

多変量解析

久保田知恵

2018 年 4 月 18 日

1 第 1 回

2 第 2 回：回帰分析 1 <単回帰と最少 2 乗法>

2.1 回帰モデル

2.2 線形解析モデル

現実的に考えれば線形で考えても問題ないのではないかというモデル。例：日本人の身長と体重
説明変数が 1 個だけの場合を線形単回帰モデル、複数個の場合を重回帰モデルという。

2.3 線形単回帰モデル

$Y = \alpha + \beta X$ (X, Y) の観測値 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ 例：熊谷の 1 日の最高気温と最低気温一
年分を調べてみよう
変数 1 個だけで表して見たい。

2.4 回帰パラメータの推定

$y_1 = \alpha + \beta x_1$ と考えて行きたいが、観測値には誤差があるので ϵ_1 を「残差」として導入してあげ
る。ハットがついたものは推定値ということは、 $\hat{y}_1 = Y - \epsilon_1$ ハットが成り立つ。

2.5 残差に関する仮定

統計には仮定がつきもの。この先の議論を進めるために、残差についていくつかの仮定を施す。

- 残差の期待値は 0 であることを認める。誤差はマイナスにもプラスにも触れるということは、その平均
を見た場合、その値は 0 になるであろうという考え。もし、0 でなかったら、その定数分を α に足し
てあげれば良い。
- 残差の出方（分散）は一定（定数）であってほしい。
- 相異なる個体の残差は無相関である。誤差同士には何も相関はないよね、という考え方。それよりも今
は Y と X の相関について調べている。

2.6 残差の評価尺度

残差の評価尺度には次の3つがあげられる。

残差：residual

- 残差平方和残差の二乗和
- 絶対残差和1つ1つ
- 最大絶対残差一番大きな残差が出ているところが小さくなるように働きかける。

2.7 最少2乗法

なぜ、RSSがよく使われるのか。凸関数であるため、極小値がひとつだけである。もう1つは、2次導関数が常に正である。

課題：RSSを α, β どちらについても偏微分に行い形を整える。求めるべき α, β に*をつけ、連立方程式として解く。

2.8 回帰分析の語源

並外れた値を集めて、もう一度サンプルをとると、平均に回帰する。一般的に、回帰分析の対象は相関が極めて強いデータに対して用いられる。

3 第3回：

予習：重回帰モデルについて、ベクトル・行列を用いて連立式を解けるように。