Causal Inference Methods and Case Studies

STAT24630

Jingshu Wang

Lecture 6

Topic: Regression for complete randomized experiment

- Using regression with no covariates
- Using regression with covariates adjustments
- Using regression with covariates adjustments and interactions
- The LRC-CPPT cholesterol data example
- Textbook Chapter 7

Linear regression and causality

- Linear regression:

$$\mathbb{E}(Y_i|\boldsymbol{X}_i) = f(\boldsymbol{X}_i) = \alpha + \boldsymbol{\beta}^T\boldsymbol{X}_i$$

- Question:
  - When can we interpret the coefficient(s) as causal effect?
  - How can we do correct inference if we take into account the randomization procedure of treatment assignments?

- Benefit of using linear regression in randomized experiments
  - Provides a straightforward and familiar way to incorporate covariates
  - More accurate estimator if covariates are predictive of potential outcomes

- Some critiques
  - In complete randomized experiments, covariates are not confounders
  - Why do we want to assume a linear model if we don't need to?

"Experiments should be analyzed as experiments, not as observational studies"

---- David A. Freedman, 2006

The LRC-CPPT cholesterol data

- An experiment to evaluate the effect of the drug cholestyramine on reducing cholesterol levels
- $N = 337$ patients are completely randomized
- **Pre-treatment covariates:** two cholesterol measurements before and after a suggestion of low-cholesterol diet, both measurements taken prior to the random assignment
  - Does chol2 – chol1 reflect the average causal effect of suggestion? Not necessarily
  - cholp = 0.25 chol1 + 0.75 chol2

Table 7.1. ***Summary Statistics for PRC-CPPT Cholesterol Data***

	Variable	Control (N_c =172)		Treatment (N_t =165)			
		Average	Sample (S.D.)	Average	Sample (S.D.)	Min	Max
Pre-treatment	chol1	297.1	(23.1)	297.0	(20.4)	247.0	442.0
	chol2	289.2	(24.1)	287.4	(21.4)	224.0	435.0
	cholp	291.2	(23.2)	289.9	(20.4)	233.0	436.8
Post-treatment	cholf	282.7	(24.9)	256.5	(26.2)	167.0	427.0
	chold	−8.5	(10.8)	−33.4	(21.3)	−113.3	29.5
	comp	74.5	(21.0)	59.9	(24.4)	0	101.0

The LRC-CPPT cholesterol data

- An experiment to evaluate the effect of the drug cholestyramine on reducing cholesterol levels
- $N = 337$ patients are completely randomized
- **Post-treatment outcomes:**
  - cholf: post-treatment average cholesterol level
  - chold = cholf – cholp
  - comp: compliance rate, the percentage of individuals follow the treatment assignment

Table 7.1. ***Summary Statistics for PRC-CPPT Cholesterol Data***

	Variable	Control (N_c =172)		Treatment (N_t =165)			
		Average	Sample (S.D.)	Average	Sample (S.D.)	Min	Max
Pre-treatment	chol1	297.1	(23.1)	297.0	(20.4)	247.0	442.0
	chol2	289.2	(24.1)	287.4	(21.4)	224.0	435.0
	cholp	291.2	(23.2)	289.9	(20.4)	233.0	436.8
Post-treatment	cholf	282.7	(24.9)	256.5	(26.2)	167.0	427.0
	chold	−8.5	(10.8)	−33.4	(21.3)	−113.3	29.5
	comp	74.5	(21.0)	59.9	(24.4)	0	101.0

The LRC-CPPT cholesterol data

- Can we evaluate the drug effect by simply look at whether chold is positive or negative?
  - No! The before-after comparison is NOT necessarily causal
  - Even for the control group, chold is significantly negative
- The patient's post-treatment cholesterol should be highly correlated with his/her pre-treatment cholesterol level
- How do we evaluate the causal effect after "adjusting for the pre-treatment cholesterol"?

Table 7.1. ***Summary Statistics for PRC-CPPT Cholesterol Data***

	Variable	Control (N_c =172)		Treatment (N_t =165)			
		Average	Sample (S.D.)	Average	Sample (S.D.)	Min	Max
Pre-treatment	chol1	297.1	(23.1)	297.0	(20.4)	247.0	442.0
	chol2	289.2	(24.1)	287.4	(21.4)	224.0	435.0
	cholp	291.2	(23.2)	289.9	(20.4)	233.0	436.8
Post-treatment	cholf	282.7	(24.9)	256.5	(26.2)	167.0	427.0
	chold	−8.5	(10.8)	−33.4	(21.3)	−113.3	29.5
	comp	74.5	(21.0)	59.9	(24.4)	0	101.0

What does "adjust for" exactly mean here?

The LRC-CPPT cholesterol data

A bit explanation about compliance

- If we compare between control and treatment group, we are evaluating the causal effect of "being assigned", not the causal effect of actually taking the drug
- Compliance lower in the treatment group possibly due to the side effect of the drug
- Can we just throw away individuals who do not follow the treatment and estimate the causal effect of taking the drug based on the rest individuals? No
- Will discuss more about compliance in later lectures

Table 7.1. ***Summary Statistics for PRC-CPPT Cholesterol Data***

	Variable	Control (N_c =172)		Treatment (N_t =165)			
		Average	Sample (S.D.)	Average	Sample (S.D.)	Min	Max
Pre-treatment	chol1	297.1	(23.1)	297.0	(20.4)	247.0	442.0
	chol2	289.2	(24.1)	287.4	(21.4)	224.0	435.0
	cholp	291.2	(23.2)	289.9	(20.4)	233.0	436.8
Post-treatment	cholf	282.7	(24.9)	256.5	(26.2)	167.0	427.0
	chold	−8.5	(10.8)	−33.4	(21.3)	−113.3	29.5
	comp	74.5	(21.0)	59.9	(24.4)	0	101.0

Linear regression with no covariates

- Causal model on the potential outcomes
$$Y_i(w) = \alpha + \tau_i w + \varepsilon_i^* = \alpha + \tau w + \varepsilon_i(w)$$
where $\mathbb{E}(\varepsilon_i^*) = 0$ and $\varepsilon_i(w) = \varepsilon_i^* + (\tau_i - \tau)w$
  - Assume that there is a super-population and the potential outcomes are random
  - Individual-level causal effect $\tau_i = Y_i(1) - Y_i(0)$ is also treated as random (as individuals are randomly sampled), and are allowed to be heterogenous
  - Define PATE: $\tau = \mathbb{E}(\tau_i) = \mathbb{E}\big(Y_i(1) - Y_i(0)\big)$
  - $\alpha = \mathbb{E}\big(Y_i(0)\big)$ and $\mathbb{E}\big(\varepsilon_i(w)\big) = 0$
  - If the treatment is binary ($w = 0,1$), then the above model essentially has no assumption on $Y_i(0)$ and $Y_i(1)$
  - If the treatment is continuous, the model assumes a linear but heterogenous causal effect on each individual

  - How to estimate τ from observed data?
  - When does the above model imply the linear regression model on observed data?
$$Y_i^{\text{obs}} = \alpha + \tau W_i + \varepsilon_i$$

Linear regression with no covariates

$$Y_i(w) = \alpha + \tau_i w + \varepsilon_i^* = \alpha + \tau w + \varepsilon_i(w)$$

We assume the following identification conditions

- Randomization of the treatment:

$$(\boldsymbol{Y}(0), \boldsymbol{Y}(1)) \perp \boldsymbol{W}$$

  - Satisfied in complete randomized experiments
  - Then, $\mathbb{E}\big(Y_i(w)\big) = \mathbb{E}\big(Y_i^{\text{obs}} | W_i = w\big) = \alpha + \tau w$ Regression model for the observed Y_i^{obs}
  - So this implies a regression model $Y_i^{\text{obs}} = \alpha + \tau W_i + \varepsilon_i$ and $\varepsilon_i = \varepsilon_i(W_i)$
  - In the regression model, we treat assignment vectors as fixed

- Random sampling of the units
  - $(\varepsilon_i(0), \varepsilon_i(1))$ are independent across i
  - This implies that ε_i in the linear regression model are independent as W_i are treated as fixed (the regression model is conditional on W_i)

Least square estimator

$$(\hat{\tau}^{\text{ols}}, \hat{\alpha}^{\text{ols}}) = \arg\min_{\tau,\alpha} \sum_{i=1}^{N} \left(Y_i^{\text{obs}} - \alpha - \tau \cdot W_i \right)^2,$$

with solutions

$$\hat{\tau}^{\text{ols}} = \frac{\sum_{i=1}^{N} \left(W_i - \overline{W}\right) \cdot \left(Y_i^{\text{obs}} - \overline{Y}^{\text{obs}}\right)}{\sum_{i=1}^{N} \left(W_i - \overline{W}\right)^2}, \quad \text{and} \quad \hat{\alpha}^{\text{ols}} = \overline{Y}^{\text{obs}} - \hat{\tau}^{\text{ols}} \cdot \overline{W},$$

where

$$\overline{Y}^{\text{obs}} = \frac{1}{N} \sum_{i=1}^{N} Y_i^{\text{obs}} \quad \text{and} \quad \overline{W} = \frac{1}{N} \sum_{i=1}^{N} W_i = \frac{N_{\text{t}}}{N}.$$

Simple linear algebra shows that $\hat{\tau}^{\text{ols}} = \overline{Y}_{\text{t}}^{\text{obs}} - \overline{Y}_{\text{c}}^{\text{obs}} = \hat{\tau}^{\text{dif}}$

- $\hat{\tau}^{\text{ols}}$ is unbiased for the estimation of τ
- How to estimate the variance of $\hat{\tau}^{\text{ols}}$?

Homoscedastic error assumption

Homoscedastic error assumption: $\mathbb{V}(\varepsilon_i(0)) = \mathbb{V}(\varepsilon_i(1)) = \sigma^2 = \mathbb{V}(Y_i^{\text{obs}} | W_i)$

- OLS estimates of the variance is

$$\hat{\sigma}^2_{Y|W} = \frac{1}{N-2} \sum_{i=1}^{N} \hat{\varepsilon}_i^2 = \frac{1}{N-2} \sum_{i=1}^{N} \left(Y_i^{\text{obs}} - \hat{Y}_i^{\text{obs}} \right)^2,$$

where the estimated residual is $\hat{\varepsilon}_i = Y_i^{\text{obs}} - \hat{Y}_i^{\text{obs}}$, and the predicted value $\hat{Y}_i^{\text{obs}}$ is

$$\hat{Y}_i^{\text{obs}} = \begin{cases} \hat{\alpha}^{\text{ols}} & \text{if } W_i = 0, \\ \hat{\alpha}^{\text{ols}} + \hat{\tau}^{\text{ols}} & \text{if } W_i = 1. \end{cases}$$

- Then from OLS, estimate of the variance of $\hat{\tau}^{\text{ols}}$

$$\hat{\mathbb{V}}^{\text{homosk}} = \frac{\hat{\sigma}^2_{Y|W}}{\sum_{i=1}^{N} \left(W_i - \overline{W} \right)^2}$$

Heteroscedastic errors

- If we don't want to assume $\mathbb{V}(\varepsilon_i(0)) = \mathbb{V}(\varepsilon_i(1))$, then the homoscedastic error assumption fails
- Robust variance estimation of $\hat{\tau}^{\text{ols}}$ allowing any heterogeneity of $\mathbb{V}\left(Y_i^{\text{obs}}|W_i\right)$ across i

$$\hat{\mathbb{V}}^{\text{hetero}} = \frac{\sum_{i=1}^{N} \hat{\varepsilon}_i^2 \cdot \left(W_i - \overline{W}\right)^2}{\left(\sum_{i=1}^{N} \left(W_i - \overline{W}\right)^2\right)^2}$$

  - $\hat{\varepsilon}_i = Y_i^{\text{obs}} - \hat{\alpha} - \hat{\tau} W_i = Y_i^{\text{obs}} - \hat{Y}_i^{\text{obs}}$ are the residuals
  - If we replace $\hat{\varepsilon}_i^2$ by $\hat{\sigma}_{Y|W}^2$, then we get back the OLS variance estimation of $\hat{\tau}^{\text{ols}}$
  - This is also called the Sandwich estimator that is robust to the violation of the homoscedastic noise assumption in linear regression

  - As $\hat{\tau}$ is estimated from Y_i^{obs}, $\mathbb{E}({\hat{\varepsilon}_i}^2) < \mathbb{V}(\varepsilon_i)$

  - When sample size N is small, $\hat{\mathbb{V}}^{\text{hetero}}$ will underestimate the true variance of $\hat{\tau}^{\text{ols}}$

Heteroscedastic errors

- The HC2 adjustment

$$\widehat{\mathbb{V}}_{HC2}^{\text{hetero}} = \frac{\sum_{i=1}^{N} (\hat{\varepsilon}_i / \sqrt{1-h_{ii}})^2 (W_i - \overline{W})^2}{\left(\sum_{i=1}^{N} (W_i - \overline{W})^2\right)^2}$$

  - h_{ii} is the leverage of unit i in the linear regression
  - If W_i is binary. $\widehat{\mathbb{V}}_{HC2}^{\text{hetero}} = \frac{s_c^2}{N_c} + \frac{s_t^2}{N_t}$
  - This is the same variance estimator as from Neyman's approach, or our two-sample testing approach

Linear regression with covariates adjustment

To summarize the logic

- We build a (linear) model on the potential outcomes
- This model implies a linear regression model on the observed outcome if $(\boldsymbol{Y}(0), \boldsymbol{Y}(1)) \perp \boldsymbol{W}$
- The coefficient on W_i in the linear regression model is the average causal effect (PATE)
- The linear regression model treat $\boldsymbol{W}$ as fixed so it works for any randomization assignment mechanism that satisfy $(\boldsymbol{Y}(0), \boldsymbol{Y}(1)) \perp \boldsymbol{W}$
- Noise in the linear regression model are independent as long as potential outcomes are independent across units

- For statistical inference
  - The OLS estimator estimator is always unbiased
  - We can apply standard linear regression inference results if we assume $\mathbb{V}(\varepsilon_i(0)) = \mathbb{V}(\varepsilon_i(1))$
  - If $\mathbb{V}(\varepsilon_i(0)) \neq \mathbb{V}(\varepsilon_i(1))$, we need to use the robust variance estimator

Linear regression with covariates adjustment

- Regress $Y_i(0)$ on the pre-treatment covariates $\boldsymbol{X}_i$

$$Y_i(w) = \alpha + \tau_i w + \boldsymbol{\beta}^T \boldsymbol{X}_i + \varepsilon_i^* = \alpha + \tau w + \boldsymbol{\beta}^T \boldsymbol{X}_i + \varepsilon_i(w)$$

where $\mathbb{E}(\varepsilon_i^* | \boldsymbol{X}_i) = 0$ and $\varepsilon_i(w) = \varepsilon_i^* + (\tau_i - \tau)w$

  - We assume that $\mathbb{E}(Y_i(0) | \boldsymbol{X}_i) = \alpha + \boldsymbol{\beta}^T \boldsymbol{X}_i$
  - Individual-level causal effect is $\tau_i = Y_i(1) - Y_i(0)$
  - Assume CATE $\tau(\boldsymbol{x}) = \mathbb{E}(\tau_i | \boldsymbol{X}_i = \boldsymbol{x}) \equiv \tau = \text{PATE}$ constant across levels of $\boldsymbol{X}_i$
  - We can allow for heterogeneous causal effect but need $\mathbb{E}(\tau_i - \tau \mid \boldsymbol{X}_i) = 0$ (individual causal effects are independent from the pre-treatment covariates)
  - The above implies that $\mathbb{E}(\varepsilon_i(w) | \boldsymbol{X}_i) = 0$

  - The above also implies that $\mathbb{E}(Y_i(w) | \boldsymbol{X}_i) = \alpha + \tau w + \boldsymbol{\beta}^T \boldsymbol{X}_i$

When does the above model imply the linear regression model on observed data?

$$Y_i^{\text{obs}} = \alpha + \tau W_i + \boldsymbol{\beta}^T \boldsymbol{X}_i + \varepsilon_i$$

Linear regression with covariates adjustment

$$Y_i(w) = \alpha + \tau w + \boldsymbol{\beta}^T \boldsymbol{X}_i + \varepsilon_i(w)$$

We assume the following identification conditions

- (Conditional) randomization of the treatment:

$$\big(\boldsymbol{Y}(0), \boldsymbol{Y}(1)\big) \perp \boldsymbol{W} \mid \boldsymbol{X}$$

  - Always satisfied in randomized experiments
  - $\mathbb{E}(Y_i(w)|\boldsymbol{X}_i = \boldsymbol{x}) = \mathbb{E}\big(Y_i^{\text{obs}}|W_i = w, \boldsymbol{X}_i = \boldsymbol{x}\big) = \alpha + \tau w + \boldsymbol{\beta}^T \boldsymbol{X}_i$

Regression model for the observed Y_i^{obs}

  - So this implies a regression model $Y_i^{\text{obs}} = \alpha + \tau W_i + \boldsymbol{\beta}^T \boldsymbol{X}_i + \varepsilon_i$ and $\varepsilon_i = \varepsilon_i(W_i)$
  - Both $\boldsymbol{X}_i$ and W_i are treated as fixed

- Random sampling of the units
  - $(\varepsilon_i(0), \varepsilon_i(1))$ are independent across i
  - This implies that $\varepsilon_i = \varepsilon_i(W_i)$ are independent across units

OLS with covariates adjustment

$$(\hat{\alpha}^{\text{ols}}, \hat{\tau}^{\text{ols}}, \hat{\beta}^{\text{ols}}) = \arg\min_{\alpha,\tau,\beta} \sum_{i=1}^{N} \left(Y_i^{\text{obs}} - \alpha - \tau \cdot W_i - X_i\beta\right)^2$$

- The estimator $\hat{\tau}^{\text{ols}}$ is unbiased for the causal estimand τ
- Even if the model is incorrect (either the violation of $\mathbb{E}(Y_i(0)|\,\boldsymbol{X}_i) = \alpha + \boldsymbol{\beta}^T\boldsymbol{X}_i$ or $\tau \equiv \mathbb{E}(\tau_i|\,\boldsymbol{X}_i = \boldsymbol{x})$), $\hat{\tau}^{\text{ols}}$ still converges to the PATE $\mathbb{E}(\tau_i)$ under complete randomization

Efficiency gain from regression

- If the model is correct, we have

$$\mathbb{V}\left(\hat{\tau}^{\text{ols}}\right) \approx \frac{\mathbb{E}\{\mathbb{V}(Y_i(1)|\,\boldsymbol{X}_i)\}}{N_t} + \frac{\mathbb{E}\{\mathbb{V}(Y_i(0)|\,\boldsymbol{X}_i)\}}{N_c} \leq \frac{\sigma_c^2}{N_c} + \frac{\sigma_t^2}{N_t}$$

  - If $\boldsymbol{X}_i$ is predictive of the (potential) outcomes, we have a more accurate estimator
- If the linear model is incorrect, the efficiency might be lost (Freedman 2008, *Adv. Appl. Math.)*

Estimate of the variance of $\hat{\tau}^{\text{ols}}$ with covariates adjustment

- Assume homoscedastic error assumption:

$$\mathbb{V}(\varepsilon_i(0)) = \mathbb{V}(\varepsilon_i(1)) = \sigma^2 = \mathbb{V}\left(Y_i^{\text{obs}} | W_i, \boldsymbol{X}_i\right)$$

We can follow standard linear regression inference and estimate variance of $\hat{\tau}^{\text{ols}}$ as

$$\hat{\mathbb{V}}_{\text{sp}}^{\text{homo}} = \frac{1}{N\left(N-1-\dim(X_i)\right)} \cdot \frac{\sum_{i=1}^{N}\left(Y_i^{\text{obs}} - \hat{\alpha}^{\text{ols}} - \hat{\tau}^{\text{ols}} - X_i\hat{\beta}^{\text{ols}}\right)^2}{\overline{W}\cdot(1-\overline{W})}$$

- The robust variance estimator (Sandwich estimator) without assuming homoscedasticity

$$\hat{\mathbb{V}}_{\text{sp}}^{\text{hetero}} = \frac{1}{N\left(N-1-\dim(X_i)\right)} \cdot \frac{\sum_{i=1}^{N}\left(W_i - \overline{W}\right)^2 \cdot \left(Y_i^{\text{obs}} - \hat{\alpha}^{\text{ols}} - \hat{\tau}^{\text{ols}} - X_i\hat{\beta}^{\text{ols}}\right)^2}{\left(\overline{W}\cdot(1-\overline{W})\right)^2}$$

Linear regression with covariates adjustment and interactions

What if the assumption $\tau \equiv \tau(\boldsymbol{x}) = \mathbb{E}(\tau_i | \ \boldsymbol{X}_i = \boldsymbol{x})$ constant across levels of $\boldsymbol{X}_i$ is incorrect?

- $Y_i(w) = \alpha + \tau_i w + \boldsymbol{\beta}^T \boldsymbol{X}_i + \varepsilon_i^* = \alpha + \tau w + \boldsymbol{\beta}^T \boldsymbol{X}_i + \varepsilon_i(w)$ where $\mathbb{E}(\varepsilon_i^* | \boldsymbol{X}_i) = 0$ and $\varepsilon_i(w) = \varepsilon_i^* + (\tau_i - \tau)w$

- Assume CATE $\tau(\boldsymbol{x}) = \mathbb{E}(\tau_i | \ \boldsymbol{X}_i = \boldsymbol{x}) = \tau + \boldsymbol{\gamma}^T(\boldsymbol{x} - \overline{\boldsymbol{X}})$
- τ is still the population average treatment effect
- Then $\mathbb{E}(\varepsilon_i(w) | \ \boldsymbol{X}_i) = \boldsymbol{\gamma}^T(\ \boldsymbol{X}_i - \overline{\boldsymbol{X}})w$

- The above also implies that $\mathbb{E}(Y_i(w) | \ \boldsymbol{X}_i) = \alpha + \tau w + \boldsymbol{\beta}^T \boldsymbol{X}_i + \boldsymbol{\gamma}^T(\ \boldsymbol{X}_i - \overline{\boldsymbol{X}})w$

- When does the above model imply the linear regression model with interactions on observed data?

$$Y_i^{\text{obs}} = \alpha + \tau W_i + \boldsymbol{\beta}^T \boldsymbol{X}_i + \boldsymbol{\gamma}^T(\ \boldsymbol{X}_i - \overline{\boldsymbol{X}})W_i + \varepsilon_i$$

  - Same assumptions as before (slide 15)
  - We can still use least square estimators

Results on the LRC-CPPT cholesterol data

- We estimate the PATE for both the post-treatment cholesterol level cholf and compliance
    - A considerable reduction of the variance of $\hat{\tau}^{\text{ols}}$ for cholf when we add the pre-treatment cholesterol levels in the regression
    - Our goal is always estimating PATE even after “covariates adjustment”
    - In randomized experiments satisfying $(\boldsymbol{Y}(0), \boldsymbol{Y}(1)) \perp \boldsymbol{W}$, adjusting for covariates or not, our estimate of PATE is always valid, we only change the efficiency of our estimate

Covariates	Effect of Assignment to Treatment on			
	Post-Cholesterol Level		Compliance	
	$\hat{\tau}$	$\widehat{\text{(s. e.)}}$	$\hat{\tau}$	$\widehat{\text{(s. e.)}}$
No covariates	−26.22	(3.93)	−14.64	(3.51)
`cholp`	−25.01	(2.60)	−14.68	(3.51)
`chol1, chol2`	−25.02	(2.59)	−14.95	(3.50)
`chol1, chol2`, interacted with W	−25.04	(2.56)	−14.94	(3.49)

Results on the LRC-CPPT cholesterol data

- We estimate the PATE for both the post-treatment cholesterol level cholf and compliance
  - A considerable reduction of the variance of $\hat{\tau}^{\text{ols}}$ for cholf when we add the pre-treatment cholesterol levels in the regression
  - Our goal is always estimating PATE even after "covariates adjustment"
  - In randomized experiments satisfying $(\boldsymbol{Y}(0), \boldsymbol{Y}(1)) \perp \boldsymbol{W}$, adjusting for covariates or not, our estimate of PATE is always valid, we only change the efficiency of our estimate
  - Which variables are "significantly" contributing to the variance reduction of $\hat{\tau}^{\text{ols}}$ for cholf ?

Covariates	Model for Levels	
	Est	$\widehat{(\text{s. e.})}$
Assignment	−25.04	(2.56)
Intercept	−3.28	(12.05)
`chol1`	0.98	(0.04)
`chol2-chol1`	0.61	(0.08)
`chol1` × Assignment	−0.22	(0.09)
(`chol2-chol1`) × Assignment	0.07	(0.14)
R-squared	0.63	

Why do we use linear regression in randomized experiments?

- Covariate adjustment can be used to improve efficiency in randomized experiments
- Under various experimental designs, linear regression models are useful methods for this purpose

- Randomization of treatment assignment protects researchers from misspecification
  - independence between treatment and covariates
  - linear regression estimators are often consistent even when the model is incorrect