Unit 7

Monday, August 10, 2020 4:39 PM

Main Topics

- Heterogeneous treatment effects (HTEs): same treatment having different effects on different subjects.
- Use of regression to measure HTEs.
    - Interaction between covariates of interest and treatment variable is important.
- Multiple comparisons problem: Fishing expeditions can cause overstatement of true statistical significance.

Assigned Reading

- Read the Chapter 9 introduction in *Field Experiments*.
- Skim Section 9.1.
    - Be sure to read the last paragraph.

Quiz Question Discussion

- Potential outcomes are hypothetical population parameters.
- In real-world samples, we get to measure only:

- Potential outcomes are hypothetical population parameters.
- In real-world samples, we get to measure only:
  - Treatment outcomes for the treatment group
  - Control outcomes for the control group
- We can measure only treatment outcomes *or* control outcomes for a single person.

Assigned Reading

- Read Section 9.3.1 of *Field Experiments*.
  - Do different groups have different treatment effects?

HTE Example: Electricity Consumption

- Test groups: below-average and above-average electricity consumers
- Treatment: social comparison (i.e., being told how much neighbors are using)

HTE Example: Congressmen

- Test groups: congressmen from the North and the South
- Treatment: being told about a meeting with a donor versus a constituent
- HTE: responsiveness of different congressmen

HTE Example: eBay Shipping Prices

- Treatment: changing the shipping price on an auction.
- Test groups: buyers of high-priced and low-priced items.
- Some measure of value should be predetermined before the experiment.
- HTE: responsiveness to shipping costs on high-value versus low-value items.

HTE Example: eBay Seller Reputation

- Treatment: changing the seller's reputation
- HTE: responsiveness for low-priced and high-priced items

HTE Example: Donation Matching

- Treatment: donation matching
- HTE: responsiveness to donation matching in liberal versus conservative states

Quiz Question

- Answer: students whose parents' literacy was above the median.
- Treatment effect not statistically significantly different between students whose parents have above-median versus below-median literacy.
- Remember: We need to know standard errors in order to evaluate how much to believe a point estimate.

Measuring HTEs: Two-Sample Test

1. Split data into two separate samples.
2. Compute means and standard errors of the treatment effect in each group.
3. Do a two-sample test of means.
    - Group one: students whose parents have above-median literacy
    - Group two: students whose parents have below-median literacy

Estimating HTEs: Regression

1. Run a regression test with:
    - Dummy variable for treatment (*I*)
    - Dummy variable for covariate (*P*)
        - *P* stands for parents.
    - Interaction term (I · P)
2. Test if the interaction coefficient is zero.
    - Explains effectiveness of treatment with the given covariate

$$Y_i = \beta_0 + \beta_1 \cdot I_i + \beta_2 \cdot P_i + \beta_3 \cdot I_i \cdot P_i$$

Reading Clarification

- Always present standard errors with point estimates.
- Always show the number of observations.
- Standard results format:

Coefficient
(Standard error)

- Present N, R2.
- Gerber and Green do an F-test for restricted regression.
    - More complicated than necessary
- If we are testing whether a particular coefficient equals zero.
    - Only a t-test is needed.

Reading Clarification (contd)

Reading Clarification (contd)

$Y_i = \beta_0 + \beta_1 \cdot I_i + \beta_2 \cdot P_i + \beta_3 \cdot I_i \cdot P_i$

- If $\beta_3 \neq 0$, treatment impact varies with different levels of covariate.
- Gerber and Green do randomization inference.
  - Unnecessary with many observations according to the central limit theorem
    - Theorem applies to regression coefficients.
  - Necessary only with small sample sizes
- A 95% confidence interval standard should generally be used.
  - Two times the standard error

Teacher Incentives Example

- Paper by Muralidharan and Sundararaman (2011).
- Full table is reproduced below. We will highlight sections of it in these slides.
- Dependent variable: student exam score.
  - Measured in units of standard deviation (also known as normalized scores)
- Treatment: students' teacher given monetary incentive based on exam scores.
- Each column represents a student characteristic.

Teacher Incentives Example

	Dependent Variable: Normalized Student Test Score
Teacher Incentive (*I*)	−.198
	(.354)
Covariate (*P*): Log School Enrollment	−.065

	(.354)
Covariate (*P*): Log School Enrollment	−.065
	(.058)
Interaction (*I*·*P*)	.083
	(.074)
Observations	29,760
R2	.244

- Teacher incentive is a dummy variable (i.e., a binary value).

Teacher Incentives: Male Covariate

Male (*P*)	Incentive (*I*)	Interaction (*I*·*P*)
1	1	1
1	0	0
0	1	0
0	0	0

Teacher Incentives: Male Covariate

	Dependent Variable: Normalized Student Test Score
Teacher Incentive (*I*)	.233***
	(.049)
Covariate (*P*): Male	.029
	(.027)
Interaction (*I*·*P*)	−.02
	(.034)

Interaction (*I*·*P*)	−.02
	(.034)
Observations	25,881
R2	.266

- .233 is the effect of teacher incentives on female students.
- Asterisks indicate statistical significance.
- Interaction: How much more effective is treatment for students with Male = 1 instead of Male = 0?
- Treatment effect for male students: .233 + −.02 = .213.

Estimation Equation

We next test for heterogeneity of the incentive treatment effect across baseline student, school, and teacher characteristics by testing whether δ3 is significantly different from zero:

$$\begin{aligned} T_{ijkm}(Y_n) = \alpha + \gamma \cdot T_{ijkm}(Y_0) + \delta_1 \cdot \text{Incentives} \\ + \delta_2 \cdot \text{Characteristic} \\ + \delta_3 \cdot (\text{Incentives} \times \text{Characteristic}) \\ + \beta \cdot Z_m + \varepsilon_k + \varepsilon_{jk} + \varepsilon_{ijk} \end{aligned}$$

- *T* = test score in year *Y*.
- Covariate improves precision.
  - Helps reduce residual variance
- Subscript indexes:
  - *i* = individual student
  - *j* = grade
  - *k* = school
  - *m* = mandal (school district)
  - *n* = year of treatment

Estimation Equation (contd)

$$\begin{aligned} T_{ijkm}(Y_n) = \alpha + \gamma \cdot T_{ijkm}(Y_0) + \delta_1 \cdot \text{Incentives} \\ + \delta_2 \cdot \text{Characteristic} \\ + \delta_3 \cdot (\text{Incentives} \times \text{Characteristic}) \\ + \beta \cdot Z_m + \varepsilon_k + \varepsilon_{jk} + \varepsilon_{ijk} \end{aligned}$$

$$+ \theta_3 \cdot (\text{Incentives} \times \text{Characteristic})$$
$$+ \beta \cdot Z_m + \varepsilon_k + \varepsilon_{jk} + \varepsilon_{ijk}$$

- Regressors of interest:
  - Incentives
  - Characteristic
  - Zm = fixed effects by mandal
- Fixed-effect dummy variables should be uncorrelated with each treatment variable.
- ε terms indicate errors correlated within a grade or school.
  - Standard errors clustered at the school level

Review: Reading Regression Results

- Each column represents a separate regression equation.
- Dependent variable: students' test score after treatment.
  - Normalized by the standard deviation across students

Review: Male Covariate and Interaction Coefficients

	(7) Male
Incentive	.233***
	(.049)
Covariate	.029
	(.027)
Interaction	−.02
	(.034)
Observations	25,881
R2	.266

- NORMSDIST(.233) = 0.59

R2	.266

- NORMSDIST(.233) = 0.59
  - If the student started at the 50th percentile, the treatment would have increased him to the 59th percentile.
- Male students get scores .029 standard deviations higher than females.
- Incentive treatment works slightly less for male than female students.
- Only the main treatment effect is statistically significant.
  - No HTEs found

Next Assignment

- Quiz question

Quiz Review

- Asterisks help reader pick out statistically significant effects.
- Most columns had statistically significant treatment effects.
  - But no statistically significant HTEs

Household Wealth Interaction Effect

- Only the household affluence interaction term was statistically significant.
- Household affluence variable is difficult to interpret.
  - Interaction coefficient = .038**
    - Treatment effect increases by .038 for each additional point of household-affluence score.
  - Ranges from 0–7
  - Assumes all categories have equal treatment-effect benefits

Alternative Specification for Household-Wealth Covariate

- Alternative specification:
  - Have seven separate dummy variables as covariates.
    - One for owning land
    - One for owning a house
    - One for having running water
    - One for owning a TV
    - Etc.
  - Have seven different interaction terms.
  - This is a saturated model.

Several Specific Regressions or One Big Regression?

- Used one column for each covariate
  - Each covariate could have been put into the same regression.
    - But they weren't able to identify the covariates separately due to collinearity.

Collinearity

- Collinearity should never be a problem between treatment variable and covariate.
  - Randomization guarantees independence between them.

Teacher Incentives: Household Affluence Covariate

	Dependent Variable: Normalized Student Test Score
Teacher Incentive (*I*)	.09
	(.073)
Covariate (*P*): Household Affluence	.017
	(.014)
Interaction (*I*·*P*)	.038**
	(.019)
Observations	25,231
R2	.272

- Adding household affluence causes the estimated main treatment effect to shrink:

		3.5			
		Average affluence			
3.5	×	0.038	=	0.133	
Average affluence		Interaction term			
0.09	+	0.133	=	0.223	
Incentive term				Average treatment effect	

- Interpretation of treatment effect depends on the definition of the interaction term.
- Need to identify the average effect of treatment.

Teacher Incentives: Parental Literacy Covariate

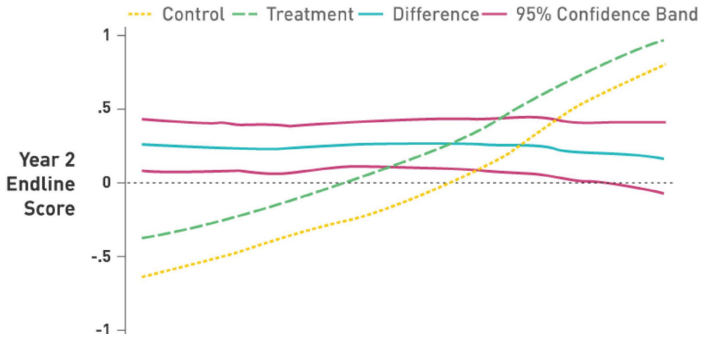
	Dependent Variable: Normalized Student Test Score
Teacher Incentive (*I*)	.224***

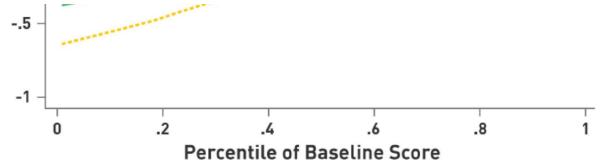
	Normalized Student Test Score
Teacher Incentive (*I*)	.224***
	(.054)
Covariate (*P*): Parental Literacy	.068***
	(.015)
Interaction (*I*·*P*)	−.003
	(.019)
Observations	25,226
R2	.273

- Always be aware of the amount of statistical uncertainty in estimates.
- Confidence interval: −.003±.038 per unit of literacy.
- The largest absolute value in that interval is -.003-.038 = -.041.
- Even using the coefficient estimate with the largest magnitude (-.041) and evaluating it at the largest parental literacy value (4), the estimated total interaction effect confidence interval is (-.041)(4) = 0.16.
- This is still smaller than the baseline treatment effect (0.224) for those with totally illiterate parents.
- Always look at magnitude of point estimate.
- Always look at size of confidence interval.

Treatment Effect Hypothesis

- Benefits of teacher incentives are broad-based.
  - Checked by sorting students by pretest scores
  - Conducted nonparametric (kernel) regression to estimate treatment effects





Review: Results on Student Characteristics

- Broad-based benefits from teacher incentive program were found.
  - No HTEs by past student test score
- Household wealth was the only significant HTE interaction.

Teacher Characteristics

- Review the table.
- Answer the quiz questions.

Quiz Review

- Most significant effects (at the 5% level) for covariates of (a) teacher training and (b) years of experience.
  - Displayed in the interaction row
- The interaction terms in columns two and three show the most significant impact.
- Teachers with more experience deliver fewer benefits based on incentives.
- Teachers with more training deliver more benefits based on incentives.
- Understanding the size of the HTEs is important.
  - Examine the magnitude of the regressors.

Teacher-Training Covariate

- Teacher training levels:
  - 1 = No training
  - 2 = Diploma
  - 3 = Bachelor's
  - 4 = Master's
- Starting with a baseline of 0 for the no training value makes coefficient interpretation easier.
- Main treatment effect in this specification is negative. Why?

	(2) Training
Incentive	−.224
	(.176)

Incentive	−.224
	(.176)

Teacher-Training Treatment Effect and Interaction

	(2) Training
Incentive	−.224
	(.176)
Covariate	−.051
	(.041)
Interaction	.138**
	(.061)
Observations	53,890
R2	.29

- −.224 is the baseline treatment effect for a teacher with training score of 0.
  - Although no teachers have a training score of 0.
  - Starting with a baseline of 0 makes coefficient interpretation easier
- Estimate is not statistically different from 0.
  - T-ratio is very low.

Estimating Confidence Intervals for Heterogeneous Treatment Effects

- Confidence interval for incentive effect (for those with training = 0) in this specification: −.224±.35
- Covariate should equal 0 (not 1) for no training.
  - Incentive coefficient would be easier to interpret.
- Few teachers in the sample had literally no training.
- What would be the estimate for a teacher who did have training?

training?

Quiz Review

- Answer: estimated treatment effect for teachers with bachelor's degrees

−.224	+	(.138)(3)	=	+.19
Incentive coefficient		Interaction coefficient × bachelor's-degree value		Treatment effect

- Positive treatment effect estimate for those with bachelor's degrees, negative treatment effect for teachers with no training.
- Easier method:
  - Split up the sample by the levels of training.
  - Estimate treatment effect for each group.
- Standard errors on the treatment effect will probably be large.
  - Both the incentive and interaction coefficients have large standard errors.

Quiz Review (contd)

	(2) Training
Incentive	−.224
	(.176)
Covariate	−.051
	(.041)
Interaction	.138**
	(.061)
Observations	53,890

	(.061)
Observations	53,890
R2	.29

- Delta method could be used to compute exact standard errors.
- If the covariate is well defined, it's good enough to:
    - Examine the size of the treatment and interaction coefficients
    - Determine if interaction coefficient is statistically significant
    - Estimate the size of associated standard errors

Conclusion

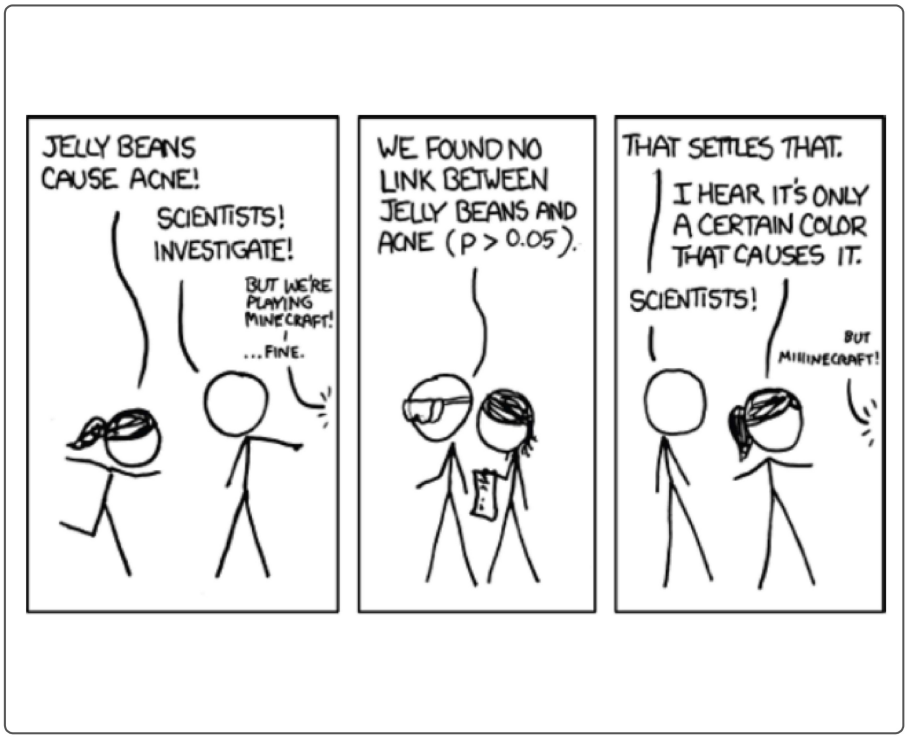
- Understand point estimates.
- Determine statistical significance of the HTE coefficient.
- Compute the point estimate of the treatment effect for different values of the covariates.

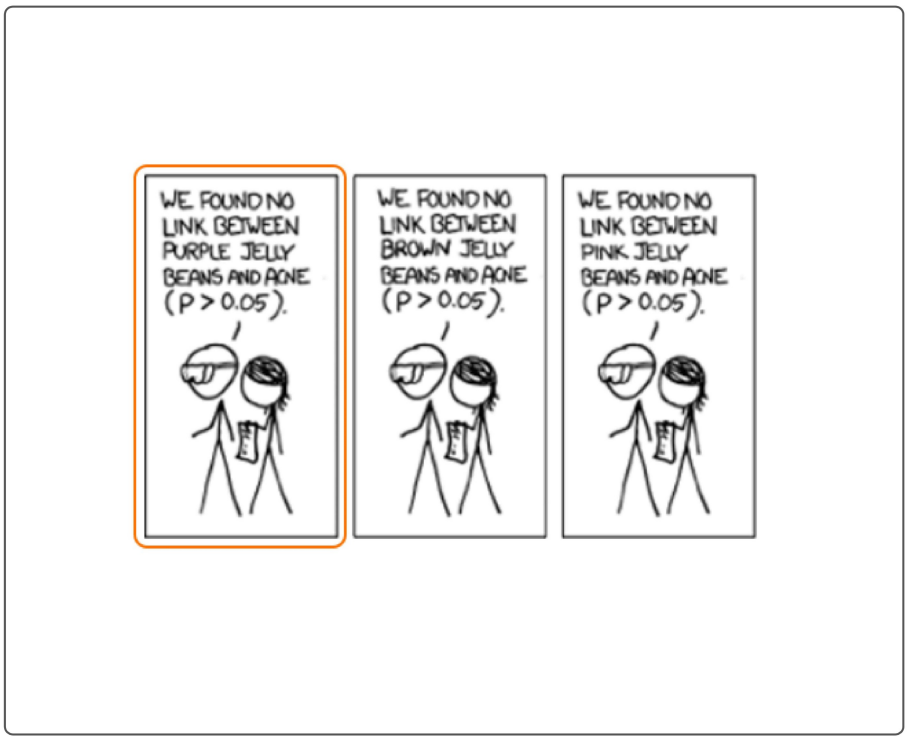
Assigned Reading

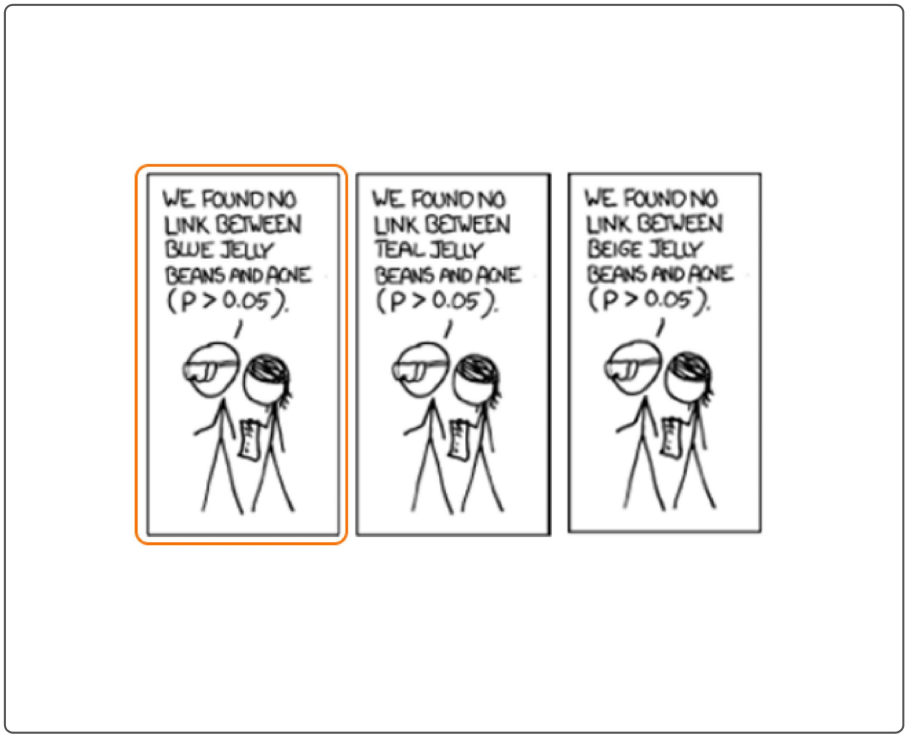
- Read Section 9.3.2 of *Field Experiments*.
- Pay attention to paragraph at the bottom of page 301.
    - Subgroup membership is nonexperimental in nature.

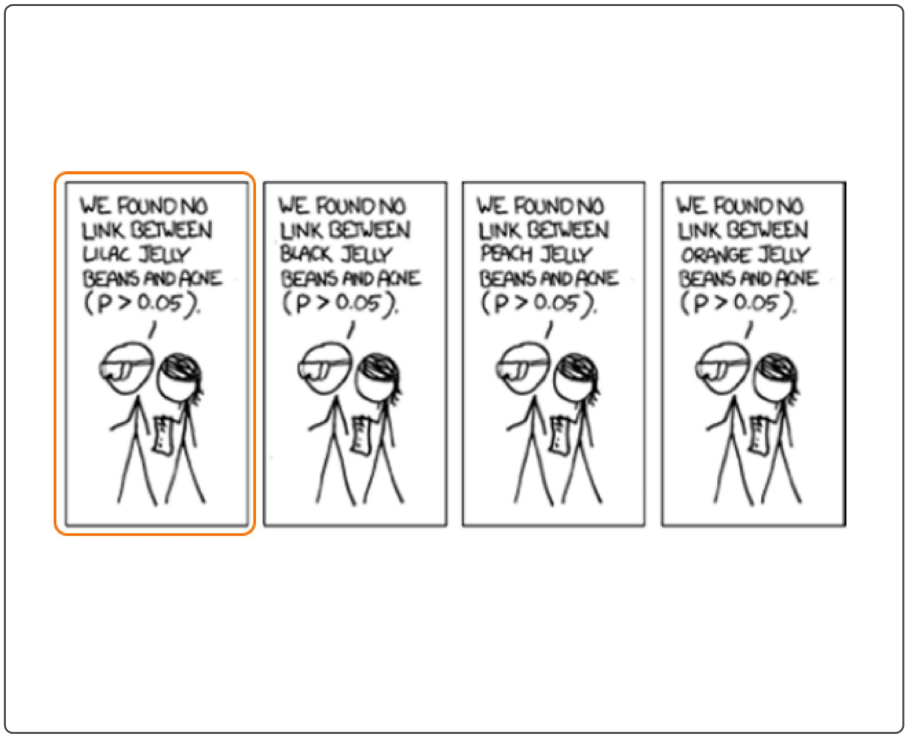
Assigned Reading

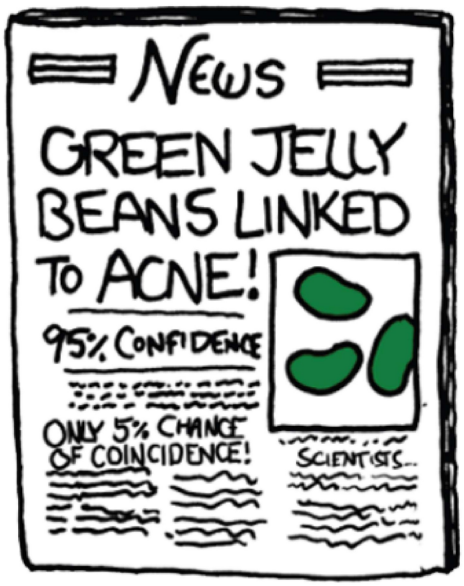
- Skim the second half of Section 9.4.
  - Example: two treatment variables (*J* and *G*) and one covariate (*H*)
    - *J*: Jose vs. Colin
    - *G*: bad grammar vs. good grammar
    - *H*: Hispanic legislator
- Read Section 9.5.
- Study Equation 9.19 and Table 9.2.
  - Equation 9.19: regression presentation of data in Table 9.2
- New scenario:
  - Three binary variables
  - Three possible two-way interactions
  - Three-way interaction


JELLY BEANS CAUSE ACNE!
SCIENTISTS! INVESTIGATE!
BUT WE'RE PLAYING MINECRAFT!
...FINE.
WE FOUND NO LINK BETWEEN JELLY BEANS AND ACNE (P > 0.05).
THAT SETTLES THAT.
I HEAR IT'S ONLY A CERTAIN COLOR THAT CAUSES IT.
SCIENTISTS!
BUT MINECRAFT!


WE FOUND NO LINK BETWEEN PURPLE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BROWN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PINK JELLY BEANS AND ACNE (P > 0.05).


WE FOUND NO LINK BETWEEN BLUE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TEAL JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BEIGE JELLY BEANS AND ACNE (P > 0.05).


WE FOUND NO LINK BETWEEN LILAC JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLACK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PEACH JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN ORANGE JELLY BEANS AND ACNE (P > 0.05).



Multiple-Comparisons Problem

- More variables means more specification searching is possible.
- Specifications can be changed until the coefficients are suitable.
- All possible covariates can be tried until statistical significance is found.
- Statistical theory assumes we know the correct model.
    - We don't always have the correct model, though.
- Some searching is inevitable.
- An effect isn't necessarily real when one coefficient was significant out of many possibilities tried.

Fishing Expeditions

- Trying out multiple hypotheses before picking a favorite.
- Violates assumptions that give valid confidence intervals.
- Analyzing data in different ways is OK.
    - But the computed confidence intervals will be too narrow.

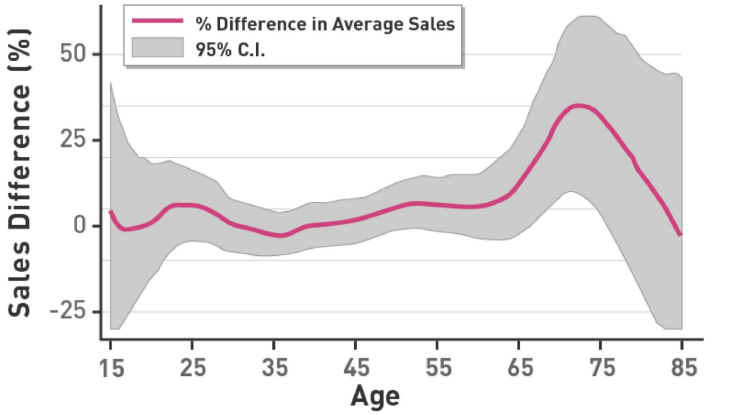
Fishing Expeditions: Solutions

- **Bonferroni correction (see Box 9.4):** helps avoid overstating statistical significance.
    - Critical values should be much higher (t=3 instead of t=2).
- Consider findings to be interesting hypotheses.
    - Test them in another experiment to ensure replication.

Lewis and Reiley Age Effect

- Found higher sales revenue from older people who watched advertisements
  - Results didn't replicate in a second experiment.

Treatment-Control Sales Difference, in Percentage



Heterogeneous treatment effects by age, from Lewis and Reiley (2014)

Teacher Incentives Effect

- Affluent households had statistically significantly larger treatment effects.
  - This was the only significant interaction term out of eight tried.
  - Bonferroni correction may have shown that this effect wasn't significant.
- Muralidharan and Sundararaman examined many covariates.
  - But properly avoided a fishing expedition
- Student ability didn't affect incentive treatment.

- Muralidharan and Sundararaman examined many covariates.
    - But properly avoided a fishing expedition
- Student ability didn't affect incentive treatment.

Quiz Recap

- Practiced converting sample means to regression coefficients.
- Examined multifactor experiments.
- Examined HTEs.
- This quiz put together HTEs and multifactor experiments.

Automated Searches for Interactions

- Main point: Fishing expeditions are a problem because of multiple comparisons.
    - Keep analysis simple.
- Machine learning can help find heterogeneous treatment effects.
    - Performs automated specification searches.
    - Multiple comparisons will still be an issue.
    - Treat any output as a hypothesis to be confirmed in another experiment.

What to Remember From This Week

- Options for reporting heterogeneous treatment effects
  - Report separate treatment effects for each subgroup.
  - Use regressions where the treatment dummy variable is multiplied by covariates of interest.
- Defining covariates so that reading output is easier
- Being able to read output of regression models with interaction terms
- Testing significance of treatment effects between subgroups
  - Can be done with a t-test on one coefficient
  - Can also be done with an F-test

What to Remember From This Week (contd)

- Interpreting HTEs
  - They explain how different subgroups respond to treatment.
  - They *don't* explain causal effects of reassigning people to new subgroups.
- Multiple-comparisons problem
  - Examining many different regression models and picking favorites makes us overstate statistical significance.
  - Results from fishing expeditions should be treated as hypotheses to be retested.
- Machine learning can help develop testable HTE hypotheses.

hypotheses to be retested.

- Machine learning can help develop testable HTE hypotheses.