15.6 Using the NLS panel data on $N = 716$ young women, we consider only years 1987 and 1988. We are interested in the relationship between ln(*WAGE*) and experience, its square, and indicator variables for living in the south and union membership. Some estimation results are in Table 15.10.

TABLE 15.10 **Estimation Results for Exercise 15.6**

	(1) OLS 1987	(2) OLS 1988	(3) FE	(4) FE Robust	(5) RE
C	0.9348	0.8993	1.5468	1.5468	1.1497
	(0.2010)	(0.2407)	(0.2522)	(0.2688)	(0.1597)
EXPER	0.1270	0.1265	0.0575	0.0575	0.0986
	(0.0295)	(0.0323)	(0.0330)	(0.0328)	(0.0220)
$EXPER^2$	−0.0033	−0.0031	−0.0012	−0.0012	−0.0023
	(0.0011)	(0.0011)	(0.0011)	(0.0011)	(0.0007)
SOUTH	−0.2128	−0.2384	−0.3261	−0.3261	−0.2326
	(0.0338)	(0.0344)	(0.1258)	(0.2495)	(0.0317)
UNION	0.1445	0.1102	0.0822	0.0822	0.1027
	(0.0382)	(0.0387)	(0.0312)	(0.0367)	(0.0245)
N	716	716	1432	1432	1432

(standard errors in parentheses)

a. The OLS estimates of the ln(*WAGE*) model for each of the years 1987 and 1988 are reported in columns (1) and (2). How do the results compare? For these individual year estimations, what are you assuming about the regression parameter values across individuals (heterogeneity)?

OLS 1987 與 1988 係數有些微差異，可能有 heterogeneity

b. The ln(*WAGE*) equation specified as a panel data regression model is

$$\ln(WAGE_{it}) = \beta_1 + \beta_2 EXPER_{it} + \beta_3 EXPER_{it}^2 + \beta_4 SOUTH_{it} + \beta_5 UNION_{it} + (u_i + e_{it}) \quad \text{(XR15.6)}$$

Explain any differences in assumptions between this model and the models in part (a).

假設每個隨時間變動的殘差為 u

c. Column (3) contains the estimated fixed effects model specified in part (b). Compare these estimates with the OLS estimates. Which coefficients, apart from the intercepts, show the most difference?

改變最多的係數：

SOUTH（從 -0.2128/-0.2384 到 -0.0361）變化最多，顯示 FE 模型認為 south 很大部分是個體特徵引起的。

UNION（從 0.1445/0.1102 到 0.0127）次之，顯示 union 在 FE 中被大幅削弱。

d. The *F*-statistic for the null hypothesis that there are no individual differences, equation (15.20), is 11.68. What are the degrees of freedom of the *F*-distribution if the null hypothesis (15.19) is true? What is the 1% level of significance critical value for the test? What do you conclude about the null hypothesis.

OLS (1) 和 (2)：樣本數 716，每年 5 個參數（常數項 + 4 個變數），自由度 = 716 - 5 = 711。

FE (3)：考慮個體效應，每個個體一個虛擬變數，自由度減少為 716 - (5 + 715) = -4（實際上 plm 會調整）。

在 1% 顯著性水平下，假設自由度約為 (715, 711)，F 臨界值約為 2.64

拒絕虛無假設，個體無差異

e. Column (4) contains the fixed effects estimates with cluster-robust standard errors. In the context of this sample, explain the different assumptions you are making when you estimate with and without cluster-robust standard errors. Compare the standard errors with those in column (3). Which ones are substantially different? Are the robust ones larger or smaller?

標準誤皆變大

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?

```
Residuals:
    Min      1Q  Median      3Q     Max
-110.05  -20.27   -4.02   14.45  189.12

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 434.52072    1.28572 337.958  < 2e-16 ***
small         5.81416    0.99437   5.847 5.28e-09 ***
aide          0.79682    0.95784   0.832    0.406
tchexper      0.51286    0.06986   7.341 2.41e-13 ***
white_asian   3.74427    0.95823   3.907 9.43e-05 ***
freelunch   -14.75206    0.89478 -16.487  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 30.34 on 5760 degrees of freedom
  (因為不存在，20 個觀察量被刪除了)
Multiple R-squared:  0.08748,   Adjusted R-squared:  0.08668
F-statistic: 110.4 on 5 and 5760 DF,  p-value: < 2.2e-16
```

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.]

```
> # 比較兩個模型的係數
> summary(model_a)$coefficients
               Estimate Std. Error     t value     Pr(>|t|)
(Intercept) 434.5207225 1.28572307 337.9582536 0.000000e+00
small         5.8141611 0.99436939   5.8470837 5.277351e-09
aide          0.7968162 0.95784135   0.8318875 4.055069e-01
tchexper      0.5128553 0.06985898   7.3412937 2.408842e-13
white_asian   3.7442696 0.95823207   3.9074768 9.433391e-05
freelunch   -14.7520583 0.89477693 -16.4868560 1.045669e-59
> summary(model_b)$coefficients
                     Estimate Std. Error     t value     Pr(>|t|)
(Intercept)       407.4698192  4.1186586  98.9326520 0.000000e+00
small               6.4814512  0.9173873   7.0651197 1.797342e-12
aide                0.9869604  0.8859669   1.1139924 2.653296e-01
tchexper            0.3000867  0.0711614   4.2169874 2.514345e-05
white_asian         7.9348573  1.5430510   5.1423169 2.804640e-07
```

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?

```
> anova(model_a, model_b)
Analysis of Variance Table

Model 1: readscore ~ small + aide + tchexper + white_asian + freelunch
Model 2: readscore ~ small + aide + tchexper + white_asian + freelunch +
    factor(schid)
  Res.Df     RSS Df Sum of Sq      F    Pr(>F)
1   5760 5302072
2   5682 4311147 78    990925 16.744 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

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.

```
Call:
lm(formula = LIQUORD ~ INCOMED, data = liquir)

Residuals:
    Min      1Q  Median      3Q     Max
-3.5012 -0.8399  0.0298  1.0077  3.5049

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.40287    0.40600  -0.992    0.324
INCOMED      0.09815    0.07487   1.311    0.194

Residual standard error: 1.417 on 78 degrees of freedom
Multiple R-squared:  0.02156,   Adjusted R-squared:  0.009012
F-statistic: 1.718 on 1 and 78 DF,  p-value: 0.1937
```

```
               2.5 %    97.5 %
INCOMED -0.05090933 0.2472087
```