Methods Supplementary Lecture 2: More Survey Design and Analysis

Department of Government
London School of Economics and Political Science

Evaluative questions

- Name an object of evaluation
- Possibly describe that object
- Ask for a transformation of the evaluation onto a set of responses

Question templates

- Ratings
  - Several varieties of rating scales
- Scales/Thermometers
- Agree-disagree
- Forced choices
- Open-ended
- Rankings (note: need alternatives to rank against)

Extended Example

- Public opinion survey in Great Britain
- Construct: Opinion toward UK involvement in air strikes on Islamic State militants in Iraq and Syria
- Think about strengths and weaknesses of each question

Example: Rating (bipolar)

Do you support or oppose Great Britain's participation in U.S.-led air strikes on Islamic State (IS) in Iraq and Syria?

- Strongly support
- Somewhat support
- Neither support nor oppose
- Somewhat oppose
- Strongly oppose

Example: Rating (branching)

Do you support or oppose Great Britain's participation in U.S.-led air strikes on Islamic State (IS) in Iraq and Syria?

- Support
- Neither support nor oppose
- Oppose

Would you say that you strongly [support|oppose] or somewhat [support|oppose] Great Britain's participation?

- Strongly
- Somewhat

Example: Rating (bipolar)

Are you favourable or unfavourable toward Great Britain's participation in U.S.-led air strikes on Islamic State (IS) in Iraq and Syria?

- Very favourable
- Somewhat favourable
- Neither favourable nor unfavourable
- Somewhat unfavourable
- Strongly unfavourable

Example: Rating (unipolar)

To what extent do you support Great Britain's participation in U.S.-led air strikes on Islamic State (IS) in Iraq and Syria?

- Strongly
- Moderately
- Somewhat
- Not at all

Example: Rating (unipolar)

How favourable are you toward Great Britain's participation in U.S.-led air strikes on Islamic State (IS) in Iraq and Syria?

- Extremely favourable
- Very favourable
- Moderately favourable
- Somewhat favourable
- Not at all favourable

Example: Numbered Scale

On a scale from 1 to 5, with 1 being "strongly oppose" and 5 being "strongly support," to what extent do you support Great Britain's participation in U.S.-led air strikes on Islamic State (IS) in Iraq and Syria?

1. Strongly oppose
2.
3.
4.
5. Strongly support

Example: Thermometer

We would like to get your feelings toward some of political policies. Please rate your support for the policy using something we call the feeling thermometer. Ratings between 50 degrees and 100 degrees mean that you feel favourable and warm toward the policy. Ratings between 0 degrees and 50 degrees mean that you don't feel favourable toward the policy. You would rate the policy at the 50 degree mark if you don't feel particularly favourable or unfavourable toward.

Great Britain's participation in U.S.-led air strikes on Islamic State (IS) in Iraq and Syria.

- 0–100 slider

Example: Agree/Disagree (bipolar)

To what extent do you agree with the following statement: I support Great Britain's participation in U.S.-led air strikes on Islamic State (IS) in Iraq and Syria.

- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

Example: Agree/Disagree (unipolar)

To what extent do you agree with the following statement: I support Great Britain's participation in U.S.-led air strikes on Islamic State (IS) in Iraq and Syria.

- Agree completely
- Agree to a large extent
- Agree to a moderate extent
- Agree a little bit
- Agree not at all

Example: Forced choice

When thinking about Great Britain's participation in U.S.-led air strikes on Islamic State (IS) in Iraq and Syria, which of the following comes closer to your opinion:

- Great Britain should participate in air strikes
- Great Britain should not participate in air strikes

Example: Open-ended

In your own words, how would you describe your opinion on Great Britain's participation in U.S.-led air strikes on Islamic State (IS) in Iraq and Syria?

Additional Considerations

- How many response categories?
- Middle category (presence and label)
- "no opinion" and/or "don't know" options
- Probe if "no opinion" or "don't know"?
  - Encourage guessing?
  - Clarify/describe object of evaluation?
- Branching format?
- Order of response categories
- Changes based on survey mode

Common Statistics in Opinion Research

- Counts (tabulation)
- Means
- Proportions
- Correlations
- Differences
- Regression coefficients

Tabulation

- Tabulation is simply the counting of numbers of observed values of a given variable, or the combination of values across multiple variables
- In practice, we almost always translate these to proportions
  - For rare subgroups, we might be interested in counts (e.g., number of drug users, etc.)

Tabulation

- Tabulation is simply the counting of numbers of observed values of a given variable, or the combination of values across multiple variables
- In practice, we almost always translate these to proportions
  - For rare subgroups, we might be interested in counts (e.g., number of drug users, etc.)
- In complex surveys, these tabulations have to be weighted
  - In practice, almost always done using specialized software

Sample mean

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i \tag{1}$$

where y_i = value for a unit, and
n = sample size

$$SE_{\bar{y}} = \sqrt{(1 - f)\frac{s^2}{n}} \tag{2}$$

where f = proportion of population sampled,
s^2 = sample (element) variance, and
n = sample size

Sample proportion

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i \tag{3}$$

where y_i = value for a unit, and
n = sample size

$$SE_{\bar{y}} = \sqrt{\frac{(1-f)}{(n-1)} p(1-p)} \tag{4}$$

where f = proportion of population sampled,
p = sample proportion, and
n = sample size

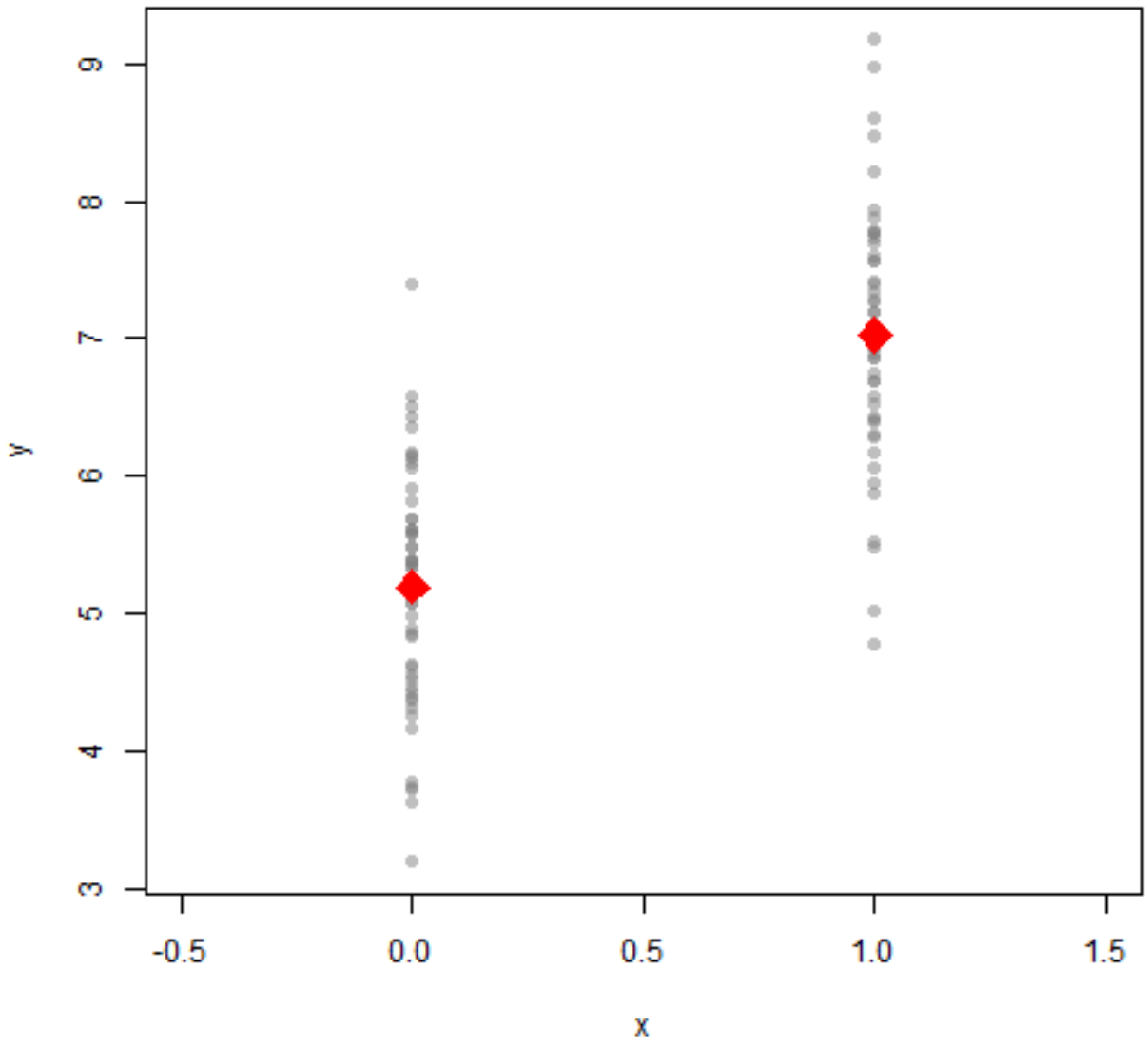
Making comparisons

- Differences of means
- Differences of proportions
- Correlation between variables
- Regression estimates

Mean Difference

- We might expect two means to differ
  - Means for two subgroups (e.g., mean and women)
  - Means across two areas (e.g., England and Scotland)
  - Means at two times (e.g., pre- and post-election)
- We calculate this as a simple difference of subgroup means:

$$\beta = \frac{\sum_{i=1}^{n_1} x_{i1}}{n_1} - \frac{\sum_{i=1}^{n_0} x_{i0}}{n_0}$$


y
x
3
4
5
6
7
8
9
-0.5
0.0
0.5
1.0
1.5

Significance

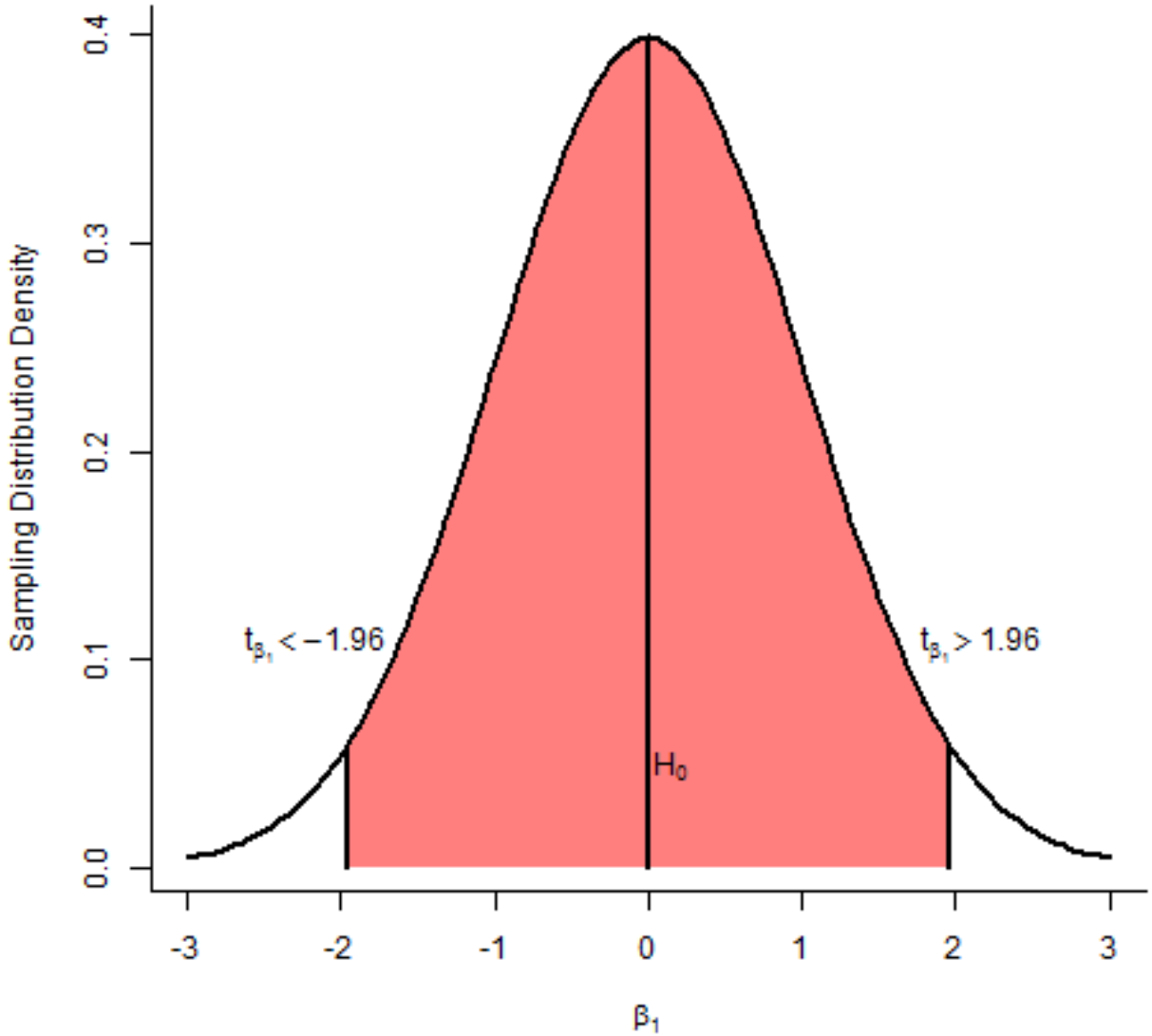
- Is this difference large?
- We can speak about that in two ways:
  - Is the difference substantively large?
  - Is the difference larger than we would expect?

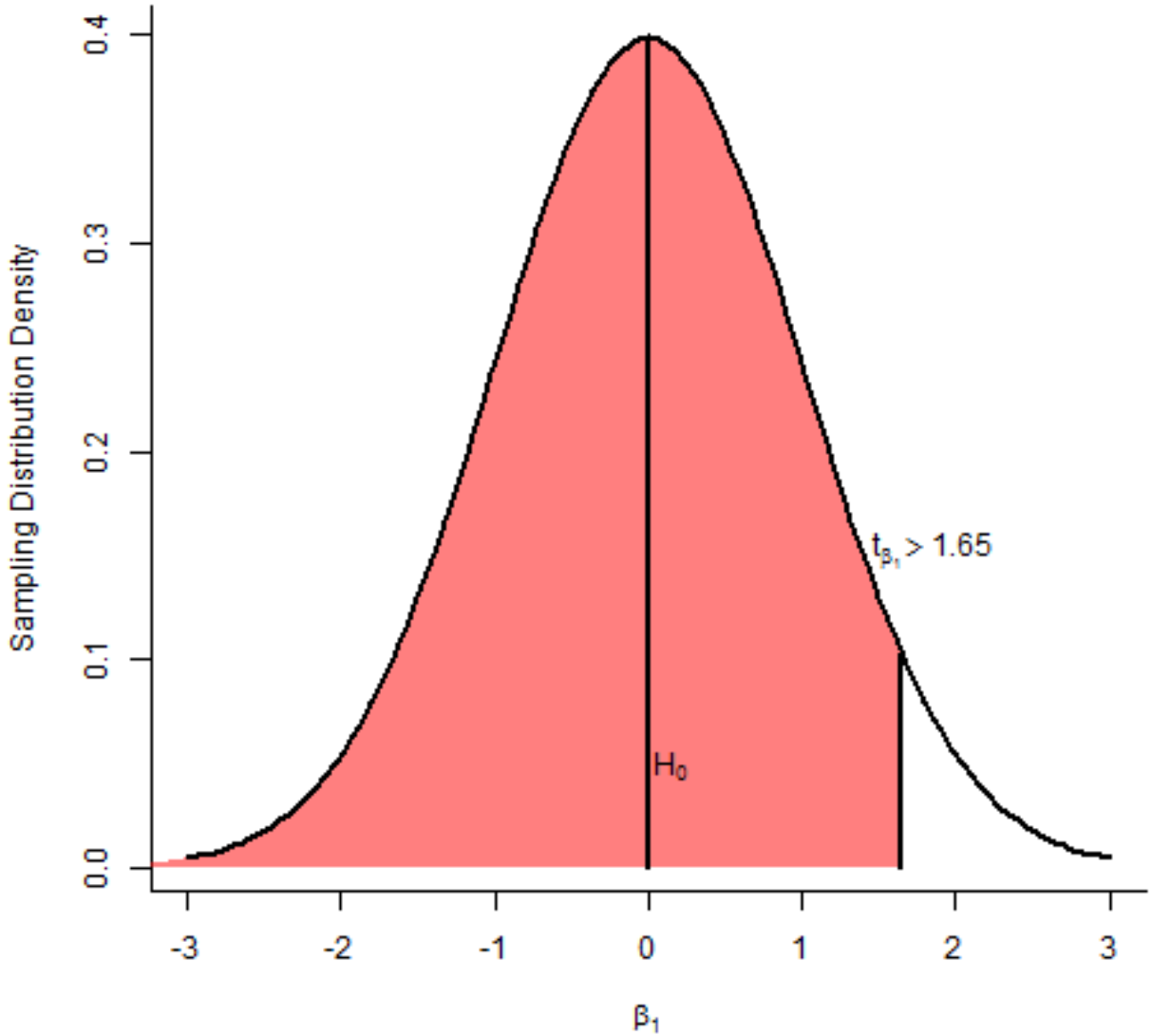
Significance

- Is this difference large?
- We can speak about that in two ways:
  - Is the difference substantively large?
  - Is the difference larger than we would expect?
- To answer the second question requires two pieces of information:
  - A *null* expectation for the difference, H_0
  - An understanding of how likely we are to see a difference this large given that null expectation

t-statistic

1. State a null expectation
   - "No difference" null: $t_{\hat{\beta}_1} = \frac{\hat{\beta}_1}{SE_{\hat{\beta}_1}}$
   - Any other null: $\frac{\hat{\beta}_1 - H_0}{SE_{\hat{\beta}_1}}$, where H_0 is our null hypothesis value
2. Assessment of variability under null comes from SE
   - Inherent (element) variance of our data
   - Sample size
   - Sampling procedures
3. Effect is "statistically significant" if the probability of seeing a t-statistic this large or larger is small


Sampling Distribution Density
0.0
0.1
0.2
0.3
0.4
$t_{\beta_1} < -1.96$
$t_{\beta_1} > 1.96$
H_0
-3
-2
-1
0
1
2
3
β_1


Sampling Distribution Density
0.0
0.1
0.2
0.3
0.4
-3
-2
-1
0
1
2
3
β1
H0
tβ1 > 1.65

We consider a difference "statistically significant" when it differs more from our null expectation than the variation in sample statistics we would expect to observe across repeated samples from a population where our null hypothesis was true (in this case, where there was actually no difference between the groups).

Common Misconceptions

The *p*-value is not:

- The probability that a hypothesis is true or false
- A reflection of our confidence or certainty about the result
- The probability that the true mean is in any particular range of values
- A statement about the importance or substantive size of the effect

Assessing Measurement Quality

1. Conceptual clarity
2. Construct validity
   - Convergent validity
   - Divergent validity
3. Accuracy and precision

Assessing Measures I

- Conceptual clarity is about knowing what we want to measure
- Sloppy concepts make for bad measures
  - Ambiguity
  - Vagueness

Assessing Measures I

- Conceptual clarity is about knowing what we want to measure
- Sloppy concepts make for bad measures
  - Ambiguity
  - Vagueness
- Revise concept definition as needed

Assessing Measures II

- Construct validity is the degree to which a variable measures a concept

Assessing Measures II

- Construct validity is the degree to which a variable measures a concept
- Construct validity is **high** if a variable is a measure of the concept we care about

Assessing Measures II

- Construct validity is the degree to which a variable measures a concept
- Construct validity is **high** if a variable is a measure of the concept we care about
- Construct validity is **low** if a variable is actually a measure of something else

Assessing Construct Validity

- Multiple measures!
- Look for:
  - Convergence (Convergent validity)
  - Discrimination (Discriminant validity)

Assessing Construct Validity

- Multiple measures!
- Look for:
  - Convergence (Convergent validity)
  - Discrimination (Discriminant validity)
- For example, the multi-trait, multi-method matrix

Accuracy and Precision

Synonyms: accurate, true, correct, unbiased, valid

Synonyms: precise, certain, exact, specific, low variance

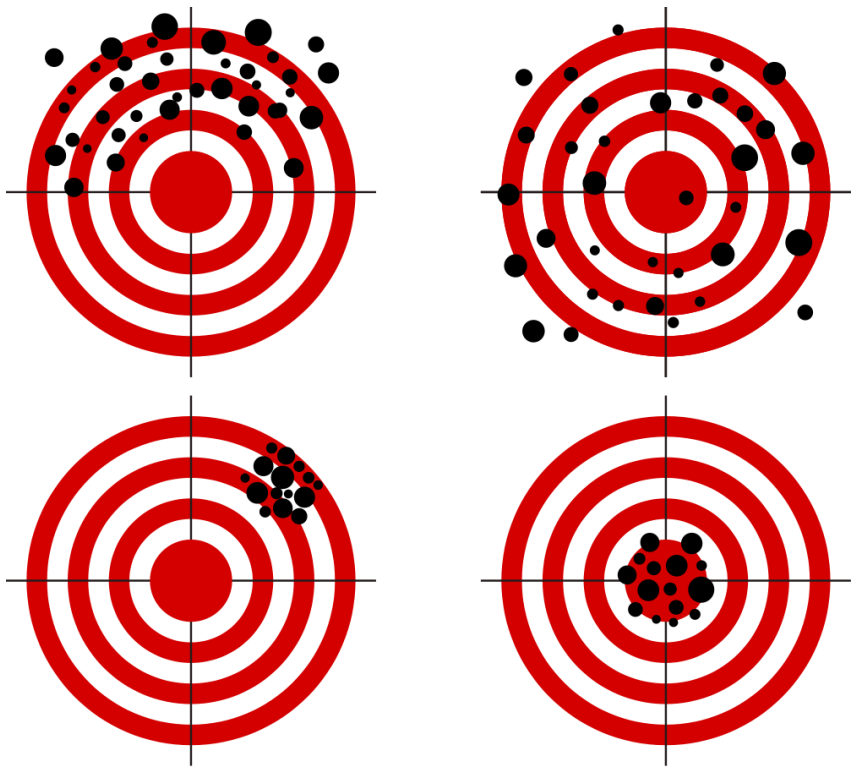


Image Source: Wikimedia, Nevit Dilmen

Scaling

- Improve precision (and accuracy) by combining multiple noisy measures
- Common methods
  - Simple sum or average
  - Factor analysis
  - Item response theory models
- In survey context, typically use prepared, time-optimized batteries of items

Cronbach's α

- Scaling only makes sense if variables "go together"
  - We can assess them *pairwise* by looking at correlations between variables
  - But it's helpful to have a way to assess the scale as a whole
- Definition: $\alpha = \dfrac{N\bar{c}}{\bar{v} + (N-1)\bar{c}}$
  - N: number of items
  - $\bar{c}$: average covariance of items
  - $\bar{v}$: average variance of items

Hierarchical Data Collection

- Sampling is just a building block
- Various elements can be combined:
  - Repeated, independent sampling
  - Repeated interviewing
  - Parallel sampling w/in different contexts
  - Experimentation (see Lecture 3)

Hierarchical Data Collection

- Sampling is just a building block
- Various elements can be combined:
  - Repeated, independent sampling
  - Repeated interviewing
  - Parallel sampling w/in different contexts
  - Experimentation (see Lecture 3)
- Big question: How do analyze complex data structures?

Research Questions

- Sampling allows us to make claims about populations
- Hierarchical data collection allows us to answer more complex research questions
  - How does a population change?
  - How do populations compare (across space and time)?
  - How do individuals change?
  - How are multiple factors related? How do those relationships differ across space, time, geography, and individuals?
- Not going to talk about qualitative or case comparisons

Repeated Cross-Section

- Simplest elaboration of a single survey
- Draw multiple, independent samples at different points in time
  - Note: in small populations, this can be complicated
- Useful for:
  - Establishing trends/changes
  - Comparing multivariate relationships over time

Example: Election Polling

- Commonly used to study elections
- Highly inconsistent procedures
  - Varying sample sizes
  - Varying sampling procedures
  - Varying question wordings
  - Irregular field periods
- Some examples:
  - Pollster
  - Lord Ashcroft Polls

Example: BES CMS

- Continuous Monitoring Study
- Procedures
  - Monthly interviewing 2004–Present
  - Modest sample sizes
  - Very short questionnaire
- Inspired "rolling thunder" study of 2015 campaign
- Data: `http://www.bes2009-10.org/cms-data.php`

Practical Considerations

Practical Considerations

- Are patterns of nonresponse consistent across periods?

Practical Considerations

- Are patterns of nonresponse consistent across periods?
- What happens if we want/need to change questions?

Practical Considerations

- Are patterns of nonresponse consistent across periods?
- What happens if we want/need to change questions?
- What happens if we want/need to change sampling methods?

Practical Considerations

- Are patterns of nonresponse consistent across periods?
- What happens if we want/need to change questions?
- What happens if we want/need to change sampling methods?
- What if we want to draw individual-level comparisons?

Panel Studies

- A panel is a single cohort, observed at multiple points in time
- Synonyms: Within-subjects studies, longitudinal studies, cohort studies
- First use: Lazarsfeld and Fiske (1938)

Ex: "Classic" Election Studies

- "Columbia School": Lazarsfeld et al. (1948) and Berelson et al. (1954)
- Empaneled small samples (less than n=1000)
- Measured vote intentions and other factors multiple times
- Largely found that campaigns didn't matter

Ex: Modern Election Studies

- American National Election Studies
  - Collected biennially (or more) since 1948
  - Questionnaire is basically unchanged
  - Most years have pre/post interviews
  - Profound source of data, survey innovations, & experiments
  - `http://electionstudies.org/`
- National Annenberg Election Survey
  - Five-wave panels during 2004 and 2008 U.S. Presidential elections
  - Online and telephone interviewing
  - Intended to mimic "classic" studies
  - Click for more info

Trade-offs

Trade-offs

- Advantages
  - Within-subjects comparisons
  - Statistical efficiency
  - Cost?

Trade-offs

- Advantages
  - Within-subjects comparisons
  - Statistical efficiency
  - Cost?
- Disadvantages
  - Initial recruitment
  - Attrition (and incentivization)
  - Panel conditioning
  - Cost?

Statistical Advantages I

- Independent samples[1]

$$\beta = \frac{\sum_{i=1}^{n_1} x_{i1}}{n_1} - \frac{\sum_{i=1}^{n_0} x_{i0}}{n_0}$$

$$SE_\beta = \sqrt{\frac{s_1^2}{n_1} + \frac{s_0^2}{n_0}}$$

[1] Note: We *do not* actually need the raw data to calculate this.

Statistical Advantages II

- Paired samples[2]

$$\beta = \frac{\sum_{i=1}^{n}(x_{i1} - x_{i0})}{n}$$

$$SE_{\beta} = \sqrt{\frac{{s_{Diff}}^2}{n}}$$

- Standard error for paired samples decreases as correlation between t_0 and t_1 observations increases.

[2] Note: We *do* need the raw data to calculate this.

Panel Variants

- Rolling panel
- Replenishment samples
- Subpanels
- “Online panels”

Multi-Level Data

- Multi-level data collection involves sampling within different geographical units
- Depending on research goal, this can be thought of as either stratified sampling or cluster sampling, or both
  - Stratified sampling: Geography correlates with variables of interest and we desire large, representative samples within each geographical stratum
  - Cluster sampling: Goal is to make claims about "super population" and units are necessary for interviewing procedures

Ex: European Social Survey

- Biennial cross-sectional survey conducted since 2001
  - Currently round 8 in field
- Implemented in parallel across most European countries
- Extremely high quality
  - CAPI interviewing
  - Probability based samples
  - Precise translation
- `http://www.europeansocialsurvey.org/`

Ex: World Values Survey

- Periodic survey since 1985
  - Currently planning Wave 8 for 2016–2018
  - Long field period (2+ years)
  - Approximately 100 countries
- Well-conducted
  - CAPI or CATI interviewing
  - Mostly probability samples
  - More cross-national protocol variations than ESS
- `http://www.worldvaluessurvey.org/wvs.jsp`

Uses of Regression

1. Description
2. Prediction
3. Causal Inference

Descriptive Inference

1. We want to understand a *population* of cases
2. We cannot observe them all, so:
    1. Draw a *representative* sample
    2. Perform mathematical procedures on sample data
    3. Use assumptions to make inferences about population
    4. Express uncertainty about those inferences based on assumptions

Parameter Estimation

- We want to observe population *parameter* θ
- If we obtain a representative sample of population units:
  - Our sample statistic $\hat{\theta}$ is an unbiased estimate of θ
  - Our sampling procedure dictates how uncertain we are about the value of θ

Causal Inference

Causal Inference

1. Everything that goes into descriptive inference

Causal Inference

1. Everything that goes into descriptive inference
2. Plus, philosophical assumptions

Causal Inference

1. Everything that goes into descriptive inference
2. Plus, philosophical assumptions
3. Plus, randomization *or* perfectly specified model

Relationship

- Covariance:

$$Cov(X, Y) = \sum_{i=1}^{n} \frac{(X_i - \bar{X})(Y_i - \bar{Y})}{n - 1}$$

Relationship

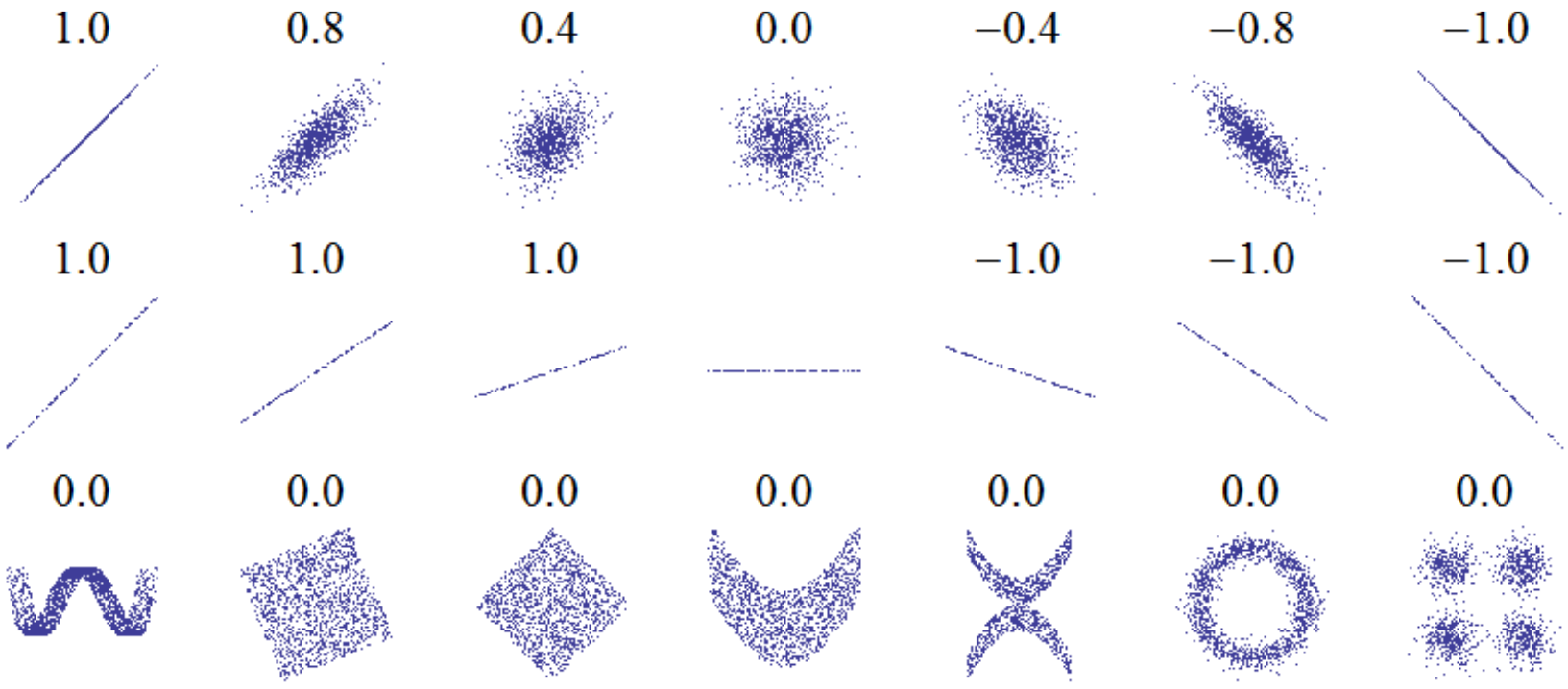
- Covariance:

$$Cov(X, Y) = \sum_{i=1}^{n} \frac{(X_i - \bar{X})(Y_i - \bar{Y})}{n - 1}$$

- Pearson's Correlation:

$$Corr(X, Y) = r_{x,y} = \sum_{i=1}^{n} \frac{(X_i - \bar{X})(Y_i - \bar{Y})}{(n - 1)s_x s_y}$$

Correlation is linear!



Source: Wikimedia

Analyzing Complex Surveys

- There's a saying: "Every simple random survey is simple in the same way, but every complex survey is complex in its own way."

Analyzing Complex Surveys

- There's a saying: "Every simple random survey is simple in the same way, but every complex survey is complex in its own way."
- Statistics courses will almost always assume simple random sampling

Analyzing Complex Surveys

- There's a saying: "Every simple random survey is simple in the same way, but every complex survey is complex in its own way."
- Statistics courses will almost always assume simple random sampling
- Any sample that is not self-weighting requires more complicated *estimators* that account for varying weights

Analyzing Complex Surveys

- There's a saying: "Every simple random survey is simple in the same way, but every complex survey is complex in its own way."
- Statistics courses will almost always assume simple random sampling
- Any sample that is not self-weighting requires more complicated *estimators* that account for varying weights
- Don't try to do this by hand
  - Stata `svy` module
  - R `survey` package

Ways of Thinking About OLS

Ways of Thinking About OLS

1. Estimating Unit-level Causal Effect

Ways of Thinking About OLS

1. Estimating Unit-level Causal Effect
2. Ratio of $Cov(X, Y)$ and $Var(X)$

Ways of Thinking About OLS

1. Estimating Unit-level Causal Effect
2. Ratio of $Cov(X, Y)$ and $Var(X)$
3. Minimizing residual sum of squares (SSR)

Ways of Thinking About OLS

1. Estimating Unit-level Causal Effect
2. Ratio of $Cov(X, Y)$ and $Var(X)$
3. Minimizing residual sum of squares (SSR)
4. Line (or surface) of best fit

Bivariate Regression I

- Y is continuous
- X is a randomized treatment indicator/dummy $(0, 1)$
- How do we know if the treatment X had an effect on Y?

Bivariate Regression I

- Y is continuous
- X is a randomized treatment indicator/dummy $(0, 1)$
- How do we know if the treatment X had an effect on Y?
- Look at mean-difference: $E[Y_i|X_i = 1] - E[Y_i|X_i = 0]$

Three Equations

1. Population: $Y = \beta_0 + \beta_1 X \ (+\epsilon)$
2. Sample estimate: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$
3. Unit:

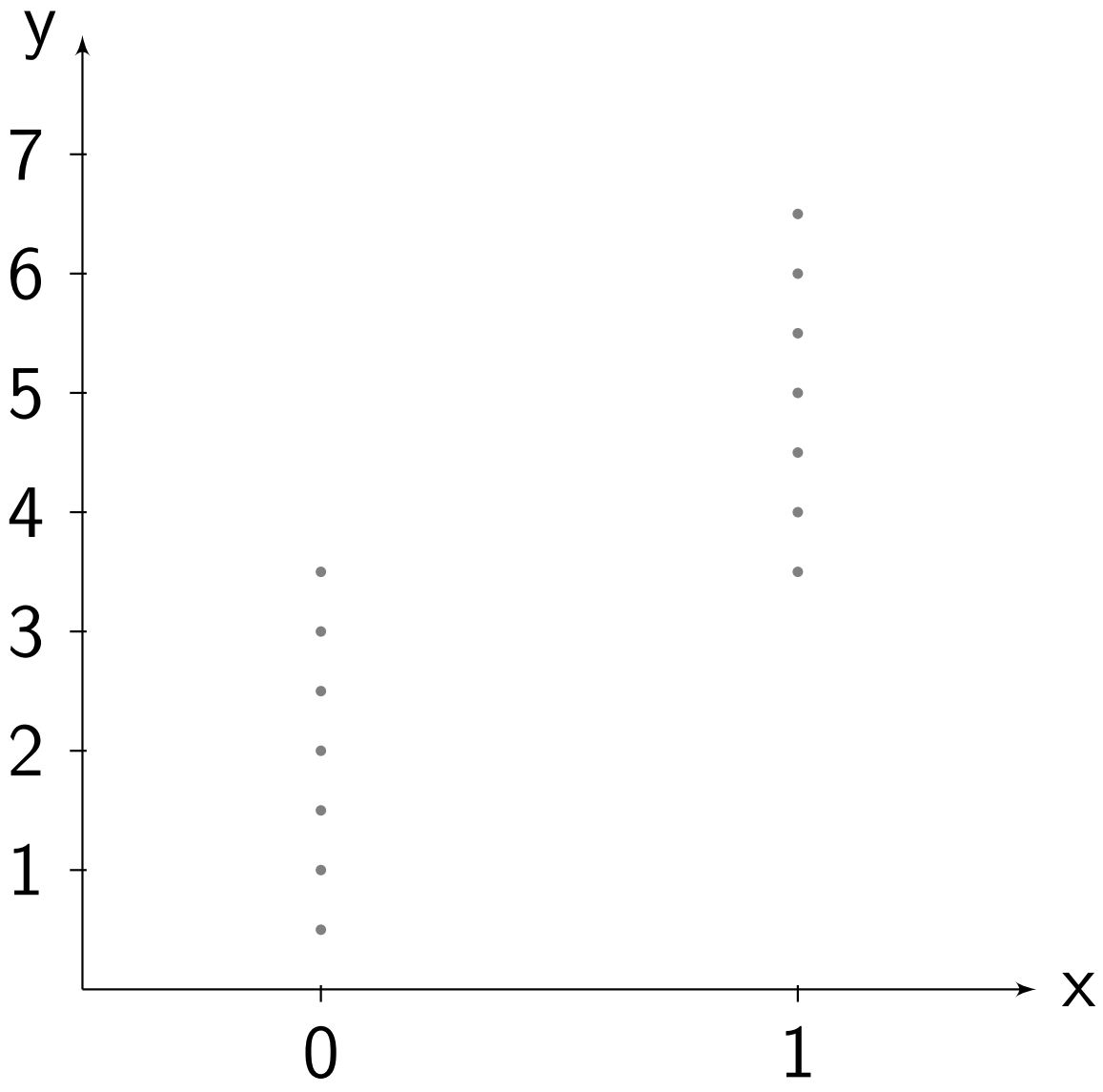
$$\begin{aligned} y_i &= \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i \\ &= \bar{y}_{0i} + (y_{1i} - y_{0i})x_i + (y_{0i} - \bar{y}_{0i}) \end{aligned}$$

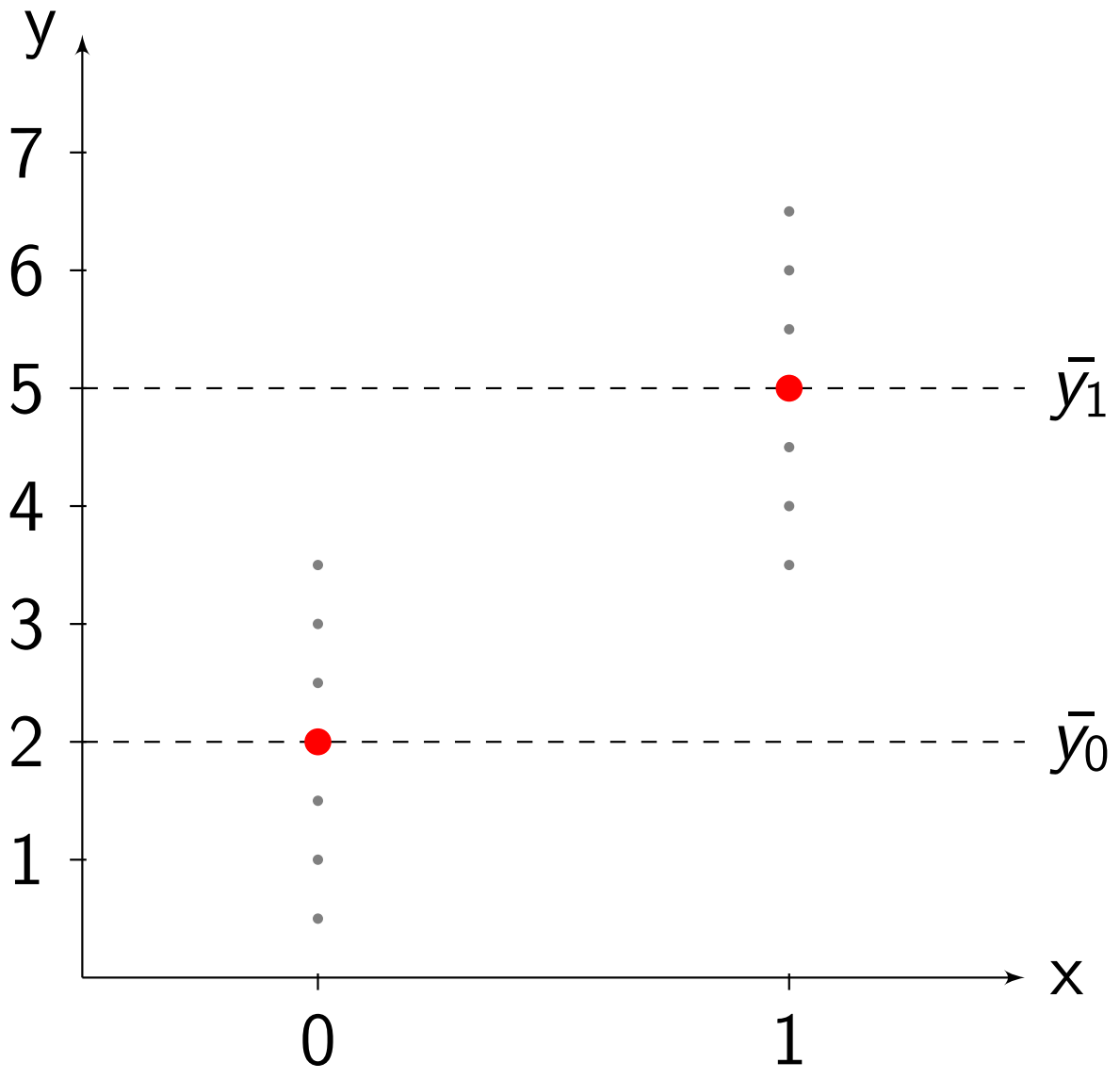
Bivariate Regression I

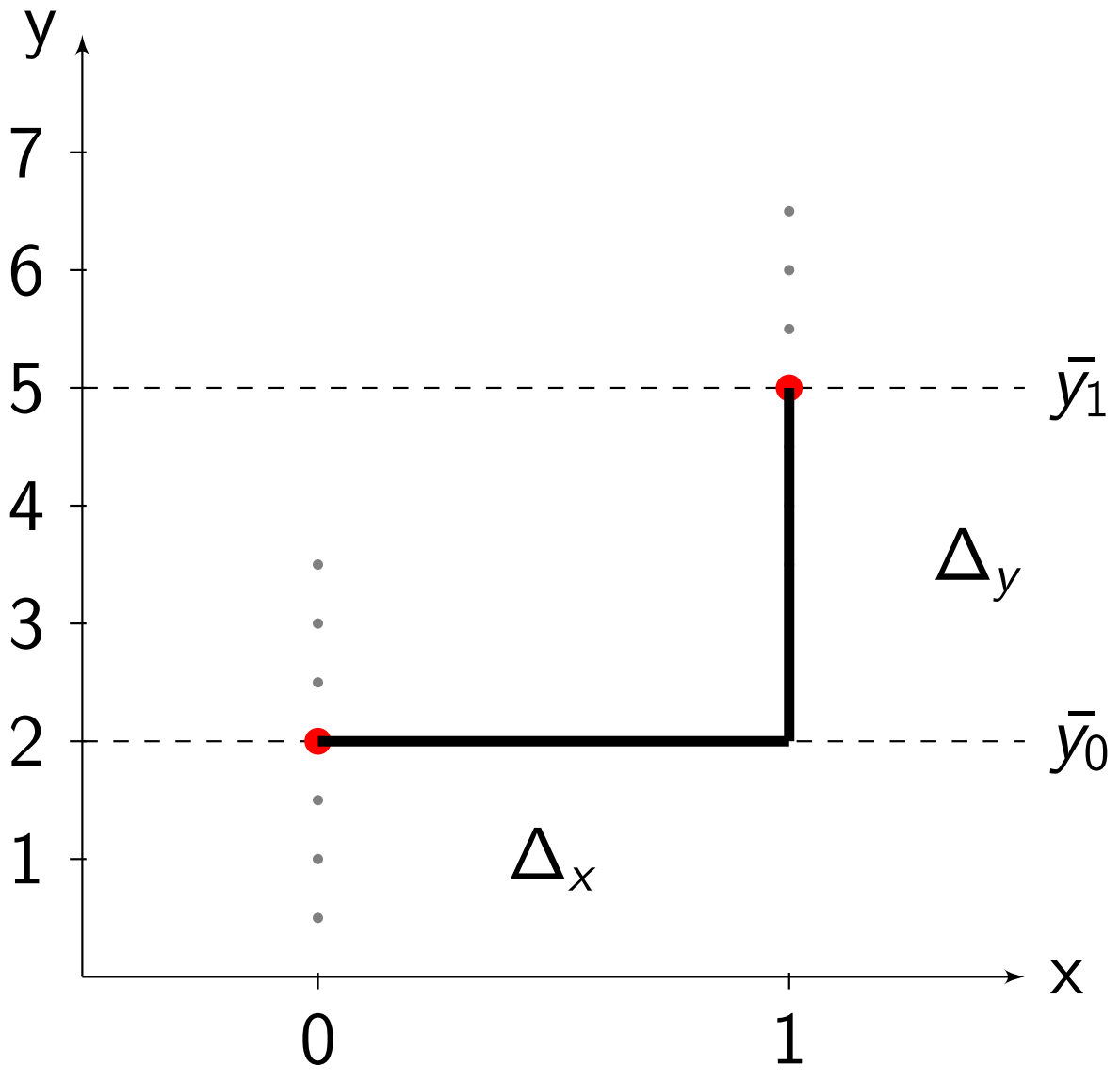
- Mean difference ($E[Y_i|X_i = 1] - E[Y_i|X_i = 0]$) is the regression line slope
- Slope (β) defined as $\frac{\Delta Y}{\Delta X}$

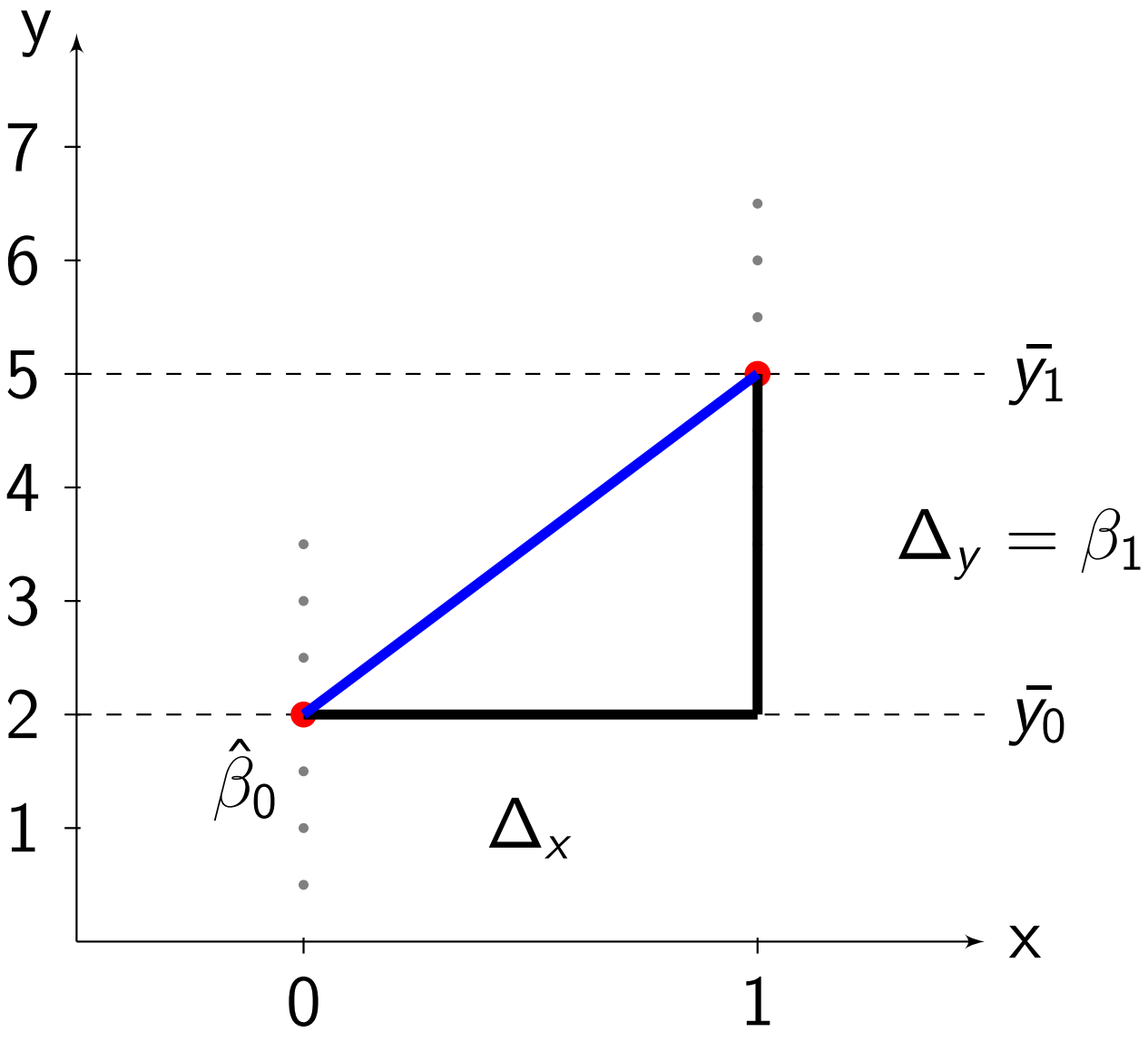
Bivariate Regression I

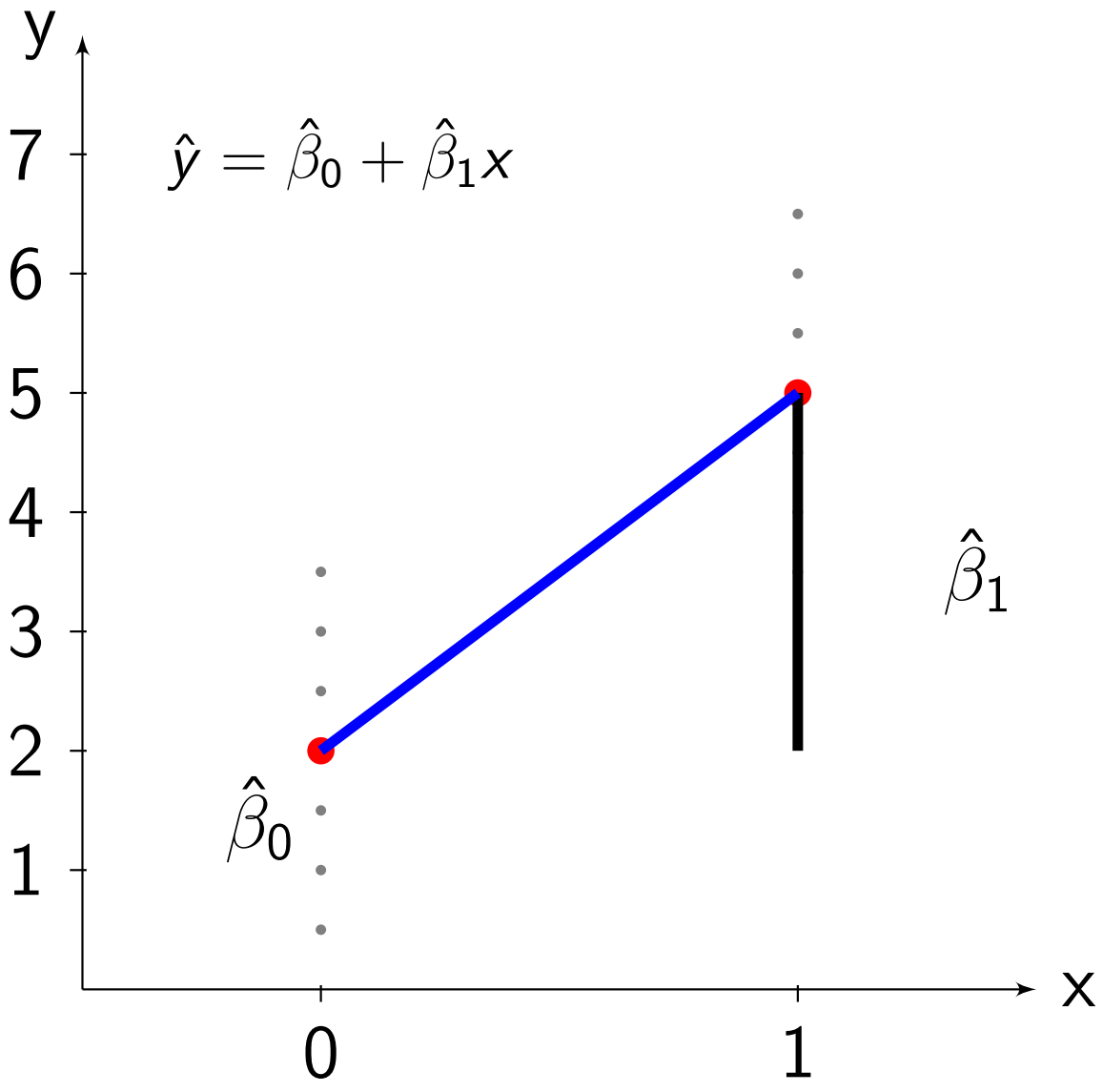
- Mean difference ($E[Y_i|X_i = 1] - E[Y_i|X_i = 0]$) is the regression line slope
- Slope (β) defined as $\frac{\Delta Y}{\Delta X}$
  - $\Delta Y = E[Y_i|X = 1] - E[Y_i|X = 0]$
  - $\Delta X = 1 - 0 = 1$

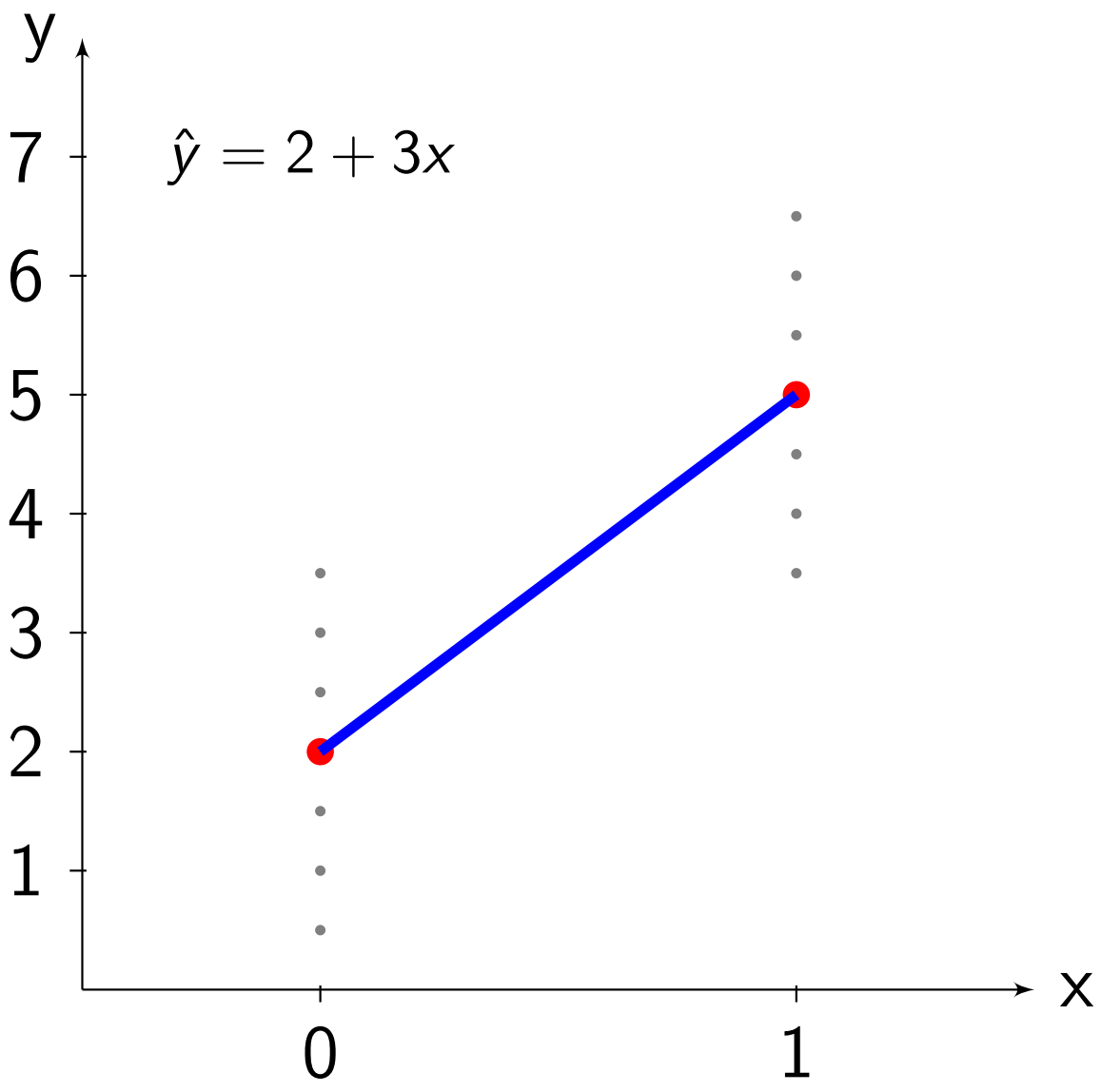

y
7
6
5
4
3
2
1
x
0
1

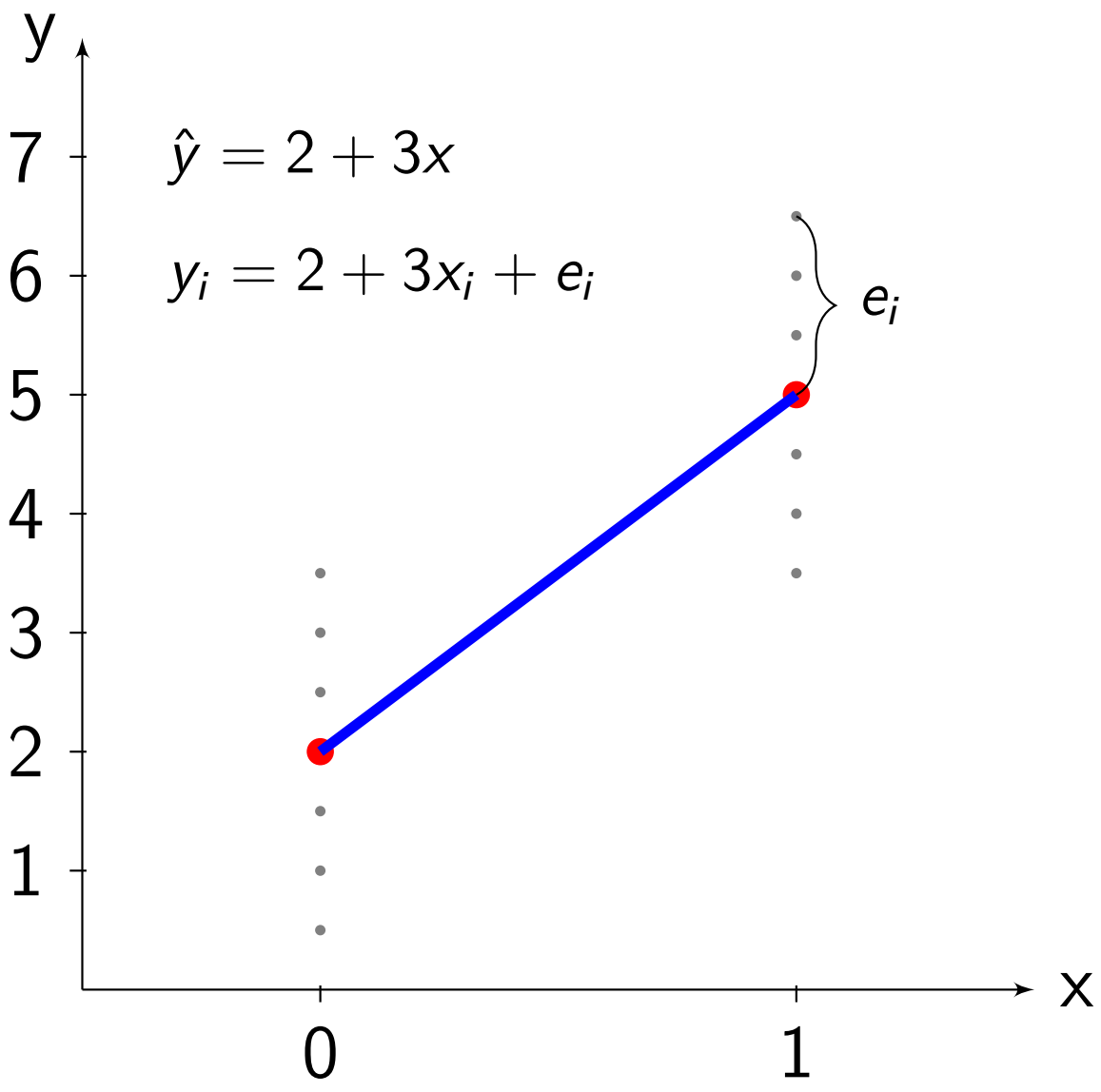

y
7
6
5
4
3
2
1
$\bar{y}_1$
$\bar{y}_0$
x
0
1


y
7
6
5
4
3
2
1
$\bar{y}_1$
$\bar{y}_0$
Δ_y
Δ_x
x
0
1


y
7
6
5
4
3
2
1
$\bar{y}_1$
$\Delta_y = \beta_1$
$\bar{y}_0$
$\hat{\beta}_0$
Δ_x
x
0
1


y
7
6
5
4
3
2
1
$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$
$\hat{\beta}_0$
$\hat{\beta}_1$
x
0
1


y
7
6
5
4
3
2
1
$\hat{y} = 2 + 3x$
x
0
1


y
7
6
5
4
3
2
1
$\hat{y} = 2 + 3x$
$y_i = 2 + 3x_i + e_i$
e_i
x
0
1

Systematic versus unsystematic component of the data

- Systematic: Regression line (slope)
  - Linear regression estimates the conditional means of the population data (i.e., $E[Y|X]$)
- Unsystematic: Error term is the deviation of observations from the line
  - The difference between each value y_i and $\hat{y}_i$ is the *residual*: e_i
  - OLS produces an estimate of the relationship between X and Y that minimizes the *residual sum of squares*

Why are there residuals?

Why are there residuals?

- Omitted variables
- Measurement error
- Fundamental randomness

Bivariate Regression I

- Mean difference ($E[Y_i|X_i = 1] - E[Y_i|X_i = 0]$) is the regression line slope
- Slope (β) defined as $\frac{\Delta Y}{\Delta X}$
  - $\Delta Y = E[Y_i|X = 1] - E[Y_i|X = 0]$
  - $\Delta X = 1 - 0 = 1$

Bivariate Regression I

- Mean difference ($E[Y_i|X_i = 1] - E[Y_i|X_i = 0]$) is the regression line slope
- Slope (β) defined as $\frac{\Delta Y}{\Delta X}$
  - $\Delta Y = E[Y_i|X = 1] - E[Y_i|X = 0]$
  - $\Delta X = 1 - 0 = 1$
- How do we know if this is a *significant* difference?
  - We'll come back to that

Ways of Thinking About OLS

1. Estimating Unit-level Causal Effect

Ways of Thinking About OLS

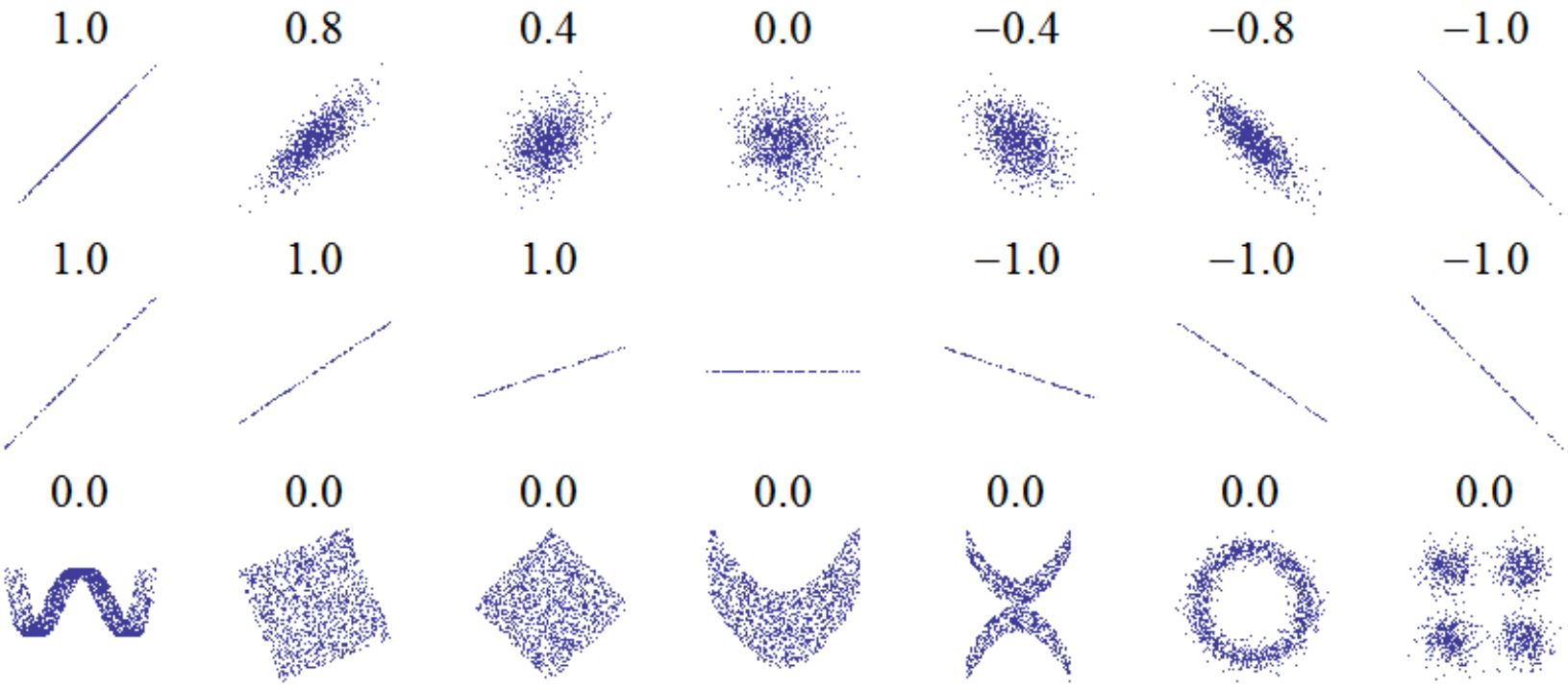
1. Estimating Unit-level Causal Effect
2. Ratio of $Cov(X, Y)$ and $Var(X)$

Bivariate Regression II

- Y is continuous
- X is continuous (and randomized)
- How do we know if the treatment X had an effect on Y?
  - Correlation coefficient (ρ)
  - Regression coefficient (slope; β_1)

Correlation Coefficient (ρ)

- Measures how well a scatterplot is represented by a straight (non-horizontal) line


1.0
0.8
0.4
0.0
−0.4
−0.8
−1.0
1.0
1.0
1.0
−1.0
−1.0
−1.0
0.0
0.0
0.0
0.0
0.0
0.0
0.0

Correlation Coefficient (ρ)

- Measures how well a scatterplot is represented by a straight (non-horizontal) line

Correlation Coefficient (ρ)

- Measures how well a scatterplot is represented by a straight (non-horizontal) line
- Formal definition: $\frac{Cov(X,Y)}{\sigma_X \sigma_y}$
- As a reminder:
  - $Cov(x,y) = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$
  - $s_x = \sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}$

OLS Coefficient $(\beta_1)^3$

- Measures ΔY given ΔX

[3] Multivariate formula involves matrices; Week 20

OLS Coefficient $(\beta_1)^3$

- Measures ΔY given ΔX
- Formal definition: $\frac{Cov(X,Y)}{Var(X)}$
- As a reminder:
  - $Cov(x,y) = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$
  - $Var(x) = \sum_{i=1}^{n}(x_i - \bar{x})^2$

[3] Multivariate formula involves matrices; Week 20

OLS Coefficient $(\beta_1)^3$

- Measures ΔY given ΔX
- Formal definition: $\frac{Cov(X,Y)}{Var(X)}$
- As a reminder:
  - $Cov(x,y) = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$
  - $Var(x) = \sum_{i=1}^{n}(x_i - \bar{x})^2$
- $\hat{\rho}$ and $\hat{\beta}_1$ are just scaled versions of $\widehat{Cov}(x,y)$

[3] Multivariate formula involves matrices; Week 20

Minimum Mathematical Requirements

1. Do we need variation in X?

Minimum Mathematical Requirements

1. Do we need variation in X?
   - Yes, otherwise dividing by zero

Minimum Mathematical Requirements

1. Do we need variation in X?
   - Yes, otherwise dividing by zero
2. Do we need variation in Y?
   - No, $\hat{\beta}_1$ can equal zero

Minimum Mathematical Requirements

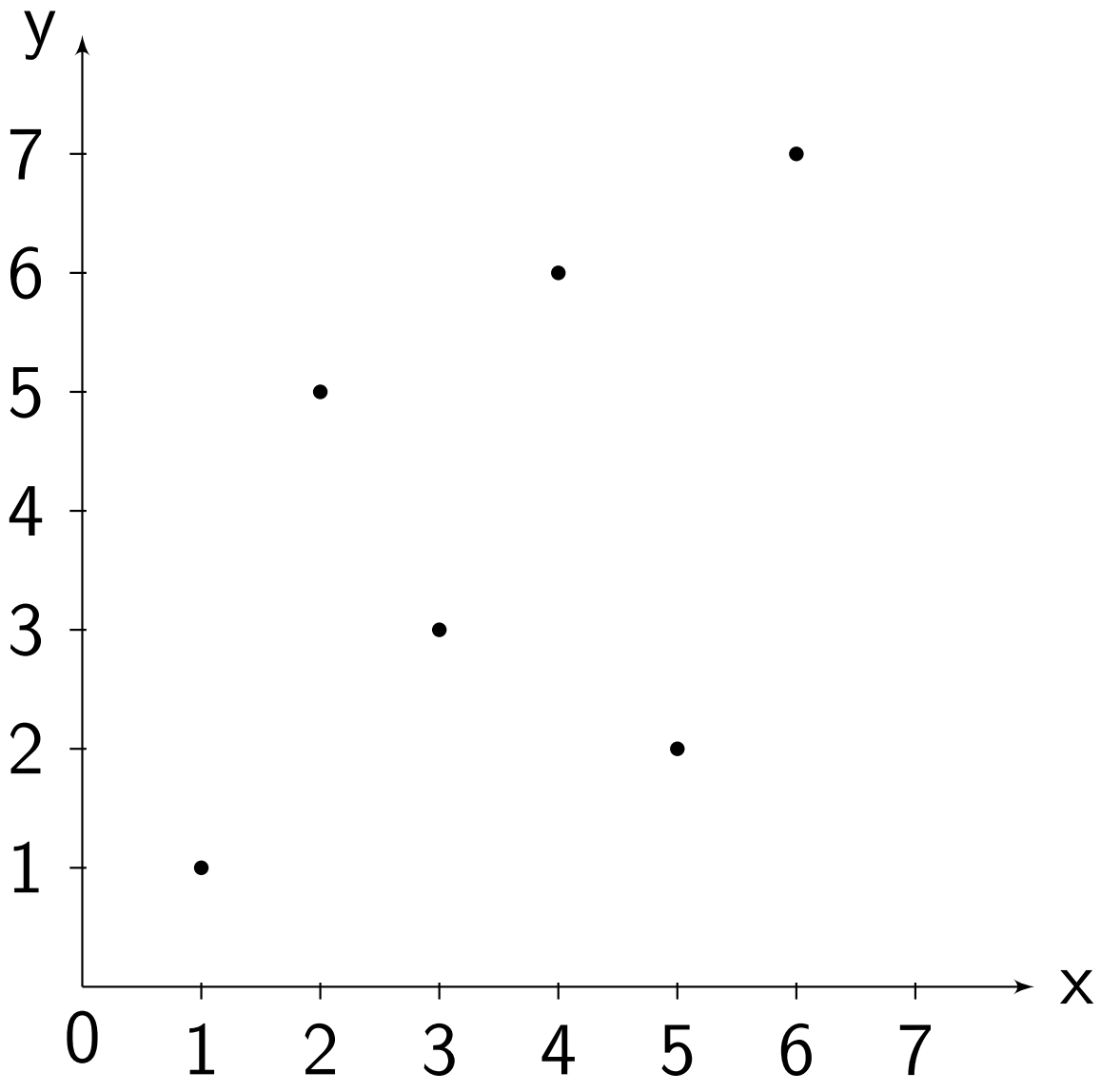
1. Do we need variation in X?
   - Yes, otherwise dividing by zero
2. Do we need variation in Y?
   - No, $\hat{\beta}_1$ can equal zero

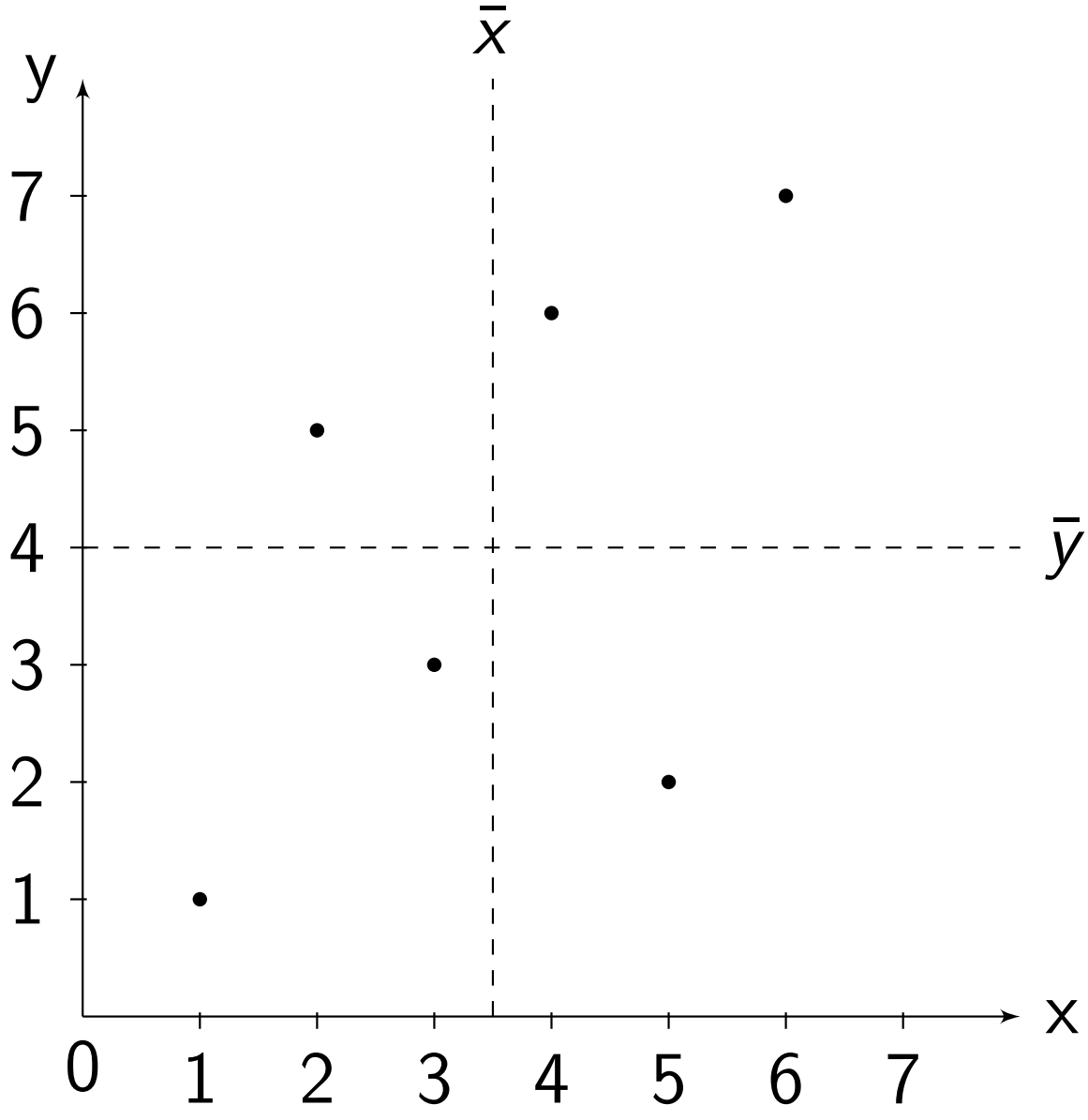
Minimum Mathematical Requirements

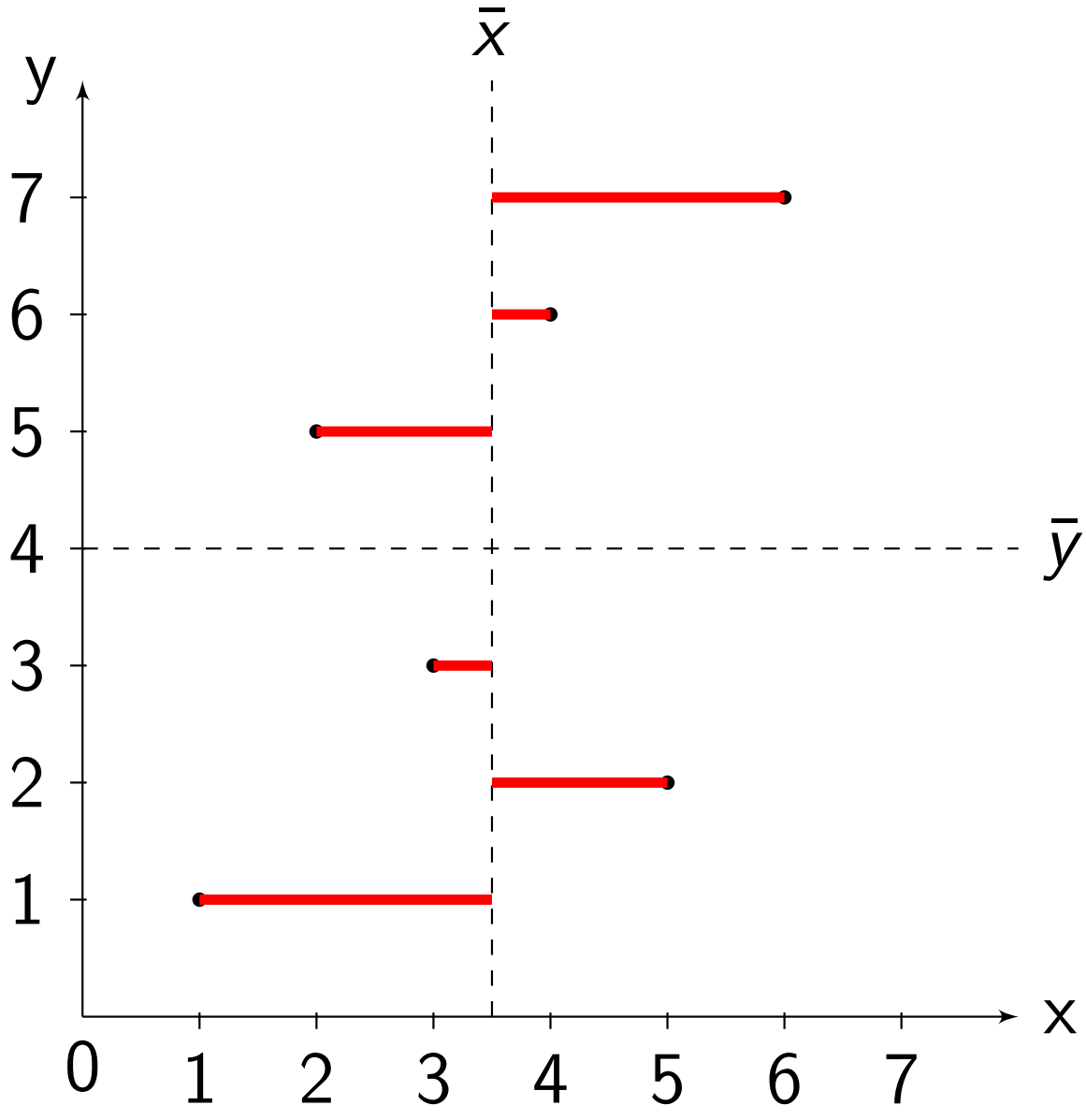
1. Do we need variation in X?
   - Yes, otherwise dividing by zero
2. Do we need variation in Y?
   - No, $\hat{\beta}_1$ can equal zero
3. How many observations do we need?

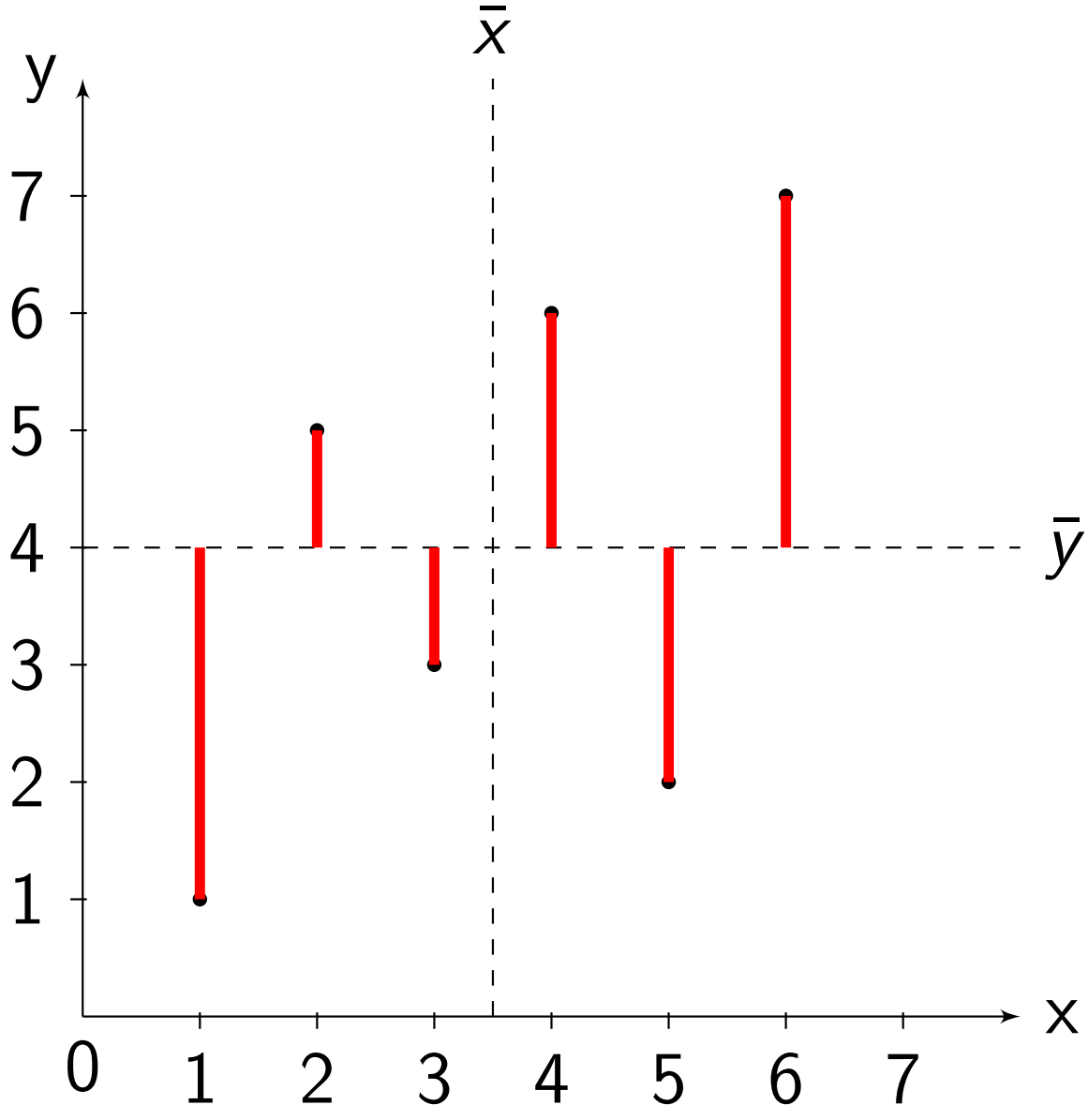
Minimum Mathematical Requirements

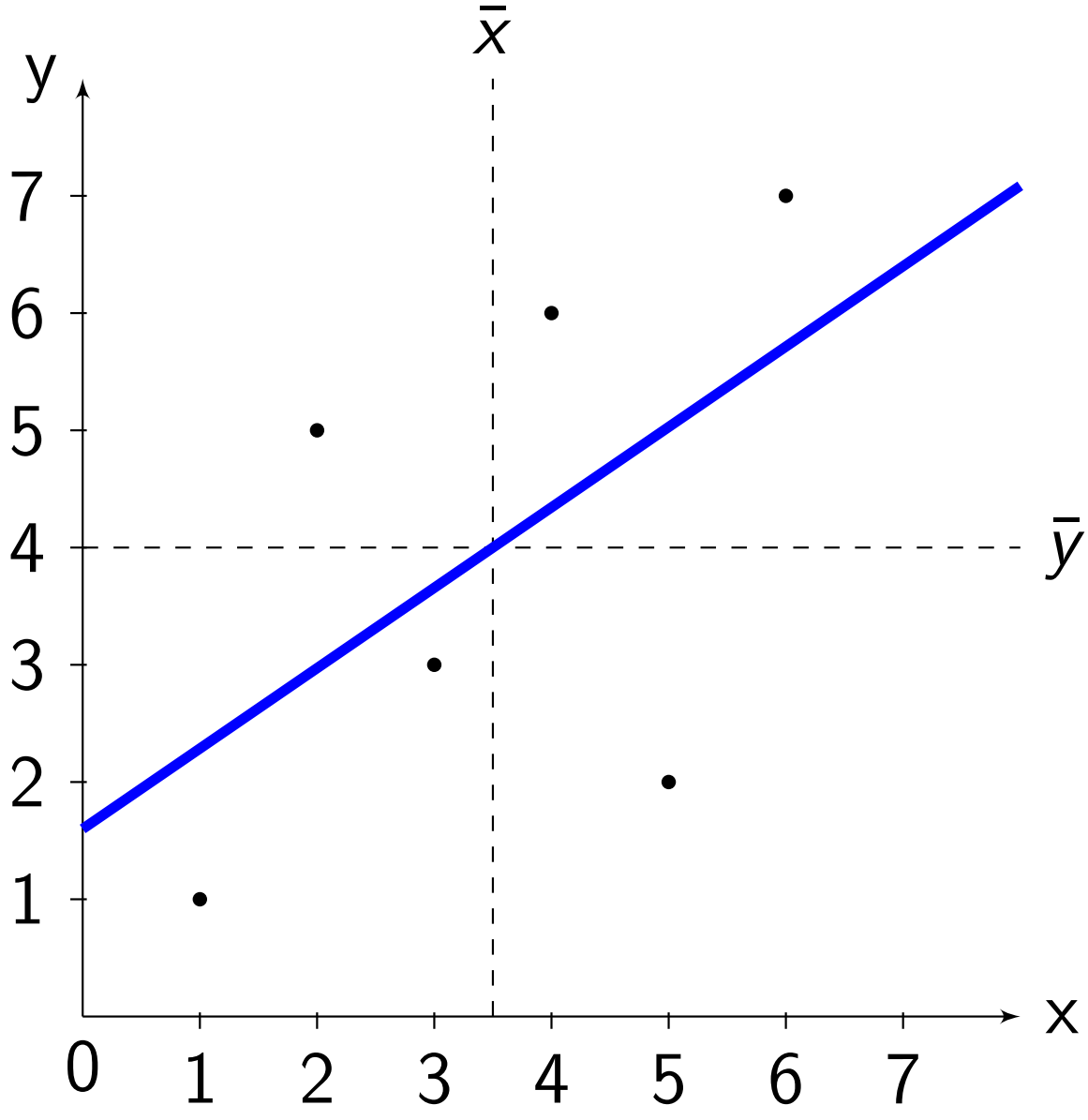
1. Do we need variation in X?
   - Yes, otherwise dividing by zero
2. Do we need variation in Y?
   - No, $\hat{\beta}_1$ can equal zero
3. How many observations do we need?
   - $n \geq k$, where k is number of parameters to be estimated


y
7
6
5
4
3
2
1
x
0
1
2
3
4
5
6
7


$\bar{x}$
y
7
6
5
4
3
2
1
$\bar{y}$
x
0
1
2
3
4
5
6
7


$\bar{x}$
y
7
6
5
4
3
2
1
$\bar{y}$
x
0
1
2
3
4
5
6
7


$\bar{x}$
y
7
6
5
4
3
2
1
$\bar{y}$
x
0
1
2
3
4
5
6
7


$\bar{x}$
y
7
6
5
4
3
2
1
$\bar{y}$
x
0
1
2
3
4
5
6
7

Calculations

x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$
1	1	?	?	?	?
2	5	?	?	?	?
3	3	?	?	?	?
4	6	?	?	?	?
5	2	?	?	?	?
6	7	?	?	?	?

Intercept $\hat{\beta}_0$

- Simple formula: $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$

Intercept $\hat{\beta}_0$

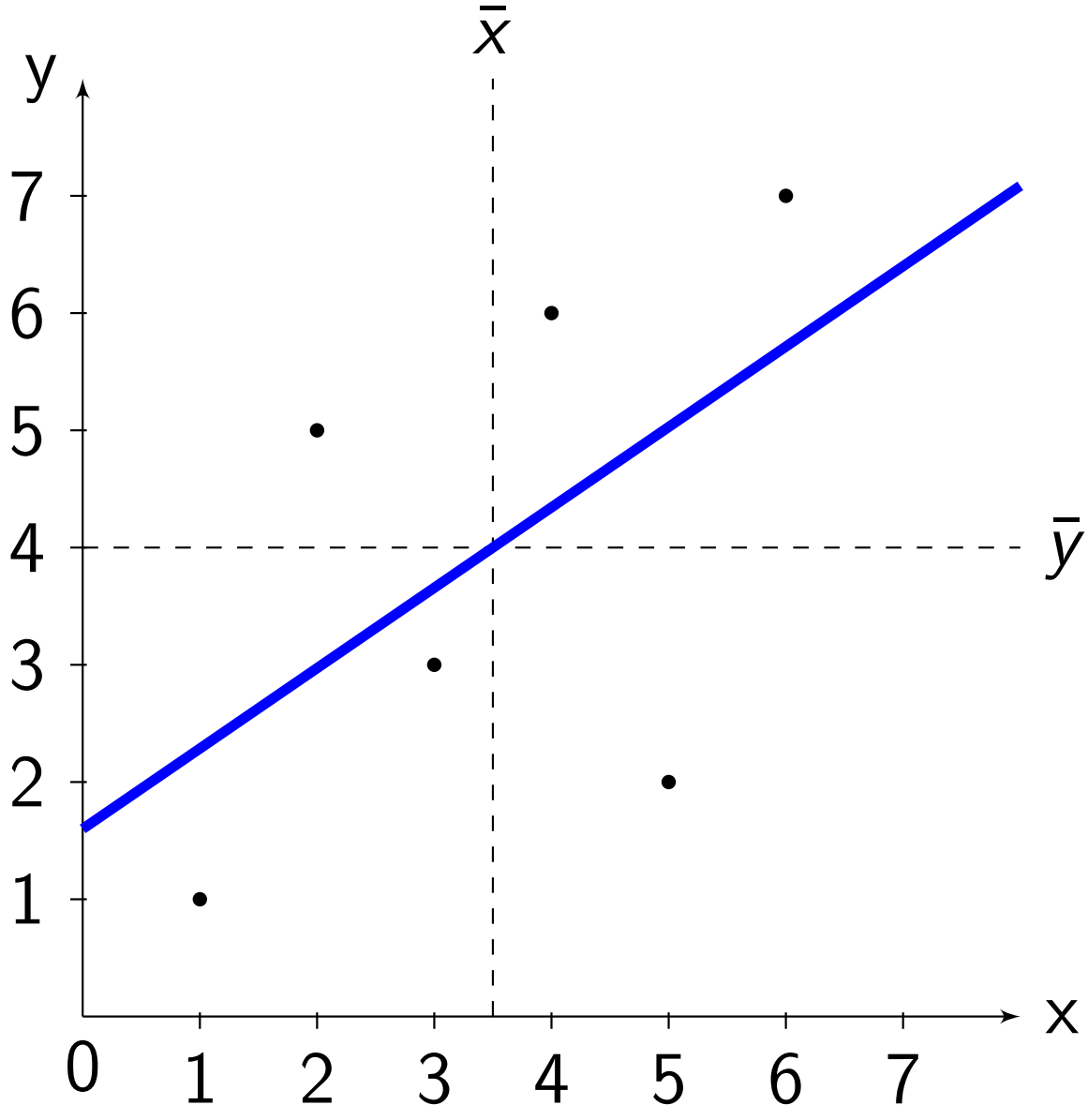
- Simple formula: $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$
- Intuition: OLS fit always runs through point $(\bar{x}, \bar{y})$

Intercept $\hat{\beta}_0$

- Simple formula: $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$
- Intuition: OLS fit always runs through point $(\bar{x}, \bar{y})$
- Ex.: $\hat{\beta}_0 = 4 - 0.6857 * 3.5 = 1.6$

Intercept $\hat{\beta}_0$

- Simple formula: $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$
- Intuition: OLS fit always runs through point $(\bar{x}, \bar{y})$
- Ex.: $\hat{\beta}_0 = 4 - 0.6857 * 3.5 = 1.6$
- $\hat{y} = 1.6 + 0.6857\hat{x}$


$\bar{x}$
y
7
6
5
4
3
2
1
$\bar{y}$
x
0
1
2
3
4
5
6
7

Ways of Thinking About OLS

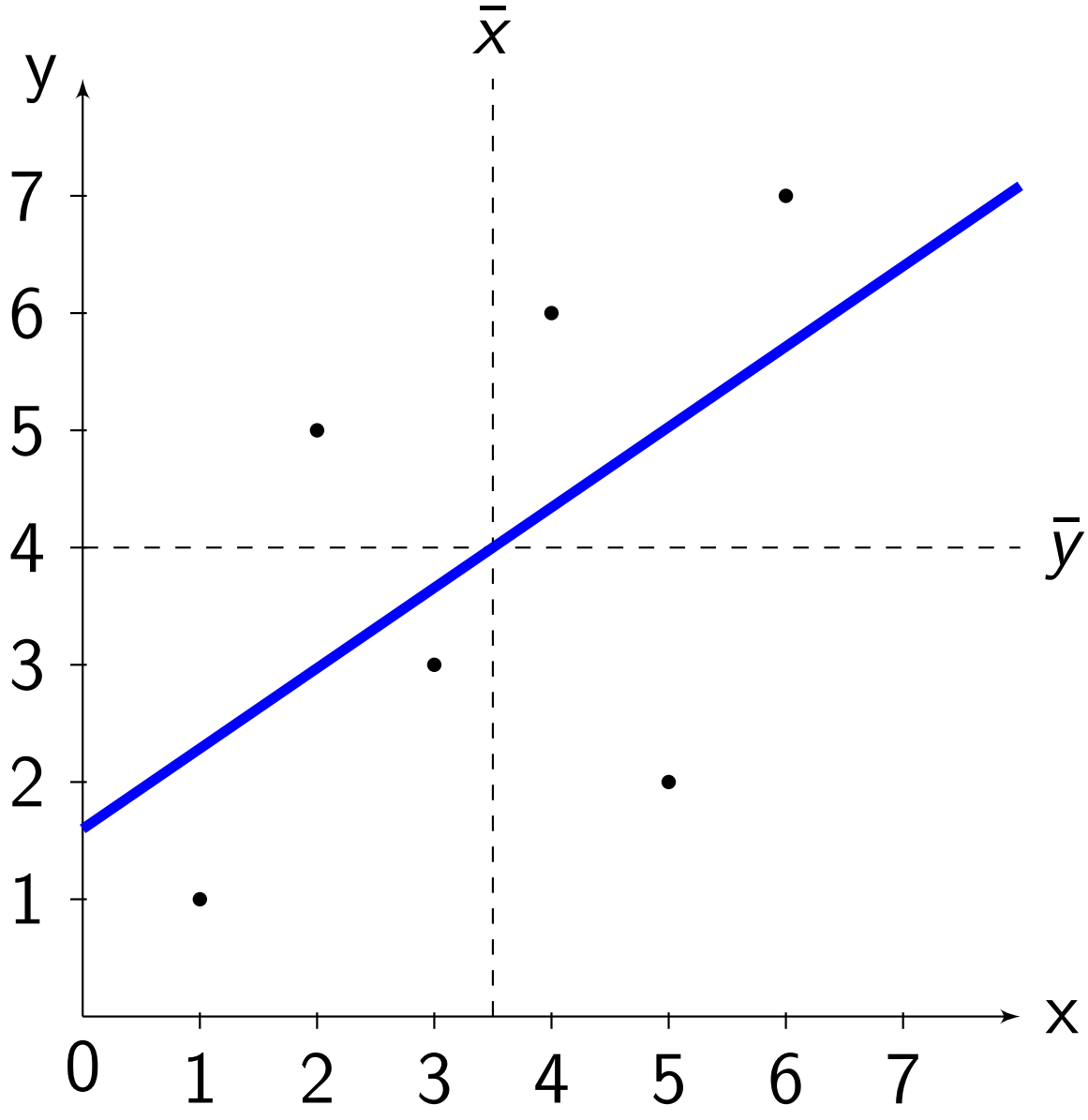
1. Estimating Unit-level Causal Effect
2. Ratio of $Cov(X, Y)$ and $Var(X)$

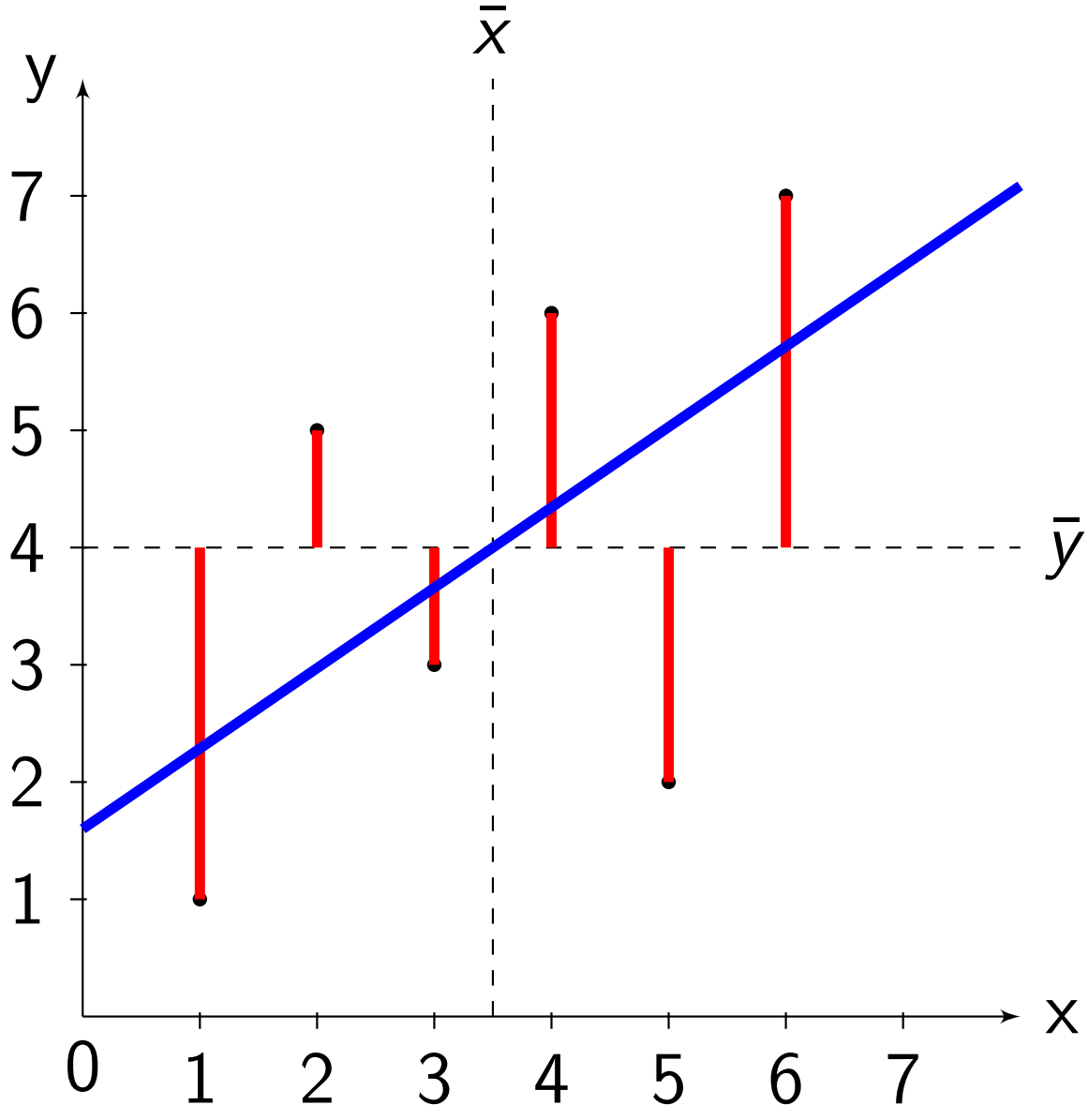
Ways of Thinking About OLS

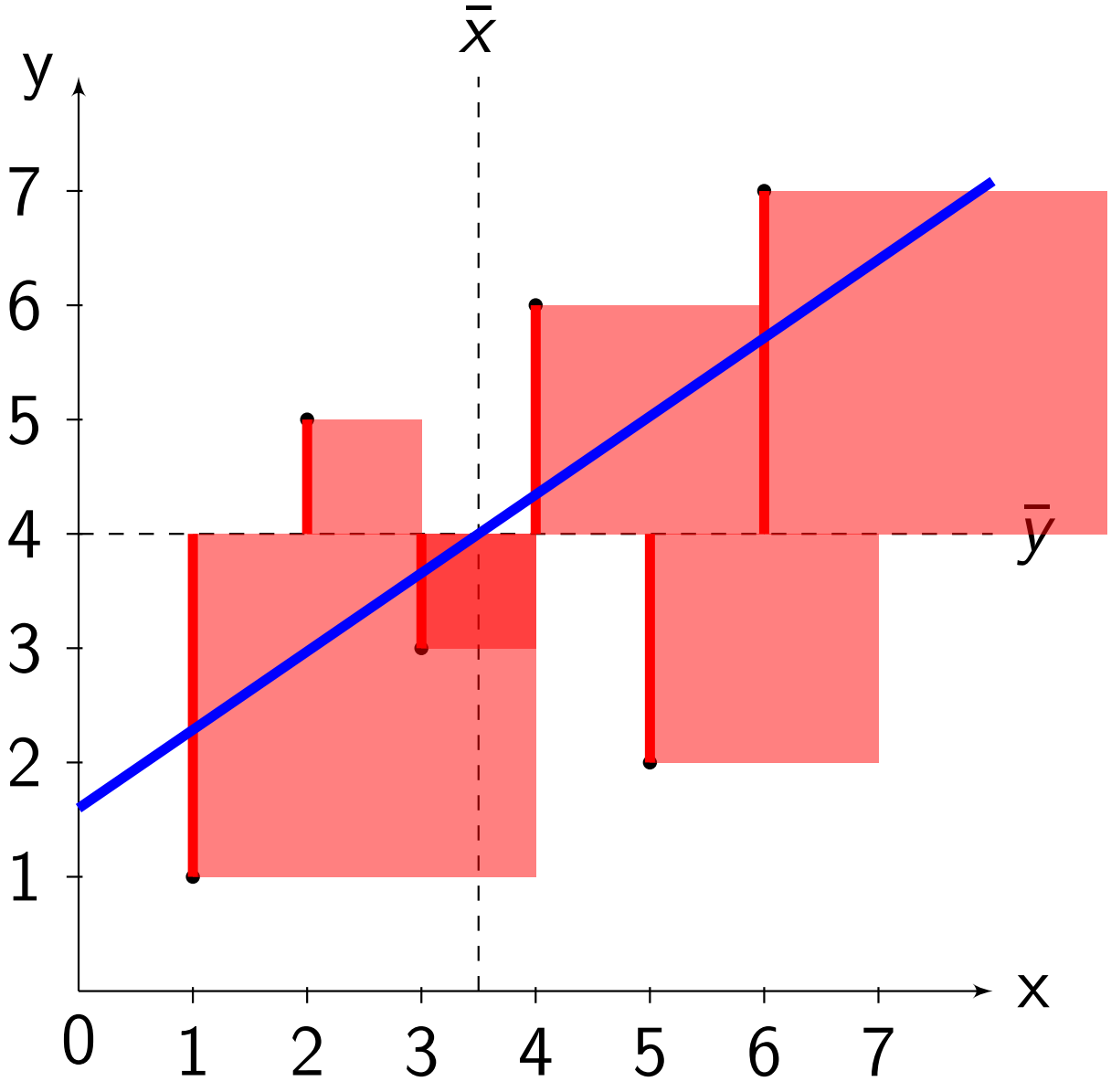
1. Estimating Unit-level Causal Effect
2. Ratio of $Cov(X, Y)$ and $Var(X)$
3. Minimizing residual sum of squares (SSR)

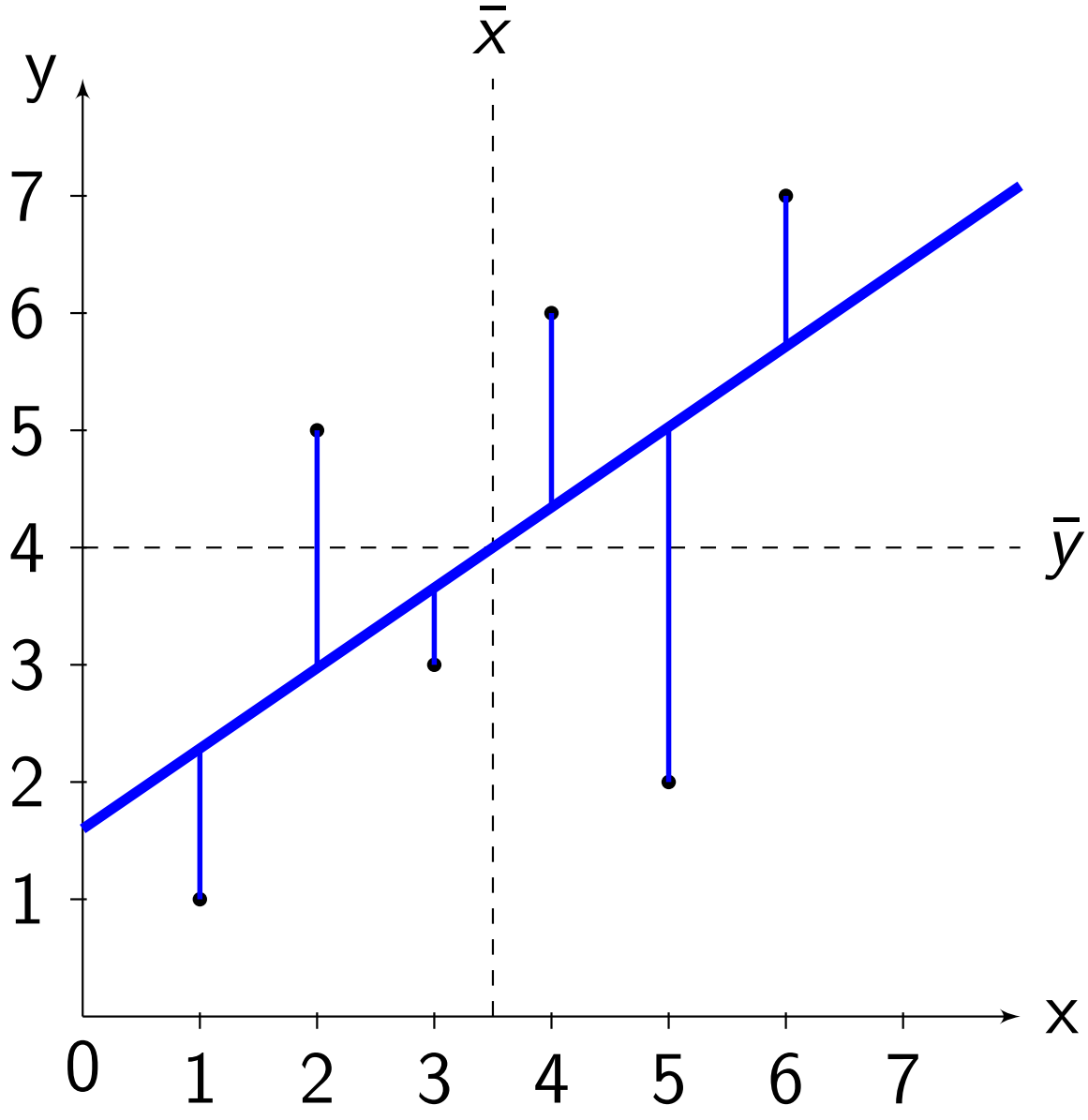
OLS Minimizes SSR

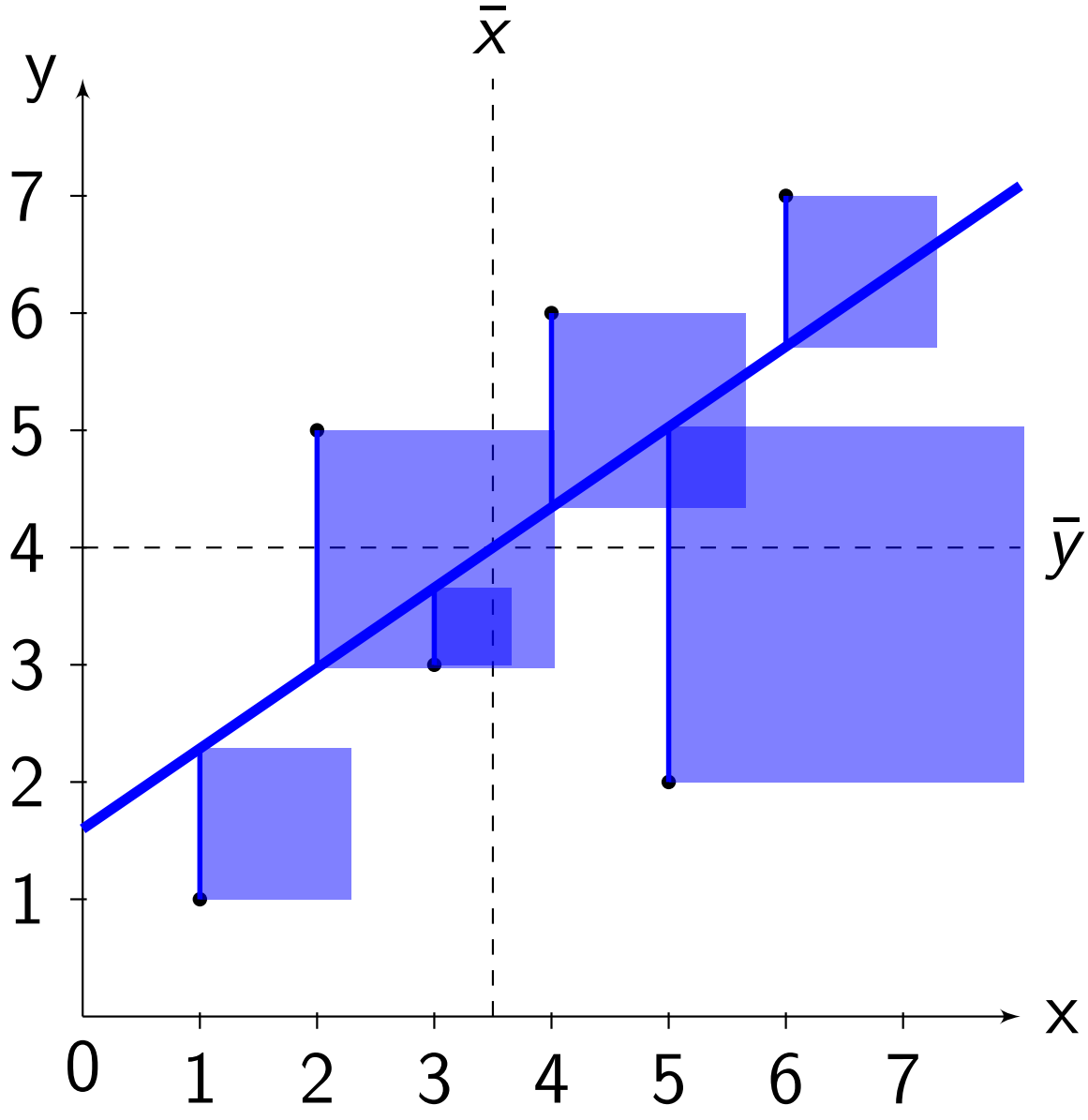
- Total Sum of Squares (SST): $\sum_{i=1}^{n}(y_i - \bar{y})^2$
- We can partition SST into two parts (ANOVA):
  - Explained Sum of Squares (SSE)
  - Residual Sum of Squares (SSR)
- $SST = SSE + SSR$
- OLS is the line with the lowest SSR

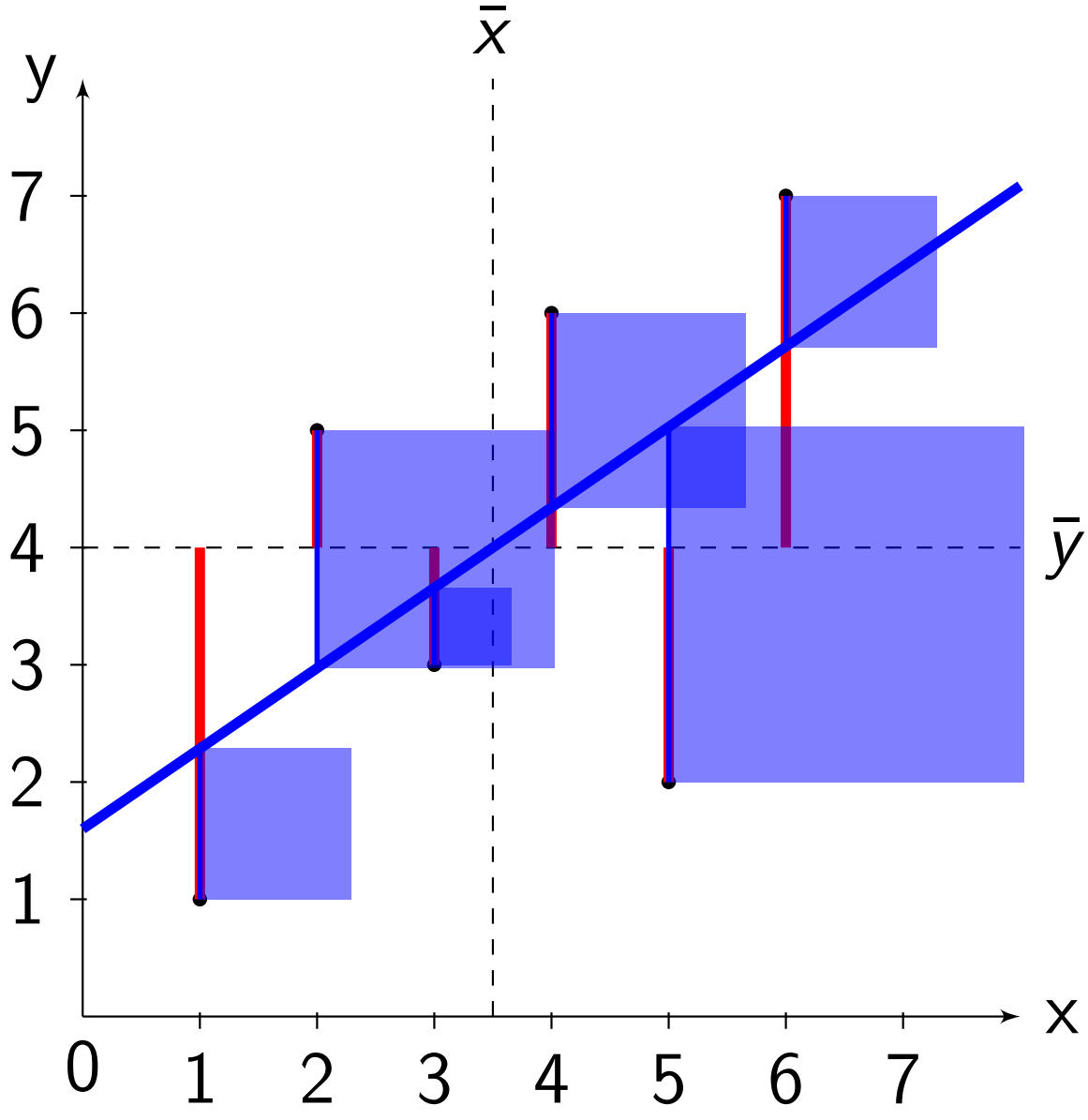

$\bar{x}$
y
7
6
5
4
3
2
1
$\bar{y}$
x
0
1
2
3
4
5
6
7

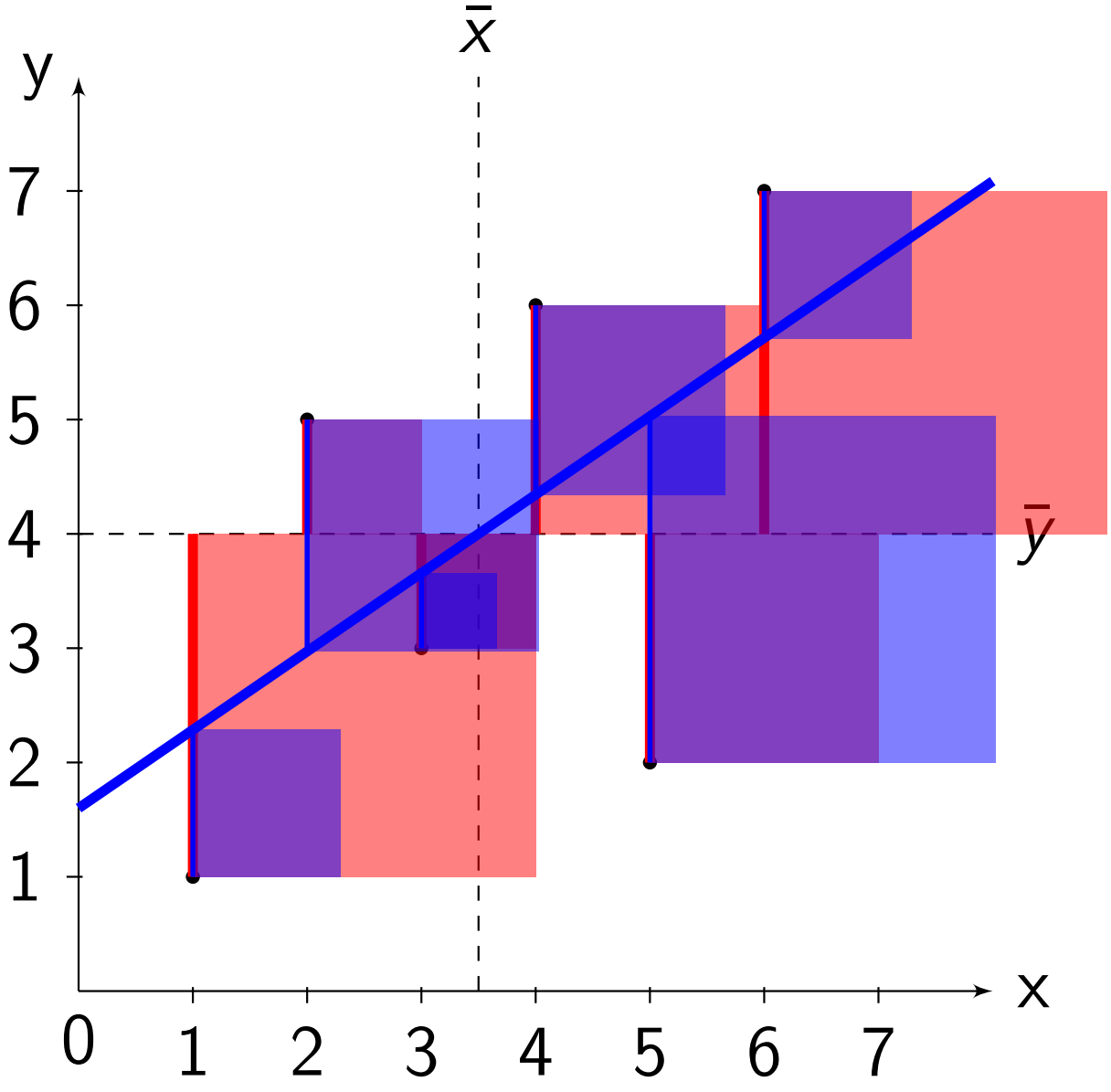

$\bar{x}$
y
7
6
5
4
3
2
1
$\bar{y}$
x
0
1
2
3
4
5
6
7


$\bar{x}$
y
7
6
5
4
3
2
1
$\bar{y}$
x
0
1
2
3
4
5
6
7


$\bar{x}$
y
7
6
5
4
3
2
1
$\bar{y}$
x
0
1
2
3
4
5
6
7


$\bar{x}$
y
7
6
5
4
3
2
1
$\bar{y}$
x
0
1
2
3
4
5
6
7


$\bar{x}$
y
7
6
5
4
3
2
1
$\bar{y}$
x
0
1
2
3
4
5
6
7


$\bar{x}$
y
7
6
5
4
3
2
1
$\bar{y}$
x
0
1
2
3
4
5
6
7

Are Our Estimates Any Good?

Yes, if:

1. Works mathematically
2. Causally valid theory
3. Linear relationship between X and Y
4. X is measured without error
5. No missing data (or MCAR)
6. No confounding

Linear Relationship

- If linear, no problems
- If non-linear, we need to transform
  - Power terms (e.g., x^2, x^3)
  - log (e.g., $log(x)$)
  - Other transformations
  - If categorical: convert to set of indicators
  - Multivariate interactions (next week)

Coefficient Interpretation

- Four types of variables:
  1. Indicator (0,1)
  2. Categorical
  3. Ordinal
  4. Interval
- How do we interpret a coefficient on each of these types of variables?

Notes on Interpretation

- Effect β_1 is constant across values of x

Notes on Interpretation

- Effect β_1 is constant across values of x
- That is not true when there are:
  - Interaction terms (next week)
  - Nonlinear transformations (e.g., x^2)
  - Nonlinear regression models (e.g., logit/probit)

Notes on Interpretation

- Effect β_1 is constant across values of x
- That is not true when there are:
  - Interaction terms (next week)
  - Nonlinear transformations (e.g., x^2)
  - Nonlinear regression models (e.g., logit/probit)
- Interpretations are sample-level
  - Sample representativeness determines generalizability

Notes on Interpretation

- Effect β_1 is constant across values of x
- That is not true when there are:
  - Interaction terms (next week)
  - Nonlinear transformations (e.g., x^2)
  - Nonlinear regression models (e.g., logit/probit)
- Interpretations are sample-level
  - Sample representativeness determines generalizability
- Remember uncertainty
  - These are *estimates*, not population parameters

Measurement Error in Regressor(s)

- We want effect of x, but we observe x^*, where $x = x^* + w$:

$$\begin{aligned} y &= \beta_0 + \beta_1 x^* + \epsilon \\ &= \beta_0 + \beta_1 (x - w) + \epsilon \\ &= \beta_0 + \beta_1 x + (\epsilon - \beta_1 w) \\ &= \beta_0 + \beta_1 x + v \end{aligned}$$

Measurement Error in Regressor(s)

- Produces *attenuation*: as measurement error increases, $\beta_1 \rightarrow 0$
- Our coefficients fit the observed data
- But they are *biased* estimates of our population equation
  - This applies to all $\hat{\beta}$ in a multivariate regression
  - Direction of bias is unknown

Measurement Error in Y

- Not necessarily a problem
- If *random* (i.e., uncorrelated with x), it costs us precision
- If *systematic*, who knows?!

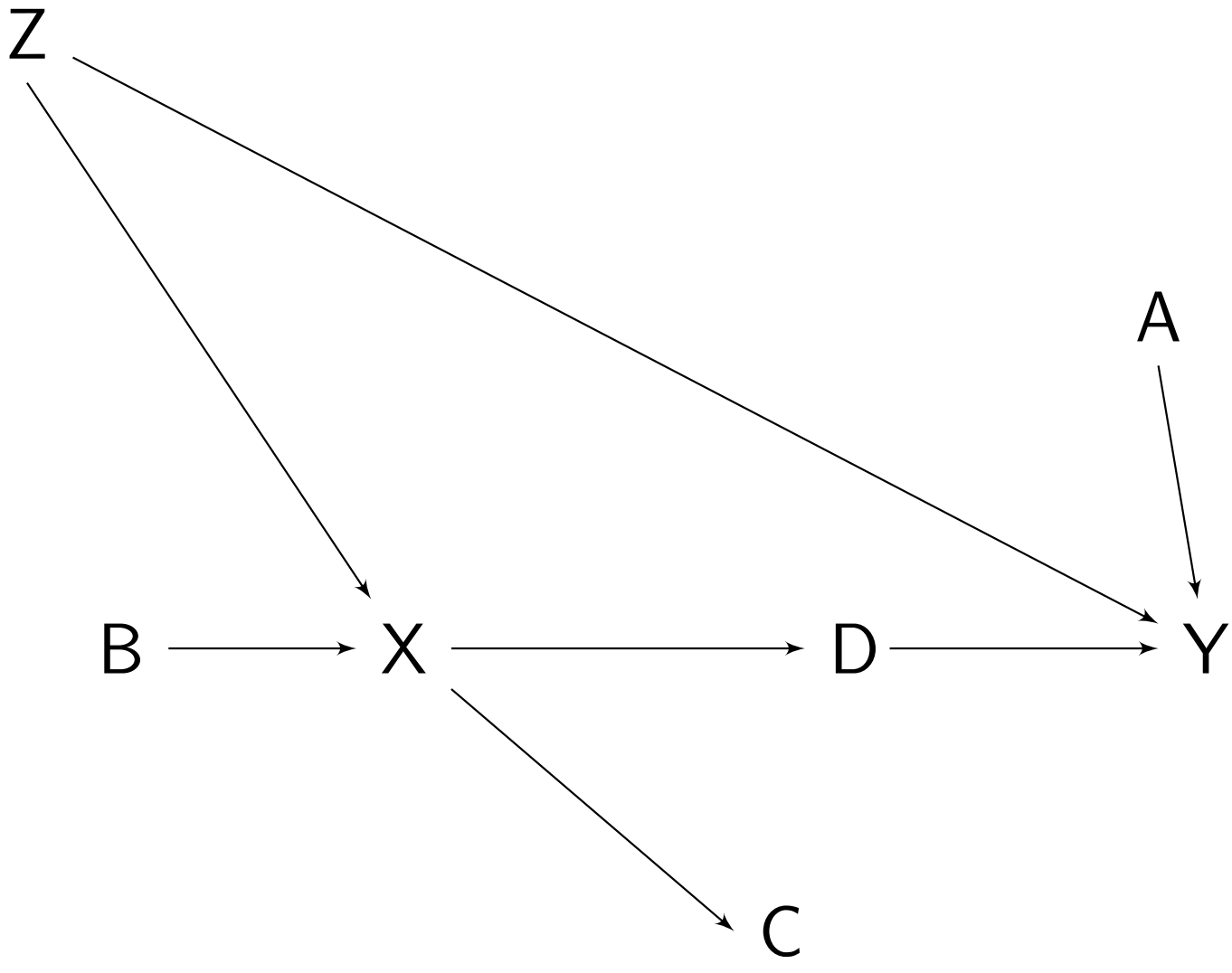
Confounding (Selection Bias)

- If x is not randomly assigned, potential outcomes are not independent of x
- Other factors explain why a unit i received their particular value x_i
- In matching, we obtain this *conditional independence* by comparing units that are identical on all confounding variables

Omitted Variables

$$\underbrace{E[Y_i|X_i = 1] - E[Y_i|X_i = 0] =}_{\text{Naive Effect}}$$

$$\underbrace{E[Y_{1i}|X_i = 1] - E[Y_{0i}|X_i = 1]}_{\text{Treatment Effect on Treated (ATT)}} + \underbrace{E[Y_{0i}|X_i = 1] - E[Y_{0i}|X_i = 0]}_{\text{Selection Bias}}$$


Z
A
B
X
D
Y
C

Omitted Variable Bias

- We want to estimate:

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + \epsilon$$

- We actually estimate:

$$\begin{aligned} \tilde{y} &= \tilde{\beta}_0 + \tilde{\beta}_1 x + \epsilon \\ &= \tilde{\beta}_0 + \tilde{\beta}_1 x + (0 * z) + \epsilon \\ &= \tilde{\beta}_0 + \tilde{\beta}_1 x + \nu \end{aligned}$$

- Bias: $\tilde{\beta}_1 = \hat{\beta}_1 + \hat{\beta}_2 \tilde{\delta}_1$, where $\tilde{z} = \tilde{\delta}_0 + \tilde{\delta}_1 x$

Size and Direction of Bias

- Bias: $\tilde{\beta}_1 = \hat{\beta}_1 + \hat{\beta}_2\tilde{\delta}_1$, where $\tilde{z} = \tilde{\delta}_0 + \tilde{\delta}_1 x$

	$Corr(x, z) < 0$	$Corr(x, z) > 0$
$\beta_2 < 0$	Positive	Negative
$\beta_2 > 0$	Negative	Positive

Aside: Three Meanings of "Endogeneity"

Formally endogeneity is when $Cov(X, \epsilon) \neq 0$

1. Measurement error in regressors
2. Omitted variables associated with included regressors
   - "Specification error"
   - Confounding
3. Lack of temporal precedence

Common Conditioning Strategies

Common Conditioning Strategies

1. Condition on nothing ("naive effect")

Common Conditioning Strategies

1. Condition on nothing ("naive effect")
2. Condition on some variables

Common Conditioning Strategies

1. Condition on nothing ("naive effect")
2. Condition on some variables
3. Condition on all observables

Common Conditioning Strategies

1. Condition on nothing ("naive effect")
2. Condition on some variables
3. Condition on all observables

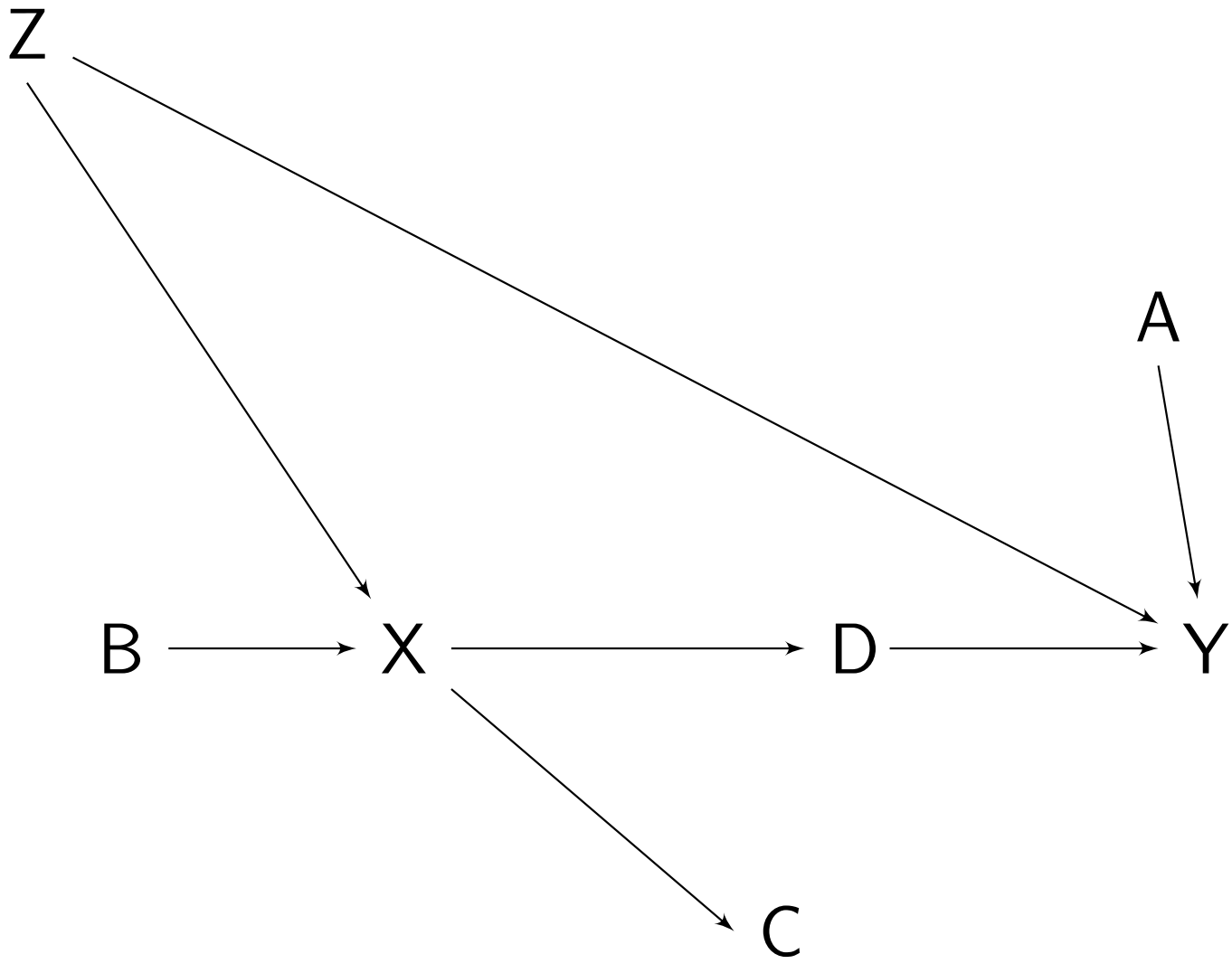
Which of these are good strategies?

What goes in our regression?

- Use theory to build causal models
  - Often, a causal graph helps
- Some guidance:

What goes in our regression?

- Use theory to build causal models
  - Often, a causal graph helps
- Some guidance:
  - Include confounding variables

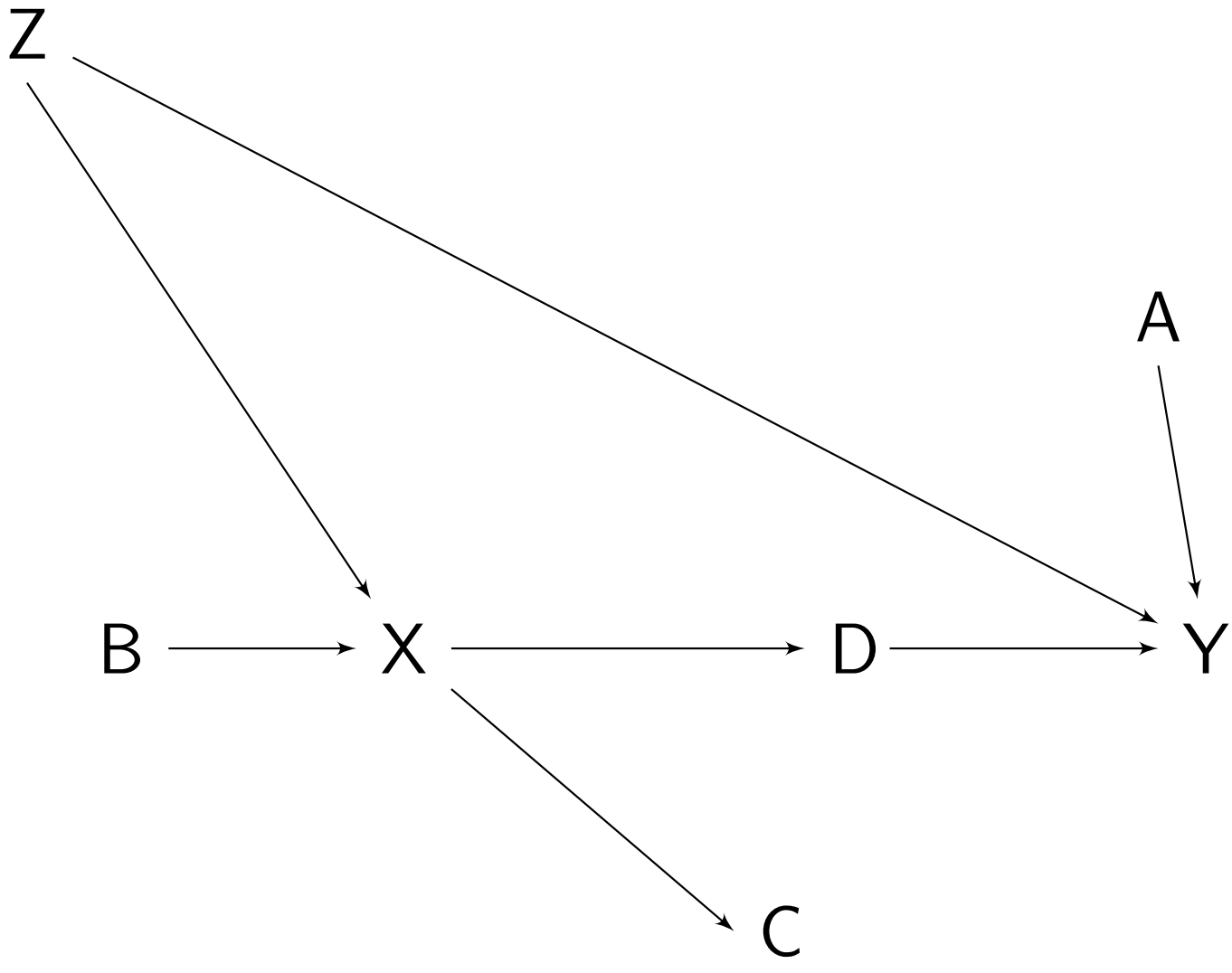

Z
A
B
X
D
Y
C

What goes in our regression?

- Use theory to build causal models
  - Often, a causal graph helps
- Some guidance:
  - Include confounding variables

What goes in our regression?

- Use theory to build causal models
  - Often, a causal graph helps
- Some guidance:
  - Include confounding variables
  - Do not include post-treatment variables


Z
A
B
X
D
Y
C

Post-treatment Bias

- We usually want to know the **total effect** of a cause
- If we include a mediator, D, of the $X \rightarrow Y$ relationship, the coefficient on X:
  - Only reflects the **direct** effect
  - Excludes the **indirect** effect of X through M
- So don't control for mediators!

What goes in our regression?

- Use theory to build causal models
  - Often, a causal graph helps
- Some guidance:
  - Include confounding variables
  - Do not include post-treatment variables

What goes in our regression?

- Use theory to build causal models
  - Often, a causal graph helps
- Some guidance:
  - Include confounding variables
  - Do not include post-treatment variables
  - Do not include *colinear* variables

Minimum Mathematical Requirements

1. Do we need variation in X?
   - Yes, otherwise dividing by zero
2. Do we need variation in Y?
   - No, $\hat{\beta}_1$ can equal zero
3. How many observations do we need?
   - $n \geq k$, where k is number of parameters to be estimated

Minimum Mathematical Requirements

1. Do we need variation in X?
   - Yes, otherwise dividing by zero
2. Do we need variation in Y?
   - No, $\hat{\beta}_1$ can equal zero
3. How many observations do we need?
   - $n \geq k$, where k is number of parameters to be estimated
4. Can we have highly correlated regressors?

Minimum Mathematical Requirements

1. Do we need variation in X?
   - Yes, otherwise dividing by zero
2. Do we need variation in Y?
   - No, $\hat{\beta}_1$ can equal zero
3. How many observations do we need?
   - $n \geq k$, where k is number of parameters to be estimated
4. Can we have highly correlated regressors?
   - Generally no (due to multicollinearity)

What goes in our regression?

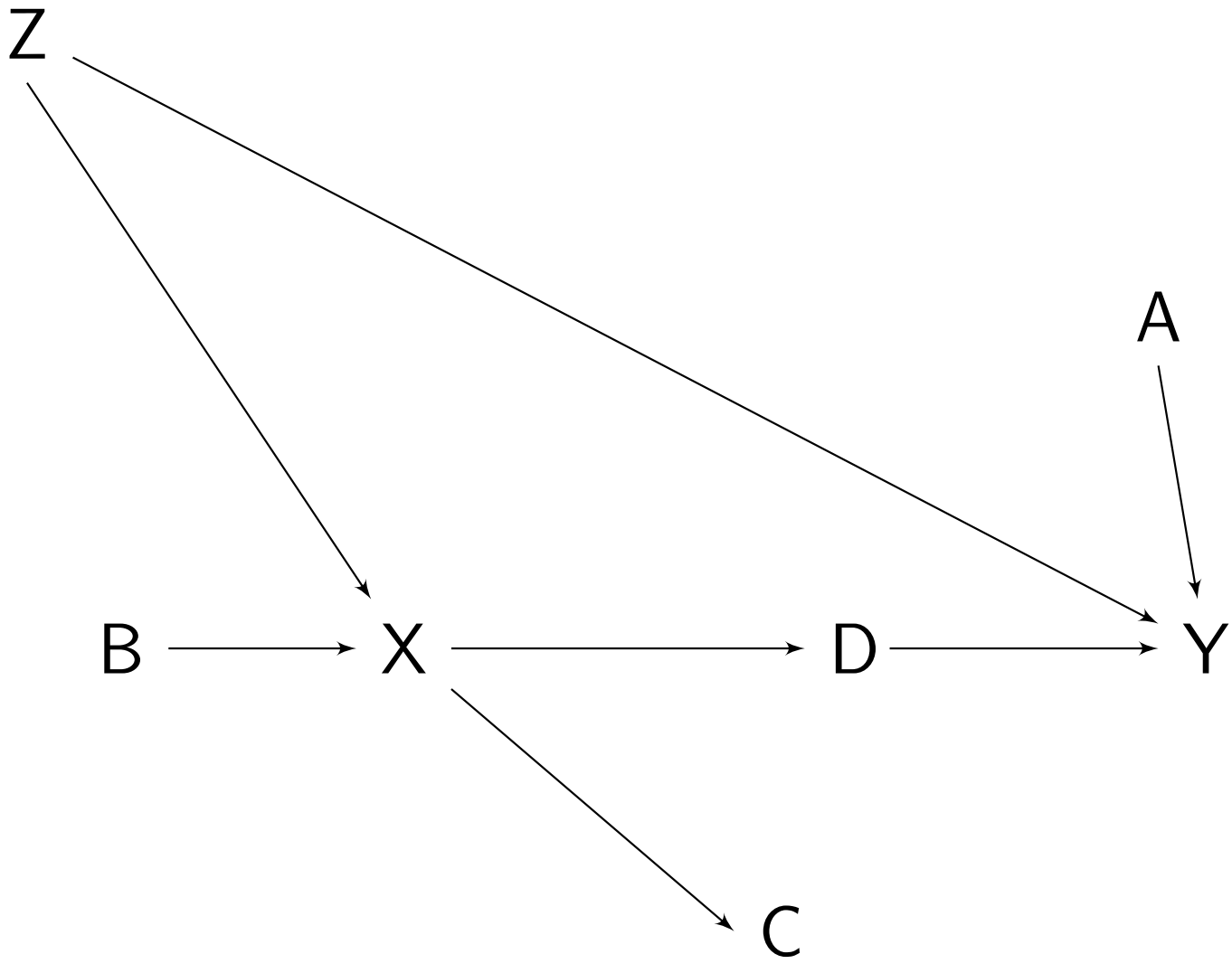
- Use theory to build causal models
  - Often, a causal graph helps
- Some guidance:
  - Include confounding variables
  - Do not include post-treatment variables
  - Do not include *colinear* variables

What goes in our regression?

- Use theory to build causal models
  - Often, a causal graph helps
- Some guidance:
  - Include confounding variables
  - Do not include post-treatment variables
  - Do not include *colinear* variables
  - Including irrelevant variables costs certainty

What goes in our regression?

- Use theory to build causal models
  - Often, a causal graph helps
- Some guidance:
  - Include confounding variables
  - Do not include post-treatment variables
  - Do not include *colinear* variables
  - Including irrelevant variables costs certainty
  - Including variables that affect Y alone increases certainty


Z
A
B
X
D
Y
C

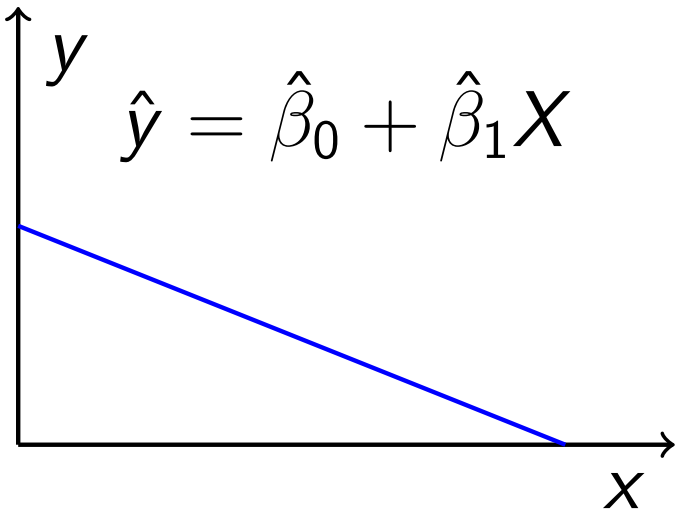
Multivariate Regression Interpretation

- All our interpretation rules from earlier still apply in a multivariate regression
- Now we interpret a coefficient as an effect "all else constant"
- Generally, not good to give all coefficients a causal interpretation
  - Think "forward causal inference"
  - We're interested in the $X \rightarrow Y$ effect
  - All other coefficients are there as "controls"

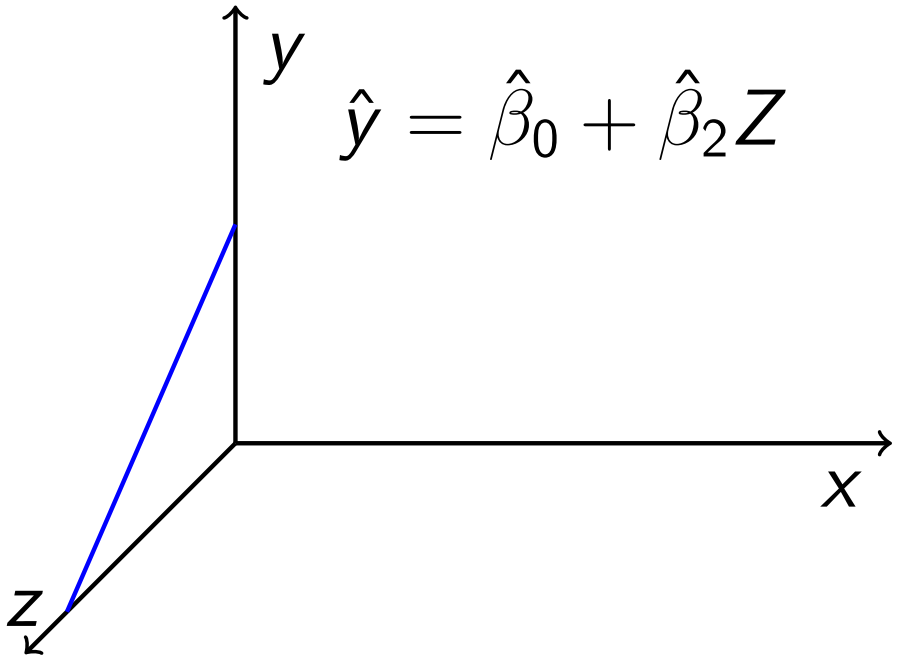
From Line to Surface I

- In simple regression, we estimate a **line**
- In multiple regression, we estimate a **surface**
- Each coefficient is the *marginal effect*, all else constant (at mean)
- This can be hard to picture in your mind

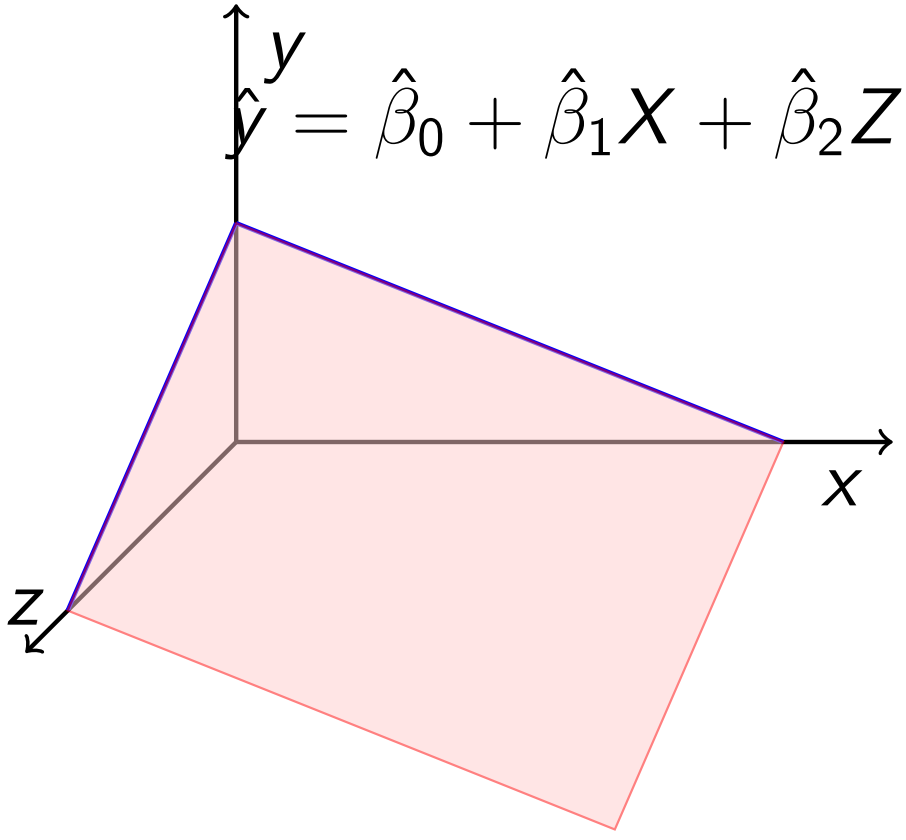
From Line to Surface II



From Line to Surface II



From Line to Surface II



Are Our Estimates Any Good?

Yes, if:

1. Works mathematically
2. Causally valid theory
3. Linear relationship between X and Y
4. X is measured without error
5. No missing data (or MCAR)
6. No confounding

OLS is BLUE

- BLUE: Best Linear Unbiased Estimator
- Gauss Markov Assumptions:
  1. Linearity in parameters
  2. Random sampling
  3. No multicollinearity
  4. Exogeneity ($E[\epsilon|\mathbf{X}] = 0$)
  5. Homoskedasticity ($Var(\epsilon|\mathbf{X}) = \sigma^2$)
- Assumptions 1–4 prove OLS is unbiased
- Assumption 5 proves OLS is the *best* estimator

Squared vs. Absolute Errors

- Conventionally use Sum of Squared Errors
- Using absolute errors is also unbiased
- Sum of Squared Errors:
  - more heavily weights outliers
  - has a smaller variance
- Thus OLS is **B**est**LUE**

Goodness-of-Fit

- We want to know: "How good is our model?"

Goodness-of-Fit

- We want to know: "How good is our model?"
- We can answer:
  "How well does our model fit the observed data?"

Goodness-of-Fit

- We want to know: "How good is our model?"
- We can answer:
  "How well does our model fit the observed data?"
- Is this what we want to know?

Correlation

- Definition: $Corr(x, y) = \hat{r}_{x,y} = \frac{Cov(x,y)}{(n-1)s_x s_y}$
- Slope $\hat{\beta}_1$ and correlation $\hat{r}_{x,y}$ are simply different scalings of $Cov(x, y)$
- Interpretation: How well the bivariate relationship is summarized by a cloud of points?
- Units: none (range -1 to 1)

Coefficient of Determination (R^2)

- Definition: $R^2 = \hat{r}^2_{x,y} = \frac{SSE}{SST} = 1 - \frac{SSR}{SST}$
- Interpretation: How much of the total variation in y is explained by the model?
- But, R^2 increases simply by adding more variables
- So, Adjusted-$R^2 = R^2 - (1 - R^2)\frac{k}{n-k-1}$, where k is number of regressors
- Units: none (range 0 to 1)

Standard Error of the Regression (SER)

- "Root mean squared error" or just σ
- Definition: $\hat{\sigma} = \sqrt{\frac{SSR}{n-p}}$, where p is number of parameters estimated
- Interpretation: How far, on average, are the observed y values from their corresponding fitted values $\hat{y}$
  - $sd(y)$ is how far, on average, a given y_i is from $\bar{y}$
  - σ is how far, on average, a given y_i is from $\hat{y}_i$
- Units: same as y (range 0 to $sd(y)$)

The F-test

- Definition: Test of whether any of our coefficients differ from zero
  - In a bivariate regression, $F = t^2$
- Interpretation: Do any of the coefficients differ from zero?
  - Not a very interesting measure
- Units: none (range 0 to ∞)

Non-continuous Outcomes

1. Why shouldn't we use OLS for a non-continuous outcome variable?

Non-continuous Outcomes

1. Why shouldn't we use OLS for a non-continuous outcome variable?
2. What do we do instead?

Non-continuous Outcomes

1. Why shouldn't we use OLS for a non-continuous outcome variable?
2. What do we do instead?
   - Use a generalized linear model (GLM)

Regression on a Latent Variable

- Consider a binary outcome y (e.g., voting)
- OLS provides a nonsensical fit to the outcome
- Think about the problem as a "latent" outcome ($y*$) that manifests in two observed categories
  - As $y*$ increases, $Pr(Y = 1) \rightarrow 1$
  - As $y*$ decreases, $Pr(Y = 1) \rightarrow 0$
- We do not observe $y*$, only y

Estimation in GLM

- In OLS, we estimate: $\hat{y} = \beta_0 + \beta_1 x + e$
- This represents the conditional mean of y
- In a GLM, we estimate: $\hat{y}* = \beta_0 + \beta_1 x + e$ where $y*$ is a transformation of y
- This is also a prediction of the conditional mean of y

Estimation in GLM

- In OLS, we estimate: $\hat{y} = \beta_0 + \beta_1 x + e$
- This represents the conditional mean of y
- In a GLM, we estimate: $\hat{y}* = \beta_0 + \beta_1 x + e$ where $y*$ is a transformation of y
- This is also a prediction of the conditional mean of y
- How do we transform y to $y*$?

Model Specification

1. Complete set of conditioning variables

Model Specification

1. Complete set of conditioning variables
2. Correctly specified model

Model Specification

1. Complete set of conditioning variables
2. Correctly specified model
3. Choice of error distribution

Model Specification

1. Complete set of conditioning variables
2. Correctly specified model
3. Choice of error distribution
4. Link function

Error Distribution

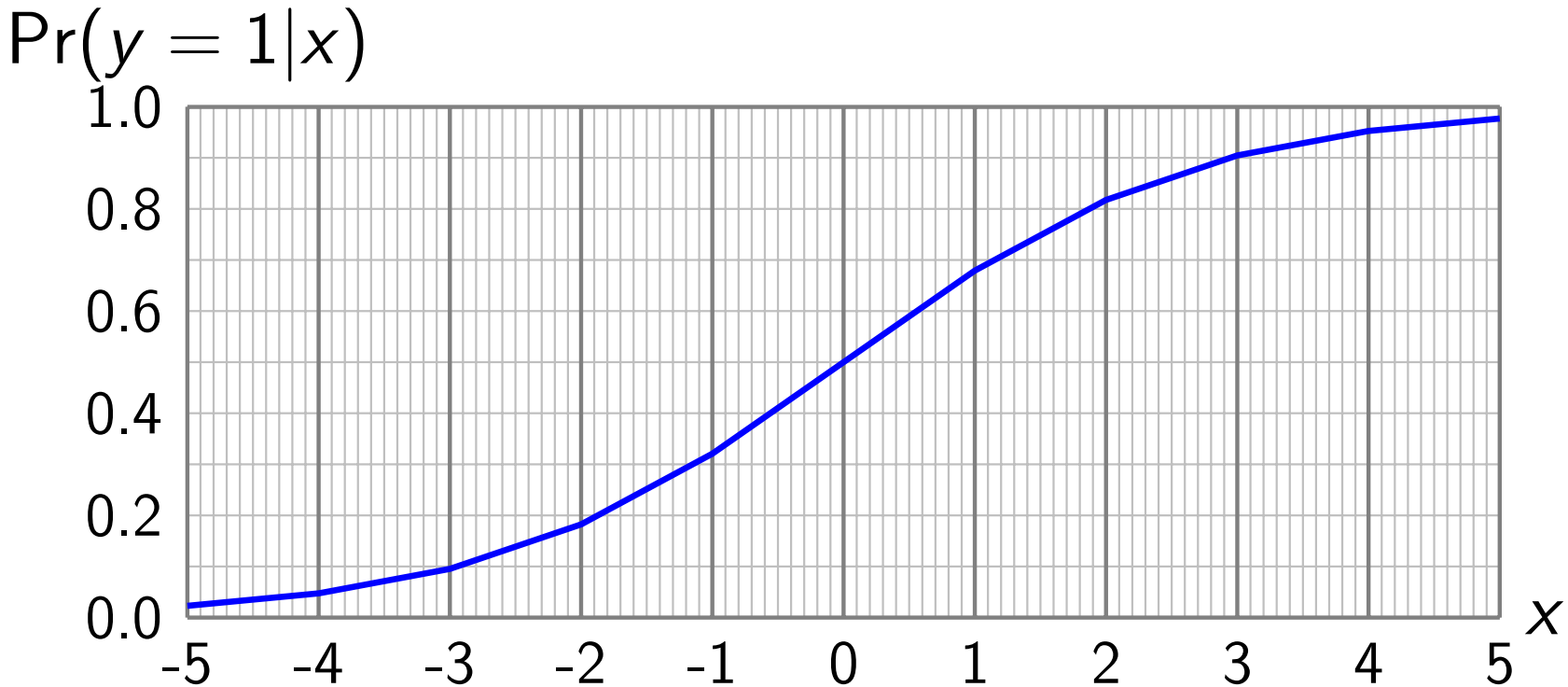
- To estimate a linear model using OLS, no distributional assumption is needed
- We can use Maximum Likelihood Estimation to obtain identical coefficient estimates as OLS by assuming errors are Normally distributed
  - $\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- For any GLM, we must assume the population distribution of the errors
  - In almost all cases, an *exponential family* distribution

Link Function

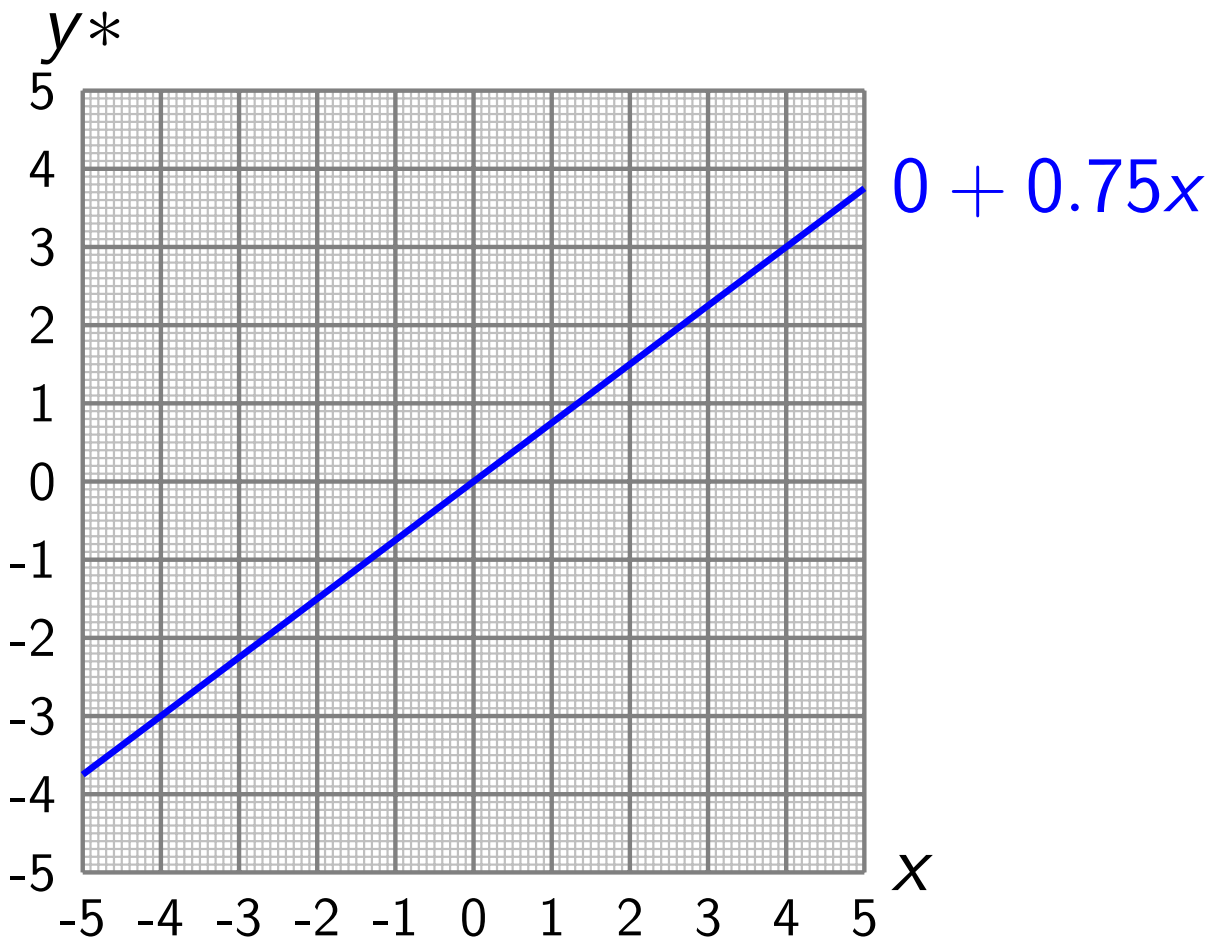
- $y* = X\beta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots$
- **Link function**: $g(\mu) = X\beta$
  - Transforms y to $y*$
- **Inverse link function**: $\mu = g^{-1}(X\beta)$
  - Transforms $y*$ back to y

Inverse Link Function

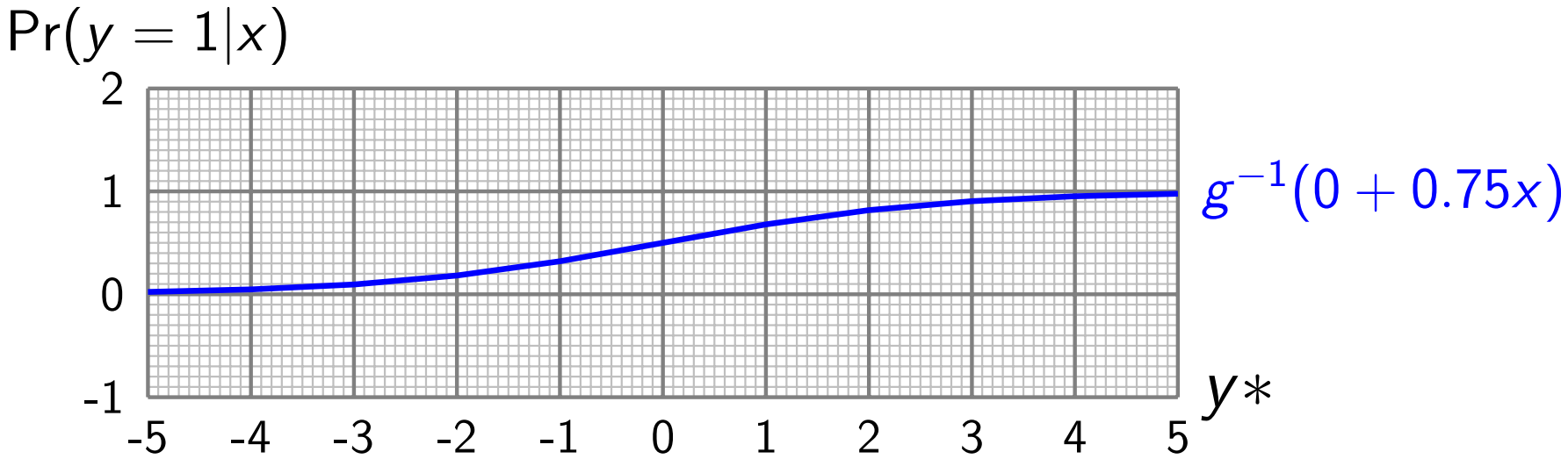
- This plot displays $g^{-1}(0.75x)$, where g^{-1} is the inverse logit link.



From x to Linear Prediction ($y*$)

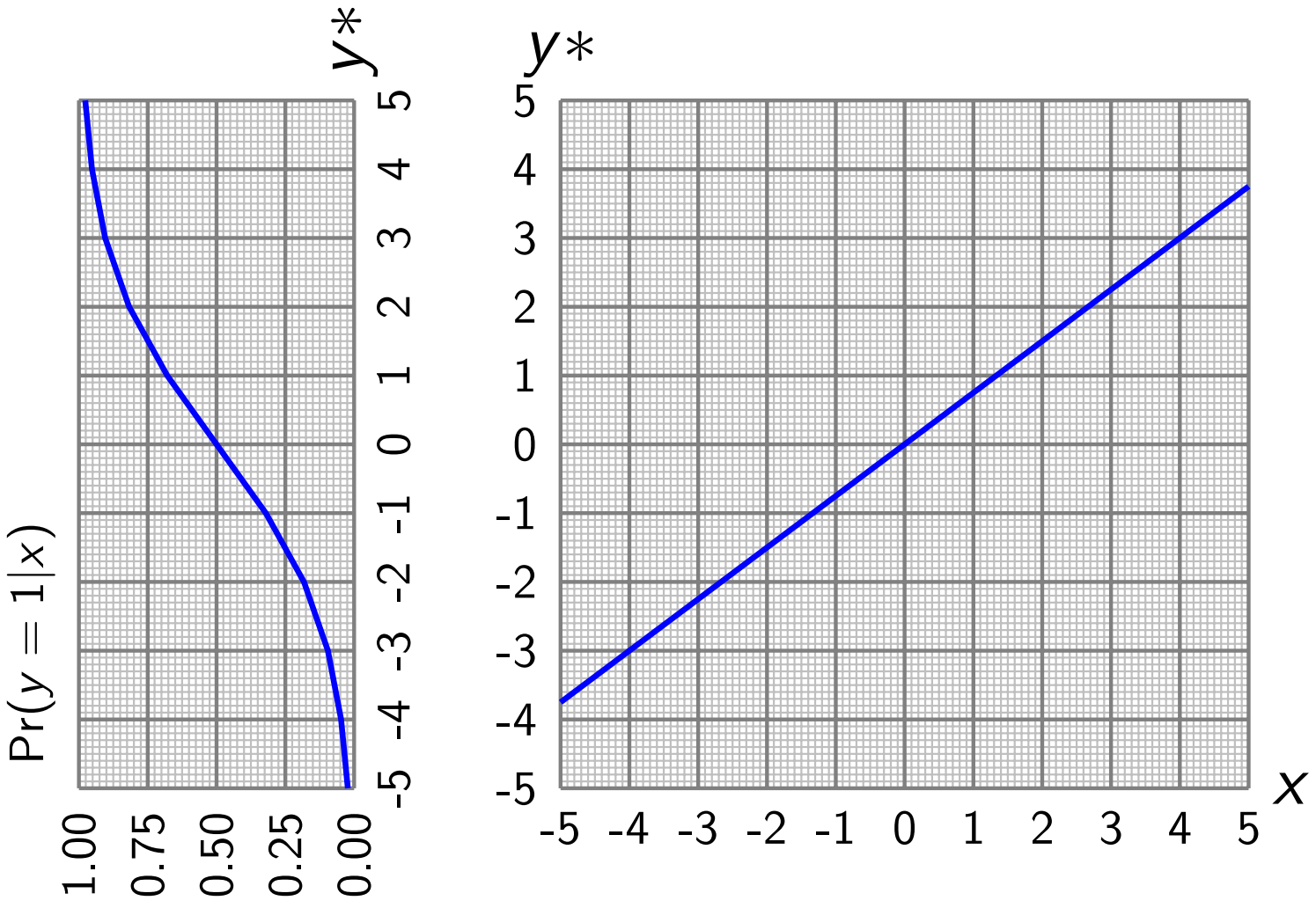


From $y*$ to $Pr(y = 1)$



- Function is monotonic
- There is a *cutpoint* in $y*$ where $\Pr(y) = 0.5$
- It is symmetric above and below the cutpoint

Putting it Together



Choosing a Link Function

- Based on expected distribution of the error term of $y*$
- Choice heavily influenced by convention rather than empirics
- Choice of link adds *model dependence*!
  - Expected influence of x on y now depends on choice of link
  - Different link functions can yield different substantive and statistical results
  - Generally, results are similar

Common Link Functions

Name	*Link*	*Inverse*
Identity	μ	$y*$
Logit	$\ln \frac{\mu}{1-\mu}$	$\frac{1}{1+e^{-y*}}$
Probit	$\Phi^{-1}(\mu)$	$\Phi(y*)$

- These are common for categorical outcomes
- Other types of outcomes will use different link functions

Beyond Binary Outcomes

- The generalized linear model works for all kinds of outcomes, not just continuous or binary
- Consider, for example, a multi-category, ordered outcome variable
- In an *ordered logit* model, we imagine a latent variable $y*$ and multiple cutpoints between categories of y

Maximum Likelihood Estimation

- The *generalized linear model* is a way of describing complex regression models
- Unlike OLS, there is no closed-form mathematical solution to GLM
  - Recall a linear model can be expressed as a GLM
- GLMs involve the big additional assumption of a distribution for the error term
- Maximum likelihood estimation is a way of estimating the GLM that makes use of that error distribution

Maximum Likelihood Estimation

- Choose an error distribution (which is described by various parameters)
- Select parameters as starting values
- Give a *probability* of seeing each observation in our sample data given that distribution
- Combine those probabilities (i.e., likelihoods)
  - Multiply the likelihoods
  - Add the log-likelihoods
- Repeat and pick the best guess from all of those that we test

Coefficients

- Coefficients express effect of x on $y*$
- In logistic regression, this is a statement about the odds-ratio: $\hat{\beta} = \frac{\frac{p_1}{1-p_1}}{\frac{p_0}{1-p_0}}$
- Coefficients are hard to interpret *substantively*
- Statistical significance is similar to OLS

Predicted Outcomes

- In OLS, fitted values from the estimated regression equation are values of y
- In GLMs, fitted values are expressed for $y*$
- To interpret logit or probit, we transform to predicted probabilities

Predicted Outcomes

- Definition: According to our coefficient estimates, what is $\Pr(\hat{y} = 1|X)$?
- To calculate this, we:
  1. Calculate a fitted value on the latent/linear scale
  2. Plug that fitted value into the inverse link function
- In Stata, use `margins` and `predict`
  - Probabilities are the default
  - Use the `, xb` option for linear predictions

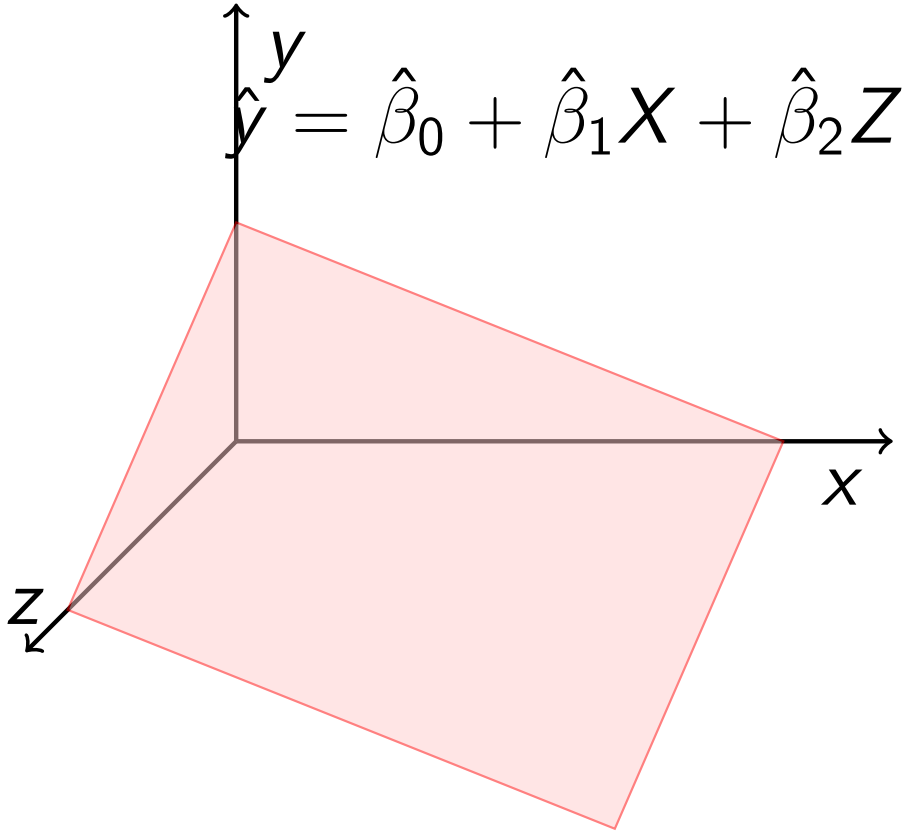
Marginal Effects

- A marginal effect refers to one of two quantities:
  - For continuous variables: the partial derivative of the regression equation with respect to a specific variable
  - For categorical variables: the difference $\Pr(y = 1|x = 1) - \Pr(y = 1|x = 0)$
- In an OLS (with no interactions or other complex terms), the marginal effect is the coefficient itself
- In GLMs, this is more complicated

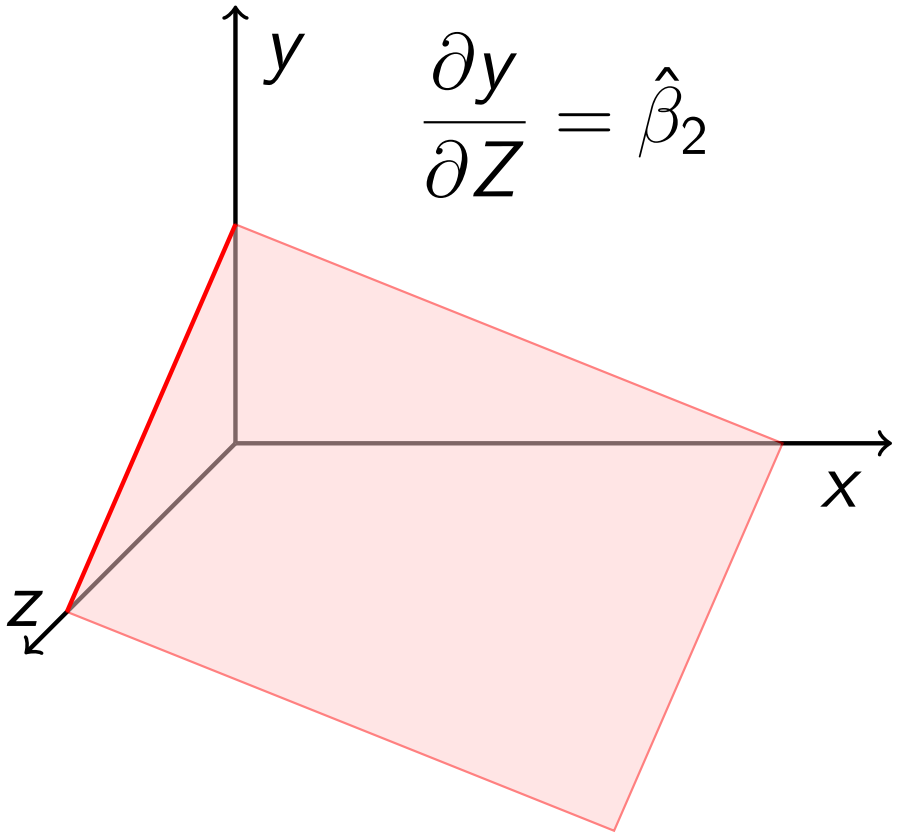
Review: Partial Derivatives

- The partial derivative is the *instantaneous* slope (or *tangent*) of a line
- With one x variable, this is just the slope of the line
- With > 1 x variable, this is the conditional slope of the regression surface
  - We hold other variables at some value see a "slice" of the regression surface
  - The marginal effect is the slope of the slice

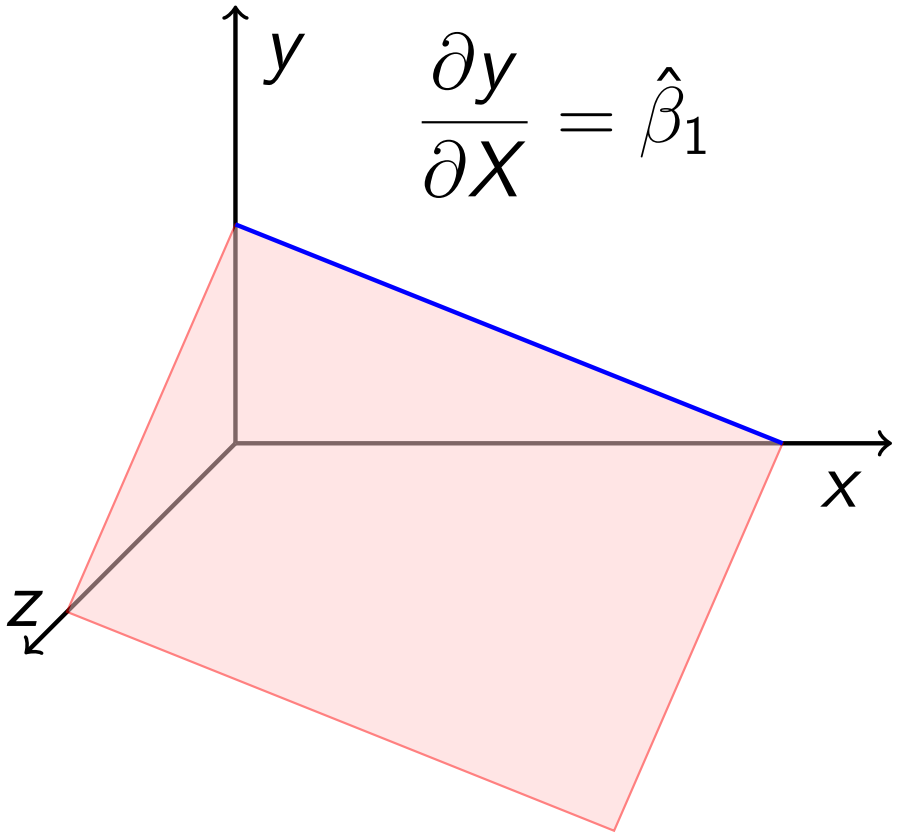
Simple regression surface



Simple regression surface



Simple regression surface



Discrete Changes

- Marginal effects are *instantaneous changes*
- This makes sense for continuous variables
- For categorical (factor) variables, we often instead calculate a discrete change
  - $Pr(y = 1|x = 1) - Pr(y = 1|x = 0)$
  - Marginal effect and discrete change are the same in OLS

Interaction Terms

- Due to the link function transformation, the marginal effect of x depends on the value of x and all other covariates
- This creates *implicit* interactions
- We still have to include *explicit* interaction terms to estimate heterogeneous effects (i.e., effect moderation)

Logit vs. Probit

- Both constrain a continuous $y*$ to (0,1)

Logit vs. Probit

- Both constrain a continuous $y*$ to (0,1)
- Probabilities are symmetric

Logit vs. Probit

- Both constrain a continuous $y*$ to (0,1)
- Probabilities are symmetric
- Logit allows us to estimate odds-ratios

Logit vs. Probit

- Both constrain a continuous $y*$ to (0,1)
- Probabilities are symmetric
- Logit allows us to estimate odds-ratios
- Logit is maybe slightly more common in political science for what are probably just historical reasons

Language of Interpretation

- How do we describe a marginal effect in OLS?

Language of Interpretation

- How do we describe a marginal effect in OLS?
- In a binary outcome model, we use different language

Language of Interpretation

- How do we describe a marginal effect in OLS?
- In a binary outcome model, we use different language
- The substantive importance of an effect may depend on the level of $\Pr(y)$ at which it occurs

Language of Interpretation

- How do we describe a marginal effect in OLS?
- In a binary outcome model, we use different language
- The substantive importance of an effect may depend on the level of $\Pr(y)$ at which it occurs
  - Small effect at $\Pr(y) = 0.01$ vs. $\Pr(y) = 0.48$

Language of Interpretation

- How do we describe a marginal effect in OLS?
- In a binary outcome model, we use different language
- The substantive importance of an effect may depend on the level of $\Pr(y)$ at which it occurs
  - Small effect at $\Pr(y) = 0.01$ vs. $\Pr(y) = 0.48$
  - Large positive effect when $\Pr(y)$ is always > 0.6

Language of Interpretation

- How do we describe a marginal effect in OLS?
- In a binary outcome model, we use different language
- The substantive importance of an effect may depend on the level of $\Pr(y)$ at which it occurs
  - Small effect at $\Pr(y) = 0.01$ vs. $\Pr(y) = 0.48$
  - Large positive effect when $\Pr(y)$ is always > 0.6
- Substantive importance depends on variability and *stickiness* of x

Summarizing Marginal Effects

- We need to decide the values for all covariates that we will use in summarizing the marginal effect of our focal variable
- If our equation is:
  $y = g^{-1}(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots)$
- And we want to know the marginal effect of x_1, we need to hold x_2 at some specified value(s)

3 Common Marginal Effect Summaries

1. Marginal Effects at the Mean (MEMs)
2. Marginal Effects at Representative Values (MERs)
3. Average Marginal Effects (AMEs)

MEM

- We are interested in the ME of x_1
- We hold all other covariates at their respective means
- For example, we interested in the ME of *knowledge*, we hold *education* at its mean

MEM

- We are interested in the ME of x_1
- We hold all other covariates at their respective means
- For example, we interested in the ME of *knowledge*, we hold *education* at its mean
- Does this make sense for categorical values (e.g., gender)?

MERs

- The means of the covariates may not be meaningful
- We hold those covariates at various interesting values
  - ME of x for a high-school educated female
  - ME of x for a university-educated male
  - etc.
- Helpful because we may also be interested in $\Pr(y = 1)$ for these cases

AMEs

- We may not only be interested in MEs at particular values
- We may want a summary measure of the effect of x for our sample as a whole
- The *average marginal effect* calculates the MER for every observation in our data, then averages those ME values
- This is Stata's default behavior when using: `margins, dydx(*)`

AMEs: An Example

- Model effects of gender & education on trust
- Calculate ME for each observation
- Average to obtain AME

AMEs: An Example

- Model effects of gender & education on trust
- Calculate ME for each observation
- Average to obtain AME

Obs.	Gender	Degree	$ME(Gender)$	$ME(Degree)$
1	1	1	0.10	0.25
2	1	1	0.10	0.25
3	1	0	0.20	0.15
4	0	1	0.30	-0.25
5	0	0	0.10	-0.40
AME	–	–		

AMEs: An Example

- Model effects of gender & education on trust
- Calculate ME for each observation
- Average to obtain AME

Obs.	Gender	Degree	$ME(Gender)$	$ME(Degree)$
1	1	1	0.10	0.25
2	1	1	0.10	0.25
3	1	0	0.20	0.15
4	0	1	0.30	-0.25
5	0	0	0.10	-0.40
AME	–	–	0.16	

AMEs: An Example

- Model effects of gender & education on trust
- Calculate ME for each observation
- Average to obtain AME

Obs.	Gender	Degree	$ME(Gender)$	$ME(Degree)$
1	1	1	0.10	0.25
2	1	1	0.10	0.25
3	1	0	0.20	0.15
4	0	1	0.30	-0.25
5	0	0	0.10	-0.40
AME	–	–	0.16	0.00

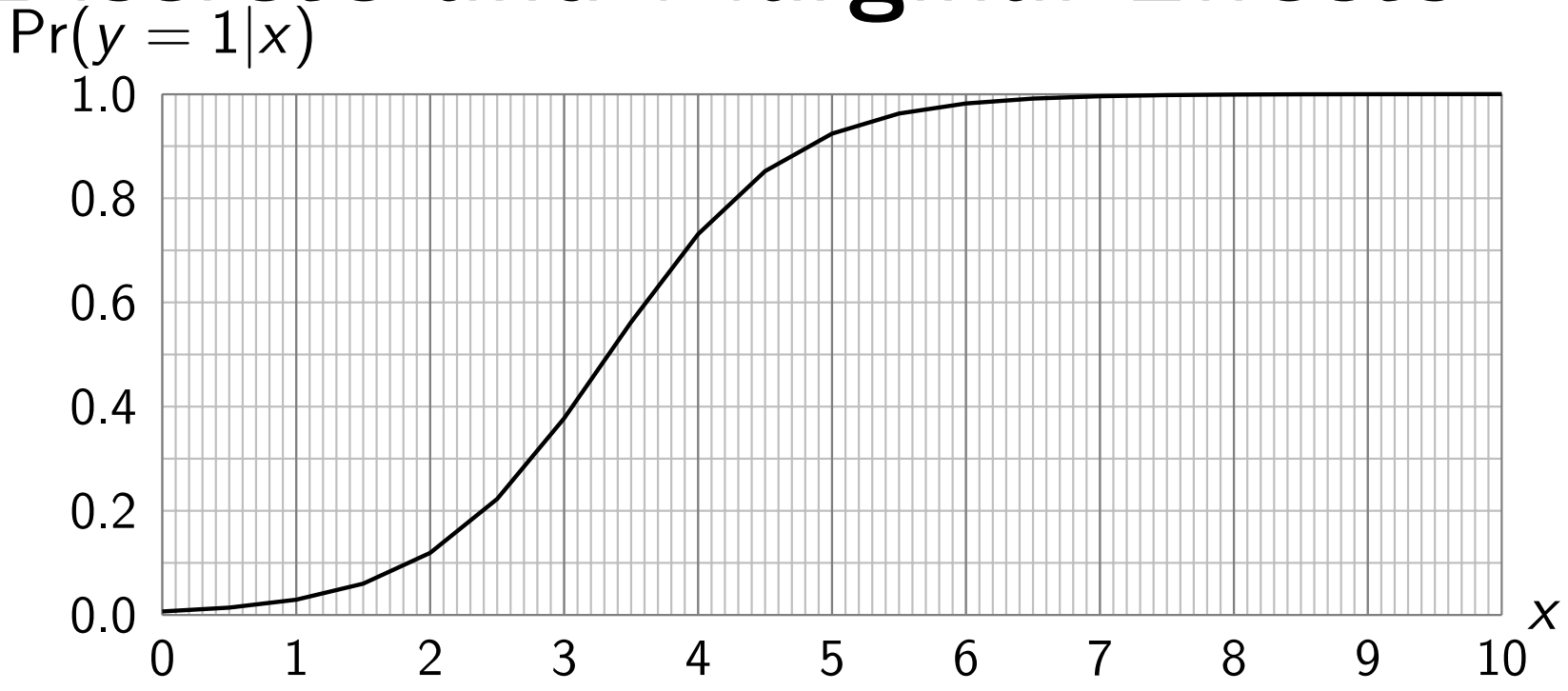
Statistical Uncertainty

- Always express statistical uncertainty for:
  - Coefficients
  - Predicted probabilities
  - Marginal effects
- Significance of coefficients and marginal effects may vary
- Marginal effects may only differ from 0 on a subset of the range of x
- Be cautious about extrapolation

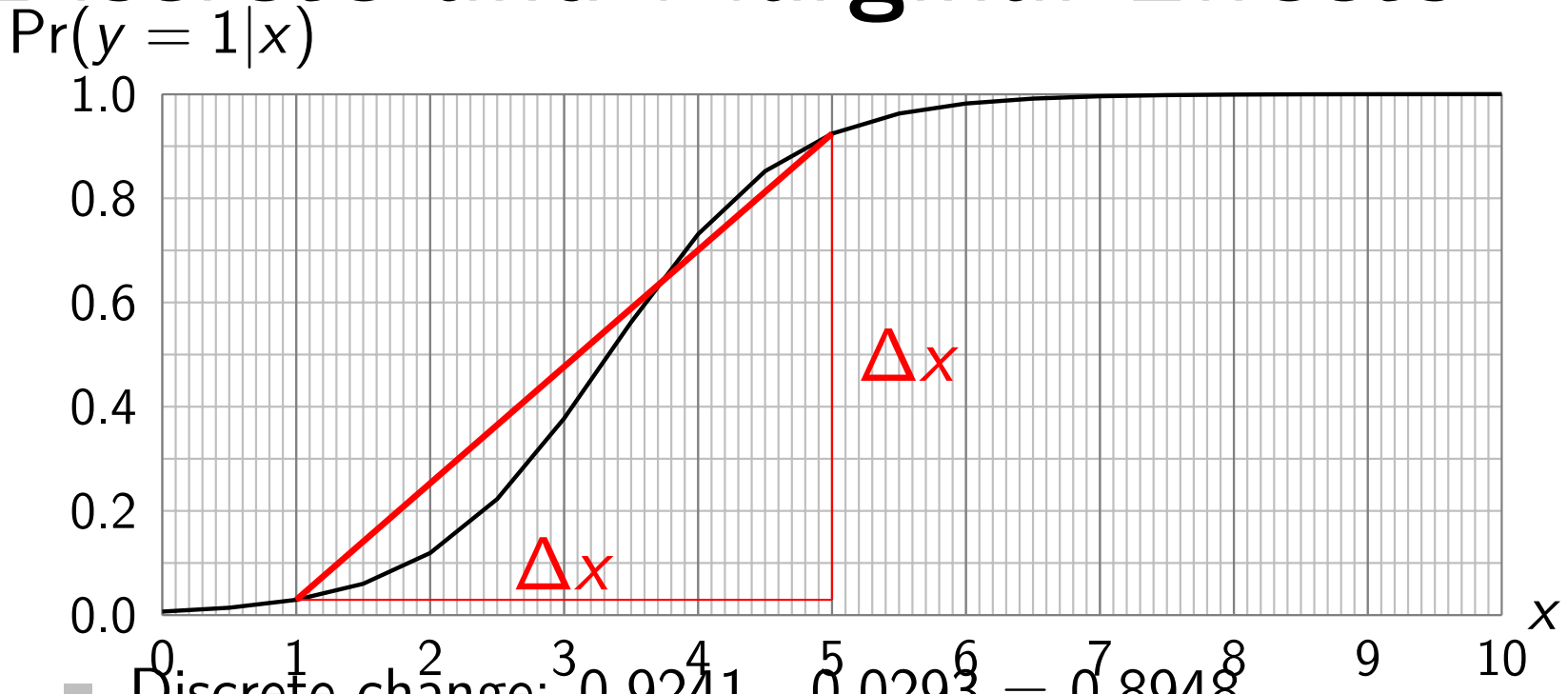
Aside: Discrete Effects

- Discrete effects can also be calculated for continuous variables
- Requires choosing a substantively meaningful change in x
- **Caution!** Not necessary equal:
  - $\Pr(y = 1|x = 5) - \Pr(y = 1|x = 1)$
  - ME at $x = 5$ (MER)
  - ME at $x = 1$ (MER)
  - ME at $x = 3$ (AME/MEM)

Discrete and Marginal Effects

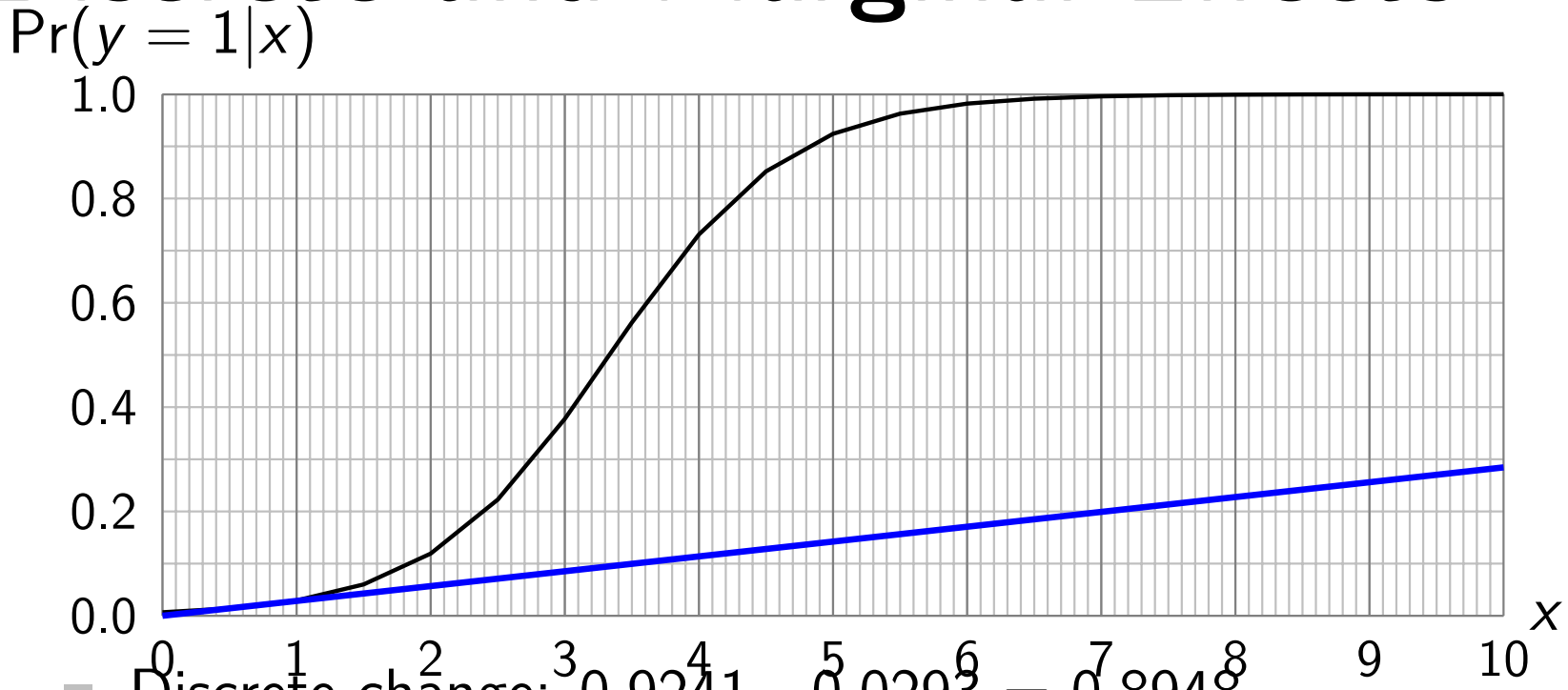


Discrete and Marginal Effects



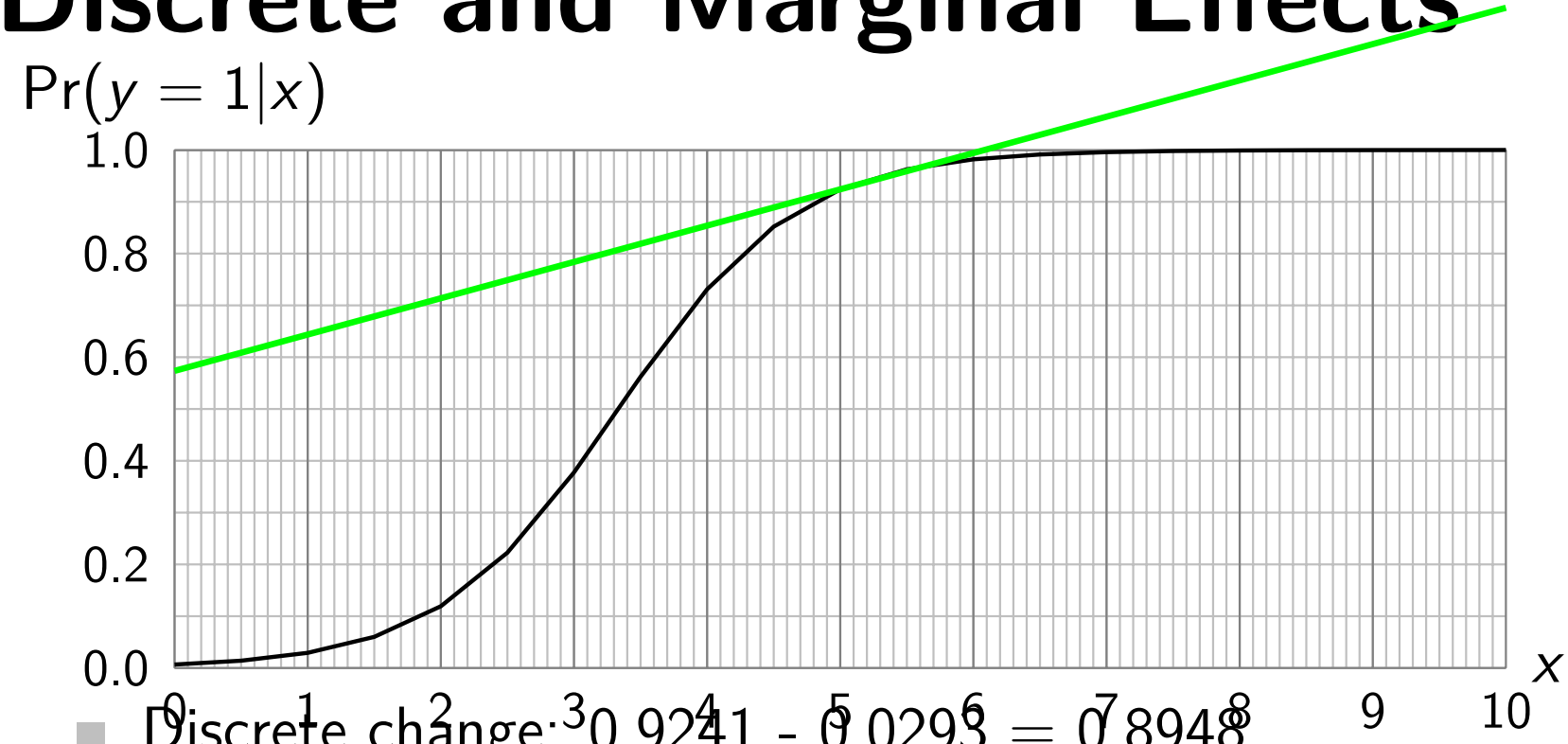
- Discrete change: 0.9241 - 0.0293 = 0.8948

Discrete and Marginal Effects



- Discrete change: 0.9241 - 0.0293 = 0.8948
- ME at $x = 1$: 0.0285

Discrete and Marginal Effects

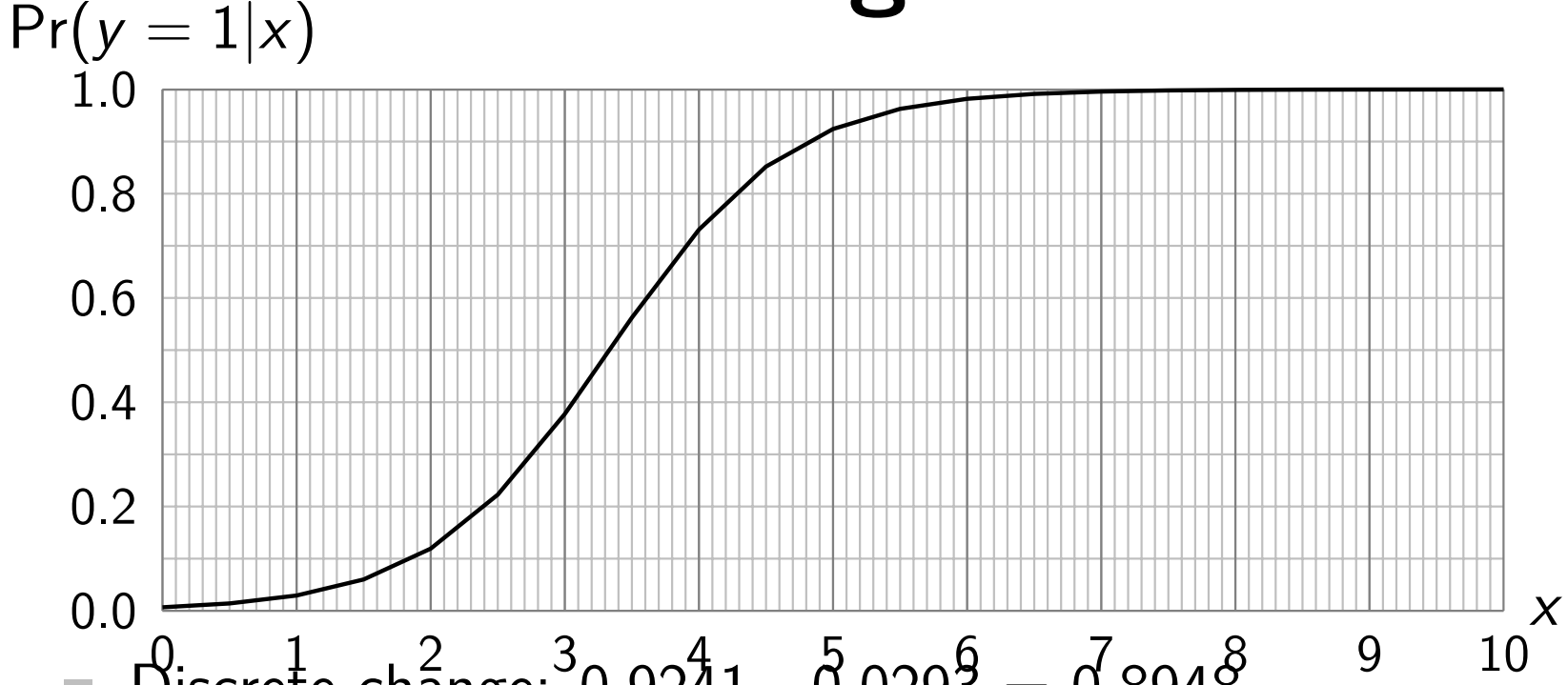


- Discrete change: 0.9241 - 0.0293 = 0.8948
- ME at $x = 1$: 0.0285
- ME at $x = 5$: 0.0701

Discrete and Marginal Effects

- Discrete change: 0.9241 - 0.0293 = 0.8948
- ME at $x = 1$: 0.0285
- ME at $x = 5$: 0.0701
- ME at $x = 3$: 0.2350

Discrete and Marginal Effects



- Discrete change: 0.9241 - 0.0293 = 0.8948
- ME at $x = 1$: 0.0285
- ME at $x = 5$: 0.0701
- ME at $x = 3$: 0.2350

Language of Interpretation

- **Discrete effect**: A change in x from a to b results in an increase in the predicted probability that y equals 1 of 0.89, which is a substantively large and statistically significant effect.
- **Marginal effect**: The marginal effect of x on the probability that $y = 1$ when x equals a is 0.07, which is a large and statistically significant effect. The predicted probability of y at this point is only 0.03, however, suggesting x may be substantively unimportant for cases with this value of x.

Summary

- Many ways to summarize GLMs
  - Coefficients
  - Predicted probabilities
  - Marginal effects
  - Discrete effects

Summary

- Many ways to summarize GLMs
  - Coefficients
  - Predicted probabilities
  - Marginal effects
  - Discrete effects
- Graphs help interpretation considerably

Summary

- Many ways to summarize GLMs
  - Coefficients
  - Predicted probabilities
  - Marginal effects
  - Discrete effects
- Graphs help interpretation considerably
- There's no single correct way of summarizing a complex model

Analyzing panel data

- Pooled analysis
- Fixed or random effects panel regression
  - Both OLS-like and GLM-like approaches possible
- Mixed effects models
  - A generalization of random effects models
  - Also works as framework for other hierarchical data structures (e.g., cross-national datasets)

Pooled Estimator

- $y_{it} = \beta_0 + \beta_1 x_{it} + \cdots + \epsilon_{it}$
- Ignores panel structure (interdependence)
- Ignores heterogeneity between units
- But, we can actually easily estimate and interpret this model!
- Estimation uses "generalized estimating equations" (GEE)
- Note: Also called *population-averaged* model

Pooled Estimator

- Continuous outcomes:
  $y_{it} = \beta_0 + \beta_1 x_{it} + \cdots + \epsilon_{it}$
- Binary outcomes:
  $y_{it}* = \beta_0 + \beta_1 x_{it} + \cdots + \epsilon_{it}$
  $y_{it} = 1$ if $y_{it}* > 0$, and 0 otherwise
- Link functions are the same in panel as in cross-sectional
  - Logit
  - Probit
- Use clustered standard errors

Respecting the Panel Structure

- With a panel structure, ϵ_{it} can be decomposed into two parts:
  - υ_{it}
  - u_i
- If we assume u_i is unrelated to X: fixed effects
- If we allow a correlation: random effects

Fixed Effects Estimator

- This gives us:

$$\begin{aligned} y_{it} &= \beta_0 + \beta_1 x_{it} + \cdots + \upsilon_{it} + u_i \\ y_{it} &= \beta_{0i} d_{it} + \beta_1 x_{it} + \cdots + \upsilon_{it} \end{aligned} \tag{5}$$

- Varying intercepts (one for each unit)
- Can generalize to other specifications (e.g., fixed period effects)

Fixed Effects Estimator

- Fixed effects terms absorb all time-invariant between-unit heterogeneity
- Effects of time-invariant variables cannot be estimated
- Each unit is its own control ("within" estimation)
- For GLMs, two ways to estimate this:
  - Unconditional maximum likelihood, or
  - Conditional maximum likelihood
  - Both are problematic

Fixed Effects Estimator

- Unconditional maximum likelihood
  - From OLS: dummy variables for each unit
  - Number of parameters to estimate increases with sample size
  - For logit/probit: *incidental parameters problem*
  - Estimate become inconsistent
- Conditional maximum likelihood
  - From OLS: "De-meaned" data to avoid estimating unit-specific intercepts
  - For logit: condition on $Pr(Y_i = 1)$ across all t periods
  - Does not work for probit!

Conditional MLE

- Estimates only based on units that change in Y
- Effects of time-invariant variables are not estimable
- Observations with time-invariant outcome are dropped

Conditional MLE

- Estimates only based on units that change in Y
- Effects of time-invariant variables are not estimable
- Observations with time-invariant outcome are dropped
- Estimation of two-wave panel using fixed-effects logistic regression is same as a pooled logistic regression where the outcome is direction of change regressed on time-differenced explanatory variables

Fixed Effects Estimator

- For linear outcomes, interpret like OLS
- For GLMs, interpretation is difficult because of the particulars of model estimation. You can:
  - In logit: predicted log-odds or log-odds marginal effects
  - In any GLM: assume fixed effect is zero

Random Effects Estimator

- If we are willing to assume that unit-specific error term is uncorrelated with other variables
- Why might this not be the case?

Random Effects Estimator

- If we are willing to assume that unit-specific error term is uncorrelated with other variables
- Why might this not be the case?
- Pooled estimator also makes this assumption
- But that estimator ignores panel structure (non-independence)

Random Effects Estimator

- Very straightforward for continuous outcomes
  - Interpretation is just like fixed effects, basically
  - Focus on predicted outcomes and/or marginal effects
- Can be used with GLMs
  - Interpretation is messy because unit-specific error terms are unobserved
  - Marginal effects either on latent scale or assume random effects are zero

Random versus Fixed Effects

- Different assumptions
- Very different estimation strategies
  - These are consequential for interpretation for non-continuous outcomes
- Common practice is to estimate multiple specifications and compare

Standard Errors

- For continuous outcomes: Works same way as OLS but typically you will cluster by unit/observation
- For GLMs, standard errors can be complicated
- For pooled model, use standard errors clustered by unit
  - `vce(robust)`
  - `vce(cluster id)`
- For random effects, you may want bootstrapped standard errors
- Always check for robustness

Intepretation: Trade-offs

- Analytic trade-off between model choice and interpretability
- Pooled estimates are interpretable in conventional ways, but use assumptions
  - Ignores panel structure
  - No unobserved confounding/heterogeneity
- Other models are harder to estimate and interpret, but may be more "correct," though:
  - RE assumes heterogeneity is not confounding
  - FE disallows effects of time-invariant variables

Mixed Effects/ Hierarchical Models

- We can also estimate mixed effects models for both continuous and non-continuous outcomes
- This can be easier analytically than other panel model specifications for non-continuous outcomes
- These models also make sense when you have data that have a true hierarchical structure (such as persons within times within countries)
- Good resource: Gelman, A. and Hill, J., 2006. *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Final thought: Weights?

- But what about weights?
- Weights have specialized use in regression modelling and any default regression modelling software is going to assume data come from an SRS
- Use specialized software if using a complex survey design

1. Basic Quantitative Analysis
   - Opinion Questions
   - Analyzing Opinion Measures
2. Hierarchical Data Collection
   - Repeated Cross-Sections
   - Survey Panels
   - Multi-Level/Comparative Data
3. Regression Analysis
   - OLS
   - Goodness-of-Fit
   - Generalized Linear Models
   - Interpreting GLMs
   - Panel Regression