

# Stata 教學

## Stata 教學

### 說明

檔案

準備

Stata 使用

Timetable (tentative)

### Stata Commands

基本操作

資料讀/寫

資料概覽

資料整理

Dummy Variable

Grouping

Reshaping

資料合併

### 論文摘要

Paying for Kidneys? A Randomized Survey and Choice Experiment

Background

Research Question

Experiment Design

Dataset

Results

Transplant Effect

Moral Concerns

## 說明

---

大家好！我是目前經研所二年級的亮丞，

2/1 以及 2/8 兩天要跟大家介紹 Stata 軟體使用以及研究資料的整理。

以下是一些資訊供大家參考～

## 檔案

下載完資料夾後，應該會含有以下的資料夾：

| NAME                           | DESCRIPTION         |
|--------------------------------|---------------------|
| <code>rawdata/</code>          | 資料夾內含課程所需的調查資料      |
| <code>do/</code>               | 放置 .do file 的資料夾（空） |
| <code>replication/</code>      | 目標要複製的AER文章研究資料夾    |
| <code>Paying4Kidney.pdf</code> | <b>AER文章本體</b>      |
| <code>appendix.pdf</code>      | AER文章附錄，內含有實驗的問題    |

## 準備

由於這一次 Stata介紹會以能夠操作實際研究為目標，因此除了用來學基本操作的資料集外，我們的主軸會放在複製一份真實的論文研究結果上面。

這邊老師選的文章是 2019年在AER上發表的 *Paying for Kidneys? A Randomized Survey and Choice Experiment*

與之後大家可能會遇到的資料相同，本篇的資料是一個隨機分配的試驗資料。

因此在 2/1 號上課之前，想先請大家先閱讀一下資料夾內的文章 *Paying for Kidneys? A Randomized Survey and Choice Experiment (2019)* （放在 `./Paying4Kidney.pdf` ）。

只需要先大概看過論文中的 Abstract 以及 Introduction 部分，

大致了解文章的研究問題以及所使用的資料即可。

本文後面也有提供我自己做的文章摘要可以給大家多個參考。

## Stata 使用

本次大致會涵蓋 Stata軟體以及do file 的基本使用，也會說明研究資料夾應該如何安排。（如果大家都會就快速跳過XD）

另外以拿到一筆新的資料集為例，從資料概覽、資料的清理以及製作欄位，到最後用清理後的資料產生基本的敘述統計表、回歸分析，並保存結果來進行。

最終會把 *Paying for Kidneys* 一文中，比較重要的圖表用Stata實現一次。

如果時間還夠的話，再多涵蓋一些用 Stata Programming (流程控制、自定義program等用法)

(如果時間還夠的話....

以下則是目前的時間安排以及使用的資料：

## Timetable (tentative)

| TIME | CONTENT                          | DATA                    |
|------|----------------------------------|-------------------------|
| 2/1  | Stata 基本使用、基本指令、資料夾整理 (20min)    | 社會變遷調查 (TSCS) 資料        |
|      | 資料概覽、資料整理 (I)、Regression         | AER replication dataset |
| 2/8  | 資料整理 (II)、Stata 圖表               | AER replication dataset |
|      | Stata for Programming (optional) |                         |

- 目標：
  - 使用 Stata 來進行資料前處理，製作能供分析的最終資料
  - 產生以及保存敘述統計結果
  - 產生以及保存回歸分析結果
- 材料：
  - 社會變遷調查的綜合調查問卷資料 (TCSC Survey) 來學習：
    - 資料清理
    - 增加欄位
    - 合併欄位
  - Paying for Kidneys AER paper Replication
    - 敘述統計
    - 回歸分析
    - 產生以及保存圖表

## Stata Commands

---

主要會涵蓋以下指令：

### 基本操作

包含資料夾操作、do-file常見指令。

e.g.

```
1 cd
2 ls
3 mkdir
4
5 log
6 capture
7 quietly
8 set
9 display
```

## 資料讀/寫

- .dta file
- .csv file
- stata 版本造成編碼問題

e.g.

```
1 use / save
2 import / export delimited
```

```
1 preserve
2 restore
```

## 資料概覽

可以檢視資料集、探索資料、檢查資料重複等指令。

e.g.

```
1 list
2 describe
3 codebook
4 assert
5
6 table
7 summary
8 duplicated
```

## 資料整理

一些常用來新增欄位、整理欄位等指令，  
以及製作虛擬變數、群組化操作的方式。

e.g.

```
1  # append variable
2  generate
3  egen
4
5  drop
6  keep
7  replace
8  rename
9  label
10 recode
11
12 # change types
13 decode
14 encode
```

## Dummy Variable

### 1. 手動生成

```
1  gen dummy = 0
2  replace dummy = 1 if (condition)
```

### 2. `xi` 指令

```
1  xi i.var
```

會自動生成 (n-1) 個虛擬變數

### 3. `tab + gen`

```
1  tab var, gen(var_<suffix>)
```

## Grouping

```
1  bysort
```

## Reshaping

```
1 reshape wide/long
```

## 資料合併

包含水平合併(merging)以及垂直合併(append)。

```
1 append using "filepath"
2 merge 1:1 `common_var' using "filepath"
```

## 論文摘要

---

### Paying for Kidneys? A Randomized Survey and Choice Experiment

- Authors: Julio J. Elías, Nicola Lacetera, Mario Macis
- Publication: AER, Aug. 2019
- [Link](#)

### Background

- 現行美國的器官捐贈制度為何？
  - 目前是沒有給捐贈者報酬的制度存在。
  - 病人需被加入等待名單，按照官方決定來決定器官移植的優先順序。
  - 美國目前一年約 19,000 個腎臟移植手術，大約滿足一半的需求；也就是說，一年等待移植的數量約為 38,000 人。

### Research Question

- 人們對於器官捐贈有報酬的制度偏好為何？
  - 研究以美國國民對於腎臟移植為例
- 研究假設
  - 對開放器官捐贈付款的偏好可能被以下因素影響：
    - 道德上的價值考量
    - 可以增加的器官移植數量 (與現行相比)
  - 可以下面的效益函數描述

- $s$  : system;  $\Pi$  : system characteristics;  $Q$  : change in transplants

$$U_{is} = \Pi_{is} + q_{is} \cdot Q_s \quad (1)$$

## Experiment Design

- 實驗共分為八個組別（也就是八個制度）
  - 付款型態：現金 / 非現金
  - 付款人：病人 / 政府官方
  - 付款額：\$30,000 / \$10,000
- 每一個組別的人皆會在 **五個不同的情境** 下做出選擇
  - 每個情境為不同制度比起現行制度可以**增加的器官移植數量**
    - $\{0, 4k, 9k, 14k, 19k\}$
  - 每一位受試者選擇 **支持** 或是 **反對** 新制度
- 每一個組別下，會在隨機抽一半的人針對六個道德議題表達意見
  - $[-10, 10]$  的分數

## Dataset

- 2,666 受試者，按照美國的人口特徵分佈

幾個重要的欄位為：

- 受試者 ID (Responseid)
- 五個情境 (level)
- 受試者特徵 (性別、年齡、族裔、婚姻狀況等)
- 受試者組別 (cond\_8)
- 受試者回答：
  - 支持 (favor) =  $\{1, 0\}$
  - 道德議題分數
- 敘述統計：
  - 組別：

TABLE 1—CHARACTERISTICS OF THE KIDNEY PROCUREMENT AND ALLOCATION SYSTEMS RANDOMLY ASSIGNED TO STUDY PARTICIPANTS

|          | Nature of compensation | Amount | Payer         | Number of respondents without morality assessment module | Number of respondents with morality assessment module |
|----------|------------------------|--------|---------------|--|---|
| System 1 | Cash                   | \$30K  | Public agency | 188  | 168   |
| System 2 | Cash                   | \$100K | Public agency | 160  | 169   |
| System 3 | Cash                   | \$30K  | Patient       | 178  | 152   |
| System 4 | Cash                   | \$100K | Patient       | 182  | 135   |
| System 5 | Noncash                | \$30K  | Public agency | 188  | 175   |
| System 6 | Noncash                | \$100K | Public agency | 167  | 162   |
| System 7 | Noncash                | \$30K  | Patient       | 165  | 159   |
| System 8 | Noncash                | \$100K | Patient       | 162  | 156   |

Table B1: Randomization check

| Outcome variables:                                | Woman               | Socially liberal    | Socially conservative | Economically liberal | Economically conservative | Atheist, agnostic   | College educated or more | Income >\$50K       | Volunteered in previous two years | Know of someone who had a transplant |
|---|---------------------|---------------------|-----------------------|----------------------|---------------------------|---------------------|--------------------------|---------------------|-----------------------------------|--------------------------------------|
|   | (1)                 | (2)                 | (3)                   | (4)                  | (5)                       | (6)                 | (7)                      | (8)                 | (9)                               | (10)                                 |
| Conditions:                                       |                     |                     |                       |                      |                           |                     |                          |                     |                                   |                                      |
| \$30K, cash, agency pays, morality module         | -0.056<br>(0.053)   | 0.011<br>(0.046)    | -0.092*<br>(0.049)    | -0.002<br>(0.042)    | -0.073<br>(0.051)         | 0.082**<br>(0.035)  | 0.015<br>(0.051)         | -0.002<br>(0.055)   | -0.038<br>(0.047)                 | -0.000<br>(0.043)                    |
| \$30K, cash, recipient pays                       | 0.023<br>(0.054)    | 0.012<br>(0.047)    | -0.086*<br>(0.050)    | -0.011<br>(0.042)    | -0.045<br>(0.051)         | 0.090**<br>(0.036)  | 0.014<br>(0.052)         | -0.040<br>(0.056)   | -0.016<br>(0.048)                 | -0.009<br>(0.044)                    |
| \$30K, cash, recipient pays, morality module      | 0.012<br>(0.053)    | 0.022<br>(0.046)    | -0.004<br>(0.049)     | 0.027<br>(0.042)     | -0.051<br>(0.051)         | 0.057<br>(0.035)    | 0.043<br>(0.051)         | -0.016<br>(0.055)   | -0.030<br>(0.047)                 | 0.022<br>(0.043)                     |
| \$100K, cash, agency pays                         | -0.032<br>(0.052)   | 0.019<br>(0.046)    | -0.004<br>(0.048)     | 0.016<br>(0.041)     | -0.040<br>(0.050)         | 0.067*<br>(0.035)   | 0.086*<br>(0.051)        | 0.098*<br>(0.055)   | -0.010<br>(0.047)                 | -0.006<br>(0.043)                    |
| \$100K, cash, agency pays, morality module        | -0.020<br>(0.055)   | -0.008<br>(0.048)   | 0.032<br>(0.050)      | -0.002<br>(0.043)    | 0.020<br>(0.052)          | 0.053<br>(0.036)    | 0.052<br>(0.053)         | 0.054<br>(0.057)    | -0.018<br>(0.049)                 | 0.020<br>(0.044)                     |
| \$100K, cash, recipient pays                      | 0.001<br>(0.052)    | -0.041<br>(0.045)   | -0.006<br>(0.048)     | -0.038<br>(0.041)    | -0.026<br>(0.050)         | 0.014<br>(0.035)    | -0.000<br>(0.050)        | 0.006<br>(0.054)    | 0.013<br>(0.046)                  | -0.021<br>(0.042)                    |
| \$100K, cash, recipient pays, morality module     | -0.060<br>(0.056)   | 0.029<br>(0.049)    | -0.041<br>(0.052)     | -0.038<br>(0.044)    | 0.012<br>(0.054)          | 0.011<br>(0.038)    | 0.063<br>(0.055)         | 0.024<br>(0.057)    | -0.010<br>(0.050)                 | 0.025<br>(0.046)                     |
| \$30K, non-cash, agency pays                      | -0.016<br>(0.052)   | 0.032<br>(0.045)    | -0.043<br>(0.048)     | 0.005<br>(0.041)     | -0.069<br>(0.049)         | 0.032<br>(0.034)    | 0.069<br>(0.050)         | 0.092*<br>(0.054)   | 0.027<br>(0.046)                  | 0.048<br>(0.042)                     |
| \$30K, non-cash, agency pays, morality module     | -0.029<br>(0.053)   | 0.001<br>(0.046)    | -0.021<br>(0.048)     | -0.003<br>(0.041)    | -0.028<br>(0.050)         | 0.052<br>(0.035)    | 0.030<br>(0.051)         | 0.036<br>(0.054)    | 0.003<br>(0.047)                  | 0.037<br>(0.043)                     |
| \$30K, non-cash, recipient pays                   | -0.024<br>(0.053)   | 0.049<br>(0.046)    | -0.030<br>(0.049)     | 0.053<br>(0.042)     | -0.059<br>(0.051)         | 0.023<br>(0.036)    | 0.071<br>(0.051)         | 0.061<br>(0.055)    | 0.020<br>(0.048)                  | -0.005<br>(0.043)                    |
| \$30K, non-cash, recipient pays, morality module  | -0.020<br>(0.054)   | -0.004<br>(0.047)   | 0.004<br>(0.049)      | 0.018<br>(0.042)     | 0.001<br>(0.051)          | 0.082**<br>(0.036)  | 0.065<br>(0.052)         | 0.021<br>(0.055)    | -0.000<br>(0.048)                 | 0.007<br>(0.044)                     |
| \$100K, non-cash, agency pays                     | -0.024<br>(0.053)   | 0.010<br>(0.046)    | 0.010<br>(0.049)      | -0.016<br>(0.042)    | -0.000<br>(0.051)         | -0.000<br>(0.036)   | 0.125**<br>(0.052)       | 0.041<br>(0.055)    | -0.014<br>(0.048)                 | -0.003<br>(0.043)                    |
| \$100K, non-cash, agency pays, morality module    | -0.067<br>(0.054)   | 0.019<br>(0.047)    | -0.028<br>(0.050)     | 0.040<br>(0.042)     | -0.011<br>(0.052)         | 0.022<br>(0.036)    | 0.142***<br>(0.052)      | 0.054<br>(0.056)    | 0.013<br>(0.048)                  | 0.036<br>(0.044)                     |
| \$100K, non-cash, recipient pays                  | -0.027<br>(0.054)   | 0.008<br>(0.047)    | -0.058<br>(0.049)     | -0.001<br>(0.042)    | -0.073<br>(0.051)         | 0.020<br>(0.036)    | 0.016<br>(0.052)         | -0.082<br>(0.056)   | 0.031<br>(0.048)                  | 0.019<br>(0.044)                     |
| \$100K, non-cash, recipient pays, morality module | 0.031<br>(0.054)    | -0.014<br>(0.047)   | -0.035<br>(0.050)     | 0.013<br>(0.043)     | -0.093*<br>(0.052)        | 0.101***<br>(0.036) | -0.029<br>(0.052)        | 0.008<br>(0.055)    | -0.024<br>(0.048)                 | 0.008<br>(0.044)                     |
| Constant  | 0.527***<br>(0.037) | 0.245***<br>(0.032) | 0.330***<br>(0.034)   | 0.186***<br>(0.029)  | 0.388***<br>(0.035)       | 0.085***<br>(0.024) | 0.330***<br>(0.035)      | 0.446***<br>(0.038) | 0.729***<br>(0.033)               | 0.197***<br>(0.030)                  |
| Observations                                      | 2,666               | 2,666               | 2,666                 | 2,666                | 2,666                     | 2,666               | 2,666                    | 2,473               | 2,666                             | 2,666                                |
| R-squared   | 0.003               | 0.002               | 0.005                 | 0.004                | 0.005                     | 0.009               | 0.008                    | 0.008               | 0.002                             | 0.002                                |

Notes: The table reports coefficient estimates from linear regressions. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

○ 基本特徴：



TABLE 2—CHARACTERISTICS OF RESPONDENTS AND COMPARISON WITH AMERICAN  
COMMUNITY SURVEY DATA (PERCENT)

|                                  | Qualtrics sample<br>( <i>N</i> = 2,666)<br>(1) | US<br>population<br>(2) |
|----------------------------------|--|-------------------------|
| Women                            | 50.8   | 51.4                    |
| Age 18–34                        | 24.2   | 30.2                    |
| Age 35–54                        | 36.4   | 33.5                    |
| Age 55+                          | 39.5   | 36.3                    |
| White (non-Hispanic)             | 63.5   | 61.3                    |
| Black                            | 12.6   | 13.3                    |
| Hispanic                         | 14.1   | 17.8                    |
| Asian                            | 6.3  | 5.7                     |
| Other race/ethnicity             | 3.4  | 1.9                     |
| Less than HS diploma             | 7.5  | 12.6                    |
| HS diploma/GED                   | 24.4   | 27.7                    |
| Some college or Associate degree | 30.3   | 31.0                    |
| Bachelor's degree                | 22.9   | 18.3                    |
| Graduate degree                  | 14.8   | 10.5                    |
| Married                          | 59.4   | 51.6                    |
| Employed                         | 60.4   | 61.4                    |
| Income < \$50,000                | 34.5   | 43.7                    |
| \$50,000 ≤ Income < \$100,000    | 35.6   | 30.0                    |
| Income ≥ 100,000                 | 29.9   | 26.2                    |

## Results

### Transplant Effect

- 沒有任何增加的情境下，57 % 支持付費制度；滿足所有需求下，70% 支持。

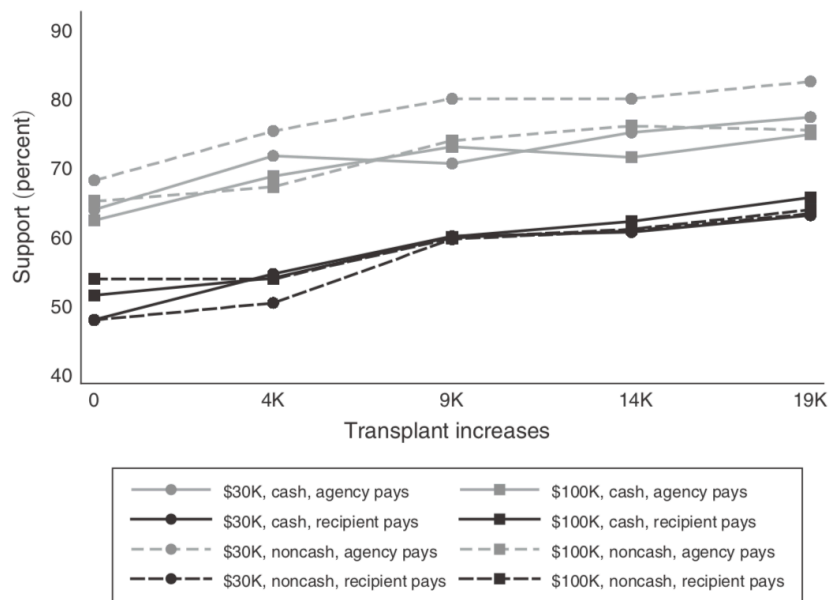


FIGURE 2. SUPPORT FOR ALTERNATIVE ORGAN PROCUREMENT SYSTEMS, BY LEVEL OF TRANSPLANT GAINS

- 隨著移植數量的增加，八個組別都呈現支持率上升

- 10% 移植量增加，帶來 2.6% 支持率上升（線性）。
- 直接付款支持率比起公部門分配低了 15-16%。
- 錢的多寡、是否是現金則沒有太多的效果。
- 交乘項不顯著：制度特徵不會對邊際支持率造成影響。

## Moral Concerns

TABLE 4—TRANSPLANT INCREASES, PROCEDURAL FEATURES, AND MORAL CONCERNS

| Outcome variable:<br>Regressors:           | Concerns<br>for ex-<br>ploitation | Concerns<br>for lack of<br>autonomous<br>choice | Concerns<br>for undue<br>influence | Concerns<br>for fairness<br>to donors | Concerns<br>for fairness<br>to patients | Concerns<br>for harm<br>to human<br>dignity | Principal component<br>of moral concerns |                   |                   |
|--|-----------------------------------|---|------------------------------------|---------------------------------------|---|---|--|-------------------|-------------------|
|  | (1)                               | (2)   | (3)                                | (4)                                   | (5)                                     | (6)   | (7)                                      | (8)               | (9)               |
| Transplant increase<br>(percentage points) | -0.009<br>(0.004)                 | -0.018<br>(0.004)                               | -0.02<br>(0.005)                   | -0.011<br>(0.004)                     | -0.024<br>(0.005)                       | -0.019<br>(0.004)                           | -0.003<br>(0.001)                        | -0.003<br>(0.001) | -0.003<br>(0.001) |
| Cash                                       | 0.706<br>(0.373)                  | 0.694<br>(0.355)                                | 0.488<br>(0.389)                   | 0.621<br>(0.393)                      | 1.115<br>(0.370)                        | 0.902<br>(0.356)                            | 0.138<br>(0.053)                         | 0.135<br>(0.054)  |                   |
| Recipient pays                             | 0.245<br>(0.374)                  | 1.058<br>(0.356)                                | 1.054<br>(0.391)                   | 0.349<br>(0.394)                      | 3.261<br>(0.374)                        | 0.897<br>(0.358)                            | 0.202<br>(0.053)                         | 0.184<br>(0.054)  |                   |
| \$100K                                     | -0.518<br>(0.371)                 | 0.158<br>(0.354)                                | 0.225<br>(0.389)                   | 0.089<br>(0.391)                      | 0.702<br>(0.369)                        | 0.051<br>(0.355)                            | 0.019<br>(0.052)                         | 0.004<br>(0.055)  |                   |
| Cash × transplant<br>increase              | 0.005<br>(0.005)                  | 0.004<br>(0.004)                                | 0.005<br>(0.005)                   | 0.012<br>(0.004)                      | -0.001<br>(0.005)                       | 0.003<br>(0.004)                            | 0.001<br>(0.001)                         | 0.001<br>(0.001)  | 0.001<br>(0.001)  |
| Recipient pays<br>× transplant increase    | 0.006<br>(0.005)                  | 0.004<br>(0.004)                                | 0.01<br>(0.005)                    | 0.009<br>(0.004)                      | -0.001<br>(0.005)                       | 0.003<br>(0.004)                            | 0.001<br>(0.001)                         | 0.001<br>(0.001)  | 0.001<br>(0.001)  |
| \$100K × transplant<br>increase            | 0.000<br>(0.005)                  | 0.000<br>(0.004)                                | 0.000<br>(0.005)                   | -0.006<br>(0.004)                     | 0.003<br>(0.005)                        | 0.002<br>(0.004)                            | -0.000<br>(0.001)                        | -0.000<br>(0.001) | -0.000<br>(0.001) |
| Constant                                   | -3.102<br>(0.357)                 | -1.389<br>(0.346)                               | 0.091<br>(0.366)                   | -3.212<br>(0.376)                     | -2.996<br>(0.340)                       | 0.459<br>(0.335)                            | -0.12<br>(0.050)                         | 0.000<br>(0.126)  | 0.053<br>(0.008)  |
| Control variables                          |                                   |   |                                    |                                       |   |   | x  |                   |                   |
| Individual fixed effects                   |                                   |   |                                    |                                       |   |   |  |                   | x                 |
| Observations                               | 6,380                             | 6,380   | 6,380                              | 6,380                                 | 6,380                                   | 6,380                                       | 6,380                                    | 5,990             | 6,380             |
| R <sup>2</sup>                             | 0.006                             | 0.012   | 0.011                              | 0.006                                 | 0.062                                   | 0.012                                       | 0.020                                    | 0.052             | 0.938             |

- 移植數量 與 道德疑慮 呈現微小的負向顯著關係。
- 現金支付 則 顯著增加道德疑慮。