Hypothesis Testing, Decisions
Prof. Chris Clifton
28 January 2020

Materials adapted from Profs. Jennifer Neville and Dan Goldwasser
What is a hypothesis?

- **Hypotheses** are tentative statements of the expected relationships between two or more variables
  - **Inductive** hypotheses are formed through inductively reasoning from many specific observations to tentative explanations (bottom-up)
  - **Deductive** hypotheses are formed through deductively reasoning implications of theory (top-down)

- Reasons for using hypotheses
  - Provides focus and directs research investigation
  - Allows the investigator to confirm or not confirm relationships
  - Provides a useful framework for organizing and summarizing results and conclusions

Types of hypotheses

**Broad categories**
- **Descriptive**: propositions that describe a characteristic of an object
- **Relational**: propositions that describe the relationship between 2+ variables
- **Causal**: propositions that describe the effect of one variable on another

**Specific characteristics**
- **Non-directional**: an differential outcome is anticipated but the specific nature of it is not known (e.g., the tuning parameter will affect algorithm performance)
- **Directional**: a specific outcome is anticipated (e.g., the use of pruning will increase accuracy of models compared to no pruning)
Ever since 1980, when Ronald Reagan inspired more men than women, the difference in the way men and women vote has been a significant part of American politics.

**Step 1:** Express data as random variables

\[ X := \text{gender} \]
\[ Y := \% \text{voted Democrat} \]

**Step 2:** Restate claim as a hypothesis about the relationship between the random variables, e.g.,

- (X=male) is associated with smaller Y

**Step 3:** Determine type of hypothesis (and consider whether you can make it stronger), e.g.,

- Directional-relational

Types of hypotheses:

- **Descriptive:** Voting practices vary throughout the population (i.e., \( Y \) varies).
- **Non-directional relational:** Voting behavior varies based on gender (i.e., \( X \) and \( Y \) are associated)
- **Directional-relational:** Women vote more for democrats (i.e., \( X=\text{female} \) is associated with larger \( Y \))
- **Causal-relational:** Women vote more for democrats because they are more likely to be poor (i.e., \( X=\text{female} \) is associated with larger \( Y \), but if you control for salaries this effect may disappear)
Example: Video games and violent behavior

- Playing video games can lead to violent acts in real life

- Video games aren’t the reason why shootings happen
  - [https://www.forbes.com/sites/erikkain/2012/04/19/as-video-game-sales-climb-year-over-year-violent-crime-continues-to-fall/#6a8eb77d4507](https://www.forbes.com/sites/erikkain/2012/04/19/as-video-game-sales-climb-year-over-year-violent-crime-continues-to-fall/#6a8eb77d4507)

- Define the random variables:
  - Descriptive:
  - Non-directional relational:
  - Directional-relational:
  - Causal-relational:
Are A and B the same color?

Issue: “biases” in the human visual system
Heuristics and biases

• Tversky & Kahneman, psychologists, propose that people often do not follow rules of probability when making decisions
• Instead, decision making may be based on heuristics
  – Lowers cognitive load but may lead to systematic errors and biases

• Examples:
  – Availability heuristic
  – Representativeness heuristic
  – Confirmation bias
  – Conjunction fallacy (we will not cover this)
  – Numerosity heuristic (we will not cover this)

Neglecting base rates

• Taxi-cab problem (Tversky & Kahneman ‘72)
  – 85% of the cabs are Green
  – 15% of the cabs are Blue
  – An accident eyewitness reports a Blue cab
  – But she is wrong 20% of the time.
• What is the probability that the cab is Blue?
  – Participants tend to overestimate probability, most answer 80%
  – They ignore baseline prior probability of blue cabs.
More on neglecting base rates

A priori (beforehand)

\[ P(\text{green}) = 0.85 \]
\[ P(\text{blue}) = 0.15 \]

\[ P(\text{see Blue}|\text{blue}) = 0.80 \]
\[ P(\text{see Blue}|\text{green}) = 0.20 \]

More on neglecting base rates

After accident (only cars reported as being blue)

85%  
15%

20% of 85%  
80% of 15%
More on neglecting base rates

• How to compute probability

\[ P(\text{blue}|\text{seeBlue}) = \frac{P(\text{blue} \land \text{seeBlue})}{P(\text{seeBlue})} = \frac{P(\text{seeBlue}|\text{blue})P(\text{blue})}{P(\text{seeBlue})} = \frac{P(\text{seeBlue}|\text{blue})P(\text{blue})}{P(\text{seeBlue}|\text{blue})P(\text{blue}) + P(\text{seeBlue}|\text{green})P(\text{green})} \]

\[ = \frac{0.80 \cdot 0.15}{(0.80 \cdot 0.15) + (0.20 \cdot 0.85)} = 0.41 \]

Most people answered 80%
Arthritis study (Redelmeier & Tversky '96)

- **Common belief:**
  - Arthritis pain is associated with changes in weather
- **Experiment:**
  - Followed 18 arthritis patients for 15 months
  - 2 x per month assessed: (1) pain and joint tenderness, and (2) weather
- **Results:**
  - No correlation between pain/tenderness and weather
  - Patients saw correlation that did not exist... why?

Arthritis study (cont)

- Patients noticed when bad weather and pain co-occurred, but failed to notice when they didn’t.
  - Better memory for times that bad weather and pain co-occurred.
  - Worse memory for times when bad weather and pain did not co-occur
- **Confirmation bias:** People often seek information that confirms rather than disconfirms their original hypothesis
Estimating probabilities (Tversky & Kahneman '73/’74)

• Question: Is the letter R more likely to be the 1st or 3rd letter in English words?

• Results: Most said R more probable as 1st letter

• Reality: R appears much more often as the 3rd letter, but it’s easier to think of words where R is the 1st letter

Estimating probabilities (cont)

• Question: Which causes more deaths in developed countries? (a) traffic accidents or (b) stomach cancer

• Typical guess: traffic accident = 4X stomach cancer

• Actual: 45,000 traffic, 95,000 stomach cancer deaths in US

• Ratio of newspaper reports on each subject: 137 (traffic fatality) to 1 (stomach cancer death)

• Availability heuristic: Tendency for people to make judgments of frequency on basis of how easily examples come to mind
Base Rate Study *(Kahneman & Tversky '73)*

- Participants told that for a set of 100 people are either:
  - 30% engineers/70% lawyers, or
  - 70% engineers/30% lawyers
- *Given*: A description of a person Jack, which is representative of a prototypical engineer (e.g., likes carpentry and mathematical puzzles, careful, conservative)
- *Question*: Is Jack more likely to be a lawyer or engineer?
- *Results*: Participants in the 30% condition judged Jack just as likely to be an engineer as participants in the 70% condition.

**Base rate study (cont)**

- People use the representative heuristic to make inferences...
  - Inferences is based solely on similarity of target to category members
  - Base rates (70%-30%) are ignored
- ...rather than using formal statistical rules to make inferences
  - Inferences should be based on similarity of target to category members AND base rates (70%-30%)
- **Representative heuristic**: categorizations made on the basis of similarity between instance and category members
Gambler’s fallacy

- **Gambler’s fallacy**: belief that if deviations from expected behavior are observed in repeated independent trials, then future deviations in the opposite direction are then more likely.
- T&K: this is an example of the **representativeness heuristic**—where the probability of an event is judged by its similarity to the population from which sample is drawn.
- The sequence “H T H T H” is seen as more representative of a prototypical coin sequence. Why?
  - When people are asked to make up random sequences, they tend to make the proportion of H and T closer to 50% than would be expected by random chance.
  - T&K interpretation: people believe that short sequences should be representative of longer ones.

Interpretation of these findings

- People do not use proper statistical/probabilistic reasoning... instead people use heuristics which can **bias** decisions.
- Heuristics can often be very effective (and efficient) for social inferences and decision-making.
  - E.g., the book “Simple Heuristics That Make Us Smart” summarizes research by Gigerenzer and Todd.
- ... but be aware that **heuristics can bias** results from exploratory data analysis and other modeling efforts.
1000+ participants were asked about their political views and also asked a series of questions to gauge their mathematical reasoning ability.

Participants were then asked to solve a fairly difficult problem that involved interpreting the results of a (fake) scientific study.

One group was given a problem involving the effectiveness of a new skin cream. The other group was given a mathematically similar problem, but the data involved the effectiveness of a gun control measure.

What was the result? Highly numerate people were more susceptible to letting politics skew their reasoning than were those with less mathematical ability.

Does this look familiar?

- Economist Dan Arieli showed that we are willing to pay more for a product when something free is involved
  - "we'll throw in X for free, if you buy Y"
- Advertisers make sophisticated use of these biases!
Relevant TED talks

- Informal, but interesting!

- Daniel Kahneman: The riddle of experience vs. memory
  - https://www.ted.com/talks/daniel_kahneman_the_riddle_of_experience_vs_memory

- Dan Ariely: Are we in control of our own decisions?
  - https://www.ted.com/talks/dan_ariely_asks_are_we_in_control_of_our_own_decisions