Name: Hajra Shahab
Andrew ID: hshahab

Homework #2- A/B Testing

Question #1:

a. Randomized experiment

Randomization at best can be described as a means to avoid selection in a given experiment. To measure average causal effect of a particular treatment on a subject, it is important to control for other factors that could have an impact on the subject. For instance, the difference in health conditions due a to a visit to the hospital should solely be determined in a subject due a change of one variable, which in this case is the visit to hospital vs. not visiting the hospital. However, if any other condition differs between the two groups (those visiting the hospital and those who don't visit the hospital), then the experiment would experience the effect of selection which means that selection would not be equal to zero. When selection is not equal to zero, it opens back door channels that allows unobservable factors to affect the independent and dependent variables. To ensure that selection is zero and to kill backdoor, the covariance between unobservable factor (ε) and dependent variable (health condition) is zero, we introduce randomization in our experiment. One way to randomize is to flip a coin and regardless of any criteria, assign each subject to control and treatment randomly. For instance, if the coin flip returns heads, there are assigned to control and treatment otherwise. This random assignment would ensure that all factors that could possibly affect the difference in health conditions due to a visit to the hospital are controlled for.

b. Fixed Effects Regression

In the absence of an experiment for multiple reasons, we resort to observational study where we follow same subject i over time, i.e., panel data. Usually, in studies like these, treatment is not randomized and we cannot infer causal effects for this reason. We thus resort to ordinary least squares regressions that helps us improve our results and brings us closer to the results that we would have achieved through randomized experiments. As we observe subject i over time, it is important to note that there are certain factors that would remain constant/ unchanged over time. For instance, in the example of providing internet access to college students, factors such as SAT score (taken at the time of admission), cohort (year applied to college) and major enrolled (also unlikely to change) would remain the same across time. These factors can be different for different students but do not change for any subject as we progress in our study over time. These factors when added in our regression model are called fixed effects regression. They help us overcome the omitted variable bias which may have an impact on the outcome of the regression. For panel data, since these variables do not change over time, we call them fixed effects.

c. Time Dummy

Time dummies helps us control for time specific fixed effects that do not vary across panel in a regression model. For instance, if we want to specifically see the impact of Year 1992 on our outcome, we would treat this as 1 and keep all other years as 0 in our regression. This is

different from time trend that focuses mainly on a sequential change over time in our outcome across all subjects.

Question #2:

a. Countries = 140; years = 5

```r
#Question 2a
#define data
length(unique(MyData$COUNTRY))
length(unique(MyData$YEAR))
```

```
> #Question 2a
> #define data
> length(unique(MyData$COUNTRY))
[1] 140
> length(unique(MyData$YEAR))
[1] 5
```

b. Interpretation (refer to screenshots below): A percent increase in per capita health expenditure leads to a 0.029% increase in composite health care attainment. We observe the same percent increase when we control for time-specific fixed effects (with time dummies (YEAR)).

c. The coefficients do not represent the causal effect of health expenditure on the measure of health care attainment because there are unobservable factors that may have an impact on the measure of health care attainment that we have not accounted for. This creates a backdoor where unobservable factors (ε) affect our independent and dependent variables in our regression model. These observable factors can be divided in three categories as things that change over time for all countries (time trends), different things that change for different countries but do not change over time (individual fixed effects) and things that change differently for different countries over time ($z_{i,t}$). We can improve our results in part (b) by accounting for fixed effects and first differences that will bring us closer to the true value of causal effects, however, we can still cannot compute causal effects as $z_{i,t}$ would remain unknown and cannot be calculated.
For this example, we can compute fixed effects and first differences (shown below in screenshots) to improve our results. As we compute first differences, we see that a percent increase in per capita health expenditure leads to a 0.0004% increase in composite health care attainment and is no longer statistically significant. When we compute fixed effects, we see that a percent increase in per capita health expenditure leads to a 0.001% increase in composite health care attainment which is also not statistically significant. This shows that per capita health expenditure doesn't lead to a statistically significant change in composite health care attainment when we account for some unobservables (through fixed effects and first differences) and there are other variables that were causing this change as the results are no longer significant.

Name: Hajra Shahab
Andrew ID: hshahab

Problem 2b

```
> #Question 2b
>
> #without time dummies
> plm_model<-plm(LCOMP~LHEXP +
+           LDALE +
+           LHC +
+           LGDPC +
+           factor(OECD) +
+           POPDEN, data = MyData,
+           index=c("COUNTRY","YEAR"),
+           model="pooling")
>
>
> #with time dummies
> plmtd<-plm(LCOMP~LHEXP +
+           YEAR +
+           LDALE +
+           LHC +
+           LGDPC +
+           factor(OECD) +
+           POPDEN, data = MyData,
+           index=c("COUNTRY","YEAR"),
+           model="pooling")
>
> stargazer(plm_model,plmtd,
+           se=list(
+             sqrt(diag(vcovHC(plm_model,
+                              method="arellano",
+                              type="HC1"))),
+             sqrt(diag(vcovHC(plmtd,
+                              method="arellano",
+                              type="HC1")))),
+           title="Panel OLS",
+           type="text",
+           model.numbers=FALSE,
+           column.labels=c("no time dummies",
+                           "time dummies"),
+           omit = c("factor[(]YEAR[)]",
+                    "factor[(]OECD[)]"),
+           add.lines = list(c("OECD Dummies","Yes","Yes"),
+                            c("Year Dummies","No","Yes"),
+                            c("LDALE Control","Yes","Yes"),
+                            c("LHC Control","Yes","Yes"),
+                            c("LGDPC Control","Yes","Yes"),
+                            c("POPDEN Control","Yes","Yes")))
```

Name: Hajra Shahab
Andrew ID: hshahab

```
Panel OLS
=================================================================
                            Dependent variable:
               ---------------------------------------------------
                                    LCOMP
                    no time dummies             time dummies
-----------------------------------------------------------------
LHEXP                    0.029*                     0.029*
                        (0.015)                    (0.016)

YEAR1994                                            0.00004
                                                   (0.001)

YEAR1995                                            0.0001
                                                   (0.001)

YEAR1996                                           -0.0005
                                                   (0.002)

YEAR1997                                           -0.001
                                                   (0.003)

LDALE                   0.408***                   0.408***
                        (0.034)                    (0.034)

LHC                     0.030**                    0.030**
                        (0.013)                    (0.013)

LGDPC                    0.014                      0.013
                        (0.017)                    (0.018)

POPDEN                  -0.00000                   -0.00000
                       (0.00000)                  (0.00000)

Constant                2.331***                   2.332***
                        (0.122)                    (0.123)

-----------------------------------------------------------------
OECD Dummies              Yes                        Yes
Year Dummies               No                        Yes
LDALE Control             Yes                        Yes
LHC Control               Yes                        Yes
LGDPC Control             Yes                        Yes
POPDEN Control            Yes                        Yes
Observations              700                        700
R2                       0.913                      0.913
Adjusted R2              0.912                      0.912
F Statistic    1,213.604*** (df = 6; 693) 724.023*** (df = 10; 689)
=================================================================
Note:                                *p<0.1; **p<0.05; ***p<0.01
>
>
```

Name: Hajra Shahab
Andrew ID: hshahab

Problem 2c

```
>
> #Question 2c
>
> #first-differences
> plmtdfd <-plm(LCOMP~LHEXP +
+           factor(YEAR) +
+           LDALE +
+           LHC +
+           LGDPC +
+           factor(OECD) +
+           POPDEN, data = MyData,
+           index=c("COUNTRY","YEAR"),
+           model="fd")
>
> stargazer(plm_model,plmtd,plmtdfd,
+           se=list(
+             sqrt(diag(vcovHC(plm_model,
+                              method="arellano",
+                              type="HC1"))),
+             sqrt(diag(vcovHC(plmtd,
+                              method="arellano",
+                              type="HC1"))),
+             sqrt(diag(vcovHC(plmtdfd,
+                              method="arellano",
+                              type="HC1")))),
+           title="Panel OLS",
+           type="text",
+           model.numbers=FALSE,
+           column.labels=c("no time dummies","time dummies","first-differences"),
+           omit = c("factor[(]YEAR[)]",
+                    "factor[(]OECD[)]"),
+           add.lines = list(c("OECD Dummies","Yes","Yes","Yes"),
+                            c("Year Dummies","No","Yes","Yes"),
+                            c("LDALE Control","Yes","Yes","Yes"),
+                            c("LHC Control","Yes","Yes","Yes"),
+                            c("LGDPC Control","Yes","Yes","Yes"),
+                            c("POPDEN Control","Yes","Yes","Yes")))
```

Name: Hajra Shahab
Andrew ID: hshahab

Panel OLS

	Dependent variable:		
	LCOMP		
	no time dummies	time dummies	first-differences
LHEXP	0.029*	0.029*	0.0004
	(0.015)	(0.016)	(0.0004)
YEAR1994		0.00004	
		(0.001)	
YEAR1995		0.0001	
		(0.001)	
YEAR1996		-0.0005	
		(0.002)	
YEAR1997		-0.001	
		(0.003)	
LDALE	0.408***	0.408***	0.580***
	(0.034)	(0.034)	(0.031)
LHC	0.030**	0.030**	0.018**
	(0.013)	(0.013)	(0.007)
LGDPC	0.014	0.013	
	(0.017)	(0.018)	
POPDEN	-0.00000	-0.00000	
	(0.00000)	(0.00000)	
Constant	2.331***	2.332***	0.001***
	(0.122)	(0.123)	(0.0001)
OECD Dummies	Yes	Yes	Yes
Year Dummies	No	Yes	Yes
LDALE Control	Yes	Yes	Yes
LHC Control	Yes	Yes	Yes
LGDPC Control	Yes	Yes	Yes
POPDEN Control	Yes	Yes	Yes
Observations	700	700	560
R2	0.913	0.913	0.906
Adjusted R2	0.912	0.912	0.905
F Statistic	1,213.604*** (df = 6; 693)	724.023*** (df = 10; 689)	893.106*** (df = 6; 553)

Note: *p<0.1; **p<0.05; ***p<0.01

Name: Hajra Shahab
Andrew ID: hshahab

```
> #using fixed-effects
> plmtdfe <-plm(LCOMP~LHEXP +
+           COUNTRY +
+           factor(YEAR) +
+           LDALE +
+           LHC +
+           LGDPC +
+           factor(OECD) +
+           POPDEN, data = MyData,
+           index=c("COUNTRY","YEAR"),
+           model="within")
>
> stargazer(plm_model,plmtd,plmtdfd,plmtdfe,
+           se=list(
+             sqrt(diag(vcovHC(plm_model,
+                              method="arellano",
+                              type="HC1"))),
+             sqrt(diag(vcovHC(plmtd,
+                              method="arellano",
+                              type="HC1"))),
+             sqrt(diag(vcovHC(plmtdfd,
+                              method="arellano",
+                              type="HC1"))),
+             sqrt(diag(vcovHC(plmtdfe,
+                              method="arellano",
+                              type="HC1")))),
+           title="Panel OLS",
+           type="text",
+           model.numbers=FALSE,
+           column.labels=c("no time dummies",
+                           "time dummies",
+                           "first-differences",
+                           "fixed-effects"),
+           omit = c("factor[(]YEAR[)]",
+                    "factor[(]OECD[)]"),
+           add.lines = list(c("OECD Dummies","Yes","Yes","Yes","Yes","Yes"),
+                            c("Year Dummies","No","Yes","Yes","Yes","Yes"),
+                            c("LDALE Control","Yes","Yes","Yes","Yes","Yes"),
+                            c("LHC Control","Yes","Yes","Yes","Yes","Yes"),
+                            c("LGDPC Control","Yes","Yes","Yes","Yes","Yes"),
+                            c("POPDEN Control","Yes","Yes","Yes","Yes","Yes")))
```

Name: Hajra Shahab
Andrew ID: hshahab

Panel OLS

	Dependent variable:			
	LCOMP			
	no time dummies	time dummies	first-differences	fixed-effects
LHEXP	0.029*	0.029*	0.0004	0.001
	(0.015)	(0.016)	(0.0004)	(0.001)
YEAR1994		0.00004		
		(0.001)		
YEAR1995		0.0001		
		(0.001)		
YEAR1996		-0.0005		
		(0.002)		
YEAR1997		-0.001		
		(0.003)		
LDALE	0.408***	0.408***	0.580***	0.573***
	(0.034)	(0.034)	(0.031)	(0.035)
LHC	0.030**	0.030**	0.018**	0.022**
	(0.013)	(0.013)	(0.007)	(0.009)
LGDPC	0.014	0.013		
	(0.017)	(0.018)		
POPDEN	-0.00000	-0.00000		
	(0.00000)	(0.00000)		
Constant	2.331***	2.332***	0.001***	
	(0.122)	(0.123)	(0.0001)	
OECD Dummies	Yes	Yes	Yes	Yes
Year Dummies	No	Yes	Yes	Yes
LDALE Control	Yes	Yes	Yes	Yes
LHC Control	Yes	Yes	Yes	Yes
LGDPC Control	Yes	Yes	Yes	Yes
POPDEN Control	Yes	Yes	Yes	Yes
Observations	700	700	560	700
R2	0.913	0.913	0.906	0.930
Adjusted R2	0.912	0.912	0.905	0.911
F Statistic	1,213.604*** (df = 6; 693)	724.023*** (df = 10; 689)	893.106*** (df = 6; 553)	1,044.673*** (df = 7; 553)
Note:				*p<0.1; **p<0.05; ***p<0.01