When Standard Methods Succeed

Lucy D’Agostino McGowan

Wake Forest University

2022-07-05 (updated: 2022-07-23)

when correlation **is** causation

When you have no confounders and there is a linear relationship between the exposure and the outcome, that **correlation is a causal relationship**

😮

When you have no confounders and there is a linear relationship between the exposure and the outcome, that correlation is a causal relationship

😮

When you have no confounders and there is a linear relationship between the exposure and the outcome, that correlation is a causal relationship

😮

When you have no confounders and there is a linear relationship between the exposure and the outcome, that correlation is a causal relationship

😮

randomized controlled trials

A/B testing

Even in these cases, using the methods you will learn here can help!

1. **Adjusting for baseline confounders can make an estimate more efficient**
2. **Propensity score weighting is more efficient that direct adjustment**
3. **Sometimes we are more comfortable with the functional form of the propensity score (predicting exposure) than the outcome model**

Example

simulated data (100 observations)

Example

simulated data (100 observations)

Treatment is **randomly** assigned

Example

simulated data (100 observations)

Treatment is **randomly** assigned

There are **two baseline covariates**: age **and** weight

Unadjusted model

```
lm(y ~ treatment, data = data)
```

Characteristic	Beta	SE[1]	95% CI[1]	p-value
treatment	1.6	0.803	-0.04, 3.1	0.056

[1] SE = Standard Error, CI = Confidence Interval

Adjusted model

```
lm(y ~ treatment + weight + age,
```

Characteristic	Beta	SE[1]	95% CI[1]	p-value
treatment	1.5	0.204	1.1, 1.9	<0.001
weight	0.18	0.103	-0.03, 0.38	0.087
age	0.20	0.005	0.19, 0.21	<0.001

[1] SE = Standard Error, CI = Confidence Interval

Propensity score adjusted model

Characteristic	Beta	SE	95% CI	p-value
treatment	1.5	0.197	1.1, 1.9	<0.001

Example

simulated data (10,000 observations)

Treatment is **randomly** assigned

There are **two baseline covariates**: age and weight

Unadjusted model

```
lm(y ~ treatment, data = data)
```

Characteristic	Beta	SE[1]	95% CI[1]	p-value
treatment	0.89	0.082	0.73, 1.1	<0.001

[1] SE = Standard Error, CI = Confidence Interval

Adjusted model

```
lm(y ~ treatment + weight + age,
```

Characteristic	Beta	SE[1]	95% CI[1]	p-value
treatment	1.0	0.020	1.0, 1.0	<0.001
weight	0.19	0.010	0.17, 0.21	<0.001
age	0.20	0.001	0.20, 0.20	<0.001

[1] SE = Standard Error, CI = Confidence Interval

Propensity score adjusted model

Characteristic	Beta	SE	95% CI	p-value
treatment	1	0.02	1, 1	<0.001

time-varying confounding