Study Design and Contingency Tables

Ryan Miller

Grinnell College
Statistics

Introduction

- In 1936, the US was in the midst of the Great Depression with 9 million unemployed and real incomes down 33% from the start of the decade
- Franklin Roosevelt was finishing his first term as president and was being opposed by Republican candidate Alfred Landon
- The *Literary Digest*, a magazine that had correctly predicted every election since 1916, sampled 2.4 million voters and found 57% said they'd vote for Landon (compared with only 43% voting for Roosevelt)
  - Does this sample provide sufficient statistical evidence to believe that Landon will win? What would such a hypothesis test look like?

Introduction (cont.)

- We could set up $H_0 : p = 0.5$ vs. $H_a : p \neq 0.5$ where p is the proportion of the electorate that votes for Landon
  - The Literary Digest found $\hat{p} = 0.57$ in their sample of $n = 2400000$
- The one-sample Z-test should be our "default" in this scenario (one-sample categorical data with a large sample size)
  - Here, $SE = \sqrt{0.5 \cdot (1 - 0.5)/2400000} = 0.000323$, so $Z = \frac{0.57 - 0.5}{0.000323} \approx 216.9$
  - $Pr(Z \geq 216.9 | H_0 \text{ is true}) \approx 0$, so there's *overwhelming evidence* against H_0
  - However, does anyone know what happened in the actual election?

The 1936 Presidential Election

- Roosevelt won reelection 62% to 38%
  - Our hypothesis test evaluated the role of *sampling variability*, which was miniscule with a sample of 2.4 million
  - It *did not* account for the role of *sampling bias*
- The *Literary Digest* collected its sample by mailing 10 million questionnaires to addresses gathered from phone books and club memberships
  - Why did this result in sampling bias?

Ideal Study Design (one-sample)

- The ideal *one-sample design*
  - Have a complete list of everyone in the population of interest
  - Can randomly sample these people
  - Equally costly to sample each person, and no one ever refuses to participate in the study
  - No systematic bias in how the outcome is measured
- Under these circumstances, only *sampling variability* can explain why the sample might differ from the population, so our hypothesis test will produce a reliable conclusion

Ideal Study Design (two-sample)

- The ideal *two-sample design*:
  - Able to collect an ideal sample from the population of interest (previous slide)
  - Can divide this sample into two groups that are identical in every way
  - Impose the treatment of interest on one group
  - No systematic bias in how the outcome is measured in each group
- Under these ideals, only *sampling variability* can explain any difference between groups, so if a hypothesis test provides convincing evidence, we can make a *causal conclusion* (ie: the treatment causes an increase/decrease in the outcome)

Ideal Study Design (two-sample)

- The "gold standard" for a two-sample study is a *randomized, placebo-controlled, double-blind experiment*
  - Random assignment of the treatment and control groups ensures the groups are approximately identical
  - Using a placebo, or fake treatment that is indistinguishable to study participants from the real one, prevents measurement bias (the placebo effect)
  - Double-blinding, or preventing both the researchers and participants from knowing whether they are receiving the real treatment or the placebo, also helps prevent measurement bias

Clofibrate

In 1980, the New England Journal of Medicine published a *randomized controlled double-blind experiment* where participants were assigned to receive clofibrate, a cholesterol lowering drug, or a placebo pill. Below are results for those who took the drug vs. those who didn't:

	Deaths	Survivors
Took Clofibrate	708	4012
Didn't Take Clofibrate	357	1071

1. Does this study provide statistical evidence of a difference between those who took the drug and those who didn't?
2. Is a causal conclusion justified?

Clofibrate (cont.)

Subjects were not randomized with respect to their adherence, so the groups shown in the previous table might differ in meaningful ways:

	Deaths (trt)	Percent	Deaths (ctl)	Percent
Adhered	708	15%	1813	15%
Didn't	357	25%	882	28%
Total	1065	18%	2695	19%

If we'd like to make a causal conclusion, we have to compare the groups as they were randomized, which are not significantly different

Observational Designs

- It is not always possible to conduct a "gold standard" study
  - Does that mean it is impossible to reach scientifically meaningful conclusions in these contexts?

Observational Designs

- It is not always possible to conduct a "gold standard" study
  - Does that mean it is impossible to reach scientifically meaningful conclusions in these contexts?
- No, we can still generate useful ideas for how different variables are *associated* using *observational study designs*
  - While *association* doesn't necessarily imply a causal relationship, it is still useful knowledge
  - Furthermore, if enough plausible alternative explanations can be ruled out, an observational study might be enough for us to act as if it provides causal evidence

Observational Study Example

Consider the following study, which tracked a cohort of 6,168 women born in the 1980s in search of risk factors for breast cancer

	Breast Cancer	No Cancer
Birth Before Age 25	65	4475
Birth After Age 25	31	1157

- We've previously described data like this using *differences in proportions*.
  - If we ignore the question of statistical significance, do the differences in proportions in these groups seem compelling?
  - Is there a better way to describe the *risk* associated with this explanatory variable?

Note: Some women in the cohort never had children and are not included in this contingency table

Relative Risk vs. Risk Difference

- The difference in proportions observed in this type of study is known as a **risk difference**
  - Because breast cancer is a relatively rare outcome, the risk difference is small $\frac{31}{1157} - \frac{65}{4475} = 0.012$ (1.2%)
- In *prospective studies*, Risk differences tend to be used less frequently than **relative risk**:

$$\text{Relative Risk} = \hat{p}_{\text{event|exposed}} / \hat{p}_{\text{event|not exposed}} = \frac{A}{A+B} / \frac{C}{C+D}$$

- The *relative risk of breast cancer* is estimated as 1.84 times higher (elevated by 84%) for women who gave birth before age 25
  - This paints a different picture than the 1.2% risk difference

Prospective Studies

- The breast cancer example, which involved following a cohort of 6,168 women born in the 1980s, is an example of a **prospective study** (sometimes called a cohort study)
- Prospective studies follow a representative sample *forward in time*, waiting for each subject to experience the exposure and experience the event of interest
  - Prospective studies are considered second only to randomized experiments when it comes to the strength of the evidence they provide

Retrospective Studies

- Tracking thousands of individuals for long periods of time is extremely resource intensive (in both time and money)
- An easier way to conduct a study on breast cancer risk factors might:
  - Recruit 100 women with breast cancer (cases)
  - Recruit 100 women without breast cancer (controls)
  - Ask each of these women about their past exposures, such as when they had their first child
- This, which looks backward in time, is called a **retrospective study** (sometimes called a case-control study)

Retrospective Study Example

In a 1986 case-control study investigating the relationship between smoking and oral cancer, researchers collected the smoking history of 304 cases with oral cancer and 139 controls without oral cancer. Data from the study are summarized below:

	Cases	Controls
< 16 cigarettes per day	49	46
≥ 16 cigarettes per day	255	93

Based upon this study design, do you believe these data can be used to estimate the risk that an individual in each population develops oral cancer? Can we estimate the relative risk of oral cancer?

Odds and Odds Ratios

- Relative risk *cannot* be used to measure association in a retrospective study, but a slightly different measure, the **odds ratio** can
  - Unlike relative risk, the odds ratio is symmetric, so it doesn't matter which category we designate as the outcome
- The odds ratio is just as it sounds: the odds of the event given the exposure divided by the odds of the event given a lack of the exposure
- The *odds* of an event is a ratio itself, it is how many times an event occurs relative to how many times it doesn't occur
  - If there is a 50% probability of an event, the odds are 1, which we often express as "1 to 1"
  - If there is a 75% probability of an event, the odds are 3, which we often express as "3 to 1"

Odds and Odds Ratios

Let's revisit the oral cancer study:

	Cases	Controls
< 16 cigarettes per day	49	46
≥ 16 cigarettes per day	255	93

1. What are the odds of oral cancer among the low-smoking subjects? What are the odds among high-smoking subjects?
2. What is the *odds ratio* relating these two groups? How might you interpret it?

Odds Ratios and Hypothesis Tests

- For odds ratios, we are usually interested in the hypothesis $H_0 : OR = 1$, which implies equal odds of the outcome in both groups.
  - We've seen odds ratios calculated in R when we encountered Fisher's Exact Test:

```
my_table = data.frame(cases = c(255, 49), controls = c(93,46))
fisher.test(my_table)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  my_table
## p-value = 9.599e-05
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  1.566573 4.213349
## sample estimates:
## odds ratio
##    2.56799
```

Prospective vs. Retrospective Studies

Advantages of prospective studies:

- Only a single sample is collected (less room for sampling bias)
- Risk factors and events are directly observed (less potential for recall bias)
- Can be used to estimate probabilities, relative risk, and odds ratios
- More reflective of nature

Advantages of retrospective studies:

- Less expensive and less time consuming
- Easier to use when studying rare events
- No loss to follow-up concerns
- Odds ratios provide a valid measure of association

Cross-Sectional Studies

- The weakest type of observational design is the **cross-sectional study**
- In this design, researchers collect a *single sample* at *a single snapshot in time* and cross-classify individuals in that sample depending upon their exposure and outcome
  - This differs from a retrospective study, which collects separate samples of cases and controls (and pays careful attention to the separate challenges of sampling these populations)

Weaknesses of Cross-Sectional Studies

- Cross-sectional studies are the easiest to perform, but because they don't pay attention to time, they struggle to establish cause-effect relationships
- Selection bias is a major issue for cross sectional designs:
  - Consider a cross-sectional sample of factory workers
  - We might want to compare their rate of asthma to the rate of asthma in the general public in order to establish an association between factory work and asthma
  - Why might this be problematic?

Weaknesses of Cross-Sectional Studies (cont.)

- Factory workers who develop asthma will likely change jobs, so they will not appear in a cross-sectional sample
  - A cohort, or a case-control study, is less likely to encounter this problem
- It is also nearly impossible to make cause-effect claims from a cross-sectional study
  - If X and Y are measured at the same time, X could cause Y, or Y could cause X, or another variable could cause both!

Practice

A study surveyed 257 hospitalized individuals, classifying whether suffered from a circulatory disease, a respiratory disease, both, or neither. The results are displayed below:

	Respiratory Disease	No Respiratory Disease
Circulatory Disease	7	29
No Circulatory Disease	13	208

1. Use an appropriate statistical test to determine whether the association between presence of a circulatory disease, and presence of a respiratory disease could be due to chance (sampling variability)
2. Considering design limitations, does this mean that you're more likely to get a respiratory disease if you have a circulatory disease?

Practice (solution)

1. Using a Z-test in `R`, the p-value is 0.005, so it is very unlikely the association is due to random chance
2. No, individuals with both types of disease are more likely to be hospitalized (and biased towards ending up in this sample). The researchers in this study also looked at a sample of non-hospitalized individuals:

	Respiratory Disease	No Respiratory Disease
Circulatory Disease	15	142
No Circulatory Disease	189	2181

Conclusion

For our second exam you are expected to be familiar with:

1. The importance of study design, and how various features of a study's design influence the conclusions that can be drawn from the sample data
2. How to measure association in contingency tables (two-way frequency tables) using risk differences, relative risks, and odds ratios, as well as how/why these measures are used, and how they are interpreted