Le Thi Phuong Thao – 413707007

HW0519: 15.17, 15.20

15.17

a) Answer

```
Call:
lm(formula = LIQUORD ~ INCOMED - 1, data = liquor_diff)

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
Multiple R-squared:  0.01295,	Adjusted R-squared:  0.0004544 
F-statistic: 1.036 on 1 and 79 DF,  p-value: 0.3118

> coef_income <- coef(summary(moda))["INCOMED", "Estimate"]
> std_income <- coef(summary(moda))["INCOMED", "Std. Error"]
> conf_income <- confint(moda, level = 0.95)
> conf_income
              2.5 %     97.5 %
INCOMED -0.02841457 0.08790818
```

The 95% interval estimate of the coefficient of INCOMED is [-0.0284146, 0.0879082]. The interval covers zero; we have no evidence against the hypothesis that income does not affect liquor expenditures

b) Answer

```
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
> conf_income_re <- confint(mod_re, level = 0.95)
> conf_income_re
                  2.5 %     97.5 %
(Intercept) -0.05211904 1.99018381
income       0.01283111 0.04031983
```

The 95% interval estimate for the coefficient of INCOME is [0.01283, 0.04032]. We estimate with 95% confidence that for each additional \$1000 income the household will spend between \$12.83 and \$40.32 more on liquor. The random effects coefficient estimate is slightly smaller than the difference estimator coefficient, but the standard error of the random effects estimator is about 25% of the standard error of the difference estimator’s standard error, yielding a statistical significance.

c) Answer

R software report LM^2 = 20.68 > Chi square (0.95,1) = 3.841

$$LM = \sqrt{\frac{NT}{2(T-1)}}\left\{\frac{\sum_{i=1}^{N}\left(\sum_{t=1}^{T}\hat{e}_{it}\right)^2}{\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{e}_{it}^2} - 1\right\} \qquad (15.35)$$

```
> mod_fe <- plm(liquor ~ income, data = liquor5, model = "within")
> lm_test <- plmtest(mod_fe, effect = "individual", type = "bp")
> print(lm_test)

        Lagrange Multiplier Test - (Breusch-Pagan)

data:  liquor ~ income
chisq = 20.68, df = 1, p-value = 5.429e-06
alternative hypothesis: significant effects
```

We reject the null hypothesis that sigma^2$_u$ = 0 and accept the alternative that sigma^2$_u$ > 0, indicating that there is statistically significant unobserved heterogeneity.

d) Answer

```
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
```

P-value of incomem = 0.767 with t-value = 0.3. There is no evidence for correlation between income and the unobserved heterogeneity based on this Mundlak test. Based on these results the random effects estimator is preferred.

15.20

a) Answer - OLS

```
Coefficients:
               Estimate Std. Error  t-value  Pr(>|t|)    
(Intercept) 437.764253   1.346221 325.1800 < 2.2e-16 ***
small         5.822816   0.989333   5.8856 4.190e-09 ***
aide          0.817837   0.952993   0.8582    0.3908    
tchexper      0.492469   0.069555   7.0803 1.611e-12 ***
boy          -6.156421   0.796128  -7.7330 1.232e-14 ***
white_asian   3.905809   0.953607   4.0958 4.264e-05 ***
freelunch   -14.771337   0.890248 -16.5924 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    5810300
Residual Sum of Squares: 5247600
R-Squared:      0.096853
Adj. R-Squared: 0.095912
F-statistic: 102.932 on 6 and 5759 DF, p-value: < 2.22e-16
```

Do students perform better in reading when they are in small classes? -> Yes
Does a teacher's aide improve scores? -> No, coefficient is insignificant
Do the students of more experienced teachers score higher on reading tests? -> Yes
Does the student's sex or race make a difference -> Yes, male is lower score

b) Answer - FE

```
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

Total Sum of Squares:    4628000
Residual Sum of Squares: 4268900
R-Squared:      0.077592
Adj. R-Squared: 0.063954
F-statistic: 79.6471 on 6 and 5681 DF, p-value: < 2.22e-16
```

The effect of being in a small class is estimated to increase average reading score by 6.49 points, which is slightly larger than the OLS estimate.
The estimated effect of teaching experience on average reading score falls to 0.29 points per additional year of experience.
The estimated difference between boys and girls average reading scores is slightly smaller than the OLS estimates.
The estimated difference in average reading scores between white or Asian students and black students roughly doubles to 8 points

c) Answer

```
> pFtest(mod_fe1520, mod_ols1520)

        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
```

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

d) Answer - RE
Test for the significance of the school random effects.
Under what conditions would we expect the inclusion of significant random effects to have little influence on the coefficient estimates of the remaining variables

```
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
Residual Sum of Squares: 4332100
R-Squared:      0.29655
Adj. R-Squared: 0.29582
Chisq: 493.205 on 6 DF, p-value: < 2.22e-16
> plmtest(readscore ~ small + aide + tchexper + boy + white_asian + freelunch,
+         data = pdata, type = "bp")

	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
```

e) Answer

```
> hausman_test <- phtest(mod_fe1520, mod_re1520)
> hausman_test

	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
```

P-vale < 5% we reject null hypothesis, we should use fixed effect

```
> boy_fe <- coef(summary(mod_fe1520))["boy", "Estimate"]
> seboy_fe <- coef(summary(mod_fe1520))["boy", "Std. Error"]
> boy_re <- coef(summary(mod_re1520))["boy", "Estimate"]
> seboy_re <- coef(summary(mod_re1520))["boy", "Std. Error"]
>
> t_val <- (boy_fe - boy_re) / (sqrt(seboy_fe^2 + seboy_re^2))
> t_val
[1] 0.05455779
```

For the variable BOY, the same approach was applied, but its t-value does not reject the null hypothesis, indicating no difference between the effects, and the random effects model is valid

f) Answer

Test that average variables are jointly different from 0 at 5% level. This suggests that the explanatory variables (x1, x2) are correlated with the unobserved heterogeneity.

In this case, the fixed effects model is more appropriate

```
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
---
```

```
Linear hypothesis test:
small_m = 0
aide_m = 0
tchexper_m = 0
boy_m = 0
white_asian_m = 0
freelunch_m = 0

Model 1: restricted model
Model 2: readscore ~ small + aide + tchexper + boy + white_asian + freelunch +
    small_m + aide_m + tchexper_m + boy_m + white_asian_m + freelunch_m

  Res.Df Df  Chisq Pr(>Chisq)
1   5674
2   5668  6 12.716    0.04778 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```