

第4章 回帰分析の諸問題（1）

第1節 多重共線性

1. 多重共線性の例

多重共線性の症状について、数値例を用いて確認してみる。次のプログラムを入力し、実行してみよう。（ファイル名は **multico.R** としておく）

```
Y <- c(13, 4, 8, 9, 6)
X <- c(4, 1, 2, 5, 3)
Y <- c(8, 2, 4, 10, 6)
#
# 完全な多重共線性
#
data1 <- data.frame(Y=Y, X=X, Z=Z)
data1
cor(data1)
reg1 <- lm(Y~X+Z, data=data1)
#
# 多重共線性その1
#
summary(reg1)
X[1] <- 4.01
data2 <- data.frame(Y=Y, X=X, Z=Z)
data2
cor(data2)
reg2 <- lm(Y~X+Z, data=data2)
summary(reg2)
#
# 多重共線性その2
#
X[1] <- 4.05
data3 <- data.frame(Y=Y, X=X, Z=Z)
data3
cor(data3)
reg3 <- lm(Y~X+Z, data=data3)
summary(reg3)
```

2. VIF値

上の最後の例について VIF 値を計算してみる。end 文の前に次のようなコマンドを入力してみよう。

```
#
# VIF 値
#
vifregx <- lm(X~Z, data=data3)
vifx <- 1/(1-summary(vifregx)$r.squared)
vifregz <- lm(Z~X, data=data3)
vifz <- 1/(1-summary(vifregz)$r.squared)
vif1 <- data.frame(vifx=vifx, vifz=vifz)
vif1
```

☒ 演習問題 : **ex3-3.R** について、同様に VIF 値を求めるコマンドを追加してみよう。そのプログラムは **ex3-3multico.R** という名前でも保存しておくこと。

※ VIF 値の計算には、DAAG パッケージを利用することも可能である。その手順は

- ① メニューバーの「パッケージ」－「パッケージのインストール」で、CRAN のミラー (JAPAN(Tsukuba)を推奨)を選び、DAAG パッケージを選ぶ。
- ② メニューバーの「パッケージ」－「パッケージの読み込み」で、DAAG パッケージを選ぶ。
- ③ プログラムで、`vif(reg3)`の 1 行を加えれば、VIF が計算できる。

第 2 節 系列相関

3. Excelをもちいた系列相関の判定と対処法

3 章で実習した `ex3-3.csv` のデータについて、ダービン・ワトソン比を計算して系列相関の判定し、対処法である、コ克蘭・オーカット法、一般化最小 2 乗法を適用してみよう。

📖 手順

- ① CE を被説明変数、YD を説明変数とする単回帰分析を、分析ツールを用いておこない、結果を `ex3-3sc.xlsx` の名前で保存する。
- ② 標準残差の右側 (E 列) に残差を 1 期ずらしてコピーし、 $e(t-1)$ を作成する。C 列の 1 期から 19 期のデータを E 列の 2 期から 20 期までにコピーする。
- ③ F 列に $e(t)-e(t-1)$ を計算する。F26 セルに「=C26-E26」と入力し、それをコピーする。
- ④ G 列に $(e(t)-e(t-1))^2$ を計算 (G26 セルなら「=F26^2」) し、その合計を求める。
- ⑤ H 列に $e(t)^2$ を計算 (H25 セルなら「=C25^2」) し、その合計を求める。
- ⑥ G 列の合計を H 列の合計で割ったものがダービン・ワトソン比となる。

<コ克蘭・オーカット法>

📖 手順

- ① $e(t)$ を $e(t-1)$ に対して回帰する。分析ツールを用いてよい。ここで「定数に 0 を使用」にチェックすること。
- ② 分析結果の X 値 1 の係数を用いて Y_t^* , X_t^* を計算する。この係数は絶対参照すること。
- ③ Y_t^* を X_t^* に対して回帰する。これがコ克蘭・オーカット法によって求めた係数推定値である。b の推定値は分析結果そのままであるが、a の推定値はこの結果を $1-\rho$ で割ったものになる。また、この結果について残差プロットやダービン・ワトソン比によって系列相関への対処がなされたことを確認すること。

<一般化最小 2 乗法>

📖 手順

- ① コ克蘭・オーカット法と同様に $e(t)$ を $e(t-1)$ に対して回帰する。コ克蘭・オーカット法の計算のすでに計算してあるので、このステップは省略できる。
- ② 分析結果の X 値 1 の係数を用いて Y_t^* , X_t^* , Z_t^* を計算する。
- ③ Y_t^* を X_t^* , Z_t^* に対して回帰する。「定数に 0 を使用」にチェックすること。このようにしてもとめられたものが一般化最小 2 乗法による係数推定値である。

4. Rをもちいた系列相関の判定と対処法

Excel では、系列相関の判定や対処法に若干の工夫が必要であった。R ではそれが容易におこなうことができる。次ページのプログラムを作成し、`ex3-3sc.R` という名前で保存せよ。

ただし、このプログラム中で使用した、ダービン・ワトソン検定をおこなう `dwtest` コマンドは、`lmtest` パッケージを読み込む必要がある。そのため、次の手順を最初におこなう必要がある。

- ① メニューバーの「パッケージ」－「パッケージのインストール」で、CRAN のミラー (JAPAN(Tsukuba)を推奨)を選び、`lmtest` パッケージを選ぶ。
- ② メニューバーの「パッケージ」－「パッケージの読み込み」で、`lmtest` パッケージを選ぶ。

```

data1 <- read.table("ex3-3.csv", header=TRUE, sep=",")
data1
#
# 単回帰と残差プロット
#
reg1 <- lm(CE~YD, data=data1)
summary(reg1)
windows()
plot(resid(reg1))
dwtest(reg1)
#
#  $\rho$  の導出とコ克蘭・オーカット法
#
resid1<-resid(reg1)
resid1_lag1<-c(0, resid(reg1)[1:length(resid1)-1])
CE_lag1<-c(0, data1$CE[1:length(data1$CE)-1])
YD_lag1<-c(0, data1$YD[1:length(data1$YD)-1])
dataset_dw <- cbind(data1, resid1, resid1_lag1, CE_lag1, YD_lag1)
rho1<-coef(summary(lm(resid1~resid1_lag1-1)))[1, 1]
CE_star<-cbind(c(data1$CE[2:length(data1$CE)]-rho1*CE_lag1[2:length(data1$CE)]))
YD_star<-cbind(c(data1$YD[2:length(data1$YD)]-rho1*YD_lag1[2:length(data1$YD)]))
dataset_dw2<-cbind(CE_star, YD_star)
cochrane_orcutt<-lm(CE_star~YD_star)
summary(cochrane_orcutt)
dwtest(cochrane_orcutt)

```

第3節 不均一分散

5. 不均一分散の対処法

hetero.csv は 27 の会社の労働者数と管理者数のデータである。このデータについて不均一分散があるかどうかを調べてみる。

残差プロットと **Breusch-Pagan test** の結果から不均一分散が示唆される。

対処法として、 Y/X を被説明変数、 $1/X$ を説明変数とする回帰分析を行うことがある。これによって不均一分散の問題は解決できたといえよう。(下のプログラムを作成し、**hetero.R** という名前で保存せよ。なお、このプログラムも実行する前に、**lmtest** パッケージを読み込む必要がある。)

```

data1 <- read.table("hetero.csv", header=TRUE, sep=",")
data1
reg1 <- lm(y~x, data=data1)
summary(reg1)
windows()
plot(resid(reg1))
bptest(reg1)
#
# 変数変換による対処
#
yy=data1$y/data1$x
xx=1/data1$x
reg2 <- lm(yy~xx)
summary(reg2)
plot(resid(reg2))
bptest(reg2)

```