4. Linear Regression and Randomized Experiments

ISS5096 || ECI

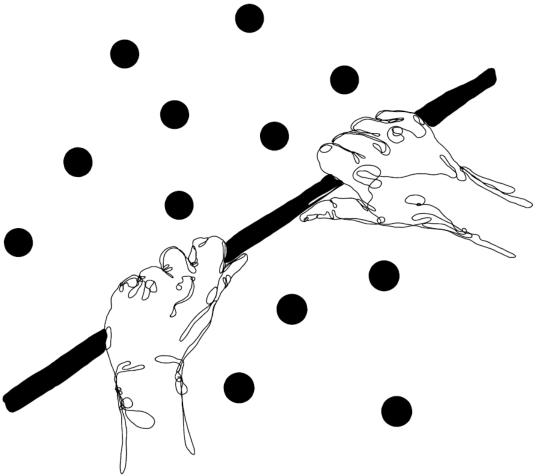
Jaewon (“Jay-one”) Yoo
National Tsing Hua University

Where are we? Where are we going?

- So far: analysis of experiments with Fisher's and Neyman's approaches.
  - Neyman: Unbiased estimators, (conservative) variances.
  - Fisher: exact test of the sharp null.
- Today: how does the workhorse estimator, OLS, fit into this story?

Where are we? Where are we going?

- So far: analysis of experiments with Fisher's and Neyman's approaches.
  - Neyman: Unbiased estimators, (conservative) variances.
  - Fisher: exact test of the sharp null.
- Today: how does the workhorse estimator, OLS, fit into this story?
- Why would we consider using regression?
  - **Simplicity:** known tool that is already very common.
  - **Increased precision:** we may want to add covariates for more precise effect estimates.



Source: *Chapter 13* of The Effect (Textbook 2) by Nick Huntington-Klein

1/ Regression with no covariates

Analysing Experiments with Regression?

- Q: Under complete randomization, can we use OLS to estimate ATEs
    - Literally, just `lm(y ~ d)`?
- Recall that the OLS estimator solves the least squares problem:

$$(\widehat{\tau}_{\text{ols}}, \widehat{\alpha}_{\text{ols}}) = \arg\min_{\tau,\alpha} \sum_{i=1}^{n} (Y_i - \alpha - \tau D_i)^2 \tag{1}$$

Analysing Experiments with Regression?

- Q: Under complete randomization, can we use OLS to estimate ATEs
    - Literally, just `lm(y ~ d)`?
- Recall that the OLS estimator solves the least squares problem:

$$(\widehat{\tau}_{\text{ols}}, \widehat{\alpha}_{\text{ols}}) = \arg\min_{\tau,\alpha} \sum_{i=1}^{n} (Y_i - \alpha - \tau D_i)^2 \tag{1}$$

- The coefficient on a binary r.v. is mechanically the diff. in means:

$$\widehat{\tau}_{\text{ols}} = \overline{Y}_1 - \overline{Y}_0 = \widehat{\tau}_{\text{diff}} \tag{2}$$

    - Standard Neyman analysis for unbiasedness, sampling variance.
- Generalized to discrete treatments with > 2 levels.

Justifying the Linear Model

- Mechanically the same, but can we justify the linear model itself?
    - Key assumptions: **linearity** and **mean independence of errors**.
- Some simple manipulation of the consistency assumption:

$$Y_i =$$

$$= \alpha + D_i \tau + \varepsilon_i$$

Justifying the Linear Model

- Mechanically the same, but can we justify the linear model itself?
  - Key assumptions: **linearity** and **mean independence of errors**.
- Some simple manipulation of the consistency assumption:

$$Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0)$$

$$= \alpha + D_i \tau + \varepsilon_i$$

Justifying the Linear Model

- Mechanically the same, but can we justify the linear model itself?
  - Key assumptions: **linearity** and **mean independence of errors**.
- Some simple manipulation of the consistency assumption:

$$
\begin{aligned}
Y_i &= D_i Y_i(1) + (1 - D_i) Y_i(0) \\
&= Y_i(0) + D_i \{ Y_i(1) - Y_i(0) \} \\
&= \alpha + D_i \tau + \varepsilon_i
\end{aligned}
$$

Justifying the Linear Model

- Mechanically the same, but can we justify the linear model itself?
    - Key assumptions: **linearity** and **mean independence of errors**.
- Some simple manipulation of the consistency assumption:

$$\begin{aligned} Y_i &= D_i Y_i(1) + (1 - D_i) Y_i(0) \\ &= Y_i(0) + D_i \{ Y_i(1) - Y_i(0) \} \\ &= Y_i(0) + D_i \tau_i \end{aligned}$$

$$= \alpha + D_i \tau + \varepsilon_i$$

Justifying the Linear Model

- Mechanically the same, but can we justify the linear model itself?
  - Key assumptions: **linearity** and **mean independence of errors**.
- Some simple manipulation of the consistency assumption:

$$
\begin{aligned}
Y_i &= D_i Y_i(1) + (1 - D_i) Y_i(0) \\
&= Y_i(0) + D_i \{Y_i(1) - Y_i(0)\} \\
&= Y_i(0) + D_i \tau_i \\
&= \mathbb{E}[Y_i(0)] + D_i \tau + \{Y_i(0) - \mathbb{E}[Y_i(0)]\} + D_i(\tau_i - \tau) \\
&= \alpha + D_i \tau + \varepsilon_i
\end{aligned}
$$

- "Linear" functional form fully justified by consistency alone with:
  - Intercept $\alpha = \mathbb{E}[Y_i(0)]$ is the average control outcome.
  - Slope $\tau = \mathbb{E}[Y_i(1) - Y_i(0)]$ is the PATE.
  - Error is deviation for control PO + treatment effect heterogeneity.

Mean Independent Errors

$$\varepsilon_i = (Y_i(0) - \mathbb{E}[Y_i(0)]) + D_i \cdot (\tau_i - \tau)$$

- What about mean independence: $\mathbb{E}[\varepsilon_i | D_i] = 0$?

Mean Independent Errors

$$\varepsilon_i = (Y_i(0) - \mathbb{E}[Y_i(0)]) + D_i \cdot (\tau_i - \tau)$$

- What about mean independence: $\mathbb{E}[\varepsilon_i | D_i] = 0$?
- Using randomization, we know D_i is independent of $Y_i(1), Y_i(0)$, so:

$$\mathbb{E}[\varepsilon_i | D_i] =$$

$$= 0$$

Mean Independent Errors

$$\varepsilon_i = (Y_i(0) - \mathbb{E}[Y_i(0)]) + D_i \cdot (\tau_i - \tau)$$

- What about mean independence: $\mathbb{E}[\varepsilon_i | D_i] = 0$?
- Using randomization, we know D_i is independent of $Y_i(1), Y_i(0)$, so:

$$\mathbb{E}[\varepsilon_i | D_i] = \mathbb{E}\left[\{Y_i(0) - \mathbb{E}[Y_i(0)]\} + D_i \cdot (\tau_i - \tau) | D_i\right]$$

$$= 0$$

Mean Independent Errors

$$\varepsilon_i = (Y_i(0) - \mathbb{E}[Y_i(0)]) + D_i \cdot (\tau_i - \tau)$$

- What about mean independence: $\mathbb{E}[\varepsilon_i | D_i] = 0$?
- Using randomization, we know D_i is independent of $Y_i(1), Y_i(0)$, so:

$$\begin{aligned}
\mathbb{E}[\varepsilon_i | D_i] &= \mathbb{E}\left[\{Y_i(0) - \mathbb{E}[Y_i(0)]\} + D_i \cdot (\tau_i - \tau) | D_i\right] \\
&= \mathbb{E}[Y_i(0) | D_i] - \mathbb{E}[Y_i(0)] + D_i(\mathbb{E}[\tau_i | D_i] - \tau) \\
\\
&= 0
\end{aligned}$$

Mean Independent Errors

$$\varepsilon_i = (Y_i(0) - \mathbb{E}[Y_i(0)]) + D_i \cdot (\tau_i - \tau)$$

- What about mean independence: $\mathbb{E}[\varepsilon_i | D_i] = 0$?
- Using randomization, we know D_i is independent of $Y_i(1), Y_i(0)$, so:

$$\begin{aligned}
\mathbb{E}[\varepsilon_i | D_i] &= \mathbb{E}\left[\{Y_i(0) - \mathbb{E}[Y_i(0)]\} + D_i \cdot (\tau_i - \tau) | D_i\right] \\
&= \mathbb{E}[Y_i(0) | D_i] - \mathbb{E}[Y_i(0)] + D_i(\mathbb{E}[\tau_i | D_i] - \tau) \\
&= \mathbb{E}[Y_i(0)] - \mathbb{E}[Y_i(0)] + D_i \underbrace{(\mathbb{E}[\tau_i] - \tau)}_{\tau = \mathbb{E}[\tau_i]} \\
&= 0
\end{aligned}$$

- Randomization + consistency $\rightsquigarrow$ linear model.
    - Does not imply homoskedasticity or normal errors, though!

Homoskedasticity

- Software default assumption: **Homoskedasticity**

$$\mathbb{V}[\varepsilon_i \mid \mathbf{D}] = \sigma^2, \quad \forall i$$

Homoskedasticity

- Software default assumption: **Homoskedasticity**

$$\mathbb{V}[\varepsilon_i \mid \mathbf{D}] = \sigma^2, \quad \forall i$$

- But in general, based on previous error definition:

$$\mathbb{V}[\varepsilon_i \mid \mathbf{D}] = \mathbb{V}[\varepsilon_i \mid D_i] = D_i\sigma_1^2 + (1 - D_i)\sigma_0^2$$

  - $\rightsquigarrow$ homoskedasticity true when $\sigma_1^2 = \mathbb{V}[Y_i(1)] = \mathbb{V}[Y_i(0)] = \sigma_0^2$
  - True under constant treatment effects!

Homoskedasticity

- Software default assumption: **Homoskedasticity**

$$\mathbb{V}[\varepsilon_i \mid \mathbf{D}] = \sigma^2, \quad \forall i$$

- But in general, based on previous error definition:

$$\mathbb{V}[\varepsilon_i \mid \mathbf{D}] = \mathbb{V}[\varepsilon_i \mid D_i] = D_i \sigma_1^2 + (1 - D_i)\sigma_0^2$$

  - $\rightsquigarrow$ homoskedasticity true when $\sigma_1^2 = \mathbb{V}[Y_i(1)] = \mathbb{V}[Y_i(0)] = \sigma_0^2$
  - True under constant treatment effects!

- Under homoskedasticity, variance of the OLS estimator is:

$$\mathbb{V}[\widehat{\tau}_{\text{ols}} \mid \mathbf{D}] = \frac{\sigma^2}{\sum_{i=1}^n (D_i - \overline{D})^2}$$

Variance Estimation

- "Standard" variance estimator under homoskedasticity:

$$\widehat{\mathbb{V}}_{\text{const}} = \frac{\frac{1}{n-2}\sum_{i=1}^{n}\hat{\varepsilon}_i^2}{\sum_{i=1}^{n}(D_i - \overline{D})^2} = \frac{\frac{1}{n-2}\sum_{i=1}^{n}(Y_i - \widehat{\alpha}_{\text{ols}} - \widehat{\tau}_{\text{ols}}D_i)^2}{\sum_{i=1}^{n}(D_i - \overline{D})^2}$$

Variance Estimation

- "Standard" variance estimator under homoskedasticity:

$$\widehat{\mathbb{V}}_{\text{const}} = \frac{\frac{1}{n-2}\sum_{i=1}^{n}\hat{\varepsilon}_i^2}{\sum_{i=1}^{n}(D_i - \overline{D})^2} = \frac{\frac{1}{n-2}\sum_{i=1}^{n}(Y_i - \widehat{\alpha}_{\text{ols}} - \widehat{\tau}_{\text{ols}} D_i)^2}{\sum_{i=1}^{n}(D_i - \overline{D})^2}$$

- We can rewrite this as a function of the **pooled** variance $\widehat{\sigma}^2_{Y|D}$:

$$\widehat{\mathbb{V}}_{\text{const}} = \widehat{\sigma}^2_{Y|D}\left(\frac{1}{n_0} + \frac{1}{n_1}\right)$$

$$\widehat{\sigma}^2_{Y|D} = \frac{1}{n-2}\left(\sum_{i:D_i=0}(Y_i - \overline{Y}_0)^2 + \sum_{i:D_i=1}(Y_i - \overline{Y}_1)^2\right)$$

Variance Estimation

- "Standard" variance estimator under homoskedasticity:

$$\widehat{\mathbb{V}}_{\text{const}} = \frac{\frac{1}{n-2}\sum_{i=1}^{n}\hat{\varepsilon}_i^2}{\sum_{i=1}^{n}(D_i - \overline{D})^2} = \frac{\frac{1}{n-2}\sum_{i=1}^{n}(Y_i - \widehat{\alpha}_{\text{ols}} - \widehat{\tau}_{\text{ols}} D_i)^2}{\sum_{i=1}^{n}(D_i - \overline{D})^2}$$

- We can rewrite this as a function of the **pooled** variance $\widehat{\sigma}^2_{Y|D}$:

$$\widehat{\mathbb{V}}_{\text{const}} = \widehat{\sigma}^2_{Y|D}\left(\frac{1}{n_0} + \frac{1}{n_1}\right)$$

$$\widehat{\sigma}^2_{Y|D} = \frac{1}{n-2}\left(\sum_{i:D_i=0}(Y_i - \overline{Y}_0)^2 + \sum_{i:D_i=1}(Y_i - \overline{Y}_1)^2\right)$$

- **Inconsistent**: $\widehat{\mathbb{V}}_{\text{const}} - \mathbb{V}[\widehat{\tau}] \xrightarrow{p} c \neq 0$ unless
  - Homoskedasticity holds: $\sigma_1^2 = \sigma_0^2$
  - Design is balanced: $n_1 = n_0$

Variance Estimation

- Can we use "standard" variance estimator: $\mathbb{V}[\varepsilon_i|\mathbf{D}] = \sigma^2$
  - Inconsistent: $\widehat{\mathbb{V}}_{\text{const}} - \mathbb{V}[\widehat{\tau}] \xrightarrow{p} c \neq 0$

Variance Estimation

- Can we use "standard" variance estimator: $\mathbb{V}[\varepsilon_i | \mathbf{D}] = \sigma^2$
  - Inconsistent: $\widehat{\mathbb{V}}_{\text{const}} - \mathbb{V}[\widehat{\tau}] \xrightarrow{p} c \neq 0$
  - Bias:

$$
\begin{aligned}
&\mathbb{E}(\widehat{\mathbb{V}}_{\text{const}}) - \mathbb{V}[\widehat{\tau}] \\
&= \underbrace{\mathbb{E}\left(\frac{\frac{1}{n-2}\sum_{i=1}^{n}\hat{\varepsilon}_i^2}{\sum_{i=1}^{n}(D_i - \overline{D})^2}\right)}_{\text{under const. effect assumption}} - \underbrace{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_0^2}{n_0}\right)}_{\text{true variance}} \\
&= \frac{(n_1 - n_0)(n-1)}{n_1 n_0 (n-2)}(\sigma_1^2 - \sigma_0^2) \neq 0
\end{aligned}
$$

Variance Estimation

- Can we use "standard" variance estimator: $\mathbb{V}[\varepsilon_i|\mathbf{D}] = \sigma^2$
  - Inconsistent: $\widehat{\mathbb{V}}_{\text{const}} - \mathbb{V}[\widehat{\tau}] \xrightarrow{p} c \neq 0$
  - Bias:

$$\begin{aligned}
&\mathbb{E}(\widehat{\mathbb{V}}_{\text{const}}) - \mathbb{V}[\widehat{\tau}] \\
&= \underbrace{\mathbb{E}\left(\frac{\frac{1}{n-2}\sum_{i=1}^{n}\hat{\varepsilon}_i^2}{\sum_{i=1}^{n}(D_i - \overline{D})^2}\right)}_{\text{under const. effect assumption}} - \underbrace{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_0^2}{n_0}\right)}_{\text{true variance}} \\
&= \frac{(n_1 - n_0)(n-1)}{n_1 n_0 (n-2)}(\sigma_1^2 - \sigma_0^2) \neq 0
\end{aligned}$$

- Unless:
  - Homoskedasticity holds: $\sigma_1^2 = \sigma_0^2$
    - Constant effect: $Y_i(1) - Y_i(0) = \text{const.}$
    - $\mathbb{V}[Y_i(1)] = \mathbb{V}[Y_i(0) + \text{const.}] = \mathbb{V}[Y_i(0)]$
  - Design is balanced: $n_1 = n_0$

Robust SEs

- Eicker-Huber-White (EHW) robust/sandwich variance estimator:

$$\underbrace{\widehat{\mathbb{V}}_{\text{EHW}}}_{\text{sandwich}} = \underbrace{\left(\sum_{i=1}^{n} \mathbf{X}_i\mathbf{X}_i'\right)^{-1}}_{\text{bread}} \underbrace{\left(\sum_{i=1}^{n} \hat{\varepsilon}_i^2\mathbf{X}_i\mathbf{X}_i'\right)}_{\text{meat}} \underbrace{\left(\sum_{i=1}^{n} \mathbf{X}_i\mathbf{X}_i'\right)^{-1}}_{\text{bread}}$$

$$= (\mathbb{X}'\mathbb{X})^{-1}\left(\sum_{i=1}^{n} \hat{\varepsilon}_i^2\mathbf{X}_i\mathbf{X}_i'\right)(\mathbb{X}'\mathbb{X})^{-1} \quad \text{where} \quad \mathbb{X} = [1 \quad \mathbf{D}]$$

Robust SEs

- Eicker-Huber-White (EHW) robust/sandwich variance estimator:

$$\underbrace{\widehat{\mathbb{V}}_{\text{EHW}}}_{\text{sandwich}} = \underbrace{\left(\sum_{i=1}^{n} \mathbf{X}_i \mathbf{X}_i'\right)^{-1}}_{\text{bread}} \underbrace{\left(\sum_{i=1}^{n} \hat{\varepsilon}_i^2 \mathbf{X}_i \mathbf{X}_i'\right)}_{\text{meat}} \underbrace{\left(\sum_{i=1}^{n} \mathbf{X}_i \mathbf{X}_i'\right)^{-1}}_{\text{bread}}$$

$$= (\mathbb{X}'\mathbb{X})^{-1} \left(\sum_{i=1}^{n} \hat{\varepsilon}_i^2 \mathbf{X}_i \mathbf{X}_i'\right) (\mathbb{X}'\mathbb{X})^{-1} \quad \text{where} \quad \mathbb{X} = [1 \quad \mathbf{D}]$$

- Recall the PATE-targeted variance of the difference-in-means:

$$\mathbb{V}(\widehat{\tau}_{\text{diff}}) = \frac{\sigma_0^2}{n_0} + \frac{\sigma_1^2}{n_1} = \frac{\mathbb{V}[Y_i(0)]}{n_0} + \frac{\mathbb{V}[Y_i(1)]}{n_1}$$

Robust SEs

- Eicker-Huber-White (EHW) robust/sandwich variance estimator:

$$\underbrace{\widehat{\mathbb{V}}_{\text{EHW}}}_{\text{sandwich}} = \underbrace{\left(\sum_{i=1}^{n} \mathbf{X}_i \mathbf{X}_i'\right)^{-1}}_{\text{bread}} \underbrace{\left(\sum_{i=1}^{n} \hat{\varepsilon}_i^2 \mathbf{X}_i \mathbf{X}_i'\right)}_{\text{meat}} \underbrace{\left(\sum_{i=1}^{n} \mathbf{X}_i \mathbf{X}_i'\right)^{-1}}_{\text{bread}}$$

$$= (\mathbb{X}'\mathbb{X})^{-1} \left(\sum_{i=1}^{n} \hat{\varepsilon}_i^2 \mathbf{X}_i \mathbf{X}_i'\right) (\mathbb{X}'\mathbb{X})^{-1} \quad \text{where} \quad \mathbb{X} = [1 \quad \mathbf{D}]$$

- Recall the PATE-targeted variance of the difference-in-means:

$$\mathbb{V}(\widehat{\tau}_{\text{diff}}) = \frac{\sigma_0^2}{n_0} + \frac{\sigma_1^2}{n_1} = \frac{\mathbb{V}[Y_i(0)]}{n_0} + \frac{\mathbb{V}[Y_i(1)]}{n_1}$$

- To see this, we can derive $\widehat{\mathbb{V}}_{\text{EHW}}$ under our case:

$$\widehat{\mathbb{V}}_{\text{EHW}} = \frac{\widetilde{\sigma}_1^2}{n_1} + \frac{\widetilde{\sigma}_0^2}{n_0}, \quad \text{where} \quad \widetilde{\sigma}_d^2 = \frac{1}{n_d} \sum_{i:D_i=d} (Y_i - \overline{Y}_d)^2$$

  - $\widetilde{\sigma}_0^2, \widetilde{\sigma}_1^2$ consistent for $\sigma_0^2, \sigma_1^2 \rightsquigarrow \widehat{\mathbb{V}}_{\text{EHW}}$ consistent for $\mathbb{V}(\widehat{\tau}_{\text{diff}})$

Better Robust SEs

- Many different "improved" versions of robust variances proposed.
  - Almost all are "finite-sample corrections" (no asympotic effects)

Better Robust SEs

- Many different "improved" versions of robust variances proposed.
  - Almost all are "finite-sample corrections" (no asympotic effects)

- HC2 estimator normalizes residuals by the leverage, h_{ii}:

$$\mathbb{V}_{\text{HC2}} = (\mathbb{X}'\mathbb{X})^{-1} \left(\sum_{i=1}^{n} \frac{\hat{\varepsilon}_i^2}{1 - h_{ii}} \mathbf{X}_i \mathbf{X}_i' \right) (\mathbb{X}'\mathbb{X})^{-1}$$

  - Leverage: $h_{ii} = \mathbf{X}_i (\mathbb{X}'\mathbb{X})^{-1} \mathbf{X}_i'$
  - In this setting: $h_{ii} = n_1^{-1}$ if $D_i = 1$ and $h_{ii} = n_0^{-1}$ if $D_i = 0$

Better Robust SEs

- Many different "improved" versions of robust variances proposed.
  - Almost all are "finite-sample corrections" (no asympotic effects)

- HC2 estimator normalizes residuals by the leverage, h_{ii}:

$$\mathbb{V}_{\text{HC2}} = (\mathbb{X}'\mathbb{X})^{-1} \left(\sum_{i=1}^{n} \frac{\hat{\varepsilon}_i^2}{1 - h_{ii}} \mathbf{X}_i \mathbf{X}_i' \right) (\mathbb{X}'\mathbb{X})^{-1}$$

  - Leverage: $h_{ii} = \mathbf{X}_i (\mathbb{X}'\mathbb{X})^{-1} \mathbf{X}_i'$
  - In this setting: $h_{ii} = n_1^{-1}$ if $D_i = 1$ and $h_{ii} = n_0^{-1}$ if $D_i = 0$

- Samii & Aronow (2012): HC2 is exactly the Neyman variance estimator:

$$\widehat{\mathbb{V}}_{\text{HC2}} = \frac{\widehat{\sigma}_0^2}{n_0} + \frac{\widehat{\sigma}_1^2}{n_1}$$

- $\rightsquigarrow$ simple OLS + HC2 = unbiased point and variance estimator.

Application in R Using (Sim) AI Exp. Data

```
> AI_data <- as_tibble(read.csv(url("https://bit.ly/3FHsusw"))); AI_data
# A tibble: 500 × 8
   treat_ind test_outcome_pre test_outcome_post student_age student_gender tutor_age
       <int>            <int>             <int>       <int>          <int>     <int>
 1         1                0                 1          11              0        27
 2         1                0                 1           6              1        45
 3         0                1                 1          15              1        54
 4         0                1                 1           6              1        50
 5         0                1                 0           8              1        64
 6         1                1                 1          15              1        43
 7         1                0                 1           7              1        47
 8         1                1                 1          16              0        49
 9         0                0                 1          12              1        42
10         1                1                 1          11              1        28
#  490 more rows
#  2 more variables: years_of_experience <int>, education_level <chr>
#  Use `print(n = ...)` to see more rows

> lm1 <- lm(test_outcome_post ~ treat_ind, data = AI_data)
> vcovM <- sandwich::vcovHC(lm1, type = 'HC2')
> sqrt(vcovM[1,1])
[1] 0.02955412
> sqrt(vcovM[2,2]) # sqrt(diag(vcovM))
[1] 0.04209018

> # Or
> estimatr::lm_robust(test_outcome_post ~ treat_ind, AI_data, se_type = 'HC2')
              Estimate Std. Error   t value     Pr(>|t|)    CI Lower  CI Upper  DF
(Intercept) 0.65134100 0.02955412 22.038927 1.233662e-75  0.59327487 0.7094071 498
treat_ind   0.03903557 0.04209018  0.927427 3.541541e-01 -0.04366065 0.1217318 498
```

2/ Linear regression with covariates

Adding Covariates

- What if we add covariates to our regression model?

$$(\widehat{\tau}_{\text{adj}}, \widehat{\alpha}_{\text{adj}}, \widehat{\beta}_{\text{adj}}) = \underset{\tau,\alpha,\beta}{\arg\min} \sum_{i=1}^{n} (Y_i - \alpha - \tau D_i - \widetilde{\mathbf{X}}_i'\beta)^2$$

  - $\widetilde{\mathbf{X}}_i = \mathbf{X}_i - \overline{\mathbf{X}}$ are **centered** covariates for notational ease.

Adding Covariates

- What if we add covariates to our regression model?

$$(\widehat{\tau}_{\text{adj}}, \widehat{\alpha}_{\text{adj}}, \widehat{\beta}_{\text{adj}}) = \underset{\tau,\alpha,\beta}{\arg\min} \sum_{i=1}^{n} (Y_i - \alpha - \tau D_i - \widetilde{\mathbf{X}}_i'\beta)^2$$

  - $\widetilde{\mathbf{X}}_i = \mathbf{X}_i - \overline{\mathbf{X}}$ are **centered** covariates for notational ease.

- Why might we do this? To increase **precision** of our estimates.
  - We hope $\mathbb{V}[\widehat{\tau}_{\text{adj}}] < \mathbb{V}[\widehat{\tau}_{\text{diff}}]$ so we have smaller CIs, more powerful tests.
  - Intuition: less residual variation in Y_i after accounting for $\mathbf{X}_i$

Adding Covariates

- What if we add covariates to our regression model?

$$(\widehat{\tau}_{\text{adj}}, \widehat{\alpha}_{\text{adj}}, \widehat{\beta}_{\text{adj}}) = \underset{\tau,\alpha,\beta}{\arg\min} \sum_{i=1}^{n} (Y_i - \alpha - \tau D_i - \widetilde{\mathbf{X}}_i' \beta)^2$$

  - $\widetilde{\mathbf{X}}_i = \mathbf{X}_i - \overline{\mathbf{X}}$ are **centered** covariates for notational ease.

- Why might we do this? To increase **precision** of our estimates.
  - We hope $\mathbb{V}[\widehat{\tau}_{\text{adj}}] < \mathbb{V}[\widehat{\tau}_{\text{diff}}]$ so we have smaller CIs, more powerful tests.
  - Intuition: less residual variation in Y_i after accounting for $\mathbf{X}_i$

- Questions:
  - Is $\widehat{\tau}$ still unbiased? Consistent?
  - Should we expect an increase in precision?
  - Controversial! Freedman (2008) "Randomization does not justify the regression model"

OLS is biased, but consistent (Freedman, 2008. Adv. in Appl. Math)

- Agnostic approach: don't assume correctness of the linear model.
    - $\mathbf{X}_i'\beta$ could be badly misspecified.
    - No assumptions about homoskedasticity.

OLS is biased, but consistent (Freedman, 2008. Adv. in Appl. Math)

- Agnostic approach: don't assume correctness of the linear model.
  - $\mathbf{X}_i'\beta$ could be badly misspecified.
  - No assumptions about homoskedasticity.

- Under minimal assumptions, OLS is consistent for the linear projection

$$(\widehat{\tau}_{\text{adj}}, \widehat{\alpha}_{\text{adj}}, \widehat{\beta}_{\text{adj}}) \xrightarrow{p} (\tau_0, \alpha_0, \beta_0) = \underset{(\tau,\alpha,\beta)}{\arg\min}\, \mathbb{E}\left[(Y_i - \alpha - \tau D_i - \widetilde{X}_i'\beta)^2\right]$$

- $\widehat{\tau}_{\text{adj}}$ now **biased** for τ though bias should be small.

- But $\widehat{\tau}_{\text{adj}}$ is **consistent**.
  - Intuition: Since $D_i \perp\!\!\!\perp \mathbf{X}_i$, including $\widetilde{\mathbf{X}}_i$ won't (asymptotically) affect coefficient on D_i.

OLS is biased, but consistent (Freedman, 2008. Adv. in Appl. Math)

- Agnostic approach: don't assume correctness of the linear model.
    - $\mathbf{X}_i'\beta$ could be badly misspecified.
    - No assumptions about homoskedasticity.

- Under minimal assumptions, OLS is consistent for the linear projection

$$(\widehat{\tau}_{\text{adj}}, \widehat{\alpha}_{\text{adj}}, \widehat{\beta}_{\text{adj}}) \xrightarrow{p} (\tau_0, \alpha_0, \beta_0) = \underset{(\tau,\alpha,\beta)}{\arg\min}\, \mathbb{E}\left[(Y_i - \alpha - \tau D_i - \widetilde{X}_i'\beta)^2\right]$$

- $\widehat{\tau}_{\text{adj}}$ now **biased** for τ though bias should be small.

- But $\widehat{\tau}_{\text{adj}}$ is **consistent**.
    - Intuition: Since $D_i \perp\!\!\!\perp \mathbf{X}_i$, including $\widetilde{\mathbf{X}}_i$ won't (asymptotically) affect coefficient on D_i.

- Freedman (2008) shows the same thing for finite-sample inference.

Variance of Adjustment Estimator

- Complete randomization + single, mean-zero covariate X_i
    - Generalizes easily to more covariates.
    - Let $\sigma_{0x} = \text{cov}(Y_i(0), X_i)$ and $\sigma_{1x} = \text{cov}(Y_i(1), X_i)$.
    - Probability of treatment $p = n_1/n$

Variance of Adjustment Estimator

- Complete randomization + single, mean-zero covariate X_i
  - Generalizes easily to more covariates.
  - Let $\sigma_{0x} = \text{cov}(Y_i(0), X_i)$ and $\sigma_{1x} = \text{cov}(Y_i(1), X_i)$.
  - Probability of treatment $p = n_1/n$

- Freedman (2008) derived gains from adjusting for X_i using OLS:

$$\mathbb{V}[\widehat{\tau}_{\text{diff}}] - \mathbb{V}[\widehat{\tau}_{\text{adj}}] = \frac{\sigma_{0x}\{\sigma_{0x} + 2(1-2p)\sigma_{1x}\}}{np(1-p)}$$

Variance of Adjustment Estimator

- Complete randomization + single, mean-zero covariate X_i
  - Generalizes easily to more covariates.
  - Let $\sigma_{0x} = \text{cov}(Y_i(0), X_i)$ and $\sigma_{1x} = \text{cov}(Y_i(1), X_i)$.
  - Probability of treatment $p = n_1/n$

- Freedman (2008) derived gains from adjusting for X_i using OLS:

$$\mathbb{V}[\widehat{\tau}_{\text{diff}}] - \mathbb{V}[\widehat{\tau}_{\text{adj}}] = \frac{\sigma_{0x}\{\sigma_{0x} + 2(1-2p)\sigma_{1x}\}}{np(1-p)}$$

- Will adjustment decrease the sampling variance?
  - If design is balanced, $p = 1/2$, then adjustment always helps.
  - Design imbalance could lead to adjustment hurting.

Variance of Adjustment Estimator

- Complete randomization + single, mean-zero covariate X_i
  - Generalizes easily to more covariates.
  - Let $\sigma_{0x} = \text{cov}(Y_i(0), X_i)$ and $\sigma_{1x} = \text{cov}(Y_i(1), X_i)$.
  - Probability of treatment $p = n_1/n$

- Freedman (2008) derived gains from adjusting for X_i using OLS:

$$\mathbb{V}[\widehat{\tau}_{\text{diff}}] - \mathbb{V}[\widehat{\tau}_{\text{adj}}] = \frac{\sigma_{0x}\{\sigma_{0x} + 2(1-2p)\sigma_{1x}\}}{np(1-p)}$$

- Will adjustment decrease the sampling variance?
  - If design is balanced, $p = 1/2$, then adjustment always helps.
  - Design imbalance could lead to adjustment hurting.

- Estimation: EHW robust variance estimators are consistent or asymptotically conservative for $\mathbb{V}[\widehat{\tau}_{\text{adj}}]$

Regression with Full Interactions

- OLS estimator from fully interacted model, $\widehat{\tau}_{\text{inter}}$:

$$Y_i = \alpha + \tau D_i + \widetilde{\mathbf{X}}_i'\beta + D_i\widetilde{\mathbf{X}}_i'\gamma + \varepsilon_i$$

  - Equivalent to running separate Y_i on $\widetilde{\mathbf{X}}_i$ in each D_i

- As with non-interacted model, $\widehat{\tau}_{\text{inter}}$ is consistent for τ and asymptotically normal.

- Lin (2013): **fully interacted model will never hurt precision asymptotically.**
  - Freedman critique was right, but Lin shows an easy way to resolve.

- EHW robust variance estimators are consistent or asymptotically conservative.

Linear Regression with Covariates

In R

```
# Step 1: Compute centered covariates
your_data$Xtilde <- NULL

# Step 2: Write down your formula
your_formula <- NULL

# Step 3: Fit the model using lm() or estimatr::lm_robust()
your_fitted_model <- lm(your_formula, data = your_data)

# Step 4: Compute robust standard errors (skip if you used lm_robust)
your_vcov <- sandwich::vcovHC(your_fitted_model, type = 'HC2')

# Step 5: Check the point and se estimate of your coefficients
# (look for tau hat!)
est <- cbind("coef" = your_fitted_model$coef,
             "se" = sqrt(diag(your_vcov)))
```

Example Code

```
> AI_data <- AI_data |>
    mutate(Xtilde = student_age - mean(student_age)) |>
    select(treat_ind, test_outcome_post, Xtilde); head(AI_data,3)
# A tibble: 3 × 3
  treat_ind test_outcome_post Xtilde
      <int>             <int>  <dbl>
1         1                 1 -0.566
2         1                 1 -5.57
3         0                 1  3.43

> estimatr::lm_robust(test_outcome_post ~ treat_ind * Xtilde, data = AI_data)
                     Estimate  Std. Error    t value     Pr(>|t|)     CI Lower    CI Upper  DF
(Intercept)       0.649860304 0.029645673 21.9209157 5.557487e-75  0.591613722 0.708106886 496
treat_ind         0.038364943 0.042245496  0.9081428 3.642438e-01 -0.044637245 0.121367131 496
Xtilde            0.008351571 0.008581795  0.9731730 3.309417e-01 -0.008509581 0.025212723 496
treat_ind:Xtilde -0.019462905 0.012003266 -1.6214674 1.055529e-01 -0.043046422 0.004120612 496

> your_fitted_model <- lm(test_outcome_post ~ treat_ind * Xtilde, data = AI_data)
> vcovM_adj <- sandwich::vcovHC(your_fitted_model, type = 'HC2'); sqrt(diag(vcovM_adj))
     (Intercept)        treat_ind           Xtilde treat_ind:Xtilde
     0.029645673      0.042245496      0.008581795      0.012003266

> est <- cbind("coef" = your_fitted_model$coef,
               "se" = sqrt(diag(your_vcov))); est
                        coef          se
(Intercept)       0.649860304 0.029645673
treat_ind         0.038364943 0.042245496
Xtilde            0.008351571 0.008581795
treat_ind:Xtilde -0.019462905 0.012003266
```

Summarizing Regression

- Regression with no covariates = standard Neyman analysis.
- Regression with (uninteracted) covariates:
    - Consistent for SATE/PATE.
    - Usually will help with precision, but can hurt.
- Regression with interacted covariates:
    - Consistent for SATE/PATE.
    - Asymptotically will never hurt precision.
- Always use robust/HC2 variance estimators unless you have good reasons.

Onto the presentations & discussions!

Contact Information:
jaewon.yoo@iss.nthu.edu.tw
https://j1yoo4.github.io/

Appendix

Linear Regression and Causality

- Regression: conditional expectation function of Y given $\mathbf{X}$

$$\mathbb{E}(Y|\mathbf{X}) = f(\mathbf{X}) = \beta^T \mathbf{X}$$

- Q: When can we interpret coefficients as causal effects?
- Causal model as structural equation model:

$$Y_i(d) = \alpha + \tau d + \varepsilon_i \quad \text{for } d = 0, 1, \text{ where } \mathbb{E}(\varepsilon_i) = 0$$

  1. No interference between units
  2. $\mathbb{E}(Y_i(0)) = \alpha$
  3. $Y_i(1) - Y_i(0) = \tau$ for all i $\rightsquigarrow$ constant unit-level causal effect

- Heterogeneous treatment effect model:

$$Y_i(d) = \alpha + \tau_i d + \epsilon_i = \alpha + \tau d + \underbrace{(\tau_i - \tau)t + \varepsilon_i}_{=\epsilon_i(d)}$$

where $\mathbb{E}(\varepsilon_i) = 0$ and $\tau = \mathbb{E}(\tau_i) = \mathbb{E}(Y_i(1) - Y_i(0))$

  - $\mathbb{E}(\varepsilon_i(d)) = 0$ for $d = 0, 1$
  - $\alpha = \mathbb{E}(Y_i(0))$

Robust SEs

- Use robust variance estimator!
  - Eicker-Huber-White (EHW) estimator: consistent for $\mathbb{V}(\widehat{\tau}_{\text{diff}})$

$$\widehat{\mathbb{V}}_{\text{EHW}} = \frac{\widetilde{\sigma}_1^2}{n_1} + \frac{\widetilde{\sigma}_0^2}{n_0}, \text{ where } \widetilde{\sigma}_d^2 = \frac{1}{n_d} \sum_{i:D_i=d} (Y_i - \overline{Y}_d)^2$$

  - HC2 estimator:

$$\widehat{\mathbb{V}}_{\text{HC2}} = \frac{\widehat{\sigma}_0^2}{n_0} + \frac{\widehat{\sigma}_1^2}{n_1}, \text{ where } \widehat{\sigma}_d^2 = \frac{1}{n_d - 1} \sum_{i:D_i=d} (Y_i - \overline{Y}_d)^2$$

  - Samii & Aronow (2012): HC2 is exactly the Neyman variance estimator.
- ⇝ Simple OLS + HC2 = unbiased point and variance estimator.

In R

```
your_fitted_model <- lm(your_formula, data = your_data)
sandwich::vcovHC(your_fitted_model, type = 'HC2')
# Or
estimatr::lm_robust(your_formula, your_data, se_type = 'HC2')
```