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估計值在1987年與1988年的迴歸分析中几乎沒差異。

OLS模型假設所有個体的母体參数(包括截距項)都是完全相同的,其背後的基本假設是不同個体之間不存在異質性。

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

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

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

在XR 15.16 中,使用的是panel data,每個可觀察変数加上代表个体與時間期別的下標 it。

此外,未被觀察到的誤差成分
- e_{it} 隨个体與時間变動的隨机誤差 → idiosyncratic error component
- u_i 每个人固定但不可觀察的特性(个体異質性),時間不变

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?

所有 FE model 的係數都與 OLS model 的有明顯差異。

FE model 建構 95% 區間:

EXPER (-0.0085, 0.1235)

EXPER^2 (-0.0034, 0.001)

SOUTH (-0.5777, -0.0745)

UNION (0.0198, 0.1446)

⇒ EXPER 的 OLS 估計值不落在 FE model 的區間估計，故有顯著差異。

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.

If the null hypothesis is true, then $\beta_{11} = \beta_{12} = \beta_{13} = \cdots = \beta_{1N} = \beta_1$, where β_1 denotes the common value, and there are no individual differences and no unobserved heterogeneity. The null hypothesis is $J = N - 1$ separate equalities, $\beta_{11} = \beta_{12}$, $\beta_{12} = \beta_{13}$, and so on. If the null hypothesis is true, then the "restricted model" is

$$y_{it} = \beta_1 + \beta_2 x_{2it} + \cdots + \beta_K x_{Kit} + e_{it}$$

Under the standard OLS assumptions, the F-test statistic is

K-1

$$F = \frac{(SSE_R - SSE_U)/(N-1)}{SSE_U/(NT - N - K_S)} \quad (15.20)$$

→ 716 - 1 = 715 方程式參數量 -1

→ 716×2 - 716 - 4 = 712

where SSE_U is the sum of squared residuals from the fixed effects model, and SSE_R is the sum of squared errors from the OLS regression that pools all the data, $y_{it} = \beta_1 + \beta_2 x_{2it} + \cdots + \beta_K x_{Kit} + e_{it}$. If the null hypothesis is true, the test statistic has the F-distribution with $J = N - 1$ numerator degrees of freedom and $NT - N - K_S$ denominator degrees of freedom. Using the α level of significance, we reject the null hypothesis if the test statistic value is greater than, or equal to, the $1 - \alpha$ percentile of the F-distribution, $F \geq F_{(1-\alpha,\, N-1,\, NT-N-K_S)}$. The test can be made "robust" to heteroskedasticity and serial correlation, topics that we consider in Section 15.3.

$F = 11.68 > 1.2 = F_{0.01}(715, 712)$

⇒ 拒絕 H_0 (不存在个体差異)，支持使用固定效果。

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?

隨机誤差變為 $\tilde{e}_{it} = e_{it} - \bar{e}_i$ → 使原本的个体特有誤差(idiosyncratic errors)彼此不相関，轉換後的誤差也會呈序列相関(serial correlation)

使用叢集穩健標準誤可讓我們在估計中考慮到誤差項 e_{it} 在不同个体與時間間的異質變異，以及跨時間的時間序列相関性。

⇒ 經驗變數(EXPER、EXPER2)差不多

SOUTH，穩健標準誤幾乎是傳統2倍

UNION，穩健標準誤是傳統標準誤的1.8倍

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 - 1, data = liquor_fd)

Residuals:
    Min      1Q  Median      3Q     Max
-3.6852 -0.9196 -0.0323  0.9027  3.3620

Coefficients:
        Estimate Std. Error t value Pr(>|t|)
INCOMED  0.02975    0.02922   1.018    0.312

Residual standard error: 1.417 on 79 degrees of freedom
  (因為不存在，40 個觀察量被刪除了)
Multiple R-squared:  0.01295,   Adjusted R-squared:  0.0004544
F-statistic: 1.036 on 1 and 79 DF,  p-value: 0.3118

> confint(fd_mod, level = 0.95)
              2.5 %     97.5 %
INCOMED -0.02841457 0.08790818
```

$$\widehat{LIQUORD}_{it} = \underset{(0.02922)}{0.02975}\, INCOMED_{it}$$
(SE)

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?

```
lm(formula = readscore ~ small + aide + tchexper + boy + white_asian +
    freelunch, data = star)

Residuals:
     Min       1Q   Median       3Q      Max
-107.220  -20.214   -3.935   14.339  185.956

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 437.76425    1.34622 325.180  < 2e-16 ***
small         5.82282    0.98933   5.886 4.19e-09 ***
aide          0.81784    0.95299   0.858    0.391
tchexper      0.49247    0.06956   7.080 1.61e-12 ***
boy          -6.15642    0.79613  -7.733 1.23e-14 ***
white_asian   3.90581    0.95361   4.096 4.26e-05 ***
freelunch   -14.77134    0.89025 -16.592  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 30.19 on 5759 degrees of freedom
  (因為不存在，20 個觀察量被刪除了)
Multiple R-squared:  0.09685,   Adjusted R-squared:  0.09591
F-statistic: 102.9 on 6 and 5759 DF,  p-value: < 2.2e-16
```

(i) 若學生在小班（small class）中學習，相較於在一般班級（regular sized class），其平均閱讀成績將高出 5.8 分。此係數在 1% 的顯著水準下與零有顯著差異。
(ii) 與一般班級相比，配置一位教師助理（teacher aide）對平均閱讀成績沒有顯著改善效果。
(iii) 每增加一年教學經驗，預計平均閱讀成績將提升 0.49 分，且該係數與零有顯著差異。
(iv) 男孩的閱讀成績預估比女孩低 6 分；白人或亞裔學生的平均閱讀成績預估比黑人學生高 3.9 分。
(v) 領取免費午餐的學生，其平均閱讀成績預估比未領取者低 14.8 分。
除了教師助理（AIDE）以外，所有變數的係數在統計上皆具有個別顯著性。

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?

```
> 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，即不同學校之間不存在顯著差異。
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