Exhibit 2 + 3

BEYOND THE NUMBERS 1.2

What's Simple about an SRS?

Name: _____ Section Number: _____

To be graded, all assignments must be completed and submitted on the original book page.

Background

The definition of a simple random sample (SRS) can be confusing: An SRS of size *n* is a sample of size *n*, chosen in such a way that all samples of size *n* have the same chance of being chosen. It doesn't help that the word "random" is used in many different ways, but when it comes to selecting a simple random sample, we have to be very careful to know its technical meaning. We will explore these issues in this set of activities.

EXHIBIT 1

Random Evolution

On November 30th, 2012, National Public Radio ran a short segment entitled "That's So Random: The Evolution of an Odd Word." You may find this segment at: http://www.npr.org/2012/11/30/166240531/thats-so-random-the-evolution-of-an-odd-word.

Question

1. List two uses of the word "random" from the audio that are different from the technical definition given above. How do you use the term "random" in your own life?

EXHIBIT 2

Careful Counting

The audio segment ends with Charlie McDonnell (of the British "Fun Science" videos) noting that "every now and then, at random, you end up with something awesome." We might take that to mean that every now and then, a simple random sample is representative of a population with respect to a certain list of demographics. Let's look at a simple example to see how likely that might be. Suppose you have a population with two men—one a Republican and one a Democrat; and two women—one a Republican and one a Democrat.

Questions

1. List all possible distinct samples of size two from this four-person population. Make sure that your notation makes it possible to distinguish all four members of the population.

① Male Republican + Male Democrat
② Female Republican + Female Democrat
③ Male Republican + Female Republican
④ Male Republican + Female Democrat
⑤ Male Democrat + Female Republican
⑥ Male Democrat + Female Democrat

2. For a simple random sample of size two, all samples of size two have the same chance of being chosen. What would the likelihood be of choosing any one of these samples?

1 out of 6

3. Suppose for a sample of size two to be "representative" of the population, it has to have exactly one man and one woman, and one Democrat and one Republican. What is the chance of selecting a simple random sample of size two from this population that is representative (in this sense of the word)?

1 out of 3

EXHIBIT 3

Social Media Sampling

Suppose you have 113 friends on Facebook and you want to choose a simple random sample of 20 of them. Answer the following:

Questions

1. What is your population?

113 friends on Facebook

2. Describe in detail how you would select your simple random sample.

① Place all 113 friend names on separate pieces of paper
② Place all papers in hat and mix
③ Choose 20 names out of the hat

Alternative: ① List all 113 names in separate cells in Excel
② Assign random numbers to names "=Rand()"
③ Sort by random number + choose top 20

Exhibit 1

BEYOND THE NUMBERS 1.3

Are Online Reviews Statistical Samples?

Name: _____ Section Number:

To be graded, all assignments must be completed and submitted on the original book page.

EXHIBIT 1

Bravos for Bucks

The VIP brand Kindle Fire cover received 4,945 reviews on Amazon by early 2012, averaging a nearly perfect 4.9 stars out of five. That's quite impressive. It is tempting to think that online reviews, especially those posted at major sites like Amazon, are representative of consumer experiences. We know, however, that voluntary responses are often biased. But are product reviews even less accurate than previously thought? In his 2012 *Time* article "9 Reasons Why You Shouldn't Trust Online Reviews," Brad Tuttle writes, "You shouldn't believe everything you read. And if you're reading online reviews of products, hotels, restaurants, or local businesses or services? Then you should believe even less." You can find Tuttle's article online at: http://business.time.com/2012/02/03/9-reasons-why-you-shouldnt-trust-online-reviews/.

Questions

1. Describe three reasons listed in the article as to why you should be very cautious about online reviews.
   ① Marketplace for fake reviews operates fairly openly
   ② Companies give freebies in exchange for reviews
   ③ Review could be real but work for the company being reviewed

2. What was VIP doing to boost the ratings of its Kindle Fire cover? Be specific.
   VIP Deals reimbursed customers for their tablet case if they posted a review

3. Compared to computer algorithms, how well did people perform in spotting fake reviews? How does this study inform your perception of online reviews?
   - People spotted fake reviews ~50% of time
   - I've always been skeptical of online reviews, so this didn't teach me anything new.

Questions 1 + 3

BEYOND THE NUMBERS 1.4

Random or Representative?

Name: _____ Section Number: _____

To be graded, all assignments must be completed and submitted on the original book page.

EXHIBIT 1

Gulliver Travels

900 people live in Gulliver, a small town in Michigan's Upper Peninsula.[1] You want to know what proportion of Gulliver's population supports legalizing marijuana. Suppose you already know the following demographic information about Gulliver's 900 citizens:

Gender
500 are Females

Income	20% make more than $80,000	70% make between $40,000 and $80,000	10% make less than $40,000
Political Affilliation	70% Democrat / 30% Republican	50% Democrat / 50% Republican	30% Democrat / 70% Republican

Gender
400 are Males

Income	40% make more than $80,000	50% make between $40,000 and $80,000	10% make less than $40,000
Political Affilliation	60% Democrat / 40% Republican	50% Democrat / 50% Republican	40% Democrat / 60% Republican

Questions

1. You have enough money to interview 90 residents. Working much the way Gallup did in the 1930s, you want your sample of 90 to mirror the distribution of subjects in the population exactly (at least along the lines of gender, income, and political affiliation). How many people would your sample place in the groups shown on the next page? If a calculation results in a partial person (e.g., 6.4 persons), leave the number as it is—don't round.

See table on next page

[1] There really is a Gulliver, MI of about this size. The demographics are completely made up, however.

TABLE 1.1 Population Data

Category	Number of Persons
Males	40
Females	50
Males making between $40,000 and $80,000 yearly	20
Females making less than $40,000 per year who are Democrats	1
Male Republicans making over $80,000 per year	6

2. Suppose the cross-sectional sample taken above represents a perfect microcosm of the larger population with respect to the legalization of marijuana. Is there any uncertainty involved in using this sample to represent the proportion of people in Gulliver who favor the legalization of marijuana? Explain.

3. Suppose you decided, instead, to take a simple random sample of Gulliver's population. Explain how you could take an SRS of size 90 from this population.

(1) Assign random number to each of the 900 citizens
(2) Sort by random number + choose first 90

4. A carefully chosen simple random sample may not be representative of the population. Explain how this could be.

BEYOND THE NUMBERS 1.5

Research Randomizer

Name: _____ Section Number: _____

To be graded, all assignments must be completed and submitted on the original book page.

Background

A simple random sample is the easiest kind of statistically viable sample to select and measure. But how do you actually select an SRS? One useful tool is the Research Randomizer, available at: http://www.randomizer.org/. The following activities are designed to allow you to get familiar with this tool.

EXHIBIT 1

No-Stumble Sampling

Data from the NHTSA's 1998 San Diego field sobriety test validation study is available at www.statconcepts.com/datasets. There are 296 participants in this study, so there are 296 case numbers displayed. Note, though, that these case numbers do not run sequentially from 1 to 296. Your job is to use Research Randomizer to select a sample of 20 cases from this data set.

Questions

1. Explain how you plan to identify the cases for Research Randomizer

2. What entries did you use for the following Research Randomizer fields?

TABLE 1.2 Research Randomizer Fields

How many sets of numbers do you want to generate?	How many numbers per set?	Number Range (e.g., 1–50)

Exhibit 2

e 20 cases selected, fill out the following chart:

LE 1.3 Results for 20 Cases

Case Number	Actual BAC	Case Number	Actual BAC

4. What is the average BAC ("Blood Alcohol Content") of the 20 selected cases?

5. What proportion of cases in your sample had BACs at or above the legal limit of 0.04?

EXHIBIT 2

Social Media Sampling Revisited

Suppose you have 113 friends on Facebook and you want to choose a simple random sample of 20 of them to ask a survey question you have constructed.

Question

1. Carefully explain how you could use Research Randomizer to select your sample.

① Assign numbers to all 113 friends

② Use Randomizer to generate 1 set of 20 numbers with a range 1 to 113

③ Select sample with student numbers + random number set

Exhibit 1

BEYOND THE NUMBERS 1.6

How Do National Polls Sample?

Name: _____ Section Number: _____

To be graded, all assignments must be completed and submitted on the original book page.

EXHIBIT 1

Getting Gallup

The following is an excerpt from Gallup's website describing their Daily Tracking Survey. The page is available at http://www.gallup.com/174155/gallup-daily-tracking-methodology.aspx.

> Gallup interviews U.S. adults aged 18 and older living in all 50 states and the District of Columbia using a dual-frame design, which includes both landline and cellphone numbers. Gallup samples landline and cell phone numbers using random-digit-dial methods. Gallup purchases samples for this study from Survey Sampling International (SSI). Gallup chooses landline respondents at random within each household based on which member had the next birthday. Each sample of national adults includes a minimum quota of 50% cellphone respondents and 50% landline respondents, with additional minimum quotas by time zone within region. Gallup conducts interviews in Spanish for respondents who are primarily Spanish-speaking.

Questions

1. What is the actual population being addressed by a Gallup telephone survey? Be very precise with your answer.

U.S adults aged 18 and oldst living in all 50 states and District of Columbia using a dual-frame design, which includes both landline + cellphone numbers

2. In what sense can a random-digit-dial sample be thought of as a simple random sample? Be very specific.

- Every number in the purchased sample has an equal chance of being dialed

BEYOND THE NUMBERS 2.1

Slippery Evidence and Confounding

Name: _____ Section Number: _____

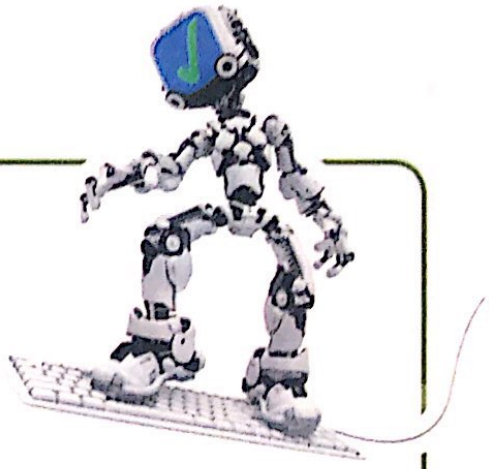
To be graded, all assignments must be completed and submitted on the original book page.

Carefully read and think about each of the exhibits below. Then, give detailed answers to the questions that accompany each exhibit.

EXHIBIT 1

Questions

1. Looking at the Results section, what conclusion are you likely to make about the effectiveness of online instruction?



Thinking Critically

Title: Learning in an Online Format versus an In-Class Format: An Experimental Study

Authors: Allan H. Schulman and Randi L. Sims

Source: *T.H.E. Journal* 26, no. 11 (1999): 54–56

Methodology Students enrolled in five different undergraduate online courses during the Fall semester 1997 participated in a voluntary test-retest study designed to measure their learning of the course material. These students were compared with students enrolled in traditional in-class courses taught by the same instructors.

Subjects In total, 40 undergraduate students were enrolled in the online courses and 59 undergraduate students were enrolled in the in-class courses during the testing period.

Pre-tests Instructors designed pre-tests to measure the level of knowledge students had of the course content prior to the start of the course. The average pre-test scores for online students was 40.70 (s.d. = 24.03). The average pre-test scores for in-class students was 27.64 (s.d. = 21.62).

Post-tests Instructors designed post-tests on a 100-point scale to test students' knowledge at the end of the course. The average post-test scores for online students was 77.80 (s.d. = 18.64). The average post-test scores for in-class students was 77.58 (s.d. = 16.93).

Results [O]ur results indicate that there were no significant differences for post-test scores. ...

Give at least two reasons why this conclusion might be compromised. Be sure your reasons come from the part of this paper that you have access to here.

2
19.1423

EXHIBIT 2

"Make Mine a Large"

In the 2009 *New York Times* piece "Excess Pounds, but Not Too Many, May Lead to Longer Life," author Roni Caryn Rabin reported:

> Being overweight won't kill you—it may even help you live longer. That's the latest from a study that analyzed data on 11,326 Canadian adults, ages 25 and older, who were followed over a 12-year period. The report ... found that overall, people who were overweight but not obese—defined as a body mass index of 25 to 29.9—were actually less likely to die than people of normal weight, defined as a B.M.I. of 18.5 to 24.9.
>
> By contrast, people who were underweight, with a B.M.I. under 18.5, were more likely to die than those of average weight. Their risk of dying was 73% higher than that of normal weight people.

Question

1. Although this article doesn't describe an experiment, it does imply that being a little overweight may lead to a longer life. Identify at least one confounding variable that may compromise the validity of this inference. Support your case.

- This is a 12 year study, so healthcare for overweight people may have improved at a faster rate than compared to healthcare for typical "normal" weight. Overweight health problems like cholesterol, type II diabetes, cardiovascular disease, & physical exercise have had a lot of attention & funding lately... so overweight people are probably living longer due to medical advances