

# Homework 1: Stata Practice

XIANG-JYUN JHANG b08303024

YI-JIE WU b08302129

May 25, 2023

## 1 Read Data

### Question 1.1

The dataset I use are Taiwan Education Panel Survey (TEPS) and the Taiwan Education Panel Survey and Beyond (TEPS-B). It's a broad panel data following three group of students born in 1984-1985 and 1988-1989, doing survey in their junior and high school period and recording their background and performance. In the TEPS-B, researchers further tracking the information after those students entering labour market. We utilize the students' family background and labour market performance to conduct our analysis.

### Question 1.2

I use the following code to read my dataset, which is Stata format:

---

```
1 if "`c(username)'" == "Administrator" {  
2     global do = "C:\111Spring\labor\termPaper\do"  
3     global rawData = "C:\111Spring\labor\termPaper\rawData"  
4     global workData = "C:\111Spring\labor\termPaper\workData"  
5     global log = "C:\111Spring\labor\termPaper\log"  
6     global pic = "C:\111Spring\labor\termPaper\pic"  
7 }  
8 cd "$rawData"  
9 use "SH\SH_2001_A_student.dta", clear
```

---

## 2 Examine Data

### Question 2.1

I choose the variable *w1cls\_pn* to see the distribution of "number of students per classroom". The result shows that mean is 44.70 and the median is 45.

---

```
1 sum w1cls_pn, detail
```

---

### Question 2.2

I choose the variable *w1s208* to see the frequency of "parents divorce".

---

```
1 tab w1s208
```

---

### Question 2.3

We choose the variable *w1s208* to see if there is any missing value.

---

```
1 inspect w1s208
```

---

### Question 2.4

By Using **duplicates** command, we can observe that every observation has their unique value.

---

```
1 duplicates report
```

---

### Question 2.5

With the *stud\_id* variable, we can specify that every observation is unique.

---

```
1 duplicates report stud_id
```

---

## 3 Create Sample For Analysis: Part I & Part II

### Question 3.1

We consider the aforementioned divorce variable *w1s208*, we can utilize it to generate the *divorce* variable indicating whether the student encountered parents divorce.

---

```
1 gen divorce = (2 <= w1s208) & (w1s208 <= 5)
```

---

### Question 3.2

We use the *divorce* variable to create the divorce rate in the sample.

---

```
1 egen divorceRate = mean(divorce)
```

---

### Question 3.3

To make the *divorce* variable not be misunderstood, we can add label on it.

---

```
1 label define mapping_divorce 1"divorced" 0"not divorced"  
2 label value divorce mapping_divorce
```

---

### Question 3.4

For some non-answered value or unreasonable value, we can recode it as missing value

---

```
1 recode w1s208 (97/99 = .)
```

---

### Question 3.5

To investigate the outcome variable, we have to merge with the TEPS-B dataset

---

```
1 merge 1:1 stud_id using "SH\SH_2009.dta"
```

---

## 4 Visualize Data

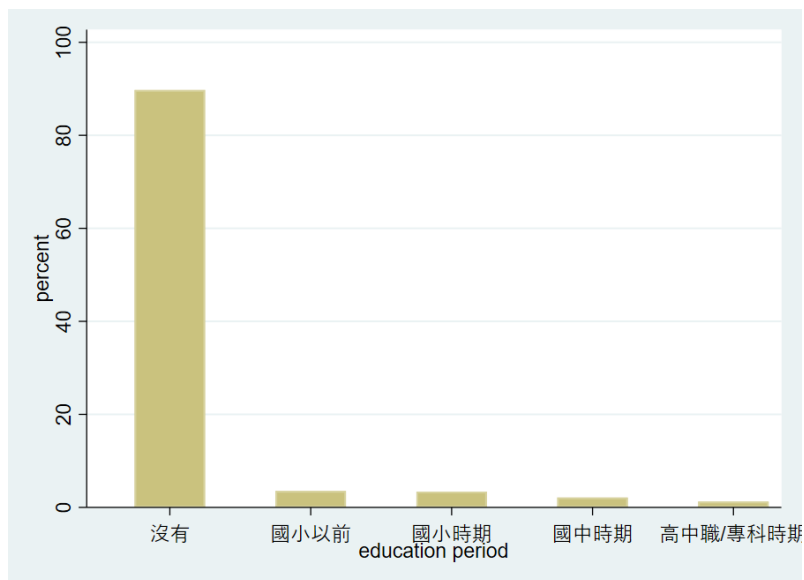
### Question 4.1

We draw the distribution of *w1s208*, which illustrates the period of parent's divorce, respectively is: not divorced, elementary school, junior high school, senior high school.

---

```
1 histogram w1s208, discrete percent width(0.5) xtitle(education period)  
   ytitle(percent) xlabel(1 2 3 4 5, valuelabel)
```

---



### Question 4.2

We use two-way graph to discuss the relationship between *divorce* and *undergraduate*, that is, whether encountering divorce reduces the probability of obtaining educational degree higher than undergraduate.

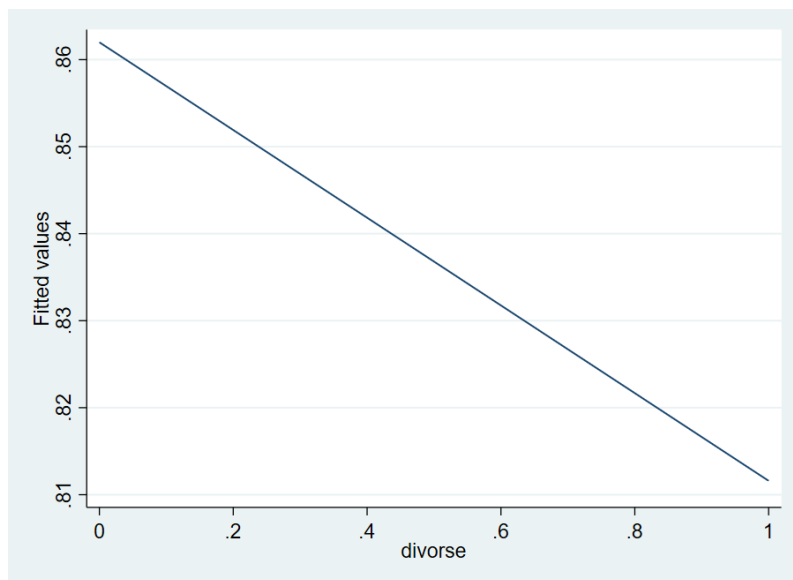
---

```

1 recode sh09v33 (9/99 = .)
2 gen undergraduate = 1 if (sh09v33>=5) & (sh09v33 != .)
3 replace undergraduate = 0 if (sh09v33<5) & (sh09v33 != .)
4 twoway (lfit undergraduate divorce)

```

---



## 5 Preliminary Analysis

### Question 5.1

We can regress the *undergraduate* on *divorce*

---

```
1 qui reg undergraduate divorce, r
```

---

The coefficient of *divorce* is negative, which means students in single parent families tend to have less human capital accumulation. This might influence their future performance and wages.

### Question 5.2

Consider the OVB formula

$$\hat{\alpha} \rightarrow \alpha + \beta \frac{Cov(X_i, D_i)}{Var(D_i)}$$

Since the children in the single parent family might have less economic situation or opportunity to accumulate their human capital, the correlation is non-zero and thus confounding the outcome.

### Question 5.3

We include other two variables as the control variable, which are the education background of both father and mother.

---

```
1 rename _merge merge_2009
2 merge 1:1 stud_id using "SH/SH_2001_G_parent.dta"
3 recode w1faedu (6/99 = .)
4 recode w1moedu (6/99 = .)
5 gen fa_undergraduate = 1 if (w1faedu>=3) & (w1faedu<=5) & (w1faedu != .)
6 replace fa_undergraduate = 0 if (w1faedu<3 | w1faedu>5) & (w1faedu != .)
7 gen ma_undergraduate = 1 if (w1moedu>=3) & (w1moedu<=5) & (w1moedu != .)
8 replace ma_undergraduate = 0 if (w1moedu<3 | w1moedu>5) & (w1moedu != .)
9
10 qui reg undergraduate divorce fa_undergraduate ma_undergraduate, r
```

---

Since parent's divorce took place after they got the educational degree, with utilizing the panel data, we can verify that the educational status should not be a confounding factor and can control for the causal inference.