

### R 入門 (3) 仮説検定

テキスト p-171 GDP 成長率の AR1 モデル(1981:1~2007:4) 1991:2 に構造変化?

構造変化あり  $y_t = (\beta_0 + \theta_0 D_t) + (\beta_1 + \theta_1 D_t)x_t + \varepsilon_t = \beta_0 + \theta_0 D_t + \beta_1 x_t + \theta_1 D_t x_t + \varepsilon_t$

構造変化なし  $y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$

#### 0 準備

(1) ホームページ上の `yabu23.xlsx` の sheet "09" を `09.csv` として `c:\$dail` に保存

(2) R を起動して、作業フォルダを `c:\$dail` にしてください。

`setwd("dail")`

(3) 拡張 package 導入

`install.packages("lmtest")`

このとき(ミラーサイト(download 先)を指定する:Japan(Tokyo))

`install.packages("dplyr")`

`library(lmtest)` waldtest コマンドで必要

`library(dplyr)` lag コマンドで必要

#### 1 データの読み込み

`dat=read.table("09.csv", header = TRUE, sep = ",")` <-の代わりに = でも OK  
`attach(dat)`

#### 2 変数作成

`d=ifelse(date>=1991.2, 1, 0)` date==1991.2 なら 1991 年 Q2 ダミー

`y1=lag(growth, n=1)` ダミー変数 1991 年 Q2 以降は 1, それ以前は 0

`y1d=y1*d` ラグ変数  $Y_{t-1}$  を  $y1$  とする

#### 3. 推定

`reg0=lm(growth~y1)` 構造変化なし

`reg1=lm(growth~y1+d+y1d)` 構造変化あり

`summary(reg0)` p-172

`summary(reg1)`

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	2.2968	0.5826	3.943	0.000147	***
y1	0.5041	0.1216	4.144	7.01E-05	***
d	-1.947	0.6105	-3.189	0.001891	**
y1d	0.1735	0.151	1.149	0.253233	
Signif. codes: ***'0.001'**'0.01'*'0.05 '.'0.1					

#### 4. 検定

`waldtest(reg0, reg1)` Wald 検定(F 検定) p-172

	Res.Df	Df	F	Pr(>F)	
1	105				
2	103	2	9.2546	0.000201	***

#### 演習

1. Sheet "04" 家賃データを利用し、家賃を専有面積、距離、築年数、階数で説明する重回帰をおこない、R の結果を貼り付けなさい (anova は不要)。

家賃 = 4.735 + 0.1651 専有面積 - 0.07787 距離 - 0.06579 築年数 + 0.2156 階数

(38.67)(64.78) (-12.05) (-25.23) (7.411) 括弧内は t 値

$R^2=0.8835 \bar{R}^2=0.8828$  標準誤差 = 0.7682

2階数に加えて2Fダミー(2Fなら1,それ以外は0),3Fダミー,4Fダミー,5F以上ダミーを加えたモデル  $家賃 = b_0 + b_1 \text{専有面積} + b_2 \text{距離} + b_3 \text{築年数} + b_4 \text{階数} + b_5 d2 + b_6 d3 + b_7 d4 + b_8 d5$  を推定し、下記の問い合わせに答えなさい。

- (1)結果を上記と同様に整理しなさい
  - (2)なぜ1Fダミーd1を除くのかを説明しなさい
  - (3)階数の違いの家賃への影響について1の結果との違いを説明しなさい
  - (4)1と2の結果に有意な違いがあるかを検定しなさい

```

setwd("/dail")
install.packages("lmtest")
library(lmtest)
dat=read.table("04.csv",header = TRUE, sep =",")
attach(dat)
reg1=lm(rent~space+distance+age+floor)
summary(reg1)
anova(reg1)
d2=ifelse(floor==2,1,0)
d3=ifelse(floor==3,1,0)
d4=ifelse(floor==4,1,0)
d5=ifelse(floor>=5,1,0)
reg2=lm(rent~space+distance+age+floor+d2+d3+d4+d5)
summary(reg2)
anova(reg2)
waldtest(reg1,reg2)

```

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	4.7352	0.122463	38.666	< 2e-16	***
space	0.165071	0.002548	64.779	< 2e-16	***
distance	-0.07787	0.006462	-12.051	< 2e-16	***
age	-0.06579	0.002607	-25.234	< 2e-16	***
floor	0.215639	0.029098	7.411	3.53e-13	***
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1					
Residual standard error: 0.7682 on 719 degrees of freedom					
R2: 0.8835	Adjusted R2: 0.8828				
F-statistic:	p-value: < 2.2e-16				
SSR: 424.26					

重回帰です。基本です。Rの結果を示していない人は減点。

2

- (1) 結果を上記と同様に整理しなさい

家賃 = 5.159 + 0.1652 専有面積 - 0.07878 距離 - 0.06582 築年数 - 0.2236 階数  
 (18.82) (63.70) (-12.23) (-25.32) (0.2470)  
 + 0.5122 階 2 + 0.8047 階 3 + 1.243 階 4 + 2.518 階 5 括弧内は t 値  
 (2.001) (1.604) (1.653) (2.076)

R2=0.8851 R<sup>-2</sup>=0.8838 標準誤差=0.7650

- (2) なぜ 1F ダミー d1 を除くのかを説明しなさい

D1±D2±D3±D4±D5=1 となり識別できないため(完全な多重共線性)

(3) 階数の違いの家賃への影響について 1 の結果との違いを説明しなさい  
 階数による家賃の差は線形とはならないことが判る。

(4) 1 と 2 の結果に有意な違いがあるかを検定しなさい  
 有意水準 5%で有意な差があり、 $H_0: \beta_6 = \beta_7 = \beta_8 = \beta_9 = 0$  は棄却される

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	5.158555	0.274046	18.824	<2e-16	***
space	0.165221	0.002594	63.695	<2e-16	***
distance	-0.07878	0.006443	-12.227	<2e-16	***
age	-0.06582	0.0026	-25.317	<2e-16	***
floor	-0.22355	0.247021	-0.905	0.3658	
d2	0.512223	0.255979	2.001	0.0458	*
d3	0.804705	0.501634	1.604	0.1091	
d4	1.243168	0.751877	1.653	0.0987	.
d5	2.517858	1.212847	2.076	0.0383	*
Signif. codes:	0 '***'	0.001 '**'	0.01 '*'	0.05 '.'	0.1 ' '
Residual standard error:	0.765	on 715 degrees of freedom			
R2:	0.8851	Adjusted R2:	0.8838		
F-statistic:	p-value: < 2.2e-16				
SSR:	418.41				

Wald test				
Model 1: rent ~ space + distance + age + floor				
Model 2: rent ~ space + distance + age + floor + d2 + d3 + d4 + d5				
	Res.Df	Df	F	Pr(>F)
Reg 1	719			
Reg2	715	4	2.4977	0.04153 *
Signif. codes:	0 '***'	0.001 '**'	0.01 '*'	0.05 '.'

ダミー変数が作れていない人がいたが、早めに R を動かしてください。試行錯誤しながら、R のくせを理解してください。

複数のダミー変数を一括して説明変数に入れるというイメージが沸かないためか、  
`lm(rent ~ space + distance + age + floor + d2)`  
`lm(rent ~ space + distance + age + floor + d3)`  
 のように 1 つづつ追加して 4 つの回帰分析をおこなっている人が複数いました。  
 一括して説明変数に入れることで、1 階を基準として、各階（2 階..5 階以上）の効果を見ることができます