

# EW MBA 296 (Fall 2015)

## Section 1

GSI: Fenella Carpena

October 22, 2015

# Agenda for Today

- ▶ Housekeeping: Announcements
  - ▶ Changes to office hours
  - ▶ Slight changes to section times
- ▶ Samples and Surveys: The Big Picture
- ▶ Sampling Methods: Review
- ▶ Common Sources of Bias in Survey Design
- ▶ Practice Problems
  - ▶ Tesla Handout Part B
  - ▶ Section Notes 1, Exercises 1 and 2
  - ▶ Stine and Foster, Chap. 13, Exercise 27 (time permitting)

# Housekeeping: Announcements

- ▶ Changes to office hours
  - ▶ Starting next week (week of 10/25), OH moved to Tuesdays, 6:30PM-8:30PM in Adobe Connect
  - ▶ <http://haasberkeley.adobeconnect.com/ewmba296-oh>
  - ▶ This Saturday, 10/24 I will still have OH 8-9AM and 1-2PM in F421. But the Sat OH will be replaced with Tue OH in succeeding weeks.
  - ▶ Please feel free to email me if the OH times don't work for you.
- ▶ Main comments from feedback survey from last week
  - ▶ Voluntary response sample: 42% response rate, not representative
  - ▶ Better pen: working on it!
  - ▶ Merging both sections to have more section time: unfortunately not possible due to Adobe Connect constraints
- ▶ Changes to section timing
  - ▶ We'll start on the hour instead of Berkeley time
  - ▶ Section Axe: 6:30PM-7:30PM
  - ▶ Section Oski: 7:40PM-8:40PM

# Samples and Surveys: The Big Picture

- ▶ We want to answer questions about the **population**.
- ▶ But collecting data from the entire population is infeasible or impractical. So we use a **sample**.
  - ▶ A sample is said to be **representative** if it reflects a snapshot of the population.
  - ▶ A sample that distorts the population is said to have **bias**.
  - ▶ A representative sample can be achieved by **randomization**.
- ▶ Using data from the sample, we can calculate a **sample statistic**.
- ▶ Because **there is only one population but many possible samples**, the value of the sample statistic we obtain would differ from sample to sample. This variability is called **sampling variation**.

# Sampling Methods: Review

- ▶ **Example:** Suppose that you are the head of a global company with a total of 1,000,000 employees based in 100 offices around the world. The company would like to assess the general mental health of its staff by conducting a one-on-one consultation with a sample of 1,000 employees. What are the different sampling methods you can use? What are the potential benefits and potential issues of each method?

# Sampling Methods: Review

## Simple Random Sample

- ▶ **Example:** Randomly draw a sample of 1,000 people from a list of all employees
- ▶ **Potential Benefit:** Easy to implement
- ▶ **Potential Issues:**
  - ▶ May not be possible to examine employee's mental health by country (b/c smaller groups may not be represented in sample)
    - ▶ We can use stratified sampling to overcome this issue.
  - ▶ Can be costly
    - ▶ We can use clustered sampling to overcome this issue.

# Sampling Methods: Review

## Systematic Sampling

- ▶ **Example:** List the names of all employees in order of their employee ID number, then include every 1000th name in the sample.
- ▶ **Potential Benefit:** Easy to implement (no need to use random number generator)
- ▶ **Potential Issues:** Need to assume/justify that the order of the list and the sampling mechanism is *\*not\** associated with responses.

# Sampling Methods: Review

## Stratified Sampling

- ▶ **Example:** Divide the sampling frame by country office and within each office, take a random sample of employees.
- ▶ **Potential Benefit:** Smaller offices can be represented in the sample.
- ▶ **Potential Issues:** Data on the characteristics we want to stratify on may not be available.



# Sampling Methods: Review

## Cluster Sampling

- ▶ **Example:** Divide the sampling frame by country office.
  - ▶ Step 1: Take a random sample of country offices
  - ▶ Step 2: Within each sampled office, take a random sample of employees.
  - ▶ Note the difference between stratified sampling and clustered sampling
- ▶ **Potential Benefit:** Less costly to interview employees since we only have to visit those offices that were sampled.
- ▶ **Potential Issues:** Gives us less information about the population than a simple random sample, since individuals who work in the same office tend to be homogenous.

# Common Sources of Bias in Survey Design

- ▶ **Voluntary Response Samples.** A group of individuals are invited to respond, but only those who respond are counted in the sample. Biased towards those with strong opinions.
- ▶ **Convenience Samples.** Sampling individuals who are readily available (e.g., surveys done at shopping malls.)
- ▶ **Survivor Bias.** Bias in the sample that arises from selecting items that are present in the sample frame for a longer period of time. Those units that have survived (and therefore made it to the sample) are likely to be very different from those who did not.

## Tesla Handout, Part B

Suppose that Tesla wanted to learn whether it should offer a new long-distance battery for the forthcoming Model X.

(a) What is the population of interest here?

(b) Is it reasonable to expect that this population of interest could be used as the sampling frame?

## Tesla Handout, Part B

(c) Suppose Tesla wished to collect detailed information using a face-to-face survey. To save costs, a member of the analytics team suggests using a clustered design. What are the benefits and drawbacks of a clustered design? *Hint: Suppose you survey customers in California; think about what happens to your results if you happen to include the cluster of San Francisco versus the cluster of Los Angeles, versus neither.*

## Tesla Handout, Part B

(d) If Tesla again decides to survey 2014 Model S owners to learn about the demand for Model X, how will this affect your views on the information collected in the survey—is it likely to be representative of the views of the population of interest?

## Section Notes 1, Exercise 1

Suppose an employer recently replaced its paycheck system with a paperless system that directly deposits payments into the employee's bank account. The employer would like to determine whether employees are satisfied with the new system. To do so, the employer mailed a questionnaire to a simple random sample (SRS) of 500 employees, and received responses from 10% of them. The responses indicated that 90% were not satisfied with the paperless system. Is this sample representative? Based on these results, do you think that the employer should return to the former paycheck system?

## Section Notes 1, Exercise 2

Suppose that a local school district has hired you as a consultant to examine the performance of its high schools. Sub-district A has a large proportion of high-income students, while sub-district B has a large proportion of low-income students. You implemented a stratified random sample for your survey, selecting 100 graduating students from each sub-district to interview. Assume that all sampled students responded to the survey, and that all students reported the truth. Using the survey data, you find that in both sub-districts, 80% of all graduating students will be attending college in the fall. With this result, would you conclude that the school district has succeeded in narrowing the achievement gap between high-income and low-income students?

## Stine and Foster, Chap. 13, Exercise 27

A bank with branches in a large metropolitan area is considering opening its offices on Saturday, but it is uncertain whether customers will prefer (1) having walk-in hours on Saturday, or (2) having extended branch hours during the week. Listed below are some of the ideas proposed for gathering data. For each, indicate what kind of sampling strategy is involved and what (if any) biases might result.

(a) Put a big ad in the newspaper asking people to log their opinions on the bank's website.



## Stine and Foster, Chap. 13, Exercise 27

For each, indicate what kind of sampling strategy is involved and what (if any) biases might result.

(b) Randomly select one of the branches and contact every customer at that bank by phone.

## Stine and Foster, Chap. 13, Exercise 27

For each, indicate what kind of sampling strategy is involved and what (if any) biases might result.

(c) Send a survey to every customer's home, and ask the customers to fill it out and return it.

## Stine and Foster, Chap. 13, Exercise 27

For each, indicate what kind of sampling strategy is involved and what (if any) biases might result.

(d) Randomly select 20 customers from each branch. Send each a survey, and follow up with a phone call if he or she does not return the survey within a week.

# Final Points

- ▶ Quiz 2 Logistics
  - ▶ Covers Intro + Chapter 13 (Lectures 1 and 2, from last Saturday)
  - ▶ You can bring a single-sided, 8.5 x 11 sheet of notes and use a calculator.
  - ▶ *Remember to check the back of the page and make sure you've answered all questions.*
- ▶ Other materials I can suggest for Quiz 2 review:
  - ▶ eBay handout from Lecture
  - ▶ Quiz 2 from 2013/2014
  - ▶ Stine and Foster, Chap 13: Exercise 31, 35, 36
  - ▶ All solutions to the above posted in bCourses. Textbook solutions are at the back of the book.
- ▶ Lecture this Saturday (10/24) will cover Chapter 14 (Section 1 and 2) and Chapter 15