f. Column (5) contains the random effects estimates. Which coefficients, apart from the intercepts, show the most difference from the fixed effects estimates? Use the Hausman test statistic (15.36) to test whether there are significant differences between the random effects estimates and the fixed effects estimates in column (3) (Why that one?). Based on the test results, is random effects estimation in this model appropriate?

(f) ① $EXPER^2$ 的係數，差距 $\frac{0.0023}{0.0012} = 1.92$ 倍

② 已知 Hausman Test 統計量 $t_j = \dfrac{\hat{\beta}_{FE} - \hat{\beta}_{RE}}{\sqrt{Var(\hat{\beta}_{FE}) - Var(\hat{\beta}_{RE})}}$

$$t_{EXPER} = \frac{0.0575 - 0.0986}{\sqrt{0.033^2 - 0.0022^2}} = -1.67$$

$$t_{EXPER^2} = \frac{-0.0012 - (-0.0023)}{\sqrt{0.0011^2 - 0.0007^2}} = 1.30$$

$$t_{SOUTH} = \frac{-0.3261 - (-0.2326)}{\sqrt{0.1258^2 - 0.0317^2}} = -0.77$$

$$t_{UNION} = \frac{0.0822 - 0.1027}{\sqrt{0.0312^2 - 0.0245^2}} = -1.06$$

⇒ 僅 EXPER 係數有顯著差異 ($\alpha = 10\%$ 時)

⇒ Random Effect 估計合理

15.17 The data file *liquor* contains observations on annual expenditure on liquor (*LIQUOR*) and annual income (*INCOME*) (both in thousands of dollars) for 40 randomly selected households for three consecutive years.

a. Create the first-differenced observations on *LIQUOR* and *INCOME*. Call these new variables *LIQUORD* and *INCOMED*. Using OLS regress *LIQUORD* on *INCOMED* without a constant term. Construct a 95% interval estimate of the coefficient.

b. Estimate the model $LIQUOR_{it} = \beta_1 + \beta_2 INCOME_{it} + u_i + e_{it}$ using random effects. Construct a 95% interval estimate of the coefficient on *INCOME*. How does it compare to the interval in part (a)?

c. Test for the presence of random effects using the LM statistic in equation (15.35). Use the 5% level of significance.

d. For each individual, compute the time averages for the variable *INCOME*. Call this variable *INCOMEM*. Estimate the model $LIQUOR_{it} = \beta_1 + \beta_2 INCOME_{it} + \gamma INCOMEM_i + c_i + e_{it}$ using the random effects estimator. Test the significance of the coefficient γ at the 5% level. Based on this test, what can we conclude about the correlation between the random effect u_i and *INCOME*? Is it OK to use the random effects estimator for the model in (b)?

(b) 使用隨機效果模型並計算 INCOME 係數的 95% 信賴區間

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

Call:
plm(formula = liquor ~ income, data = pdat, model = "random")

Balanced Panel: n = 40, T = 3, N = 120

Effects:
                 var std.dev share
idiosyncratic 0.9640  0.9819 0.571
individual    0.7251  0.8515 0.429
theta: 0.4459

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-2.263634 -0.697383  0.078697  0.552680  2.225798

Coefficients:
             Estimate Std. Error z-value  Pr(>|z|)
(Intercept) 0.9690324  0.5210052  1.8599 0.0628957 .
income      0.0265755  0.0070126  3.7897 0.0001508 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    126.61
Residual Sum of Squares: 112.88
R-Squared:      0.1085
Adj. R-Squared: 0.10095
Chisq: 14.3618 on 1 DF, p-value: 0.00015083

 2.5 %.income 97.5 %.income
   0.01283111    0.04031983
```

(c) LM 檢定（判斷是否存在隨機效果）

```
        Lagrange Multiplier Test - (Breusch-Pagan)

data:  liquor ~ income
chisq = 20.68, df = 1, p-value = 5.429e-06
alternative hypothesis: significant effects
```

(d) 加入各戶平均 INCOME，估計模型並檢定 γ

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

Call:
plm(formula = liquor ~ income + INCOMEM, data = pdat2, model = "random")

Balanced Panel: n = 40, T = 3, N = 120

Effects:
                 var std.dev share
idiosyncratic 0.9640  0.9819 0.571
individual    0.7251  0.8515 0.429
theta: 0.4459

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max. 
-2.300955 -0.703840  0.054992  0.560255  2.257325 

Coefficients:
             Estimate Std. Error z-value Pr(>|z|)  
(Intercept) 0.9163337  0.5524439  1.6587  0.09718 .
income      0.0207421  0.0209083  0.9921  0.32117  
INCOMEM     0.0065792  0.0222048  0.2963  0.76700  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    126.61
Residual Sum of Squares: 112.79
R-Squared:      0.10917
Adj. R-Squared: 0.093945
Chisq: 14.3386 on 2 DF, p-value: 0.00076987
```

Estimate	Std. Error	z value	p value
0.006579241	0.022204776	0.296298479	0.767002148

15.20 This exercise uses data from the STAR experiment introduced to illustrate fixed and random effects for grouped data. In the STAR experiment, children were randomly assigned within schools into three types of classes: small classes with 13–17 students, regular-sized classes with 22–25 students, and regular-sized classes with a full-time teacher aide to assist the teacher. Student scores on achievement tests were recorded as well as some information about the students, teachers, and schools. Data for the kindergarten classes are contained in the data file *star*.

a. Estimate a regression equation (with no fixed or random effects) where *READSCORE* is related to *SMALL*, *AIDE*, *TCHEXPER*, *BOY*, *WHITE_ASIAN*, and *FREELUNCH*. Discuss the results. Do students perform better in reading when they are in small classes? Does a teacher's aide improve scores? Do the students of more experienced teachers score higher on reading tests? Does the student's sex or race make a difference?

b. Reestimate the model in part (a) with school fixed effects. Compare the results with those in part (a). Have any of your conclusions changed? [*Hint*: specify *SCHID* as the cross-section identifier and *ID* as the "time" identifier.]

c. Test for the significance of the school fixed effects. Under what conditions would we expect the inclusion of significant fixed effects to have little influence on the coefficient estimates of the remaining variables?

d. Reestimate the model in part (a) with school random effects. Compare the results with those from parts (a) and (b). Are there any variables in the equation that might be correlated with the school effects? Use the LM test for the presence of random effects.

e. Using the t-test statistic in equation (15.36) and a 5% significance level, test whether there are any significant differences between the fixed effects and random effects estimates of the coefficients on *SMALL*, *AIDE*, *TCHEXPER*, *WHITE_ASIAN*, and *FREELUNCH*. What are the implications of the test outcomes? What happens if we apply the test to the fixed and random effects estimates of the coefficient on *BOY*?

f. Create school-averages of the variables and carry out the Mundlak test for correlation between them and the unobserved heterogeneity.

(d) School Random Effects 模型，並 LM 檢定是否有隨機效果

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

Call:
plm(formula = readscore ~ small + aide + tchexper + boy + white_asian +
    freelunch, data = star.pdata, effect = "individual", model = "random")

Unbalanced Panel: n = 79, T = 34-137, N = 5766

Effects:
                 var std.dev share
idiosyncratic 751.43   27.41 0.829
individual    155.31   12.46 0.171
theta:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.6470  0.7225  0.7523  0.7541  0.7831  0.8153

Residuals:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-97.483 -17.236  -3.282   0.037  12.803 192.346

Coefficients:
              Estimate Std. Error  z-value  Pr(>|z|)
(Intercept) 436.126774   2.064782 211.2217 < 2.2e-16 ***
small         6.458722   0.912548   7.0777 1.466e-12 ***
aide          0.992146   0.881159   1.1260    0.2602
tchexper      0.302679   0.070292   4.3060 1.662e-05 ***
boy          -5.512081   0.727639  -7.5753 3.583e-14 ***
white_asian   7.350477   1.431376   5.1353 2.818e-07 ***
freelunch   -14.584332   0.874676 -16.6740 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    6158000
Residual Sum of Squares: 4332100
R-Squared:      0.29655
Adj. R-Squared: 0.29582
```

```
        Lagrange Multiplier Test - (Breusch-Pagan)

data:  readscore ~ small + aide + tchexper + boy + white_asian + freelunch
chisq = 6677.4, df = 1, p-value < 2.2e-16
alternative hypothesis: significant effects
```

(e) 檢驗 FE 與 RE 係數的顯著差異（t-test）

```
                variable       t_stat   p_value
small              small  0.024409819 0.9805257
aide                aide  0.003162011 0.9974771
tchexper        tchexper -0.171463571 0.8638593
white_asian white_asian  0.322746040 0.7468876
freelunch      freelunch -0.007447715 0.9940576
>
```

(f) Mundlak Test for Endogeneity of Random Effect

```
Linear hypothesis test:
small_avg = 0
aide_avg = 0
tchexper_avg = 0
boy_avg = 0
white_asian_avg = 0
freelunch_avg = 0

Model 1: restricted model
Model 2: readscore ~ small + aide + tchexper + boy + white_asian + freelunch +
    small_avg + aide_avg + tchexper_avg + boy_avg + white_asian_avg +
    freelunch_avg

Note: Coefficient covariance matrix supplied.

  Res.Df Df      F  Pr(>F)
1   5695
2   5689  6 2.2541 0.03557 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```