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)

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?

15.6

f. $H_0: \beta_{fe,k} = \beta_{re,k}$ no endogeniety

$H_a: \beta_{fe,k} \neq \beta_{re,k}$ endogeniety

Exper: $\dfrac{0.0575 - 0.0986}{\sqrt{0.033^2 - 0.022^2}} = -1.67$

$Exper^2$: $\dfrac{-0.0012 - (-0.0023)}{\sqrt{0.0011^2 - 0.0007^2}} = 1.3$

South: $\dfrac{-0.3261 - (-0.2326)}{\sqrt{0.1258^2 - 0.0317^2}} = -0.77$

UNION: $\dfrac{0.0822 - 0.1027}{\sqrt{(0.0312)^2 - 0.0245^2}} = -1.06$

$t_{0.05} = 1.65$

exper 可能是 endogenous v.r.

但大部分都是 exogenous v.r.

使用 random effect 是可以的

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.

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. interval is narrower than fixed effect in part a.

a suggest beta2 is not significant, while b suggest it is significant.

```
> cat("a. 95% re_CI：", conf_interval, "with se", fe_sum$coefficients[2])
a. 95% re_CI： -0.02841457 0.08790818 with se 0.02922021
> cat("b. 95% re_CI：", round(re_ci, 4),"with se", re_se)
b. 95% re_CI： 0.0121 0.0429 with se 0.007867104
```

c. H0: there's no indivisual difference

Ha: there's indivisual difference

p-value< 0.05. It’s statistical significant that there’s indivisual difference.

We should consider random effect.

```
        Lagrange Multiplier Test - (Honda)

data:  liquor ~ income
normal = 2.4786, p-value = 0.006594
alternative hypothesis: significant effects
```

d. H0: there’s no endogenous

Ha: there’s endogenous

p-value=0.767>0,05, there’s no statistical significant endogenous.

We should consider random effect.

```
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
```

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*.

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. H0: no individual difference

Ha: individual difference

p-value < 0.05, there's statistical significant individual difference exist.

We should consider random effects.

```
        Lagrange Multiplier Test - (Honda)

data:  readscore ~ small + aide + tchexper + boy + white_asian + freelunch
normal = 81.715, p-value < 2.2e-16
alternative hypothesis: significant effects
```

e. H0: no endogenous

Ha: endogenous

p-value 都大於 0.05，不能顯著說 endogenous

```
small       :  t =  1.15,  p = 0.252
aide        :  t =  0.13,  p = 0.898
tchexper    :  t = -1.94,  p = 0.053
white_asian :  t =  1.22,  p = 0.223
freelunch   :  t = -0.10,  p = 0.924
boy         :  t =   NaN,  p = NaN
```

Boy 的 fixed effect se^2 < random effect se^2,無法計算。

f. 平均過後的 dependent variables 大多沒有顯著影響，不拒絕 H0，代表無法顯著說明存在 endogenous 問題。

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	