

# 重回帰分析

10/05/12

M1 石田彩夏

# 目次

- 回帰分析とは
  - 単回帰分析
  - 重回帰分析
- SPSSとR
  - SPSSによる重回帰分析
  - Rによる重回帰分析

# 回帰分析とは

## 回帰分析(regression analysis)

- 「独立変数 $x$ によって、従属変数 $y$ がどれくらい説明できるか」を予測・説明するための方法
  - 式を用いて、 $x$ から $y$ を予測する
  - $x$ と $y$ が直線的な関係であることが前提となる

独立変数が1つのとき：単回帰分析

独立変数が2つ以上のとき：重回帰分析

# 単回帰分析(single regression analysis )

- 回帰直線

$$\hat{y}=a+bx$$

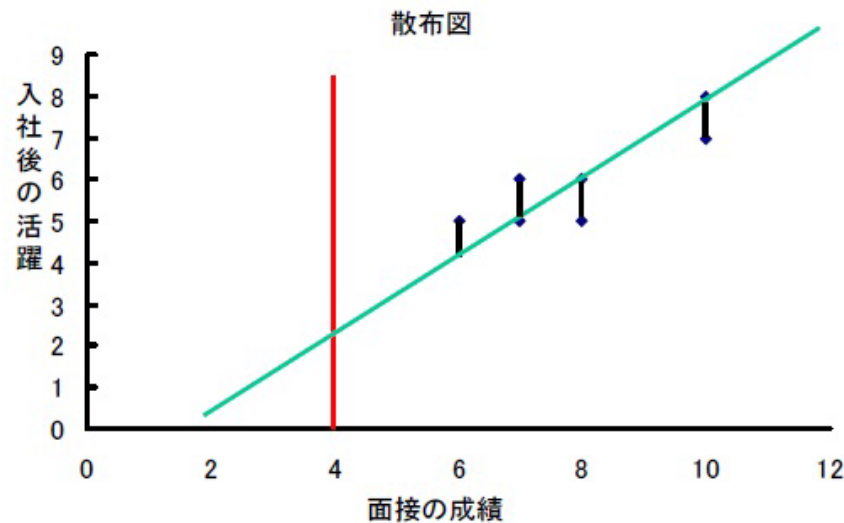
$x$ : 独立変数 (説明変数)

$y$ : 従属変数 (基準変数)

$\hat{y}$ : 予測値

$a$ : 切片  $b$ : 傾き (回帰係数)

- 独立変数 $x$ について、従属変数 $y$ との関係を調べる

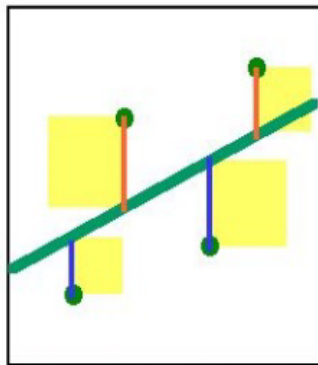


ex. 左の図では、面接の成績と入社後の活躍にどのような関係があるかがわかる

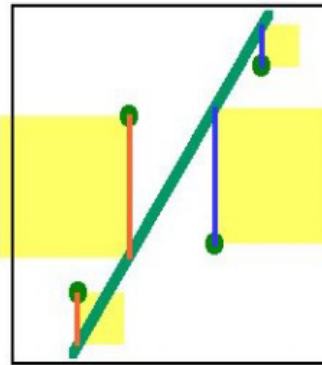
面接の成績が4点の人の入社後の活躍の予測ができる

# 最小二乗法による回帰直線への当てはめ

- 最小二乗法  
残差の2乗和を最小にする  
ような直線を求める方法



残差の2乗和が小さい直線



残差の2乗和が大きい直線

実際には、「適合の悪さ」を以下の指標で定義する

$$Q = \sum_{i=1}^N [y_i - (a + bx_i)]^2$$

aとbは、aとbを偏微分し、結果を0とした連立方程式の解によって求められる

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

$$b = r \frac{s_y}{s_x}$$

- $\hat{y} = \bar{y} + b(x - \bar{x})$

(xが平均 $\bar{x}$ に等しいとき、それに対応するyの予測値 $\hat{y}$ は、yの平均 $\bar{y}$ に等しくなる)

# 単回帰モデルの評価 - 決定係数

- $S_y = S_{\hat{y}}^2 + S_e^2$   
     $S_y$ のうち、 $r^2$  : 予測値の分散  
                     $1 - r^2$  : 残差の分散       $(r = \frac{S_{xy}}{S_x S_y})$

- $r^2$  : 分散説明率

全体の分散に対する、回帰式によって説明される分散の割合

1に近い: 全体の分散が回帰式で説明される

0に近い: 全体の分散が誤差で占められている

= 回帰式の予測の精度を示す指標として、  
決定係数と呼ばれる

# 重回帰分析(multiple regression analysis )

- **回帰平面** (独立変数が2つの場合。3つ以上の場合には「回帰超平面」「予測超平面」と言い、視覚的把握は困難だが、手続きは同じ)

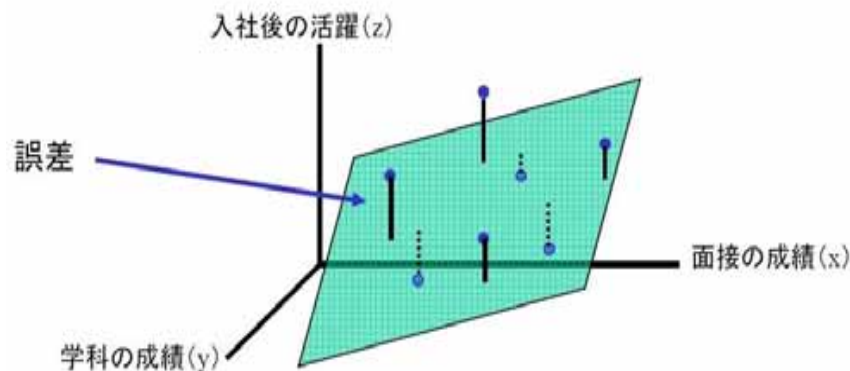
$$\hat{y} = a + b_1x_1 + b_2x_2$$

$x$  : 独立変数 (説明変数)

$y$  : 従属変数 (基準変数)

$\hat{y}$  : 予測値  $a$  : 切片  $b$  : 偏回帰係数

( $b_1x_1$  : 回帰平面の $x_1$ 方向の傾き)



- すべての独立変数について、ほかのすべての独立変数の影響を除いた成分を考え、それらの従属変数との関係を調べる
- すべての独立変数に対する偏回帰係数が一挙に推定できる

# 偏回帰係数と標準偏回帰係数

- 最小二乗法によって  
偏回帰係数を求める

$$Q = \sum_{i=1}^N [y_i - (a + b_1 x_{1i} + b_2 x_{2i})]^2$$

が最小になるaとbの値

$$a = \bar{y} - (b_1 \bar{x}_1 + b_2 \bar{x}_2)$$

$$b_1 = \frac{s_y}{s_{x1}} \frac{r_{yx1} - r_{yx2} r_{x1x2}}{1 - r_{x1x2}^2}$$

$$b_2 = \frac{s_y}{s_{x2}} \frac{r_{yx2} - r_{yx1} r_{x1x2}}{1 - r_{x1x2}^2}$$

- 独立変数同士の測定単位  
が異なるとき、切片が0に  
なる重回帰モデルで表す。  
このときの回帰式の偏回帰  
係数が、標準偏回帰係数

$$b_1^* = \frac{r_{yx1} - r_{yx2} r_{x1x2}}{1 - r_{x1x2}^2}$$

$$b_2^* = \frac{r_{yx2} - r_{yx1} r_{x1x2}}{1 - r_{x1x2}^2}$$



# 重回帰モデルの評価 重決定係数

- 重相関係数

$$R_{yy} = \sqrt{\frac{r_{yx1} + r_{yx2} - 2r_{yx1}r_{yx2}r_{x1x2}}{1 - r_{x1x2}^2}}$$

…各独立変数と従属変数との相関係数  
より少し大きくなる  
= 複数の独立変数を用いた効果

- 重決定係数

$$R^2 = \frac{r_{yx1} + r_{yx2} - 2r_{yx1}r_{yx2}r_{x1x2}}{1 - r_{x1x2}^2}$$

…回帰方程式の  
「当てはまりのよさ」を示している

- 自由度調整済み重決定係数

説明変数が増える 予測精度が高まる  $R^2$ が大きくなる

しかし！説明変数が増える F値が小さくなる 重回帰モデルの有意性低下

説明変数の数  $p$  を考慮してモデルを評価する指標

$$R_a^2 = -\frac{N-1}{N-p-1}(1-R^2) \quad (N \text{ は標本数})$$

# 重相関係数の検定

- 重相関係数が統計的に有意かどうか
  - 帰無仮説：母集団におけるモデル全体としての説明力が0

$$F = \frac{R^2 / p}{(1 - R^2) / (N - p - 1)}$$

- …帰無仮説が正しいとき、 $F$ は自由度  $p$  と  $N-p-1$  の  $F$  分布に従う

# 多重共線性の問題

- 独立変数間の相関が高すぎる時、  
偏回帰係数の推定量が不安定になる

相関の強い独立変数を取り除く

( $VIF > 10$ および $CI > 15$ を指標にする)

新しい変数を加える

相関する複数の変数を合成する

…等の工夫が必要になる

# SPSSとR

- SPSSとRは、  
いずれも統計ソフトの1つであり、以下のような特徴がある。

SPSS	R
GUI: マウスとボタンで直感的な作業が可能	CUI: コマンドを覚える / 調べる 打ち込むという作業が必要
高価: 個人所有は難しく、 学外環境では使えないことがほとんど	無料: いつでも・どこでも使える
マニュアルが豊富	マニュアルは英語が多く、 探すのがやや大変 (近年はだいぶ増えてきている)
	* オープンソースであり、研究開発が盛ん

# 例題

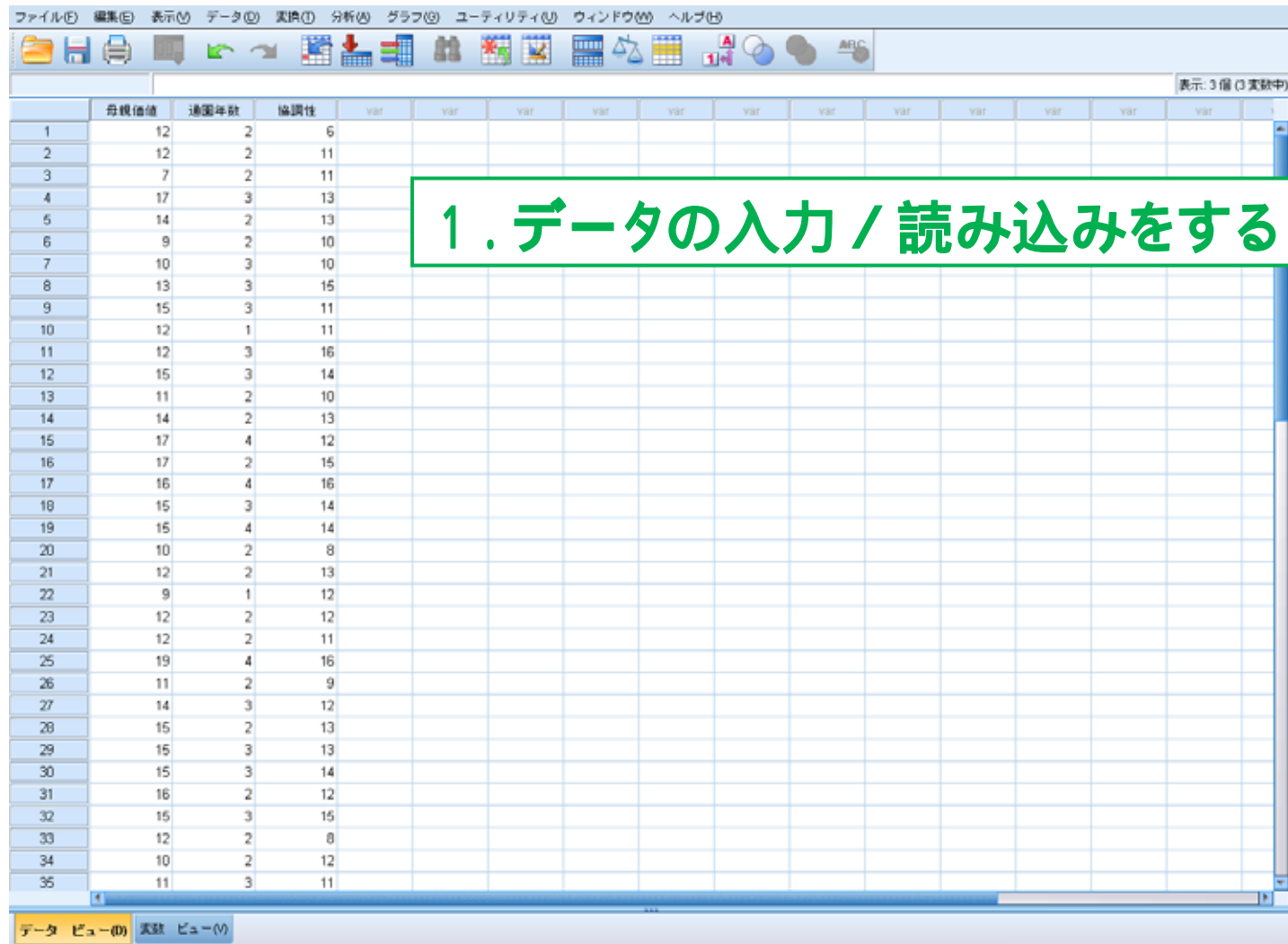
『心理統計学の基礎』(南風原;2002) p.226のデータより

- 子供の協調性の発達に関する研究で、
  - 「母親が子供の協調性に価値を置いているほど、子供の協調性の発達を促進するだろう」
  - 「幼稚園などでの友達との相互作用の経験が、子供の協調性の発達を促進するだろう」という二つの仮説を検討する。

# 重回帰分析の際に重要なデータ

- 1 . 記述統計量
- 2 . 相関行列
- 3 . 標準偏回帰係数
- 4 . 分散説明率
- 5 . 重相関係数の検定 (F値)
- 6 . VIF / CI

# SPSSによる重回帰分析



表示: 3 層 (3 変数中)

	母親価値	通園年数	協調性	V&R	V&R	V&R	V&R	V&R	V&R	V&R	V&R	V&R	V&R	V&R	V&R	V&R
1	12	2	6													
2	12	2	11													
3	7	2	11													
4	17	3	13													
5	14	2	13													
6	9	2	10													
7	10	3	10													
8	13	3	15													
9	15	3	11													
10	12	1	11													
11	12	3	16													
12	15	3	14													
13	11	2	10													
14	14	2	13													
15	17	4	12													
16	17	2	15													
17	16	4	16													
18	15	3	14													
19	15	4	14													
20	10	2	8													
21	12	2	13													
22	9	1	12													
23	12	2	12													
24	12	2	11													
25	19	4	16													
26	11	2	9													
27	14	3	12													
28	15	2	13													
29	15	3	13													
30	15	3	14													
31	16	2	12													
32	15	3	15													
33	12	2	8													
34	10	2	12													
35	11	3	11													

データ ビュー(0) 変数 ビュー(0)

ファイル(F) 編集(E) 表示(V) データ(D) 変換(T) 分析(A) グラフ(G) ユーティリティ(U) ウィンドウ(W) ヘルプ(H)

表示: 3 個 (2 実数中)

	母親価値	通園年齢	協調性
1	12	2	
2	12	2	
3	7	2	
4	17	3	
5	14	2	
6	9	2	
7	10	3	
8	13	3	
9	15	3	
10	12	1	
11	12	3	
12	15	3	
13	11	2	
14	14	2	
15	17	4	
16	17	2	15
17	16	4	16
18	15	3	14
19	15	4	14
20	10	2	8
21	12	2	13
22	9	1	12
23	12	2	12
24	12	2	11
25	19	4	16
26	11	2	9
27	14	3	12
28	15	2	13
29	15	3	13
30	15	3	14
31	16	2	12
32	15	3	15
33	12	2	8
34	10	2	12
35	11	3	11

分析(A) のメニュー:

- 報告書(R)
- 記述統計(E)
- 平均の比較(M)
- 一般線型モデル(G)
- 一般化線型モデル(G)
- 混合モデル(M)
- 相関(C)
- 回帰(R)
  - 線形(L)...**
  - 曲線推定(C)...
  - 偏相関最小 2 乗法(S)...
  - 二項ロジスティック(G)...
  - 多項ロジスティック(M)...
  - 順序(O)...
  - プロビット(P)...
  - 非線形(N)...
  - 重み付け推定(W)...
  - 2 段階最小 2 乗(Z)...
- 対応線型(O)
- 分類(F)
- 次元分解
- 尺度(A)
- ノンパラメトリック検定(N)
- 見込み(T)
- 生存分析(S)
- 多重回帰(J)
- 品質管理(Q)
- ROC 曲線(V)...

データ ビュー(D) 実数 ビュー(M)

## 2. 分析 回帰 線形を選択



ファイル(F) 編集(E) 表示(V) データ(D) 変換(T) 分析(A) グラフ(G) ユーティリティ(U) ウィンドウ(W) ヘルプ(H)

表示: 3 値 (3 変数中)

	母親価値	通園年数	協調性	V3F	V3F	V3F	V3F	V3F	V3F	V3F	V3F	V3F	V3F	V3F	V3F
1	12	2	6												
2	12	2	11												
3	7	2	11												
4	17	3	13												
5	14	2	13												
6	9	2	10												
7	10	3	10												
8	13	3	15												
9	15	3	11												
10	12	1	11												
11	12	3	16												
12	15	3	14												
13	11	2	10												
14	14	2	13												
15	17	4	12												
16	17	2	15												
17	16	4	16												
18	15	3	14												
19	15	4	14												
20	10	2	8												
21	12	2	13												
22	9	1	12												
23	12	2	12												
24	12	2	11												
25	19	4	16												
26	11	2	9												
27	14	3	12												
28	15	2	13												
29	15	3	13												
30															
31															
32															
33	12	2	9												
34	10	2	12												
35	11	3	11												

線型回帰

従属変数: 協調性

独立変数: 母親価値, 通園年数

方法: 最小二乗法

ケース選択変数:

ケースのラベル:

WLS 重み:

統計量

作成

保存

オプション

線型回帰: 統計

統計量

☒ 決定係数

☒ 決定係数の変化量

☒ 決定係数の標準誤差

☒ 決定係数の信頼区間

☒ 決定係数の検定

レベル(%): 95

☒ 分散共分散行列

☒ モデルの適合度

☒ R 2 乗の変化量

☒ 決定係数の標準誤差

☒ 決定係数の信頼区間

☒ 決定係数の検定

検定

☐ Durbin-Watson の検定

☐ ケースごとの検定

外れ値: 3 標準偏差

すべてのケース

実行

キャンセル

ヘルプ

データ ビュー(D) 変換 ビュー(V)

3. 変数を決め、必要な統計量にチェックを入れる

REGRESSION  
 /DESCRIPTIVES MEAN STDDEV CORR SIG N  
 /MISSING LISTWISE  
 /STATISTICS COEFF OUTS CI(95) BOOY R ANOVA COLLIN TOL CHANGE ZPP  
 /CRITERIA=PIN(.05) POUT(.10)  
 /NOORIGIN  
 /DEPENDENT 協調性  
 /METHOD=ENTER 母親価値 通園年数.

回帰

[データセット0] F:\V大学院\データ解析\5月12日発表\重回帰分析PPVデータセット0.sav

**記述統計**

	平均値(ラン 検定)	標準偏差	N
協調性	11.96	2.507	50
母親価値	13.20	2.449	50
通園年数	2.48	.762	50

**相関**

	協調性	母親価値	通園年数
Pearsonの相関			
協調性	1.000	.480	.501
母親価値	.480	1.000	.538
通園年数	.501	.538	1.000
有意確率(片側)			
協調性		.000	.000
母親価値	.000		.000
通園年数	.000	.000	
N			
協調性	50	50	50
母親価値	50	50	50
通園年数	50	50	50

**投入済み変数または除去された変数<sup>a</sup>**

モデル	投入済み変数	除去された変数	方法
1	通園年数, 母親価値 <sup>a</sup>		入力

a. 要求された変数がすべて入力されました。  
 b. 従属変数 協調性

\*出力4 [ドキュメント4] - PASW Statistics ユーザー

ファイル(F) 編集(E) 表示(V) データ(D) 変換(T) 挿入(I) 書式(O) 分析(A) グラフ(G) ユーティリティ(U) ウィンドウ(W) ヘルプ(H)

出力 ログ 回帰 表題 記録 アクティブ データ 記述統計 相関 投入済み変数またはモデル集計 分散分析 係数 相関係数 共線性の診断

### 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	R2 乗変化量	F 変化量	自由度 1	自由度 2	有意確率 F 変化量
1	.560 <sup>a</sup>	.314	.284	2.120	.314	10.740	2	47	.000

a. 予測値: (定数)、通園年数、母親価値。

### 分散分析<sup>b</sup>

モデル	平方和 (分散成分)	自由度	平均平方	F 値	有意確率
1 回帰	96.587	2	48.294	10.740	.000 <sup>a</sup>
残差 (分散分析)	211.333	47	4.496		
合計 (ピボットテーブル)	307.920	49			

a. 予測値: (定数)、通園年数、母親価値。  
b. 従属変数 協調性

### 係数<sup>a</sup>

モデル		標準化されていない係数		標準化係数	t 値	有意確率	B の 95.0% 信頼区間		相関			共線性の統計量	
		B	標準誤差	ベータ			下限	上限	ゼロ次	偏	部分	許容度	VIF
1	(定数)	5.172	1.665		3.107	.003	1.823	8.520					
	母親価値	.303	.147	.296	2.064	.045	.008	.598	.480	.288	.249	.711	1.407
	通園年数	1.126	.471	.342	2.389	.021	.178	2.074	.501	.329	.289	.711	1.407

a. 従属変数 協調性

### 相関係数<sup>a</sup>

モデル		通園年数	母親価値
1	相関	1.000	-.538
	母親価値	-.538	1.000
	共分散	.222	-.037
	母親価値	-.037	.022

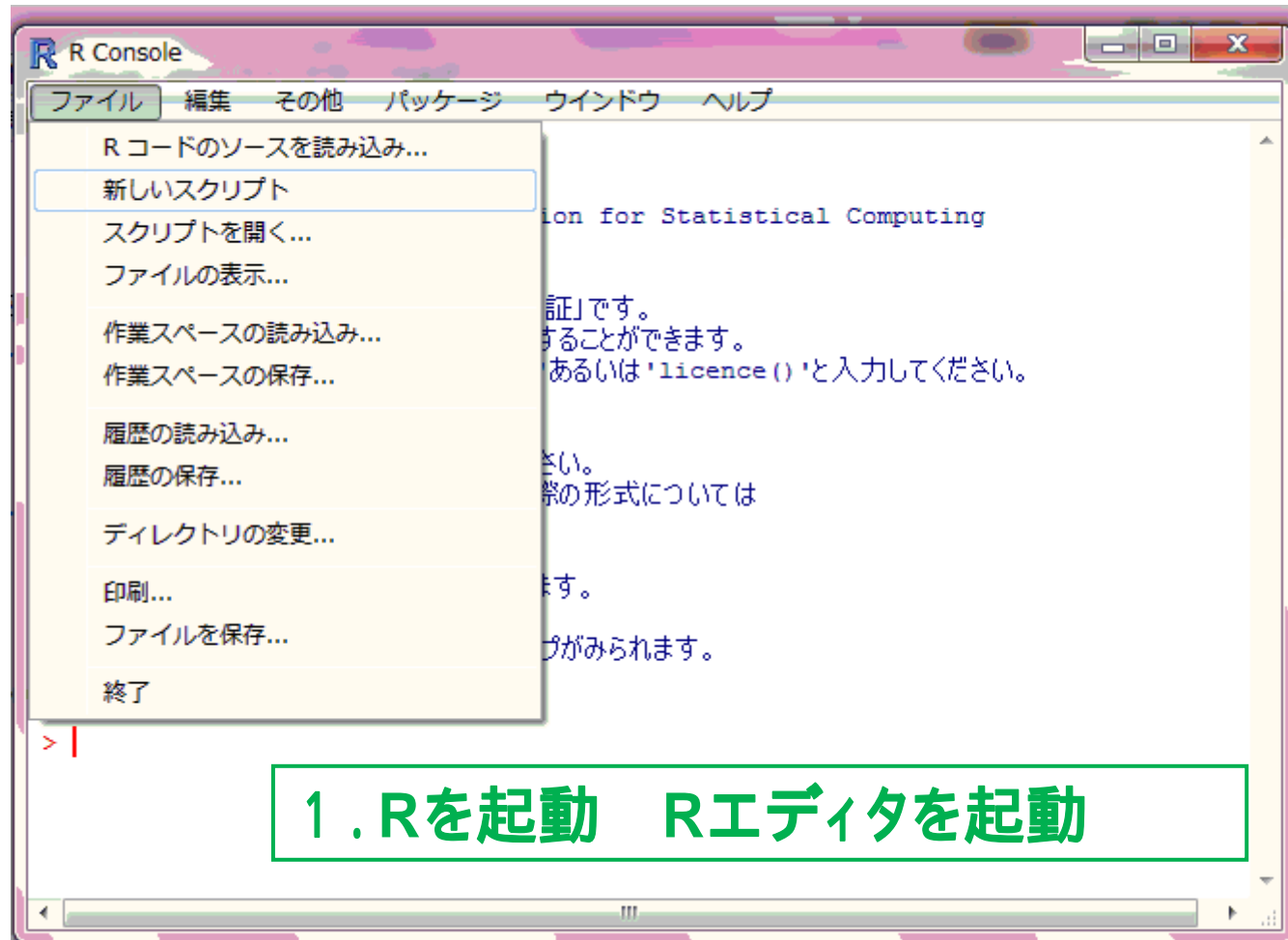
a. 従属変数 協調性

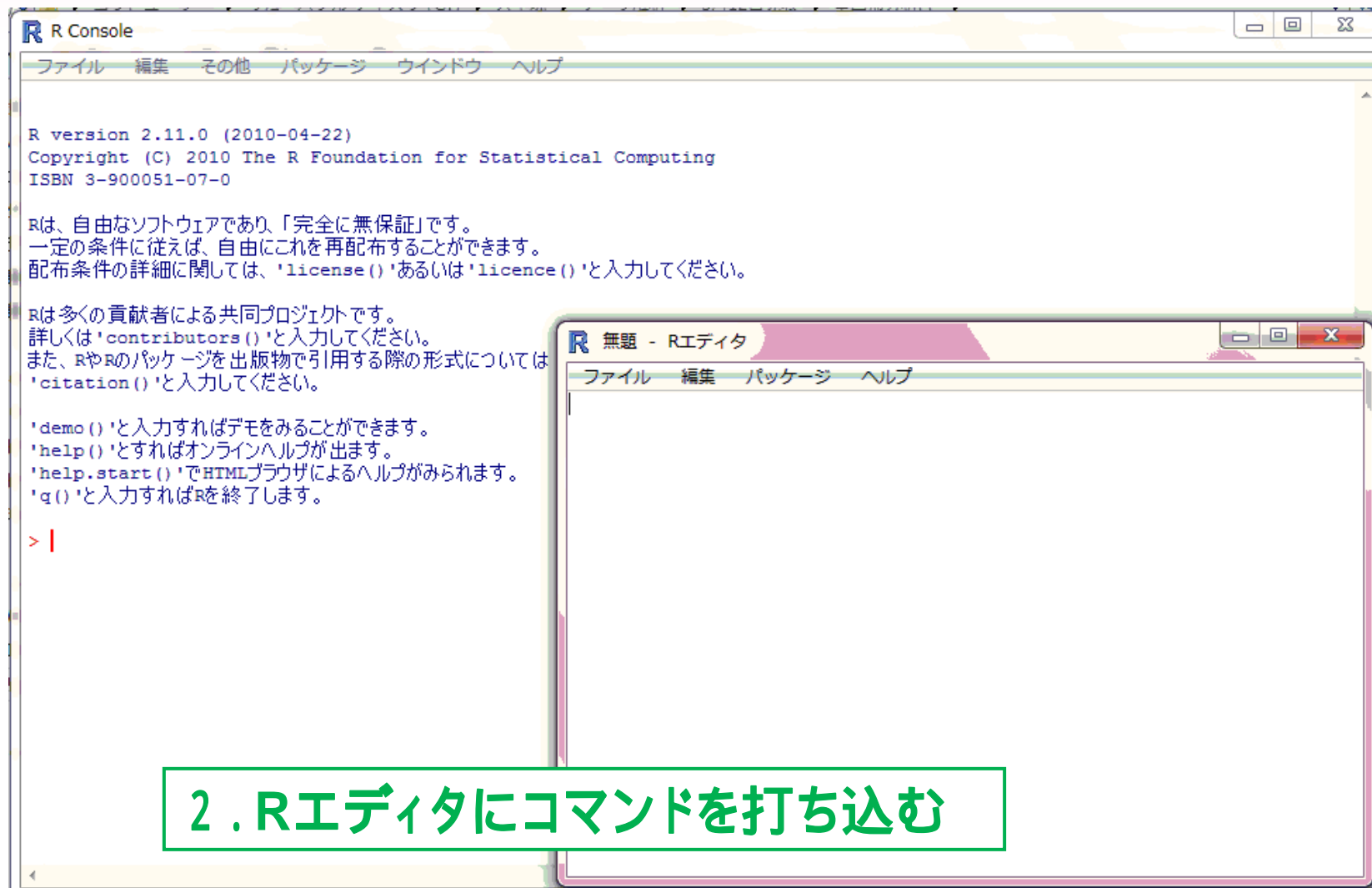
### 共線性の診断<sup>a</sup>

モデル	次元	固有値	条件指数	分散プロパティ		
				(定数)	母親価値	通園年数
1	1	2.940	1.000	.00	.00	.01
	2	.046	8.022	.23	.03	.83
	3	.015	14.166	.77	.97	.16

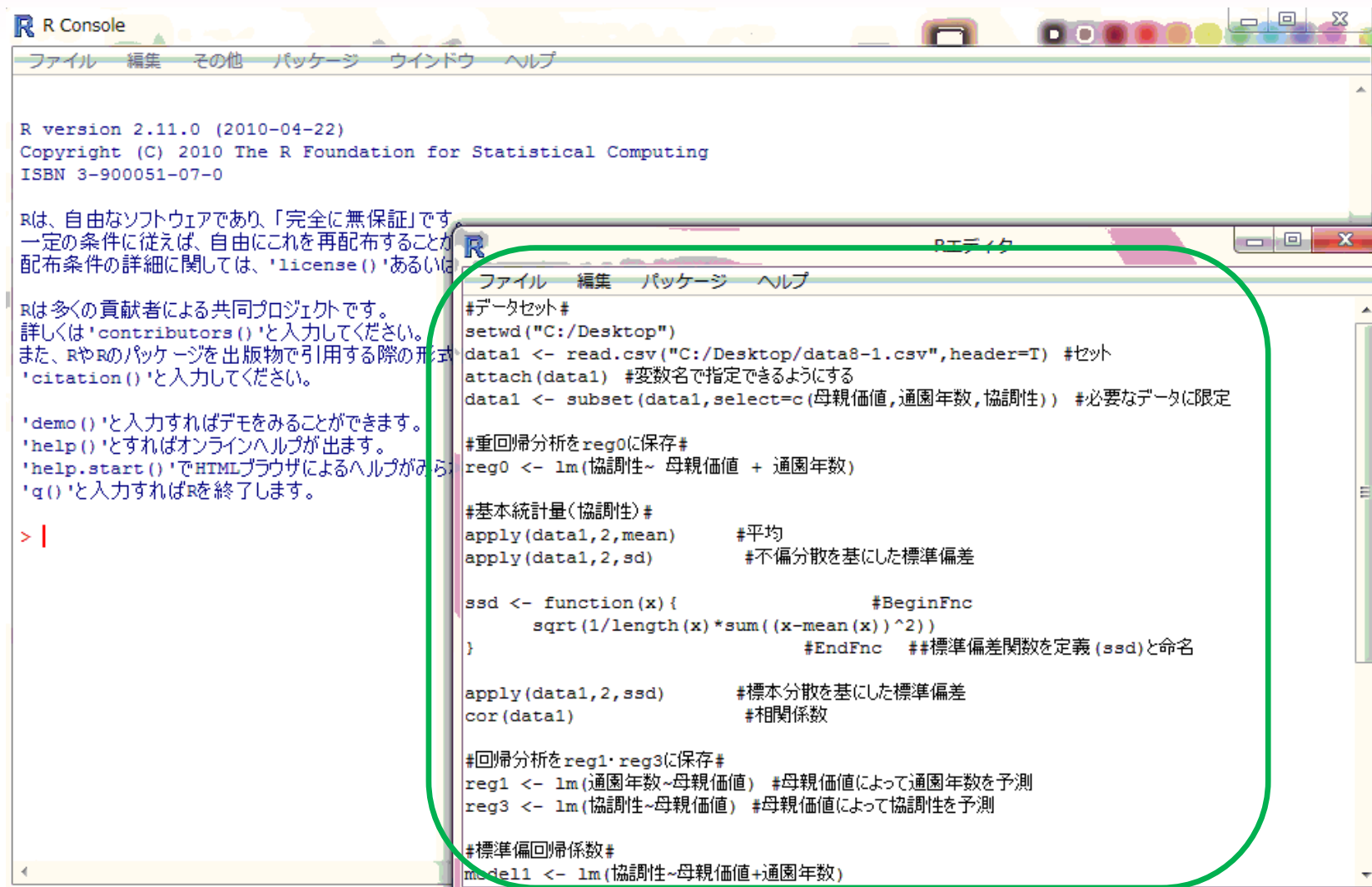
a. 従属変数 協調性

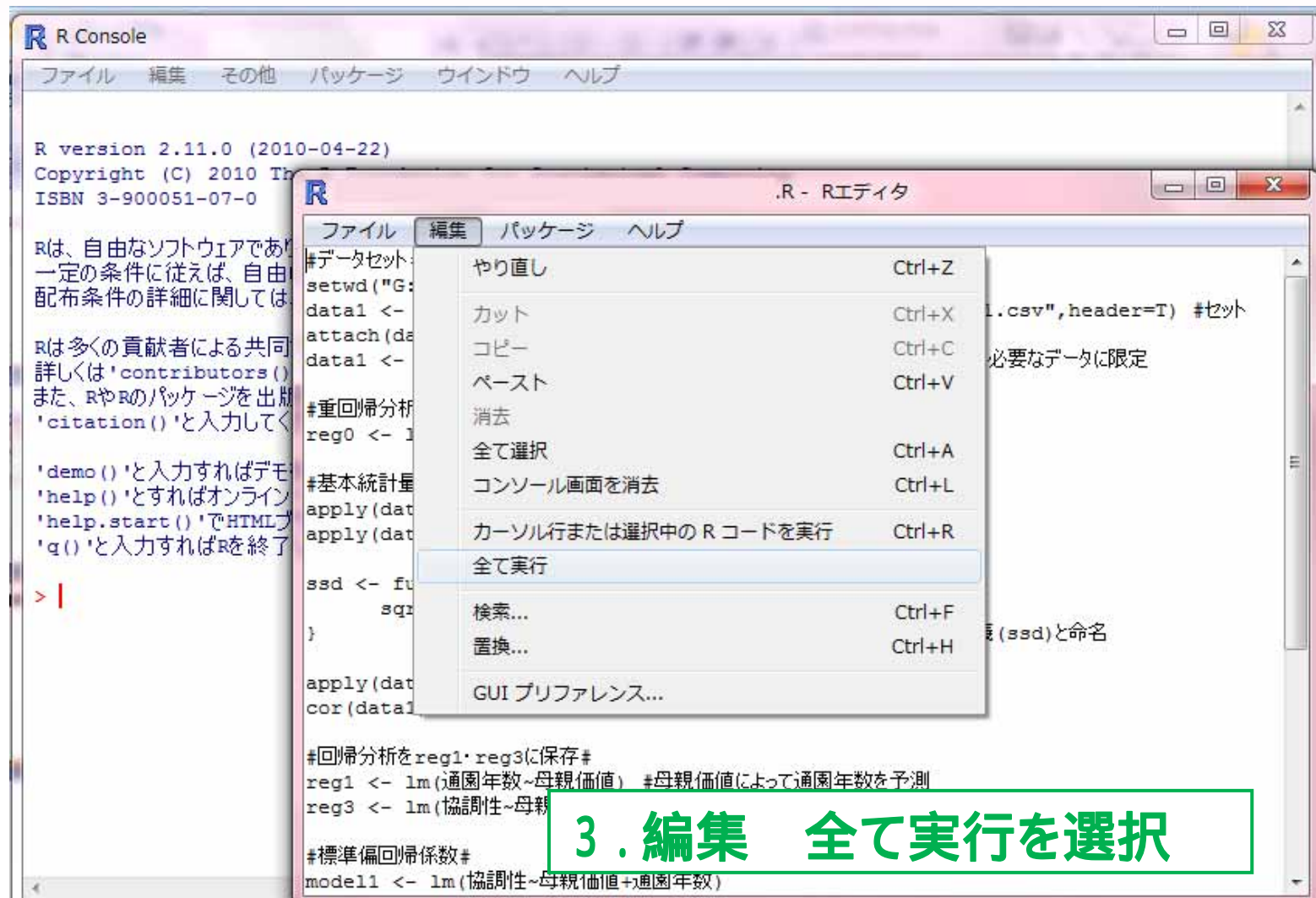
# Rによる重回帰分析





## 2. Rエディタにコマンドを打ち込む





### 3. 編集 全て実行を選択

```
R Console
ファイル 編集 その他 パッケージ ウィンドウ ヘルプ

>
> #基本統計量(協調性)#
> apply(data1,2,mean)#平均
母親価値 通園年数 協調性
13.20 2.48 11.96
> apply(data1,2,sd) #不偏分散を基にした標準偏差
母親価値 通園年数 協調性
2.4494897 0.7623808 2.5068071
>
> ssd <- function(x){#BeginFnc
+ sqrt(1/length(x)*sum((x-mean(x))^2))
+ }#EndFnc ##標準偏差関数を定義(ssd)と命名
>
> apply(data1,2,ssd)#標本分散を基にした標準偏差
母親価値 通園年数 協調性
2.4248711 0.7547185 2.4816124
>
> cor(data1) #相関係数
母親価値 母親価値 通園年数 協調性
母親価値 1.0000000 0.5376770 0.4799263
通園年数 0.5376770 1.0000000 0.5014633
協調性 0.4799263 0.5014633 1.0000000
>
> #回帰分析をreg1・reg3に保存#
> reg1 <- lm(通園年数~母親価値) #母親価値によって通園年数を予測
> reg3 <- lm(協調性~母親価値) #母親価値によって協調性を予測
>
> #標準偏回帰係数#
> model1 <- lm(協調性~母親価値+通園年数)
> std.est <- function(model){
+ sdd <- c(0,sd(model$model[-1]))
+ std.est <- coef(model)*sdd/sd(model$model[1])[1]
+ print(std.est)
+ }
> std.est(model1)
(Intercept) 母親価値 通園年数
0.0000000 0.2958223 0.3424065
```

1

2

3



```
R Console
ファイル 編集 その他 パッケージ ウィンドウ ヘルプ

> #分散説明率(決定係数)#
> summary(reg0)

Call:
lm(formula = 協調性 ~ 母親価値 + 通園年数)

Residuals:
    Min       1Q   Median       3Q      Max
-5.05629 -1.47805 -0.05629  1.47510  4.94371

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    5.1716     1.6647   3.107  0.00320 **
母親価値        0.3027     0.1467   2.064  0.04456 *
通園年数        1.1259     0.4713   2.389  0.02096 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.12 on 47 degrees of freedom
Multiple R-squared: 0.3137,    Adjusted R-squared: 0.2845
F-statistic: 10.74 on 2 and 47 DF,  p-value: 0.0001440

> summary(reg0)$r.squared#決定係数だけを取り出す
[1] 0.3136772
> summary(reg0)$adj.r.squared#自由度調整済み決定係数だけ取り出す
[1] 0.2844719
> summary(reg0)$fstatistic      #F値と自由度を取り出す
      value      numdf      dendf
10.74045    2.00000    47.00000
>
> #attach解除#
> detach(data1)
> |
```

4  
5

```
R Console
ファイル 編集 その他 パッケージ ウィンドウ ヘルプ

Residuals:
      Min       1Q   Median       3Q      Max
-5.05629 -1.47805 -0.05629  1.47510  4.94371

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    5.1716     1.6647   3.107  0.00320 **
母親価値       0.3027     0.1467   2.064  0.04456 *
通園年数       1.1259     0.4713   2.389  0.02096 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.12 on 47 degrees of freedom
Multiple R-squared:  0.3137,    Adjusted R-squared:  0.2845
F-statistic: 10.74 on 2 and 47 DF,  p-value: 0.0001440

> summary(reg0)$r.squared#決定係数だけを取り出す
[1] 0.3136772
> summary(reg0)$adj.r.squared#自由度調整済み決定係数だけ取り出す
[1] 0.2844719
> summary(reg0)$fstatistic      #F値と自由度を取り出す
      value    numdf    dendif
10.74045    2.00000    47.00000
>
> #VIF(DAAGインストール・読み込み後)#
> result <- lm(協調性~母親価値+通園年数) #従属変数が2つの回帰モデル
> vif(result) # VIF チェックの関数
母親価値 通園年数
   1.4067   1.4067
>
> #attach解除#
> detach(data1)
> |
```

6

# 参考文献

- 南風原朝和(2002) 心理統計学の基礎 統合的理解のために 有斐閣アルマ
- 渡部洋(2002) 心理統計の技法 福村出版
- 重回帰分析(魚野;2006) <http://kyoumu.educ.kyoto-u.ac.jp/cogpsy/personal/Kusumi/datasem06/uono.pdf>
- フリー統計ソフトRを用いた分散分析(田中;2007) <http://kyoumu.educ.kyoto-u.ac.jp/cogpsy/personal/Kusumi/datasem07/tanaka.pdf>
- Rによる重回帰分析(猪原;2007) <http://www.educ.kyoto-u.ac.jp/cogpsy/personal/Kusumi/datasem07/inohara.pdf>
- R-Tips <http://cse.naro.affrc.go.jp/takezawa/r-tips/r2.html>
- Passepied ( **奥村泰之の情報公開** ) <http://blue.zero.jp/yokumura/index.html>
- Use R! 回帰分析編 [http://forest-environment.cocolog-nifty.com/forest/files/use\\_r.pdf](http://forest-environment.cocolog-nifty.com/forest/files/use_r.pdf)
- 中部大学小塩研究室 <http://psy.isc.chubu.ac.jp/~oshiolab/>
- アイスクリーム屋さんで学ぶ楽しい統計学  
<http://kogolab.jp/elearn/icecream/index.html>

Webページは2010/05/09最終参照