# Quasi Experiment EE13.1

# EE12.2

我們直接從課本的 Empirical Exercise 13.1 來複習如何用 Stata 實作 quasi-experimental analysis，並做出解釋。

# EE13.1 Questions

**E13.1** A prospective employer receives two resumes: a resume from a white job applicant and a similar resume from an African American applicant. Is the employer more likely to call back the white applicant to arrange an interview? Marianne Bertrand and Sendhil Mullainathan carried out a randomized controlled experiment to answer this question. Because race is not typically included on a resume, they differentiated resumes on the basis of "white-sounding names" (such as Emily Walsh or Gregory Baker) and "African American–sounding names" (such as Lakisha Washington or Jamal Jones). A large collection of fictitious resumes was created, and the presupposed "race" (based on the "sound" of the name) was randomly assigned to each resume. These resumes were sent to prospective employers to see which resumes generated a phone call (a callback) from the prospective employer. Data from the experiment and a detailed data description are on the text website, **http://www.pearsonglobaleditions.com**, in the files **Names** and **Names_Description**.[8]

# EE13.1 Questions

**a.** Define the *callback rate* as the fraction of resumes that generate a phone call from the prospective employer. What was the callback rate for whites? For African Americans? Construct a 95% confidence interval for the difference in the callback rates. Is the difference statistically significant? Is it large in a real-world sense?

**b.** Is the African American/white callback rate differential different for men than for women?

**c.** What is the difference in callback rates for high-quality versus low-quality resumes? What is the high-quality/low-quality difference for white applicants? For African American applicants? Is there a significant difference in this high-quality/low-quality difference for whites versus African Americans?

**d.** The authors of the study claim that race was assigned randomly to the resumes. Is there any evidence of nonrandom assignment?

# EE13.1 Data Descrptions

## Names Data

1. Observations: 4870 resumes
2. Time Period : 2001

**Variable Descriptions**

| Variable Name | Description |
|---|---|
| | ***Key Variables*** |
| *firstname* | applicant's first name |
| *female* | 1 = female |
| *black* | 1 = black |
| *high* | 1= high quality resume |
| *call_back* | 1= applicant was called back |
| *chicago* | 1 = data from Chicago |
| | |
| | ***Detailed Information on Resume*** |
| *ofjobs* | number of jobs listed on resume |
| *yearsexp* | number of years of work experience on the resume |
| *honors* | 1=resume mentions some honors |
| *volunteer* | 1=resume mentions some volunteering experience |
| *military* | 1=applicant has some military experience |
| *empholes* | 1=resume has some employment holes |
| *workinschool* | 1=resume mentions some work experience while at school |
| *email* | 1=email address on applicant's resume |
| *computerskills* | 1=resume mentions some computer skills |
| *specialskills* | 1=resume mentions some special skills |
| *college* | applicant has college degree or more |
| | |
| | ***Detailed Information Concerning Employer*** |
| *expminreq* | min experience required, if any |
| *eoe* | 1=ad mentions employer is EOE |
| *manager* | 1=manager wanted |
| *supervisor* | 1=supervisor wanted |
| *secretary* | 1=secretary wanted |
| *offsupport* | 1=office support |
| *salesrep* | 1=sales representative wanted |
| *retailsales* | 1=retail sales worker wanted |
| *req* | 1=ad mentions any requirement for job |
| *expreq* | 1=ad mentions some experience requirement |
| *comreq* | 1=ad mentions some communication skills requirement |
| *educreq* | 1=ad mentions some educational requirement |
| *compreq* | 1=ad mentions some computer skill requirement |
| *orgreq* | 1=ad mentions some organizational skills requirement |
| *manuf* | 1=employer industry is manufacturing |
| *transcom* | 1=employer industry is transport/communication |
| *bankreal* | 1=employer industry is finance, insurance, real estate |
| *trade* | 1=employer industry is wholesale or retail trade |
| *busservice* | 1=employer industry is business and personal services |
| *othservice* | 1=employer industry is health, educ. and social services |
| *missind* | 1=employer industry is other/unknown |

# EE13.1 a

The command:

```
mean callback
```

gives us the fraction of resumes that generate a phone call from prospective employer.

The expression:

```
ci mean callback
```

gives us the same result.

# EE13.1 a

For whites, simply add a condition:

```
ci mean callback if black==0
```

For blacks:

```
ci mean callback if black==1
```

# EE13.1 a conti.

To understand how Stata gives us the 95% C.I., recall the point estimator in the previous semester.

Let

$$y_i = \begin{cases} 1, \text{if receive phone call} \\ 0, \text{o.w.} \end{cases}$$

Clearly, $y_i$ follows $Bernoulli(p)$ where $p$ is the unknown fraction of resumes that generate a phone call from prospective employer.

Naturally, we would like to use $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ to estimate $p$. Where $n = 4870$.

We need to know $se(\bar{y})$ in order to build up a interval estimator.

# EE13.1 a conti.

Now, our concern would be: What is the sampling distribution of $\bar{y}$?

Given

$$y_i \overset{i.i.d.}{\sim} Bernoulli(p)$$

then

$$\sum_{i=1}^{n} y_i \sim Binomial(n, p)$$

which can be approximated by normal distribution

$$\sum_{i=1}^{n} y_i \overset{a}{\sim} N(np, np(1-p))$$

as $n \to \infty$. And let $X = \frac{1}{n}\sum_{i=1}^{n} y_i$, then

$$X \sim N(p, \frac{p(1-p)}{n})$$

That is, $se(\bar{y})$ can be calculated easily from the approximated normal distribution.

# EE13.1 a conti.

With the normal quantile $Z_{0.05} = -1.96$, we know the interval estimator for $p$ is:

$[\bar{y} - 1.96 \times se(\bar{y}), \bar{y} + 1.96 \times se(\bar{y})]$

And we know
$se(\bar{y}) \approx (0.0804928(1 - .0804928)/4870)^{\frac{1}{2}} = 0.003898446728$

Which is nearly the `Std. Err.` calculated by Stata.

After that, we may construct the C.I. under every given quantile.

# EE13.1 a conti.

Some may know that the command:

`ci proportion callback`

gives us a very similar result, and wonder the difference between the two.

Actually, you may see the text "Binomial Exact" in the latter command.

That is, Stata calculate the exact sampling distribution from "categories" we're interested in, instead of using the approximated normal distribution.

```
. ci proportion call_back
```

| Variable | Obs | Proportion | Std. Err. | — Binomial Exact — [95% Conf. | Interval] |
|---|---|---|---|---|---|
| call_back | 4,870 | .0804928 | .0038984 | .0730025 | .0884904 |

# EE13.1 a conti.

Also, if the variable is binary, and it is valued at 1 and 0, then the two command will yield the same result as the number of observations is large.

And the option `proportion` is designed for category variables. You may notice that there are Binomial distribution, Trinomial distribution, and Multinomial distribution.

For example,

$$(X, Y) \sim Trinomial(n, p_1, p_2)$$

$$f_{XY}(x, y) = \frac{n!}{x!y!(n-x-y)!} p_1^x p_2^y (1 - p_1 - p_2)^{n-x-y}$$

# EE13.1 b

Given the model:

$$call\_back = \beta_0 + \beta_1 black + \beta_2 female + \beta_3 blackfemale + u$$

If $black = 0$ & $female = 0$, then the effect is $\beta_0$

If $black = 1$ & $female = 0$, then the effect is $\beta_0 + \beta_1$

If $black = 0$ & $female = 1$, then the effect is $\beta_0 + \beta_2$

If $black = 1$ & $female = 1$, then the effect is $\beta_0 + \beta_1 + \beta_2 + \beta_3$

# EE13.1 c

Given the model:

$$call\_back = \beta_0 + \beta_1 black + \beta_2 high + \beta_3 blackhigh + u$$

If $black = 0$ & $high = 0$, then the effect is $\beta_0$

If $black = 1$ & $high = 0$, then the effect is $\beta_0 + \beta_1$

If $black = 0$ & $high = 1$, then the effect is $\beta_0 + \beta_2$

If $black = 1$ & $high = 1$, then the effect is $\beta_0 + \beta_1 + \beta_2 + \beta_3$

# EE13.1 Table

| VARIABLES | (1)<br>a | (2)<br>b | (3)<br>c |
|---|---|---|---|
| black | -0.0320*** | -0.0304* | -0.0231** |
| | (0.00778) | (0.0155) | (0.0106) |
| female | | 0.0102 | |
| | | (0.0137) | |
| blackFemale | | -0.00224 | |
| | | (0.0179) | |
| high | | | 0.0229* |
| | | | (0.0120) |
| blackHigh | | | -0.0178 |
| | | | (0.0156) |
| Constant | 0.0965*** | 0.0887*** | 0.0850*** |
| | (0.00599) | (0.0119) | (0.00801) |
| Observations | 4,870 | 4,870 | 4,870 |
| R-squared | 0.003 | 0.004 | 0.004 |

Robust standard errors in parentheses