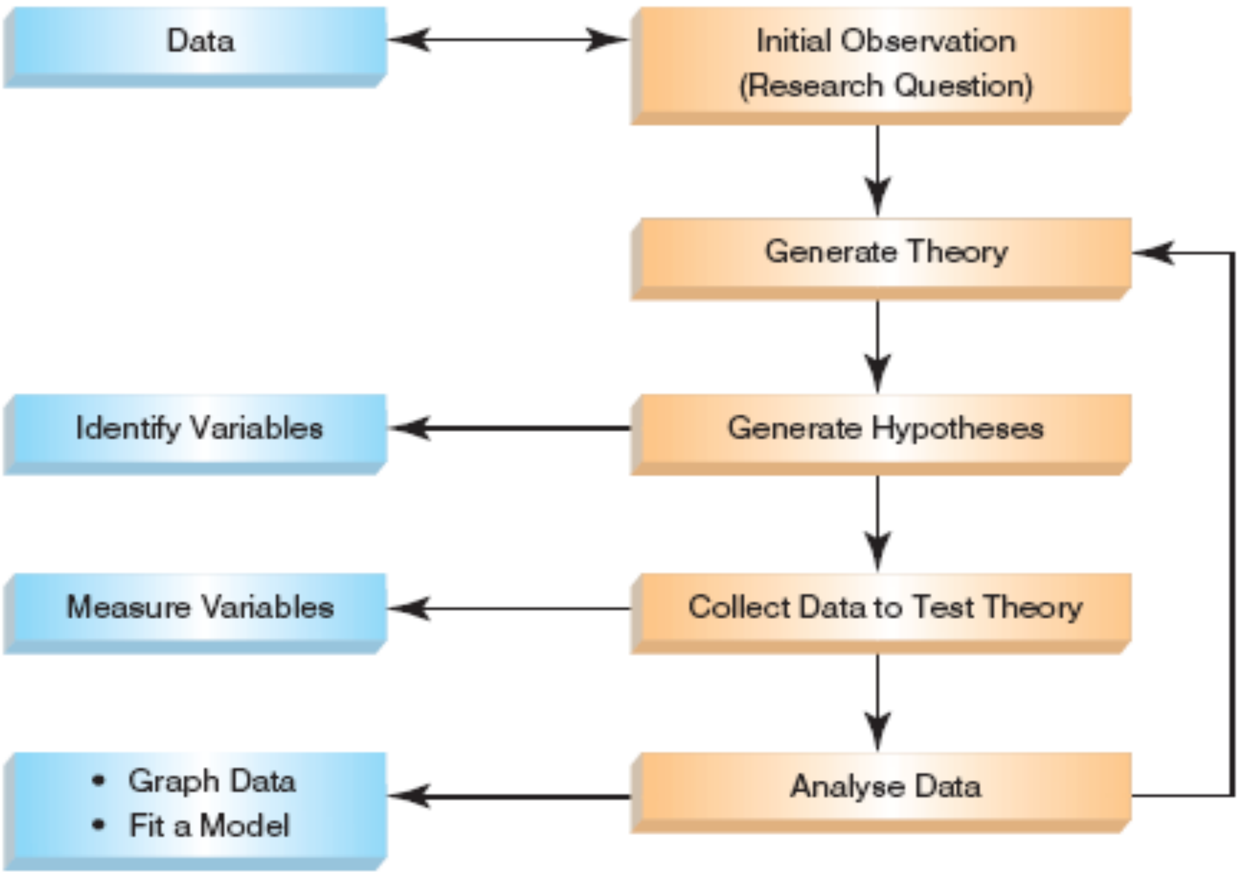
Types of Studies and Inferences

Lecture Plan

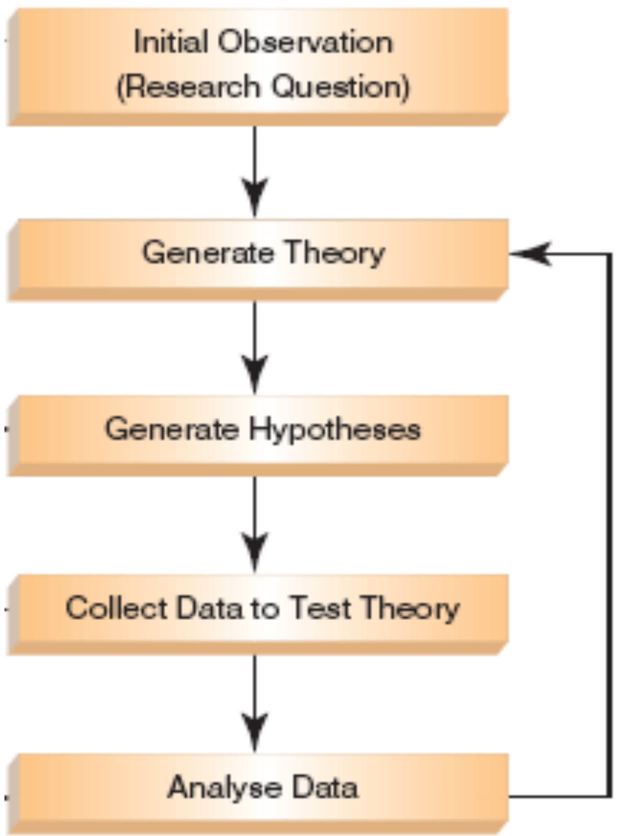
- The Research Process
- Testing a theory
  - Statistical Inference
  - Data collection: what to collect
    - Outcome vs. Predictors
  - Data collection: how to collect
    - Scope of Inference
    - Types of studies
  - Causal Inference

THE RESEARCH PROCESS

The Research Process

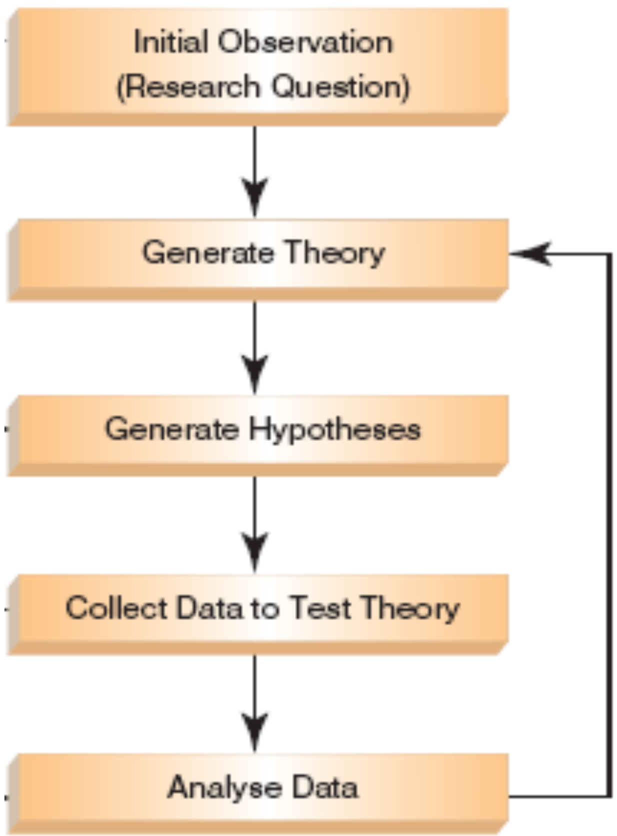


The Research Process



Graduate students groan when they find out that they must take my class.

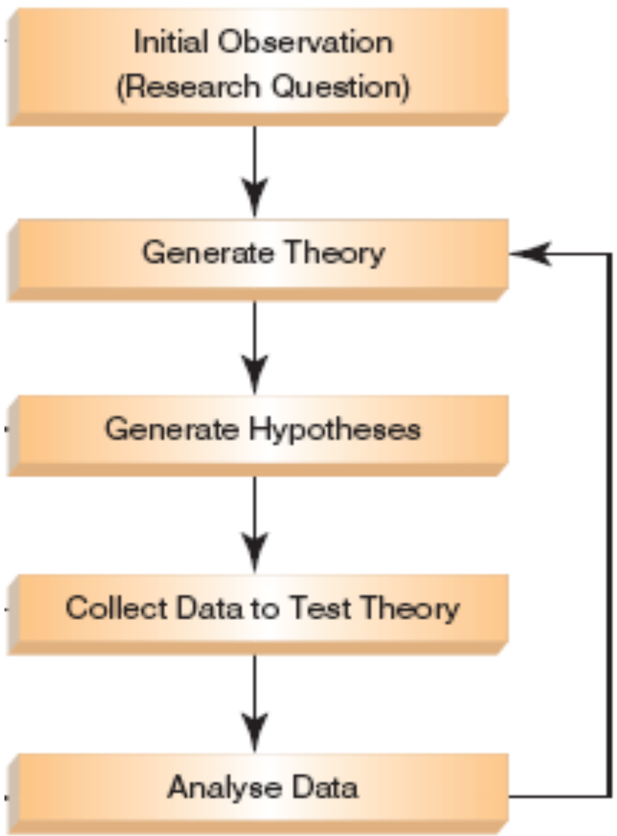
The Research Process



Graduate students groan when they find out that they must take my class.

Graduate students hate statistics

The Research Process



Graduate students groan when they find out that they must take my class.

Graduate students hate statistics

75% of graduate students hate statistics

TESTING A THEORY

Population and Samples

Researchers are typically interested in finding results that apply to an entire population of people or things.

- Population
  - The collection of units (be they people, plankton, plants, cities, suicidal authors, etc.) to which we want to generalize a set of findings or a statistical model.
- Sample
  - A smaller (but hopefully representative) collection of units from a population used to determine truths about that population.

Inference

- An <u>**inference**</u> is a conclusion that patterns in the data (sample) are present in a larger context (population).

<u>Observation</u> | <u>Inference</u>

Out of the 24 students in this class, 18 hate statistics → 75% of all graduate students hate statistics

Statistical Inference

- **<u>Statistical inference</u>** is an inference justified using statistical methods.
  - Statistical inference allows us to quantify the uncertainty in our conclusions.

<u>Observation</u>

Out of the 24 students in this class, 18 hate statistics

<u>Statistical Inference</u>

It is probable that at least 60% and less than 90% of all graduate students hate statistics

DATA COLLECTION: WHAT TO COLLECT

Data collection: what to collect

To test a theory/hypothesis we must collect data. Typically, hypotheses can be expressed in terms of two types of variables:

1. **Explanatory variables** (predictor, independent)
   – Variables that we believe explain or predict the behavior of another variable
2. **Response variables** (outcome, dependent)
   – Variables whose behavior can be explained (predicted) by the explanatory variable.
   – This variable is typically of primary interest.

Data collection: how to collect

Scope of Inference

The **scope of inference** is the group of individuals to whom the statistical conclusions can be extended.

- Inferences to populations can be only drawn from random sampling studies.
  - A **random sampling** study is when units are randomly selected from a well-defined population.
  - Random sampling typically ensures that all subpopulations are represented in the sample in roughly the same proportion as the population.
  - Our statistical procedures take into account that sometimes the sample may not be a very good mix of the population.
  - When subjects are not obtained through random sampling the results or model extend to the sampled group but not the larger population.

Simple Random Sample

- The most basic form of random sampling is **simple random sample**.

Each individual is chosen entirely by chance and each member of the population has an equal chance of being included in the sample.

Fertility Rate Example

Selection Biases in Pharmaceutical Sciences and Toxicology Studies

- Cell cultures

- Animal studies

- Clinical trials

Types of Studies: Randomized Experiments

- **Experiments** are studies in which we manipulate one or more explanatory variables (e.g., treatments) to see the effect they have on another variable.
  - In a **randomized experiment** the investigator uses a chance mechanism to assign experimental units to various treatment groups

What does randomization mean in bench experiments?

Types of Studies: Observational Studies

- **Observational studies** are studies in which the data are measured through observation of the world as it naturally occurs.
  - Grouping (i.e., explanatory variable) occurs naturally and is not assigned

Causal Inference

Causal inference is drawing a cause-and-effect relationship between an explanatory variable and a response variable.

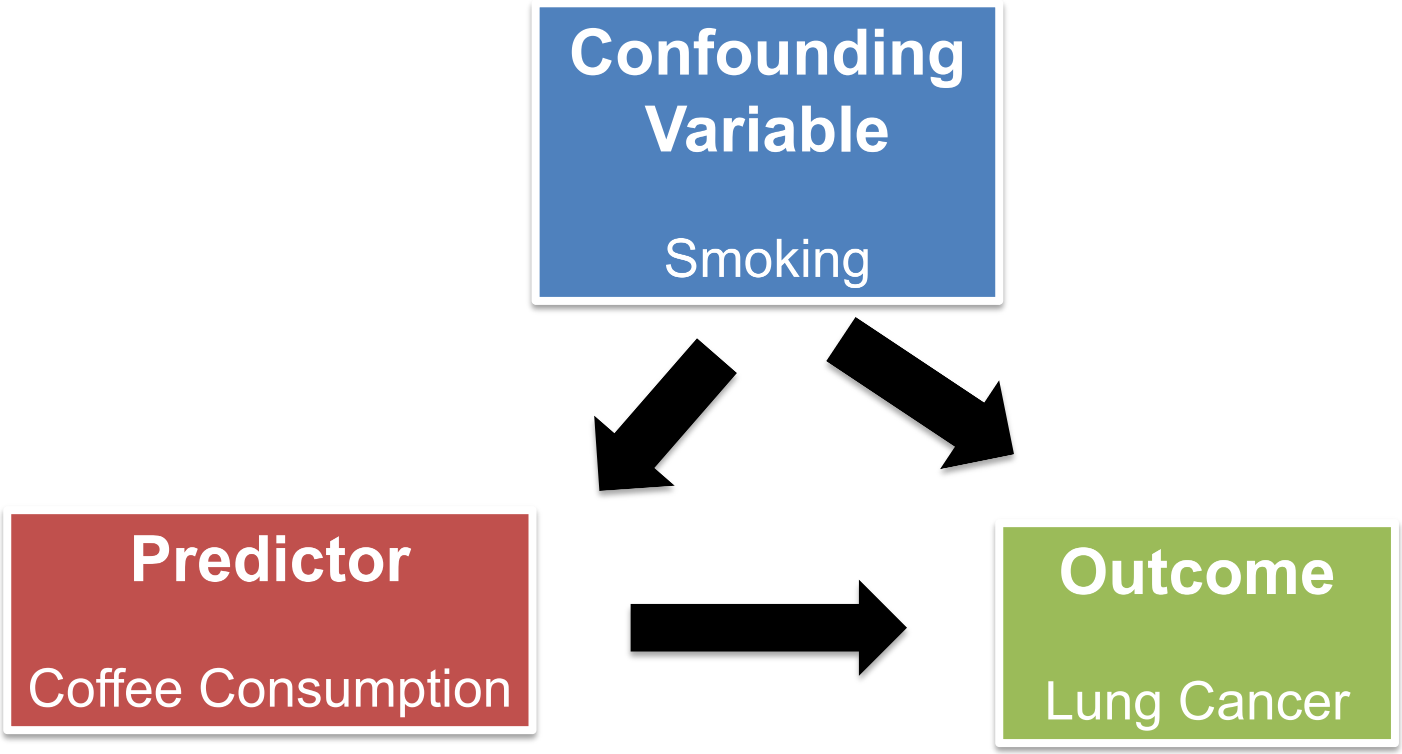
- e.g., putting my hand on a hot stove caused me to feel pain

Causal Inference in Observational Studies

- Causal inference is impossible in observational studies because confounding variables may cause the differences in the behavior of the response variable for different groups.
  - A **confounding variable** is a variable that explains the group a person is in (e.g., is related to the explanatory variable) and also the outcome of interest.

Confounding Variable

Confounding Variable



Causal Inference in Randomized Experiments

Causal inference <u>can</u> be made from randomized experiments but not from observational studies.

- Randomization ensures that subjects with different features (i.e., confounding variables) are mixed up evenly among the treatment groups.
- Randomized experiments seek to create groups that are similar with the exception of treatment status.
- The possibility that treatment groups may not end up being very "random" (i.e., groups are not mixed very well) is incorporated into the statistical tools used to express our uncertainty.

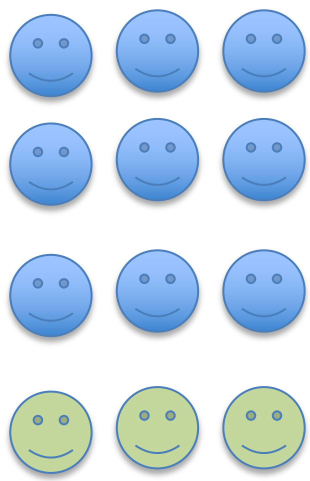
Smokers

2/3 will develop Lung Cancer

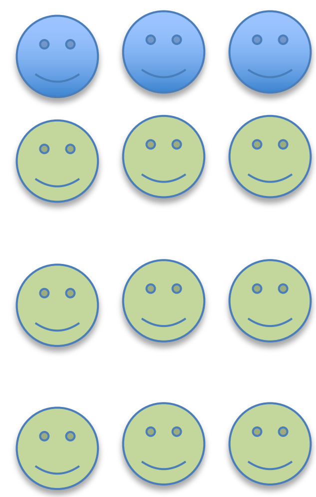
Non-smokers

1/3 will develop Lung Cancer

Coffee Drinkers



No Coffee



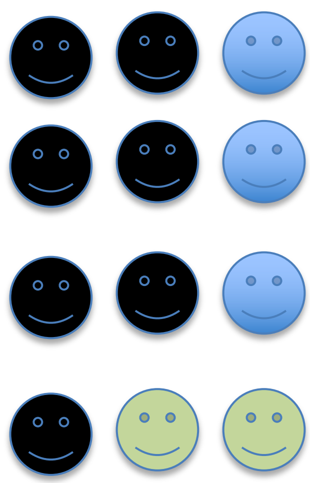
Smokers

2/3 will develop Lung Cancer

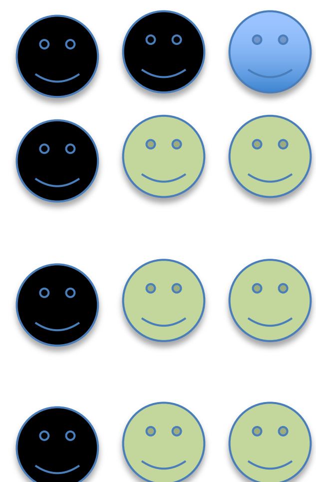
Non-smokers

1/3 will develop Lung Cancer

Coffee Drinkers



No Coffee



The NEW ENGLAND JOURNAL of MEDICINE

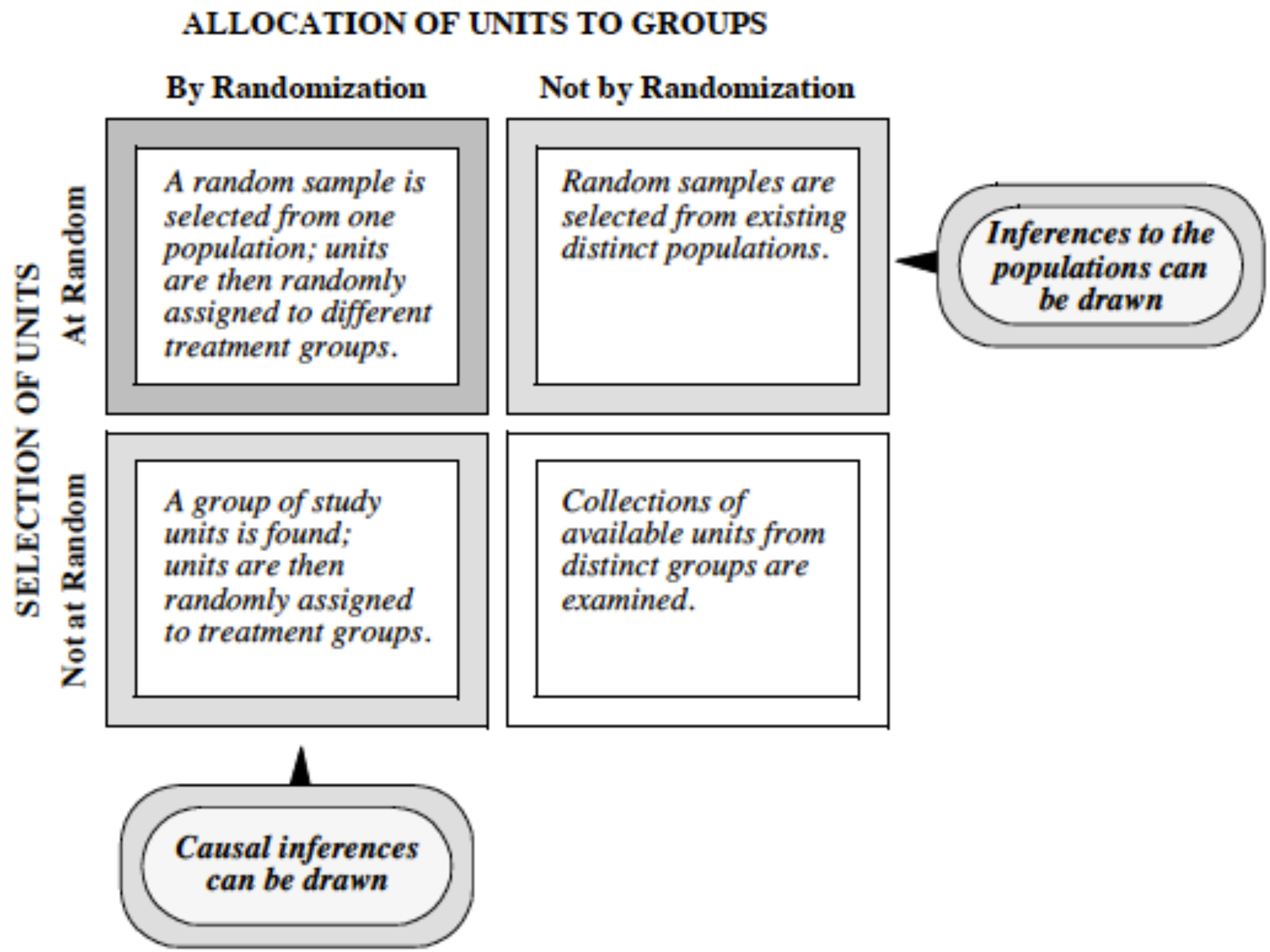
OCCASIONAL NOTES

Chocolate Consumption, Cognitive Function, and Nobel Laureates

Franz H. Messerli, M.D.

"Switzerland was the top performer in terms of both the number of Nobel laureates and chocolate consumption. The slope of the regression line allows us to estimate that it would take about 0.4 kg of chocolate per capita per year to increase the number of Nobel laureates in a given country by 1. For the United States, that would amount to 125 million kg per year. The minimally effective chocolate dose seems to hover around 2 kg per year, and the dose–response curve reveals no apparent ceiling on the number of Nobel laureates at the highest chocolate-dose level of 11 kg per year."

Statistical inferences permitted by study designs



What did we learn

- Population vs. Sample
- Statistical Inference
- Explanatory vs Response Variables
- Scope of Inference
- Simple Random Sample
- Observational vs. Randomized Experiments
- Causal Inference