Experimental Design

**Controlled Experiment:** Researchers assign treatment and control groups and examine any resulting changes in the response variable. (cause-and-effect conclusion)

**Observational Study:** Researchers observe differences in the treatment and control groups and notice any related differences in the response variable. (association between variables)

How to Handle Confounding?

**Controlled Experiments:** Researchers randomly assign treatment and control groups so that possible confounding will “even out” across groups.

**Observational Study:** Researchers measure effects of “lurking variables” and determine if they have an impact on the response.

Experiments: Basic Principles

**Randomization:** to balance out lurking variables across treatment and control groups

**Placebo:** to control for the power of suggestion

**Control Group:** to understand changes not related to the treatment of interest

Double-Blind Experiments

If an experiment is conducted in such a way that neither the subjects nor the investigators working with them know whether the patient is receiving a treatment or was placed in the control group then the experiment is **double-blinded.**

- to control response bias (from respondent or experimenter)
Paired Comparison Designs

• **Matched Pairs** Design: Compares responses of paired subjects

• Technique:
  – choose *pairs* of subjects that are as closely matched as possible
  – randomly assign one subject to the treatment group and the control group to the other subject

• Sometimes a “pair” could be a single subject receiving both treatments. This is called *repeated measures* design.
  – randomize the order of the treatments for each subject
  – longitudinal by definition

Blocked Design

• A *block* is a group of individuals that are known before the experiment to be similar in some way that is expected to affect the response to the treatments.

• In a *block design*, the random assignment of individuals to treatments is carried out separately within each block.
  – a single subject could serve as a block if the subject receives each of the treatments (in random order)
  – matched pairs designs are block designs

Statistical Significance

• If an experiment finds a difference in two (or more) groups, is this difference really important?

• If the observed difference is larger than what would be expected just by chance, then it is labeled *statistically significant*.

• Rather than relying solely on the basis of statistical significance, also look at the actual results to determine if they are *practically important*.

Experimental Design

Scientists who study human growth use different measures of the size of an individual. Weight, height, and weight divided by height are three of the most common measures. If you were interested in studying the short-term effects of a digestive illness, which of these three variables would you study? Why?
Experimental Design

Height would be a rather silly variable to study for a short term digestive illness – weight and the weight-to-height ratio are more informative.

There are two ways the scientist could measure a weight change.
1. Difference:
2. Relative Percent Change:

Rule of Thumb: In this class, a Percent Change ≥ 5% is significant.

Experimental Design

Let’s discuss the conversation below from the point of view of establishing a valid conclusion of cause and effect.

Overheard at a coffee shop:

Person 1: “I’m convinced that eating cottage cheese makes people fat.”

Person 2: “What makes you say that?”

Person 1: “Have you looked at the people who eat it?”

Experimental Design

One study in the 1940’s found that by comparison with the general population, a high-percentage of delinquents are middle children – that is, neither the first-born nor last-born. This association remained even when race, religion, and family income were controlled for.

Bottom Line: Just because you’ve noticed an association between two variables doesn’t mean you can automatically conclude which direction causality goes.

Being a middle child, therefore, seems to be a contributing factor to delinquency.

OR IS IT?
SAMPLE SURVEYS

Next, we will consider the problem of sampling from a finite population. This is usually referred to as a survey. The goal of the survey is to learn about some parameters of a population, like averages or proportions. A well designed survey avoids incurring in systematic biases. The three most typical sources of bias are selection bias, response and non-response bias.

Collecting data: Sample Surveys

A population is a class of individuals that an investigator is interested in. Examples of populations are:

- All eligible voters in a presidential election.
- All potential consumers of a given product.
- The female elephant seals that mate at Año Nuevo State Reserve during the winter.
- The bottles of beer that are produced at a certain brewery.

A full examination of a population requires a CENSUS. Usually this is impractical. If only one part of the population is examined, then we are looking at a SAMPLE. The goal is to make INFERENCES from the sample to the whole population.

Q: Why was the Literary Digest so wrong?  
A: Because their poll was badly designed.

The Literary Digest poll had a strong bias against the poor, since they were unlikely to belong to clubs or have phones (in the '30s). The outcome of the election showed a split that followed a clear economic line: the poor voted for Roosevelt and the rich were with Landon.

The sampling procedure systematically tended to exclude one kind of person. This type of bias is called selection bias.

Usually, the more data, the less uncertainty in your results, however:

Taking a large number of samples with a biased procedure does not improve the results. This just repeats the basic mistake on a larger scale.
Collecting Data: Sample Surveys

Another source of bias in the Digest’s poll is that there was a large number of non-respondents. Only 2.4 million people bothered to reply, out of the 10 million who received the questionnaire. Studies have shown that people from the middle class are more likely to respond than people from the upper or the lower classes. So in a survey with a high non-response rate, middle class people may be over-represented.

These 2.4 million don’t even represent the 10 million people who were polled, let alone the population of all voters.

Non-respondents can be very different from respondents. When there is a high non-response rate, look out for non-response bias.

Quota Sampling

Consider the following scheme to obtain a sample. You send an interviewer to the field and ask him or her to get a fixed number of interviews within certain categories. For example:

- Interview 13 subjects
- Exactly 6 from the suburbs, 7 from the central city.
- Exactly 7 men and 6 women
- Of the men, 3 have to be under forty, 4 above forty.
- Of the men, 1 has to be black and 6 white.

The list of restrictions could go on. The goal is to achieve a sample that is fairly indicative of all demographic and social characteristics of the population to make it representative.

This is called a quota sampling scheme.

But, in the end, the interviewer has the freedom of deciding who gets interviewed, that is, the ultimate selection is left to human wisdom.

Gallup polls were conducted using the quota system for more than a decade, these are the results regarding the Republican vote:

<table>
<thead>
<tr>
<th>Year</th>
<th>Prediction</th>
<th>Results</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1936</td>
<td>44%</td>
<td>38%</td>
<td>+6%</td>
</tr>
<tr>
<td>1940</td>
<td>48%</td>
<td>45%</td>
<td>+3%</td>
</tr>
<tr>
<td>1944</td>
<td>48%</td>
<td>46%</td>
<td>+2%</td>
</tr>
<tr>
<td>1948</td>
<td>50%</td>
<td>45%</td>
<td>+5%</td>
</tr>
</tbody>
</table>

The sample sizes are around 50,000.

In the 1948 election, Gallup predicted the wrong winner.

Gallup had a systematic bias in favor of the Republican candidate in all elections from ’36 to ’48.

The reason for the bias is:

- The interviewers CHOSE who they interviewed! There could be an unintentional bias of the interviewers. The interviewers chose more Republicans to interview because they owned telephones and lived on nicer blocks. This is an example of Selection bias.
Collecting data: Sample Surveys

Two surveys are conducted to measure the effect of an advertising campaign for a certain brand of detergent. In the first survey, interviewers ask housewives whether they use that brand of detergent. In the second, the interviewers ask to see what detergent is being used.

Q: Would you expect the two surveys to reach similar conclusions? What type of bias is present and will the sample result be systematically above or below the true population result?

USING CHANCE

To eliminate the selection bias in a sample we use CHANCE in choosing the individuals to be included in the sample.

How does it work?

1. Set the size of the sample
2. Choose subject using chance
3. Delete subject from list and choose a second subject by chance
4. Continue process until we have a complete sample

This is called simple random sampling-(SRS). The subjects have been drawn at RANDOM WITHOUT REPLACEMENT. Using a sample based on chance eliminates selection bias.

A REAL POLL

A simple random sample can be difficult and costly when the population is large. For example, taking an SRS survey from people of voting age in America.

A better idea is to consider a sampling scheme that consists of multiple stages, each one subject to chance.

The Gallup poll after the 1948 is an example. The poll is taken as follows:

1. The Nation is split in 4 regions: W, NW, NE and S. All population centers of similar size are grouped together.
2. A random sample of the towns is selected. No interviews are conducted in the towns not in the sample.
3. Each town is divided in wards and the wards are subdivided into precincts.
4. Some wards are selected at random within the selected towns.
5. Some precincts are selected at random within the selected wards.
6. Some households are selected at random within the selected precincts.
7. Some members of the selected households are interviewed.

This is called a MULTISTAGE CLUSTER SAMPLING scheme.
The results

The following table presents the results of Gallup’s predictions for some elections from 1952 to 1992.

<table>
<thead>
<tr>
<th>Year</th>
<th>Sample size</th>
<th>Won</th>
<th>Prediction</th>
<th>Result</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1952</td>
<td>5,385</td>
<td>Eisenhower</td>
<td>51%</td>
<td>55.4%</td>
<td>4.4%</td>
</tr>
<tr>
<td>1960</td>
<td>8,015</td>
<td>Kennedy</td>
<td>51%</td>
<td>50.1%</td>
<td>.9%</td>
</tr>
<tr>
<td>1968</td>
<td>4,414</td>
<td>Nixon</td>
<td>43%</td>
<td>43.5%</td>
<td>.5%</td>
</tr>
<tr>
<td>1976</td>
<td>3,439</td>
<td>Carter</td>
<td>49.5%</td>
<td>51.1%</td>
<td>1.6%</td>
</tr>
<tr>
<td>1984</td>
<td>4,089</td>
<td>Reagan</td>
<td>59%</td>
<td>59.2%</td>
<td>.2%</td>
</tr>
<tr>
<td>1992</td>
<td>2,019</td>
<td>Clinton</td>
<td>49.0%</td>
<td>43.2%</td>
<td>5.8%</td>
</tr>
</tbody>
</table>

We observe a much smaller error (except for the 1992 election), no bias in favor of the Republican candidate and much smaller sample sizes.

Problems

Investigators doing polls have to face several problems that can bias the results of the survey even after considering a probabilistic sample.

Non-voters: Usually between 30% and 50% of the eligible voters don’t vote. But many of these are tempted to respond affirmatively when asked about their voting intentions. Interviewers ask indirect questions that allow to check if the person is genuinely a voter or not.

Undecided: Polls ask questions that give information about the political attitudes of the interviewed person in order to forecast the vote of undecided voters.

Response bias: Questions can be posed in a way that bias the response. A useful tool is to have the interviewed person deposit a ballot in a box.

Non-response bias: Non-respondents are different from the respondents. This is usually corrected by giving more weight to people who are difficult to get, since they, somehow, represent a subpopulation which is closer to the non-respondents.

Check data: Some subpopulations are likely to be overrepresented in the sample than others. This is usually corrected during the analysis of the sample using demographic data by weighting the subgroups accordingly.

Control: Interviewers are controlled either by direct supervision or by the cross-validation provided by redundant information in the survey.

Telephone surveys

Conducting a survey by phone saves money. It can also be done in less time.

How do you select sample? Phone numbers look like this

<table>
<thead>
<tr>
<th>Area code</th>
<th>Exchange</th>
<th>Bank</th>
<th>Digits</th>
</tr>
</thead>
<tbody>
<tr>
<td>415</td>
<td>767</td>
<td>26</td>
<td>76</td>
</tr>
</tbody>
</table>

The Gallup poll in ‘88 used a multistage cluster sample using area codes, exchanges, banks and digits as a hierarchy.

The Gallup poll in ‘92 was simpler and worked like this:

1. There are 4 time zones in the US. Each zone is divided in 3 types of areas: heavy, medium and lightly populated areas. This produced 12 STRATA.
2. They sampled numbers at random within each stratum.
Example Problem

Problem 1: A survey organization is planning to an opinion survey of 2,500 people of voting age in the U.S.. True or false and explain: the organization will choose people to interview by taking a simple random sample (SRS).

This is false. Taking a SRS survey of a population of about 200 million voters is impractical. First because a list of all the voters is not available. Second because taking a simple random sample of such list is a big problem in itself and third because interviewing 2,500 people all scattered around the map will be very costly.

Example Problem

Problem 2: A sample of Japanese-American residents in San Francisco is taken by considering the four most representative blocks in the Japanese area of the town and interviewing all the residents in those areas. However, a comparison with Census data shows that the sample did not include a high enough proportion of Japanese with college degrees. How can this be explained?

SELECTION BIAS: This was not a good way to draw the sample because you would expect that people living in the more traditional areas have very specific characteristics. In particular, it is likely that people with college degrees were living in more suburban neighborhoods.
A Source of Bias: Volunteers
Welsh coal mining town example.

Example Problem
A flour company wants to know what fraction of Minneapolis households bake their own bread. An SRS of 500 residential addresses is drawn and interviewers are sent to these addresses. The interviewers are employed during regular working hours on weekdays and they interview only during those hours.

1. What type of bias is present?
2. Are the interviewers more likely to under- or over-estimate the percentage of bread baking households in Minneapolis?

When considering the quality of a survey keep in mind three possible sources of bias:
- Selection bias
- Non-response bias
- Response bias
- Hawthorne Effect - People change their behavior when they know they’re being watched.
- Sample of Convenience - “first come, first served” sampling