15.6

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?

① $EXPER^2$ 係數差了 $\frac{-0.00023}{-0.00012} = 1.92$ 倍

② Hausman test $= \frac{\hat{\beta}_{FE} - \hat{\beta}_{RE}}{\sqrt{Var(\hat{\beta}_{FE}) - Var(\hat{\beta}_{RE})}}$

$t_{EXPER} = \frac{0.0575 - 0.0984}{\sqrt{0.033^2 - 0.022^2}} = -1.67$

$t_{EXPER} = \frac{-0.00012 - (-0.00023)}{\sqrt{0.00011^2 - 0.00007^2}} = 1.296$

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

$t_{EXPER} = \frac{0.0822 - 0.1029}{\sqrt{0.0312^2 - 0.0245^2}} = -1.06$

只有EXPER的係數在10%顯著水準上有顯著差異

⇒ 因此內生性證據相當薄弱

⇒ 故在此例採用隨机效果估計是合適的，

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)?

15.17

```
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
```

$\widehat{LIQUOR}_{it} = 0.969 + 0.0265 INCOME_{it}$

```
> confint(random_mod)
                    2.5 %      97.5 %
(Intercept) -0.05211904 1.99018381
income       0.01283111 0.04031983
```

INCOME(a) 係數的 95% 信賴區間非常寬，且包含 0，表示統計上並不顯著。

而在（b）部分的隨機效果（Random Effects）模型下，標準誤大幅下降至 0.00701，對應的 95% 信賴區間也明顯縮小為 [0.0128, 0.0403]，且不包含 0。

因此，隨機效果模型提供了強而有力的證據，我們以 95% 的信心水準估計，每增加 1000 美元的收入，家庭在酒類上的支出將增加 12.83 至 40.32 美元之間。

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

$\begin{cases} H_0: u_i = 0 \ / \ \sigma_u^2 = 0 \\ H_1: u_i \neq 0 \ / \ \sigma_u^2 > 0 \end{cases}$

在5%顯著水準下臨界值為 ±1.96
把絕 $H_0: \sigma_u^2 = 0$，接受 $H_1: \sigma_u^2 > 0$
表示存在 統計上顯著的未觀察異質性。

```
> plmtest(random_mod,effect = "individual")

        Lagrange Multiplier Test - (Honda)

data:  liquor ~ income
normal = 4.5475, p-value = 2.714e-06
alternative hypothesis: significant effects
```

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)?

```
Call:
plm(formula = liquor ~ income + INCOMEM, data = pdat1, 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
```

INCOMEM 係數 γ 不顯著，故根據 Mundlak 檢定結果沒有證據顯示收入與未觀察的異質性之間存在相關性。
∴隨機效果估計量是合適的選擇。

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.]

```
> summary(FE_model)
Oneway (individual) effect Within Model

Call:
plm(formula = readscore ~ small + aide + tchexper + boy + white_asian +
    freelunch, data = pa_data, effect = "individual", model = "within")

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

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-102.6381  -16.7834   -2.8473   12.7591  198.4169

Coefficients:
             Estimate Std. Error  t-value  Pr(>|t|)
small        6.490231   0.912962   7.1090 1.313e-12 ***
aide         0.996087   0.881693   1.1297    0.2586
tchexper     0.285567   0.070845   4.0309 5.629e-05 ***
boy         -5.455941   0.727589  -7.4987 7.440e-14 ***
white_asian  8.028019   1.535656   5.2277 1.777e-07 ***
freelunch  -14.593572   0.880006 -16.5835 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

控制了學校特定的因素，小班教學的優勢實際上變得更明顯，且仍然具有高度顯著性。
教學經驗的效果在數值上有所下降，表示 OLS 所觀察到的一部分效果可能來自於學校之間的差異，但其估計仍為正值且高度顯著。
男孩的成績仍然比女孩低，這是在各校內部也持續存在的一個高度顯著差距。
控制學校的異質性之後，種族 / 族群的優勢幾乎加倍，這顯示出在同一所學校內部的差異甚至比合併樣本所呈現的估計還要大。

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? 哪些情況下我們預期納入顯著的固定效果對其他係數估計影響小?

```
> pFtest(FE_model, pool_model)

        F test for individual effects

data:  readscore ~ small + aide + tchexper + boy + white_asian + freelunch
F = 16.698, df1 = 78, df2 = 5681, p-value < 2.2e-16
alternative hypothesis: significant effects
```

拒絕虛無假設，拒絕學校之間沒有顯著差異的說法。
如果學校指標變數與已包含的解釋變數之間沒有相關性，那麼將這些學校指標變數納入或排除回歸模型，對回歸估計結果應該影響不大

15.20

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.

```
Call:
plm(formula = readscore ~ small + aide + tchexper + boy + white_asian +
    freelunch, data = pa_data, 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
```

整體而言，隨機效果估計結果看起來與 OLS 估計結果大致相同，但與與固定效果估計結果更為相似

```
> plmtest(
+   readscore ~ small + aide + tchexper + boy + white_asian + freelunch,
+   data   = pa_data,
+   type   = "bp",
+   effect = "individual"
+ )

        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
```

Breusch–Pagan 拉格朗日乘數檢定顯示出顯著的統計結果，因此我們拒絕「沒有隨機效果」的虛無假設。換句話說，閱讀成績中存在高度顯著的學校層級異質性，這證實相較於單純的合併 OLS，應使用隨機效果模型。

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*?

```
        Hausman Test

data:  readscore ~ small + aide + tchexper + boy + white_asian + freelunch
chisq = 13.809, df = 6, p-value = 0.03184
alternative hypothesis: one model is inconsistent
```

```
> print(t_values)
      small         aide     tchexper white_asian   freelunch
1.14600764  0.12843803 -1.93771666  1.21807432 -0.09555102
```

只有教師經驗係數的差異具有顯著性。
而這項檢定可能出現的問題可以從 BOY 變數的檢定中看出，固定效果的標準誤小於隨機效果的標準誤，因此分母會包含負數的平方根，這種情況產生的原因是標準誤為估計變異數的平方根，而像所有估計量一樣，變異數估計值會呈現抽樣變動，有時大於，有時小於真實變異數。
Hausman 對比統計量（Hausman contrast statistic），用來聯合檢定所有係數，其統計量大於這六個係數的比較臨界值為12.592
因此我們拒絕虛無假設－即「未觀察異質性與解釋變數之間沒有相關性」。根據這項檢定結果，不建議採用隨機效果估計量。

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

```
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
```

BOYM的係數在5%顯著水準下是顯著的。對6個擴增平均數的聯合顯著性檢定
$Pr(>F) = 0.03557$，因此拒絕 H_0，
結論將與e題部分同。

```
Coefficients:
              Estimate Std. Error  z-value  Pr(>|z|)
(Intercept)  459.462989  20.529888  22.3802 < 2.2e-16 ***
small          6.637460   0.922068   7.1985 6.090e-13 ***
aide           1.157620   0.889542   1.3014    0.1931
tchexper       0.289286   0.071754   4.0316 5.539e-05 ***
boy           -5.386109   0.735063  -7.3274 2.346e-13 ***
white_asian    8.081423   1.550155   5.2133 1.855e-07 ***
freelunch    -14.699025   0.892109 -16.4767 < 2.2e-16 ***
small_m      -18.410060  22.273923  -0.8265    0.4085
aide_m        16.811358  20.793685   0.8085    0.4188
tchexper_m     1.006007   0.625690   1.6078    0.1079
boy_m        -53.353521  25.221654  -2.1154    0.0344 *
white_asian_m -6.648191   6.320012  -1.0519    0.2928
freelunch_m   -3.318853   8.779553  -0.3780    0.7054
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    6007200
Residual Sum of Squares: 4281300
```