Chapter 6
Compare Two Groups

Section 6.6:
Infer Causality

The School of Business
Portland State University

- Infer Causality
  - Relationships and Causality
  - Mean Difference and Causality
  - The Experiment
  - Design of Experiments
  - Applications

6.6a
Relationships and Causality

Relationship Between Variables

A relationship is an association, positive or negative

- **Relationship** of two variables: As the values of one variable increase, the values of the other variable tend to either systematically increase, or systematically decrease
- **Positive relationship**: As values of one variable increase, the values of the other variable tend to increase
  - Food quality increases, customer satisfaction increases
  - Occupancy rate increases, needed staff increases
- **Negative (inverse) relationship**: As values of one variable increase, the values of the other variable tend to decrease
  - Price decreases, sales volume increases
  - Time brushing teeth increases, cavities decrease
- An association between two variables is also called a correlation, assessed numerically with what is called a correlation coefficient, introduced in a later chapter

Causal Relationship

Causality results in correlation

- These previous examples are of a specific type of relationship called a causal relationship
- **Causal Relationship**: A change in the value of one variable, X, directly results in a subsequent change in the value of another variable, Y
- The variable Y is called the outcome variable, or dependent variable, or response variable
- To indicate that X causes Y: $X \rightarrow Y$
  - Price → Sales Volume
  - Food Quality → Customer Satisfaction
  - Quality of Employee Instructions → Quality of Output
- The *result* of the casual impact of X on Y is association, that is, correlation, between X and Y

Correlation Does Not Imply Causality

Correlation is not necessarily the result of direct causation

- To assert that variables X and Y are related does not imply a causal relationship unless additional information is provided
- **Key Concept**: Correlation does not imply causality
- **Spurious Correlation**: Relation of two variables due to a common cause
- **Confounding Variable**: A variable that causes both variables X and Y so that the relation between them is at least partly spurious
- The goal of much research is to demonstrate causality, X → Y, as the primary explanation for a relationship between variables, ruling out potential confounding variables as an explanation

Example of Spurious Correlation

Hypothetical but intuitive

- Measure amount of Damage in $ done by a fire (D) and the Number of Fire Trucks at a fire (F)
- The relationship is real: Fires that result in more damage tend to be serviced by more fire trucks
- Yet no one would claim that the fire trucks *cause* the damage, $F \rightarrow D$, or, conversely, that more damage *causes* more fire trucks, $D \rightarrow F$
- Instead, both variables, D and F, share a common cause, a confounding variable, the Severity of the Fire (S), so their correlation is spurious: $D \leftarrow S \rightarrow F$
- Two variables can be correlated without one causing the other

Passive Observation

Observe relationships without specifying causality

- **Passive observational study**: A study in which the researcher passively observes reality without actively influencing the environment experienced by the respondents
- By definition, all measurement results from observation, but the reference here is to *passive* observation
- Sometimes this type of study is called a correlational study, but this term confuses a type of statistical analysis with the method by which the data are collected
- **Key Concept**: Data from passive observational studies demonstrate correlations without providing the direct basis of causal explanations

6.6b
Mean Difference and Causality

Causal Variable with Few Number of Levels

Corresponding to each level is a group of respondents

- Consider an X variable, a potential causal influence on response variable Y, sampled at a relatively small number of levels, usually 2, 3 or 4 levels
  - A categorical variable naturally provides a discrete set of levels, such as Male and Female for Gender
  - A small number of levels can be sampled from the many possibilities provided by a continuous variable, such as 0ml, 1ml and 2ml drug dosages
- If the X variable has 2 levels, such as Sales according to Red and Yellow package colors, then the primary result is a mean difference of the two groups on the response variable Y
- **Key Concept**: A mean difference is an expression of the relationship between two variables, Grouping variable X and Response variable Y

Two Primary Questions Regarding a Mean Difference

Step 1: Does a difference (association) exist between groups?

- The question of a mean difference is a statistical issue
- Detect a difference between means with statistical inference: a confidence interval and/or hypothesis test of the mean difference

Step 2: Does group membership cause a detected difference?

- Detecting a difference is distinct from delineating the *cause* of the difference, in this case $X \to Y$
- Detecting a difference does *not* imply that being a member of one group or the other caused the difference
- Causality is best established as a methodological issue, specifically, how group membership was assigned

Assess a Causal Outcome According to a Mean Difference

A group of respondents for each level of the treatment variable

- **Equivalent groups**: Two or more groups of respondents that are equivalent in the population on all variables
- Three conditions must be met to show X → Y
  - Obtain equivalent groups, by eliminating or controlling other potential confounding variables, a variable other than X that causes Y, leaving X as the relevant variable
  - Expose the respondents in each group to the same level of the treatment variable, with each group is exposed to a different level
  - Therefore observed differences between the groups are likely due to differences of the level of the treatment variable
- The methodology for achieving these conditions is the experiment, the next topic

6.6c
The Experiment

Investigating Causality: The Experiment

Understanding causes as a way to understand reality

- **Experiment**: A procedure for data collection in which the values of one or more variables are explicitly manipulated so as to study the resulting effects on another variable
- **Treatment Variable**: The categorical X variable, or potential cause of which the different levels are manipulated
- **Treatment Level**: One of usually a small number of possible values of the treatment variable, each value corresponds to a group
- **Respondent**: The customer, employee, company or other unit whose environment is manipulated by assignment to a specific group that corresponds to a level of the treatment variable
- **Response Variable**: The Y variable or potential effect, the variable whose values have changed for the respondents in response to the change in the values of the treatment variable

Examples of Experiments

What accounts for improved customer satisfaction and sales?

- Which of two menu designs, if either, leads to more sales?
  - Treatment variable: Menu design, with two levels, i.e., two designs
  - Respondents: Customers at the restaurant, each experiences one treatment level, i.e., one of the two menu designs
  - Response variable: Sales
- Which of two package colors, if either, leads to more sales?
  - Treatment variable: Package color, with two levels, Red and Blue
  - Respondents: Customers shopping at a specific store stocked with either the Red or the Blue packaging
  - Response variable: Sales

Assess Causality with Experiments – Treatment Variables

Treatment is the potential cause

- Other terms for the treatment variable: X, independent variable, predictor variable, explanatory variable
- Identify different variables that affect favorable business outcomes, which directly or indirectly lead to improved sales
  - Price levels
  - Different ads
  - Package appearance: Size, color, shape
  - Selling technique
  - Product placement
  - In-store advertising

Assess Causality with Experiments – Response Variables

Response variable is the potential effect

- Other terms for the response variable: Y, dependent variable, outcome variable, criterion variable
- Examples include . . .
  - Sales
  - Market share
  - Purchase intention
  - Product awareness
  - Perceived quality

Source of Group Membership

Group membership for treatment levels set one of two ways

- Assigned by the researcher at beginning of the study
  - B2B Orders: Assigned to Supplier 1 or Supplier 2
  - Employees: Assigned to Training Method 1 or 2
- Pre-existing, that is, naturally existing before the study begins
  - Gender: Men or Women
  - MBA Program: Students at School 1 or School 2

Randomization: Assignment to Groups

Crucial aspect of a true experiment

- **Randomization**: Random assignment of people (or whatever) to groups
- Flip a coin to determine which of two suppliers gets the order
- Use the computer, such as the R `sample` function, to assign the next employee to either Training Session 1 or Training Session 2

Randomization is not random sampling

- First randomly sample the people (or whatever)
- Then randomly assign the sampled people to groups

Randomization the Key to Attaining Equivalent Groups

Randomization is the great equalizer

- To facilitate a causal inference that any detected population mean difference is attributable to group membership, randomly assign group membership
- **Key Concept**: Randomization by itself assures that *no* population average differences between the two groups exist on any variable, that is, the groups are equivalent
- After performing the study, any statistically significant difference between randomly assigned groups is then attributable only to group membership

Experimental Control

Eliminate confounding variables

- Goal is to isolate a potential cause of the response variable
- Respondents in different groups are presented with an identical environment that differs only the in the levels of the variable(s) studied as a potential cause for response Y
- For example in a study of the effect of package color on sales, the only distinction between the two types of packages presented to two groups of customers is package color
- If the packages also differed by size, or labeling, or shelf placement, these other variables are potential confounders
- If a difference is detected between the two groups that experience environments that differ on more than one feature, attribution of the difference to package color could not be properly accomplished

Spurious Relationships Exist with Pre-Existing Groups

Alternative explanations usually exist

- An entire slew of known and unknown ancillary differences characterize pre-existing groups
  - How do men and women differ before the study begins?
  - For example, the proportion of those with a doctorate in mathematics who are male is higher than for those who have a doctorate in education
- A mean difference between pre-existing groups on some response variable may be due to either the corresponding value of the treatment variable, or a spurious relationship

Experimental Manipulation

Another key consideration for inferring causality

- The experimenter deliberately and systematically creates an environmental change for respondents
- **Manipulation**: The deliberate exposure of different groups of respondents to different levels of the treatment (independent) variable
- Each treatment level corresponds to a different group of respondents
- **Experimental study**: A research study in which the researcher randomly assigning participants to groups and then deliberately exposes each group to a different level of environmental variables
- Ex: Some customers of a laundry detergent are presented with a yellow package and customers in the other group are presented a red package

Double Blind: Eliminate Unwanted Effects of Expectations

Both participants and evaluator can expect certain outcomes

- Expectations can affect behavior, so it is preferable that . . .
  - Respondents do not know the treatment to which they are assigned
  - Those administering the study do not know the treatment to which each respondent has been assigned
- **Double blind experiment**: Respondents and evaluators are unaware of the treatment group to which the respondent has been assigned
- Example is in the study of the effectiveness of a drug, participants in the control group are administrated a **placebo**, an inert substance administered such that neither the recipient nor the administrator can determine if the drug is real or not

6.6d
Design of Experiments

Experimental Design

Simplest design

- **Experimental Design**: Sampling and randomization plan that organizes participants into groups according to one or more treatment variable(s)
- Basic Question: What is the causal impact of the treatment variable on the response variable?
- Simplest design: two levels of one treatment variable
  - Could be a control group that receives no manipulation, and an experimental group
  - Could be two experimental groups, which receive different levels of the treatment
- Analyze the difference of the two group means statistically with a *t*-test

Experimental Design

More advanced designs

- Can move beyond the simplest design for more sophisticated causal analysis
  - Can be more than one treatment variable, often two or three
  - Each treatment variable can have two or more levels
- Goal remains the same, to investigate the causal impacts of the treatment variables on the response variable
- Analyze the differences of the resulting group means statistically with Analysis of Variance or ANOVA
- For the analysis of a single mean difference, either a *t*-test or an ANOVA provide equivalent results

Validity of an Experiment

To what extent are any conclusions of causality true?

- In the context of an experiment a confounding variable is an alternative explanation of the changes in Y other than from changes in X
- **Internal Validity**: The extent to which X is the actual cause of Y in the experiment
- A study is internally valid when confounding variables have been eliminated, best assured by random sampling and randomization
- **External Validity**: The extent to which X is the actual cause of Y in the world beyond the experiment, the target population, and not just a laboratory artifact
- **Construct Validity**: The extent to which X and Y are properly identified and measured such that the names of the variables reflect the actual content of the measurements

How Much Control Exists in a Study?

True experiment: Most control of any study

- Randomization [R] of respondents into a control group (CG) and at least one experimental group (EG), or two or more experimental groups, to compare with each other
- Notation: X represents exposure to the experimental level, such as the new package color Red, and O represents an observation, a measurement of some type, such as Sales
- Classic example is Postest-Only Control Group

  (EG): [R] X $\rightarrow O_1$
  (CG): [R] O_2
- Usually applied in a controlled, laboratory situation, with maximum internal validity
- Unfortunately, laboratory settings can be artificial to some extent, reducing external validity

Randomization Not Always Possible

Try for the next best thing, matching

- Depending on the situation, respondents cannot always be randomized into groups
  - Impossible: Randomly assign respondents to a level of Gender, Male or Female
  - Unethical: Randomly assign pregnant women to consume different amounts of alcohol
- Even without randomization, obtaining equivalent groups can at least be attempted
- The key is to match respondents in the different groups on potentially confounding variables
  - Put the blue package in one store and the red package in another store, but choose the stores so that all the shoppers are from comparable income levels

How Much Control Exists in a Study?

Quasi-Experiment: Intermediate level of control

- Resembles a true experiment, but attempts to obtain equivalent groups without randomization, so the groups are as similar as possible, such as shoppers from similar stores
- Example is Nonequivalent Control Group

  (EG): $O_1 \rightarrow X \rightarrow O_2$
  (CG): $O_3 \qquad\qquad O_4$
- O_1 and O_3 are pre-experimental observations meant to assess some form of equivalence between the two groups in the absence of the preferred randomization
- Usually applied in a field study, such as a set of actual stores with real shoppers where randomization is not practical
- Ideally, successful field studies maximize external validity, though the lack of randomization can diminish internal validity

Threats to Internal Validity without Randomization

Consider the usual test market analysis

- **Test Market**: A real shopping experience for customers who, usually unknowingly, are participating in a evaluation of product/service characteristics systematically varied across different retail outlets and/or locations
- Stores in the different experimental groups can be matched for similar characteristics such as income and/or ethnic composition, without random assignment of shoppers
- Unfortunately, in the absence of randomization, the experimenter cannot control potential confounding influences, so that, for example, in just one of the test markets
  - New competing products could be introduced
  - A competitor could introduce a new advertising campaign
  - A competitor could change prices or even go out of business
  - The local government could change taxes

How Much Control Exists in a Study?

Pre-Experiment: Least amount of control

- In the pre-experiment, there is no randomization and no assessment of equivalence, what is called a static group comparison

  (EG): $X \rightarrow O_1$
  (CG): $\qquad O_2$
- Selection bias a major problem because the equivalence of the groups is unknown, but any difference that might be found could be due to one or more confounding variables
- Ex: Compare those who use a product vs those who do not use the product to assess if the groups differ on some variable, but no assessment of how the groups may have differed before choosing the respective products
- Problem: Very low internal validity

6.6e
Applications

APPLICATION: Assignment of Orders to Suppliers

Randomization eliminates confounding variables

- Variables
  - Grouping variable (X): Supplier
  - Response variable (Y): Ship time of order
- To attribute any detected difference in ship time to supplier, randomly assign orders to each supplier
- Randomization assures equivalent groups, that no population average differences between the two groups exist on any variable except for choice of supplier
- Randomization rules out potential confounding variables . . .
  - Weight of shipments
  - Dollar value of shipments
  - Day of week orders were placed

APPLICATION: Spurious Correlation

Epidemiology: Observational result

- Many *observational* studies demonstrated that women taking hormone replacement therapy (HRT) also had a lower than average incidence of coronary heart disease (CHD)
- As an aggregate, the observational studies demonstrated a well-established, genuine inverse relationship between HRT and CHD: as HRT goes up, CHD tends to go down
- To *explain* this true relationship, a casual explanation was offered: HRT protects against CHD[1]
- However, an observation of a relationship does *not* imply a direct causal relation

[1] Lawlor DA, Davey Smith G, Ebrahim S (June 2004). "Commentary: the hormone replacement-coronary heart disease conundrum: is this the death of observational epidemiology?". Int J Epidemiol 33 (3): 464-7

Presumed Causal Structure

Epidemiology: Causality inferred from observed correlation

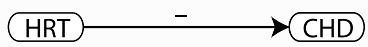

Figure: Hypothesized negative causal impact of HRT on CHD

Example of Spurious Correlation

Epidemiology: Experimental outcome

- On the contrary, subsequent *experimental research* demonstrated that HRT causes a small, but detectable, *increase* in risk of CHD
- How could the causal relationship be the *opposite* of the observed correlation?
- Women undertaking HRT were more likely to be from higher socio-economic groups, a confounding variable, with better diets and exercise
- The use of HRT and decrease in CHD were joint effects of a common cause, the benefits that result from a higher socioeconomic status (SES), rather than cause and effect

Example of Spurious Correlation

Epidemiology: Experimental outcome (continued)

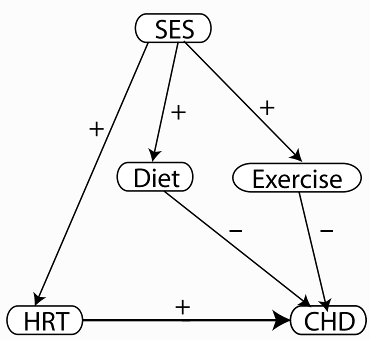

Figure: The correlation of HRT with CHD is negative, even though the causal impact of HRT on CHD is positive, because of the spurious relation with SES and its stronger intermediate causal paths

- The End