

# 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

# Linear regression and causality

- Linear regression:

$$\mathbb{E}(Y_i|W_i, \mathbf{X}_i) = \alpha + \gamma W_i + \boldsymbol{\beta}^T \mathbf{X}_i$$

- 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”

# 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  $\text{chol2} - \text{chol1}$  reflect the average causal effect of suggestion? Not necessarily
  - $\text{cholp} = 0.25 \text{ chol1} + 0.75 \text{ chol2}$

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

	Variable	Control ( $N_c = 172$ )		Treatment ( $N_t = 165$ )		Min	Max
		Average	Sample (S.D.)	Average	Sample (S.D.)		
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:**
  - chol<sub>f</sub>: post-treatment average cholesterol level
  - chol<sub>d</sub> = chol<sub>f</sub> – chol<sub>p</sub>
  - 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 ( $N_t = 165$ )		Min	Max
		Average	Sample (S.D.)	Average	Sample (S.D.)		
Pre-treatment	chol <sub>1</sub>	297.1	(23.1)	297.0	(20.4)	247.0	442.0
	chol <sub>2</sub>	289.2	(24.1)	287.4	(21.4)	224.0	435.0
	chol <sub>p</sub>	291.2	(23.2)	289.9	(20.4)	233.0	436.8
Post-treatment	chol <sub>f</sub>	282.7	(24.9)	256.5	(26.2)	167.0	427.0
	chol <sub>d</sub>	-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 ( $N_t = 165$ )		Min	Max
		Average	Sample (S.D.)	Average	Sample (S.D.)		
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 ( $N_t = 165$ )		Min	Max
		Average	Sample (S.D.)	Average	Sample (S.D.)		
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

- Neyman's approach

$$\hat{\tau}^{\text{dif}} = \bar{Y}_t^{\text{obs}} - \bar{Y}_c^{\text{obs}}$$

$$\bar{Y}_c^{\text{obs}} = \frac{1}{N_c} \sum_{i:W_i=0} Y_i^{\text{obs}} \quad \text{and} \quad \bar{Y}_t^{\text{obs}} = \frac{1}{N_t} \sum_{i:W_i=1} Y_i^{\text{obs}}$$

- An alternative way to get  $\hat{\tau}^{\text{dif}}$  is by linear regression

$$(\hat{\alpha}, \hat{\beta}) = \arg \min_{(a,b)} \sum_{i=1}^n (Y_i - a - bZ_i)^2$$

- It is easy to show that  $\hat{\beta} = \hat{\tau}^{\text{dif}}$

- However, if we use R function `lm` to compute  $\text{var}(\hat{\beta})$ , the variance estimator is different from  $\frac{s_c^2}{N_c} + \frac{s_t^2}{N_t}$



# Linear regression with no covariates

- A causal interpretation of the linear regression model

- Linear 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 (i.i.d. samples)
- Notice that the observed outcomes  $Y_i = W_i Y_i(1) + (1 - W_i) Y_i(0)$  are not i.i.d. samples under complete randomization

- For binary treatment:

$$Y_i(0) = \alpha + \varepsilon_i^*, Y_i(1) = Y_i(0) + \tau_i$$

- Define PATE:  $\tau = \mathbb{E}(\tau_i) = \mathbb{E}(Y_i(1) - Y_i(0))$
- $\alpha = \mathbb{E}(Y_i(0))$  and  $\mathbb{E}(\varepsilon_i(w)) = 0$

# 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  $\varepsilon_i(w) = \varepsilon_i^* + (\tau_i - \tau)w$

- 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  $\tau$  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:**

$$(Y(0), Y(1)) \perp W$$

- Satisfied in complete randomized experiments
  - Then,  $\mathbb{E}(Y_i(w)) = \mathbb{E}(Y_i(w) | W_i = w) = \mathbb{E}(Y_i^{\text{obs}} | W_i = w) = \alpha + \tau w$
  - So this implies a regression model  $Y_i^{\text{obs}} = \alpha + \tau W_i + \varepsilon_i$  where  $\varepsilon_i = \varepsilon_i(W_i)$
  - 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  $\varepsilon_i$  in the linear regression model are **independent** as  $W_i$  are treated as fixed (the regression model is conditional on  $W_i$ )

Regression model for the observed  $Y_i^{\text{obs}}$

# 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

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

where

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

Simple linear algebra shows that

$$\hat{\tau}^{\text{ols}} = \bar{Y}_t^{\text{obs}} - \bar{Y}_c^{\text{obs}} = \hat{\tau}^{\text{dif}},$$

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

# Homoscedastic error assumption

Homoscedastic error assumption:  $\mathbb{V}(\varepsilon_i(0)) = \mathbb{V}(\varepsilon_i(1)) = \sigma^2$

- Then  $\mathbb{V}(Y_i^{\text{obs}} | W_i) = \varepsilon_i = \varepsilon_i(W_i)$  always has variance  $\sigma^2$
- Notice that in general the homoscedastic error assumption on the observed data linear regression model is not true
- Under homoscedasticity, OLS estimates of the variance is

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

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

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

# Homoscedastic error assumption

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

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

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

- For binary treatment, one can show that

$$\hat{\mathbb{V}}^{\text{homosk}} \approx \frac{s_c^2}{N_t} + \frac{s_t^2}{N_c}$$

- This is NOT the same as the Neyman's approach variance  $\frac{s_c^2}{N_c} + \frac{s_t^2}{N_t}$  even for large-sample size

# 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  $\mathbb{V}(Y_i^{\text{obs}} | W_i)$  across  $i$

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

- $\hat{\varepsilon}_i = Y_i^{\text{obs}} - \hat{\alpha} - \hat{\tau}W_i = Y_i^{\text{obs}} - \hat{Y}_i^{\text{obs}}$  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  $Y_i^{\text{obs}}$ ,  $\mathbb{E}(\hat{\varepsilon}_i^2) < \mathbb{V}(\varepsilon_i)$
- 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

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

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



# Linear regression without 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  $(Y(0), Y(1)) \perp W$
- The coefficient on  $W_i$  in the linear regression model is the average causal effect (PATE)
- The linear regression model treat  $W$  as fixed so it works for any randomization assignment mechanism that satisfy  $(Y(0), Y(1)) \perp W$
- 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  $V(\varepsilon_i(0)) = V(\varepsilon_i(1))$
  - If  $V(\varepsilon_i(0)) \neq V(\varepsilon_i(1))$ , we need to use the robust variance estimator

# Linear regression with covariates adjustment

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

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

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

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

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

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

# Linear regression with covariates adjustment

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

We assume the following identification conditions

- (Conditional) randomization of the treatment:

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

- Always satisfied in completely randomized experiments as long as  $\mathbf{X}$  are pre-treatment covariates

- $\mathbb{E}(Y_i(w) \mid \mathbf{X}_i = \mathbf{x}) = \mathbb{E}(Y_i^{\text{obs}} \mid W_i = w, \mathbf{X}_i = \mathbf{x}) = \alpha + \tau w + \boldsymbol{\beta}^T \mathbf{X}_i$

Regression model for the observed  $Y_i^{\text{obs}}$

- So this implies a regression model  $Y_i^{\text{obs}} = \alpha + \tau W_i + \boldsymbol{\beta}^T \mathbf{X}_i + \varepsilon_i$  and  $\varepsilon_i = \varepsilon_i(W_i)$
- Both  $\mathbf{X}_i$  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  $\tau$
- Even if the model is **incorrect** (either the violation of  $\mathbb{E}(Y_i(0) | \mathbf{X}_i) = \alpha + \boldsymbol{\beta}^T \mathbf{X}_i$  or  $\tau \equiv \mathbb{E}(\tau_i | \mathbf{X}_i = \mathbf{x})$ ),  $\hat{\tau}^{\text{ols}}$  still converges to the PATE  $\mathbb{E}(\tau_i)$  under complete randomization

## Efficiency gain from regression

- If the model is correct, we have

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

- If  $\mathbf{X}_i$  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:

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

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

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

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

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

# Linear regression with covariates adjustment and interactions

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

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

- Assume CATE  $\tau(\mathbf{x}) = \mathbb{E}(\tau_i | \mathbf{X}_i = \mathbf{x}) = \tau + \boldsymbol{\gamma}^T (\mathbf{x} - \bar{\mathbf{X}})$

- $\tau$  is still the population average treatment effect

- Then  $\mathbb{E}(\varepsilon_i(w) | \mathbf{X}_i) = \boldsymbol{\gamma}^T (\mathbf{X}_i - \bar{\mathbf{X}})w$

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

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

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

- Same assumptions as before (slide 15)

- With the interaction terms, we can always guarantee no efficiency loss (Peng's book section 6.2.2)

# Results on the LRC-CPPT cholesterol data

- We estimate the PATE for both the post-treatment cholesterol level `cholp` and compliance
  - A considerable reduction of the variance of  $\hat{\tau}^{ols}$  for `cholp` 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  $(Y(0), Y(1)) \perp W$ , 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)
<code>cholp</code>	-25.01	(2.60)	-14.68	(3.51)
<code>chol1</code> , <code>chol2</code>	-25.02	(2.59)	-14.95	(3.50)
<code>chol1</code> , <code>chol2</code> , 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 chol<sub>f</sub> and compliance
  - A considerable reduction of the variance of  $\hat{\tau}^{ols}$  for chol<sub>f</sub> 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  $(Y(0), Y(1)) \perp W$ , 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  $\hat{\tau}^{ols}$  for chol<sub>f</sub> ?

Covariates	Model for Levels	
	Est	(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