Applications of Statistical Inference

Ryan Miller

Introduction

- Statistics is an application driven discipline
  - We study statistics as a tool to better understand the world around us
- My goal for this presentation is to review and connect several statistical concepts
  - Measuring associations between variables
  - Understanding possible reasons for these associations
  - Recognizing where *statistical testing* fits in this process

Application - Surgical Site Infections

- Surgical site infections (SSIs) are the most common and most costly of all hospital-acquired infections
  - 2-5% of all surgical patients
  - $3.5 billion to $10 billion in annual costs

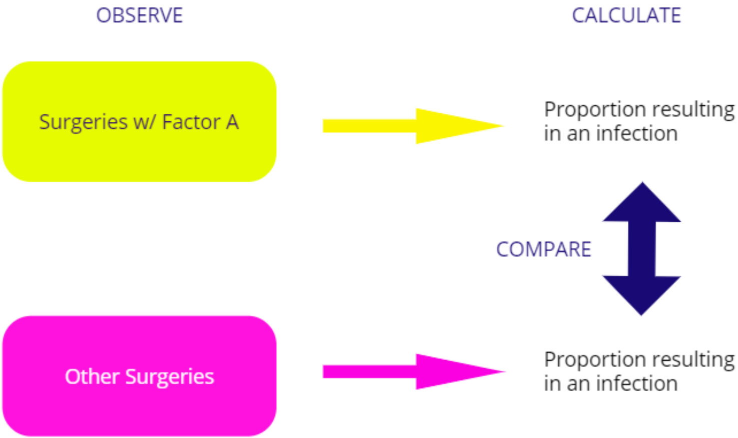
Application - Surgical Site Infections

- Surgical site infections (SSIs) are the most common and most costly of all hospital-acquired infections
  - 2-5% of all surgical patients
  - $3.5 billion to $10 billion in annual costs
- SSIs are *much less common* today than in the past, thanks (in part) to statistics!

Quantifying an Association

- Suppose you work at the Glasgow Royal Infirmary (a large teaching hospital in Scotland)
  - From your experiences, you suspect surgeries involving “Factor A” have an *increased risk* of surgical site infection (relative to surgeries without Factor A)
  - What are the explanatory and response variables here? How might you *measure* (quantify) their relationship?

Quantifying an Association



Explaining an Association

- Now suppose we collect data and observe *higher proportion* of surgeries involving “Factor A” resulted in an SSI (compared to surgeries where “Factor A” was not present)
- What might *explain* the increased rate of SSI seen in *these data*?

Explaining an Association

- Now suppose we collect data and observe *higher proportion* of surgeries involving "Factor A" resulted in an SSI (compared to surgeries where "Factor A" was not present)
- What might *explain* the increased rate of SSI seen in *these data*?

1. Bias (ie: SSI was measured differently in the "Factor A" surgeries)
2. Confounding (ie: A third variable causes both "Factor A" and increased risk of SSI)
3. Random chance (ie: These data aren't all of the data, there is randomness in who was observed)
4. Factor A *causes* an increased risk of SSI

The Role of Statistics

- Statistical thinking centers around understanding these possible explanations
  - Ideally, we'd like to rule out explanations #1-3, thereby establishing that Factor A causes SSI
  - What are some strategies for ruling out each of aforementioned factors?

The Role of Statistics

- Statistical thinking centers around understanding these possible explanations
  - Ideally, we'd like to rule out explanations #1-3, thereby establishing that Factor A causes SSI
  - What are some strategies for ruling out each of aforementioned factors?

1. Bias - Representative sampling, placebos, blinding, etc.
2. Confounding - Randomized experiments
3. Random chance - Statistical inference

Some History on SSIs

- In the 1860s surgical site infections were extremely common and often resulted in death
- Most experts believed exposure to "bad air" was responsible for these infections
    - Many hospital wards opened their windows at midday to air out

Some History on SSIs

- In the 1860s surgical site infections were extremely common and often resulted in death
- Most experts believed exposure to “bad air” was responsible for these infections
  - Many hospital wards opened their windows at midday to air out
- During this time, it customary for surgeons to move from patient to patient without taking any sanitary precautions
  - In fact, many took pride in the accumulated stains on their gowns as a measure of experience

Louis Pasteur

- In 1862, Louis Pasteur showed that food spoilage was caused by the growth of harmful micro-organisms
  - This discovery led Pasteur to the idea that ingesting these micro-organisms caused human illnesses
- Pasteur identified three methods for eliminating these micro-organisms: heat, filtration, and chemical solutions
  - The method of heating became known as pasteurization and is widely applied to milk and beer

Joseph Lister

- In the mid 1860's, Joseph Lister, a Professor of Surgery at Glasgow Royal Infirmary, became aware of Louis Pasteur's work
  - Lister theorized that similar micro-organisms might be responsible for the infections frequently occurred after surgery
  - But how might Lister evaluate his theory?

Lister's Experiment

- Lister proposed a new protocol where surgeons were required to wash their hands, wear clean gloves, and disinfect their instruments with a carbolic acid solution
  - Lister randomly assigned 75 patients to receive either his new "sterile surgery" or the old standard of care
  - He recorded how many of each group survived until their discharge from the hospital

	Died	Survived
Control	16	19
Sterile	6	34

Analyzing Lister's Experiment

Notice how a higher proportion of the sterile surgery group survived (85% vs. 54%), but can all of the following explanations be ruled out?

1. Bias
2. Confounding
3. Random chance

	Died	Survived
Control	16	19
Sterile	6	34

Analyzing Lister's Experiment

1. Bias - Some biases might be possible, but there isn't anything that really jumps out
2. Confounding - Can be ruled out since the explanatory variable (sterile surgery or control surgery) was randomly assigned
3. Random chance - IDK? . . . we need to use *statistical inference* to evaluate this possibility!

Statistical Inference

The logic behind **statistical testing**:

1. Enter a *hypothetical world* where a particular *null hypothesis* is true

Statistical Inference

The logic behind **statistical testing**:

1. Enter a *hypothetical world* where a particular *null hypothesis* is true
2. Explore how a experiment/study *might unfold* in this hypothetical world

Statistical Inference

The logic behind **statistical testing**:

1. Enter a *hypothetical world* where a particular *null hypothesis* is true
2. Explore how a experiment/study *might unfold* in this hypothetical world
3. Compare *how different* the results of the real study are from the hypothetical ones

Statistical Inference

The logic behind **statistical testing**:

1. Enter a *hypothetical world* where a particular *null hypothesis* is true
2. Explore how a experiment/study *might unfold* in this hypothetical world
3. Compare *how different* the results of the real study are from the hypothetical ones
4. If the real study is *sufficiently different*, we might *reject* the null hypothesis (eliminating "random chance" as a possible explanation)

For Lister's experiment, how would you describe the hypothetical world we are interested in? (ie: step 1)

Statistical Inference - Step 1 (setting up hypotheses)

- For Lister's experiment, we are interested in the null hypothesis: "the new sterile surgery procedure has no effect on survival"
- Mathematically, statisticians might state this hypothesis as:

$$H_0 : p_{died|sterile} = p_{died|control}$$
$$\rightarrow H_0 : p_{died|sterile} - p_{died|control} = 0$$

- For simplicity, it can be convenient to label these two proportions p_1 and p_2

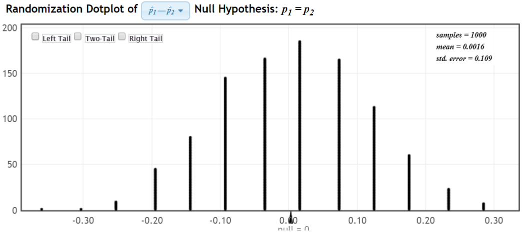
Statistical Inference - Step 2 (determining the null distribution)

After we've established a null hypothesis, we can *simulate* possible outcomes of our study in this hypothetical world. In doing so, there are two key aspects of original study we need to preserve:

1. 40 individuals were randomly assigned to the sterile surgery group and 35 were randomly assigned to the control group
2. Overall (regardless of group), 22 of 75 individuals died before being discharged

Statistical Inference - Step 2 (determining the null distribution)

StatKey provides a tool for carrying out these simulations:

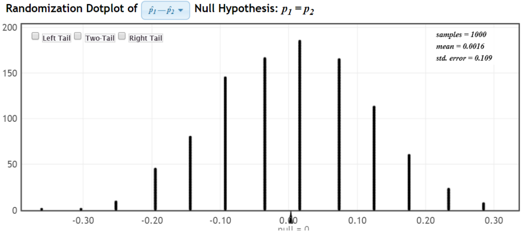


Original Sample

Group	Count	Sample Size	Proportion
Group 1	6	40	0.150
Group 2	16	35	0.457
Group 1-Group 2	-10	n/a	-0.307

Statistical Inference - Step 3 (estimating the *p*-value)

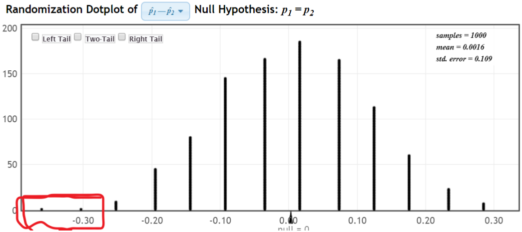
In the *real experiment*, 6/40 individuals in the sterile group died while 16/35 in the control group died (a difference in proportions of -0.31). Would this be a likely outcome *if the null hypothesis were true*?



Original Sample

Group	Count	Sample Size	Proportion
Group 1	6	40	0.150
Group 2	16	35	0.457
Group 1-Group 2	-10	n/a	-0.307

No! Only 2 of 1000 simulations had results like the real experiment!



Original Sample

Group	Count	Sample Size	Proportion
Group 1	6	40	0.150
Group 2	16	35	0.457
Group 1-Group 2	-10	n/a	-0.307

Statistical Inference - Step 4 (making a decision)

- The 2/1000 (or 0.002) of simulations *at least as unusual* as the actual result represent an estimate of the *p-value* of this statistical test

Statistical Inference - Step 4 (making a decision)

- The 2/1000 (or 0.002) of simulations *at least as unusual* as the actual result represent an estimate of the *p-value* of this statistical test
    - This tells us that, had the null hypothesis been true, there'd only be an $\sim 0.2\%$ of seeing a difference in survival like the one seen in Lister's Experiment
    - This is *strong evidence* that random chance isn't a likely explanation for the association we observed

Lister's Experiment - Recap

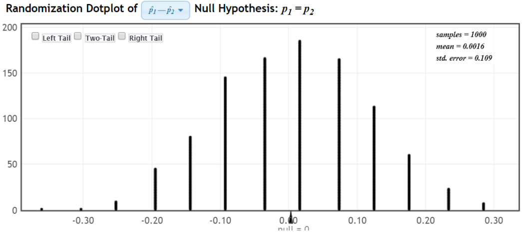
Using the results of this statistical test, we can now rule-out the following explanations for the improved survival in the sterile surgery group

1. Bias - Unlikely, the sample is representative, and the experiment used a control group
2. Confounding - Very unlikely, the explanatory variable was randomly assigned
3. Random chance - Very unlikely, the p-value is 0.002

By ruling out these explanations we can be reasonably certain that Lister's sterilization protocol *caused* the improvement in survival

Next Steps - Do we need to simulate?

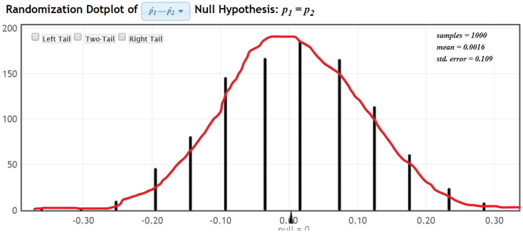
In this example we relied on StatKey to simulate the **null distribution**, but could we have come up with it mathematically?



Original Sample

Group	Count	Sample Size	Proportion
Group 1	6	40	0.150
Group 2	16	35	0.457
Group 1-Group 2	-10	n/a	-0.307

Yes, notice how the null distribution appears to resemble a *normal curve* (the area under that curve can approximate the *p*-value):

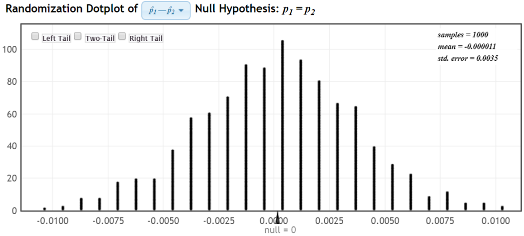


Original Sample

Group	Count	Sample Size	Proportion
Group 1	6	40	0.150
Group 2	16	35	0.457
Group 1-Group 2	-10	n/a	-0.307

Practice #1

Using a cohort born in the 1980's, the CDC tracked 6,168 women in hopes of finding risk factors for breast cancer. In this cohort, 65/4540 (1.4%) women who gave birth to a child before age 25 went on to develop breast cancer, while 31/1628 (1.9%) of the remaining women developed breast cancer. What might explain this difference?



Original Sample

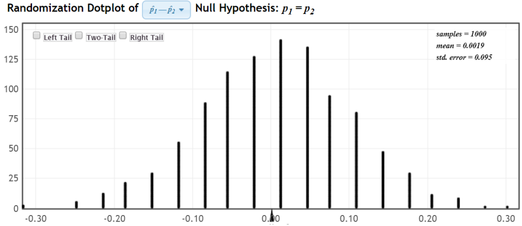
Group	Count	Sample Size	Proportion
Group 1	65	4540	0.014
Group 2	31	1628	0.019
Group 1-Group 2	34	n/a	-0.0047

Practice #1 - Solution

1. Bias - Possible but unlikely, both groups of women came from the same population and we might reasonably assume that breast cancer was diagnosed in the same way for each group
2. Confounding - Possible and likely, the women in this cohort chose their group, so there might be other factors related with both the age they gave birth and their breast cancer risk
3. Random chance - Also possible, roughly 20% of the simulated replications showed a difference at least as extreme as the one observed in this study

Practice #2

In a Florida State University study, 123 college men were randomly paired with a woman who was in either a fertile phase of her cycle or a not fertile phase. For men paired with a woman in a less fertile stage, 38 of the 61 men (62%) copied their partner's sentence construction. For men paired with a woman at peak fertility, 30 of the 62 men (48%) copied their partner's sentence construction. What might explain this difference?



Original Sample

Group	Count	Sample Size	Proportion
Less Fertile	38	61	0.623
Peak Fertile	30	62	0.484
Less Fertile-Peak Fertile	8	n/a	0.139

Practice #2 - Solution

1. Bias - Unlikely, both groups came from the same population and presumably the outcome was measured the same way for each
2. Confounding - Very unlikely, the explanatory variable (less fertile vs. peak fertility) was randomly assigned
3. Random Chance - Possible, about 10% of the time a difference this extreme occurs when the experiment is simulated under a true null hypothesis

Conclusion

- Statistical tests represent one piece of the puzzle when it comes to statistical thinking
  - How data were collected often has a much larger impact on the conclusions you can draw
- In attempting to understand the world, you should apply statistical thinking before jumping to conclusions
  - This means you think about: the data source and possible biases, the study design and confounding variables, and the sample size and possibility random chance.
  - To do this effectively, it is necessary to learn the tools and terminology of statisticians