

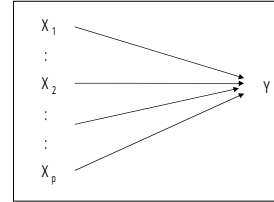
単回帰分析と重回帰分析

M1 栗本達児

1

回帰分析とは

- 変数 Y を、 p 個の変数 X_1, X_2, \dots, X_p により説明したり予測するための統計的方法



X : 独立変数、説明変数、あるいは予測変数
 Y : 従属変数または基準変数

2

回帰分析とは(2)

- 独立変数が1個($p = 1$)の場合を単回帰分析、複数個($p \geq 2$)の場合を重回帰分析と呼ぶ
 - 単回帰分析

独立変数 X 説明・予測 従属変数 Y

3

重回帰分析とは

- 2つ以上の独立変数により、1つの従属変数を予測・説明しようとする
- 多変量解析の一手法
 - 従属変数、独立変数ともに量的データである場合に用いる
 - ダミー変数(0と1、など)を含んでもよい

4

単回帰・重回帰分析の関係

- 単回帰分析を、説明変数が2つ以上の場合に拡張したものが重回帰分析
- 基本的な部分は共通しているが、拡張することで重回帰分析特有の考え方、注意点も発生する

まずは単回帰分析を理解しよう

5

単回帰分析

6

単回帰分析の流れ

- 2変数間に回帰モデル(回帰式あるいは予測式)を想定
 - 最小二乗法により、YのXへの回帰直線を得る
- 回帰式の精度(予測の精度)を評価する

7

回帰式

- 単回帰分析では、独立変数xと従属変数yの間に以下の一次式のような線形関係を考える

$$\hat{y} = a + bx \quad (\text{回帰式})$$

- \hat{y} : 予測値
- a : 切片
- b : 傾き(回帰係数)

8

定数・係数の決定

- 2変数間に完全な直線関係が存在する場合はa, bの値は簡単に決まるが、通常ありえない
- 従属変数の実測値と予測値との誤差($y - \hat{y}$)に着目。この2乗和が最小になるようなa, bの値を求める

最小二乗法。a, bは次式により求められる

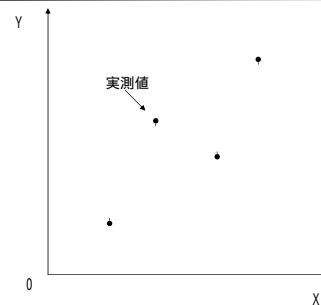
$$a = \bar{y} - b\bar{x}$$

$$b = S(x, y) / S^2(x) \quad \text{または} \quad r(x, y) S(y) / S(x)$$

こうして得られた直線をyのxへの回帰直線という

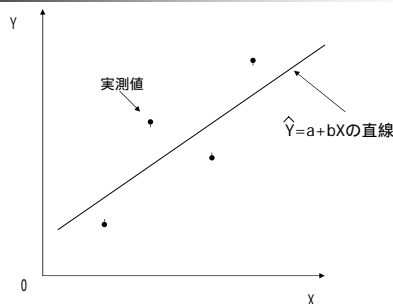
9

YのXへの回帰直線



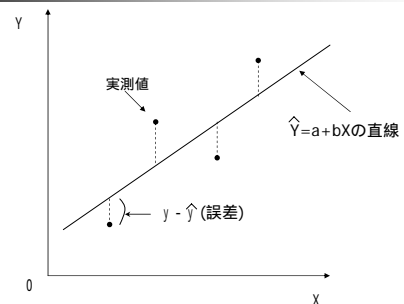
10

YのXへの回帰直線



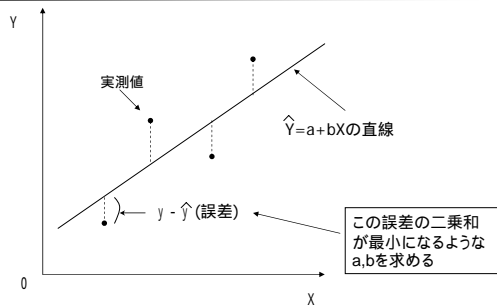
11

YのXへの回帰直線



12

YのXへの回帰直線



13

予測の精度

- 回帰式により予測値を得ても、実測値とかけ離れているようでは役に立たない
 - 回帰式の精度をあらわすのは決定係数
決定係数 = $S^2(\hat{y}) / S^2(y) = r(x, y)^2$
(cf. $S^2(y) = S^2(\hat{y}) + S^2(e)$, eは誤差)
- 全変動のうち回帰によって説明できる変動の割合

重回帰分析へ

14

重回帰分析

15

重回帰分析の流れ

- 独立-従属変数間に重回帰モデル(重回帰式あるいは予測式)を想定
- 重回帰式の精度(予測の精度)を評価する
- 各独立変数の影響の大きさを検討する
 - 標準偏回帰係数の算出

16

重回帰式

- 重回帰分析では、複数個の独立変数 x_i ($i = 2$)と、従属変数 y の間に、以下の一次式のような関係を考える

$$\hat{y} = a + b_1x_1 + b_2x_2 + \dots + b_ix_i \text{ (重回帰式)}$$

- \hat{y} : 予測値
- a : 回帰定数
- b : 偏回帰係数

17

定数・係数の決定

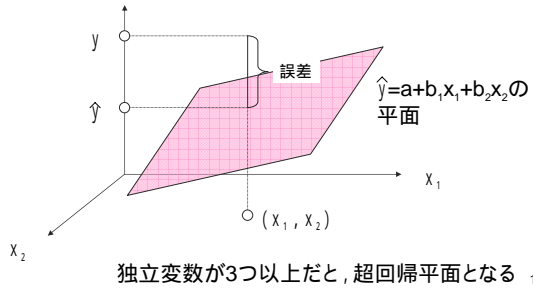
- 最小二乗法を用いて、従属変数の実測値と予測値との誤差($y - \hat{y}$)の2乗和が最小になるような回帰定数、偏回帰係数を求める

独立変数が2つの場合、 y の x への回帰平面が得られる

独立変数が3つ以上だと超回帰平面

18

yのxへの回帰平面



予測の精度

- 回帰式の精度をあらわす指標
 - 重相関係数 (R)
 - 最小二乗法により回帰定数、偏回帰係数が決定されたが、このとき実測値と観測値の相関は最大になっている
 - このとき、両者の相関係数の絶対値を示すのがR (0 R 1)
 - 独立変数全体での影響力の大きさを示す
 - 重決定係数 = $S^2(\hat{y})/S^2(y) = R^2$
 - 全変動のうち回帰によって説明できる変動の割合

20

独立変数の影響の大きさ

- 重回帰分析では、独立変数相互の相関関係についても考慮し、その影響を取り除いて、各独立変数の従属変数に対する影響を純粋な形で取り出そうとする

21

独立変数の影響の大きさ(2)

- 偏回帰係数は、他の独立変数の影響を除いた残差変数の回帰係数
 - 測定単位の影響を受けており、この大きさから従属変数への影響力をいうことはできない

そこで求められるのが標準偏回帰係数

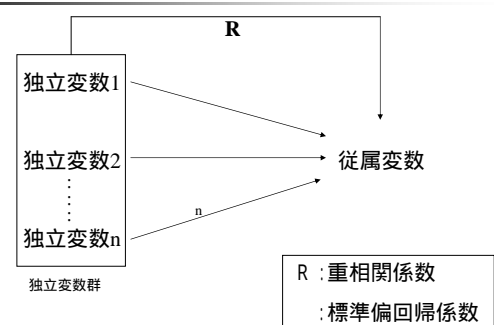
22

独立変数の影響の大きさ(3)

- 標準偏回帰係数 ()
 - 偏回帰係数に対し標準化を行ったもの
 - これにより他の変数の影響を取り除いたときの、独立変数の影響の大きさと向きを定量的に判断できる

23

モデル図



24

重回帰分析を行う際の注意点

- 多重共線性の問題
 - 安定した推定値が得られない
 - ブートストラップ法で確認できる
 - 抑制変数の原因となることもある
 - (従属変数との)単相関係数と(標準)偏回帰係数の符号が一致しない独立変数

25

重回帰分析を行う際の注意点(2)

- 独立変数間の相関が高いとき、
- 厳密には次式が成り立つとき、多重共線性があるという

$$a_0 + a_1x_{i1} + a_2x_{i2} + \dots + a_px_{ip} = 0$$

(ただし「 $a_1 \sim a_p$ のすべてが0」ではない)

 - a_jx_{ij} が他の重みつき変数の和で表現できる
 - ある独立変数がなくてもよいことを示す

26

重回帰分析を行う際の注意点(3)

- 多重共線性の検出方法
 - 独立変数間の重回帰分析
 - 独立変数の主成分分析

27

重回帰分析を行う際の注意点(4)

- 多重共線性が疑われるときは
 - 問題のある独立変数を除く
 - 相関係数の大きい独立変数を合成する
 - 研究目的に照らし、妥当な解釈が可能であればそのままでもよいこともある

28

例題: 中古マンションの価格

- 従属変数: 価格
- 独立変数: 広さ、築年数
 - 表1のデータにもとづき、
 - 価格は広さと築年数によって予測できるか
 - どちらの変数のほうが説明力があるか
 - 予測できるならその精度はどのくらいか
 - 同じ地区で広さ70m²、築年数10年のマンションが5800万円といわれたとき、この価格は妥当かなどを検討してみよう

29

表1 中古マンションのデータ

No.	広さ x_1 (m ²)	築年数 x_2 (年数)	価格 y (万円)
1	51	16	3000
2	38	4	3200
3	57	16	3300
4	51	11	3900
5	53	4	4400
6	77	22	4500
7	63	5	4500
8	69	5	5400
9	72	2	5400
10	73	1	6000

永田・棟近(2001)『多変量解析法入門』より

30

参考文献

- 服部環・海保博之(1996) Q & A心理データ解析 福村出版
- 大野高裕(1998) 多変量解析入門 同友館
- 永田靖・棟近雅彦(2001) 多変量解析法入門 サイエンス社
- 森敏昭・吉田寿夫(1990) 心理学のためのデータ解析テクニカルブック 北大路書房

31

参考HP

- 前原由喜夫 偏相関と重回帰分析 URL:
<http://kyoumu.educ.kyoto-u.ac.jp/cogpsy/personal/Kusumi/datasem04/mae.files/frame.htm>
- 堀啓造 SPSSときど記(211) URL:
<http://www.ec.kagawa-u.ac.jp/~hori/spss/tokidoki21.html>

32