Behavioral and Experimental Economics
Econometrics Primer

Professor Jonathan M.V. Davis

This class is about evidence

EC311 was about microeconomic **theory**

This class is about **testing** whether that theory reflects the real world

Crucial that we know how to assess the quality of **empirical evidence**

How to test theory (Friedman, 1966)

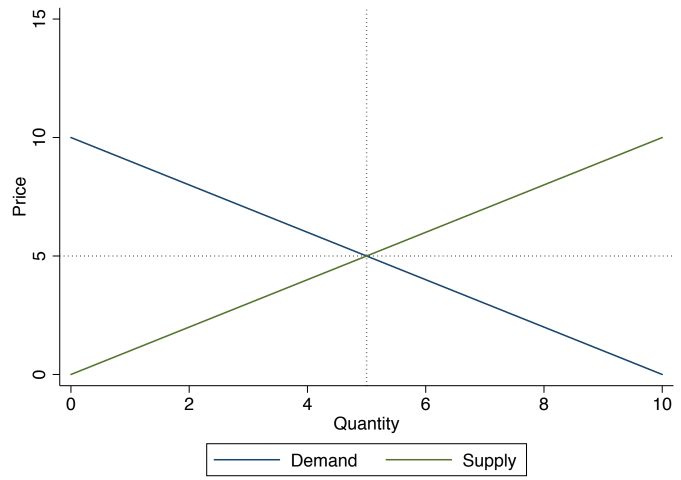
"The only relevant test of the *validity* of a hypothesis is comparison of its predictions with experience."

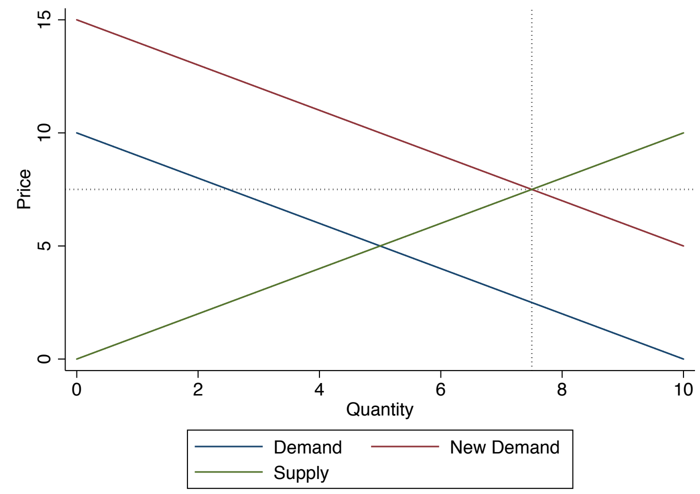
"The hypothesis is rejected if its predictions are contradicted; it is accepted if its predictions are not contradicted"

"Factual evidence can never 'prove' a hypothesis; it can only fail to disprove it, which is what we generally mean when we say, somewhat inexactly, that the hypothesis has been 'confirmed' by experience."

How to test theory

1. Need predictions from the theory, e.g. Supply and Demand





Prediction: Increase in Demand, holding supply constant -> Higher Price and Higher Quantity

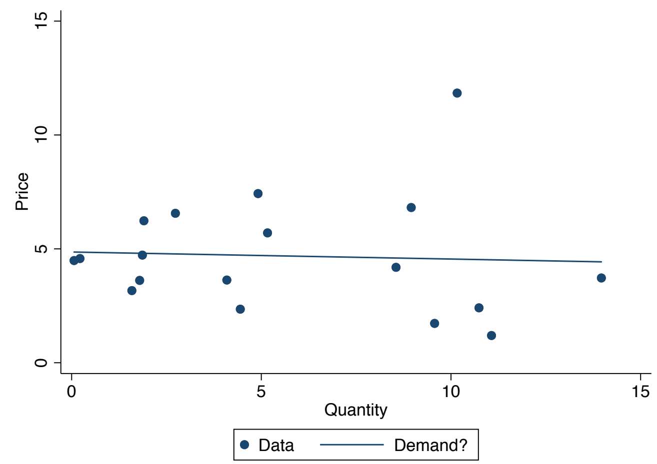
How to test theory

1. Need predictions from the theory
   - Usually: Change in X -> Some Change in Y

Topic	X	Y
Supply and Demand	Increase in Demand	Increase in Price and Quantity
Salience	Sales Tax Included or Not	No change in quantity
Endowment Effect	Loss or Gain Framing	No change in choice
Social Preferences	Donations made public	No change in donations

How to test theory

1. Need predictions from the theory
    a. Usually: Change in X -> Some Change in Y
2. Need empirical evidence to test prediction

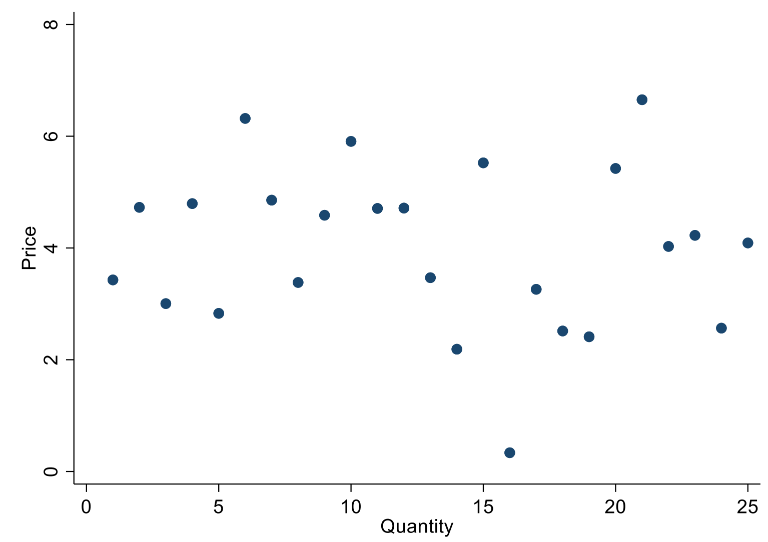


How to describe relationship between X and Y

1. Scatterplot: Graph of Y versus X

Advantage: Shows all data

Disadvantage: Difficult to summarize



How to describe relationship between X and Y

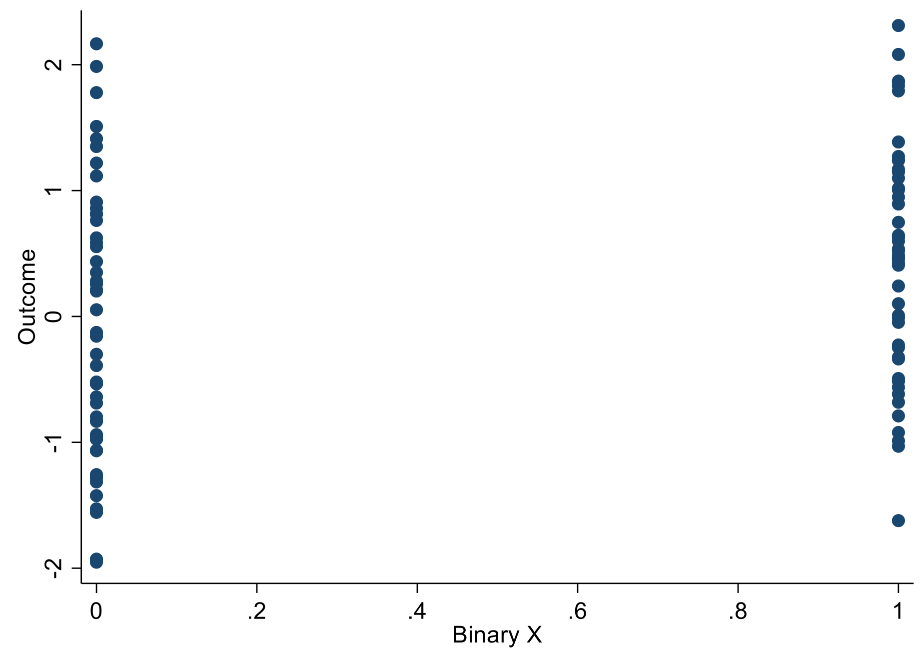
1. Scatterplot: Graph of Y versus X

Advantage: Shows all data

Disadvantage: Difficult to summarize

Especially when X is “binary” or a 0/1 variable

Most experiments are treatment (1) v. Control (0) comparisons!



How to describe relationship between X and Y

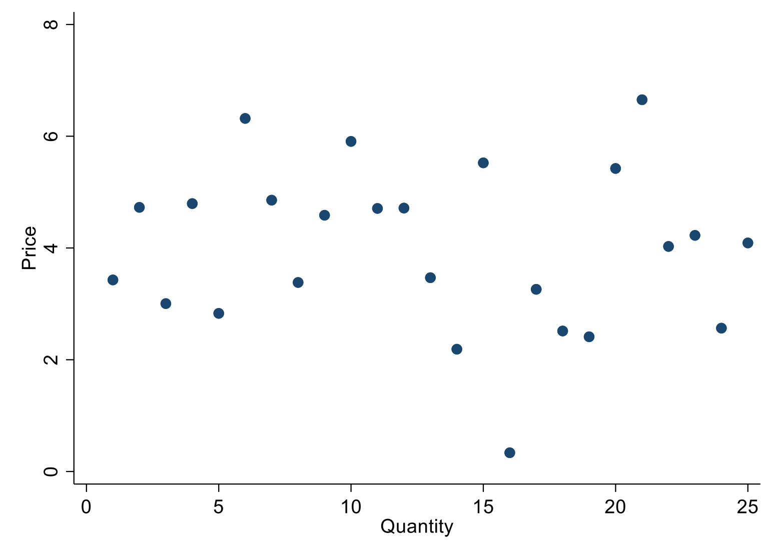
1. Scatterplot: A graph of Y versus X

2. Correlation

$$\rho_{xy} = \frac{Cov(X,Y)}{SD(X)SD(Y)}$$

Between -1 and 1

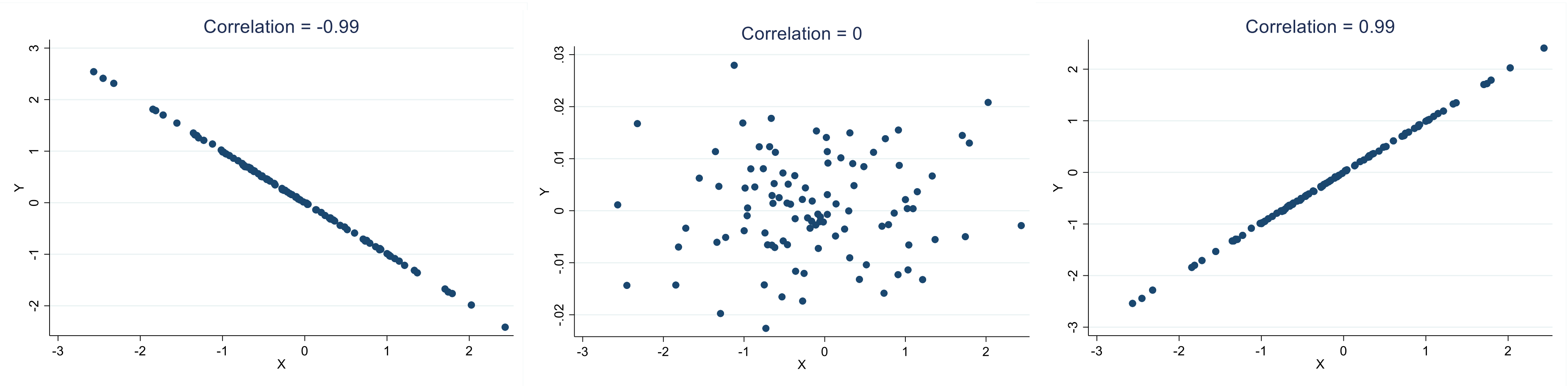
In this case, $\rho_{xy} = -0.11$



How to describe relationship between X and Y

1. Scatterplot: A graph of Y versus X

2. Correlation $\rho_{xy} = \frac{Cov(X,Y)}{SD(X)SD(Y)}$



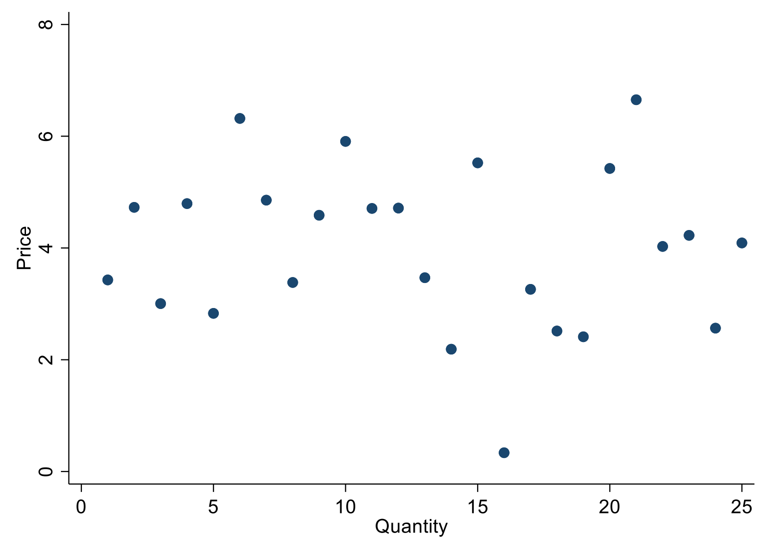
How to describe relationship between X and Y

1. Scatterplot: A graph of Y versus X

2. Correlation

$$\rho_{xy} = \frac{Cov(X,Y)}{SD(X)SD(Y)}$$

Advantage: Summarizes relationship
Disadvantage: Not a very intuitive scale



How to describe relationship between X and Y

1. Scatterplot: A graph of Y versus X

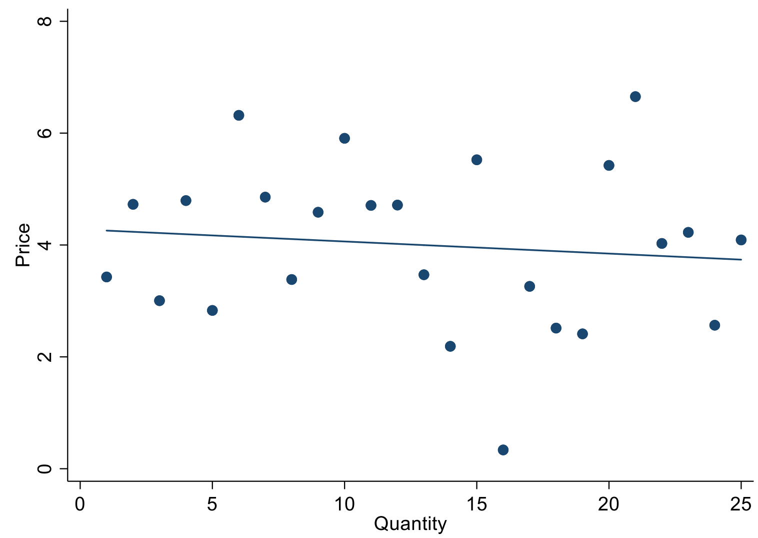
2. Correlation

3. Line of best fit (regression)

$$E[Y|X = x] = \alpha + \beta x$$

β describes how much Y changes, on average, when X increases by 1.

Here, β is -.55.



How to describe relationship between X and Y

1. Scatterplot: A graph of Y versus X

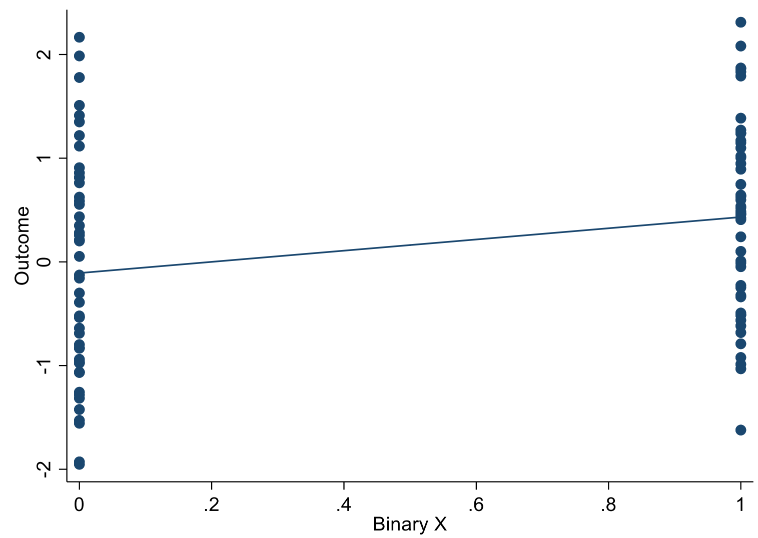
2. Correlation

3. Line of best fit (regression)

$$E[Y|X = x] = \alpha + \beta x$$

β describes how much Y changes, on average, when X increases by 1.

In this case, β is 0.54.



Interpreting Regression Coefficients

$$E[Y|X = x] = \alpha + \beta x$$

Most evidence we see in this class will be **regression estimates**

Usually care about β, not so much about α. Why?

Interpreting Regression Coefficients

$$E[Y|X = x] = \alpha + \beta x$$

Most evidence we see in this class will be **regression estimates**

Usually care about β, not so much about α. Why?

Predictions we test are about **changes** and β is a slope term $(\frac{\Delta Y}{\Delta X})$

Interpreting Regression Coefficients

$$E[Y|X = x] = \alpha + \beta x$$

β tells us by how much Y usually changes when X increases by 1

Does that mean Y will increase by β if we increase x by 1?

Interpreting Regression Coefficients

$$E[Y|X = x] = \alpha + \beta x$$

β tells us by how much Y usually changes when X increases by 1

Does that mean Y will increase by β if we increase x by 1?

Let’s think about this using a real example.

Example: Returns to Education (Card 1999)

$$E[Y|X = x] = \alpha + \beta x$$

- Y is log hourly earnings
- X is years of education
- α is not shown
- β is 0.109
- Standard error is 0.001
  (more on this later)

Table 1
Estimated education coefficients from standard human capital earnings function fit to hourly wages, annual earnings, and various measures of hours for men and women in March 1994–1996 Current Population Survey[a]

	Dependent variable				
	Log hourly earnings (1)	Log hours per week (2)	Log weeks per year (3)	Log annual hours (4)	Log annual earnings (5)
A. Men					
Education coefficient	0.100 (0.001)	0.018 (0.001)	0.025 (0.001)	0.042 (0.001)	0.142 (0.001)
R-squared	0.328	0.182	0.136	0.222	0.403
B. Women					
Education coefficient	$\boxed{0.109 \; (0.001)}$	0.022 (0.001)	0.034 (0.001)	0.056 (0.001)	0.165 (0.001)
R-squared	$\boxed{0.247}$	0.071	0.074	0.105	0.247

[a] Notes: Table reports estimated coefficient of linear education term in model that also includes cubic in potential experience and an indicator for non-white race. Samples include men and women age 16–66 who report positive wage and salary earnings in the previous year. Hourly wage is constructed by dividing wage and salary earnings by the product of weeks worked and usual hours per week. Data for individuals whose wage is under \$2.00 or over \$150.00 (in 1995 dollars) are dropped. Sample sizes are: 102,639 men and 95,309 women.

Example: Returns to Education (Card 1999)

Return to college for women

$$\log(w_{16}) - \log(w_{12}) = 4 * 0.109$$
$$\log(w_{16}) - \log(w_{12}) = 0.436$$

Table 1
Estimated education coefficients from standard human capital earnings function fit to hourly wages, annual earnings, and various measures of hours for men and women in March 1994–1996 Current Population Survey[a]

	Dependent variable				
	Log hourly earnings (1)	Log hours per week (2)	Log weeks per year (3)	Log annual hours (4)	Log annual earnings (5)
A. Men					
Education coefficient	0.100 (0.001)	0.018 (0.001)	0.025 (0.001)	0.042 (0.001)	0.142 (0.001)
R-squared	0.328	0.182	0.136	0.222	0.403
B. Women					
Education coefficient	0.109 (0.001)	0.022 (0.001)	0.034 (0.001)	0.056 (0.001)	0.165 (0.001)
R-squared	0.247	0.071	0.074	0.105	0.247

[a] Notes: Table reports estimated coefficient of linear education term in model that also includes cubic in potential experience and an indicator for non-white race. Samples include men and women age 16–66 who report positive wage and salary earnings in the previous year. Hourly wage is constructed by dividing wage and salary earnings by the product of weeks worked and usual hours per week. Data for individuals whose wage is under $2.00 or over $150.00 (in 1995 dollars) are dropped. Sample sizes are: 102,639 men and 95,309 women.

Example: Returns to Education (Card 1999)

Why logs?

$$\log(a) - \log(b) = \log(\frac{a}{b})$$

So:

$$\log\left(\frac{w_{16}}{w_{12}}\right) = 0.436$$

Or:

$$\frac{w_{16}}{w_{12}} = \exp(0.436) = 1.55$$

55% higher hourly earnings!

Table 1
Estimated education coefficients from standard human capital earnings function fit to hourly wages, annual earnings, and various measures of hours for men and women in March 1994–1996 Current Population Survey[a]

	Dependent variable				
	Log hourly earnings (1)	Log hours per week (2)	Log weeks per year (3)	Log annual hours (4)	Log annual earnings (5)
A. Men					
Education coefficient	0.100 (0.001)	0.018 (0.001)	0.025 (0.001)	0.042 (0.001)	0.142 (0.001)
R-squared	0.328	0.182	0.136	0.222	0.403
B. Women					
Education coefficient	0.109 (0.001)	0.022 (0.001)	0.034 (0.001)	0.056 (0.001)	0.165 (0.001)
R-squared	0.247	0.071	0.074	0.105	0.247

[a] Notes: Table reports estimated coefficient of linear education term in model that also includes cubic in potential experience and an indicator for non-white race. Samples include men and women age 16–66 who report positive wage and salary earnings in the previous year. Hourly wage is constructed by dividing wage and salary earnings by the product of weeks worked and usual hours per week. Data for individuals whose wage is under $2.00 or over $150.00 (in 1995 dollars) are dropped. Sample sizes are: 102,639 men and 95,309 women.

But is this true effect of going to college on earnings?

$$Estimate = Estimand + Bias + Noise$$
$$55\% = True\ Return + Bias + Noise$$

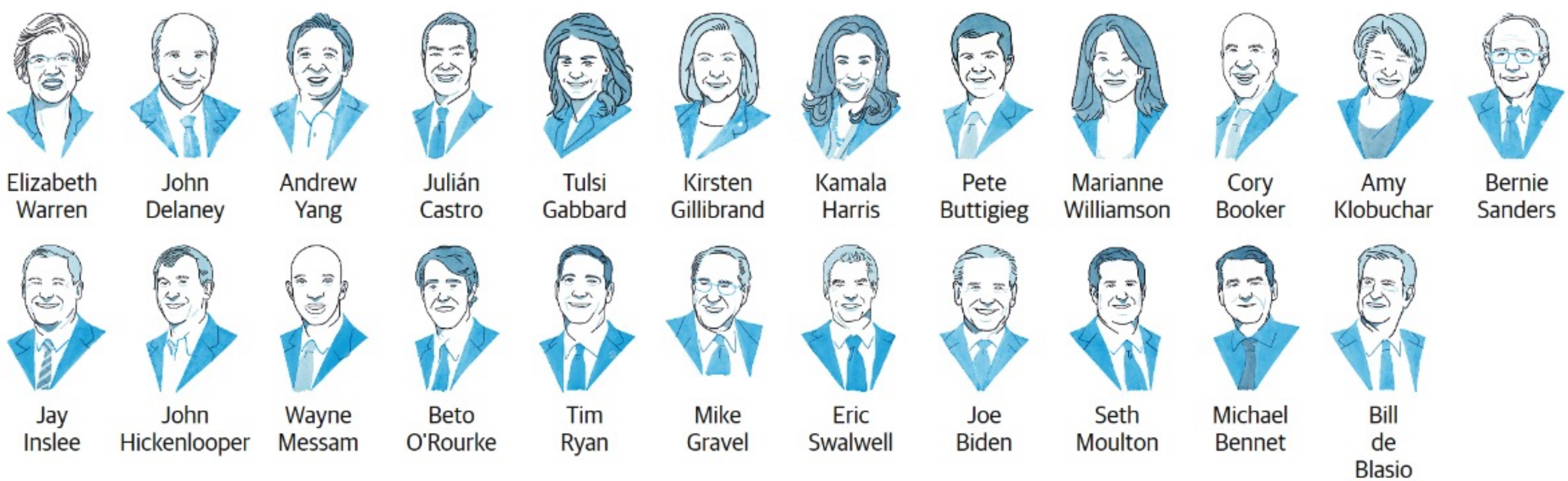
$$55\% = True\ Return + Bias + Noise$$

Two reasons an additional year of schooling might not increase earnings by 55%

1. Bias
   - E.g. Are we only changing education or is there **omitted variable bias**?
   - This is where the **ceteris paribus** assumption shows up. Is anything else changing besides the “treatment” we are trying to measure.

2. Noise

Historical Application: Who is going to win the Democratic primary?



How can we know?

How can we know?

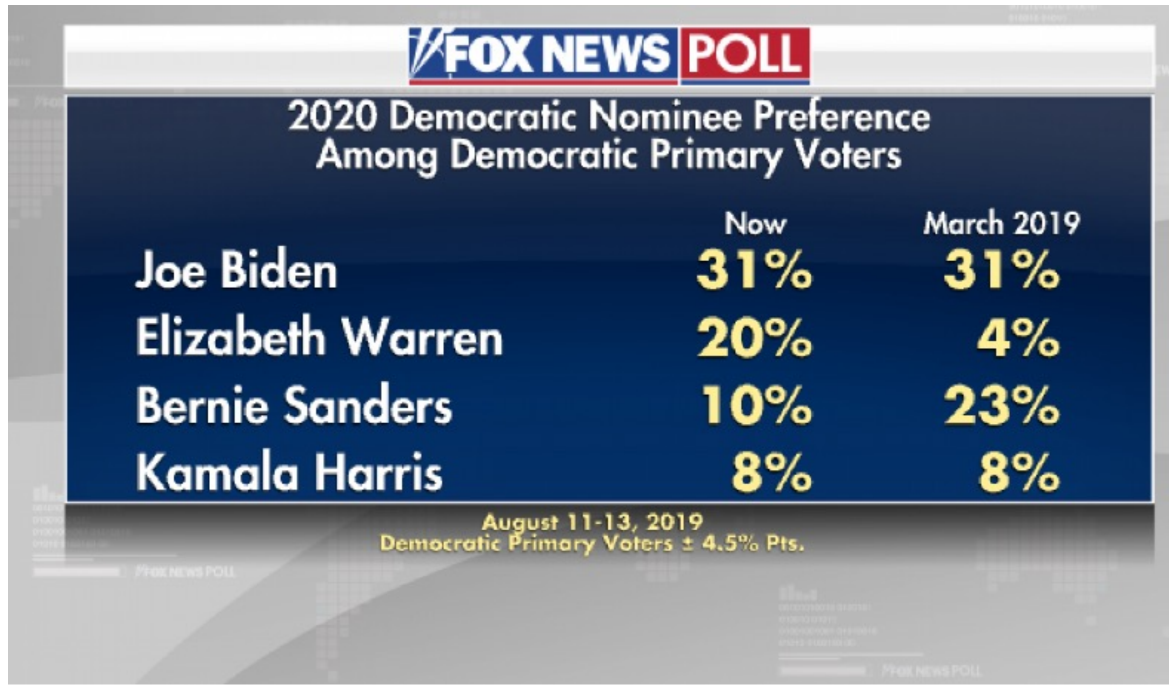
Conducted August 11-13, 2019

1,013 registered voters

222 landline phones

791 cellphone numbers

Randomly selected phone numbers



Defining Terms

$$Estimate = Estimand + Bias + Noise$$

Estimand: True percentage of voters who support Biden (or whoever)

Noise: Chance variation based on randomly selecting people

Bias: Any systematic differences between estimate and estimand

-Are Biden supporters more likely to pick up an unknown number?

Literary Digest and the Election of 1936

Landon v. Roosevelt

Literary Digest was an early pollster
- Created a list of 10 million potential voters
- Used phonebooks, magazine subscribers, club memberships, etc
- Sent all 10 million people a mock ballet

Literary Digest and the Election of 1936

Landon v. Roosevelt

Literary Digest was an early pollster
- Created a list of 10 million potential voters
- Used phonebooks, magazine subscribers, club memberships, etc
- Sent all 10 million people a mock ballet

Predicted Outcome: Landon 57%, Roosevelt 43%

What do the history buffs among you think of President Landon?

Literary Digest and the Election of 1936

Landon v. Roosevelt

Literary Digest was an early pollster
- Created a list of 10 million names
- Used phonebooks, magazine subscribers, club memberships, etc
- Sent all 10 million people a mock ballet

Predicted Outcome: Landon 57%, Roosevelt 43%
Actual Outcome: Roosevelt 61%, Landon 37%

What didn’t happen

1. Noise

The prediction was based on over 1,000,000 people.

Therefore there is very little noise in the estimate.

So that leaves: bias

1. Sample Selection

   The 10 million people were not selected randomly from the populaton. Biased towards middle and upper classes, more likely to prefer Landon.

2. Non-Response

   Only 25% of respondents returned mock ballet. Who responds could be correlated with support for Landon. For example, older people might be more likely to fill out the mock ballet and to support Landon.

Literary Digest's Failure

$$Estimate = Estimand + Bias + Noise$$

Literary Digest Estimate of Landon Vote Share: 57%

Estimand: 37% (true share of voters supporting Landon)

Noise: ~0

Bias: +20% from selection and non-response

$$Estimate = Estimand + Bias + Noise$$

How do we measure noise in data?

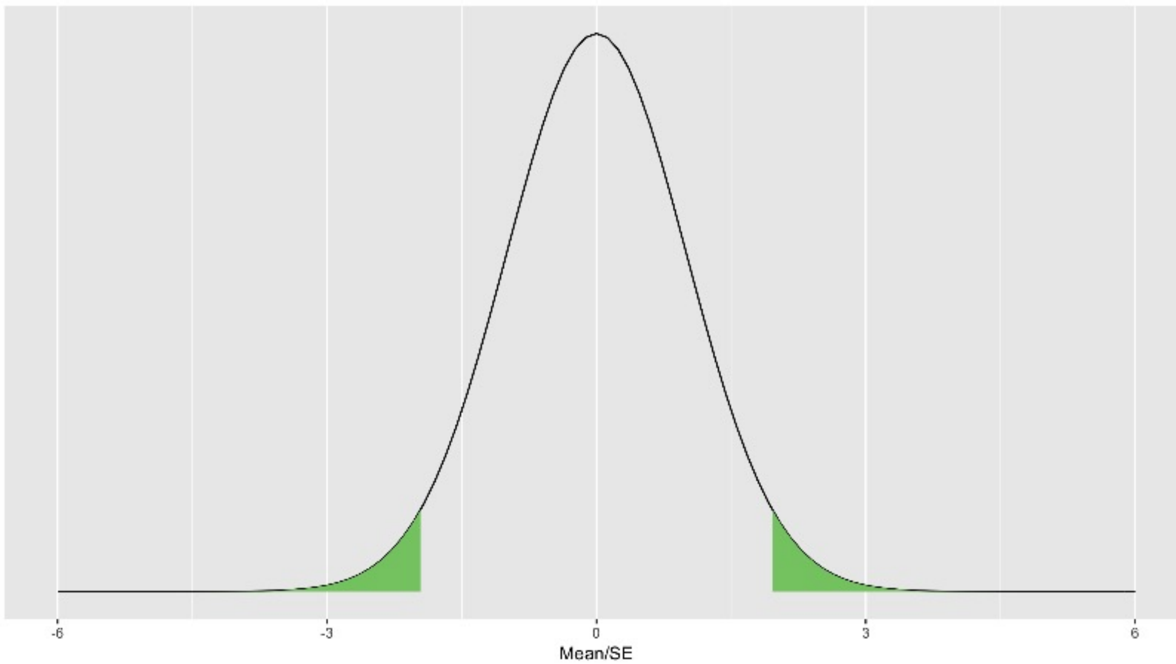
We quantify noise with Standard Errors

Standard errors quantify how much we expect estimates to vary across different samples

This is a plot of

$$\frac{Estimate}{Standard\ Error}$$

if effect is actually 0



How do we estimate standard errors?

Formulas for standard errors based on **central limit theorems**.

The standard error of an average is given by:

$$SE = \frac{\sigma}{\sqrt{N}}$$

Where:

- σ is the true standard deviation of the data
- N is the number of observations in the average

How do we estimate standard errors?

Formulas for standard errors based on **central limit theorems**.

The standard error of an average is given by:

$$SE = \frac{\sigma}{\sqrt{N}}$$

Complication: We are estimating the mean. So how would we know the true standard deviation?

How do we estimate standard errors?

The standard error of an average is given by:

$$SE = \frac{\sigma}{\sqrt{N}}$$

Complication: We are estimating the mean. So how would we know the standard deviation?

Solution: Estimate it with our best guess:

$$\widehat{SE} = \frac{\hat{\sigma}}{\sqrt{N}}$$

Exercise in R

Statistical Significance

We call it **hypothesis testing** when we test if something is just noise

Our **null hypothesis** is usually no effect. $H_0: \beta = 0$

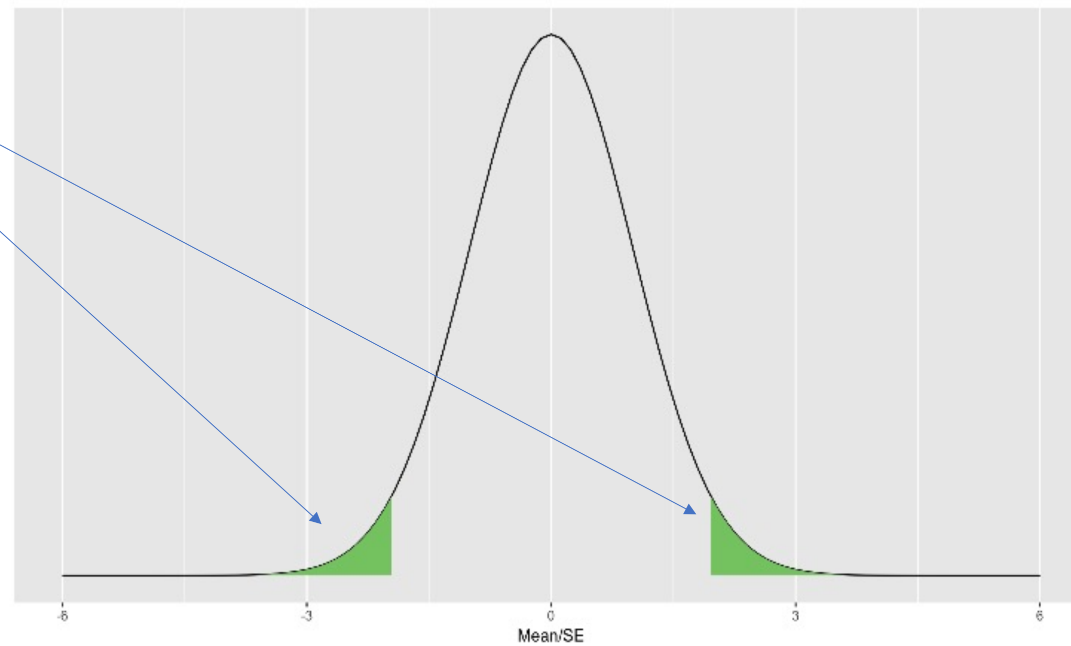
We **reject the null** if the effect is big enough relative to noise

Expect to see $|\frac{\hat{\beta}}{s.e.(\beta)}| > 1.96$ less than 5% of the time if β is actually 0

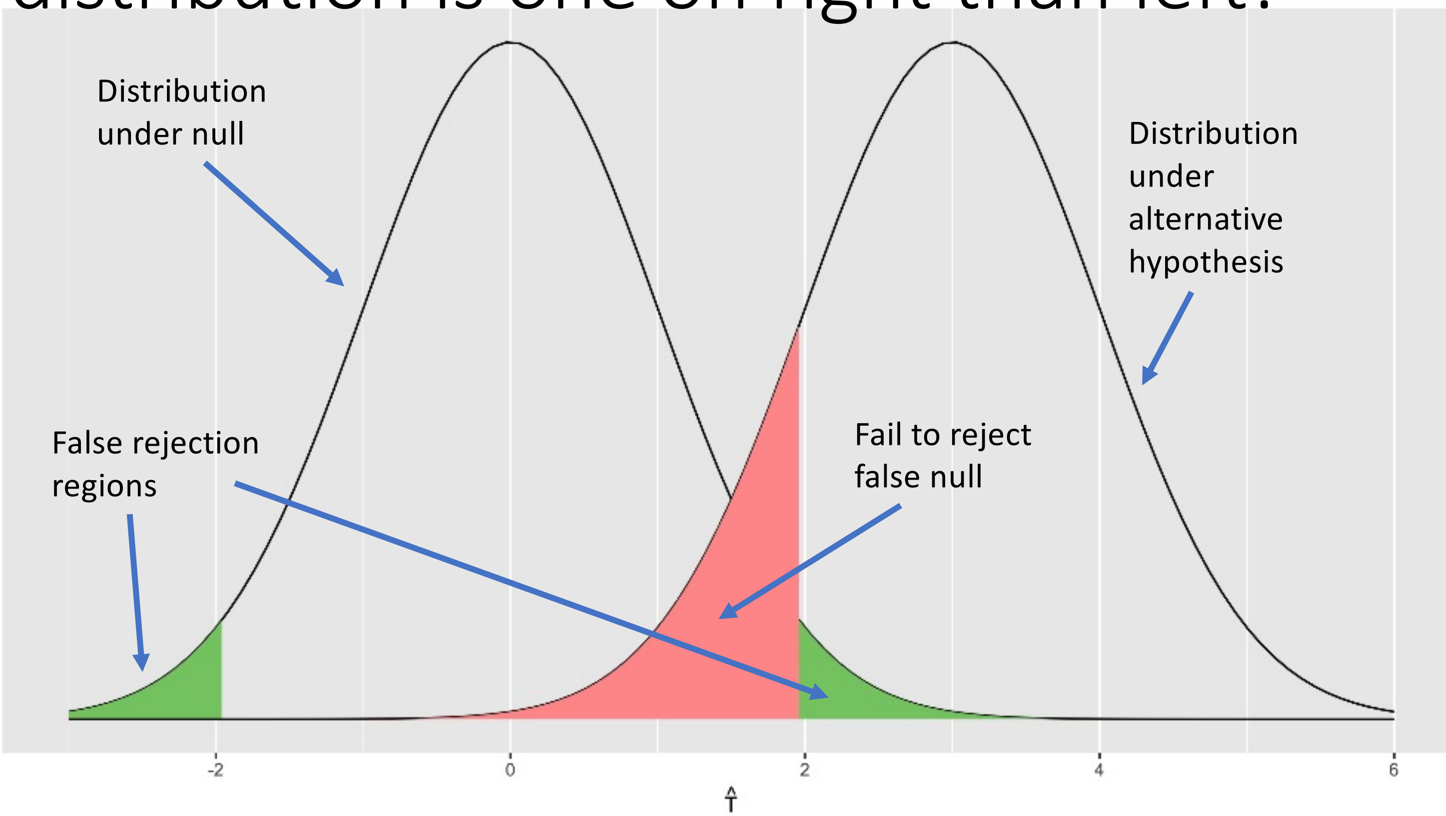
Statistical Significance

$$\left|\frac{\hat{\beta}}{s.e.(\beta)}\right| > 1.96$$

"Significant at 5% level"



Just saying it is more likely that the true distribution is one on right than left!



Caution in Hypothesis Testing

1. Remember: $Estimate = Estimand + Bias + Noise$.

   -Can be precisely estimated around Estimand+Bias!

Caution in Hypothesis Testing

1. Remember: $Estimate = Estimand + Bias + Noise$.

   -Can be precisely estimated around Estimand+Bias!

2. Based on probability of rejecting 1 test. If test 20 things in a paper:

   P(Reject at least 1 hypothesis at 5% level|20 tests)=$1 - 0.95^{20} = 0.64$

Economic/Substantive Significance

Statistical Significance: Indicates that a result was unlikely to have arisen by chance.

Substantive Significance: How big or small or important is an effect.

$$Statistical\ Significance \neq Substantive\ Significance$$

Example: Social Pressure to Vote

2012 article in *Nature* (a very good journal)

Result: People statistically significantly more likely to vote in 2010 US midterm elections if their Facebook page displayed a banner indicating which of their friends votes ($p=0.02$).

- 61 million voting age-users

Example: Social Pressure to Vote

2012 article in *Nature* (a very good journal)

Result:

> “Those who saw the social message were… 0.4% more likely to head to the polls than either other group.”

Or in other words: 4 extra people voted for every 1,000 who saw the message

Example: Second Reform Act of 1867

Berlinksi and Dewarn (QJPS, 2011)

Study Second Reform Act of 1867 in UK which doubled size of electorate

Result: “There is no evidence relating Liberal electoral support to changes in the franchise rules.”

Example: Second Reform Act of 1867

Berlinksi and Dewarn (QJPS, 2011)

Study Second Reform Act of 1867 in UK which doubled size of electorate

Result Revisited: The Second Reform Act increased liberal vote share by 8 percentage points but this effect was not statistically significant.

Example: Second Reform Act of 1867

Berlinksi and Dewarn (QJPS, 2011)

Study Second Reform Act of 1867 in UK which doubled size of electorate

Result Revisited: The Second Reform Act increased liberal vote share by 8 percentage points but this effect was not statistically significant.

Takeaway: There can be noisy evidence about an important effect!

$$Estimate = Estimand + Bias + Noise$$

Sources of Bias

Let's think back to returns to education

55% estimate is likely biased because lots of other things changing

Kids who go to college:

- More likely have parents who went to college
- Like school more
- Have better prior academic performance
- Live closer to a college or university
- Etc.

Controlling for Differences

$$E[Y|X = x] = \alpha + X\beta + \gamma College$$

Regressions make it easy to “control” for other characteristics

X here can include anything we observe about people

Impact of college is the impact holding everything in X constant

Problem solved?

Controlling for differences was the standard for research 30 years ago

Today: No serious researcher believes we can control for all differences.

There are too many things we can’t observe and therefore can’t control for.

Solution: Randomized Experiments

Randomized experiments are the gold standard

Use randomization to hold everything constant across groups

If I split you in groups by flipping a coin, each group should reflect the characteristics of the whole class, on average

Compare average outcomes across groups that only differ in result of a coin flip

Advantages of Experiments (Falk and Heckman, 2009; Fehr notes, 2004)

Subjects randomly assigned so no selection bias:

- i.e. high/low valuations random in Chamberlin (1948)

Variables that cannot be observed are observable in lab

- i.e. assigned valuations in Chamberlin (1948)

More replicable than naturally occurring data

Critiques of Experiments (Falk and Heckman, 2009)

1. Lack of realism
2. Samples not representative
3. Stakes are trivial
4. Sample sizes too small
5. Participants are inexperienced
6. “Hawthorne effects” distort results
7. Self-selection into experiment biases results

1. Lack of realism

"it seems to be extraordinarily optimistic to assume that behavior in an artificially constructed "market" game would provide direct insight into actual market behavior."

-Cross (1980)

Internal versus External Validity

Internal Validity: Can the researcher credibly estimate the effect of the feature of interest within the context of the experiment?

- Randomized experiments have high internal validity if implemented and analyzed correctly

External Validity: Can the researcher credibly estimate the effect of the feature of interest in a new population or a new environment?

- Is the sample representative of the population of interest?
- Is the situation similar to other contexts?
- Etc.

The reply…

“Experiments are sometimes criticized for not being ‘realistic’….are there field data to support the criticism, i.e., data suggesting that there may be differences between laboratory and field behavior. If not, then the criticism is pure speculation.”

-Smith (AER, 1980)

Potential Solution: Move the lab to the field and test generalizability…

2. Samples not representative

“even though these results appear prevalent, they are suspiciously drawn…. by methods similar to scientific numerology…. because of students…. who are not ‘real’ people”

Same Solution: Test generalizability of results in broader samples…

3. Stakes are trivial

What are the right stakes? How often do you make decisions involving a few dollars? Your monthly income?

Can measure impact of stakes on estimates experimentally

Can conduct experiments in developing world to have higher relative stakes at similar cost

3. Stakes are trivial

Camerer and Hogarth (1999) conduct meta-analysis of impact of stakes:

> We review 74 experiments with no, low, or high performance-based financial incentives. The modal result has no effect on mean performance (though variance is usually reduced by higher payment). Higher incentive does improve performance often, typically judgment tasks that are responsive to better effort. Incentives also reduce "presentation" effects (e.g., generosity and risk-seeking). Incentive effects are comparable to effects of other variables, particularly "cognitive capital" and task "production" demands, and interact with those variables, so a narrow-minded focus on incentives alone is misguided. We also note that no replicated study has made rationality violations disappear purely by raising incentives.

3. Stakes are trivial

Andersen et al. (2011) replicated famous “ultimatum game” in 8 villages in Northeast India

Randomly assigned to 4 levels of stakes: 20 (\$0.41), 200 (\$4.1), 2,000 (\$41), and 20,000 (\$410) Indian rupees

2,000 rupees equates to about 160 hours of work

Find that results differ when stakes are very large

4. Sample sizes too small

Can measure this effect by running experiments with bigger sample sizes.

Importantly, appropriate inference accounts for sample size so this critique is a bit misdirected.

5. Participants are inexperienced

Another critique that can be tested directly

Will discuss List (2003) “Does market experience eliminate anomalies?” this quater

6. “Hawthorne effects” distort results

“Hawthorne effect” is when individuals modify their behavior because they are being observed.

The experiment might have an effect because people realize a researcher is monitoring them.

One Response: Data collection almost always involves observation. Experiment or not.

7. Self-selection into experiment biases results

Self-selection can be informative about preferences

Will discuss econometric methods for dealing with non-compliance, attrition, and (to a lesser extent) randomization bias

A Spectrum of Experiments

______AFE_________FFE_________NFE______________________________

Lab **[field experiments]** **Non-experimental Methods**

- conventional lab experiment (Lab)
  - employs a standard subject pool of students, an abstract framing, and an imposed set of rules
- artefactual field experiment (AFE)
  - same as a conventional lab experiment but with a non-standard subject pool
  - Sometimes called “lab in the field”
- framed field experiment (FFE)
  - same as an artefactual field experiment but with field context in the commodity, task, information, stakes, time frame, etc.
- natural field experiment (NFE)
  - same as a framed field experiment but where the environment is the one that the subjects naturally undertake these tasks, such that the subjects do not know that they are in an experiment

Conventional Lab Experiment
(Coller and Williams, *Exp. Econ.*, 1999)

Each of six experimental sessions consisted of approximately 35 graduate and undergraduate students recruited from various School of Business classes at the University of South Carolina. For five of the six sessions, subjects were presented with the following initial information:[9]

> *One* person in this room will be randomly chosen to receive a large sum of money. If you are the individual chosen to receive this money (the "Assignee"), you will have a choice of two payment options; option A or option B. If you choose option B you will receive a sum of money 3 months from today. If you choose option A you will receive a sum of money 1 month from today, but this option (A) will pay a smaller amount than option B.

The remaining session differed only in that option A paid a sum of money on the day of the experiment, while option B paid the larger sum two months later. Research budget constraints dictated that only one person in each session could be paid. This person (the "Assignee") was chosen at random at the end of the experiment.[10] To ensure credibility of the payment instrument, a notarized payment certificate was given as a guarantee of payment.[11]

Option A was \$500 in all treatments. Option B paid $\$500 + \$x$ where $\$x$ ranged from \$1.67 (reflecting a 2% annual rate of return on the \$500 principal compounded daily) to \$90.54 (reflecting a 100% annual rate of return compounded daily). \$500 was chosen as the minimum payment in order to ensure that *all* subjects would have an opportunity to arbitrage in the field, regardless of whether or not they had an established investment vehicle.[12]

Artefactual Field Experiment

(Harrison, Lau, and Williams, *AER,* 2002)

In 1996 the Danish Ministry of Business and Industry contracted with the Danish Social Research Institute (SFI, after the Danish name *Socialforskningsinstituttet*) to undertake the field surveys.[9] The final surveys were conducted between June 16 and July 8, 1997, throughout Denmark.

The sample population consisted of a random selection from individuals 19–75 years old who had participated all three times in the European Community Household Panel Survey (ECHP) previously conducted by SFI. These persons were chosen because they had some experience with respect to economic surveys, and because we could expect a high response rate. The sample was constructed in two steps.

The 275 municipalities in Denmark were proportionally stratified with respect to the number of persons between 19 and 75 years of age on January 1, 1997. Copenhagen and Aarhus, the two largest municipalities, had their own stratum due to their size. Most of the other municipalities were divided into 23 strata. Some remote municipalities, primarily tiny islands, were not represented in the sample because the population is relatively small and the subjects would spend too much time traveling to the experimental session.

The 27 sessions were divided equally across geographic locations with 5, 10, and 15 participants in each experiment. In turn, the 27 sessions were located such that the number of participants at the experiments correspond to the relative size of the population in the given stratum. For example, approximately 11 percent of the population between 19 and 75 years of age live in Copenhagen, and three sessions with a total of 30 participants were held in Copenhagen, which corresponds to 11.1 percent of the total sample size.[10]

Upon arrival at the experimental session, subjects were given the following information:

> *One* person in this room will be randomly chosen to receive a large sum of money. If you are the individual chosen to receive this money (the "Assignee"), you will have a choice of *two* payment options; Option A or Option B. If you choose Option B you will receive a sum of money 7 months from today. If you choose Option A, you will receive a sum of money 1 month from today, but this Option (A) will pay a smaller amount than Option B.

Framed Field Experiment (List, *AER*, 2001)

and Shogren, 1998a). I conducted the treatments in December 1998 at a sportscard show in Tampa, FL, where there was a large supply of card collectors.

For all treatments, I auctioned off a Cal Ripken, Jr. 1982 *Topps Traded* baseball card, which had a book value in the range of \$200–\$250 (depending on which magazine publication was consulted). All auctions displayed the same sportscard to all bidders—a Cal Ripken, Jr. PSA-graded "PSA 8 near-mint/mint" baseball card. To perform the simplest possible test of hypothetical bias, I chose an allocation mechanism—William Vickrey's (1961) second-price auction—which has proved straightforward in other field experiments. While not impeccable, the second-price auction has performed reasonably

B. *Experimental Procedure*

Each participant's experience followed four steps: (1) inspection of the good, (2) learning the auction rules, (3) placing one bid, and (4) conclusion of the transaction. In Step 1, a potential subject approached the administrator's table and inquired about the sale of the sportscard displayed on the table. The monitor then invited the potential subject to participate in an auction for the sportscard that would take about five minutes. If the individual agreed, the monitor briefly explained that we were auctioning off the sportscard displayed on the table. The participant could pick up and visually examine the card. The card was sealed with the appropriate grade clearly marked on the cardholder. The monitor worked one-on-one with the participant and no time limit was imposed on his or her inspection of the card.

Natural Field Experiment
(List and Lucking-Reiley, 2002)

We design a field experiment to test two theories of fund-raising for threshold public goods: Andreoni predicts that publicly announced "seed money" will increase charitable donations, whereas Bagnoli and Lipman predict a similar increase for a refund policy. Experimentally manipulating a solicitation of 3,000 households for a university capital campaign produced data confirming both predictions. Increasing seed money from 10 percent to 67 percent of the campaign goal produced a nearly sixfold increase in contributions, with significant effects on both participation rates and average gift size. Imposing a refund increased contributions by a more modest 20 percent, with significant effects on average gift size.