



データ解析実習

2012.11.19 月曜、4・5限

担当教員：高木英至

1

重回帰分析への補足説明

- ▶ 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

▶ 2

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

$$y_i = a + b_1x_{1i} + b_2x_{2i} + \cdots + b_px_{pi} + \varepsilon_i$$

- ▶ 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

▶ 3

重回帰モデル

$$y_i = \hat{a} + \hat{b}_1x_{1i} + \hat{b}_2x_{2i} + \cdots + \hat{b}_px_{pi} + \varepsilon_i$$

$$\hat{y}_i = \hat{a} + \hat{b}_1x_{1i} + \hat{b}_2x_{2i} + \cdots + \hat{b}_px_{pi}$$

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

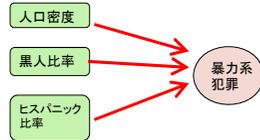
▶ 4

偏回帰係数の意味

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

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

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, \dots, x_p の影響を除いた、 \hat{y} に対する x_1 の効果を表す



▶ 5

標準偏回帰係数の意味

$$\hat{Y}_i = \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \dots + \hat{\beta}_p X_{pi} \quad \beta_j \text{ は標準偏回帰係数}$$

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, \dots, X_p の影響を除いた、 \hat{Y} に対する X_1 の効果を表す
3. $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p$ は値を相互に比較できる

▶ 6

相関係数と(標準)偏回帰係数の関係

$$r_{yi} = \sum_{j=1}^p \hat{\beta}_j r_{ij} = \hat{\beta}_1 + \sum_{j \neq 1}^p \hat{\beta}_j r_{ij}$$

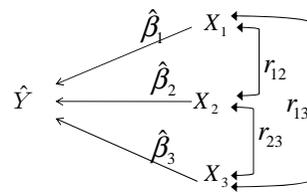
例えば

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

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



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

▶ 7

Quiz1

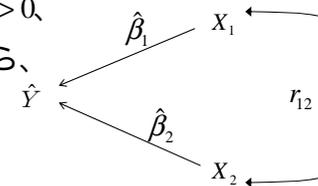
- ▶ 独立変数 x_1, x_2 と従属変数 y の重回帰分析において、 b_1 が負であるとする。このとき、 y と x_1 の相関係数が正になることは? 答: あり得る

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

$\hat{\beta}_1 < 0$ 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。



▶ 8

Quiz2

問3：変数を標準化した重回帰モデル

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

において、 $\hat{\beta}_1 = r_{X_1Y}$, $\hat{\beta}_2 = r_{X_2Y}$, $\hat{\beta}_3 = r_{X_3Y}$

となるのはいかなる場合か？

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

独立変数間の相関がゼロのとき ($r_{ij} = 0$)

▶ 9

重相関係数と自由度 重相関係数の問題

重相関係数 R $SS_y = SS_{\hat{y}} + SS_e$

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

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

- ▶ R^2 が大きいほど回帰式による説明率が高い。
- ▶ 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。

▶ 10

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、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	$V_{\hat{y}} = \frac{SS_{\hat{y}}}{p}$	$F_0 = \frac{V_{\hat{y}}}{V_e}$
誤差	SS_e	$n-p-1$	$V_e = \frac{SS_e}{(n-p-1)}$	
全体	SS_y	$n-1$	$V_y = \frac{SS_y}{(n-1)}$	

▶ 11

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 $R^2 = r_{y\hat{y}}^2 = \frac{SS_{\hat{y}}}{SS_y} = 1 - \frac{SS_e}{SS_y}$

自由度調整済みの重相関係数 R_{adj}

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

モデル集計

モデル	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。

▶ 12

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： $F(a,b,a)$ ：自由度 a,b で有意水準 a でのF値（F分布表の値）

▶ 13

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
 - ▶ $F_j > F(1,n-p-1,a)$ 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 $n-p-1,a$) 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

▶ 14

偏回帰係数の検定（分散分析）

係数^a

モデル	非標準化係数		標準化係数	t	有意確率
	B	標準誤差	ベータ		
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	-1.179	.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

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

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

▶ 15

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。

▶ 16

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。

▶ 17

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）

▶ 18

例：変数選択

- ▶ 従属変数：bsri男性
- ▶ 独立変数
 - ▶ 性別 - 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

▶ 19

SPSS：強制投入法の結果

分散分析^a

モデル	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 ^b
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

モデル		非標準化係数		標準化係数		t	有意確率
		B	標準誤差	ベータ			
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

▶ 20

ステップワイズ法の結果 (1)

分散分析^d

モデル		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

▶ 21

ステップワイズ法の結果 (2)

係数^a

モデル		非標準化係数		標準化係数	t	有意確率
		B	標準誤差	ベータ		
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

▶ 22



講義は おしまい

▶ 23