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)?

1987 年 OLS 結果➡$\widehat{ln\,WAGE} = 0.9348 + 0.1270\,\text{EXPER} - 0.0033\,\text{EXPER}^2 + 0.0536\,\text{SOUTH} + 0.1445\,\text{UNION}$

1988 年 OLS 結果➡$\widehat{ln\,WAGE} = 0.8993 + 0.1265\,\text{EXPER} - 0.0031\,\text{EXPER}^2 + 0.0542\,\text{SOUTH} + 0.1102\,\text{UNION}$

EXPER 和 EXPER^2 係數在這兩年幾乎一致，而且 SOUTH 與 UNION 的係數也很接近，但截距略有不同

針對 OLS 估計，假設「各個體間 u_i 效果不變或可忽略」，不考慮面板結構下的個體異質。

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).

考慮去除每位女性固定的 u_i ，只估計「變動部分」，可以從欄 3 看出 EXPER 係數大幅從 ~0.127 降到 0.0575 ， 代表部分工齡影響其實來自於個人不變特質，而 UNION 也從 0.14 降到 0.0822，影響縮小

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?

OLS 與 FE 比較下，最大差異就是 EXPER 係數降得最厲害，說明 OLS 高估了 EXPER 的邊際效果

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.

$H_0: u_i = 0$　$H_1: u_i \neq 0$ ➡而 F 統計量 = 11.68 分子 df= 716−1=715，分母 df= 1432−5=1427，而$F_{0.99;715,1427} \sim 1.14$

而 $11.68 \gg 1.14 \Longrightarrow$拒絕 H0$\Longrightarrow$固定效果顯著必要

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?

從欄 3 跟欄 4 來看 SOUTH 的 SE 從 0.1258 上升到 0.2495，而 UNION 的 SE 從 0.0312 上升到 0.0367

因此考慮同一個體內 e_{it} 相關後，SE 變大讓統計檢定更保守

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?

從欄 3 跟欄 5 的估計值中差異最大的是 SOUTH（+0.0935），其次為 EXPER（+0.0411），而 Hausman 檢定下 $\chi 2$ 統計量=15.36，df=4，而臨界值 $\chi 2_{0.95,42} \approx 9.49$，而 15.36>9.49，拒絕虛無假設 H_0：RE 與 FE 估計顯著不同代表隨機效果模型所需的「個體效果 u_i 與解釋變數不相關」假設不成立，RE 估計不一致、不適用；應採用 固定效果模型

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

```
                2.5 %     97.5 %
incomed -0.02841457 0.08790818
```

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?

```
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 ***
```

小班教學顯著提高學生的閱讀成績，教師助理的效果並不顯著，教學年資對學生成績有正向貢獻，男生顯著低於女生，白人與亞裔學生則顯著高於其他族群，低收入家庭的學生表現顯著較差

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.]

```
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
```

控制學校之後，小班效果變得更強，教師經驗效果縮小，種族影響變大，性別與貧窮效應維持穩定，顯示部分 OLS 結果受到學校差異干擾

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?

```
        F test for individual effects

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

H_0：所有學校的固定效果皆為 0、H_1：至少有一間學校的固定效果不為 0

p 值遠小於 0.05：表示檢定顯著，拒絕 H_0，因此，學校之間的差異顯著，必須納入「學校固定效果（FE）」來避免模型偏誤