

### 第3回連載 演習

#### 解説：標準化平均値差による群間比較

連載第2回の演習2では、シナリオごとに政策の対象者・非対象者間で所得、居住地、学習意欲の平均値を比較した。演習に取り組んだ読者は、いずれのシナリオにおいても群間でこれらの属性に差が生じていることが確認できたであろう。

注目している指標（ $X$ ）について群間における差の有無を確認する方法としては、平均値の差の検定（ $t$ 検定）がこれまで広く用いられてきた。平均値の差の検定を行う際の $t$ 統計量は以下のように表すことができる。

$$t_{\bar{x}_1 - \bar{x}_2} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

ここで、 $\bar{X}_1$ は対象者の標本平均値、 $s_1^2$ は対象者の不偏標本分散、 $n_1$ は対象者のサンプルサイズである。同様に、 $\bar{X}_2$ 、 $s_2^2$ 、 $n_2$ は非対象者の標本平均値、不偏標本分散、サンプルサイズを示している。平均値の差の検定では、この $t$ 統計量がある有意水準（例えば1%や5%など）のもとでの臨界値を超えるか否かを確認することで、群間で統計的に有意な差があるかどうかを判断する。

しかしながら近年、仮説検定によって群間差の有無を機械的に調べることに對して懸念が寄せられるようになってきている。その理由の一つが、 $t$ 統計量の分母にサンプルサイズが含まれていることである。サンプルサイズが小さい場合は平均値に実質的に大きな差があったとしても、仮説検定の結果は「統計的に有意な差があるとはいえない」という結論を導いてしまう。一方でサンプルサイズが大きい場合は、群間の平均値の差が小さかったとしても「統計的に有意な差がある」という結論を導いてしまう<sup>1</sup>。

こうした仮説検定の問題点を踏まえて、群間比較を行う際の代替的手法として標準化平均値差（standardized (mean) difference）を用いることが提唱されている。標準化平均値差

---

<sup>1</sup> 同様の理由によって、近年は $p$ 値の安易な利用にも警鐘が鳴らされるようになってきている。例えばアメリカ統計学会は“The ASA Statement on p-Values: Context, Process, and Purpose”という声明を出している（Wasserstein and Lazar 2016）。

は以下の様に定義される<sup>2</sup>。

$$SMD_{\bar{x}_1 - \bar{x}_2} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{(s_1^2 + s_2^2)/2}}$$

標準化平均値差はサンプルサイズに依存しないため、前述の懸念が発生しない。標準化平均値差については、例えば 0.1 を超えるかどうかで群間に差が生じているかを判断することが多い。

#### 参考文献

- Imbens, G. W., & Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.
- Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's statement on p-values: context, process, and purpose. *The American Statistician*, 70(2), 129-133.

---

<sup>2</sup> 標準化平均値差の詳細については、Imbens and Rubin (2015)などを参照。ただし同書では「normalized difference」と呼称している。

## 演習 1：政策の対象者・非対象者の所得、居住地、学習意欲の標準化平均値差による比較

連載第 3 回の表 2 では、ランダム割付を行った場合の政策対象者および非対象者の所得、居住地、学習意欲を比較しているが、シナリオ 1・2・3 について標準化平均値差を用いた群間比較を行ってみよう。

Stata を用いて標準化平均値差を計算するには、「stddiff」というコマンドを使うのが便利である。Stata に stdiff がインストールされていない場合は、「findit stdiff」とコマンドを打つことでインストールできる。

例えばシナリオ 1 における各属性の標準化平均値差は以下のように計算できる。

```
stddiff income area motivation, by(area)
```

実際に、このコマンドを入力すると以下のような結果が得られる。

|            | area=0    |           | area=1    |           |          |
|------------|-----------|-----------|-----------|-----------|----------|
|            | Mean or N | SD or (%) | Mean or N | SD or (%) | Std Diff |
| income     | 884.6     | 335.16    | 554.1     | 221.9     | 1.16298  |
| area       | 0         | 0         | 1         | 0         | .        |
| motivation | -.134     | 6.7524    | -.564     | 6.7511    | 0.06369  |

グループ（居住地区）ごとに平均値や標準偏差が計算され、一番右の列には標準化平均値差（Std Diff）が計算されている。シナリオ 1 ではそもそもグループごとに居住地区が異なっているため area に関する標準化平均値差は算出できないが、所得の標準化平均値差は 1.16 となっており大きな差が生じていることが確認できる。

### 演習課題 1

シナリオ 2 および 3 の標準化平均値差を計算してみよう。

## 演習 2：ランダム割付における所得、居住地、学習意欲の標準化平均値差による比較

### 演習課題 2

ランダム割付における標準化平均値差を計算してみよう。計算結果から、ランダム割付によって群間差がなくなっているとみなせるか考えてみよう。

### 演習 3：回帰分析による平均因果効果の推定

連載の第 2 節では回帰分析を行うことで平均因果効果の推定を試みている。Stata を用いて回帰分析を行うためには、「regress」というコマンドを使う。例えばシナリオ 1 について、試験結果を B 地区ダミーのみに回帰するには以下のようなコマンドを入力すれば良い。

```
regress outcome_area area
```

実際に、このコマンドを入力すると以下のような結果が得られる。

| Source   | SS        | df  | MS         | Number of obs = 1000 |   |        |
|----------|-----------|-----|------------|----------------------|---|--------|
| Model    | 9872.164  | 1   | 9872.164   | F( 1, 998)           | = | 165.61 |
| Residual | 59491.292 | 998 | 59.610513  | Prob > F             | = | 0.0000 |
| Total    | 69363.456 | 999 | 69.4328889 | R-squared            | = | 0.1423 |
|          |           |     |            | Adj R-squared        | = | 0.1415 |
|          |           |     |            | Root MSE             | = | 7.7208 |

  

| outcome_area | Coef.  | Std. Err. | t      | P> t  | [95% Conf. Interval] |          |
|--------------|--------|-----------|--------|-------|----------------------|----------|
| area         | 6.284  | .4883053  | 12.87  | 0.000 | 5.325777             | 7.242223 |
| _cons        | 53.746 | .345284   | 155.66 | 0.000 | 53.06843             | 54.42357 |

ここで B 地区ダミーの係数は 6.284 であるが、真の平均因果効果が 10.0 であるにも関わらず 3.7 点も過少評価されていることが確認できる。本文でも説明したように、これは群間で所得に偏りが生じているためである。その他の属性については群間に大きな違いはない。したがって所得を説明変数に加えることでその影響を統制できる。そのためには、以下のようなコマンドを入力すれば良い。

```
reg outcome_area area income
```

このコマンドを入力すると以下のような結果が得られる。

| Source   | SS         | df  | MS         |                 |        |  |
|----------|------------|-----|------------|-----------------|--------|--|
| Model    | 18401.0881 | 2   | 9200.54403 | Number of obs = | 1000   |  |
| Residual | 50962.3679 | 997 | 51.1157151 | F( 2, 997) =    | 179.99 |  |
| Total    | 69363.456  | 999 | 69.4328889 | Prob > F =      | 0.0000 |  |
|          |            |     |            | R-squared =     | 0.2653 |  |
|          |            |     |            | Adj R-squared = | 0.2638 |  |
|          |            |     |            | Root MSE =      | 7.1495 |  |

  

| outcome_area | Coef.    | Std. Err. | t     | P> t  | [95% Conf. Interval] |          |
|--------------|----------|-----------|-------|-------|----------------------|----------|
| area         | 9.683803 | .5231983  | 18.51 | 0.000 | 8.657107             | 10.7105  |
| income       | .0102852 | .0007962  | 12.92 | 0.000 | .0087227             | .0118477 |
| _cons        | 44.64754 | .7735385  | 57.72 | 0.000 | 43.12959             | 46.16549 |

この推定では B 地区ダミーの係数は 9.684 となっており、適切な変数（ここでは所得）を統計することによって、真の平均因果効果である 10.0 に近い値が得られたことが確認できる。