Precept 8: Diagnostics, presenting results, and causal inference

Soc 500: Applied Social Statistics

Ian Lundberg

Princeton University

November 10, 2016

Learning Objectives

1. Review Problem Set 6
2. Prepare for Problem Set 8
3. Causal inference
4. Power analysis
5. Visualization example in R

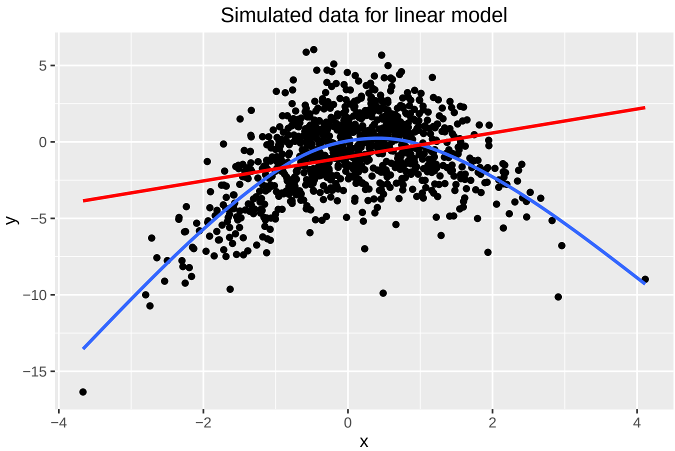
Residual plots: Linearity and homoskedasticity

- On the homework, there was some confusion about how to interpret a residual plot to evaluate the linearity and homoskedasticity assumptions
- These are **assumptions:** they are almost never 100% true in practice. But often they are reasonable enough to yield a useful model.
- If **linearity** is violated, the *mean* of the residuals will show a pattern with the fitted values (i.e. the trend line should be above 0 in some places and below 0 in others)
- If **homoskedasticity** is violated, the spread of the residuals *around that mean* will vary with the predicted values.
- These are **distinct** assumptions; though one plot tells you about both, they are not mechanically linked.

Violation of linearity

I generated some data that violates linearity:

$$Y = X - X^2 + u, u \sim N(0, 2)$$

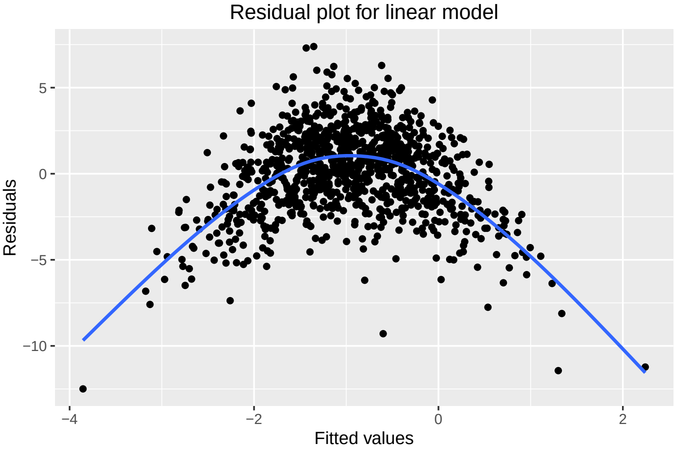


Violation of linearity

Then I fit a model that is linear in X to this data.

$$Y = \beta_0 + \beta_1 X + v$$

The residual plot is below. It shows nonlinearity.

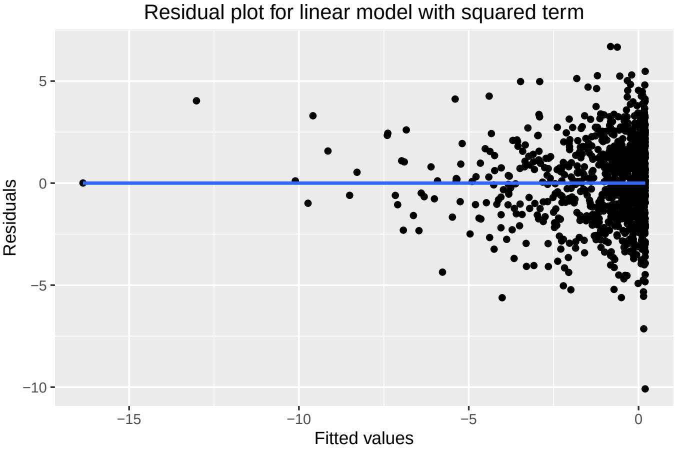


Violation of linearity

Then I fit a model that includes a quadratic in X.

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + v$$

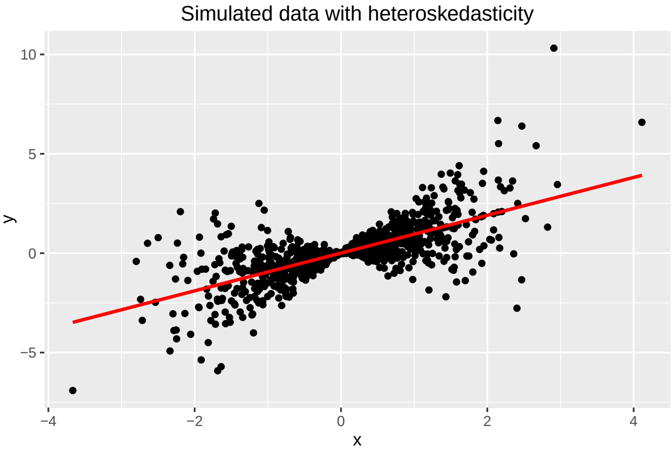
The residual plot is below. Linearity appears satisfied!



Violation of homoskedasticity

I generated some data that violates homoskedasticity (the error variance depends on X):

$$Y = X + u, u \sim N(0, |X|)$$

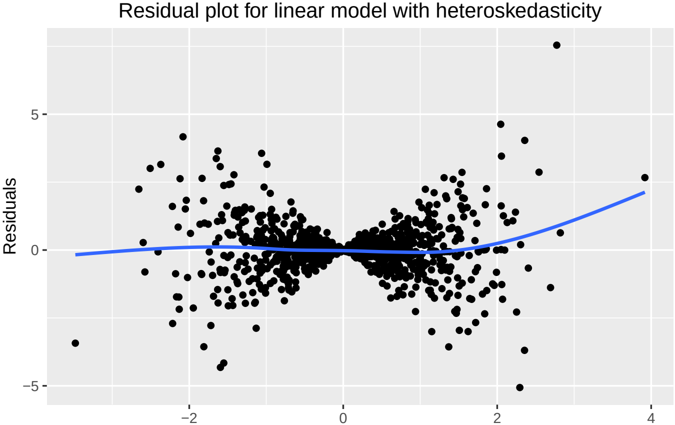


Violation of homoskedasticity

Then I fit a linear model.

$$Y = \beta_0 + \beta_1 X + v$$

The residual plot is below. The error variance is related to the fitted values - homoskedasticity is volated!



Interpreting interactions

Suppose

- Y is earnings
- Z is an indicator for getting a college degree
- X is years since completing schooling.

What is wrong with this model?

$$Y = \beta_0 + \beta_1 X + \beta_2 XZ + u$$

Interpreting interactions

Suppose

- Y is earnings
- Z is an indicator for getting a college degree
- X is years since completing schooling.

Consider the model:

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ + u$$

1. How do we interpret β_3?
2. What is the association between years since completed schooling and earnings, among those **without** a college degree?
3. What is the association between years since completed schooling and earnings, among those **with** a college degree?

Marginal "effects"

Consider the model

$$Y = \beta_0 + X\beta_1 + Z\beta_2 + XZ\beta_3 + u$$

The **marginal "effect"** of X on Y is defined to be the association between X and Y holding the other variables constant. It is also the partial derivative:

$$\frac{\partial Y}{\partial X} = \beta_1 + Z\beta_3$$

If Z is binary, this says that,

- when $Z = 0$, the association between X and Y is β_1
- when $Z = 1$, the association between X and Y is $\beta_1 + \beta_3$

Marginal "effects"

$$Y = \beta_0 + X\beta_1 + Z\beta_2 + XZ\beta_3 + u$$
$$\frac{\partial Y}{\partial X} = \beta_1 + Z\beta_3$$

What is the variance of the marginal effect?

$$\begin{aligned} Var\left(\frac{\partial Y}{\partial X}\right) &= Var(\hat{\beta}_1 + Z\hat{\beta}_3) \\ &= Var(\hat{\beta}_1) + Z^2 Var(\hat{\beta}_3) + 2ZCov(\hat{\beta}_1, \hat{\beta}_3) \end{aligned}$$

If this model is fit using the `lm()` function, we can use `vcov(fit)` to extract the variance covariance matrix that has these variance and covariance elements.

Marginal "effects"

Similarly, consider a model with a quadratic term:

$$Y = \beta_0 + X\beta_1 + X^2\beta_2 + u$$

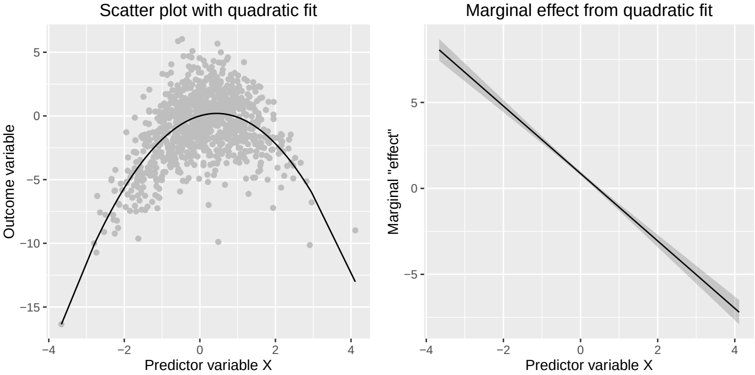
What is the marginal "effect" of X? What is its variance?

$$\frac{\partial Y}{\partial X} = \beta_1 + 2X\beta_2$$

$$\begin{aligned} Var\left(\frac{\partial Y}{\partial X}\right) &= Var(\hat{\beta}_1 + 2X\hat{\beta}_2) \\ &= Var(\hat{\beta}_1) + (2X)^2 Var(\hat{\beta}_2) + 2 * 2X * Cov(\hat{\beta}_1, \hat{\beta}_2) \end{aligned}$$

Plotting marginal effects

Given estimated coefficients, we could plot the marginal effect of X on Y as a function of X



Visualizing marginal effects
(guidelines from lecture slides)

Visualizations should:

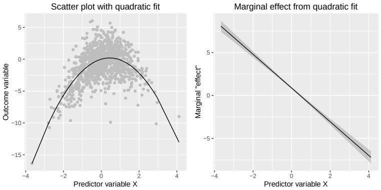
(a) Convey numerically precise estimates of the quantities of greatest substantive interest,

(b) Include reasonable measures of uncertainty about those estimates,

(c) Require little specialized knowledge to understand, and

(d) Include no superfluous information, no long lists of coefficients no one understands, no star gazing, etc.

Which was good in the plots above? How could they be improved?



Exploring causal inference with a running example

Research question

Does college education cause higher earnings?

Potential problems:

- Fundamental problem: We cannot observe both states
- Selection bias: Higher ability people may select into college
- SUTVA violation: Your sister's college education might affect your earnings
- Heterogeneity: Different effects on different people

Neyman-Rubin Model: Potential outcomes
(from lecture slides)

Two possible conditions:

- Treatment condition $T = 1$
- Control condition $T = 0$

Suppose that we have an **i**ndividual i.
Key assumption: we can imagine a world where individual i is assigned to treatment and control conditions
Potential Outcomes: responses under each condition, $Y_i(T)$

- Response under treatment $Y_i(1)$
- Response under control $Y_i(0)$

Potential outcomes in our college-earnings example

In our example examining how college affects future earnings, what do the numbers in the table below represent?

	Treatment ($Y_i(1)$)	Control ($Y_i(0)$)
Person 1	45,000	32,000
Person 2	54,000	45,000
Person 3	34,000	34,000

Potential outcomes in our college-earnings example

- $Y_i(1)$ is the earnings person i would make if they attended college
- $Y_i(0)$ is that person's potential earnings if they did not attend college
- The **individual causal effect** is $\tau_i = Y_i(1) - Y_i(0)$

Why can we never estimate τ_i? Because we can never observe an individual in both the treatment and control states.
This is the **fundamental problem of causal inference.**

Fundamental Problem of Causal Inference

Holland 1986: Only one outcome can be observed

	Treatment ($Y_i(1)$)	Control ($Y_i(0)$)
Person 1	45,000	32,000
Person 2	54,000	45,000
Person 3	34,000	34,000

Fundamental Problem of Causal Inference

Holland 1986: Only one outcome can be observed

	Treatment ($Y_i(1)$)	Control ($Y_i(0)$)
Person 1	**?**	32,000
Person 2	54,000	**?**
Person 3	34,000	**?**

Causal inference with a strong assumption

$$T \to Y$$

- T is the **treatment** (college education)
- Y is the **outcome** (earnings)

To make causal inferences, we must assume **ignorability**:

$$\{Y_i(1), Y_i(0)\} \perp\!\!\!\perp T_i$$

Then the **population average treatment effect** is

$$\begin{aligned}\bar{\tau} = E[Y_i(1) - Y_i(0)] &= E[Y_i(1)] - E[Y_i(0)] \\ &= E[Y_i(1) \mid T_i = 1] - E[Y_i(0) \mid T_i = 0]\end{aligned}$$

Meaning, if college education is ignorable with respect to potential earnings, then the average causal effect of college on earnings is just the **average earnings difference between the two groups**.

Identification and the role of assumptions

- A causal effect is **identified** if we could pin it down with an infinite amount of data.
- In any causal study we must state the assumptions under which a causal effect is identified.
- **Law of Decreasing Credibility (Manski)**: The credibility of inference decreases with the strength of the assumptions maintained

Assuming ignorability, the causal effect of college on earnings is **identified** by the mean difference between the two groups.
But this is a **heroic assumption**, so inference is not very credible!

Weakening assumptions as much as possible

- One goal of your work will be to find ways to weaken the needed assumptions.
- Instead of ignorability, let's try **conditional ignorability**!

$$\{Y_i(1), Y_i(0)\} \perp\!\!\!\perp T_i \mid X_i$$

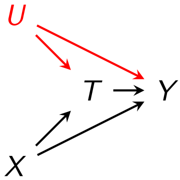
But first, let's step into it in the framework of DAGs.

Causal inference and Directed Acyclic Graphs (DAGs)

$$
\begin{tikzcd}
 & T \arrow[r] & Y \\
X \arrow[ur] \arrow[urr] & &
\end{tikzcd}
$$

- X is a **confounder** (ability)
- T is the **treatment** (college education)
- Y is the **outcome** (earnings)
- Ability causes both college education and earnings.

Causal inference and Directed Acyclic Graphs (DAGs)



To make causal inferences, we must assume there are no unobserved variables U that cause both T and Y, net of X. This is the assumption of **conditional ignorability**:

$$\{Y_i(1), Y_i(0)\} \perp\!\!\!\perp T_i \mid X_i$$

Once we account for X, the treatment T is as-if random.
Question: Is this study credible? Why or why not?

Blocking backdoor paths

Treatment assignment is **confounded** if there exists an **unblocked backdoor path** connecting T and Y.

$$\begin{tikzcd} & T \arrow[r] & Y \\ X \arrow[ur] \arrow[urr] & & \end{tikzcd}$$

Ian's informal definition: Imagine the arrows are rivers.
An **unblocked backdoor path** is one in which you can

1. Start at T
2. Swim upstream to some point (or walk out the back door)
3. Then swim downstream to Y (or walk around the house)

Conditioning on X is like building a dam or locking the back door: you block the path

Blocking backdoor paths: College and earnings

What do we need to include to block all backdoor paths between college and earnings?

$$\begin{tikzcd} & T \arrow[r] & Y \\ X \arrow[ur] \arrow[urr] & & \end{tikzcd}$$

Ability, parents' income, parents' education, extended family who pay for college and help you find a job, neighborhood characteristics that affect high school quality and also the availability of local jobs, ... **lots of things!**

Non-causal paths: Part 2

Now consider this graph. Is there an unblocked backdoor path from T to Y?

$$\begin{tikzcd}
X_1 \arrow[dr] \arrow[ddrrr, bend left] & & & \\
& X_2 & & \\
X_3 \arrow[ur] \arrow[rr] & & T \arrow[r] & Y
\end{tikzcd}$$

Remember, an **unblocked backdoor path** is one in which you can

1. Start at T
2. Swim upstream to some point
3. Then swim downstream to Y

No need to condition! X_2 already blocks this path. it is a **collider**.

Colliders: Be careful!

$$\begin{tikzcd} X_1 \arrow[dr] & \\ & Y \\ X_2 \arrow[ur] & \end{tikzcd}$$

Y is a **collider**. X_1 and X_2 are not associated, but they are when we hold Y constant.
What situations might produce this?

- X_1 being in a car accident. X_2 is having cancer. Y is being in a hospital.
- X_1 is living in a warm climate. X_2 is being an elite swimmer. Y is going swimming in January.
- X_1 is family income. X_2 is religiosity. Y attendance at a Catholic high school.

Colliders: When drawing a DAG helps

Example extended from Elwert & Winship 2014

Hypothetical substantive question:

Does acting ability causally affect the probability of marriage?

Hypothetical approach: Estimate on a sample of Hollywood actors and actresses.

We want to estimate:

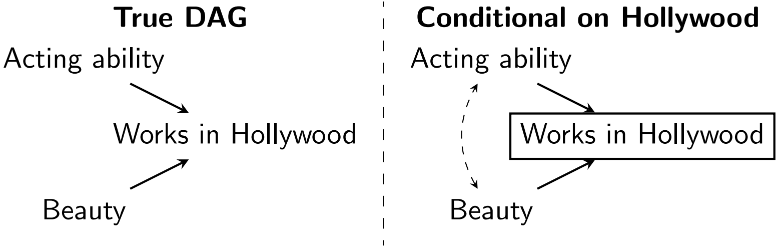
Acting ability $\longrightarrow$ Marriage

Should we worry about this design? It depends on our **theory** about how these variables are related. We can argue about identification with a DAG.

Colliders: When drawing a DAG helps

Example extended from Elwert & Winship 2014

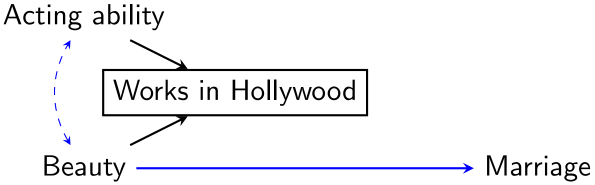
Suppose working in Hollywood is a function of two factors: acting ability and beauty. In the general population, these two are uncorrelated. However, among those who work in Hollywood, those who are bad at acting *must* be beautiful.



Colliders: When drawing a DAG helps

Example extended from Elwert & Winship 2014

This is an example of conditioning on a collider! We induce a negative association between acting ability and beauty.



Under the assumptions above, our results are driven by collider conditioning!

Stable Unit Treatment Value Assumption (SUTVA)

1. There is only **one** version of the treatment, not T_1, T_2
2. Potential outcomes depend only on **my** treatment status ($Y(1)$, not $Y(1, 0, 0, 1, 0, ..., 0, 1)$ or $Y(\boldsymbol{T})$) (no interference)

In our college-earnings example, what would violate each of these assumptions?

1. Elite college (Princeton) might be a different treatment from public universities. Only one treatment allowed.
2. My sister's college education might affect my earnings. Interference not allowed.

Definitions of treatment effects

Average treatment effect (ATE): Average over the whole population

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

Average treatment effect on the treated (ATT): Average among those who take the treatment

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

Average treatment effect on the control (ATC): Average among those who do not take the treatment

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

We could define an average treatment effect for any subpopulation of interest.

Definitions of treatment effects

In the case of college education and earnings,

- the ATT is the effect of college among those who attend
- the ATC is the effect among those who do not attend
- the ATE is the average effect over the whole population

In what situations might we care about each one?

Immutable characteristics

from lecture slides

There are three problems with race as a treatment in the causal inference sense

1. Race cannot be **manipulated**
   - without the capacity to manipulate the question is arguably ill-posed and the estimand is unidentified
2. Everything else is **post-treatment**
   - everything else comes after race which is perhaps unsatisfying
   - this also presumes we are only interested in the total effect
3. Race is **unstable**
   - there is substantial variance across treatments which is a SUTVA violation

Objections to potential outcomes
(Morgan and Winship concluding chapter)

- Cannot be used for nonmanipulable causes
  - but we can get at related things other ways
- Appropriate for effects of causes, not causes of effects
  - but we have to start with small steps before reaching for grand claims
- Relies on metaphysical quantities that cannot be observed
  - but we try to get at these with as minimal assumptions as possible

Concepts with which we want you to feel comfortable:

- Regression diagnostics and presentation
  - Residual plots and assumptions
  - Interpreting interactions
  - Marginal effects
- Causal inference
  - Potential outcomes
  - Identification vs. estimation
  - Assumptions and the Law of Decreasing Credibility
  - DAGs
    - Blocking backdoor paths
    - Recognizing confounders and colliders
  - SUTVA and interference
  - ATE, ATT, ATC