Research Methods I

MICHAEL BERNSTEIN

SPRING 2013

cs376.stanford.edu

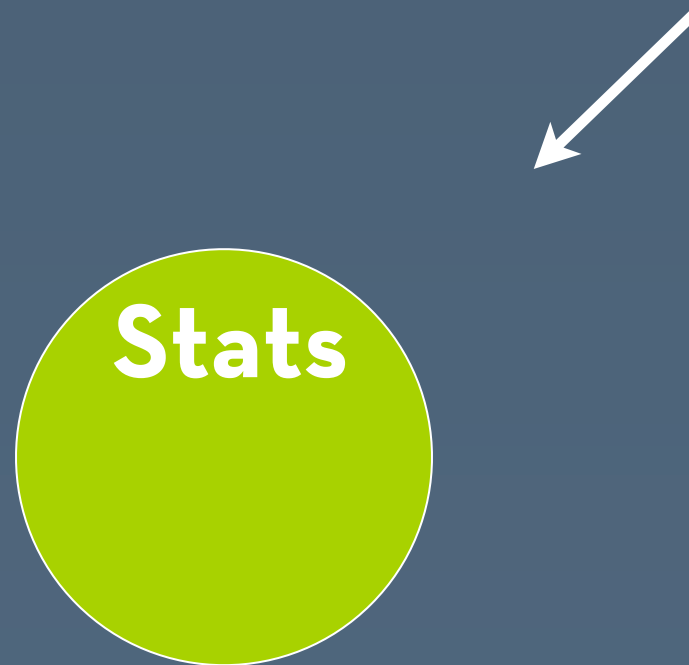
Scoping

Scoping

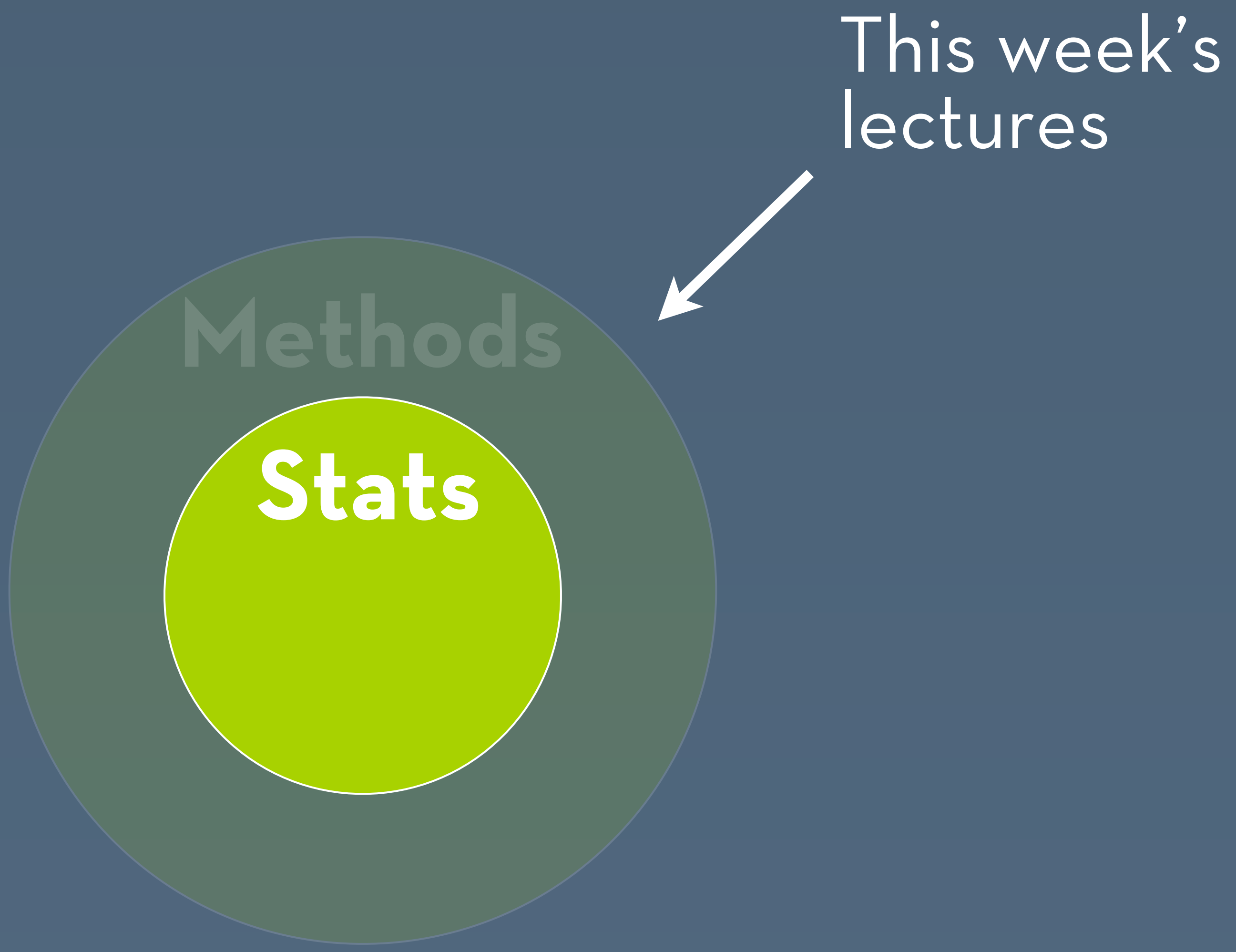
Scoping

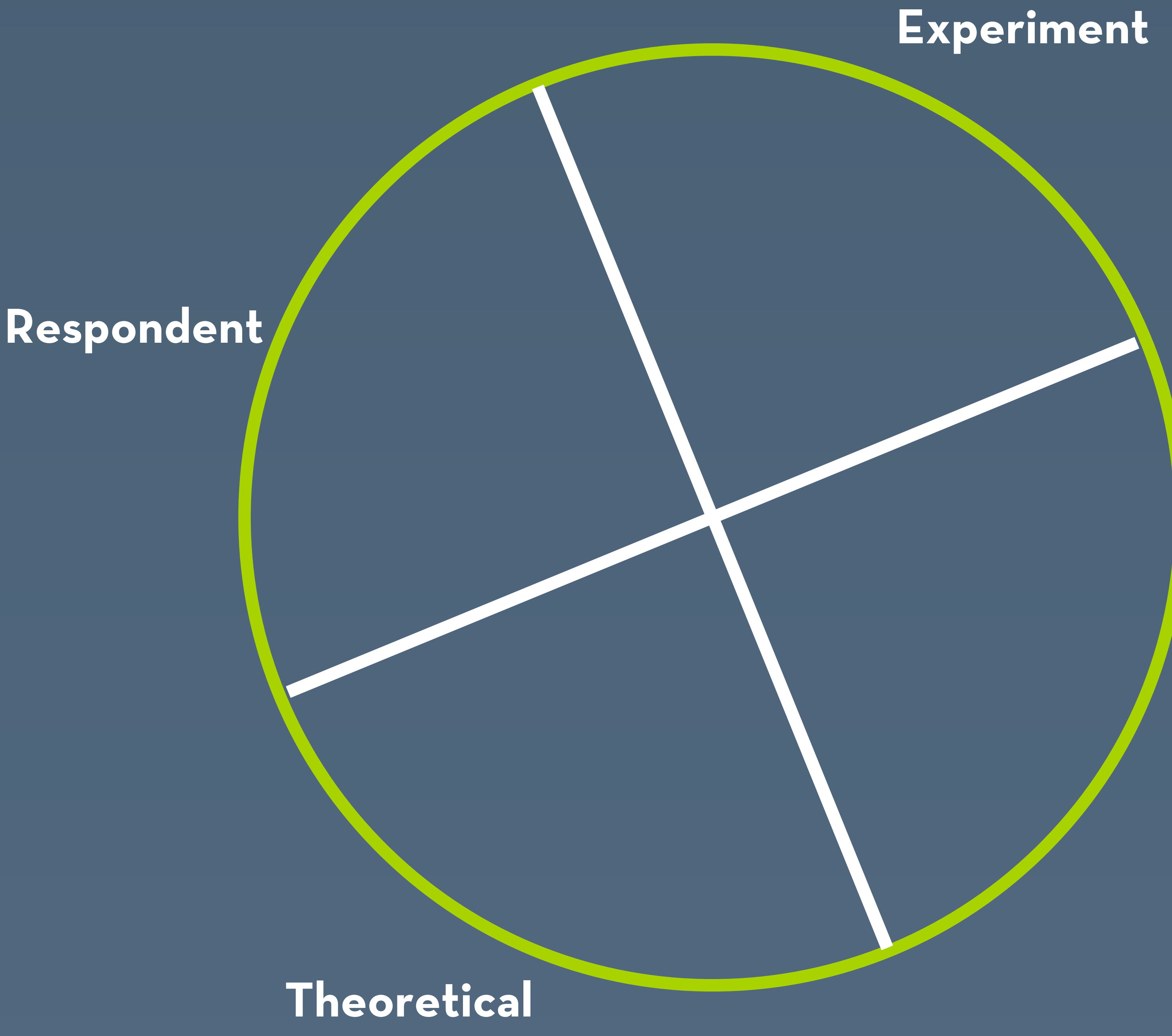
Scoping

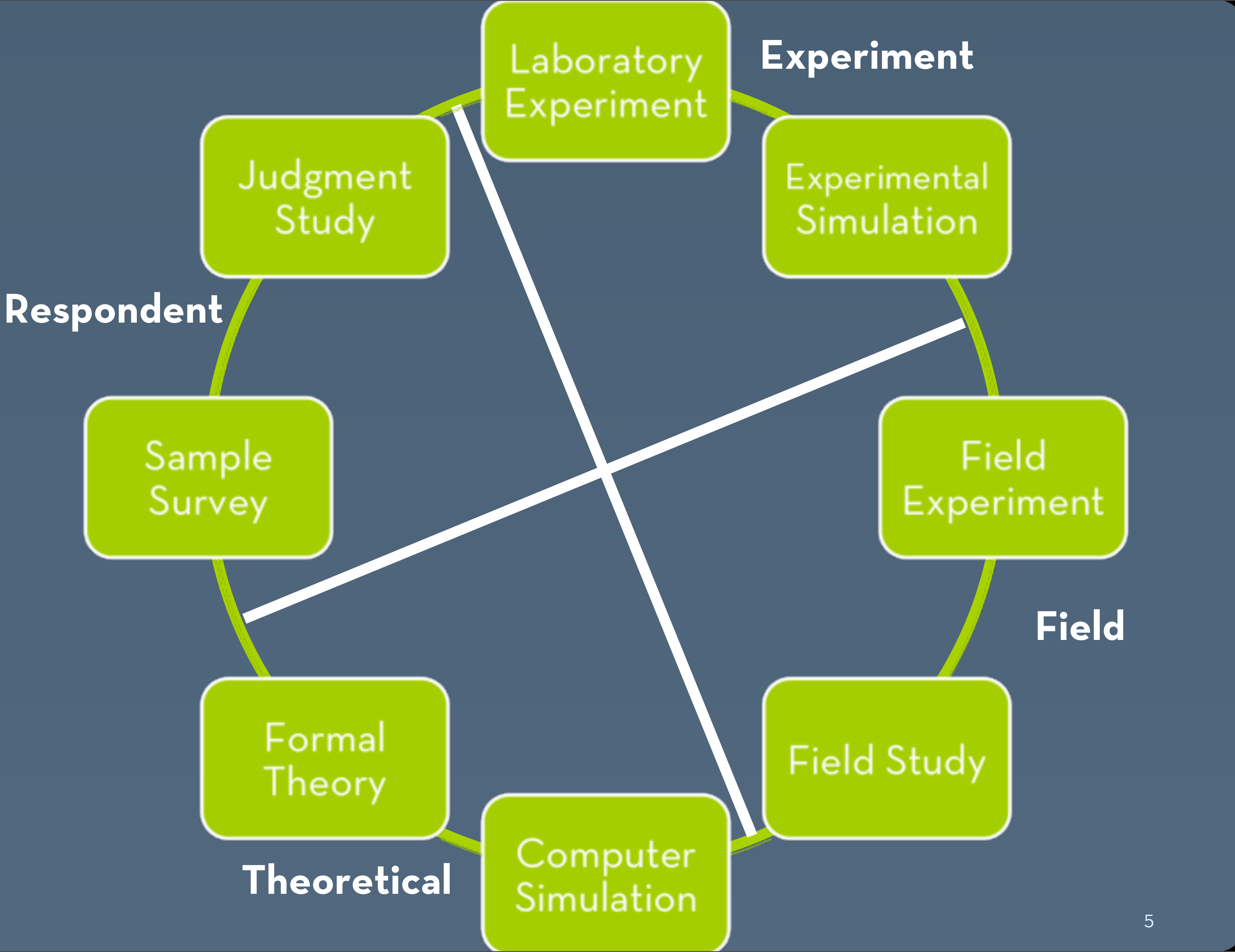
This week's lectures



Scoping




Experiment
Respondent
Field
Theoretical


Laboratory Experiment
Experiment
Judgment Study
Experimental Simulation
Respondent
Sample Survey
Field Experiment
Field
Formal Theory
Field Study
Theoretical
Computer Simulation

Method triangulation

- All methods are flawed
- Thus, your argument becomes far stronger if you can demonstrate the same phenomenon using multiple methods
  - Complement your statistics with semi-structured interviews
  - Complement qualitative work with primary source evidence or log data

Objectivity in reporting

- Readers are more cynical if that paper is presenting a one-sided argument
- Which argument do you buy?
  - "Ellipsoidal windows were better for all tasks."
    vs.
    "Ellipsoidal windows were better for all tasks we measured. However, users found them to be confusing."

Framing an evaluation

- The difficulty: defining and isolating the construct that you are trying to maximize
- Tempting to aim for something easy: time, task completion, number of clicks
- But, testing the easily quantifiable often **misses the point**.

Framing an evaluation

- Reflect on your implicit thesis about why your contribution is a good idea.
  - Skinput is a good idea because…
  - Parallel designs are a good idea because…
  - Soylent is a good idea because…
- This thesis can **directly imply the claim that you need to test**. (It may or may not be comparative in nature.)

Example theses

- Enable previously difficult/impossible tasks
- Improve task performance or outcome
- Modify/influence behavior
- Improve ease-of-use, user satisfaction
- User experience

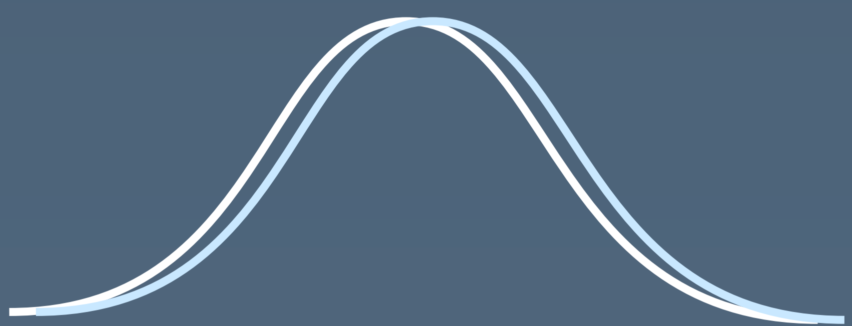
Statistics

Goal: you are more confident in the logic behind the tests you are using

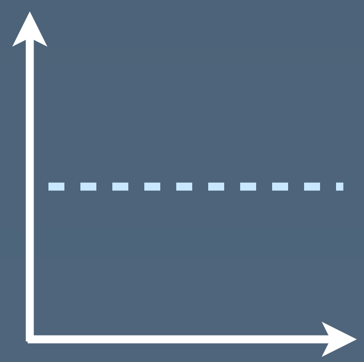
Anatomy of a statistical test

Things you know already

- If your change had no effect, what would the world look like?



No difference in means



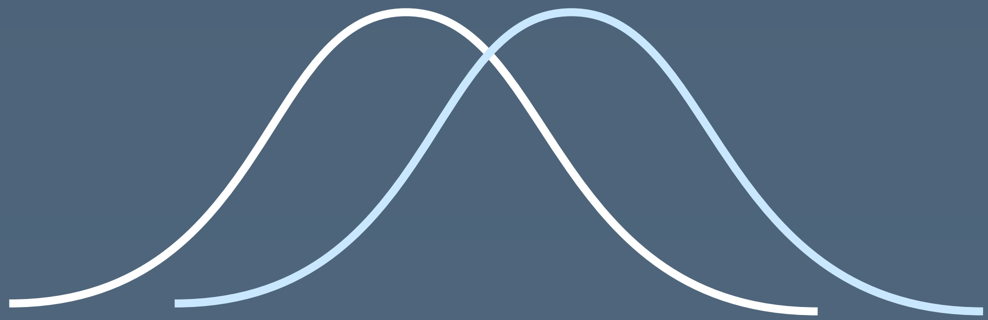
No slope in relationship

- This is known as the **null hypothesis**

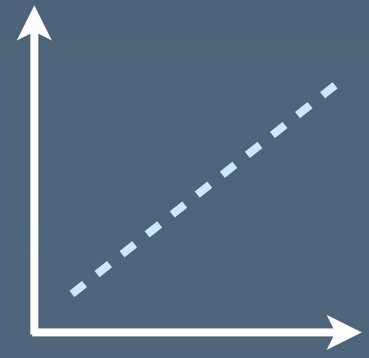
Anatomy of a statistical test

Things you know already

- Given the difference you observed, how likely is it to have occurred by chance?



Probability of seeing a mean difference at least this large, by chance, is 0.012



Probability of seeing a slope at least this large, by chance, is 0.012

Errors

Things you know already

Difference detected? \ Difference exists?	Y	N
Y	True positive	Type 1 error (publish false findings)
N	Type 2 error (get more data)	True negative

p-value

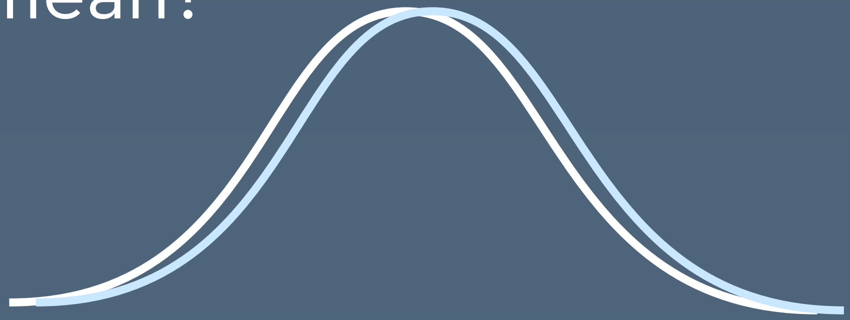
Things you know already

- The probability of seeing the observed difference by chance – e.g., P(Type I error)
- Typically accepted levels: 0.05, 0.01, 0.001

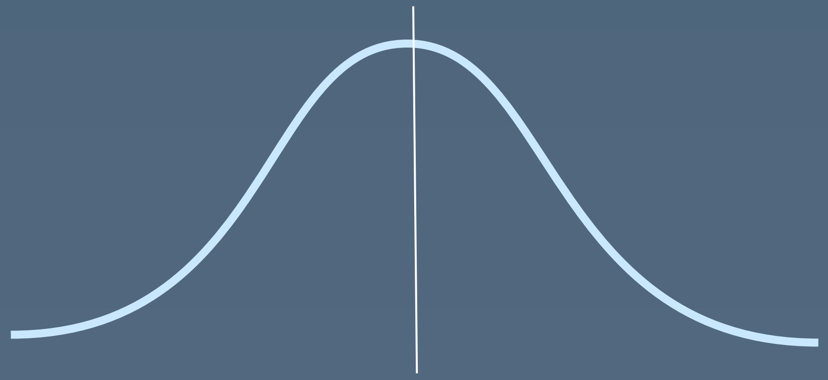
Student's t-test

Things you know already

- Do two normal distributions have the same mean?



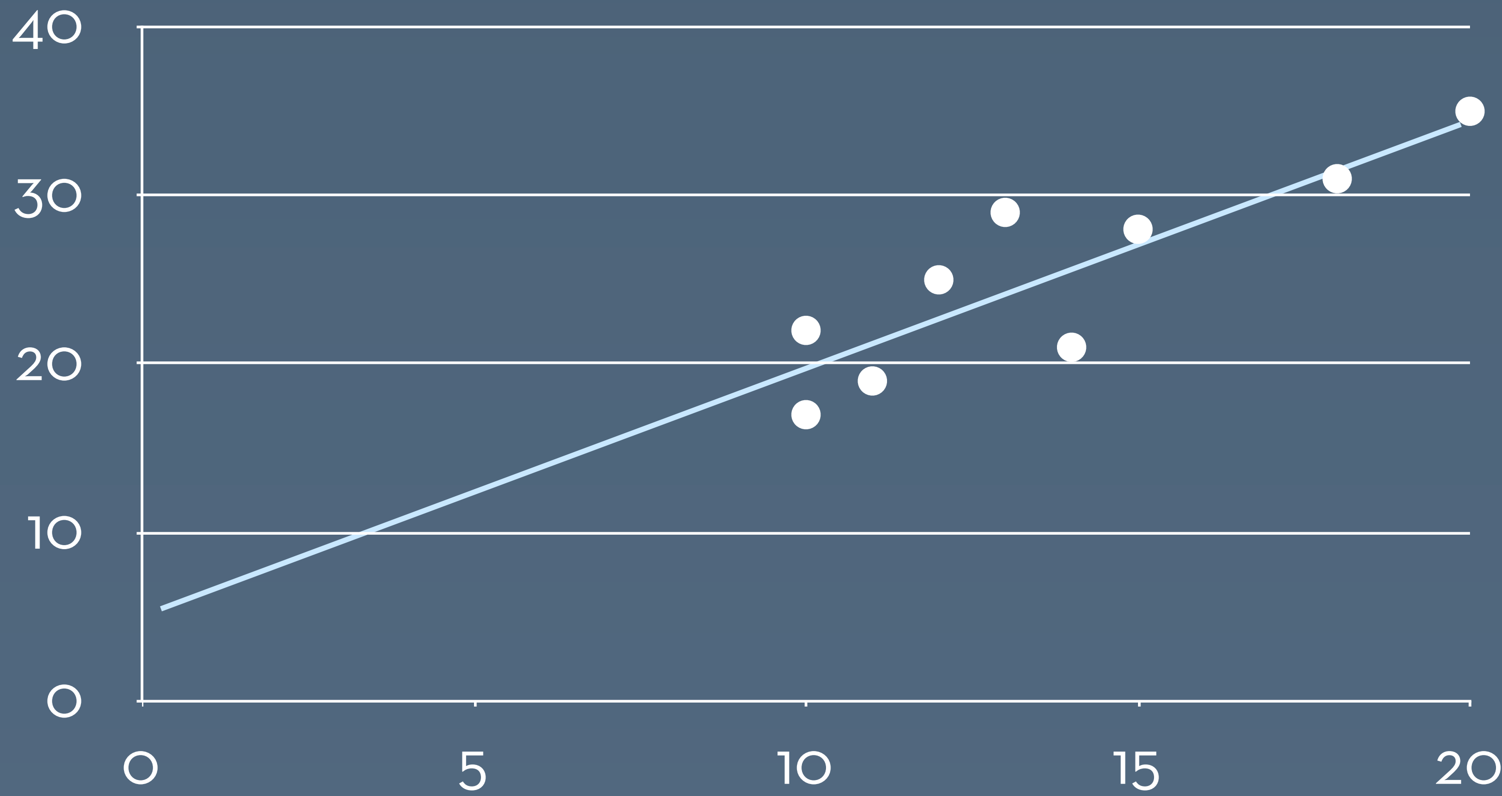
- Paired t-test: does the distribution of (after - before) have mean 0?



Demo

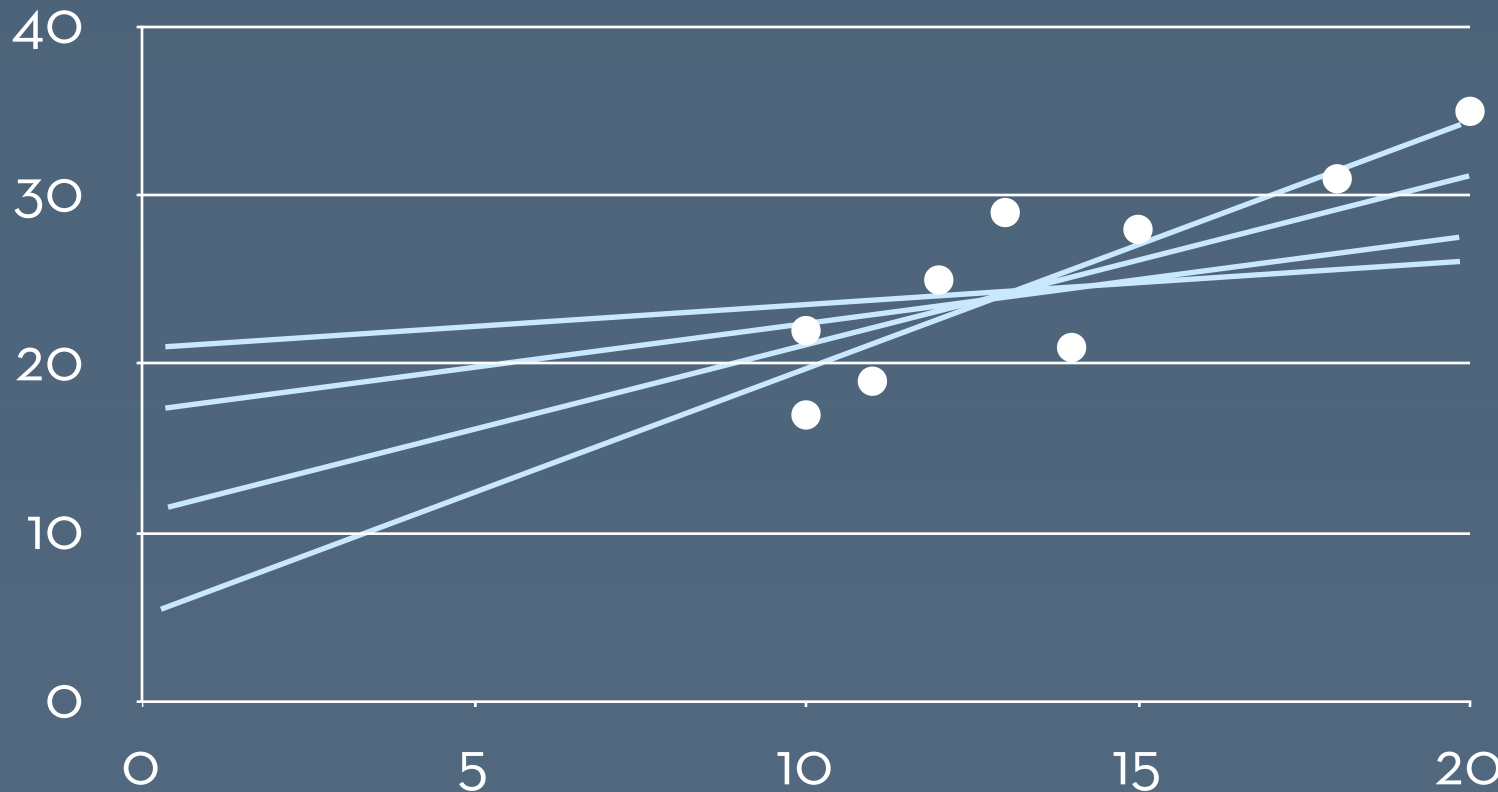
Linear regression

- Is the slope of the relationship between X and Y significantly different than 0?



Linear regression

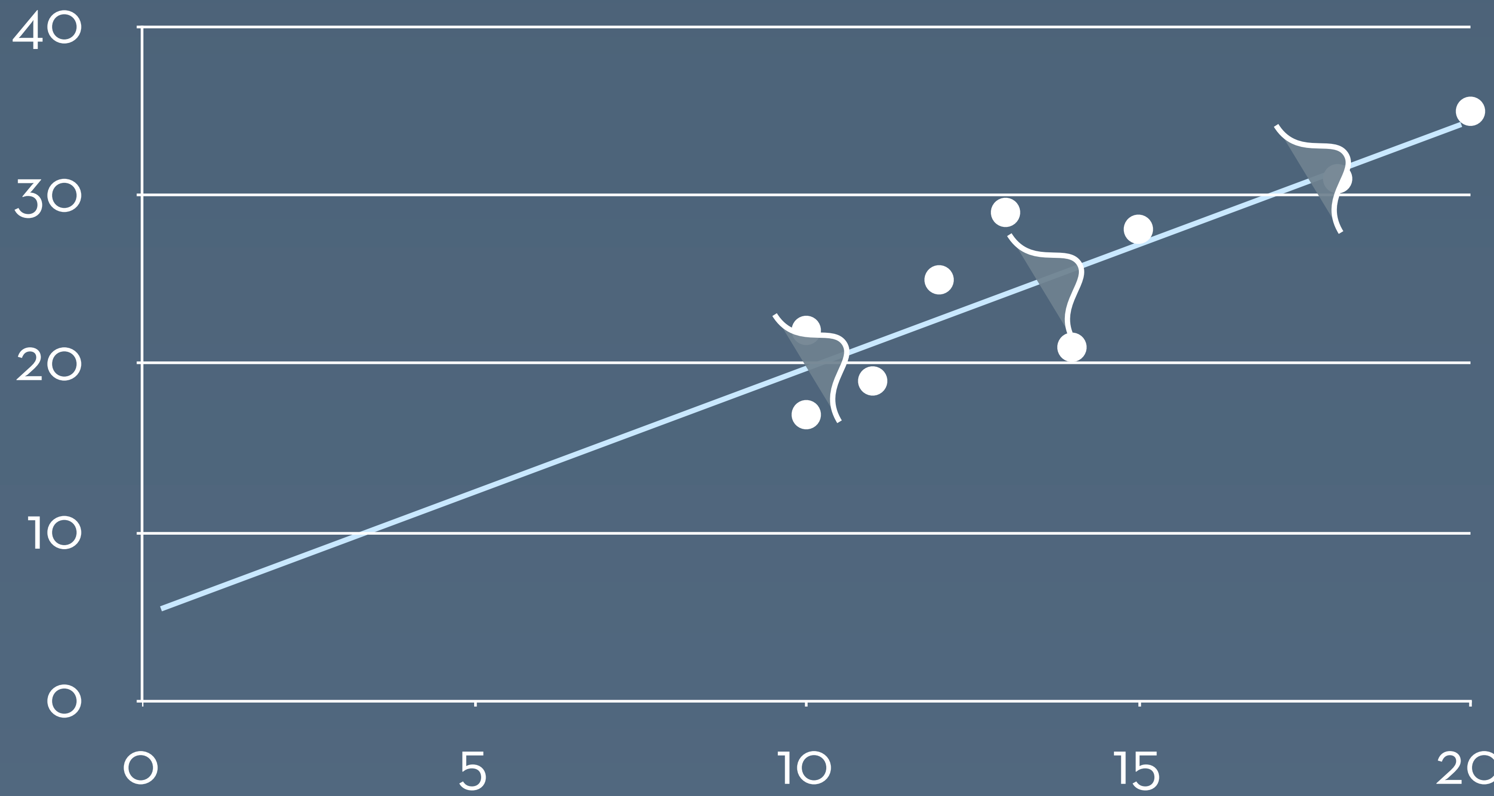
- Is the slope of the relationship between X and Y significantly different than 0?



Linear regression model

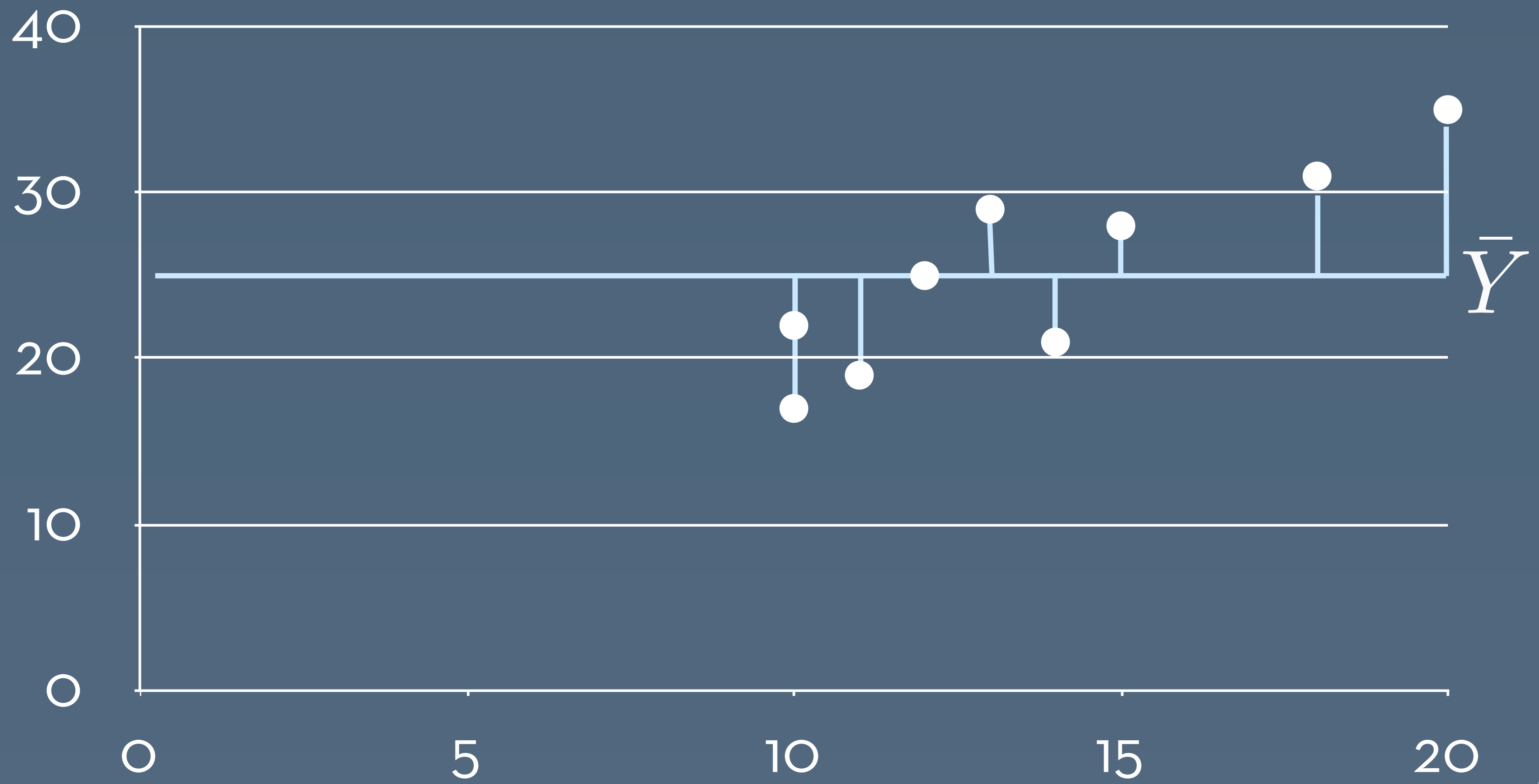
$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

intercept — **slope: ΔY for one-unit ΔX** — **error** — **n data points**



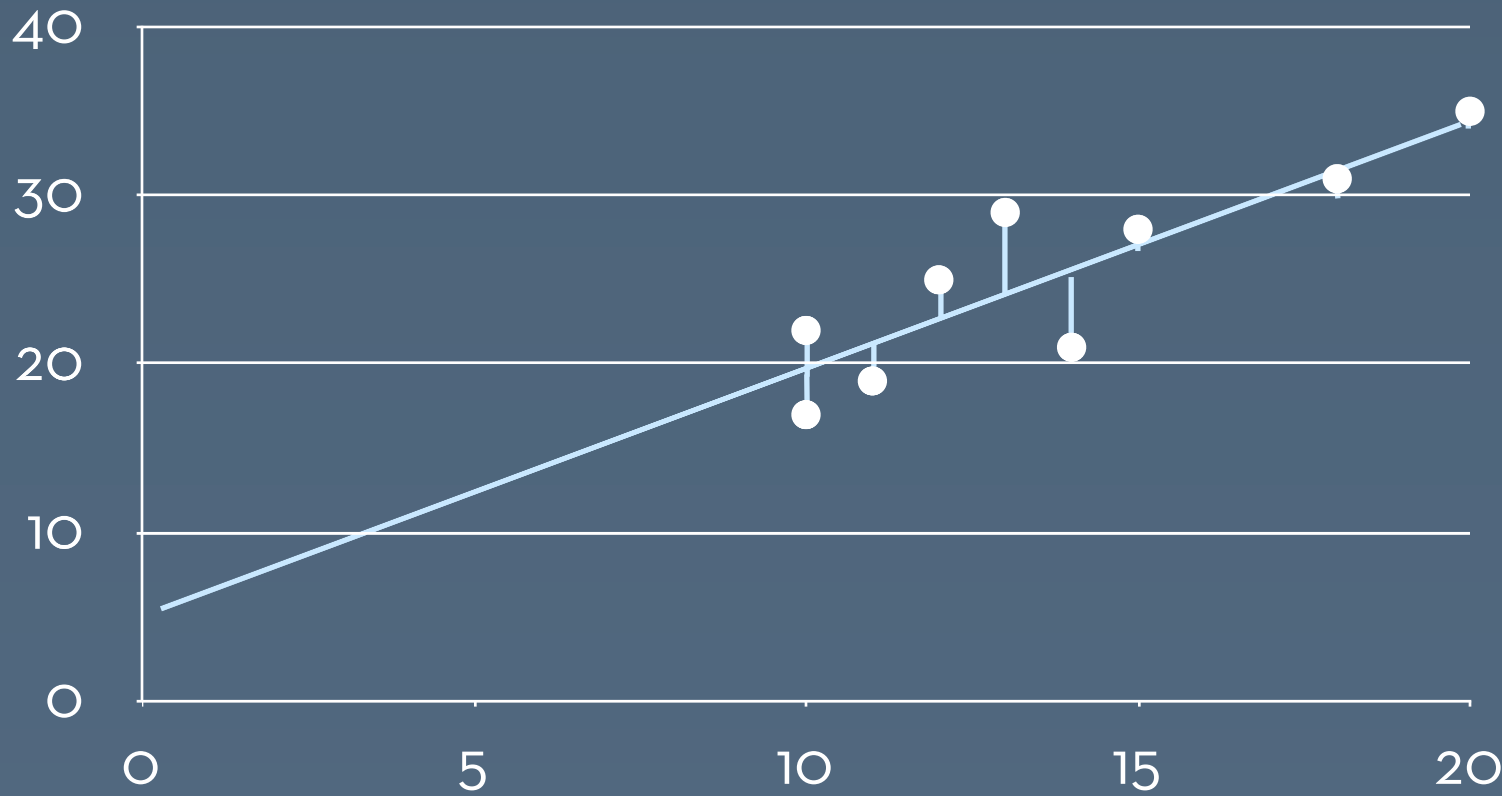
Linear regression model

SSTO: sum of squared deviations versus grand mean $\bar{Y}$



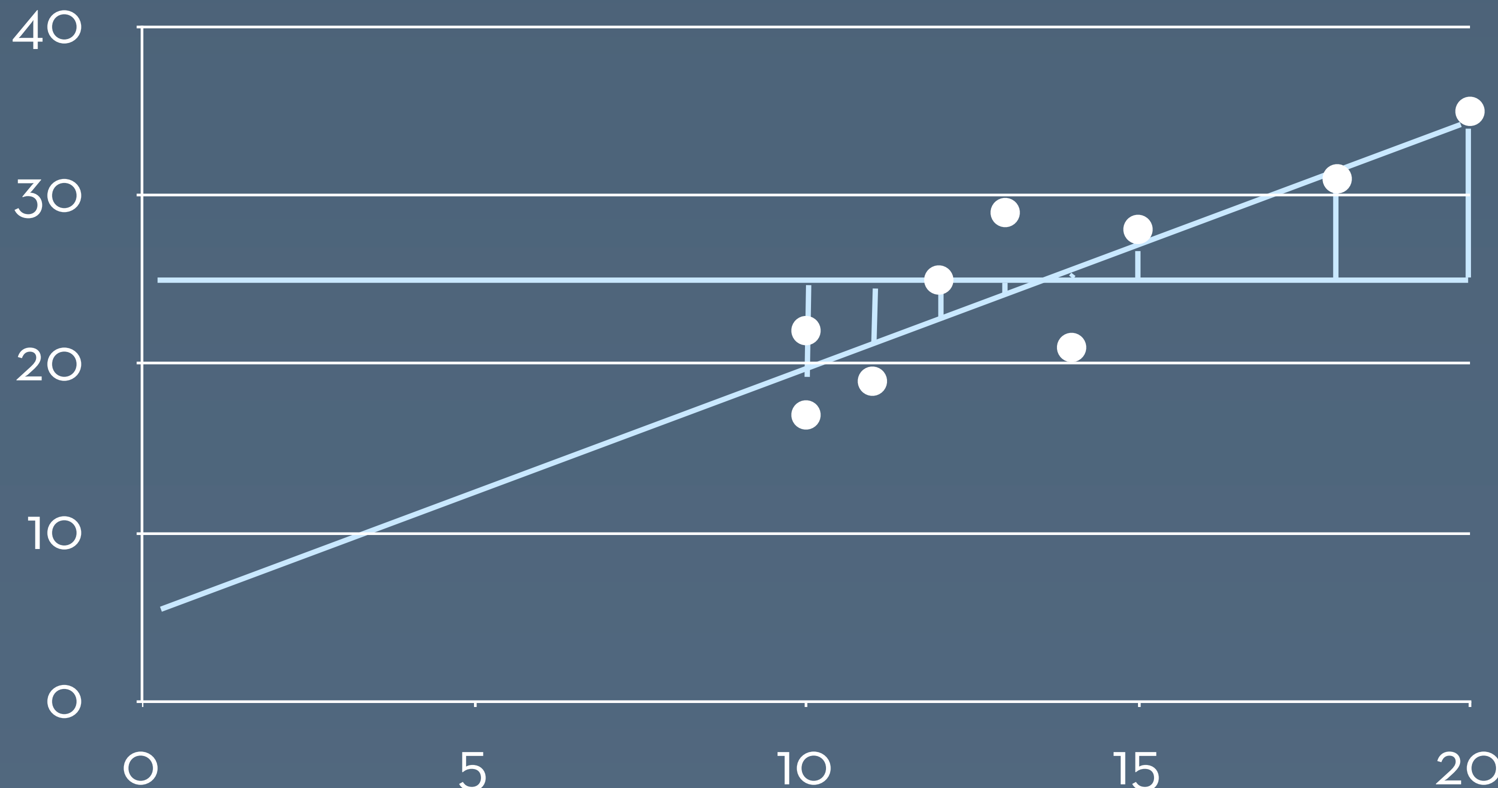
Linear regression model

SSTO: sum of squared deviations versus grand mean $\bar{Y}$
SSE: sum of squared errors versus regression line



Linear regression model

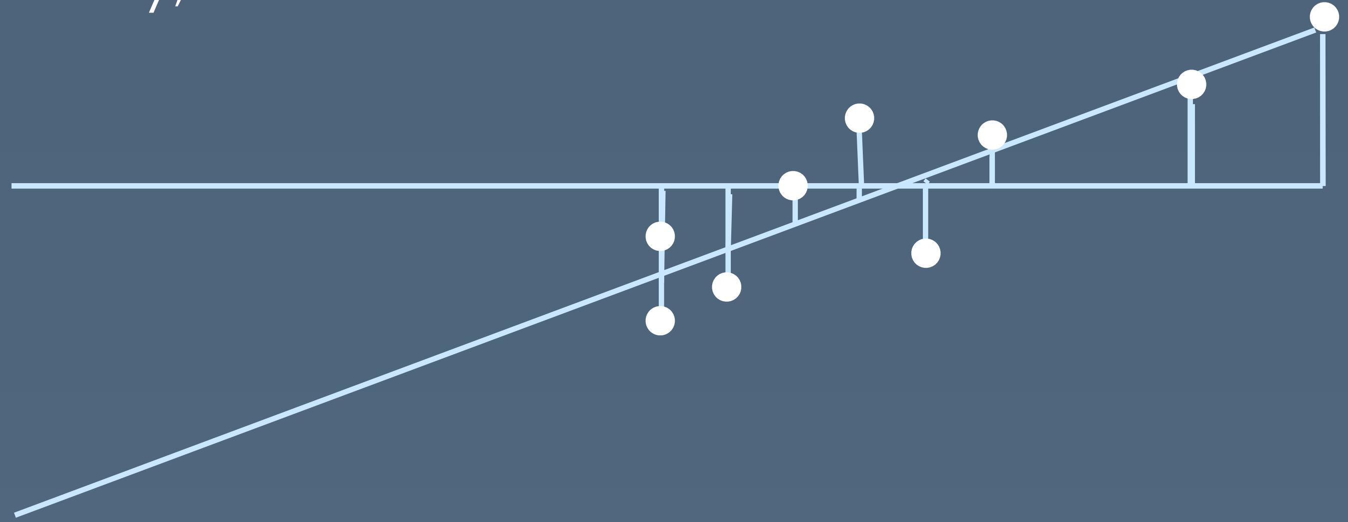
SSTO: sum of squared deviations versus grand mean $\bar{Y}$
SSE: sum of squared errors versus regression line
SSR: sum of squared deviations from regression line
to the grand mean



Linear regression model

SSTO: sum of squared deviations versus grand mean $\bar{Y}$
SSE: sum of squared errors versus regression line
SSR: sum of squared deviations from regression line to the grand mean

Remarkably, SSTO = SSE + SSR



Linear regression description

- Coefficient of determination:
  how related are X and Y?

- Put another way: what proportion of the variance in Y does a regression line explain?

$$R^2 = \frac{SSR}{SSTO} = 1 - \frac{SSE}{SSTO}$$

Demo

Linear regression tests

- R^2 does not test the relationship. So, we ask: Does $\beta_1 \neq 0$?

Linear regression tests

- R^2 does not test the relationship. So, we ask: Does $\beta_1 \neq 0$?

  $\frac{b_1 - \beta_1}{s\{b_1\}}$, the standardized difference between the sampled and true beta, is distributed as $t(n-2)$

Linear regression tests

- R^2 does not test the relationship. So, we ask: Does $\beta_1 \neq 0$?

  $\frac{b_1 - \beta_1}{s\{b_1\}}$, the standardized difference between the sampled and true beta, is distributed as $t(n-2)$

  So, test if our sampled b_1 is beyond that range:

Linear regression tests

- R^2 does not test the relationship. So, we ask: Does $\beta_1 \neq 0$?

$\frac{b_1 - \beta_1}{s\{b_1\}}$, the standardized difference between the sampled and true beta, is distributed as $t(n-2)$

So, test if our sampled b_1 is beyond that range:

$$|\frac{b_1}{s\{b_1\}}| > t(1 - \alpha/2; n-2)$$

standardized statistic **t** **confidence range (.05)** **degrees of freedom**

Reporting a linear regression

- "The bonding social capital scale was a significant predictor of sharing volume ($b=1.98$, $t(13)=12.18$, $p<0.01$).

  This single predictor explained much of the variance in sharing volume ($R^2=.91$)."

Multiple regression

- More than one predictor variable: double (or N-tuple) the fun!
- Model:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_m X_{im} + \epsilon_i$$

factor	β	p
Seeking	0.74	< .001
Bridging social capital	0.22	< .05
Bonding social capital	0.01	0.33

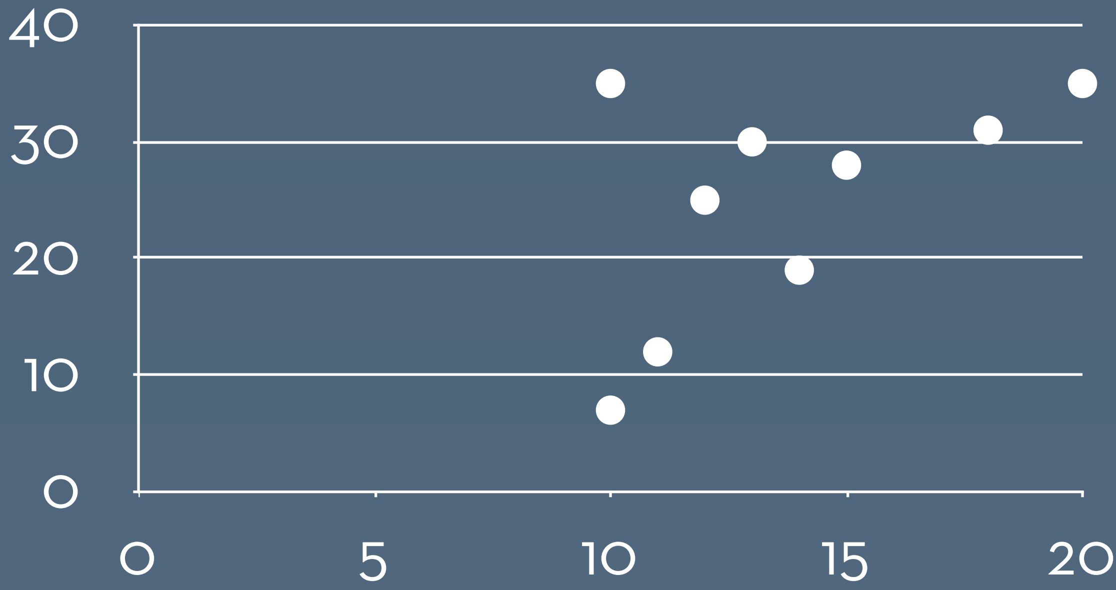
Adj. R^2 = 0.56

Multiple regression

- If a predictor is not part of your theory, you can use multiple regression to **control** for it
  - e.g., predict sharing interest by using seeking scale, bridging and bonding social capital, and controlling for age
- This has no impact on the regression mechanics — it is a reflection of your theory

Important assumptions

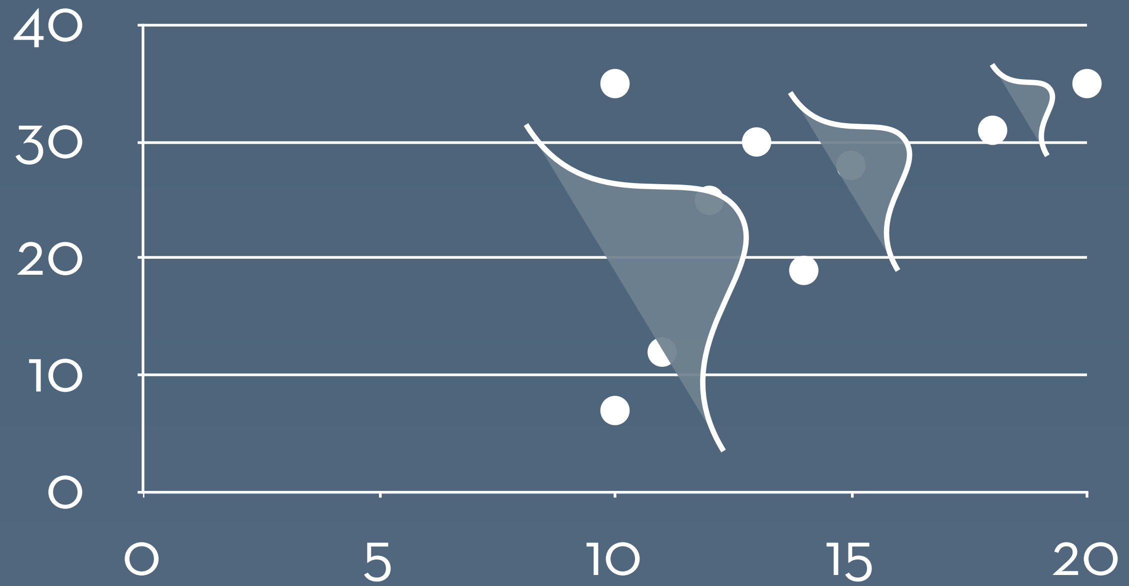
- Errors are normally distributed around regression line
- No heteroskedasticity (use Levene's test)



- Regression predictors are uncorrelated

Important assumptions

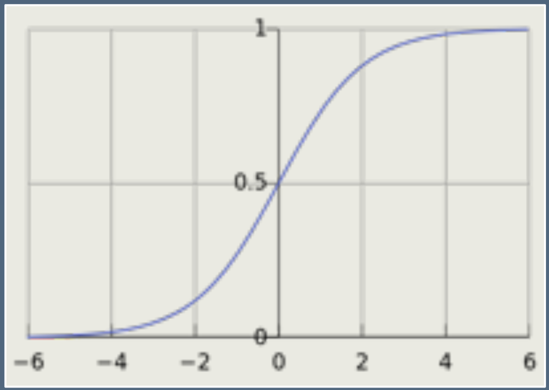
- Errors are normally distributed around regression line
- No heteroskedasticity (use Levene's test)



- Regression predictors are uncorrelated

Logistic regression

- Predicting a **binary** outcome when you are trying to control for other variables
  - e.g., predict user abandonment using training level and age
- Instead of fitting a line, the system fits a **logistic curve**: more weight toward 0 and 1
- Warning: beta coefficient interpretation is now in terms of **odds**



Next: ANOVA