応用社会統計学
回帰分析編

2012.11.20　　高木英至

■このプリントの範囲

- 回帰分析の原理
  - 単回帰分析
  - 重回帰分析
- 補足
  - ダミー変数
  - 相関係数と（標準）偏回帰係数の関係
  - 重相関係数と自由度
  - 偏回帰係数の検定－誤差の問題
  - 多重共線性
  - 変数選択

■回帰分析

- 単に相関をとることとは異なる
- 「**独立変数**」から「**従属変数**」を「説明」ないし「予測」する

$$x_1, x_2, \ldots, x_p \quad \rightarrow \quad y$$

独立変数 説明変数	従属変数 目的変数

- 回帰分析
  - 単回帰分析：独立変数が１個
  - 重回帰分析：独立変数が２個以上

回帰モデル

$$x_1, x_2, \ldots, x_p \quad \rightarrow \quad y$$

$$y = f(x_1, x_2, \ldots, x_p) + \varepsilon$$

- 線形回帰モデル

$$y = a + b_1 x_1 + b_2 x_2 + \ldots + b_p x_p + \varepsilon$$

- 線形単回帰モデル

$$y = a + bx + \varepsilon$$

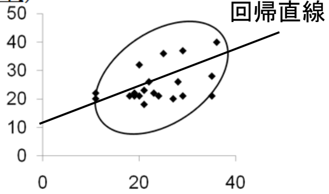
線形単回帰モデル

- n: 観測数　　i=1,…,n とすると、

$$y_i = a + bx_i + \varepsilon_i$$

a : 定数項　　b : 回帰係数

ε_i : 誤差（残差）



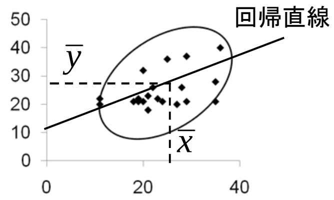
「最小二乗法」で係数の推定値を求めると

$$\hat{b} = \frac{S_{xy}}{S_{xx}}, \quad \hat{a} = \bar{y} - \frac{S_{xy}}{S_{xx}}\bar{x}$$

$$\hat{y}_i = \bar{y} - \frac{S_{xy}}{S_{xx}}\bar{x} + \frac{S_{xy}}{S_{xx}}x_i$$

$$\hat{y}_i - \bar{y} = \frac{S_{xy}}{S_{xx}}(x_i - \bar{x})$$

つまり、回帰直線は平均$(\bar{x}, \bar{y})$を通る。



決定係数

$$r_{xy} = r_{\hat{y}y} = \frac{\sum (y_i - \bar{y})(\hat{y}_i - \bar{y}) \Big/ (n-1)}{\sqrt{\frac{\sum (y_i - \bar{y})^2}{n-1} \cdot \frac{\sum (\hat{y}_i - \bar{y})^2}{n-1}}}$$

$$= \sqrt{\frac{\sum (\hat{y}_i - \bar{y})^2 / (n-1)}{\sum (y_i - \bar{y})^2 / (n-1)}}$$

決定係数　　$r_{xy}{}^2 = \frac{\sum (\hat{y}_i - \bar{y})^2 / (n-1)}{\sum (y_i - \bar{y})^2 / (n-1)}$

yの分散のうち、xによって説明される割合

分散の分割

従属変数yの分散は、予測値の分散と残差の分散の和になる

$s_y{}^2 = s_{\hat{y}}{}^2 + s_\varepsilon{}^2$　　(3･22)、p.61

決定係数を書き換えると、

決定係数　　$r_{xy}{}^2 = \frac{s_{\hat{y}}{}^2}{s_y{}^2}$

$s_y{}^2$は、r^2と$(1-r^2)$の比率で分割される。

$s_{\hat{y}}{}^2 = s_y{}^2 r^2$　　(3･23)

$s_e{}^2 = s_y{}^2(1-r^2)$　　(3･24)

■重回帰分析：基本的な考え方

- 単回帰モデル

$$y_i = a + bx_i + \varepsilon_i$$

- 重回帰モデル

$$y_i = a + b_1 x_{1i} + b_2 x_{2i} + \cdots + b_p x_{pi} + \varepsilon_i$$

- 基本的には単回帰と同じ
- a：定数項
- b_j：偏回帰係数
  - x_j 以外の変数の影響を統計的に除去している

重回帰モデル

$$y_i = \hat{a} + \hat{b}_1 x_{1i} + \hat{b}_2 x_{2i} + \cdots + \hat{b}_p x_{pi} + \varepsilon_i$$

$$\hat{y}_i = \hat{a} + \hat{b}_1 x_{1i} + \hat{b}_2 x_{2i} + \cdots + \hat{b}_p x_{pi}$$

$$\varepsilon_i = y_i - \hat{y}_i$$

最小二乗法（1）

$$y_i = \hat{a} + \hat{b}_1 x_{1i} + \hat{b}_2 x_{2i} + \cdots + \hat{b}_p x_{pi} + \varepsilon_i$$

$$\hat{y}_i = \hat{a} + \hat{b}_1 x_{1i} + \hat{b}_2 x_{2i} + \cdots + \hat{b}_p x_{pi}$$

$$\varepsilon_i = y_i - \hat{y}_i$$

$\hat{a}, \hat{b}_1, \hat{b}_2, \cdots, \hat{b}_p$の求め方＝最小二乗法

下記のQが最小になるように$\hat{a}, \hat{b}_1, \hat{b}_2, \cdots, \hat{b}_p$を求める。

$$Q(\hat{a}, \hat{b}_1, \hat{b}_2, \cdots, \hat{b}_p) \equiv \sum_{i=1}^{n} \varepsilon_i^{\,2} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$= \sum_{i=1}^{n} \{ y_i - (\hat{a} + \hat{b}_1 x_{1i} + \hat{b}_2 x_{2i} + \cdots + \hat{b}_p x_{pi}) \}^2$$

Qが最小となるには、次を充たすことが必要。

$$\partial Q / \partial \hat{a} = 0, \ \partial Q / \partial \hat{b}_1 = 0, \ \partial Q / \partial \hat{b}_2 = 0, \cdots, \partial Q / \partial \hat{b}_p = 0.$$

最小二乗法（2）

$$\partial Q / \partial \hat{a} = 0, \ \partial Q / \partial \hat{b}_1 = 0, \ \partial Q / \partial \hat{b}_2 = 0, \cdots, \partial Q / \partial \hat{b}_p = 0.$$

上の条件を整理すると：

$$\hat{a} = \bar{y} - (\hat{b}_1 \bar{x}_1 + \hat{b}_2 \bar{x}_2 + \cdots + \hat{b}_p \bar{x}_p) \qquad (1)$$

$S_{jl} = \sum_{i=1}^{n} (x_{ji} - \bar{x}_j)(x_{li} - \bar{x}_l), S_{yj} = \sum_{i=1}^{n} (y_i - \bar{y})(x_{ji} - \bar{x}_j)$　　とおくと、

$$S_{11}\hat{b}_1 + S_{12}\hat{b}_2 + \cdots + S_{1p}\hat{b}_p = S_{y1}$$

$$S_{21}\hat{b}_1 + S_{22}\hat{b}_2 + \cdots + S_{2p}\hat{b}_p = S_{y2} \qquad (2)$$

………

$$S_{p1}\hat{b}_1 + S_{p2}\hat{b}_2 + \cdots + S_{pp}\hat{b}_p = S_{yp}$$

(2)のp元連立１次方程式を解いて$\hat{b}_1, \hat{b}_2 \cdots, \hat{b}_p$を求める。

さらに(1)から$\hat{a}$を求める。

偏回帰係数の意味

$\hat{y}_i = \hat{a} + \hat{b}_1 x_{1i} + \hat{b}_2 x_{2i} + \cdots + \hat{b}_p x_{pi}$　　b_j は偏回帰係数

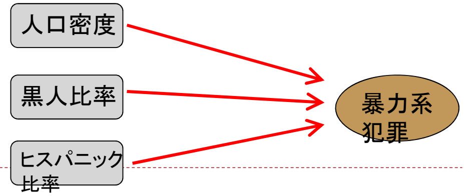
$\hat{y} = \hat{a} + \hat{b}_1 x_1 + \hat{b}_2 x_2 + \cdots + \hat{b}_p x_p$

1.　x_1が１単位増加（減少）すれば、平均して、

$\hat{y}$は$\hat{b}_1$だけ増加（減少）する

2.　　$\hat{b}_1$は、$x_2, \cdots, x_p$の影響を除いた、

$\hat{y}$　に対するx_1の効果を表す



変数を標準化した場合

- **標準化**　$Y_i = (y_i - \bar{y}) / s_y, X_{ji} = (x_{ji} - \bar{x}_j) / s_{x_j}$

$y_i = \hat{a} + \hat{b}_1 x_{1i} + \hat{b}_2 x_{2i} + \cdots + \hat{b}_p x_{pi} + \varepsilon_i$　　b_j は偏回帰係数

$Y_i = \hat{\beta}_1 X_{1j} + \hat{\beta}_2 X_{2j} + \cdots + \hat{\beta}_p X_{pj} + \varepsilon_i$　　β_j は標準偏回帰係数

定数項は消える。

$$\hat{Y}_i = (\hat{y}_i - \bar{y}) / s_y$$

$$= \frac{(\hat{a} + \hat{b}_1 x_{1i} + \hat{b}_2 x_{2i} + \cdots + \hat{b}_p x_{pi}) - (\hat{a} + \hat{b}_1 \bar{x}_1 + \hat{b}_2 \bar{x}_2 + \cdots + \hat{b}_p \bar{x}_p)}{s_y}$$

$$= \frac{\sum_{j=1}^{p} \hat{b}_j (x_{ji} - \bar{x}_j)}{s_y} = \sum_{j=1}^{p} \hat{b}_j \cdot \frac{(x_{ji} - \bar{x}_j)}{s_y} = \sum_{j=1}^{p} \hat{b}_j \cdot \frac{s_{x_j}}{s_y} \cdot \frac{(x_{ji} - \bar{x}_j)}{s_{x_j}}$$

$$= \sum_{j=1}^{p} \hat{b}_j \cdot \frac{s_{x_j}}{s_y} \cdot X_{ji}$$　　つまり、$\hat{\beta}_j = \hat{b}_j \cdot \frac{s_{x_j}}{s_y}$

標準偏回帰係数の意味

$\hat{Y} = \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \cdots + \hat{\beta}_p X_{pi}$　　β_j は標準偏回帰係数

1.　（標準化した値で）X_1が１単位増加（減少）

すれば、平均して、$\hat{Y}$は$\hat{\beta}_1$だけ増加（減少）する

2.　　$\hat{\beta}_1$は、$X_2, \cdots, X_p$の影響を除いた、

$\hat{Y}$　に対するX_1の効果を表す

3.　　$\hat{\beta}_1, \hat{\beta}_2, \cdots, \hat{\beta}_p$は値を相互に比較できる

平方和の分割

■従属変数の平方和（分散の分子）は、
独立変数によって説明できる平方和と、
誤差による平方和とに分割される
(テキスト、p.250)。

$SS_y = SS_{\hat{y}} + SS_{\varepsilon}$

SS_y : yの平方和

$SS_{\hat{y}}$: 独立変数から説明される平方和

（回帰モデルの平方和）

SS_{ε}: 誤差による平方和

重相関係数と決定係数

■重相関係数R

決定係数　$R^2 = r_{y\hat{y}}{}^2 = {SS_{\hat{y}}}\big/{SS_y} = 1 - {SS_{\varepsilon}}\big/{SS_y}$,

$$1 \geq R^2 \geq 0.$$

R^2は、従属変数の平方和のうち、回帰モデル（に導入した独立変数）によって説明される比率を表す。

■重回帰分析における検定：ポイント

- **よく使うのは２種類の検定**
- **重回帰式の検定**(pp.250-255)
  - **Ｆ検定**
- **偏回帰係数の検定**
  - **独立変数の寄与の検定**(F検定、pp.256-260)
  - **標準誤差を用いた検定**(t検定､pp.260-262)
- **検定結果は、変数を標準化しても同じ**
- **用語：** *F(a,b,α)* **：自由度*a,b*で有意水準*α*でのＦ値（Ｆ分布表の値）**

重回帰式の検定(Ｆ検定)

- **重相関係数の検定**
- **帰無仮説**$H_0: b_1 = b_2 = \ldots = b_p = 0$.
  - 独立変数は従属変数に説明に寄与していない。
- $F_0 > F(p, n\text{-}p\text{-}1, \alpha)$　→　H_0を棄却する

分散分析表（ANOVA）

変動要因	平方和	自由度	平均平方	Ｆ比
回帰モデル	$SS_{\hat{y}}$	p	$V_{\hat{y}} = {SS_{\hat{y}}}\big/{p}$	$F_0 = {V_{\hat{y}}}\big/{V_{\varepsilon}}$
誤差	SS_{ε}	$n\text{-}p\text{-}1$	$V_{\varepsilon} = {SS_{\varepsilon}}\big/{(n-p-1)}$	
全体	SS_y	$n\text{-}1$	$V_y = {SS_y}\big/{(n-1)}$	

偏回帰係数の検定

- **帰無仮説**$H_0: b_j = 0$.
  - **独立変数x_jは従属変数yの説明に寄与しない。**
- **Ｆ検定を使う方法**
  - **x_jを回帰式に導入することによる従属変数の説明の増加分が有意か否かを検定する**
  - $F_j > F(1, n\text{-}p\text{-}1, \alpha)$**なら、帰無仮説**$H_0$**を棄却**
- **ｔ検定を使う方法**
  - t_j **：**b_j**の推定値のｔ分布での値**
  - t_j > t値(自由度$n\text{-}p\text{-}1, \alpha$)　→　H_0**を棄却**
- **偏回帰係数の場合、Ｆ検定とｔ検定の結果は一致する**

■重回帰分析：例

- 埼玉県内の市町村への流入者数（人口当たりの）は何によって説明されるか？
- 分析単位：市町村
- 従属変数(dependent variable)
  - 流入者数（人口千人当たりの）
- 独立変数（independent variables)
  - 平均所得
  - 金融業（従事者数／千人）
  - サーヴィス業（従事者数／千人）
  - 公務（従事者数／千人）

主な結果：記述統計量

記述統計量

	平均値	標準偏差	N
流入	46.1638	9.3900	92
所得千円	6021.76	1516.39	92
金融業	12.9692	5.6147	92
サービス	122.3055	18.2727	92
公務	18.6792	5.7928	92

相関係数

		流入	所得千円	金融業	サービス	公務
Pearson の相関	流入	1.000	.300	.276	.332	.131
	所得千円	.300	1.000	.166	−.023	−.123
	金融業	.276	.166	1.000	.735	−.018
	サービス	.332	−.023	.735	1.000	.153
	公務	.131	−.123	−.018	.153	1.000
有意確率（片側）	流入	.	.002	.004	.001	.107
	所得千円	.002	.	.057	.416	.122
	金融業	.004	.057	.	.000	.432
	サービス	.001	.416	.000	.	.072
	公務	.107	.122	.432	.072	.
N	流入	92	92	92	92	92
	所得千円	92	92	92	92	92
	金融業	92	92	92	92	92
	サービス	92	92	92	92	92
	公務	92	92	92	92	92

重相関係数(R)、決定係数(R^2)

ﾓﾃﾞﾙ集計

ﾓﾃﾞﾙ	R	R2 乗	調整済み R2 乗	推定値の標準誤差
1	.468[a]	.219	.183	8.4869

a. 予測値: (定数)、公務, 金融業, 所得千円, サービス。

回帰モデルにより
従属変数の22%が
説明された

回帰式の検定（分散分析、 F 検定）

分散分析[b]

ﾓﾃﾞﾙ		平方和	自由度	平均平方	F 値	有意確率
1	回帰	1757.188	4	439.297	6.099	.000[a]
	残差	6266.418	87	72.028		
	全体	8023.606	91			

a. 予測値: (定数)、公務, 金融業, 所得千円, サービス。

b. 従属変数: 流入

偏回帰係数の検定（分散分析）

係数[a]

モデル		非標準化係数 B	非標準化係数 標準誤差	標準化係数 ベータ	t	有意確率
1	（定数）	9.594	8.573		1.119	.266
	所得千円	2.022E-03	.001	.327	3.310	.001
	金融業	-4.413E-02	.246	-.026	-.179	.858
	サービス	.175	.075	.340	2.317	.023
	公務	.191	.159	.118	1.204	.232

a. 従属変数: 流入

$$\hat{y} = 9.594 + 0.002x_1 - 0.044x_2 + 0.175x_3 + 0.191x_4$$

$$\hat{Y} = 0.327X_1 - 0.026X_2 + 0.340X_3 + 0.118X_4$$

偏回帰係数の検定（分散分析）

有意なのは、所得千円とサービスだけである

係数[a]

モデル		非標準化係数 B	非標準化係数 標準誤差	標準化係数 ベータ	t	有意確率
1	（定数）	9.594	8.573		1.119	.266
	所得千円	2.022E-03	.001	.327	3.310	.001
	金融業	-4.413E-02	.246	-.026	-.179	.858
	サービス	.175	.075	.340	2.317	.023
	公務	.191	.159	.118	1.204	.232

a. 従属変数: 流入

従属変数に対する各独立変数の効果（偏回帰係数）はここに出ている

検定の帰無仮説は「その偏回帰係数が母集団ではゼロ」

検定量（t値）はここに出ている。帰無仮説が間違っているほど、t値は高い。

ここで一応の区切りです

重回帰分析への補足説明

- ダミー変数
- 相関係数と（標準）偏回帰係数の関係
- 重相関係数と自由度
- 偏回帰係数の検定－誤差の問題
- 多重共線性
- 変数選択

■ダミー変数(dummy variables)

- 値は１か０
- 質的／定性的データを扱う際に使う
  - カテゴリー変数
  - 例：男女、職業、国籍…
- カテゴリーがｎ個なら n-1 個のダミー変数で表現する
  - なぜ１つ減らすか？　→　「一次従属」問題
- ダミー変数の偏回帰係数は、欠けたカテゴリーからの差を表す

■標準偏回帰係数と相関係数
【復習】重回帰分析：基本的な考え方

- 単回帰モデル

$$y_i = a + bx_i + \varepsilon_i$$

- 重回帰モデル

$$y_i = a + b_1 x_{1i} + b_2 x_{2i} + \cdots + b_p x_{pi} + \varepsilon_i$$

- 基本的には単回帰と同じ
- a：定数項
- b_j：偏回帰係数
  - x_j 以外の変数の影響を統計的に除去している

重回帰モデル

$$y_i = \hat{a} + \hat{b}_1 x_{1i} + \hat{b}_2 x_{2i} + \cdots + \hat{b}_p x_{pi} + \varepsilon_i$$

$$\hat{y}_i = \hat{a} + \hat{b}_1 x_{1i} + \hat{b}_2 x_{2i} + \cdots + \hat{b}_p x_{pi}$$

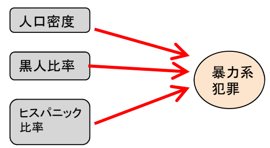
$$\varepsilon_i = y_i - \hat{y}_i$$

偏回帰係数の意味

$$\hat{y}_i = \hat{a} + \hat{b}_1 x_{1i} + \hat{b}_2 x_{2i} + \cdots + \hat{b}_p x_{pi}$$　　b_j は偏回帰係数

$$\hat{y} = \hat{a} + \hat{b}_1 x_1 + \hat{b}_2 x_2 + \cdots + \hat{b}_p x_p$$

1. x_1が１単位増加（減少）すれば、平均して、$\hat{y}$は$\hat{b}_1$だけ増加（減少）する
2. b_1は、$x_2,\cdots,x_p$の影響を除いた、$\hat{y}$ に対するx_1の効果を表す



標準偏回帰係数の意味

$\hat{Y}_i = \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \cdots + \hat{\beta}_p X_{pi}$　　β_j **は標準偏回帰係数**

1. **（標準化した値で）X_1が１単位増加（減少）すれば、平均して、$\hat{Y}$は$\hat{\beta}_1$だけ増加（減少）する**
2. **$\hat{\beta}_1$は、$X_2, \cdots, X_p$の影響を除いた、$\hat{Y}$　に対するX_1の効果を表す**
3. **$\hat{\beta}_1, \hat{\beta}_2, \cdots, \hat{\beta}_p$は値を相互に比較できる**

相関係数と(標準)偏回帰係数の関係

$$r_{yi} = \sum_{j=1}^{p} \hat{\beta}_j r_{ij} = \hat{\beta}_i + \sum_{j \neq i}^{p} \hat{\beta}_j r_{ij}$$

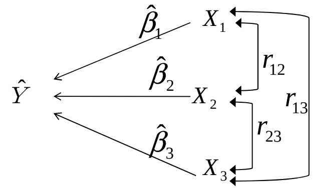
例えば

$$r_{y1} = \hat{\beta}_1 + \hat{\beta}_2 r_{12} + \hat{\beta}_3 r_{13}$$

Yに対するX1の直接効果

X2を経由したX1の間接効果

X3を経由したX1の間接効果



$$\hat{Y}_i = \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i}$$

Quiz1

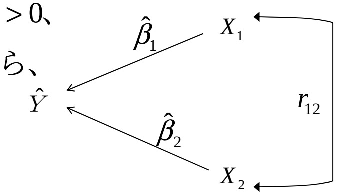
- **独立変数*x_1*,*x_2*と従属変数*y* の重回帰分析において、*b_1*が負であるとする。このとき、*y*と*x_1*の相関係数が正になることは？　→　答：あり得る**

$r_{y1} = \hat{\beta}_1 + \hat{\beta}_2 r_{12}$

$\hat{\beta}_1 < 0$でも、$r_{12} > 0$かつ$\hat{\beta}_2 > 0$、

ないし$r_{12} < 0$かつ$\hat{\beta}_2 < 0$なら、

$r_{y1} > 0$があり得る。



Quiz2

問３：変数を標準化した重回帰モデル

$\hat{Y}_i = \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \cdots + \hat{\beta}_p X_{pi}$

において、$\hat{\beta}_1 = r_{X_1 Y}$，　$\hat{\beta}_2 = r_{X_2 Y}, \hat{\beta}_3 = r_{X_3 Y}$

となるのはいかなる場合か？

$r_{yi} = \sum_{j=1}^{p} \hat{\beta}_j r_{ij} = \hat{\beta}_i + \sum_{j \neq i}^{p} \hat{\beta}_j r_{ij}$だから、

独立変数間の相関がゼロのとき($\underset{i \neq j}{r_{ij}} = 0$)

■重相関係数と自由度
重相関係数の問題

■重相関係数R　　$SS_y = SS_{\hat{y}} + SS_\varepsilon$

決定係数　$R^2 = r_{y\hat{y}}^2 = SS_{\hat{y}} / SS_y = 1 - SS_\varepsilon / SS_y$,

$1 \ge R^2 \ge 0.$

- R^2が大きいほど回帰式による説明率は高い。
- しかし、R^2が大きくても回帰式は無意味なこともある。
- 独立変数の数は、観測数との関係で、低く抑える必要がある。

重相関係数と自由度

- 独立変数の数pが n-1と等しくなるとき、自動的にR=1.0となる。
  - このとき、誤差の自由度(n-p-1)はゼロになる。
  - Rが高くても回帰式は無意味。

変動要因	平方和	自由度	平均平方	F比
回帰モデル	$SS_{\hat{y}}$	p	$V_{\hat{y}} = SS_{\hat{y}} / p$	$F_0 = V_{\hat{y}} / V_\varepsilon$
誤差	SS_ε	n-p-1	$V_\varepsilon = SS_\varepsilon / (n-p-1)$	
全体	SS_y	n-1	$V_y = SS_y / (n-1)$	

自由度調整済みの重相関係数（決定係数）

■重相関係数R

決定係数　$R^2 = r_{y\hat{y}}^2 = SS_{\hat{y}} / SS_y = 1 - SS_\varepsilon / SS_y$

■自由度調整済みの重相関係数R_{adj}

$$R_{adj}^2 = 1 - \frac{\frac{SS_\varepsilon}{(n-p-1)}}{\frac{SS_y}{n-1}} = 1 - \frac{n-1}{n-p-1}(1-R^2)$$

ﾓﾃﾞﾙ集計

ﾓﾃﾞﾙ	R	R2 乗	調整済み R2 乗	推定値の標準誤差
1	.468[a]	.219	.183	8.4869

a. 予測値: (定数)、公務, 金融業, 所得千円, サービス。

■偏回帰係数の検定－誤差の問題
重回帰分析における検定：ポイント

- よく使うのは２種類の検定
- 重回帰式の検定(pp.250-255)
  - Ｆ検定
- 偏回帰係数の検定
  - 独立変数の寄与の検定(F検定、pp.256-260)
  - 標準誤差を用いた検定(t検定､pp.260-262)
- 検定結果は、変数を標準化しても同じ
- 用語： *F(a,b,α)*：自由度*a,b*で有意水準*α*でのＦ値（Ｆ分布表の値）

偏回帰係数の検定

- 帰無仮説H_0:b_j=0.
  - 独立変数x_jは従属変数yの説明に寄与しない。
- Ｆ検定を使う方法
  - x_jを回帰式に導入することによる従属変数の説明の増加分が有意か否かを検定する
  - $F_j > F(1,n-p-1,\alpha)$なら、帰無仮説H_0を棄却
- ｔ検定を使う方法
  - t_j ：b_jの推定値のｔ分布での値
  - t_j > t値(自由度$n-p-1,\alpha$)　→　H_0を棄却
- 偏回帰係数の場合、Ｆ検定とｔ検定の結果は一致する

偏回帰係数の検定（分散分析）

係数[a]

モデル		非標準化係数 B	標準誤差	標準化係数 ベータ	t	有意確率
1	(定数)	9.594	8.573		1.119	.266
	所得千円	2.022E-03	.001	.327	3.310	.001
	金融業	-4.413E-02	.246	-.026	-.179	.858
	サービス	.175	.075	.340	2.317	.023
	公務	.191	.159	.118	1.204	.232

a. 従属変数: 流入

$$\hat{y} = 9.594 + 0.002x_1 - 0.044x_2 + 0.175x_3 + 0.191x_4$$

$$\hat{Y} = 0.327X_1 - 0.026X_2 + 0.340X_3 + 0.118X_4$$

Quiz

- 独立変数x_1,x_2と従属変数y の重回帰分析において、標準偏回帰係数では$\beta_1 > \beta_2 > 0$ であったとする。このとき、もしβ_2 が有意（H_0: $\beta_2=0$　を棄却できる)であればβ_1も有意であるといえるか？
- 結論：必ずしもいえない。
- β_1 や β_2は推定値であり、一種の期待値である。
- β_1 や β_2は誤差を含む。(テキスト、pp.260-262)
- β_1 が大きくてもその誤差が大きければ、検定のｔ値は低くなる。

■多重共線性

- Multicollinearity
- 独立変数間の相関が高すぎると、偏回帰係数の推定値は不安定になる。
- 相関が高過ぎる独立変数を同時に加えないように配慮する必要。
  - 次の「変数選択」の問題
- 一般に、相関が高い独立変数を加えても、回帰式の説明率はあまり向上しない。

■変数選択

- 自由度を考慮すると、独立変数は選択的に導入する必要がある。
- 変数選択：説明力のある独立変数だけを選ぶ（あるいは、説明力の無い独立変数を排除する）解析的な方法
  - ＳＰＳＳなどでは備わっている
    - 「方法」：デフォルトは「強制投入法」（すべての独立変数を投入）
  - 多重共線性の問題が生じにくくなる。
- 方法
  - 変数増加法
  - 変数減少法
  - ステップワイズ法（推奨）

例：変数選択

- 従属変数：bsri男性
- 独立変数
  - 性別　ー　ダミー変数（0/1）
  - パーソナリティ特性： Big Five
    - 外向性
    - 情緒不安性
    - 開放性
    - 誠実性
    - 調和性

ＳＰＳＳ：強制投入法の結果

分散分析[b]

ﾓﾃﾞﾙ		平方和	自由度	平均平方	F 値	有意確率
1	回帰	31613.313	6	5268.886	43.631	.000[a]
	残差	29344.931	243	120.761		
	全体	60958.244	249			

a. 予測値: (定数)、調和, 性別, 情緒不安, 誠実, 外向, 開放。
b. 従属変数: bsri男性

係数[a]

ﾓﾃﾞﾙ		非標準化係数 B	非標準化係数 標準誤差	標準化係数 ﾍﾞｰﾀ	t	有意確率
1	(定数)	5.139	16.183		.318	.751
	性別	-2.477	1.422	-.079	-1.741	.083
	外向	.108	.066	.090	1.645	.101
	情緒不安	-.265	.066	-.203	-4.018	.000
	開放	.783	.086	.511	9.156	.000
	誠実	.028	.141	.010	.197	.844
	調和	.508	.142	.210	3.582	.000

a. 従属変数: bsri男性

ステップワイズ法の結果（１）

分散分析[d]

ﾓﾃﾞﾙ		平方和	自由度	平均平方	F 値	有意確率
1	回帰	27406.198	1	27406.198	202.573	.000[a]
	残差	33552.046	248	135.291		
	全体	60958.244	249			
2	回帰	28604.526	2	14302.263	109.189	.000[b]
	残差	32353.718	247	130.987		
	全体	60958.244	249			
3	回帰	31033.243	3	10344.414	85.037	.000[c]
	残差	29925.001	246	121.646		
	全体	60958.244	249			

a. 予測値: (定数)、開放。
b. 予測値: (定数)、開放, 情緒不安。
c. 予測値: (定数)、開放, 情緒不安, 調和。
d. 従属変数: bsri男性

ステップワイズ法の結果（2）

係数[a]

モデル		非標準化係数 B	非標準化係数 標準誤差	標準化係数 ベータ	t	有意確率
1	(定数)	29.208	3.818		7.650	.000
	開放	1.029	.072	.671	14.233	.000
2	(定数)	40.511	5.299		7.645	.000
	開放	1.014	.071	.661	14.230	.000
	情緒不安	−.184	.061	−.141	−3.025	.003
3	(定数)	1.661	10.083		.165	.869
	開放	.830	.080	.541	10.350	.000
	情緒不安	−.289	.063	−.221	−4.579	.000
	調和	.590	.132	.244	4.468	.000

a. 従属変数: bsri男性

今日は

おしまい