Question 15.6

The file *mexican.dat* contains data collected in 2001 from the transactions of 754 Mexican sex workers. There is information on four transactions per worker. The labels ID and TRANS are used to describe a particular woman and a particular transaction. There are three categories of variables.

1. Sex worker characteristics: (i) AGE, (ii) an indicator variable ATTRACTIVE equal to 1 if the worker is attractive, and (iii) an indicator variable SCHOOL if she has completed secondary school or higher.
2. Client characteristics: (i) an indicator variable REGULAR equal to one if the client is a regular, (ii) an indicator variable RICH equal to one if the client is rich, and (iii) an indicator variable ALCOHOL if the client has consumed alcohol before the transaction.
3. Transaction characteristics: (i) the log of the price of the transaction LNPRICE, (ii) an indicator variable NOCONDOM equal to one if a condom was not used, and (iii) two indicator variables for location, BAR equal to one if the transaction originated in a bar and STREET if the transaction originated in the street.

(a) Estimate a fixed effects model with LNPRICE as the dependent variable, and as explanatory variables the client characteristics, and the remaining transaction characteristics.

```{r}
pdim(mexican)
```

```
Balanced Panel: n = 754, T = 4, N = 3016
```

```{r}
price.within <- plm(lnprice~regular+rich+alcohol+nocondom+bar+street, data =mexican ,model="within")
tbl <- tidy(price.within)
kable(tbl, digits=5, caption=
"Fixed effects using 'within' with full sample")
```

term	estimate	std.error	statistic	p.value
regular	0.03722	0.01685	2.20897	0.02728
rich	0.08264	0.02053	4.02544	0.00006
alcohol	-0.05686	0.02614	-2.17510	0.02973
nocondom	0.17028	0.02582	6.59570	0.00000
bar	0.29846	0.13445	2.21982	0.02653
street	0.45516	0.13046	3.48875	0.00049

(i) Why did we omit the sex worker characteristics?

Sex worker characteristics which presenting with 1 ~ 4 transactions are time-invariant over the time. Their effect cannot be separated from the individual effects given by the coefficients of the fixed-effects dummy variables.

(ii) What coefficient estimates are significantly different from zero at a 5% level of significance?

All coefficient estimates are significantly different from zero at a 5% level.

(iii) Gertler, Shah, and Bertozzi argue that the coefficient of NOCONDOM is a risk premium. Some sex workers are willing to take the risk of having unprotected sex because of the extra price some clients are willing to pay to avoid using a condom. What is your estimate of the risk premium? Interpret each of the other coefficient estimates. How is the price affected when clients are rich, are regular, and have consumed alcohol? How does the location of the transaction influence the price?

- According on the above output, the estimate of the risk premium for not using a condom is 17.03%. The exact estimate of the risk premium for not using a condom is ($e^{0.1703}$-1) = 0.1857 => 18.57%
- The price is approximately 3.72% higher for regular clients, 8.26% higher for rich clients, and the price is approximately 5.96% lower for clients who have consumed alcohol.
- Transaction characteristics have a large effect on the price. The risk premium is approximately 28.85% for the transaction is originated in a bar and 45.52% for the transaction is originated in the street.

(b) Estimate the model assuming random effects and with the characteristics of the sex workers added to the model. Compare the estimates with those from fixed effects. How have the coefficients of the common variables changed? How do the sex worker characteristics affect the price of commercial sex? How much extra does a client have to pay to have unprotected sex with an attractive secondary educated sex worker?

````
```{r}
price.random <- plm(lnprice~regular+rich+alcohol+nocondom+bar+street+age+attractive+school,
                    data = mexican, random.method ="swar", model ="random")
kable(tidy(price.random), digits=4, caption="The random effects results for the price")
```
````

term	estimate	std.error	statistic	p.value
(Intercept)	5.9104	0.1303	45.3529	0.0000
regular	0.0236	0.0162	1.4599	0.1443
rich	0.1160	0.0200	5.7903	0.0000
alcohol	0.0149	0.0250	0.5966	0.5507
nocondom	0.1390	0.0250	5.5535	0.0000
bar	0.4642	0.0999	4.6475	0.0000
street	0.1033	0.1011	1.0219	0.3068
age	-0.0258	0.0028	-9.3574	0.0000
attractive	0.2768	0.0602	4.5956	0.0000
school	0.2161	0.0453	4.7673	0.0000

From the above table, we can know that using random effects model and adding sex worker characteristics would lead to a huge change in the common estimate coefficients.

- For regular clients, the estimated coefficient is slightly declined, from 3.72% to 2.36%.

- The risk premium of not using a condom is slightly declined to from 17.03% to 13.9%.
- For rich clients, the price is approximately 11.6% higher, instead of 8.26% using the fixed effects model.
- For clients who have consumed alcohol, the price is from -5.69% using the fixed effects model to 1.49% using the random effects model. It suggests that these clients would pay higher price rather than the lower price, even though the coefficient is not significantly different with zero by using the random effects model.
- The estimated coefficients of indicator variable BAR and STREET change dramatically. The result is converse. For the fixed effects model, the transaction originated in the bar is less expensive than the transaction originated in the street. However, using the random model, the transaction originated in the bar is more expensive than the transaction originated in the street. The estimated coefficient of BAR is from 29.85% (fixed effects model) to 46.42% (random effects model). The estimated coefficient of STREET is from 45.52% (fixed effects model) to 10.33% (random effects model).
- The price of commercial sex is lower for old sex workers, and higher for attractive and secondary-educated sex workers, because the estimated coefficient of AGE is -0.0258 < 0, the estimated coefficients of ATTRACTIVE and SCHOOL are 0.2768 and 0.2161, respectively.
- A client must pay for unprotected sex with an attractive secondary-educated sex worker is approximately $(0.139 + 0.2768 + 0.2161)$*100% = 63.19% higher than for protected sex with an unattractive uneducated sex worker. The exact amount is 100*($e^{0.6319}$-1) % = 88.12%.

(c) Using the t-test statistic in (15.37) and a 5% significance level, test whether there are any significant differences between the fixed effects and random effects estimates of the coefficients on NOCONDOM, RICH, REGULAR, ALCOHOL, BAR, and STREET. If there are significant differences between any of the coefficients, should we rely on the fixed effects estimates or on the random effects estimates? Explain your choice.

$$t = \frac{b_{FE,k} - b_{RE,k}}{\left[\widehat{\text{var}}(b_{FE,k}) - \widehat{\text{var}}(b_{RE,k})\right]^{1/2}} = \frac{b_{FE,k} - b_{RE,k}}{\left[\text{se}(b_{FE,k})^2 - \text{se}(b_{RE,k})^2\right]^{1/2}} \qquad (15.37)$$

```{r}
fix_coef = summary(price.within)$coefficients
ran_coef = summary(price.random)$coefficients
fix_minus_ran = fix_coef[, 1]-ran_coef[c(2:4, 8:10), 1]
sd_fix_minus_ran = sqrt(fix_coef[, 2]^2-ran_coef[c(2:4, 8:10), 2]^2)
t_value = fix_minus_ran/sd_fix_minus_ran
round(t_value,4)
```

```
 regular     rich  alcohol nocondom      bar   street
  2.9012  -7.4565  -9.2249   7.6372   0.1799   1.9538
```

Using the formula to calculate the t-value for each variable. If Using the 5% significance level of critical value is 1.96, we reject the hypothesis that the estimators yield identical results. Our conclusion is that the random effects estimator is inconsistent, and that we should use the fixed effects estimator, or should attempt to improve the model specification.

Question 15.10

What is the relationship between crime and punishment? This important question has been examined by Cornwell and Trumbull using a panel of data from North Carolina. The cross sections are 90 counties, and the data are annual for the years 1981–1987. The data are in the file *crime.dat.* In these models the crime rate is explained by variables describing the deterrence effect of the legal system, wages in the private sector (which represents returns to legal activities), socioeconomic conditions such as population density and the percentage of young males in the population, and annual dummy variables to control for time effects. The authors argue that there may be heterogeneity across counties (unobservable county specific characteristics).

(a) What do you expect will happen to the crime rate if (i) deterrence increases, (ii) wages in the private sector increase, (iii) population density increases, (iv) the percentage of young males increases?

(i) If deterrence increases crime rates should drop.
(ii) If wages in the private sector increase the return to legal activities increases relative to the return to illegal activities. Therefore, crime rates should drop.
(iii) Higher population density is linked with a higher residential crime rate.
(iv) An increase in the percentage of young males should increase the crime rate, because young males are the most likely to be involved in illegal activities.

(b) Consider a model in which the crime rate (LCRMRTE) is a function of the probability of arrest (LPRBARR), the probability of conviction (LPRBCONV), the probability of a prison sentence (LPRBPRIS), the average prison sentence (LAVGSEN), and the average weekly wage in the manufacturing sector (LWMFG). Note that the logarithms of the variables are used in each case.

Estimate this model by least squares. (i) Discuss the signs of the estimated coefficients and their significance. Are they as you expected? (ii) Interpret the coefficient on LPRBARR.

```{r}
crimelm <- lm(lcrmrte~lprbarr+lprbconv+lprbpris+lavgsen+lwmfg, data = crime )
tb3 <- tidy(crimelm)
kable(tb3, digits=5, caption=
"The estimated equation with full sample")
```

term	estimate	std.error	statistic	p.value
(Intercept)	-6.08610	0.36536	-16.65788	0.00000
lprbarr	-0.65658	0.04035	-16.27383	0.00000
lprbconv	-0.44658	0.02774	-16.09793	0.00000
lprbpris	0.20823	0.07267	2.86541	0.00430
lavgsen	-0.05863	0.06060	-0.96747	0.33368
lwmfg	0.29206	0.06190	4.71830	0.00000

The estimated equation is:

LCRMRTE=-6.0861-0.65658*LPRBARR-0.44658*LPRBCONV+0.20823*LPRBPRIS-0.05863*LAVGSEN+0.29206*LWMFG

se(b1) =0.36536 se(b2) =0.04036 se(b3) =0.02774

se(b4) =0.07267 se(b5) =0.06060 se(b6) =0.06190

(i) LPRBARR, LPRBCONV, LPRBPRIS and LAVGSEN are explanatory variables that describe the deterrence effect of the legal system. We expect the coefficients of these variables to be negative. All these coefficients are negative except for the coefficient of LPRBPRIS. The variable LWMFG, which represents wages in the private sector, has a positive coefficient that is not consistent with our expectations. All coefficients are significantly different from zero at a 5% level of significance except for the coefficient of LAVGSEN.

(ii) The coefficient of LPRBARR suggests that a 1% increase in the probability of being arrested results in a 0.66% decrease in the crime rate.

(c) Estimate the model in (b) using a fixed effects estimator. (i) Discuss the signs of the estimated coefficients and their significance. Are they as you expected? (ii) Interpret the coefficient on LPRBARR and compare it to the estimate in (b). What do you conclude about the deterrent effect of the probability of arrest? (iii) Interpret the coefficient on LAVGSEN. What do you conclude about the severity of punishment as a deterrent?

```{r}
crime.within <- plm(lcrmrte~lprbarr+lprbconv+lprbpris+lavgsen+lwmfg, data = crime, model = "within")
summary(crime.within)
```

```
Oneway (individual) effect Within Model

Call:
plm(formula = lcrmrte ~ lprbarr + lprbconv + lprbpris + lavgsen +
    lwmfg, data = crime, model = "within")

Balanced Panel: n = 90, T = 7, N = 630

Residuals:
      Min.    1st Qu.     Median    3rd Qu.       Max.
-0.9948710 -0.0776321 -0.0020173  0.0789328  1.0771076

Coefficients:
          Estimate Std. Error t-value  Pr(>|t|)
lprbarr  -0.231271   0.037648 -6.1429 1.582e-09 ***
lprbconv -0.137803   0.022187 -6.2110 1.058e-09 ***
lprbpris -0.143137   0.039303 -3.6418  0.000297 ***
lavgsen   0.018281   0.030950  0.5907  0.554994
lwmfg    -0.166641   0.055267 -3.0152  0.002690 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    17.991
Residual Sum of Squares: 16.149
R-Squared:      0.10238
Adj. R-Squared: -0.05533
F-statistic: 12.2044 on 5 and 535 DF, p-value: 3.2267e-11
```

the estimated equation is:
LCRMRTE = α - 0.2313*LPRBARR - 0.1378*LPRBCONV - 0.1431*LPRBPRIS + 0.0183*LAVGSEN - 0.1666*LWMFG
se(b2) =0.0376, se(b3) =0.0222, se(b4) =0.0393, se(b5) =0.0309, se(b6) =0.0553
(i) All estimated coefficients have the expected sign except for LAVGSEN. All estimated coefficients are significantly different from zero at α =0.05 except for the coefficient for LAVGSEN.

(ii) The coefficient on LPRBARR suggests that a 1% increase in the probability of being arrested results in a 0.23% decrease in the crime rate. Comparing with the result of part b, increasing 1% in the probability of arrest would decrease the crime rate from 0.6566% to 0.2313%. Therefore, the deterrent effect of the probability of arrest would be decline once existing the county heterogeneity.

(iii) The estimated coefficient of LAVGESEN is 0.0183. Since the model is the log-log from, the estimated coefficient of LAVGSEN suggests that increasing 1% in the probability of arrest would lead to increase 0.0183% in the crime rate. However, a *t*-test on the significance this estimate yields a *t*-statistic of 0.5906 and a *p*-value of 0.555. Thus, a null hypothesis is not rejected which the coefficient of *LAVGSEN = 0.* There is no evidence to show that longer prison sentences would lead to a deterrent to crime.

(d) In the fixed effects estimation from part (c), test whether the county level effects are all equal.

H0: C1,1 = C1,2 = C1,3 …… = C1,90

H1: one or more is not the same

```
```{r}
crime.r <- plm(lcrmrte~lprbarr+lprbconv+lprbpris+lavgsen+lwmfg, data = crime, model = "pooling")
summary(crime.r)
```

Pooling Model

Call:
plm(formula = lcrmrte ~ lprbarr + lprbconv + lprbpris + lavgsen +
    lwmfg, data = crime, model = "pooling")

Balanced Panel: n = 90, T = 7, N = 630

Residuals:
     Min.   1st Qu.   Median   3rd Qu.   Max.
-1.549132 -0.244083 0.021836 0.260664 2.229848

Coefficients:
             Estimate Std. Error t-value Pr(>|t|)
(Intercept) -6.086098   0.365358 -16.6579 < 2.2e-16 ***
lprbarr   -0.656576   0.040345 -16.2738 < 2.2e-16 ***
lprbconv   -0.446581   0.027742 -16.0979 < 2.2e-16 ***
lprbpris   0.208227   0.072669   2.8654 0.004305 **
lavgsen   -0.058628   0.060599 -0.9675 0.333684
lwmfg   0.292061   0.061900   4.7183 2.937e-06 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares: 206.38
Residual Sum of Squares: 106.81
R-Squared: 0.48244
Adj. R-Squared: 0.47829
F-statistic: 116.331 on 5 and 624 DF, p-value: < 2.22e-16
```

```
```{r}
fcr <- qf(.95, df1=89, df2=535)
fcr
```

[1] 1.287491
```

F (0.95,89,535) = 1.287

F = [(106.81-16.149)/89]/[16.149/630-90-5] = 33.75

So, 33.75>1.287, Thus, we reject H0 and conclude that the county level effects are not all zero.

(e) To the specification in part (b) add the population density (LDENSITY) and the percentage of young males (LPCTYMLE), as well as dummy variables for the years 1982–1987 (D82–D87).

```{r}
crimelm2 <- lm(lcrmrte~lprbarr+lprbconv+lprbpris+lavgsen+lwmfg+ldensity+lpctymle
               +d82+d83+d84+d85+d86+d87, data = crime)
tblm2 <- tidy(crimelm2)
kable(tblm2, digits=4, caption=
"The coefficient estimates and standard errors from least squares")
```

term	estimate	std.error	statistic	p.value
(Intercept)	-3.6769	0.4662	-7.8869	0.0000
lprbarr	-0.4245	0.0419	-10.1289	0.0000
lprbconv	-0.2827	0.0288	-9.8185	0.0000
lprbpris	0.0877	0.0694	1.2647	0.2064
lavgsen	-0.1083	0.0577	-1.8764	0.0611
lwmfg	0.0160	0.0705	0.2267	0.8208
ldensity	0.3052	0.0274	11.1521	0.0000
lpctymle	0.1591	0.0840	1.8927	0.0589
d82	-0.0176	0.0574	-0.3063	0.7595
d83	-0.0669	0.0579	-1.1555	0.2483
d84	-0.1194	0.0585	-2.0386	0.0419
d85	-0.1056	0.0600	-1.7610	0.0787
d86	-0.0657	0.0612	-1.0747	0.2829
d87	-0.0101	0.0617	-0.1639	0.8699

```{r}
crimeno81.within <- plm(lcrmrte~lprbarr+lprbconv+lprbpris+lavgsen+lwmfg+ldensity+lpctymle
                   +d82+d83+d84+d85+d86+d87, data = crime, model = "within")
tbno81 <- tidy(crimeno81.within)
kable(tbno81, digits=4, caption=
"The coefficient estimates and standard errors from fixed effects (FE) estimation")
```

term	estimate	std.error	statistic	p.value
lprbarr	-0.1952	0.0367	-5.3169	0.0000
lprbconv	-0.1113	0.0217	-5.1238	0.0000
lprbpris	-0.0977	0.0384	-2.5418	0.0113
lavgsen	-0.0240	0.0315	-0.7617	0.4466
lwmfg	-0.5762	0.1330	-4.3342	0.0000
ldensity	0.7694	0.3377	2.2781	0.0231
lpctymle	1.2460	0.4346	2.8669	0.0043
d82	0.0253	0.0273	0.9261	0.3548
d83	0.0216	0.0352	0.6144	0.5392
d84	0.0121	0.0426	0.2831	0.7772
d85	0.0589	0.0528	1.1151	0.2653
d86	0.1586	0.0652	2.4319	0.0154
d87	0.2782	0.0772	3.6033	0.0003

(i) Compare the results obtained by using least squares (with no county effects) and the fixed effects estimator.

According to the result, the coefficient estimates obtained by using least squares and the fixed effects method are very different. The extents change impressively and there are some sign inversions. If we ignored county's effect, it would cause mistakes and bad conclusions.

(ii) Test the joint significance of the year dummy variables. Does there appear to be a trend effect?

The outcome of the test for the joint significance of the dummy variables is different for each of the two models.

```{r}
Hnull <- c("d82=0", "d83=0", "d84=0",'d85=0','d86=0','d87=0')
linearHypothesis(crimelm2,Hnull)
```

```
Linear hypothesis test

Hypothesis:
d82 = 0
d83 = 0
d84 = 0
d85 = 0
d86 = 0
d87 = 0

Model 1: restricted model
Model 2: lcrmrte ~ lprbarr + lprbconv + lprbpris + lavgsen + lwmfg + ldensity +
    lpctymle + d82 + d83 + d84 + d85 + d86 + d87

  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1    622 87.426
2    616 86.313  6    1.1128 1.3237 0.2442
```

In the least-squares estimated model with no fixed effects the F = 1.324 and p-values = 0.2442, so we don't reject H0 and conclude that there is no evidence of time effects.

```{r}
Hnull <- c("d82=0", "d83=0", "d84=0",'d85=0','d86=0','d87=0')
linearHypothesis(crimeno81.within,Hnull)
```

```
Linear hypothesis test

Hypothesis:
d82 = 0
d83 = 0
d84 = 0
d85 = 0
d86 = 0
d87 = 0

Model 1: restricted model
Model 2: lcrmrte ~ lprbarr + lprbconv + lprbpris + lavgsen + lwmfg + ldensity +
    lpctymle + d82 + d83 + d84 + d85 + d86 + d87

  Res.Df Df  Chisq Pr(>Chisq)
1    533
2    527  6 54.707  5.313e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

chi-squared = (numerator degrees of freedom) * F
54.707 = 6*F, so F = 9.118
In the fixed effects model above, the F = 9.118 and p-values = 0.000, so we reject H0 and conclude that there is evidence of time effects.

However, in 15.10 part(d) conclude that the county effects, and we do a further F-test to the model in order to conclude the time effects. As the result from part(d) and part(e), we concluded that the least-squares test result is misleading, and the year effects are important.

Based on the result above, using the fixed effects estimation. p-values of between D82 and D85 are all greater than $\alpha=0.05$, it implies that they are not significantly different with zero. However, the estimated coefficients of D86 and D87 are large and dramatical increase. It means that the crime rate increases in 1986 and in 1987.

(iii) Interpret the coefficient of LWMFG in both estimations.
Using the least square estimation, the estimated coefficient of LWMFG is 0.016. It indicates that increasing 1% the average weekly wage in the manufacturing sector, the crime rate would increase 0.016%, even though it is not significantly different with zero at 5% significance level since p-value=0.821 is greater than a=5%. Using the fixed effects estimation, the estimated coefficient of LWMFG is -0.5762. It implies that when the average weekly wage in the manufacturing sector increase 1%, the crime rate would decrease 0.5762%.

(f) Based on these results, what public policies would you advocate to deal with crime in the community?

According to the fixed effects estimates, LPRBARR, LPRBCONV, LPRBPRIS, LWMFG, LEDNSITY and LPCTYMLE have a significant effect on the crime rate, we can use this statistic information to improve the public policies.
LPRBARR: Hire more police in order to increase the arrest rate.
LPRBCONV, LPRBPRIS: Increase conviction and imprisonment might lower the crime rate.
LAVGSEN: lengthening the term of imprisonment is not working.
LEDNSITY, LWMFG and LPCTYMLE: Higher wages and the avoidance of high-density population areas are also important for public policy execution.