

変数変換

落とし穴

変数変換

重回帰 & 二元以上（二元も含む）の配置の分散分析
独立変数（要因）が複数

変数変換



存在しない交互作用の検出
変数の効果の強さ（偏回帰係数）：誤って推定
一般にサンプル数が大きいほど症状は悪化

しつこいですが・・・

統計的方法は 検定だけではない

検定・推定（有意確率、危険率、信頼度）

不等分散とU検定・ランダムイゼーション検定

予測の最適化、モデル選択

変数変換

変数変換する 2 つの理由

分散安定化である変換が選ばれる理由 デルタ法

角度変換

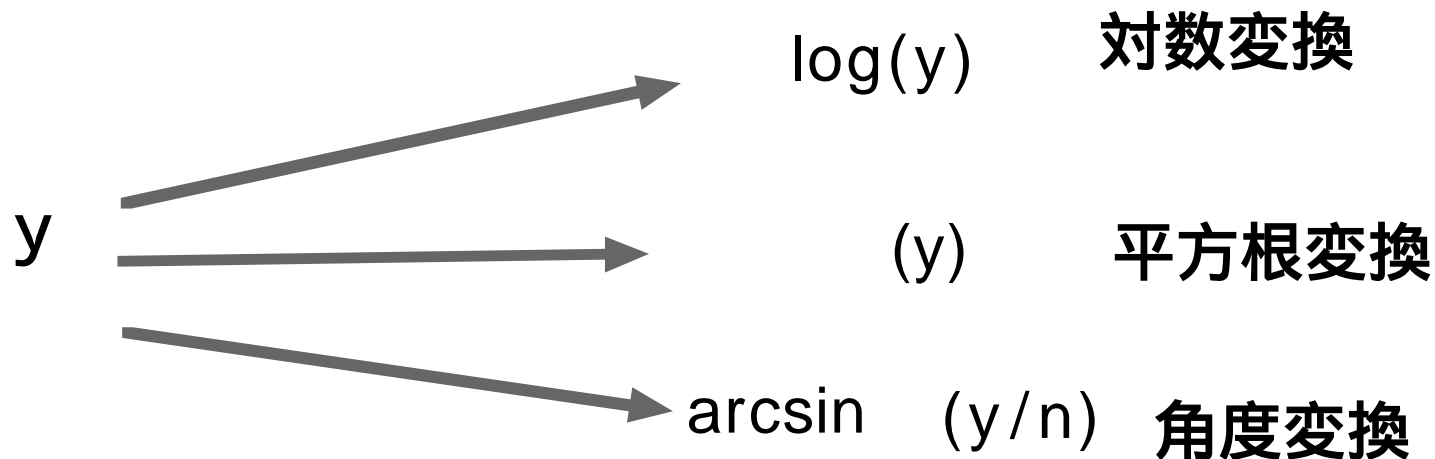
平方根変換

結果とその意味

対策

変数変換 (transformation) とは

データ **そのまま使わず**
データのある関数を使う



変数変換の目的

想定している関係

$$y = e^{b_1 x_1 + b_2 x_2} = e^{b_1 x_1} e^{b_2 x_2}$$

$y = x$ の一次式しか解析に使えない

対数をとる

x y の期待値

独立変数 従属変数の期待値

変数変換の目的

y のばらつきが y の値によってちがう

y のばらつきが一定でないとは解析に使えない

分散安定化の変数変換

y のばらつきを一定にする

従属変数の分散を一定にする

変数変換の目的

独立変数 従属変数の期待値
という関係を変える

従属変数の分散を一定にする

$y = x$ の一次式、しか解析に使えない

y のばらつきが一定でないとは解析に使えない

パラメトリックな方法

重回帰

分散分析

t 検定

デルタ法

テイラ - 展開に基づく一次（線形）近似

本当は曲線関係
なんだけど



直線関係で近似
してしまおう

分散などの値を近似的に求める

デルタ法

一次式と分散の値

$$(cy + d \text{の分散}) = c^2 (y \text{の分散})$$

y の関数 $f(y)$ 分散などの値を近似的に求める

$$(f \text{の分散}) \quad \left(\frac{f}{y} \right)^2 \times (y \text{の分散})$$

角度変換

二項比率 確率 と 確率 (1 -)

2つの結果

n : 全体

y : 部分

割合 (y/n) 分散は (1 -) / n

が変化しても分散が一定になる
変換を考える

角度変換

二項比率

n : 全体

y : 部分

$$\arcsin (y/n)$$

角度変換 逆正弦変換

逆正弦平方根変換

角変換

angular transformation

$$\text{分散が } \frac{1}{4n}$$

角度変換

n : 全体

y : 部分

$$\arcsin (y/n)$$

分散はだいたい $1 / (4n)$

nが不揃いなら、不等分散

ポアソン変数

単位時間あたりのイベント発生確率が一定
一定時間あたりのイベント発生回数

平均値 = 分散

$$(\text{一定}) = \left(\frac{f}{y} \right)^2 y$$

となる f を考える

平方根変換

y : 事象の回数

ポアソン変数

(y)

角度変換

$$\arcsin (y / n)$$

変換してから、分散分析・回帰
(パラメトリックな方法)

割合・比率などの分析に広く使われてきた