5.1.2 R による計算（原始的な方法）

5.1.1 節で見た方法をそのまま用いれば、α、β の推定量を求めることが可能である。ここでは、下に示す仮想的なデータである「データセット A」について、OLS 推定量を求める問題を考える。

データセット A は以下のようなデータであり、個体を識別する ID（Ind）、時間（time）ならびに仮想の変数である X と Y からなるパネルデータとなっている。以下の R による実行例では、"Ydf" と名付けられたデータフレームに格納して取り扱う。

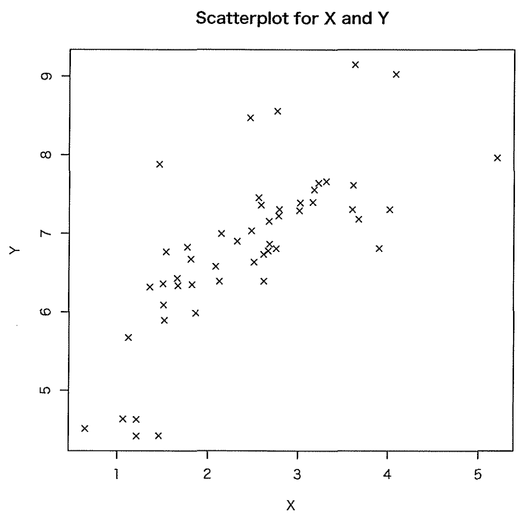
データセット A(データフレーム Ydf) の内容

```
> Ydf
   Ind time        Y        X
1    1    1 6.318302 1.358518
2    1    2 6.765785 1.538198
3    1    3 6.673199 1.812872
4    1    4 7.361044 2.589778
5    1    5 7.305603 2.791172
6    2    1 4.512218 0.652563
7    2    2 4.636834 1.068677
8    2    3 4.630790 1.213461
9    2    4 4.419180 1.215412
10   2    5 4.423203 1.461623
11   3    1 6.633481 2.513192
12   3    2 6.392095 2.623966
13   3    3 6.863346 2.684149
14   3    4 7.303656 4.019452
15   3    5 7.961180 5.209157
16   4    1 5.674991 1.122794
17   4    2 6.089810 1.512576
18   4    3 5.894734 1.518251
19   4    4 5.986789 1.865590
20   4    5 6.809329 3.904122
21   5    1 7.880392 1.463115
22   5    2 8.470850 2.468216
23   5    3 8.556446 2.769034
24   5    4 9.147444 3.630287
25   5    5 9.024920 4.083941
26   6    1 6.348810 1.826399
27   6    2 6.390482 2.129157
28   6    3 6.733463 2.619354
29   6    4 6.777782 2.668968
30   6    5 6.806236 2.758063
31   7    1 6.996998 2.147962
32   7    2 7.032473 2.483352
33   7    3 7.397037 3.164326
34   7    4 7.663331 3.312535
35   7    5 7.613489 3.613653
36   8    1 6.824914 1.773091
37   8    2 7.454234 2.563154
38   8    3 7.390236 3.023997
39   8    4 7.555790 3.179776
40   8    5 7.641978 3.227601
41   9    1 6.358868 1.505719
42   9    2 6.578637 2.088625
43   9    3 7.154565 2.677834
44   9    4 7.222814 2.783386
45   9    5 7.182551 3.674878
46  10    1 6.428335 1.663196
47  10    2 6.335629 1.670820
48  10    3 6.898983 2.324702
49  10    4 7.290997 3.015024
50  10    5 7.305412 3.604515
```

データセット A の変数 X と Y の関係をプロットすると以下のようになる。

パッケージとプロット

```
plot(Ydf$X, Ydf$Y, type="p", pch=4,
 main = "Scatterplot for X and Y",
 xlab = "X", ylab = "Y")
```



次に、5.1.1 節で述べた理論式を直接用いる「原始的な方法」によって、α、β の推定量を求めると以下の通りである。

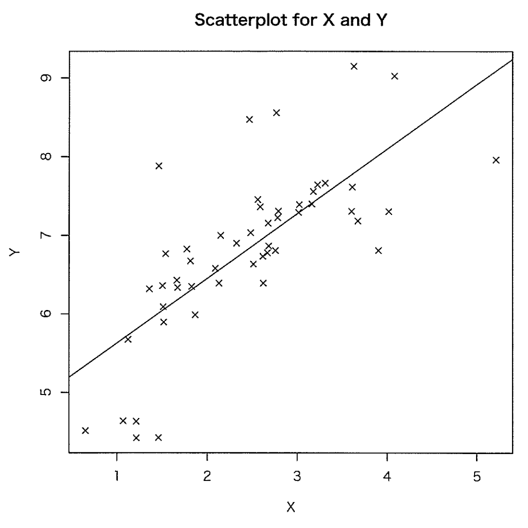
R による計算（原始的な方法）

```
V_XY <- cov(cbind(Ydf$X, Ydf$Y))
beta_hat <- V_XY[1,2] / V_XY[1,1]
x_bar <- mean(Ydf$X)
y_bar <- mean(Ydf$Y)
alpha_hat <- y_bar - x_bar * beta_hat
print(c(alpha_hat, beta_hat))

plot(Ydf$X, Ydf$Y, type="p", pch=4,
    main = "Scatterplot for X and Y",
    xlab = "X", ylab = "Y")
abline(alpha_hat, beta_hat)
```

出力結果

```
> print(c(alpha_hat, beta_hat))
[1] 4.807486 0.821806
```



5.1.3 R による計算（関数 lm を利用する方法）

R には線形回帰モデルを推定するための関数 lm が用意されている。これを利用すれば、α、β の推定量のみならず、線形回帰モデルに関する様々な推定量を得ることが可能である。

lm の中には、モデル式といわれる形式で回帰式を記述する。この場合、Ydf というデータフレームの Y を X で説明するという意味になる。説明変数が二つ以上あるときは"+"で結ぶ。結果のサマリーは summary 関数で得られる。なお、ここで得られる OLS 推定量は、同一個体が複数回含まれるデータをすべてプールして推定していることからプーリング推定量（pooling estimates）と呼ばれる。

R による計算（関数 lm を利用する方法）

```
lm.ols <- lm(Y ~ X, data = Ydf)
summary(lm.ols)
plot(Ydf$X, Ydf$Y, type="p", pch=4,
    main = "Scatterplot for X and Y",
    xlab = "X", ylab = "Y")
abline(lm.ols)
```

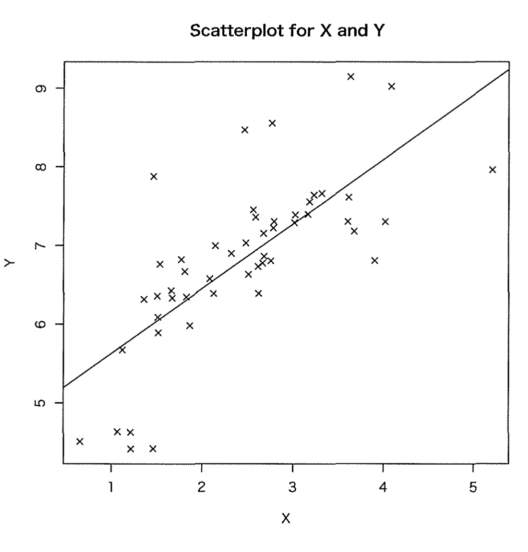
出力結果

```
Call:
lm(formula = Y ~ X, data = Ydf)

Residuals:
     Min       1Q   Median       3Q      Max
-1.58545 -0.26072  0.04754  0.29899  1.87051

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.8075     0.2888   16.65  < 2e-16 ***
X             0.8218     0.1100    7.47 1.40e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7288 on 48 degrees of freedom
Multiple R-squared: 0.5376,Adjusted R-squared: 0.5279
F-statistic:  55.8 on 1 and 48 DF,  p-value: 1.405e-09
```



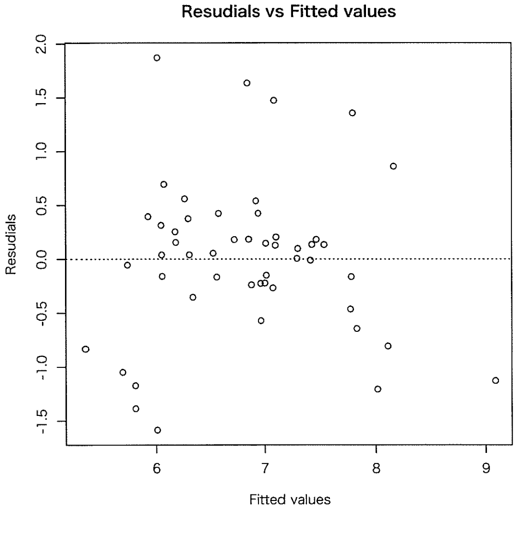
5.2 回帰モデルの残差

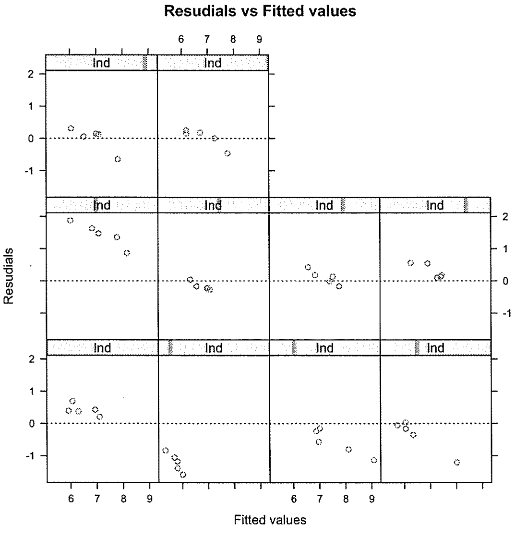
データセット A の変数 Y を変数 X で説明する回帰モデルについて、モデルにより推定された値に対する残差をプロットしてみよう。

モデルにより推定された値に対する残差をプロット

```
plot(predict(lm.ols), residuals(lm.ols),
    main = "Resudials vs Fitted values",
    xlab = "Fitted values", ylab = "Resudials")
abline(h=0, lty=3)

library(lattice)
xyplot(residuals(lm.ols) ~ predict(lm.ols) | Ind, data=Ydf,
    panel = function(x,y){
        panel.xyplot(x,y)
        panel.abline(h=0, lty=3)
        },
    main = "Resudials vs Fitted values",
    xlab = "Fitted values", ylab = "Resudials")
```





ここから、各個体別に残差を観察した場合、推定された値が大きくなるほど残差が減少する傾向が見られることがわかる。これは、観察不可能な個体の効果の存在を示唆している。

5.3 パネルデータの表示法

パネルデータは、同一の個体に対して複数時点での観察を行うことから、クロスセクションデータと時系列データの両方を併せ持ったデータであるといえる。したがって、パネルデータに対する回帰分析は、通常のクロスセクションデータや時系列データと異なり、変数に個体番号と時刻の2つの添字を併せ持っている。

いま、N 個の個体を $i=1,\cdots N$ で表し、時刻を $t=1,\cdots,T$ で表す。変数としては、被説明変数 y_{it} と、これに対する K 個の説明変数 $X'_{it}=\left[\begin{array}{cccc} x_{1,it} & x_{2,it} & \cdots & x_{K,it}\end{array}\right]$ を考える。このとき、定数項を α、定数項以外の回帰係数を $\beta'=\left[\begin{array}{cccc}\beta_1 & \beta_2 & \cdots & \beta_K\end{array}\right]$、誤差項を u_{it} として、回帰式は、

$$y_{it}=\alpha+X'_{it}\beta+u_{it} \qquad i=1,\cdots N;\, t=1,\cdots,T$$

と書くことができる。本章では、一元配置誤差構成要素回帰モデル (One-way Error Component Regression Model) を対象とする。これは、誤差項が

$$u_{it}=\mu_i+\nu_{it}$$

と表されるモデルである。ここで、μ_i は観察不可能な個体の効果、ν_{it} は撹乱項である。

5.4 固定効果モデル

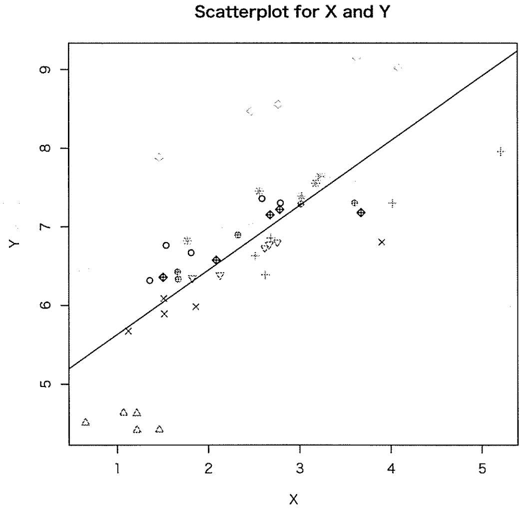
固定効果モデルとは、μ_i: 観察不可能な個体の効果をパラメータとして推定するモデルである。すなわち、傾きは同一だが、個体毎に切片が異なる直線で回帰をするモデルといえる。推定にあたっては、まず、異なる切片の影響を除去するため、X と Y から個体毎の平均値を引き去った変数を考えて、これに回帰を施すことにより傾きを推定する。

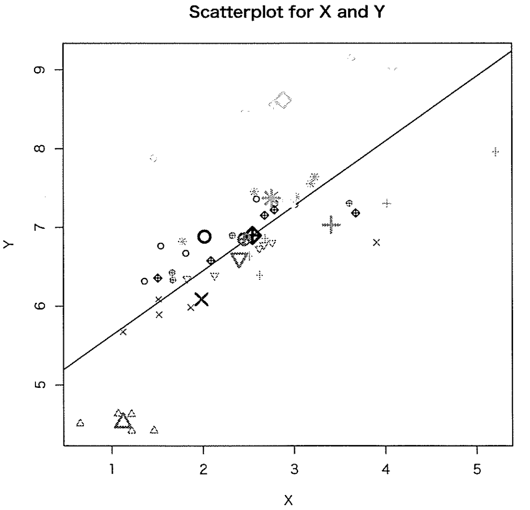
この原理を、データセット A を使ってグラフ上で考えてみよう。まず、データセット A の散布図を個体毎に色分けして描き直す。

データセット A の散布図（個体別）

```
plot(Ydf$X, Ydf$Y, type="p", col = Ydf$Ind, pch= Ydf$Ind,
    main = "Scatterplot for X and Y",
    xlab = "X", ylab = "Y")
abline(lm.ols)

# 個体毎の平均値を表示
Yb <- tapply(Ydf$Y,Ydf$Ind,mean)
Xb <- tapply(Ydf$X,Ydf$Ind,mean)
points(Xb,Yb, pch = seq_along(Yb), cex = 2, col = seq_along(Yb), lwd=3)
```



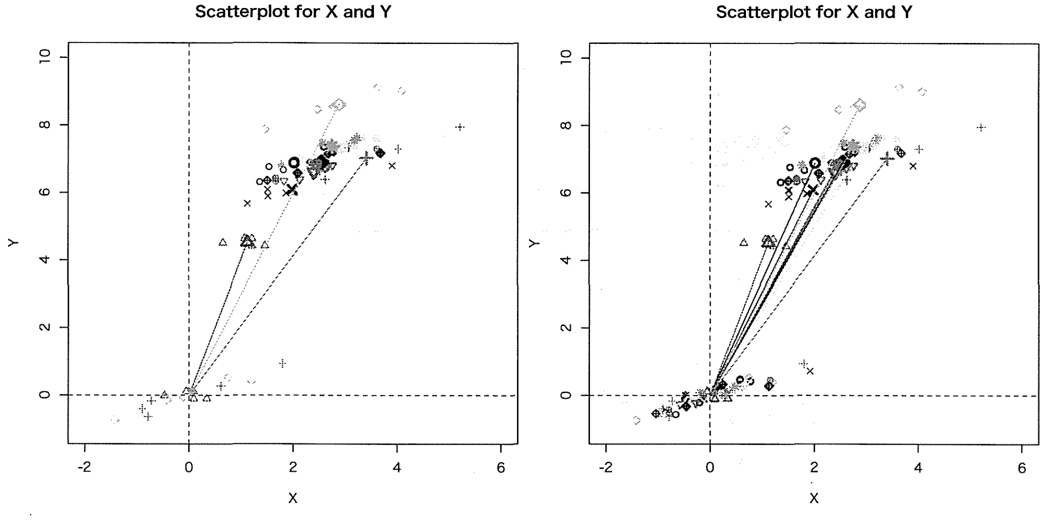


全データに当てはめた OLS の傾きが、同一個体毎に当てはめた OLS の傾きよりもやや大きいように見える。右は、個体毎に 5 時点の X,Y の平均値のポイントをを少し大きめのマーカーで表示したものである。個体毎の異質性を除いた傾きを推定するために、この大きなマーカーを原点に移すように移動する。

個体毎の平均値を引く

```
# 個体毎の平均値を引いたデータを作成
Yw <- Ydf$Y - Yb[Ydf$Ind]
Xw <- Ydf$X - Xb[Ydf$Ind]

plot(Ydf$X, Ydf$Y, col = Ydf$Ind, pch= Ydf$Ind, ylim=c(-1,10), xlim=c(-2,6),
    main = "Scatterplot for X and Y",
    xlab = "X", ylab = "Y")
points(Xw, Yw , col = Ydf$Ind, pch= Ydf$Ind)
points(Xb,Yb, pch = seq_along(Yb), cex = 1.5, col = seq_along(Yb), lwd=3)
abline(h=0, v=0, lty=2, lwd = 0.5)
arrows(Xb, Yb, 0, 0, col = seq_along(Yb), length = 0.1)
```



左の図は、水色・オレンジ・緑で示された個体についての原点への移動の様子を示したものである。点線は X 軸と Y 軸を示すことから、点線の交点が原点を示している。各個体の大きなマーカーから原点へ向かう矢印が、その個体が移動されるベクトルを示しており、原点の周りには移動されたデータが示されている。右の図はこの操作を全ての個体に対して適用した結果である。

このデータに線形回帰モデルを当てはめたときの傾きが固定効果モデルの X の回帰係数となる。

固定効果モデル (原始的な方法)

```
lm.w <- lm(Yw ~ Xw)
summary(lm.w)

plot(Ydf$X, Ydf$Y, col = Ydf$Ind, pch= Ydf$Ind, ylim=c(-1,10), xlim=c(-2,6),
    main = "Scatterplot for X and Y",
    xlab = "X", ylab = "Y")
points(Xw, Yw , col = Ydf$Ind, pch= Ydf$Ind)
points(Xb,Yb, pch = seq_along(Yb), cex = 1.5, col = seq_along(Yb), lwd=3)
abline(h=0, v=0, lty=2, lwd = 0.5)
abline(lm.w)
```

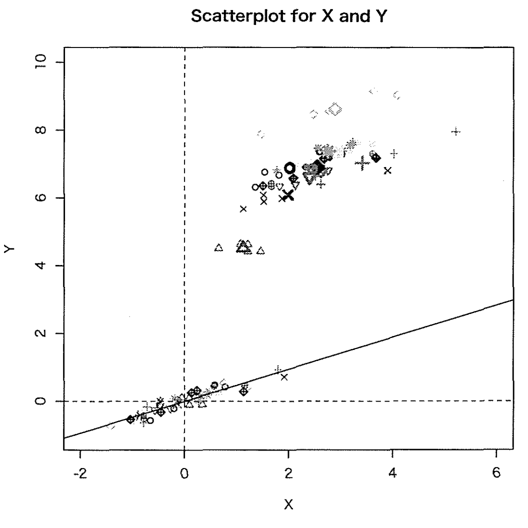
出力結果

```
Call:
lm(formula = Yw ~ Xw)

Residuals:
      Min        1Q    Median        3Q       Max
-0.267974 -0.090150  0.001145  0.100978  0.221316

Coefficients:
             Estimate Std. Error  t value Pr(>|t|)
(Intercept) 2.115e-16  1.919e-02 1.10e-14        1
Xw          4.716e-01  2.673e-02    17.64   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1357 on 48 degrees of freedom
Multiple R-squared: 0.8664,Adjusted R-squared: 0.8636
F-statistic: 311.3 on 1 and 48 DF,  p-value: < 2.2e-16
```



上の図では、X と Y の値によるプロットならびに X と Y の各値について個体内平均値との差分をとった値（偏差）によるプロットをそれぞれ示しており、後者に OLS 直線を当てはめている。

固定効果モデルは、図でわかるように、個体間の差異を無視して、個体内の変動に着目した推定であることから、Within(ウィズイン) 推定とも呼ばれる。これに対して、個体毎の平均値、すなわち、大きなマーカーに対して線形回帰モデルを当てはめたものを、Between(ビトウィーン) 推定と呼ぶ。なお、繰り返しとなるが、各個体についての全データをプールして OLS を得る場合はプーリング推定となる。

Between 推定 (原始的な方法)

```
lm.b <- lm(Yb ~ Xb)
summary(lm.b)


plot(Ydf$X, Ydf$Y, col = Ydf$Ind, pch= Ydf$Ind, ylim=c(-1,10), xlim=c(-2,6),
    main = "Scatterplot for X and Y",
    xlab = "X", ylab = "Y")
points(Xw, Yw , col = Ydf$Ind, pch= Ydf$Ind)
points(Xb,Yb, pch = seq_along(Yb), cex = 1.5, col = seq_along(Yb), lwd=3)
abline(h=0, v=0, lty=2, lwd = 0.5)
abline(lm.ols, lty=2, col=2, lwd=1.5)
abline(lm.w, lty=3, col=3, lwd=1.5)
abline(lm.b, lty=4, col=4, lwd=1.5)
legend("topleft", c("OLS(Pooled)", "Fixed Effect(Withih)", "Between"),
    lty=seq(2,4), col=seq(2,4), lwd = 1.5, cex=0.65)
```

出力結果

```
Call:
lm(formula = Yb ~ Xb)

Residuals:
    Min      1Q  Median      3Q     Max
-1.0564 -0.1404 -0.0810  0.1207  1.2245

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.5860     0.8635   4.153  0.00320 **
Xb            1.3200     0.3420   3.859  0.00481 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.651 on 8 degrees of freedom
Multiple R-squared: 0.6506,Adjusted R-squared: 0.6069
F-statistic: 14.89 on 1 and 8 DF,  p-value: 0.004814
```

これらの結果から、X の回帰係数について、プールド OLS では 0.82、固定効果モデルでは 0.47、Between 推定では 1.32 となることがわかった。実は、OLS の回帰係数は、固定効果モデルと Between 推定の回帰係数の加重平均となっている。

OLS で復習した S_{xx} と S_{xy} を、パネルデータの形式で書き直し (S_{xx}^p などと表す)、Within 項 (S_{xx}^w などと表す) と Between 項 (S_{xx}^b などと表す) に分けると、

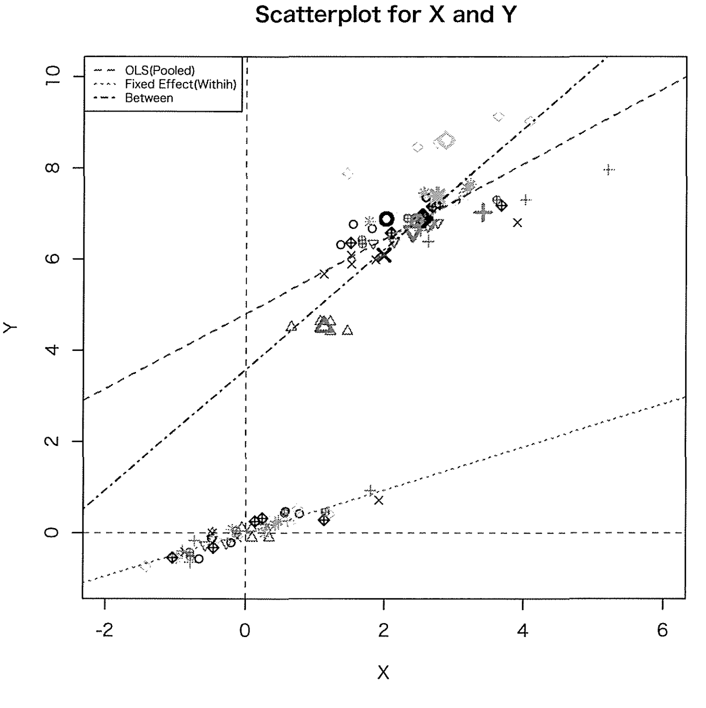
$$
\begin{aligned}
S_{xx}^p &= \sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\bar{x})^2 \\
&= \sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\bar{x}_i)^2 + \sum_{i=1}^{N} T(\bar{x}_i-\bar{x})^2 \\
&= S_{xx}^w + S_{xx}^b
\end{aligned}
$$

ただし、$\bar{x} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}$、$\bar{x}_i = \frac{1}{T}\sum_{t=1}^{T} x_{it}$ である。S_{xy}^p についても同様の式が成り立つ。

このとき、

$$\begin{aligned}\beta^p &= \frac{S^p_{xy}}{S^p_{xx}} && (5.1)\\ &= \frac{S^w_{xy}+S^b_{xy}}{S^p_{xx}} && (5.2)\\ &= \frac{S^w_{xy}}{S^p_{xx}}\frac{S^w_{xx}}{S^w_{xx}}+\frac{S^b_{xy}}{S^p_{xx}}\frac{S^b_{xx}}{S^b_{xx}} && (5.3)\\ &= \frac{S^w_{xx}}{S^p_{xx}}\beta^w+\left(1-\frac{S^w_{xx}}{S^p_{xx}}\right)\beta^b && (5.4)\end{aligned}$$

となるので、Within 項が相対的に大きくなれば、OLS 推定は固定効果モデルに近くなる。



なお、プールド OLS、固定効果モデル、そして Between 推定の回帰直線を示したものが上図である。推定方法によって、回帰直線の傾きには大きな差があることが確認できる。

OLS, 固定効果,Between 推定の関係

```
nN <- max(Ydf$Ind)
nT <- max(Ydf$time)

Sp <- cov(cbind(Ydf$X, Ydf$Y)) * (nN * nT - 1)
Sw <- cov(cbind(Xw,Yw)) * (nN * nT - 1)
Sb <- cov(cbind(Xb,Yb)) * (nN - 1) * nT

beta_p <- Sp[1,2] / Sp[1,1]
beta_w <- Sw[1,2] / Sw[1,1]
beta_b <- Sb[1,2] / Sb[1,1]

print(c(beta_p, beta_w, beta_b))
print(c(Sw[1,1]/Sp[1,1], Sb[1,1]/Sp[1,1]))
print(Sw[1,1]/Sp[1,1] * beta_w + Sb[1,1]/Sp[1,1] * beta_b)
```

出力結果

```
> print(c(beta_p, beta_w, beta_b))
[1] 0.8218060 0.4715958 1.3199901
> print(c(Sw[1,1]/Sp[1,1], Sb[1,1]/Sp[1,1]))
[1] 0.5872082 0.4127918
> print(Sw[1,1]/Sp[1,1] * beta_w + Sb[1,1]/Sp[1,1] * beta_b)
[1] 0.821806
```

上の出力結果は、式 5.4 で示したプールド OLS、固定効果モデル、Between 推定の回帰係数の関係を確認したものである。出力結果の一つ目はそれぞれの回帰直線の傾きである β^p, β^w, β^b を示しており、二つ目は加重平均のウエイトとなる $\frac{S^w_{xx}}{S^p_{xx}}$ と $\frac{S^b_{xx}}{S^p_{xx}}$ が示されている。三つ目はこのウエイトを使って β^w と β^b を加重平均した結果が示されており、β^p と一致していることが確認できる。

ここまで、固定効果モデルについて、その理論的背景を確認するために原始的な方法を用いて説明を行ったが、OLS 推定に lm 関数があったように、パネル推定のためにも plm というライブラリがある。そこで、次にデータセット A の固定効果モデル等について、plm ライブラリの関数を用いて分析する方法を示す。なお、分析に当たって、パッケージ plm 及びそれに依存するパッケージを全てインストールしておくことが必要となる。

plm 関数による推定 (OLS)

```
library(plm)

plm.p <- plm(Y ~ X, data = Ydf, model = "pooling")
summary(plm.p)
```

出力結果

```
Oneway (individual) effect Pooling Model

Call:
plm(formula = Y ~ X, data = Ydf, model = "pooling")

Balanced Panel: n=10, T=5, N=50

Residuals :
   Min. 1st Qu.  Median 3rd Qu.    Max.
-1.5900 -0.2610  0.0475  0.2990  1.8700

Coefficients :
            Estimate Std. Error t-value  Pr(>|t|)
(Intercept)  4.80749    0.28875  16.649 < 2.2e-16 ***
X            0.82181    0.11001   7.470 1.405e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    55.131
Residual Sum of Squares: 25.494
R-Squared      :  0.53758
      Adj. R-Squared :  0.51608
F-statistic: 55.8014 on 1 and 48 DF, p-value: 1.4046e-09
```

plm 関数による推定 (固定効果モデル)

```
plm.w <- plm(Y ~ X, data = Ydf, model = "within")
summary(plm.w)
summary(fixef(plm.w))
```

出力結果

```
> summary(plm.w)
Oneway (individual) effect Within Model

Call:
plm(formula = Y ~ X, data = Ydf, model = "within")

Balanced Panel: n=10, T=5, N=50

Residuals :
    Min.  1st Qu.   Median  3rd Qu.     Max.
-0.26800 -0.09020  0.00114  0.10100  0.22100

Coefficients :
  Estimate Std. Error t-value  Pr(>|t|)
X 0.471596   0.029652  15.904 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    6.6146
Residual Sum of Squares: 0.88363
R-Squared       :  0.86641
      Adj. R-Squared :  0.6758
F-statistic: 252.944 on 1 and 39 DF, p-value: < 2.22e-16

> summary(fixef(plm.w))
   Estimate Std. Error t-value  Pr(>|t|)
1  5.933055   0.090069  65.872 < 2.2e-16 ***
2  3.995151   0.075093  53.202 < 2.2e-16 ***
3  5.422618   0.121472  44.641 < 2.2e-16 ***
4  5.155170   0.089413  57.655 < 2.2e-16 ***
5  7.256438   0.108808  66.690 < 2.2e-16 ***
6  5.479342   0.097967  55.930 < 2.2e-16 ***
7  5.952115   0.110245  53.990 < 2.2e-16 ***
8  6.074880   0.105820  57.408 < 2.2e-16 ***
9  5.698762   0.101150  56.340 < 2.2e-16 ***
10 5.693797   0.099164  57.418 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

plm 関数による推定 (Between 推定)

```
plm.b <- plm(Y ~ X, data = Ydf, model = "between")
summary(plm.b)
```

出力結果

```
Oneway (individual) effect Between Model

Call:
plm(formula = Y ~ X, data = Ydf, model = "between")

Balanced Panel: n=10, T=5, N=50

Residuals :
   Min.  1st Qu.  Median 3rd Qu.    Max.
 -1.060  -0.140  -0.081   0.121   1.220

Coefficients :
            Estimate Std. Error t-value Pr(>|t|)
(Intercept)  3.58604    0.86351  4.1529 0.003196 **
X            1.31999    0.34204  3.8592 0.004814 **
---
Signif. codes:  0  '***'  0.001  '**'  0.01  '*'  0.05  '.'  0.1  ' '  1

Total Sum of Squares:    9.7033
Residual Sum of Squares: 3.3908
R-Squared      :  0.65055
      Adj. R-Squared :  0.52044
F-statistic: 14.8934 on 1 and 8 DF, p-value: 0.0048144
```

5.5 ランダム効果モデル

固定効果モデルでは個体間の差異をパラメータとして推定するため、個体数が多い場合、推定しなければならないパラメータ数が大きくなることがある。しかしながら、分析上、より関心があるのは、目的変数に対する説明変数の影響、データセット A の例でいえば、変数 X の回帰係数であることが往々にしてある。このような場合、個体間の差はある確率分布に従っていると考え、個別にパラメータ推定をしないという考え方もできる。

そこで、回帰式の誤差項 $u_{it} = \mu_i + \nu_{it}$ において、ν_{it} だけでなく、μ_i も確率変数と考え、μ_i は平均 0、分散 σ_μ^2 の、ν_{it} は平均 0、分散 σ_ν^2 の、互いに独立な同一確率分布に従うとする。

すると、誤差項 u_{it} の分散・共分散行列は以下のようになる。

$$cov(u_{it}, u_{js}) = \begin{cases} \sigma_\mu^2 + \sigma_\nu^2 & (i = j \text{ and } t = s) \\ \sigma_\mu^2 & (i = j \text{ and } t \neq s) \\ 0 & (\text{otherwise}) \end{cases}$$

これをランダム効果モデルと呼ぶ。この推定には一般化最小二乗法（GLS：generalized least square）を用いることができるが、データセット A において行った、説明変数が一つのケースについては、ランダム効果モデルで関心のある推定量である回帰直線の傾き β_r は、再び、固定効果モデルの傾き β^w と Between 推定の傾き β^b の加重平均として、以下のように表される。

$$\beta^r = W\beta^w + (1 - W)\beta^b$$

ただし、$W = \frac{S_{xx}^w}{S_{xx}^w + \phi^2 S_{xx}^b}$ であり、この式の中に現れる ϕ^2 は、ν_{it} の分散 σ_ν^2 と、Between 推定における誤差項の分散に時点数 T を乗じた σ_l^2 の比、すなわち、$\phi^2 = \frac{\sigma_\nu^2}{\sigma_l^2}$ である。

なお、この式から、$\phi^2 = 0$ のとき β^r は β^w に一致し、$\phi^2 = 1$ のときは $W = \frac{S_{xx}^w}{S_{xx}^p}$ となることから、5.4 節の結果からこれは OLS 推定量 β^p に一致することもわかる。

さて、具体的な推定を行う。ここまでで既に S_{xx}^w と S_{xx}^b は求めてあるので、あとは、ϕ^2 の推定値さえあれば計算可能である。

σ_ν^2 の推定量 $\hat{\sigma_\nu}^2$ には、固定効果モデルで行った回帰式の誤差分散の推定量を用いることができる。これは、固定効果モデルの回帰式 (原点に移動をした後のもの) の残差 e_{it} の二乗平方和を自由度 $(NT - N - K = NT - N - 1 = 39)$ で割ることによって推定される。すなわち、$\hat{\sigma_\nu}^2 = \frac{\sum_{i=1}^N \sum_{t=1}^T e_{it}^2}{NT-N-K}$ である。

また、σ_μ^2 は、Between 推定の誤差分散の T 倍として定義される σ_l^2 と、σ_ν^2 から、$\sigma_\mu^2 = \frac{\sigma_l^2 - \sigma_\nu^2}{T}$ として求めることができる。

σ_l^2 の推定量 $\hat{\sigma_l}^2$ は、Between 推定の誤差分散の推定量の T 倍であるから、その平方根を 5 で割ったものは既に Between 推定の出力結果に標準偏差の推定量として示されている (Residual standard error: 0.651)。式で表せば、$\hat{\sigma_l}^2 = T\frac{\sum_{i=1}^N e_i^2}{N-K-1}$ である。

以上を用いてランダム効果モデルの推定を行ってみると以下のようになる。

ランダム効果モデルの推定

```
K <- 1
sgm2_nu <- sum(residuals(lm.w)^2)/(nN * nT - nN - K)
sgm2_l  <- sum(residuals(lm.b)^2)/(nN - K - 1) * nT
sgm2_mu <- (sgm2_l - sgm2_nu) / nT
print(c(sgm2_nu, sgm2_mu))

phi2 <- sgm2_nu / sgm2_l
W <- Sw[1,1] / (Sw[1,1] + phi2 * Sb[1,1])
beta_r <- W * beta_w + (1 - W) * beta_b
print(beta_r)
```

出力結果

```
> print(c(sgm2_nu, sgm2_mu))
[1] 0.02265726 0.41931510

> print(beta_r)
[1] 0.4779245
```

出力結果の一つ目は、σ_ν^2 と σ_μ^2 の推定量を示したものである。また、二つ目は、ϕ^2 から加重平均のウエイト W を推定し、これを用いてランダム効果モデルの回帰直線の傾き β_r を推定した結果である。

plm 関数はランダム効果の推定も可能である。以下はそのコードと推定結果を示したものであるが、出力結果の Coefficients の部分を見ると、X の係数の推定値が 0.477925 となっており、先の結果と一致していることがわかる。

plm 関数による推定 (ランダム効果モデル)

```
plm.r <- plm(Y ~ X, data = Ydf, model = "random")
summary(plm.r)
```

出力結果

```
Oneway (individual) effect Random Effect Model
   (Swamy-Arora's transformation)

Call:
plm(formula = Y ~ X, data = Ydf, model = "random")

Balanced Panel: n=10, T=5, N=50

Effects:
                  var std.dev share
idiosyncratic 0.02266 0.15052 0.051
individual    0.41932 0.64755 0.949
theta:  0.8966

Residuals :
    Min.  1st Qu.   Median  3rd Qu.     Max.
-0.43500 -0.08180 -0.00466  0.11100  0.33800

Coefficients :
            Estimate Std. Error t-value  Pr(>|t|)
(Intercept) 5.650616   0.229560  24.615 < 2.2e-16 ***
X           0.477925   0.031073  15.381 < 2.2e-16 ***
---
Signif. codes:  0  '***'  0.001  '**'  0.01  '*'  0.05  '.'  0.1  ' '  1

Total Sum of Squares:    7.1333
Residual Sum of Squares: 1.2032
R-Squared        :  0.83132
      Adj. R-Squared :  0.79807
F-statistic: 236.563 on 1 and 48 DF, p-value: < 2.22e-16
```

5.6 具体的な分析例

ここでは、Balatagi (2005) において用いられている、Grunfeld(1958) の投資関数による分析例を示すとともに、推定法の選択に関する仮説検定について補足する。このモデルは以下のように記述される。

$$I_{it} = \alpha + \beta_1 F_{it} + \beta_2 C_{it} + u_{it}$$

I_{it}: 会社 i の t 年における投資 (real gross investment)
F_{it}: 発行済株式数 (real value of the firm(shares outstanding))
C_{it}: 資本金 (capital stock)
使用するのは plm パッケージに含まれる"Grunfeld"というデータであり、これに、プーリング推定、固定効果推定、Between 推定、ランダム効果推定を適用してみよう。

Grunfeld の投資関数による分析例

```
data("Grunfeld", package = "plm")

grun.p <- plm(inv ~ value + capital, data = Grunfeld, model = "pooling")
grun.w <- plm(inv ~ value + capital, data = Grunfeld, model = "within")
grun.b <- plm(inv ~ value + capital, data = Grunfeld, model = "between")
grun.r <- plm(inv ~ value + capital, data = Grunfeld, model = "random")

result <- rbind(
        summary(grun.p)$coef[2:3,1],
        summary(grun.b)$coef[2:3,1],
        summary(grun.w)$coef[,1],
        summary(grun.r)$coef[2:3,1])
dimnames(result)[[1]] <- c("OLS", "Between", "Within", "Random")
print(result)
```

出力結果

```
            value    capital
OLS     0.1155622 0.23067849
Between 0.1346461 0.03203147
Within  0.1101238 0.31006534
Random  0.1097812 0.30811298
```