STA 235H - Potential Outcomes II

Fall 2022

McCombs School of Business, UT Austin

Housekeeping

- Homework 2 is due this Friday
    - Note on Chatter: In Task 2, use *price000s* from 2.2 onwards.
    - Send your questions until Friday 5.00pm.
- Remember to send re-grading requests for Homework 1 until Thursday
- No OH this Thursday (changed to Wed and Friday; check OH calendar).

Homework 3 will be posted this Friday

(Half the length of a homework)

Last week

- Finished our chapter on multiple regression.
  - How to handle binary outcomes: Linear Probability Models.
  - Video posted online about heteroskedasticity.
- Introduced Causal Inference

Today

- Continue with **causal inference**:
  - Potential outcomes
  - Ignorability assumption
- **Introduction to Randomized Controlled Trials**

Before we start, let's check this week's JITT

"You are trying to analyze what variables contribute to someone donating or not to a charity. You are running the following linear probability mode to analyze the association between different covariates and a binary response of whether someone responds with a charitable gift or not."

"The variables used in this model are the following:

- *respond*: Binary variable of whether the person responded with a gift (1) or not (0).
- *mailsyear*: number of mailings per year
- *propresp*: response rate to mailings. Continuous variable, measured from 0 to 1.
- *avggift*: average amount of past gifts (in US$)."

```
Call:
lm_robust(formula = respond ~ mailsyear + propresp + avggift,
    data = charity)

Standard error type:  HC2

Coefficients:
              Estimate Std. Error t value   Pr(>|t|)  CI Lower  CI Upper   DF
(Intercept) -0.1355169  2.294e-02  -5.908  3.742e-09 -1.805e-01 -0.090544 4264
mailsyear    0.0600391  1.064e-02   5.642  1.788e-08  3.918e-02  0.080901 4264
propresp     0.8449998  2.430e-02  34.771 1.939e-233  7.974e-01  0.892644 4264
avggift      0.0001718  5.826e-05   2.948  3.214e-03  5.754e-05  0.000286 4264

Multiple R-squared:  0.204 ,	Adjusted R-squared:  0.2034
F-statistic: 460.5 on 3 and 4264 DF,  p-value: < 2.2e-16
```

TRUE OR FALSE: "For one additional mailing a year, the probability of donating increases by 0.06%"

Before we start, let's check this week's JITT

"You are trying to analyze what variables contribute to someone donating or not to a charity. You are running the following linear probability mode to analyze the association between different covariates and a binary response of whether someone responds with a charitable gift or not."

"The variables used in this model are the following:

- *respond*: Binary variable of whether the person responded with a gift (1) or not (0).
- *mailsyear*: number of mailings per year
- *propresp*: response rate to mailings. Continuous variable, measured from 0 to 1.
- *avggift*: average amount of past gifts (in US$)."

```
Call:
lm_robust(formula = respond ~ mailsyear + propresp + avggift,
    data = charity)

Standard error type:  HC2

Coefficients:
              Estimate Std. Error t value   Pr(>|t|)   CI Lower  CI Upper   DF
(Intercept) -0.1355169  2.294e-02  -5.908  3.742e-09 -1.805e-01 -0.090544 4264
mailsyear    0.0600391  1.064e-02   5.642  1.788e-08  3.918e-02  0.080901 4264
propresp     0.8449998  2.430e-02  34.771 1.939e-233  7.974e-01  0.892644 4264
avggift      0.0001718  5.826e-05   2.948  3.214e-03  5.754e-05  0.000286 4264

Multiple R-squared:  0.204 ,    Adjusted R-squared:  0.2034
F-statistic: 460.5 on 3 and 4264 DF,  p-value: < 2.2e-16
```

TRUE OR FALSE: "For one additional dollar donated in the past, the probability of donating increases by 0.0002, on average, holding other variables constant."

Before we start, let's check this week's JITT

"You are trying to analyze what variables contribute to someone donating or not to a charity. You are running the following linear probability mode to analyze the association between different covariates and a binary response of whether someone responds with a charitable gift or not."

"The variables used in this model are the following:

- *respond*: Binary variable of whether the person responded with a gift (1) or not (0).
- *mailsyear*: number of mailings per year
- *propresp*: response rate to mailings. Continuous variable, measured from 0 to 1.
- *avggift*: average amount of past gifts (in US$)."

```
Call:
lm_robust(formula = respond ~ mailsyear + propresp + avggift,
    data = charity)

Standard error type:  HC2

Coefficients:
              Estimate Std. Error t value   Pr(>|t|)   CI Lower  CI Upper   DF
(Intercept) -0.1355169  2.294e-02  -5.908  3.742e-09 -1.805e-01 -0.090544 4264
mailsyear    0.0600391  1.064e-02   5.642  1.788e-08  3.918e-02  0.080901 4264
propresp     0.8449998  2.430e-02  34.771 1.939e-233  7.974e-01  0.892644 4264
avggift      0.0001718  5.826e-05   2.948  3.214e-03  5.754e-05  0.000286 4264

Multiple R-squared:  0.204 ,    Adjusted R-squared:  0.2034
F-statistic: 460.5 on 3 and 4264 DF,  p-value: < 2.2e-16
```

TRUE OR FALSE: "Holding mailings per year and average past gift constant, increasing the response rate of mailings from 0% to 100% is associated with an average 84 percentage point increase in the probability of donating."

Before we start, let's check this week's JITT

"You are trying to analyze what variables contribute to someone donating or not to a charity. You are running the following linear probability mode to analyze the association between different covariates and a binary response of whether someone responds with a charitable gift or not."

"The variables used in this model are the following:

- *respond*: Binary variable of whether the person responded with a gift (1) or not (0).
- *mailsyear*: number of mailings per year
- *propresp*: response rate to mailings. Continuous variable, measured from 0 to 1.
- *avggift*: average amount of past gifts (in US$)."

```
Call:
lm_robust(formula = respond ~ mailsyear + propresp + avggift,
    data = charity)

Standard error type:  HC2

Coefficients:
              Estimate Std. Error t value   Pr(>|t|)  CI Lower  CI Upper   DF
(Intercept) -0.1355169  2.294e-02  -5.908  3.742e-09 -1.805e-01 -0.090544 4264
mailsyear    0.0600391  1.064e-02   5.642  1.788e-08  3.918e-02  0.080901 4264
propresp     0.8449998  2.430e-02  34.771 1.939e-233  7.974e-01  0.892644 4264
avggift      0.0001718  5.826e-05   2.948  3.214e-03  5.754e-05  0.000286 4264

Multiple R-squared:  0.204 ,    Adjusted R-squared:  0.2034
F-statistic: 460.5 on 3 and 4264 DF,  p-value: < 2.2e-16
```

TRUE OR FALSE: "Holding mailings per year and average past gift constant, a 1% increase in the response rate of mailings is associated with an average 84% increase in the probability of donating."

Before we start, let's check this week's JITT

"You are trying to analyze what variables contribute to someone donating or not to a charity. You are running the following linear probability mode to analyze the association between different covariates and a binary response of whether someone responds with a charitable gift or not."

"The variables used in this model are the following:

- *respond*: Binary variable of whether the person responded with a gift (1) or not (0).
- *mailsyear*: number of mailings per year
- *propresp*: response rate to mailings. Continuous variable, measured from 0 to 1.
- *avggift*: average amount of past gifts (in US$)."

```
Call:
lm_robust(formula = respond ~ mailsyear + propresp + avggift,
    data = charity)

Standard error type:  HC2

Coefficients:
              Estimate Std. Error t value   Pr(>|t|)  CI Lower  CI Upper   DF
(Intercept) -0.1355169  2.294e-02  -5.908  3.742e-09 -1.805e-01 -0.090544 4264
mailsyear    0.0600391  1.064e-02   5.642  1.788e-08  3.918e-02  0.080901 4264
propresp     0.8449998  2.430e-02  34.771 1.939e-233  7.974e-01  0.892644 4264
avggift      0.0001718  5.826e-05   2.948  3.214e-03  5.754e-05  0.000286 4264

Multiple R-squared:  0.204 ,    Adjusted R-squared:  0.2034
F-statistic: 460.5 on 3 and 4264 DF,  p-value: < 2.2e-16
```

TRUE OR FALSE: "In this model, standard errors are larger than they would have been if we used lm() instead."

Causal Inference: Terminology and Notation

Potential Outcomes

- Last week we discussed potential outcomes, (e.g. $Y_i(1)$ and $Y_i(0)$):

"The outcome that *we would have observed* in the case of difference scenarios"

- Potential outcomes are related to your choices/possible conditions:
  - One for each path!
  - Do not confuse them with the **values** that your outcome variable can take.
- Definition of **Individual Causal Effect**:

$$ICE_i = Y_i(1) - Y_i(0)$$

What was the Fundamental Problem of Causal Inference?

Estimands vs Estimates vs Estimators

Estimand

A quantity we want to estimate

Estimator

A rule for calculating an estimate based on data

Estimate

The result of an estimation

Estimands vs Estimates vs Estimators

Estimand

A quantity we want to estimate

E.g.: Population mean

μ

Estimator

A rule for calculating an estimate based on data

E.g.: Sample mean

$$\frac{1}{n}\sum_i Y_i$$

Estimate

The result of an estimation

E.g.: Result of the sample mean for a given sample S

$\hat{\mu}$

Estimands vs Estimates vs Estimators

Source: Deng, 2022

Estimands vs Estimates vs Estimators

- Some important **estimands** that we need to keep in mind:

Average Treatment Effect (ATE)

Average Treatment Effect on the Treated (ATT)

Conditional Average Treatment Effect (CATE)

Estimands vs Estimates vs Estimators

- Some important **estimands** that we need to keep in mind:

$$\boxed{ATE = E[Y(1) - Y(0)]}$$

$$\boxed{ATT = E[Y(1) - Y(0)|Z = 1]}$$

$$\boxed{CATE = E[Y(1) - Y(0)|X]}$$

Getting around the fundamental problem of causal inference

- Let's go back to our original example: Does a pill help reduce headaches?

i	Z	Y	Y(1)	Y(0)	Y(1)-Y(0)
1	0	1	?	1	?
2	1	1	1	?	?
3	1	0	0	?	?
4	0	0	?	0	?
5	0	1	?	1	?
6	1	0	0	?	?

Getting around the fundamental problem of causal inference

- We have a **missing data problem**

i	Z	Y	Y(1)	Y(0)	Y(1)-Y(0)
1	0	1	?	1	?
2	1	1	1	?	?
3	1	0	0	?	?
4	0	0	?	0	?
5	0	1	?	1	?
6	1	0	0	?	?

Getting around the fundamental problem of causal inference

- Compare those who took the pill to the ones did not take it.

i	Z	Y	Y(1)	Y(0)	Y(1)-Y(0)
1	0	1	?	1	?
2	**1**	**1**	**1**	**?**	**?**
3	**1**	**0**	**0**	**?**	**?**
4	0	0	?	0	?
5	0	1	?	1	?
6	**1**	**0**	**0**	**?**	**?**

Getting around the fundamental problem of causal inference

- Compare those who took the pill to the ones did not take it.

i	Z	Y	Y(1)	Y(0)	Y(1)-Y(0)
1	0	1	?	1	?
2	1	1	1	?	?
3	1	0	0	?	?
4	0	0	?	0	?
5	0	1	?	1	?
6	1	0	0	?	?

Getting around the fundamental problem of causal inference

- Compare those who **took the pill** to the ones **did not take it.**

i	Z	Y	Y(1)	Y(0)	Y(1)-Y(0)
1	0	1	?	1	?
2	1	1	1	?	?
3	1	0	0	?	?
4	0	0	?	0	?
5	0	1	?	1	?
6	1	0	0	?	?

$$\hat{\tau} = \frac{1}{3}\Big(\sum_{i \in Z=1} Y_i - \sum_{i \in Z=0} Y_i\Big) = -0.333$$

Getting around the fundamental problem of causal inference

$$\hat{\tau} = \frac{1}{3}(\sum_{i \in Z=1} Y_i - \sum_{i \in Z=0} Y_i) = -0.333$$

- What is the **estimand**?

Getting around the fundamental problem of causal inference

$$\hat{\tau} = \frac{1}{3}(\sum_{i \in Z=1} Y_i - \sum_{i \in Z=0} Y_i) = -0.333$$

- What is the estimand?

Average Treatment Effect

Getting around the fundamental problem of causal inference

$$\hat{\tau} = \frac{1}{3}(\sum_{i \in Z=1} Y_i - \sum_{i \in Z=0} Y_i) = -0.333$$

- What is the **estimand**?

Average Treatment Effect

- What is the **estimator**?

Getting around the fundamental problem of causal inference

$$\hat{\tau} = \frac{1}{3}(\sum_{i \in Z=1} Y_i - \sum_{i \in Z=0} Y_i) = -0.333$$

- What is the **estimand**?

Average Treatment Effect

- What is the **estimator**?

Difference in sample means

Getting around the fundamental problem of causal inference

$$\hat{\tau} = \frac{1}{3}(\sum_{i \in Z=1} Y_i - \sum_{i \in Z=0} Y_i) = -0.333$$

- What is the estimand?

Average Treatment Effect

- What is the estimator?

Difference in sample means

- What is the estimate and *how do we interpret it*?

Getting around the fundamental problem of causal inference

$$\hat{\tau} = \frac{1}{3}(\sum_{i \in Z=1} Y_i - \sum_{i \in Z=0} Y_i) = -0.333$$

- What is the estimand?

Average Treatment Effect

- What is the estimator?

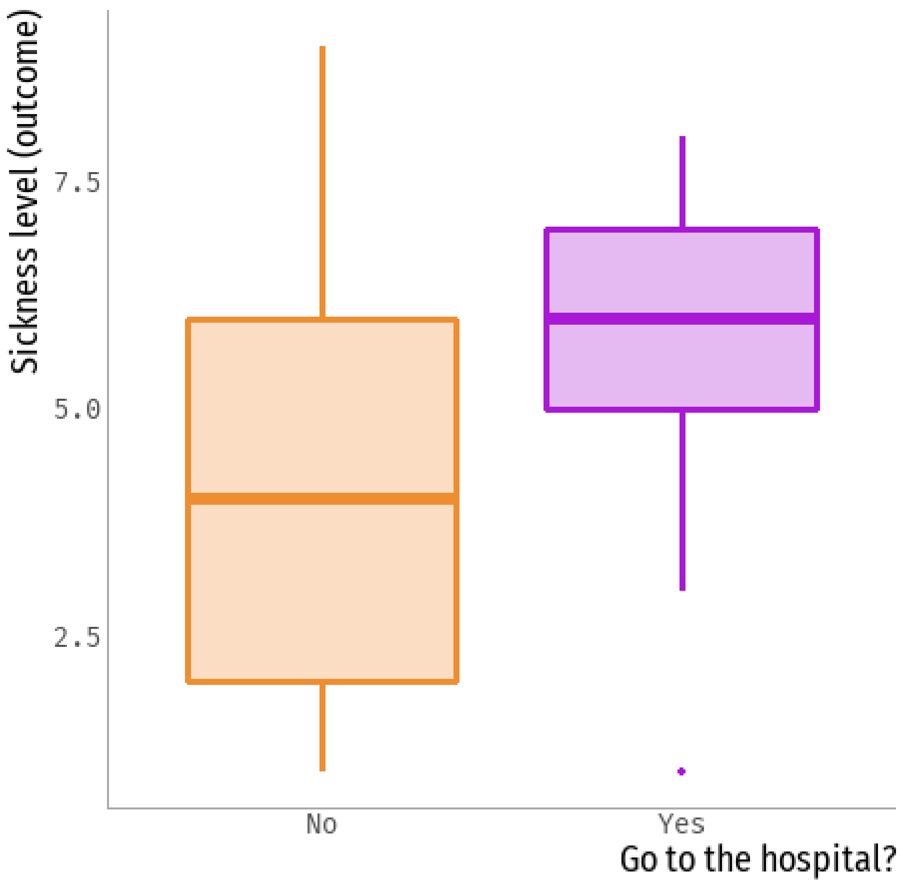
Difference in sample means

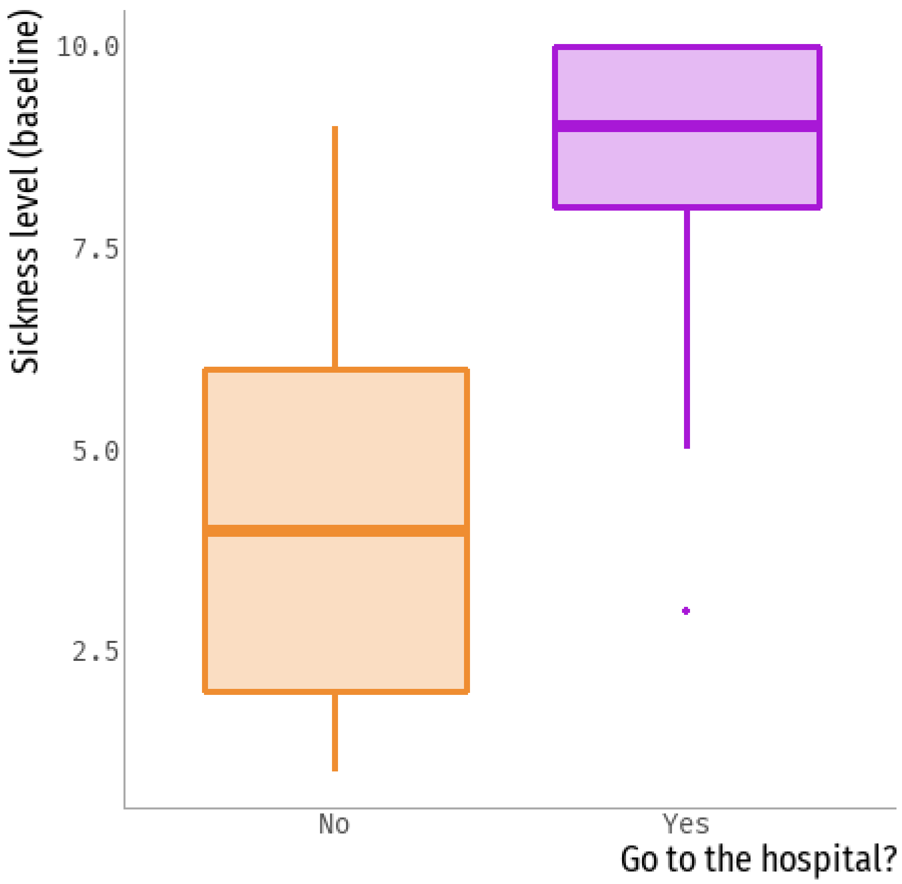
- What is the estimate and *how do we interpret it*?

33.3 percentage point decrease in probability of having a headache

What could be the problem with comparing the sample means?

Remember our exercise last week!





Under what assumptions is our estimate causal?

We are using:

$$\hat{\tau} = \frac{1}{3}(\sum_{i \in Z=1} Y_i - \sum_{i \in Z=0} Y_i)$$

to estimate:

$$\tau = E[Y_i(1) - Y_i(0)]$$

Under what assumptions is our estimate causal?

We are using:

$$\hat{\tau} = \frac{1}{3}\left(\sum_{i \in Z=1} Y_i - \sum_{i \in Z=0} Y_i\right)$$

to estimate:

$$\tau = E[Y_i(1) - Y_i(0)]$$

Let's do some math

Under what assumptions is our estimate causal?

$$\tau = E[Y_i(1) - Y_i(0)]$$

$$= E[Y_i(1)] - E[Y_i(0)]$$

Key assumption:

Ignorability

- Ignorability means that the potential outcomes $Y(0)$ and $Y(1)$ are independent of the treatment, e.g. $(Y(0), Y(1)) \perp\!\!\!\perp Z$.

Under what assumptions is our estimate causal?

$$\tau = E[Y_i(1) - Y_i(0)]$$

$$= E[Y_i(1)] - E[Y_i(0)]$$

Key assumption:

Ignorability

- Ignorability means that the potential outcomes $Y(0)$ and $Y(1)$ are independent of the treatment, e.g. $(Y(0), Y(1)) \perp\!\!\!\perp Z$.
    - *Remember that if* $A \perp\!\!\!\perp B \ \rightarrow \ E[A|B] = E[A]$

Under what assumptions is our estimate causal?

$$\tau = E[Y_i(1) - Y_i(0)]$$

$$= E[Y_i(1)] - E[Y_i(0)]$$

Key assumption:

Ignorability

- Ignorability means that the potential outcomes $Y(0)$ and $Y(1)$ are independent of the treatment, e.g. $(Y(0), Y(1)) \perp\!\!\!\perp Z$.

$$\tau = E[Y_i(1)] - E[Y_i(0)] = E[Y_i(1)|Z = 1] - E[Y_i(0)|Z = 0]$$

Under what assumptions is our estimate causal?

$$\tau = E[Y_i(1) - Y_i(0)]$$

$$= E[Y_i(1)] - E[Y_i(0)]$$

Key assumption:

Ignorability

- Ignorability means that the potential outcomes $Y(0)$ and $Y(1)$ are independent of the treatment, e.g. $(Y(0), Y(1)) \perp\!\!\!\perp Z$.

$$\tau = E[Y_i(1)] - E[Y_i(0)] = \underbrace{E[Y_i(1)|Z = 1]}_{\text{Obs. Outcome for T}} - \overbrace{E[Y_i(0)|Z = 0]}^{\text{Obs. Outcome for C}}$$

Under what assumptions is our estimate causal?

$$\tau = E[Y_i(1) - Y_i(0)]$$

$$= E[Y_i(1)] - E[Y_i(0)]$$

Key assumption:

Ignorability

- Ignorability means that the potential outcomes $Y(0)$ and $Y(1)$ are independent of the treatment, e.g. $(Y(0), Y(1)) \perp\!\!\!\perp Z$.

$$\tau = E[Y_i(1)] - E[Y_i(0)] = E[Y_i(1)|Z=1] - E[Y_i(0)|Z=0] =$$

$$= E[Y_i|Z=1] - E[Y_i|Z=0]$$

References

- Angrist, J. & S. Pischke. (2015). "Mastering Metrics". *Chapter 1*.
- Cunningham, S. (2021). "Causal Inference: The Mixtape". *Chapter 4: Potential Outcomes Causal Model*.
- Neil, B. (2020). "Introduction to Causal Inference". *Fall 2020 Course*