SPIA 586a: Machine Learning for Policy Analysis
Week 6: Experiments

Brandon Stewart[1]

Princeton

March 3–5, 2025

[1]Huge thanks to Matt Blackwell, Jens Hainmueller and especially Matt Salganik for slides.

Where We've Been and Where We're Going...

Where We've Been and Where We're Going...

- Last Week
  - prediction in a counterfactual population
  - neural networks

Where We've Been and Where We're Going...

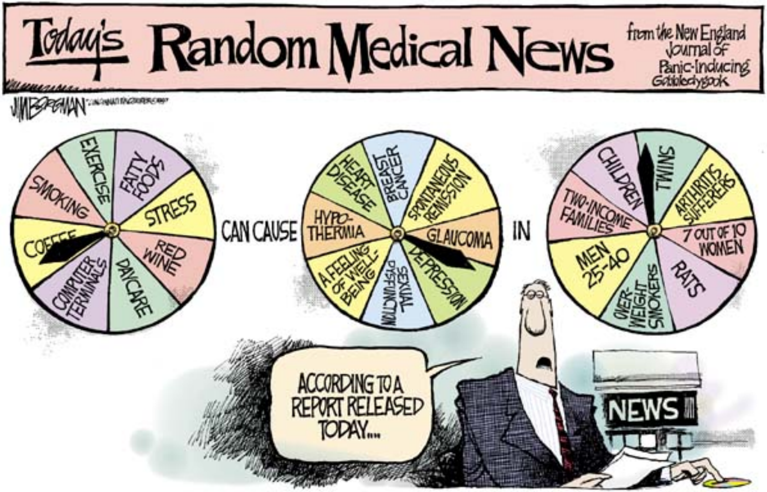
- Last Week
  - prediction in a counterfactual population
  - neural networks
- This Week
  - experiments
  - post-stratification

Where We've Been and Where We're Going...

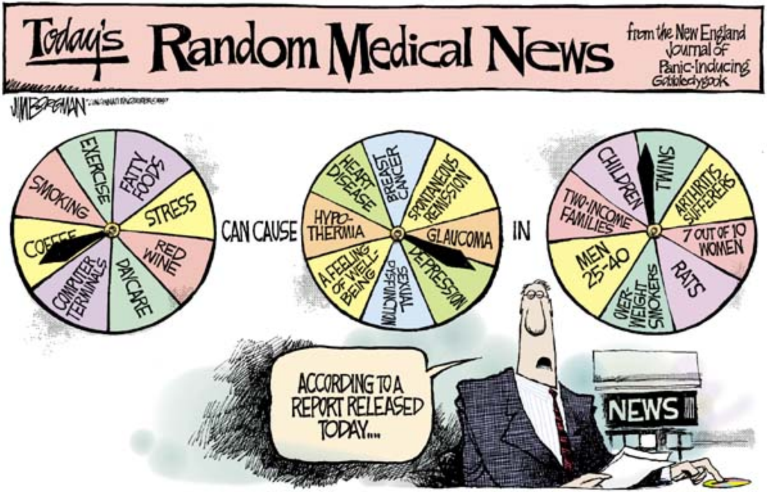
- Last Week
  - prediction in a counterfactual population
  - neural networks
- This Week
  - experiments
  - post-stratification
- Next Week
  - spring break
  - (then, designed data)

Where We've Been and Where We're Going...

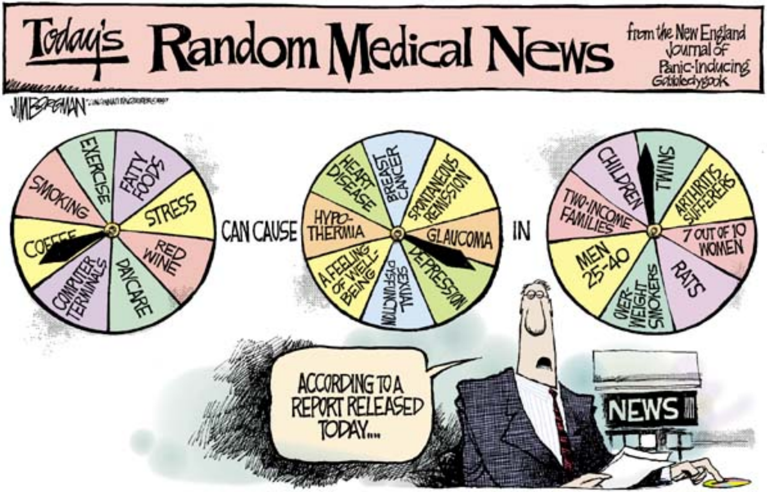
- Last Week
  - prediction in a counterfactual population
  - neural networks
- This Week
  - experiments
  - post-stratification
- Next Week
  - spring break
  - (then, designed data)
- Long Run
  - target tasks → data sources → guiding values



Lancet 2001: negative correlation between coronary heart disease mortality and level of vitamin C in bloodstream (controlling for age, gender, blood pressure, diabetes, and smoking)



Lancet 2002: no effect of vitamin C on mortality in controlled placebo trial (controlling for nothing)

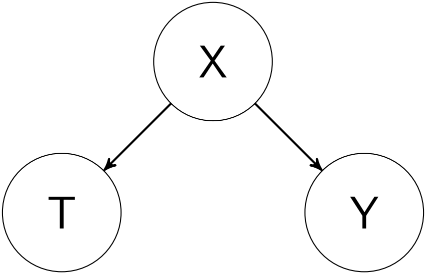


Lancet 2003: comparing among individuals with the same age, gender, blood pressure, diabetes, and smoking, those with higher vitamin C levels have lower levels of obesity, lower levels of alcohol consumption, are less likely to grow up in working class, etc.

Why So Much Variation?

Why So Much Variation?

Confounders



Observational Studies and Experimental Ideal

- Randomization forms gold standard for causal inference, because it balances observed and unobserved confounders

Observational Studies and Experimental Ideal

- Randomization forms gold standard for causal inference, because it balances observed and unobserved confounders
- Cannot always randomize so we do observational studies, where we adjust for the observed covariates and hope that unobservables are balanced

Observational Studies and Experimental Ideal

- Randomization forms gold standard for causal inference, because it balances observed and unobserved confounders
- Cannot always randomize so we do observational studies, where we adjust for the observed covariates and hope that unobservables are balanced
- Better than hoping: design observational study to approximate an experiment

Observational Studies and Experimental Ideal

- Randomization forms gold standard for causal inference, because it balances observed and unobserved confounders
- Cannot always randomize so we do observational studies, where we adjust for the observed covariates and hope that unobservables are balanced
- Better than hoping: design observational study to approximate an experiment
  - "The planner of an observational study should always ask himself [sic]: How would the study be conducted if it were possible to do it by controlled experimentation" (Cochran 1965)

Basic Experimental Logic

- An experiment is a study where assignment to treatment is controlled by the researcher.

Basic Experimental Logic

- An experiment is a study where assignment to treatment is controlled by the researcher.
  - $P(T_i = 1)$ is controlled and known by researcher.

Basic Experimental Logic

- An experiment is a study where assignment to treatment is controlled by the researcher.
  - $P(T_i = 1)$ is controlled and known by researcher.
- In the ideal randomized experiment, two assumptions hold by design:

Basic Experimental Logic

- An experiment is a study where assignment to treatment is controlled by the researcher.
  - $P(T_i = 1)$ is controlled and known by researcher.
- In the ideal randomized experiment, two assumptions hold by design:
  1. Positivity: assignment is probabilistic: $0 < P(T_i = 1) < 1$

Basic Experimental Logic

- An experiment is a study where assignment to treatment is controlled by the researcher.
  - $P(T_i = 1)$ is controlled and known by researcher.
- In the ideal randomized experiment, two assumptions hold by design:
  1. Positivity: assignment is probabilistic: $0 < P(T_i = 1) < 1$
     - No deterministic assignment.

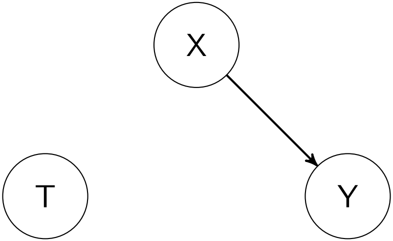
Basic Experimental Logic

- An experiment is a study where assignment to treatment is controlled by the researcher.
  - $P(T_i = 1)$ is controlled and known by researcher.
- In the ideal randomized experiment, two assumptions hold by design:
  1. Positivity: assignment is probabilistic: $0 < P(T_i = 1) < 1$
     - No deterministic assignment.
  2. Ignorability: $T_i \perp\!\!\!\perp (\mathbf{Y}(1), \mathbf{Y}(0))$

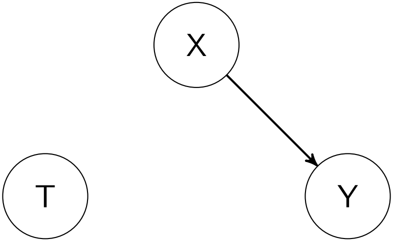
Basic Experimental Logic

- An experiment is a study where assignment to treatment is controlled by the researcher.
  - $P(T_i = 1)$ is controlled and known by researcher.
- In the ideal randomized experiment, two assumptions hold by design:
  1. Positivity: assignment is probabilistic: $0 < P(T_i = 1) < 1$
     - ⋆ No deterministic assignment.
  2. Ignorability: $T_i \perp\!\!\!\perp (\mathbf{Y}(1), \mathbf{Y}(0))$
     - ⋆ Treatment assignment does not depend on any potential outcomes.

Randomization Removes the Arrows into the Treatment

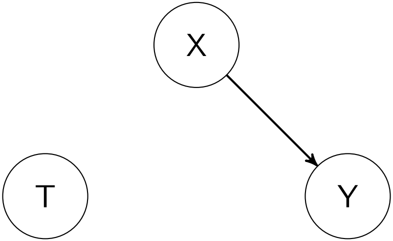


Randomization Removes the Arrows into the Treatment



This ensures any observed or unobserved pretreatment covariates have the same distribution in the treatment and the control group.

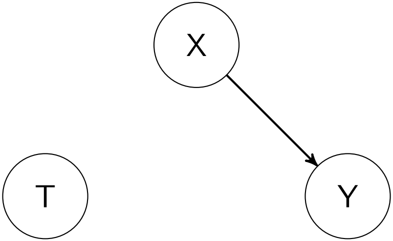
Randomization Removes the Arrows into the Treatment



This ensures any observed or unobserved pretreatment covariates have the same distribution in the treatment and the control group.

That is, they are balanced.

Randomization Removes the Arrows into the Treatment



This ensures any observed or unobserved pretreatment covariates have the same distribution in the treatment and the control group.

That is, they are balanced.

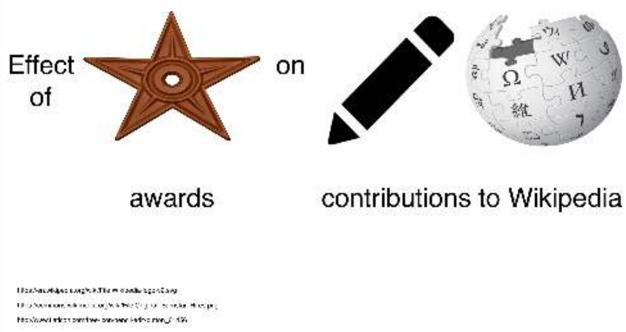
The benefit of experiments is that key assumptions hold by design and that you have balance along observed AND unobserved covariates.

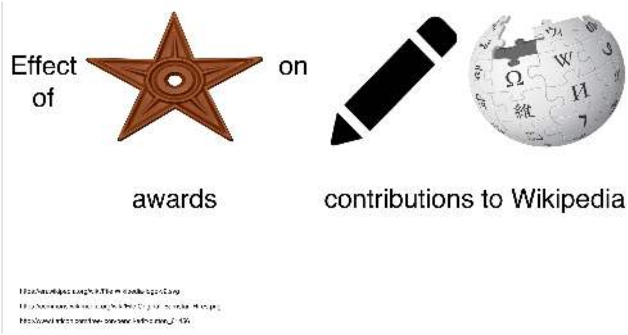
1) Introduction
2) Observing behavior
3) Asking questions
4) Running experiments
5) Mass collaboration
6) Ethics
7) The future

Experimental Study of Informal Rewards in Peer Production

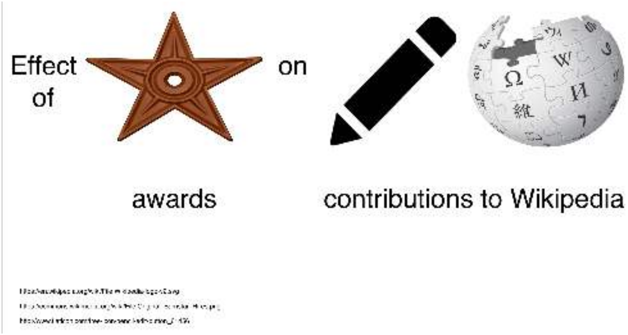
Michael Restivo*, Arnout van de Rijt

https://doi.org/10.1371/journal.pone.0034358

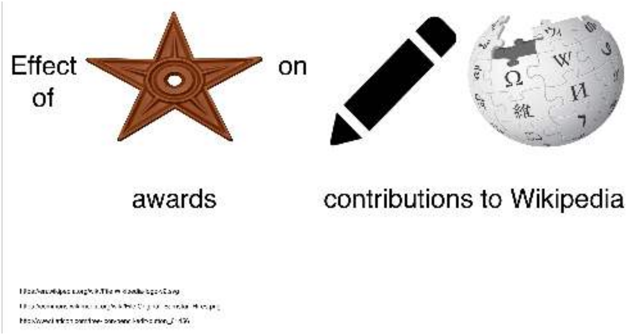

Effect of
on
awards
contributions to Wikipedia



- People who received barnstar edited less afterwards



- People who received barnstar edited less afterwards
- People in control group edited even less



- People who received barnstar edited less afterwards
- People in control group edited even less
- Not having a control group would lead to exactly the wrong conclusion (why did this happen?)

What's mostly the same?

- recruiting participants
- randomization treatment
- delivering treatment and control
- measuring outcomes

What's mostly the same?

- recruiting participants
- randomization treatment
- delivering treatment and control
- measuring outcomes

What's different?

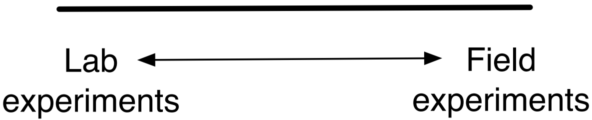
- Fully digital experiment leads to $\approx$ zero variable cost data.

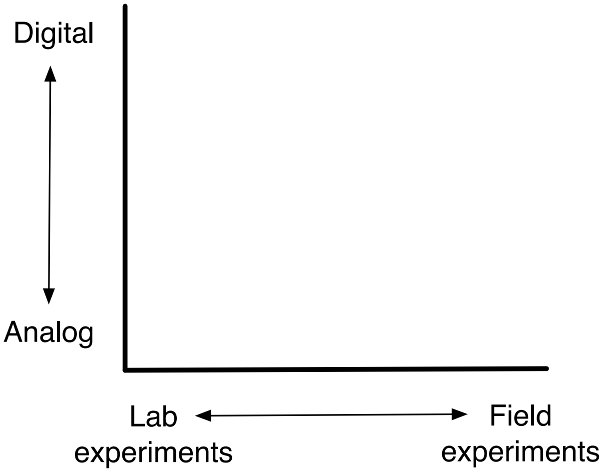
What's mostly the same?

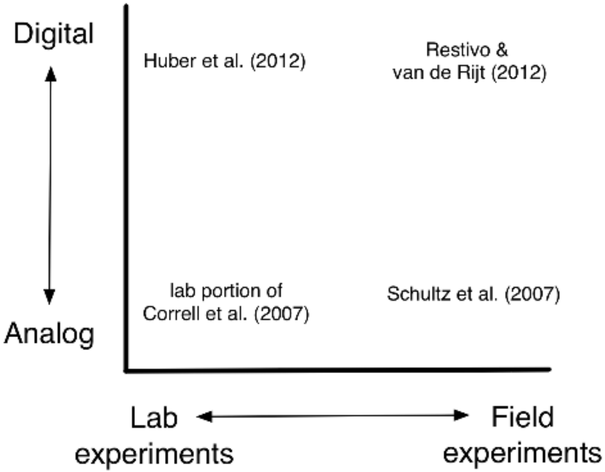
- recruiting participants
- randomization treatment
- delivering treatment and control
- measuring outcomes

What's different?

- Fully digital experiment leads to $\approx$ zero variable cost data.
- Constraint on size is not cost, it is ethics.


Lab
experiments
Field
experiments


Digital
Analog
Lab experiments
Field experiments


Digital
Analog
Huber et al. (2012)
Restivo &
van de Rijt (2012)
lab portion of
Correll et al. (2007)
Schultz et al. (2007)
Lab
experiments
Field
experiments

Challenging Ethics in the Age of Abundance

Challenging Ethics in the Age of Abundance

- Ethical frameworks for experiments were designed in social science's era of scarcity

Challenging Ethics in the Age of Abundance

- Ethical frameworks for experiments were designed in social science's era of scarcity
- We think about how to trade off moderate harms to a small number of people against potentially big benefits

Challenging Ethics in the Age of Abundance

- Ethical frameworks for experiments were designed in social science's era of scarcity
- We think about how to trade off moderate harms to a small number of people against potentially big benefits
  - key principles: respect for persons, beneficence, justice

Challenging Ethics in the Age of Abundance

- Ethical frameworks for experiments were designed in social science's era of scarcity
- We think about how to trade off moderate harms to a small number of people against potentially big benefits
  - key principles: respect for persons, beneficence, justice
- Now think about how to trade off minor harms to millions of people against potential benefits. What about minor risk of major harms?

Challenging Ethics in the Age of Abundance

- Ethical frameworks for experiments were designed in social science's era of scarcity
- We think about how to trade off moderate harms to a small number of people against potentially big benefits
  - key principles: respect for persons, beneficence, justice
- Now think about how to trade off minor harms to millions of people against potential benefits. What about minor risk of major harms?
- We will talk more about these aspects in future weeks!

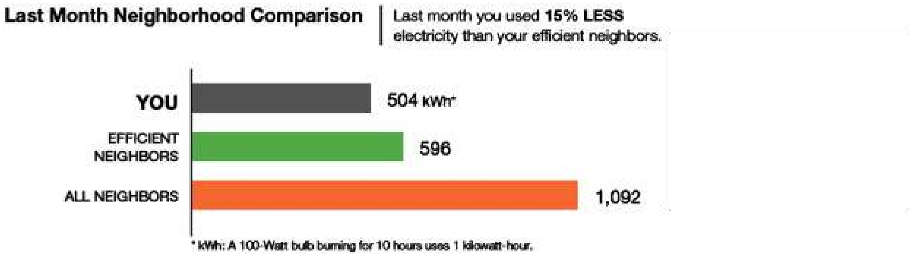
1. Experiments
   - Basic Experiments
   - Digital Experiments
   - Moving Beyond Simple Experiments
   - Roles for Machine Learning

2. Post-Stratification
   - Stratified Sampling
   - Poststratification

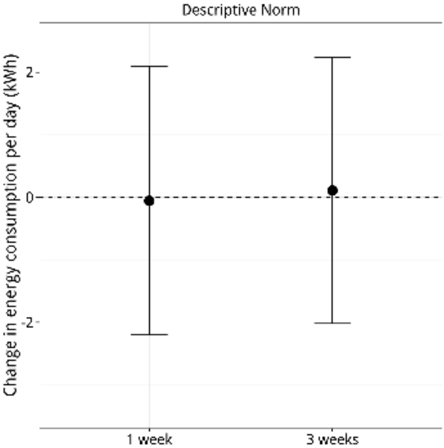
The Constructive, Destructive, and Reconstructive Power of Social Norms

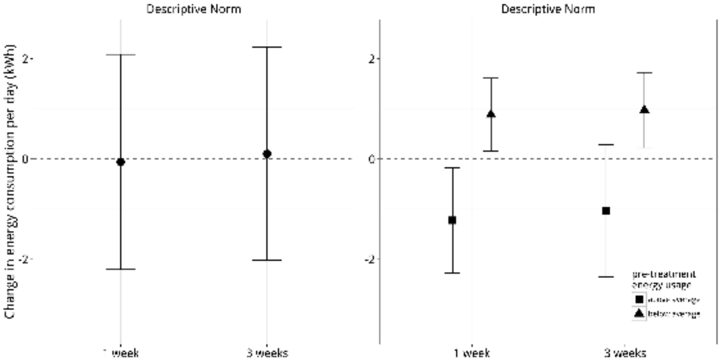
P. Wesley Schultz,[1] Jessica M. Nolan,[2] Robert B. Cialdini,[3] Noah J. Goldstein,[3] and Vladas Griskevicius[3]

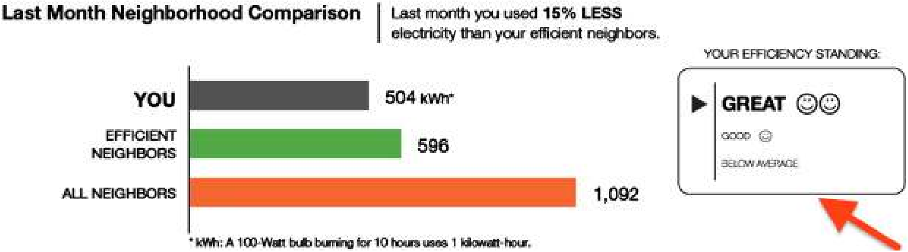
http://dx.doi.org/10.1111/j.1467-9280.2007.01917.x



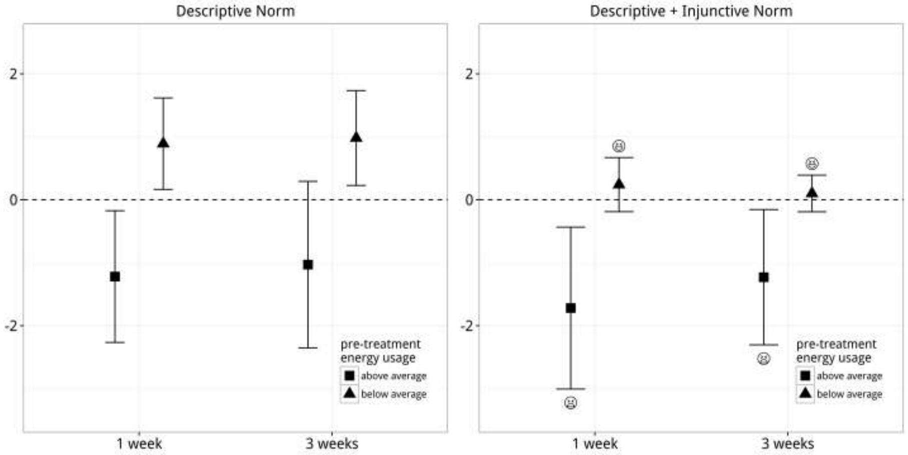
Figures from Allcott (2011)


Descriptive Norm
Change in energy consumption per day (kWh)
2
0
-2
1 week
3 weeks


Descriptive Norm
Descriptive Norm
Change in energy consumption per day (kWh)
2
0
-2
2
0
-2
1 week
3 weeks
1 week
3 weeks
pre-treatment
energy usage



Figures from Allcott (2011)


Descriptive Norm
Descriptive + Injunctive Norm
2
0
-2
1 week
3 weeks
pre-treatment
energy usage
above average
below average

If you want to move beyond simple experiments:

- Validity
- Heterogeneity of treatment effects
- Mechanisms

Validity

Validity

- statistical conclusion validity
- internal validity
- construct validity
- external validity

Heterogeneous treatment effects

Heterogeneous treatment effects

- lots of people, lots of pre-treatment information: stop treating people like widgets

Heterogeneous treatment effects

- lots of people, lots of pre-treatment information: stop treating people like widgets
- beware of fishing (preregistration, data splits)

Heterogeneous treatment effects

- lots of people, lots of pre-treatment information: stop treating people like widgets
- beware of fishing (preregistration, data splits)
- machine learning can help with this!

Mechanisms

Mechanisms

- Eating citrus prevents scurvy

Mechanisms

- Eating citrus prevents scurvy
- The mechanism through which citrus prevents scurvy vitamin C

Mechanisms

- Eating citrus prevents scurvy
- The mechanism through which citrus prevents scurvy vitamin C
- Not as easy as scurvy and limes, “Enough already about black box experiments” Green et al (2010)

Mechanisms

- Eating citrus prevents scurvy
- The mechanism through which citrus prevents scurvy vitamin C
- Not as easy as scurvy and limes, "Enough already about black box experiments" Green et al (2010)
- Experiments specifically designed to test mechanisms: Ludwig et al (2011), Imai et al (2012), Pirlott and MacKinnon (2016)

Where is the Machine Learning?

An advantage to experiments is simplicity, the estimator is often just a difference in means!

Where is the Machine Learning?

An advantage to experiments is simplicity, the estimator is often just a difference in means!

Does machine learning have any role to play?

Where is the Machine Learning?

An advantage to experiments is simplicity, the estimator is often just a difference in means!

Does machine learning have any role to play? Yes!

Roles for Machine Learning

Roles for Machine Learning

- Treatment effect heterogeneity $\rightsquigarrow$ personalized medicine settings

Roles for Machine Learning

- Treatment effect heterogeneity $\rightsquigarrow$ personalized medicine settings
- Dynamic treatment regimes $\rightsquigarrow$ understanding sequences

Roles for Machine Learning

- Treatment effect heterogeneity $\rightsquigarrow$ personalized medicine settings
- Dynamic treatment regimes $\rightsquigarrow$ understanding sequences
- Discovering treatments $\rightsquigarrow$ searching/learning treatments

Roles for Machine Learning

- Treatment effect heterogeneity $\rightsquigarrow$ personalized medicine settings
- Dynamic treatment regimes $\rightsquigarrow$ understanding sequences
- Discovering treatments $\rightsquigarrow$ searching/learning treatments
- Improving efficiency $\rightsquigarrow$ assigning participants optimally

Why Sample?

Why Sample?

- Infeasible or costly to measure every unit

Why Sample?

- Infeasible or costly to measure every unit
- Ex:
    - Surveys
    - Large datasets with complex computation
    - Urgent need for data

Why Sample?

- Infeasible or costly to measure every unit
- Ex:
  - Surveys
  - Large datasets with complex computation
  - Urgent need for data
- Sampling (and weighting) are important for drawing inferences about a population estimand.

Why Sample?

- Infeasible or costly to measure every unit
- Ex:
  - Surveys
  - Large datasets with complex computation
  - Urgent need for data
- Sampling (and weighting) are important for drawing inferences about a population estimand.
- Distinguised from experiments since we often do not perturb anything to identify estimand
  - i.e. we care about descriptive inference

Basic Idea of Sampling

Basic Idea of Sampling

- Start with a population that consists of a finite number of units

Basic Idea of Sampling

- Start with a population that consists of a finite number of units
  - With an associated variable of interest (our Y)
  - The sequence of $y_1, y_2, \ldots, y_N$ in the population is fixed
  - Estimand of interest is usually some summary characteristic of the population

Basic Idea of Sampling

- Start with a population that consists of a finite number of units
  - With an associated variable of interest (our Y)
  - The sequence of $y_1, y_2, \ldots, y_N$ in the population is fixed
  - Estimand of interest is usually some summary characteristic of the population
- Observe the outcome for a selected sample

Basic Idea of Sampling

- Start with a population that consists of a finite number of units
  - With an associated variable of interest (our Y)
  - The sequence of $y_1, y_2, \ldots, y_N$ in the population is fixed
  - Estimand of interest is usually some summary characteristic of the population
- Observe the outcome for a selected sample
  - The sampling design is the procedure by which the sample is selected

Basic Idea of Sampling

- Start with a population that consists of a finite number of units
  - With an associated variable of interest (our Y)
  - The sequence of $y_1, y_2, \ldots, y_N$ in the population is fixed
  - Estimand of interest is usually some summary characteristic of the population
- Observe the outcome for a selected sample
  - The sampling design is the procedure by which the sample is selected
  - We must pay careful attention to the sampling units

Concrete Example

Concrete Example

- Imagine if we want to estimate the average height of all students in a high school.

Concrete Example

- Imagine if we want to estimate the average height of all students in a high school.
- The school has 4 grades with the following sizes:
  - 9th Grade: 400 students
  - 10th Grade: 300 students
  - 11th Grade: 200 students
  - 12th Grade: 100 students

Concrete Example

- Imagine if we want to estimate the average height of all students in a high school.
- The school has 4 grades with the following sizes:
  - 9th Grade: 400 students
  - 10th Grade: 300 students
  - 11th Grade: 200 students
  - 12th Grade: 100 students
- We can sample 20 students at random from the full population and use their heights to estimate the average height.

Concrete Example

- Imagine if we want to estimate the average height of all students in a high school.
- The school has 4 grades with the following sizes:
  - 9th Grade: 400 students
  - 10th Grade: 300 students
  - 11th Grade: 200 students
  - 12th Grade: 100 students
- We can sample 20 students at random from the full population and use their heights to estimate the average height.
- We could easily miss some grades entirely, is there a better way?

Concrete Example

- Imagine if we want to estimate the average height of all students in a high school.
- The school has 4 grades with the following sizes:
  - 9th Grade: 400 students
  - 10th Grade: 300 students
  - 11th Grade: 200 students
  - 12th Grade: 100 students
- We can sample 20 students at random from the full population and use their heights to estimate the average height.
- We could easily miss some grades entirely, is there a better way?
- What if we sampled students by grade?

Concrete Example: Mechanically

- 9th Grade: 400 students
- 10th Grade: 300 students
- 11th Grade: 200 students
- 12th Grade: 100 students

Concrete Example: Mechanically

- 9th Grade: 400 students
- 10th Grade: 300 students
- 11th Grade: 200 students
- 12th Grade: 100 students

Sample 8 students from 9th grade, 6 students from 10th grade, 4 students from 11th grade, 2 student from 12th grade.

Concrete Example: Mechanically

- 9th Grade: 400 students
- 10th Grade: 300 students
- 11th Grade: 200 students
- 12th Grade: 100 students

Sample 8 students from 9th grade, 6 students from 10th grade, 4 students from 11th grade, 2 student from 12th grade.

To estimate just average the results!

Concrete Example: Mechanically

- 9th Grade: 400 students
- 10th Grade: 300 students
- 11th Grade: 200 students
- 12th Grade: 100 students

Concrete Example: Mechanically

- 9th Grade: 400 students
- 10th Grade: 300 students
- 11th Grade: 200 students
- 12th Grade: 100 students

Sample 5 students per grade.

Concrete Example: Mechanically

- 9th Grade: 400 students
- 10th Grade: 300 students
- 11th Grade: 200 students
- 12th Grade: 100 students

Sample 5 students per grade.

$$\hat{\mu} = \frac{400}{1000}\bar{Y}_9 + \frac{300}{1000}\bar{Y}_{10} + \frac{200}{1000}\bar{Y}_{11} + \frac{100}{1000}\bar{Y}_{12}$$

Stratified Sampling

Stratified Sampling

- Divide the population in to H groups in indexed by $h = 1, 2, \ldots, H$ called strata

Stratified Sampling

- Divide the population in to H groups in indexed by $h = 1, 2, \ldots, H$ called strata
  - Done using auxiliary information about the target population

Stratified Sampling

- Divide the population in to H groups in indexed by $h = 1, 2, \ldots, H$ called strata
  - Done using auxiliary information about the target population
  - hth strata contains N_h elements and each element is in one and only one strata

Stratified Sampling

- Divide the population in to H groups in indexed by $h = 1, 2, \ldots, H$ called strata
  - Done using auxiliary information about the target population
  - hth strata contains N_h elements and each element is in one and only one strata
- Simple random sample within the strata.

Stratified Sampling

- Divide the population in to H groups in indexed by $h = 1, 2, \dots, H$ called strata
  - Done using auxiliary information about the target population
  - hth strata contains N_h elements and each element is in one and only one strata
- Simple random sample within the strata.
- Estimate the mean with $\hat{\mu} = \sum_{h=1}^{H} \frac{N_h}{N} \bar{y}_H$

Stratified Sampling

- Divide the population in to H groups in indexed by $h = 1, 2, \ldots, H$ called strata
  - Done using auxiliary information about the target population
  - hth strata contains N_h elements and each element is in one and only one strata
- Simple random sample within the strata.
- Estimate the mean with $\hat{\mu} = \sum_{h=1}^{H} \frac{N_h}{N} \bar{y}_H$
- Principle of stratification–the goal is to partition the population in such a way that the units within a stratum are as similar as possible. This will decrease the variance of the samples.

Stratified Sampling

- Divide the population in to H groups in indexed by $h = 1, 2, \ldots, H$ called strata
  - Done using auxiliary information about the target population
  - hth strata contains N_h elements and each element is in one and only one strata
- Simple random sample within the strata.
- Estimate the mean with $\hat{\mu} = \sum_{h=1}^{H} \frac{N_h}{N} \bar{y}_H$
- Principle of stratification–the goal is to partition the population in such a way that the units within a stratum are as similar as possible. This will decrease the variance of the samples.
- How many units per strata?
  proportional allocation vs. optimal allocation.

Why Stratify?

Why Stratify?

- We believe the variable we care about takes on different mean values in different subpopulations

Why Stratify?

- We believe the variable we care about takes on different mean values in different subpopulations
- We want to increase the precision of our estimate

Why Stratify?

- We believe the variable we care about takes on different mean values in different subpopulations
- We want to increase the precision of our estimate
- We want to protect against the possibility of obtaining a “really bad sample”

Why Stratify?

- We believe the variable we care about takes on different mean values in different subpopulations
- We want to increase the precision of our estimate
- We want to protect against the possibility of obtaining a "really bad sample"
- We want to ensure some level of precision for subgroups of the population

Why Stratify?

- We believe the variable we care about takes on different mean values in different subpopulations
- We want to increase the precision of our estimate
- We want to protect against the possibility of obtaining a "really bad sample"
- We want to ensure some level of precision for subgroups of the population
- May be more convenient to administer and may result in a lower cost

Defining Strata

Defining Strata

- Stratification is most effective when the strata means are very different.

Defining Strata

- Stratification is most effective when the strata means are very different.
- Allocating the units optimally across strata requires knowledge of the within strata variances as well as the costs of sampling.

Defining Strata

- Stratification is most effective when the strata means are very different.
- Allocating the units optimally across strata requires knowledge of the within strata variances as well as the costs of sampling.
- Requires prior knowledge which is hard to have exactly but also comes implicitly with a good proxy for the outcome.

Defining Strata

- Stratification is most effective when the strata means are very different.
- Allocating the units optimally across strata requires knowledge of the within strata variances as well as the costs of sampling.
- Requires prior knowledge which is hard to have exactly but also comes implicitly with a good proxy for the outcome.

Rule of Thumb:
the more information you have, the more likely you should stratify

1. Experiments
   - Basic Experiments
   - Digital Experiments
   - Moving Beyond Simple Experiments
   - Roles for Machine Learning

2. Post-Stratification
   - Stratified Sampling
   - Poststratification

Poststratification: The Core Idea

Poststratification: The Core Idea

Reweight your sample to look like your population.

Poststratification: The Core Idea

Reweight your sample to look like your population.

Analyze your sample as though you had done stratified sampling (even though you didn't!).

Forecasting elections with non-representative polls

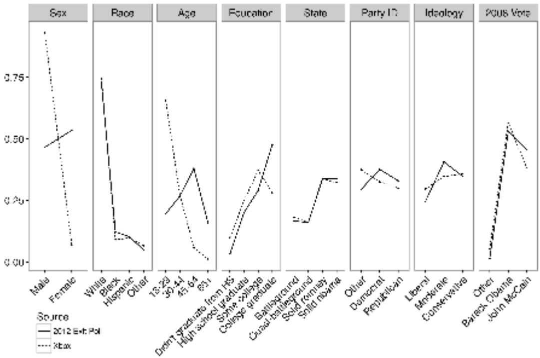
Wei Wang [a,*], David Rothschild [b], Sharad Goel [b], Andrew Gelman [a,c]

[a] Department of Statistics, Columbia University, New York, NY, USA

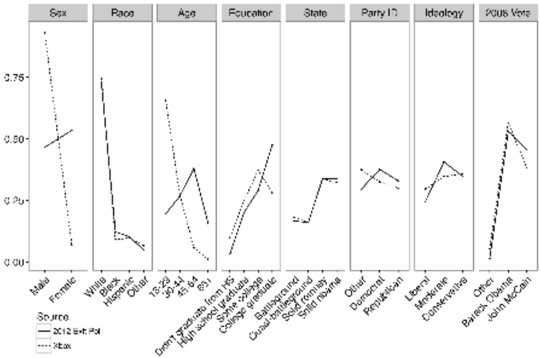
[b] Microsoft Research, New York, NY, USA

[c] Department of Political Science, Columbia University, New York, NY, USA

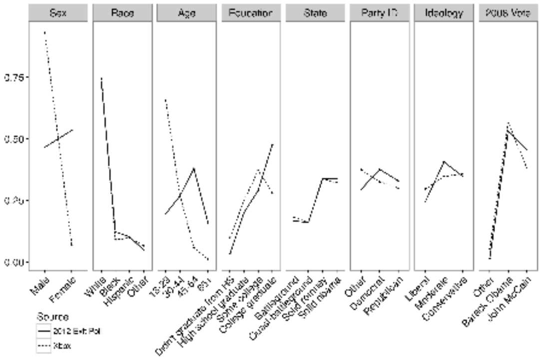
Wang et al (2015)



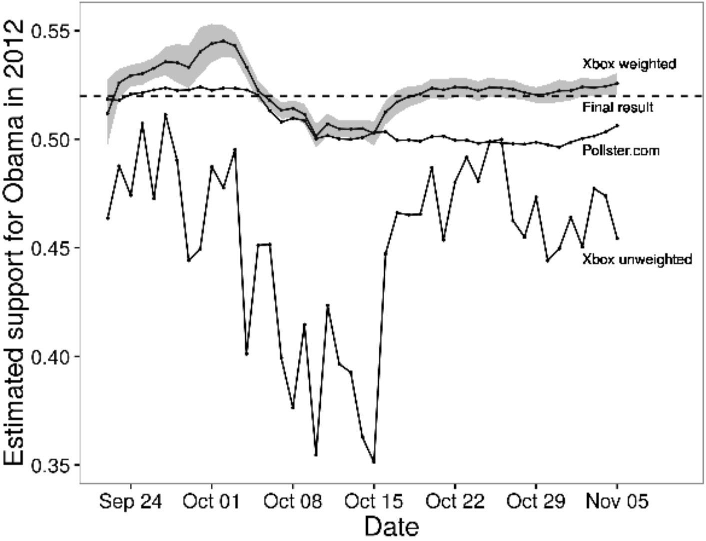
- about 750,000 interviews
- about 350,000 unique respondents



- about 750,000 interviews
- about 350,000 unique respondents
- Problem: 18-29 year olds make up 19% of the electorate but 65% of Xbox users.



- about 750,000 interviews
- about 350,000 unique respondents
- Problem: 18-29 year olds make up 19% of the electorate but 65% of Xbox users.
- Core idea: within these subgroups, pretend that it is a random sample from the population
  (would this work with technology literacy as an outcome?)


Estimated support for Obama in 2012
0.55
0.50
0.45
0.40
0.35
Xbox weighted
Final result
Pollster.com
Xbox unweighted
Sep 24
Oct 01
Oct 08
Oct 15
Oct 22
Oct 29
Nov 05
Date

Returning to Our Height Example

Returning to Our Height Example

- Imagine now that we had just been standing in the cafeteria and measuring the height of anyone who walked by.

Returning to Our Height Example

- Imagine now that we had just been standing in the cafeteria and measuring the height of anyone who walked by.
- Problem: not a random sample.

Returning to Our Height Example

- Imagine now that we had just been standing in the cafeteria and measuring the height of anyone who walked by.
- Problem: not a random sample.
- If we are willing to assume that among people within strata it is a random sample (or at least that the sense in which it is not is uncorrelated with height), we can reweight as though we had chosen to stratify.

Returning to Our Height Example

- Imagine now that we had just been standing in the cafeteria and measuring the height of anyone who walked by.
- Problem: not a random sample.
- If we are willing to assume that among people within strata it is a random sample (or at least that the sense in which it is not is uncorrelated with height), we can reweight as though we had chosen to stratify.
- Key assumption:homogenous response probabilities within strata.

Returning to Our Height Example

- Imagine now that we had just been standing in the cafeteria and measuring the height of anyone who walked by.
- Problem: not a random sample.
- If we are willing to assume that among people within strata it is a random sample (or at least that the sense in which it is not is uncorrelated with height), we can reweight as though we had chosen to stratify.
- Key assumption:homogenous response probabilities within strata.
- This is analogous to the ideas in causal inference!

Uses of Poststratification

Uses of Poststratification

- Collect data cheaper.

Uses of Poststratification

- Collect data cheaper.
- Correct for non-response.

Uses of Poststratification

- Collect data cheaper.
- Correct for non-response.
- Produce estimates for finer-grained subsets of the data

Uses of Poststratification

- Collect data cheaper.
- Correct for non-response.
- Produce estimates for finer-grained subsets of the data

Challenges: strong assumptions and difficult estimation with many cells

Multilevel Regression with Poststratification

Multilevel Regression with Poststratification

- Core challenge is that the cells in the survey might be very sparsely populated: e.g. how many Hispanic women above 65 with a PhD live in Alaska?

Multilevel Regression with Poststratification

- Core challenge is that the cells in the survey might be very sparsely populated: e.g. how many Hispanic women above 65 with a PhD live in Alaska?
- Gelman and coauthors introduce Multilevel Regression with Poststratification. Often called “MRP” or “Mister P”

Multilevel Regression with Poststratification

- Core challenge is that the cells in the survey might be very sparsely populated: e.g. how many Hispanic women above 65 with a PhD live in Alaska?
- Gelman and coauthors introduce Multilevel Regression with Poststratification. Often called "MRP" or "Mister P"
- This strategy allows you to smartly borrow information from related cells when trying to estimate the mean in the cell of interest.

Multilevel Regression with Poststratification

- Core challenge is that the cells in the survey might be very sparsely populated: e.g. how many Hispanic women above 65 with a PhD live in Alaska?
- Gelman and coauthors introduce Multilevel Regression with Poststratification. Often called "MRP" or "Mister P"
- This strategy allows you to smartly borrow information from related cells when trying to estimate the mean in the cell of interest.
- That is, it is a prediction problem!

Multilevel Regression with Poststratification

- Core challenge is that the cells in the survey might be very sparsely populated: e.g. how many Hispanic women above 65 with a PhD live in Alaska?
- Gelman and coauthors introduce Multilevel Regression with Poststratification. Often called "MRP" or "Mister P"
- This strategy allows you to smartly borrow information from related cells when trying to estimate the mean in the cell of interest.
- That is, it is a prediction problem!
- Historically such models have been additive and simple, but we can use machine learning.

Multilevel Regression with Poststratification

- Core challenge is that the cells in the survey might be very sparsely populated: e.g. how many Hispanic women above 65 with a PhD live in Alaska?
- Gelman and coauthors introduce Multilevel Regression with Poststratification. Often called “MRP” or “Mister P”
- This strategy allows you to smartly borrow information from related cells when trying to estimate the mean in the cell of interest.
- That is, it is a prediction problem!
- Historically such models have been additive and simple, but we can use machine learning.
- Once we have something that can predict the conditional expectations for us, we can just reweight like we did in our high school example.

Connecting Back to Tasks

Connecting Back to Tasks

- Poststratification is a quite general technique which can help us address those problems we faced earlier like how do we make inference to a different population?

Connecting Back to Tasks

- Poststratification is a quite general technique which can help us address those problems we faced earlier like how do we make inference to a different population?
- Insight for us: surveys with non-response or non-probabilistic surveys are always inference to a different population.

Connecting Back to Tasks

- Poststratification is a quite general technique which can help us address those problems we faced earlier like how do we make inference to a different population?
- Insight for us: surveys with non-response or non-probabilistic surveys are always inference to a different population.
- As with many of our cases: key to use machine learning was reducing the problem to a simple prediction task.

Summary

Summary

- stratified sampling
- poststratification
- machine learning with poststratification

Summary

- stratified sampling
- poststratification
- machine learning with poststratification

Going Deeper:

Park, Gelman and Bafumi (2004) "Bayesian Multilevel Estimation with Poststratification: State-Level Estimates from National Polls" *Political Analysis*.

This Week in Review

This Week in Review

- Why experiments work
- Digital experiments
- Moving beyond simple experiments
- Roles for machine learning

This Week in Review

- Why experiments work
- Digital experiments
- Moving beyond simple experiments
- Roles for machine learning
- Great additional material in *Bit by Bit*:
    - making it happen
    - implications of zero variable cost data
    - improving ethics

This Week in Review

- Why experiments work
- Digital experiments
- Moving beyond simple experiments
- Roles for machine learning
- Great additional material in *Bit by Bit*:
  - making it happen
  - implications of zero variable cost data
  - improving ethics

Next Class Week

- Salganik Chapter 3: "Asking Questions"
- Bradley, V.C., Kuriwaki, S., Isakov, M. et al. 2021. "Unrepresentative big surveys significantly overestimated US vaccine uptake." *Nature* 600, 695–700.