

Introduction

- MN Department of Revenue tax-compliance experiment
 - Anticipated heterogeneous treatment effects
 - People with high income have more incentive to hide income
 - People with more opportunity to hide income (self-employed) more likely to underreport
 - Without oversampling, treatment group of 1500 taxpayers contained only 7 high-income, high-opportunity earners
 - Oversampling increased this to 80

When to Oversample

- When we have a particularly interesting subpopulation
 - High-opportunity taxpayers
 - Girls in computer-programming course
 - Minority legislators in study about response to minority constituents
- When interesting subpopulation represents small fraction of overall population
- When important subpopulation has relatively high variance of Y
 - If standard deviation is many times the mean
 - Large samples can overcome huge variance

Adaptive Sampling

- Adjust experiment based on responses
 - E.g., A/B test reveals one subpopulation has high variance of purchases
 - Oversample high-variance group

Summary

- If a group is either very small or has high variance, oversampling is prudent

Up Next

- Measuring outcomes

Please read section 12.3 on catalog mailing experiment of Simester *et al.* (2009).

Simester, *et al.* (2009) Experiment

- Catalog mailing experiment
 - Long run effects differ from short run effects
 - Extra 5 catalogs in 8-month experiment produced positive return on investment
 - Measured effect smaller with more broadly-defined outcome
 - Majority of effect from accelerating purchases forward in time
 - Negative treatment effect on purchases for 8 months following experiment
 - Negative treatment effect for website purchases for same retailer

Comparison 1: Lewis and Reilly Advertising Experiment

- Outcome defined more broadly
- Online ads had positive treatment effect on online sales, as well as sales in brick-and-mortar stores
- Study included examination of purchases for 10 weeks after conclusion of 2-week ad campaign
 - No evidence of acceleration of purchases
- Advice: pay attention to outcomes of actual concern (in this case, total sales)

Comparison 2: Employee Reciprocity Study

- Labor economics experiment
- Lab experiments with two people in job market scenario
 - Employer, employee
 - Employer decides how much to pay employee
 - Employee decides how hard to work for employer
 - Effort level between 0 and 10
 - Cash payout to employee subtracts effort from what employer paid
 - Game theory predictions:
 - Employee has received wage; no incentive to engage in costly effort
 - Employer has no incentive to pay more than minimum wage

Comparison 2: Employee Reciprocity Study (cont.)

- Actual results:
 - Most employers set wages well above minimum
 - Employees who receive higher wage offers tend to work harder for employer than employees who receive low wage offers
- Treatment effect of offer received
- Evidence of reciprocity effect between workers and wage-setting employees

Comparison 3: Gneezy and List (2006)

- Field Experiment
 - Students hired to solicit charitable donations
 - Half received higher-than-advertised wage
 - In first hour, higher-paid students brought in more donations (reciprocity treatment effect)
 - After a few hours (or days), treatment effect diminished
 - Redefined outcome leads to different conclusion

Conclusion

- Long-run effects are difficult to measure in laboratory experiments
- Treatment effects may wear off after longer period of time

Assignment

- Please read 12.4 in Field Experiments textbook (Bertrand & Mullainathan Audit Study)
 - Field experiment where researcher doesn't necessarily follow through with a completed transaction

Recap

- Bertrand and Mullainathan job discrimination study
 - Drew from identical resumes with white-sounding names vs. black-sounding names
 - Showed clear evidence of discrimination against black-sounding candidate names
 - Proves causal effect of race

Bertrand and Mullainathan Audit Study: Advantages

- Drawing from variety of names
- Drawing from variety of high- and low-quality resumes
 - Yields more generalizable treatment effect
- Sending four resumes (black-sounding, white-sounding, high- and low-quality) per firm to identify some within-employer effects
 - Which candidates were called back?
 - What did distribution of discrimination look like?
- Is double-blind (requires absence of actual interviews)

Aside: Car Dealer Audit Study

- Scenario: black and white consumers visited car dealers
 - Followed script to counter with lower price than price offered by dealer
- Result: Black consumers ended up with a higher negotiated price
- Problems:
 - Actors behaved differently from one another
 - Actors were small in number
 - Treatment effect could erroneously be ascribed if one white actor had exceptional negotiating skills

Bertrand and Mullainathan Audit Study: Advantages (cont.)

- Used wide variety of resumes (inexpensive way to expand the study)
- Used factorial design
 - 2x2 design (two races, two quality levels)
 - More discrimination for high-quality or low-quality resumes?

Next: Schneider Audit Study (2012)

- Does expectation of repeat business cause an auto mechanic to behave differently?

Audit Study Overview

- Study published by Schneider in 2012
- Study investigated effect of expected repeat business on behavior of auto mechanics
- Hypothesis: a firm is likely to work harder when expecting repeat business or referrals
- Schneider broke his own car in ways that should be repaired for safety during a road trip, then visited mechanics
 - Loose battery cable
 - Low coolant
 - Missing tail light
 - Etc...

Audit Study Details

- Schneider assigned 40 repair shops to one of two treatments:
 - Low reputation: "I'm moving from New Haven to Chicago in two weeks and want to make sure my car is in good condition before I go."
 - High reputation: "I'm taking a vacation to Montreal in two weeks."
 - Destinations are similar distances from origination city
- Schneider requested estimate for repairs

Audit Study Results

- Schneider was charged more than twice as much for estimate in low reputation treatment vs. high reputation treatment
- Schneider saw no significant difference in diagnosing problems between treatments

Bandiera *et al.* Experiment

- British farm employed eastern Europeans to pick fruit in Summer 2005
- Initial employee wages tied to *relative* performance
 - Farm wanted to implement productivity incentives
 - Farm manager reluctant to set piece rate (pay per unit picked)
 - Picking conditions vary considerably
 - Payments relative to quantities picked by others

Bandiera *et al.* Experiment (cont.)

- Researchers wanted to try *absolute* performance payments
 - Fixed amount per piece picked
 - Implementation half-way through Summer
 - Deliberate intervention designed to measure causal effects
 - Before/After design constraint related to concern regarding spill-over affects from employees talking to one another

Bandiera *et al.* Results

- Productivity increased substantially from first to second half of Summer
- Productivity per worker increased under *absolute* vs. *relative* payment
- Variance of output per worker much higher
- Proposed explanation:
 - High-ability workers were holding back under *relative* payment
 - Lived and socialized almost exclusively with co-workers
 - Had fear of co-workers' opinions
 - Worried that out-picking others lowered compensation for all
- Evidence that relative performance scheme can encourage people to work less

Evaluation of Bandiera *et al.*

- Study was not randomized
- Study included three comparisons:
 - Comparison of late part of Summer to early part
 - Sharp increase in productivity
 - Comparison to previous year, same time frame
 - *Absolute* payment outperformed *relative*, had higher variance
 - Comparison of pickers of strawberries to pickers of raspberries
 - Raspberries on tall bushes with briars limited visibility
 - Previous results held for pickers of strawberries but not raspberries
 - Raspberries functioned as control

Response to Incentives

Are people more likely to exercise if they are paid to go to the gym?

- Charness and Gneezy (2009) worked with UCSD student gymnasium
 - Introduced three conditions
 - Control
 - Low: \$25 for one visit in a week
 - High: \$100 for 8 additional visits
 - Collected data on gym attendance for 7 weeks post-experiment
 - Collected data on gym attendance pre-experiment

Charness and Gneezy: Results

- People in high-incentive condition developed exercise habit that extended beyond experiment
- Heterogeneous treatment effect
 - Highest effect in participants who did not previously exercise at gym
 - Low effect in participants who were already going to gym

Charness and Gneezy: Limitations

- Examination of carry-over effect to subsequent semesters was not possible due to inaccessibility of data
- Attempt to replicate similar results for the long term didn't work
 - Places limit on generalizability regarding habit formation

Additional Incentive Examples

"Can we pay people to. . ."

- Lose weight?
 - From behavioral economics perspective, what constitutes sufficient incentive?
- Quit smoking?
- Take blood pressure medication consistently?
 - From behavioral economics perspective, what constitutes sufficient incentive?
- Get better grades?
 - Fryer's attempt to pay students to get better scores was ineffective

Incentives With Unintended Consequences

- Gneezy and Rustichini (2000) experiment with Israeli daycare centers
 - Imposed fine for late pick-ups
 - Number of late pick-ups increased
 - Added 10 additional centers to experiment
 - Six centers had fines imposed for late pick-ups
 - Four functioned as controls
 - Number of late pick-ups *increased*

Incentives With Unintended Consequences (cont.)

- Behavioral economics explanation:
 - Prior to fine, parents were loath to impose on daycare workers
 - Fine was interpreted as a price
 - Fine was insufficient to deter unwanted behavior

Experimental Design

- Strive for originality and creativity
- Think more broadly
- Consider ideal experiment to address a research question
- Draw from other experiments
 - Audit studies
 - Factorial design
 - Double blindness
 - Variety in treatment to create more generalizability
 - Subpopulation overampling