

Rによるやさしい統計学

回帰分析と因子分析

著者

山田剛史

杉澤武俊

村井潤一郎

回帰分析

- 一つもしくは複数の変数の値をつかって他の変数を求めるための式をつくるためのもの

例 娘の身長

- 身長というのは遺伝する傾向があるので父の身長と母の身長から娘の身長を求める式を作りたいとする
- 父と母の身長から娘の身長を求める式
娘の身長 $= a + b_1 \times \text{父の身長} + b_2 \times \text{母の身長} + e$
- この時 a を切片、 b_1 、 b_2 を編回帰係数(傾き)、 e を残差(誤差)とする。この様な式を多くのデータを用いて求めることが回帰分析の目的である

語句説明

- 独立変数(予測変数、説明変数とも)
 - 「父の身長」のように「娘の身長」を求めるために使われる変数のこと
- 従属変数(基準変数、目的変数とも)
 - 「娘の身長」のことで求めたい変数のこと
- 重回帰分析
 - 変数を複数つかってする回帰分析のこと
- 単回帰分析
 - 変数を一つのみ用いてする回帰分析のこと
- 回帰直線
 - 求めた式をデータの図に引いたときの直線

標準編回帰係数

- これはどの独立変数が最も従属変数に影響を与えているかというものを調べるための数値で
- 編回帰係数(傾き) $\times \frac{\text{独立変数の標準偏差}}{\text{従属変数の標準偏差}}$
- という式で求めることができるこの値が高いほどその独立変数は従属変数に大きく影響を与えていることが理解できる

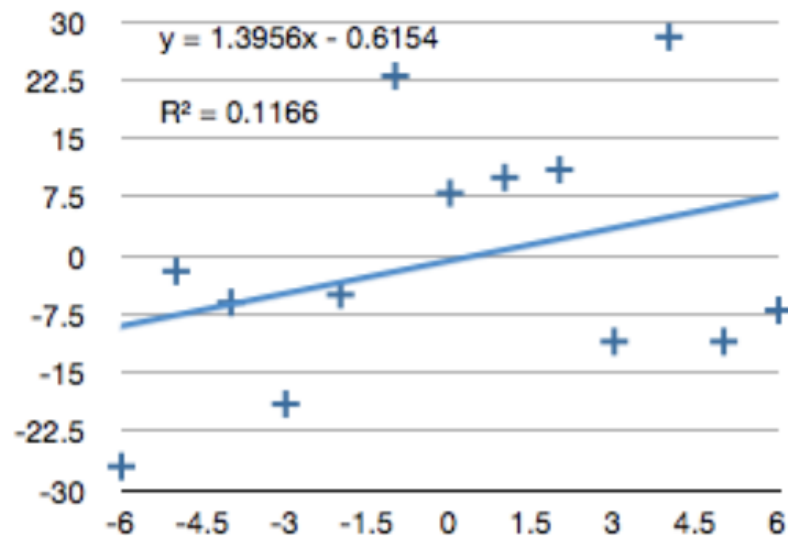
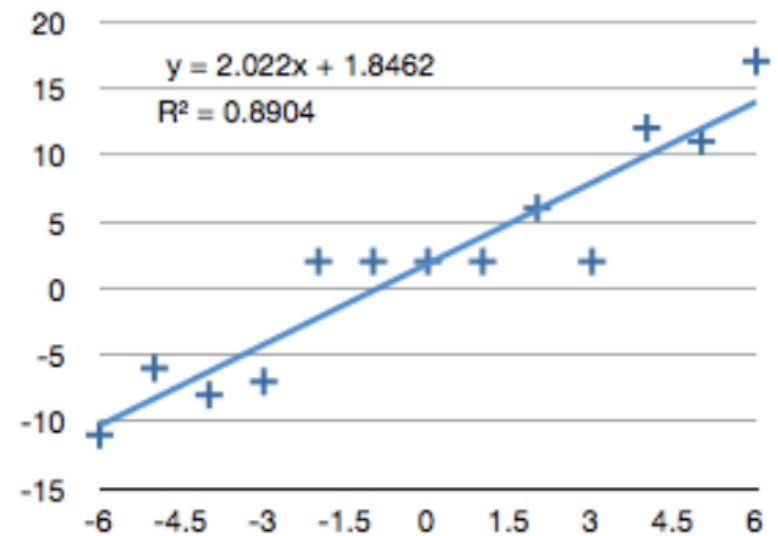
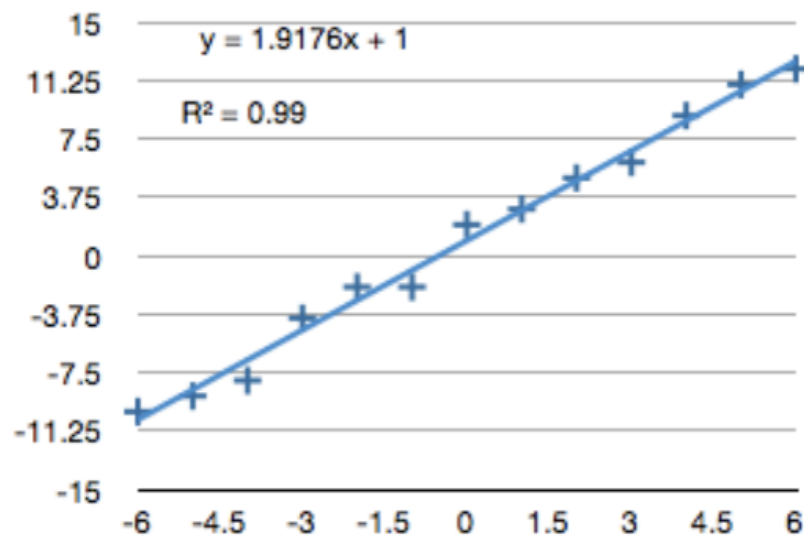
相関係数

- 独立変数が従属変数に一か＋にどの位の影響を与えているのかを-1から+1の値で数値化したもの

- 相関係数 =
$$\frac{\text{共分散}}{x\text{の標準偏差} * y\text{の標準偏差}}$$

決定係数

- 相関係数の二乗で求めることができ、回帰分析を行って求めた式がどの程度説得力があるかどうかを見るための指標
- 0から1の値で決められていてこの値が1に近いほど全体のデータに近い式を求めることができる



因子分析

- 複数の変数がありその複数の変数の一部分に共通して影響する因子を仮定し、複数の変数間の相関関係を説明すること

例 五教科のテスト成績

- (国語、社会、数学、理科、英語)の点数をデータとして羅列していくと一般的に国語の成績がよければ社会の成績がよい傾向があり、数学の成績がよければ理科の成績がよい傾向があった
- これは国語、社会の間に文系因子、数学、理科の間に理系因子があるのではないかと考え分析を行っていくこと

語句説明2

- 因子負荷
 - 求めた因子がそれぞれの変数にどれくらいの影響があるのかを数値化したもので値域は-1から1の間の値
- 独自性
 - 国語などの調べたい変数がどれだけ独立しているか（因子の影響を受けないか）の程度を示したもので値域は0から1
- 共通性
 - 1－独自性で求められる値で変数がどの程度因子の影響を受けているかの程度を示したものの値域は独自性と同じ0から1

語句説明3

- 因子寄与

- 変数それぞれの因子負荷を二乗した和であり全体の
変数に対して因子がどの位影響しているか求めるため
のもの
- 値域は0から変数の要素数、テストの例で言えば五科
目の要素を使ったので最大で5になる

- 因子寄与率

- 因子寄与を変数の要素で割り%に直したもの、この値
を見ればこの因子が全体の何%影響してるかわかる

まとめ

- 回帰分析はいくつかの変数からデータを求めるための式を求めること
- 因子分析はいくつかの変数にある関係性(因子)があるのではないかと推測し、その推測が正しいのかをもとめるために検証すること