎 |
| 花子 | 国語花子 | 英語花子 | 数学花子 |

変数が3つではわかりにくいので（ということにして）、変数（スコア）2つで成績を表現したい。

考え方は2とおり。

### 主成分分析の考え方

主成分スコア{1, 2}{太郎, 花子}＝主成分{1, 2}国語×国語{太郎, 花子}＋主成分{1, 2}英語×英語{太郎, 花子}＋主成分{1, 2}数学×数学{太郎, 花子}

行列で書けば次のようになる。

### 因子分析の考え方

{国語, 英語, 数学}{太郎, 花子}＝因子負荷量1{国語, 英語, 数学}×因子スコア1{太郎, 花子}

＋因子負荷量2{国語, 英語, 数学}×因子スコア2{太郎, 花子}

行列で書けば次のようになる。

|  |  |  |
| --- | --- | --- |
|  | 主成分分析 | 因子分析 |
| 考え方の共通点 | 少数の変数でデータを説明する試み。 | |
| 考え方の相違点 | データを外から見る | データを内から見る |
| Rでの解析方法 | prcomp() | factanal() |
| 得られる指標 | 主成分  myResult$rotation | 因子負荷量  myResult$loadings |
| 変換されたデータ | 主成分得点（スコア）  myResult$x | 因子得点（スコア）  myResult$scores |

## Rでの因子分析

「テスト結果」のデータを因子分析してみよう。

setwd("c:/cit")

myData <- read.csv("テスト結果（5科目28人）.csv")

head(myData)

#因子分析を行う

myResult<-factanal(myData[,-1],factors=2,scores="regression",rotation="promax")

#因子負荷量

#結果に含まれている情報は「attributes(myResult)」で確認できる。

myResult$loadings

#因子負荷量の図示（この結果を解釈する。）

barplot(sort(myResult$loadings[,1]))

barplot(sort(myResult$loadings[,2]))

#因子得点

myResult$scores

#因子得点の図示

biplot(myResult$scores, myResult$loadings)