a. Demand : $Q_i = \alpha_1 + \alpha_2 P_i + \alpha_3 PS_i + \alpha_4 DI_i + e_{di}$ (11.1)

Supply : $Q_i = \beta_1 + \beta_2 P_i + \beta_3 PF_i + e_{si}$ (11.2)

DI : Disposable Income

PF : Price of Factors of Production

Demand Equation : $\alpha_2 P_i = Q_i - \alpha_1 - \alpha_3 PS_i - \alpha_4 DI_i - e_{di}$

$$P_i = \frac{Q_i}{\alpha_2} - \frac{\alpha_1}{\alpha_2} - \frac{\alpha_3}{\alpha_2} PS_i - \frac{\alpha_4}{\alpha_2} DI_i - \frac{e_{di}}{\alpha_2}$$

$$= \delta_1 + \delta_2 Q_i + \delta_3 PS_i + \delta_4 DI_i - u^d$$

$\delta_2 < 0$

$\delta_3 > 0$

$\delta_4 > 0$

Supply Equation : $\beta_2 P_i = Q_i - \beta_1 - \beta_3 PF_i + e_{si}$

$$P_i = \frac{Q_i}{\beta_2} - \frac{\beta_1}{\beta_2} - \frac{\beta_3}{\beta_2} PF_i + \frac{e_{si}}{\beta_2}$$

$$P_i = \pi_1 + \pi_2 Q_i + \pi_3 PF_i + u^s$$

$\pi_2 > 0$

$\pi_3 > 0$

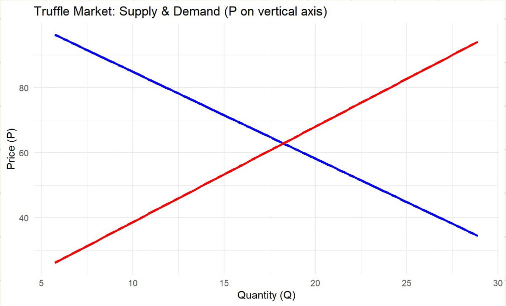
b.

```
2SLS estimates for 'demand' (equation 1)
Model Formula: p ~ q + ps + di
Instruments: ~ps + di + pf

              Estimate Std. Error  t value   Pr(>|t|)
(Intercept) -11.42841   13.59161 -0.84084  0.4081026
q            -2.67052    1.17495 -2.27287  0.0315350 *
ps            3.46108    1.11557  3.10252  0.0045822 **
di           13.38992    2.74671  4.87490 4.6752e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 13.165551 on 26 degrees of freedom
Number of observations: 30 Degrees of Freedom: 26
SSR: 4506.625289 MSE: 173.331742 Root MSE: 13.165551
Multiple R-Squared: 0.556717 Adjusted R-Squared: 0.505569


2SLS estimates for 'supply' (equation 2)
Model Formula: p ~ q + pf
Instruments: ~ps + di + pf

               Estimate Std. Error  t value   Pr(>|t|)
(Intercept) -58.798223   5.859161 -10.0353 1.3165e-10 ***
q             2.936711   0.215772  13.6103 1.3212e-13 ***
pf            2.958486   0.155964  18.9690 < 2.22e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.399078 on 27 degrees of freedom
Number of observations: 30 Degrees of Freedom: 27
SSR: 522.500877 MSE: 19.351884 Root MSE: 4.399078
Multiple R-Squared: 0.948605 Adjusted R-Squared: 0.944798
```

correct, yes

c. −1.272464

d.



e.

```
> num <- (b0 + b2*23) - (a0 + a2*22 + a3*3.5)
> den <- lambda1 - theta1
> Q_eq <- num / den
> P_eq <- a0 + lambda1*Q_eq + a2*22 + a3*3.5
> c(Q_eq = Q_eq, P_eq = P_eq)
Q_eq.(Intercept) P_eq.(Intercept)
        18.25021         62.84257
>
> rf_Q <- lm(q ~ ps + di + pf, data = truffles)
> rf_P <- lm(p ~ ps + di + pf, data = truffles)
> newX <- data.frame(ps = 22, di = 3.5, pf = 23)
> Q_hat <- predict(rf_Q, newdata = newX)   # π^1
> P_hat <- predict(rf_P, newdata = newX)   # π^1
> c(Q_hat = Q_hat, P_hat = P_hat)
 Q_hat.1  P_hat.1
18.26040 62.81537
```

f.

model	term	estimate	std.error	statistic	p.value
<chr>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>
Demand 2SLS	(Intercept)	-11.4	13.6	-0.841	4.08e- 1
Demand OLS	(Intercept)	-13.6	9.09	-1.50	1.46e- 1
Supply 2SLS	(Intercept)	-58.8	5.86	-10.0	1.32e-10
Supply OLS	(Intercept)	-52.9	5.02	-10.5	4.68e-11
Demand 2SLS	di	13.4	2.75	4.87	4.68e- 5
Demand OLS	di	12.4	1.83	6.77	3.48e- 7
Supply 2SLS	pf	2.96	0.156	19.0	3.88e-17
Supply OLS	pf	2.92	0.148	19.7	1.47e-17
Demand 2SLS	ps	3.46	1.12	3.10	4.58e- 3
Demand OLS	ps	1.36	0.594	2.29	3.03e- 2
Demand 2SLS	q	-2.67	1.17	-2.27	3.15e- 2
Demand OLS	q	0.151	0.499	0.303	7.64e- 1
Supply 2SLS	q	2.94	0.216	13.6	1.32e-13
Supply OLS	q	2.66	0.171	15.5	5.42e-15

```
lm(formula = p ~ q + ps + di, data = truffles)

Residuals:
     Min       1Q   Median       3Q      Max 
-25.0753  -2.7742  -0.4097   4.7079  17.4979 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -13.6195     9.0872  -1.499   0.1460    
q             0.1512     0.4988   0.303   0.7642    
ps            1.3607     0.5940   2.291   0.0303 *  
di           12.3582     1.8254   6.770 3.48e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.814 on 26 degrees of freedom
Multiple R-squared:  0.8013,	Adjusted R-squared:  0.7784 
F-statistic: 34.95 on 3 and 26 DF,  p-value: 2.842e-09

> summary(sup_ols)

Call:
lm(formula = p ~ q + pf, data = truffles)

Residuals:
    Min      1Q  Median      3Q     Max 
-8.4721 -3.3287  0.1861  2.0785 10.7513 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -52.8763     5.0238  -10.53 4.68e-11 ***
q             2.6613     0.1712   15.54 5.42e-15 ***
pf            2.9217     0.1482   19.71  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.202 on 27 degrees of freedom
Multiple R-squared:  0.9531,	Adjusted R-squared:  0.9496 
F-statistic: 274.4 on 2 and 27 DF,  p-value: < 2.2e-16
```

11.30 Example 11.3 introduces Klein's Model I. Use the data file *klein* to answer the following questions.

a. Estimate the investment function in equation (11.18) by OLS. Comment on the signs and significance of the coefficients.

b. Estimate the reduced-form equation for profits, P_t, using all eight exogenous and predetermined variables as explanatory variables. Test the joint significance of all the variables except lagged profits, P_{t-1}, and lagged capital stock, K_{t-1}. Save the residuals, $\hat{v}_t$ and compute the fitted values, $\hat{P}_t$.

c. The Hausman test for the presence of endogenous explanatory variables is discussed in Section 10.4.1. It is implemented by adding the reduced-form residuals to the structural equation and testing their significance, that is, using OLS estimate the model

$$I_t = \beta_1 + \beta_2 P_t + \beta_3 P_{t-1} + \beta_4 K_{t-1} + \delta \hat{v}_t + e_{2t}$$

Use a t-test for the null hypothesis $H_0 : \delta = 0$ versus $H_1 : \delta \neq 0$ at the 5% level of significance. By rejecting the null hypothesis, we conclude that P_t is endogenous. What do we conclude from the test? In the context of this simultaneous equations model what result should we find?

d. Obtain the 2SLS estimates of the investment equation using all eight exogenous and predetermined variables as IVs and software designed for 2SLS. Compare the estimates to the OLS estimates in part (a). Do you find any important differences?

e. Estimate the second-stage model $I_t = \beta_1 + \beta_2 \hat{P}_t + \beta_3 P_{t-1} + \beta_4 K_{t-1} + e_{2t}$ by OLS. Compare the estimates and standard errors from this estimation to those in part (d). What differences are there?

f. Let the 2SLS residuals from part (e) be $\hat{e}_{2t}$. Regress these residuals on all the exogenous and predetermined variables. If these instruments are valid, then the R^2 from this regression should be low, and none of the variables are statistically significant. The Sargan test for instrument validity is discussed in Section 10.4.3. The test statistic TR^2 has a chi-square distribution with degrees of freedom equal to the number of "surplus" IVs if the surplus instruments are valid. The investment equation includes three exogenous and/or predetermined variables out of the total of eight possible. There are $L = 5$ external instruments and $B = 1$ right-hand side endogenous variables. Compare the value of the test statistic to the 95th percentile value from the $\chi^2_{(4)}$ distribution. What do we conclude about the validity of the surplus instruments in this case?

```
13.255556 16.577368 19.282347 20.960143 19.766509 18.238731 17.573065 19.541720 20.375101 17.180415
12.705026  8.999780  9.054102 12.671263 14.421338 14.711907 19.796405 19.206691 17.419605 20.305654
22.657273
```

a.

```
Call:
lm(formula = i ~ p + plag + klag, data = klein)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.56562 -0.63169  0.03687  0.41542  1.49226 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 10.12579    5.46555   1.853 0.081374 .  
p            0.47964    0.09711   4.939 0.000125 ***
plag         0.33304    0.10086   3.302 0.004212 ** 
klag        -0.11179    0.02673  -4.183 0.000624 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’

Residual standard error: 1.009 on 17 degrees of freedom
  (因為不存在，1 個觀察量被刪除了)
Multiple R-squared:  0.9313,	Adjusted R-squared:  0.9192 
F-statistic: 76.88 on 3 and 17 DF,  p-value: 4.299e-10
```

b.

```
Call:
lm(formula = p ~ g + w2 + tx + time + plag + klag + elag, data = klein)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.9067 -1.3050  0.3226  1.3613  2.8881 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 50.38442   31.63026   1.593   0.1352
g            0.43902    0.39114   1.122   0.2820
w2          -0.07961    2.53382  -0.031   0.9754
tx          -0.92310    0.43376  -2.128   0.0530 .
time         0.31941    0.77813   0.410   0.6881
plag         0.80250    0.51886   1.547   0.1459
klag        -0.21610    0.11911  -1.814   0.0928 .
elag         0.02200    0.28216   0.078   0.9390
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.183 on 13 degrees of freedom
  (因為不存在，1 個觀察量被刪除了)
Multiple R-squared:  0.8261,	Adjusted R-squared:  0.7324 
F-statistic: 8.821 on 7 and 13 DF,  p-value: 0.0004481
```

d.

```
Call:
lm(formula = i ~ p + plag + klag + vhat, data = df)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.04645 -0.56030  0.06189  0.25348  1.36700 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 20.27821    4.70179   4.313 0.000536 ***
p            0.15022    0.10798   1.391 0.183222    
plag         0.61594    0.10147   6.070 1.62e-05 ***
klag        -0.15779    0.02252  -7.007 2.96e-06 ***
vhat         0.57451    0.14261   4.029 0.000972 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7331 on 16 degrees of freedom
Multiple R-squared:  0.9659,	Adjusted R-squared:  0.9574 
F-statistic: 113.4 on 4 and 16 DF,  p-value: 1.588e-11
```

$H_0: \delta = 0$

p-value: $0.000972 < 0.05$

$\Rightarrow$ reject H_0

P_t 为内生

e.

```
Call:
lm(formula = i ~ phat + plag + klag, data = df)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.8778 -1.0029  0.3058  0.7275  2.1831 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)   
(Intercept) 20.27821    9.97663   2.033  0.05802 . 
phat         0.15022    0.22913   0.656  0.52084   
plag         0.61594    0.21531   2.861  0.01083 * 
klag        -0.15779    0.04778  -3.302  0.00421 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.556 on 17 degrees of freedom
Multiple R-squared:  0.837,	Adjusted R-squared:  0.8082 
F-statistic: 29.09 on 3 and 17 DF,  p-value: 6.393e-07
```

Standard Errors change

f. Sargan test: $TR^2 < \chi^2(0.95, 4)$

non-reject H_0

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)?

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).

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?

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.

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?

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?

a. OLS在1987、1988年估計量差異不大，
个体間无差異

b. 加入个体、時間下標

u_i：只隨个体改變的誤差

e_{it}：隨个体 時間改變的誤差

c. EXPER （不在信賴區間）

d. $F^* = 11.68 > F_{0.05}(715, 712)$

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

H_0：沒个体差異

reject H_0

e. within transformation $\tilde{e}_{it} = e_{it} - \bar{e}_{it}$

column (4) SE 变大

f.

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.

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)?

a

```
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
```

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?

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.]

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?

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.

a

```
Call:
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
```

Yes, small class → perform better

No, p-value = 0.391, not significant

Yes

b.

```
Call:
plm(formula = readscore ~ small + aide + tchexper + boy + white_asian +
    freelunch, data = pdata, 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

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
```

small：更高

aide：都不显著

tchexper：效果下降

boy：效果上升

white_asian：效果上升

freelunch：一致

c.

```
        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
```

H_0：所有学校固定效果 $=0$

H_1：至少有一学校固定效果 $\neq 0$

$\alpha = 0.05$

$df_1 = n-1 = 79-1 = 78$

$df_2 = N-n-k = 5766-79-6 = 5681$

$F^* = 16.698 > \chi^2(0.95, 78, 5681) = 1.298$

reject H_0