Stat 13, Intro. to Statistical Methods for the Life and Health Sciences.

1. ANOVA conditions.
2. Comparing more than two proportions, chapter 8.

For the final, bring a CALCULATOR and any books or notes you want. No computers.
Read ch8.
http://www.stat.ucla.edu/~frederic/13/F18.

1. ANOVA validity conditions.

- Just as with the simulation-based method, we are assuming we have independent observations.
- Two extra conditions must be met to use traditional ANOVA:
  - Normality: We assume in ANOVA the observations are all draws from a normal distribution. This was not stated clearly last time. If there are obvious departures from normality, then ANOVA should not be used.
  - Equal variation: Largest standard deviation should be no more than twice the value of the smallest.

Comparing More Than Two Proportions

Chapter 8

Chapter Overview

- In chapter 5, our explanatory variable had two outcomes (like parents smoke or not) and the response variable also had two outcomes (like baby is boy or girl).
- In this chapter we will allow the explanatory variable to have more than two outcomes (like both parents smoke, only mother smokes, only father smokes, or neither parent smokes).
- This is all a special case of comparing multiple means, which we explored in chapter 9, only now the responses are 0 or 1, so the sample mean = sample proportion.

Section 8.1: Comparing Multiple Proportions: Simulation-Based Approach

Comparing Multiple Proportions: Simulation-Based Approach

Section 8.1

Stopping

- Virginia Tech students investigated which vehicles came to a stop at a intersection where there was a four-way stop.
- While the students examined many factors for an association with coming to a complete stop, we will investigate whether intersection arrival patterns are associated with coming to a complete stop:
  - Vehicle arrives alone
  - Vehicle is the lead in a group of vehicles
  - Vehicle is a follower in a group of vehicles

Hypotheses

- **Null hypothesis**: There is *no association* between the arrival pattern of the vehicle and if it comes to a complete stop.
- **Alternative hypothesis**: There is *an association* between the arrival pattern of the vehicle and if it comes to a complete stop.

Hypotheses

- Another way to write the null uses the probability that a single vehicle will stop is *the same as* the probability a lead vehicle will stop, which is *the same as* the long-term probability that a following vehicle will stop
- Or $\pi_{Single} = \pi_{Lead} = \pi_{Follow}$ where π is the probability a vehicle will stop.
- The alternative hypothesis is that not all these probabilities are the same (at least one is different).

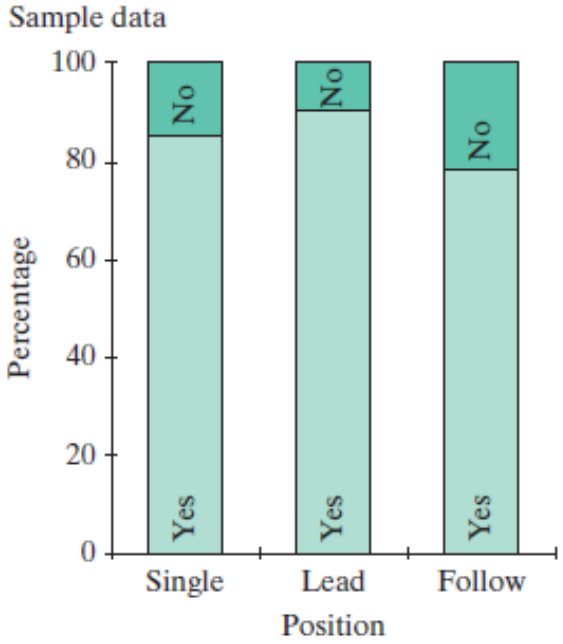
Hypotheses

We can write out the null hypothesis in symbols, but we can't do that (easily) for the alternative

- H_0: $\pi_{Single} = \pi_{Lead} = \pi_{Follow}$
- H_a: Not all the probabilities of stopping are the same (at least one is different).

Explore the Data

- Percentage of vehicles stopping:
  - 85.8% of single vehicles
  - 90.5% of lead vehicles
  - 77.6% of vehicles following in a group



	Single vehicle	Lead vehicle	Following vehicle	Total
Complete stop	151 (85.8%)	38 (90.5%)	76 (77.6%)	265
Not complete stop	25 (14.2%)	4 (9.5%)	22 (22.4%)	51
Total	176	42	98	316

No Association

- Remember that no association implies the proportion of vehicles that stop in each category should be the same.
- Our question is, if the same proportion of vehicles come to a complete stop in all the three categories in the long run, how unlikely would we get sample proportions at least as far apart as we did?

Statistic

To find a statistic we start by finding the three **differences** in proportions.

$\hat{p}_S$ proportion of single vehicles that stopped = 0.858

$\hat{p}_L$ proportion of lead vehicles that stopped = 0.905

Difference in proportions = 0.858 − 0.905 = **-0.047**

$\hat{p}_F$ proportion of following vehicles that stopped = 0.776

$\hat{p}_S$ proportion of single vehicles that stopped = 0.858

Difference in proportions = 0.776 − 0.858 = **-0.082**

$\hat{p}_L$ proportion of lead vehicles that stopped = 0.905

$\hat{p}_F$ proportion of following vehicles that stopped = 0.776

Difference in proportions = 0.905 − 0.776 = **0.129**

Statistic

- How can we combine 3 differences (-0.047, -0.082 and 0.129) into a single statistic?
- Add them up? Average them?
  - -0.047 + (-0.082) + 0.129 = 0.
  - [-0.047 + (-0.082) + 0.129]/3 = 0.

Instead, we will use the MAD.

MAD Statistic

1. Statistic

- We are going to use the mean of the absolute value of the differences (MAD)
- (0.047 + 0.082 + 0.129)/3 = **0.086**.
- What would have to be true for the average of absolute differences to equal 0?

Simulate

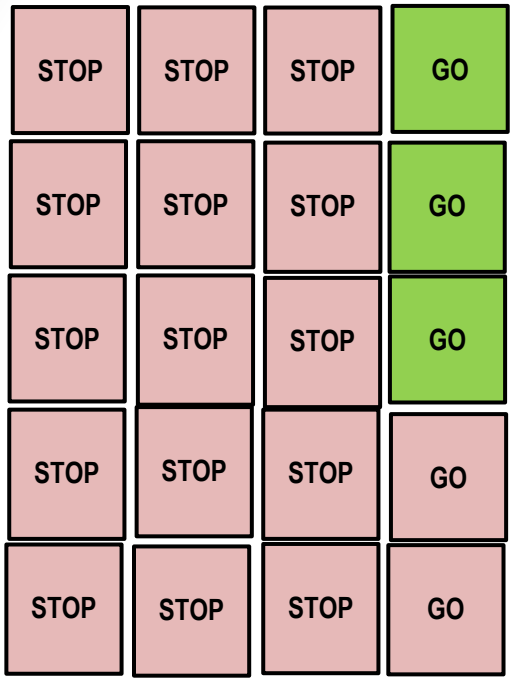
2. Simulate

- If there is no association between arrival pattern and whether or not a vehicle stops it basically means it doesn't matter what the arrival pattern is. Some vehicles will stop no matter what the arrival pattern and some vehicles won't.
- We can model this by shuffling the response variable again.
- We could also use cards, with the response outcomes (stop/no stop) on the cards, shuffle, and place them in three piles (single/lead/following).

Simulation

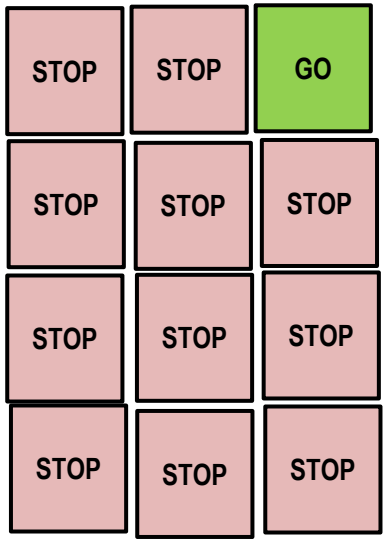
- Similar to testing two proportions we can shuffle values of the response variable (stop or not) and randomly place them into piles representing the categories of the explanatory variable (the arrival pattern).
- This time we have three piles instead of two.
- After each shuffle, we calculate the MAD statistic of the shuffled data and that will be a point in the null distribution.
- We will now see what happens with a smaller data set.

Single



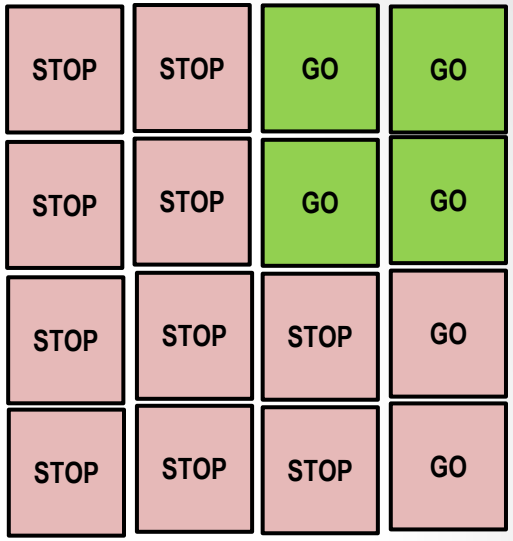
prop = 0.85

Lead



prop = 0.83

Following

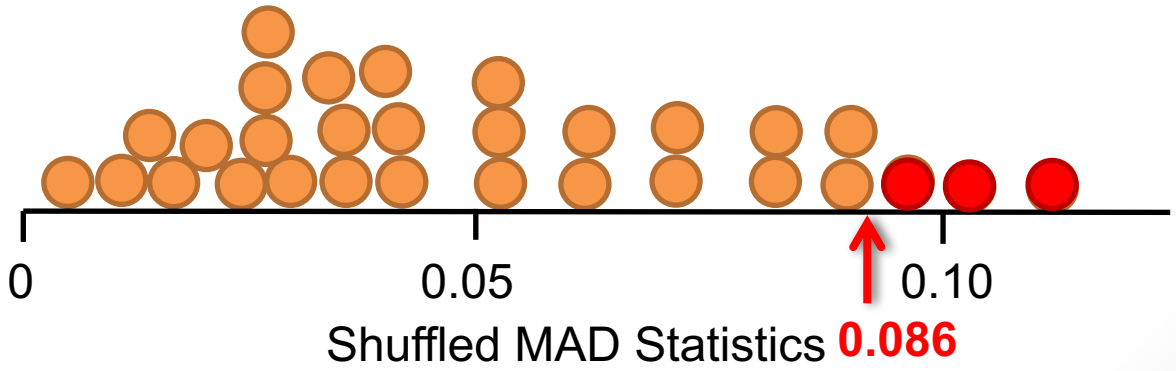


prop = 0.81

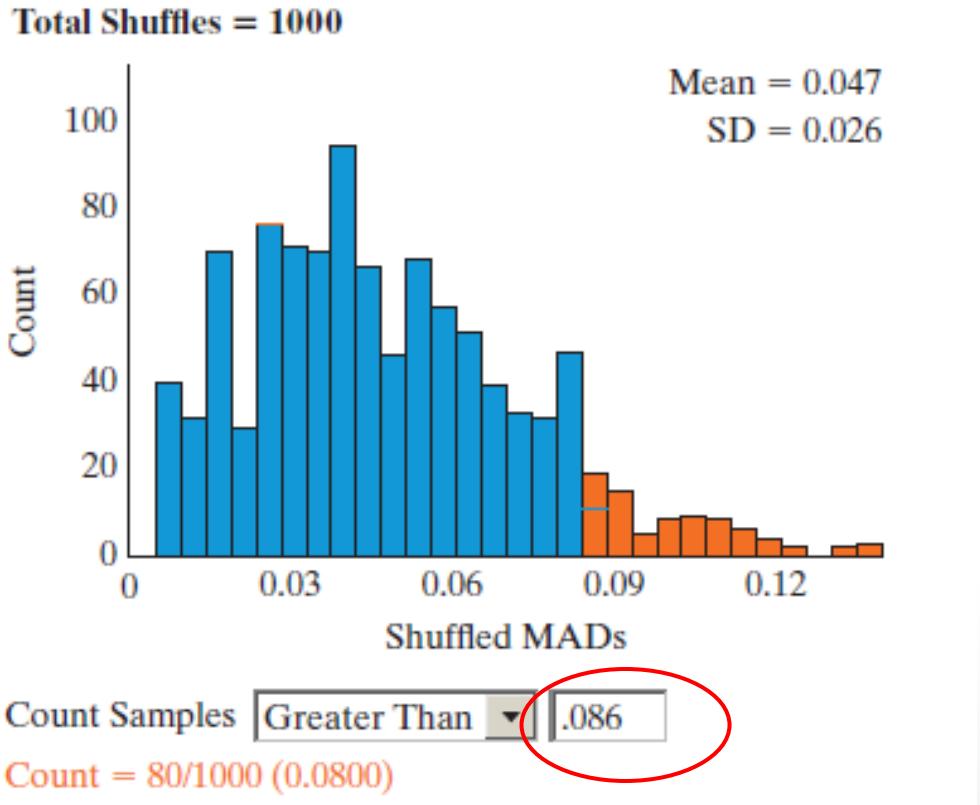
MAD = (|0.85 – 0.83| + |0.85 – 0.81| + |0.83 – 0.81|)/3 = 0.027

More Simulations

With 30 repetitions of creating simulated MAD statistics, we had 3 MAD statistics that were as large or larger than 0.086.



- Simulated values of the statistic for 1000 shuffles
- Bell-shaped?
- Centered at 0?



Strength of Evidence

3. Strength of evidence

- If the MAD statistic is larger, then the differences between the groups is greater.
- **Hence to calculate our p-value we will always count the simulations that are as large or larger than the MAD statistic.**

Conclusions

- We had a p-value of 0.08 so there is moderate evidence against the null hypothesis, but not strong enough for statistical significance.
- Do the results generalize to intersections beyond the one used?
  - Probably not since intersections have different factors that influence stopping.
- Can we draw any cause-and-effect conclusions?
  - No, because this was an observational study. What might be confounders?
- If we had stronger evidence of a difference in groups, we could follow with pairwise tests to see which proportions are significantly different from each other.

Learning Objects for Section 8.1

- Compute the MAD (mean absolute value of the differences) statistic from a data set when comparing multiple proportions.
  - What is the MAD for data with means: 0.3, 0.4, 0.9?
- Understand that larger values of the MAD statistic suggest stronger evidence against the null hypothesis.
- Understand that the simulated null distribution of the MAD statistic looks different from other simulated null distributions presented thus far.

Comparing Multiple Proportions: Theory-Based Approach

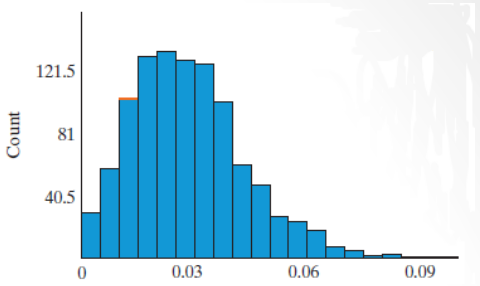
Section 8.2

Theory based vs. Simulation based

As always:

- Simulation based methods
  - do not require validity conditions such as normality or large sample size.
  - require the ability to simulate.
- Theory-based methods
  - avoid the need for simulation
  - require additional validity conditions to be met.

Other Distributions



- The null distributions in chapters 5-7 were bell-shaped and centered at 0.
- For these, we used normal and t-distributions to predict what the null distribution would look like.
- In this chapter our null distribution was neither bell-shaped nor centered at 0.
- We can, however, use a chi-squared distribution to predict the shape of the null distribution.
- When doing this, we will not use the MAD statistic, but a chi-square statistic.

Sham Acupuncture

Example 8.2

Sham Acupuncture

- A randomized experiment was conducted exploring the effectiveness of acupuncture in treating chronic lower back pain in Germany (Haake et al. 2007).
- Acupuncture inserts needles into the skin of the patient at acupuncture points to treat a variety of ailments.

Sham Acupuncture

- 3 treatment groups
  - Verum acupuncture: traditional Chinese
  - Sham acupuncture: needles inserted into the skin, but not deeply and not at acupuncture points
  - Traditional, non-acupuncture, therapy of drugs, physical therapy and exercise.
- 1162 patients were randomly assigned to one of the three treatment groups
- 10 therapy sessions
- Did substantial reduction in back pain occur after 6 months as measured on one of two scales?

Sham Acupuncture

- Explanatory Variable: Which type of treatment the subject received. (Categorical with 3 categories)
- Response Variable: Did the subject get substantial relief from their lower back pain? (Categorical with 2 categories)

- Is this an experiment or an observational study?

Sham Acupuncture

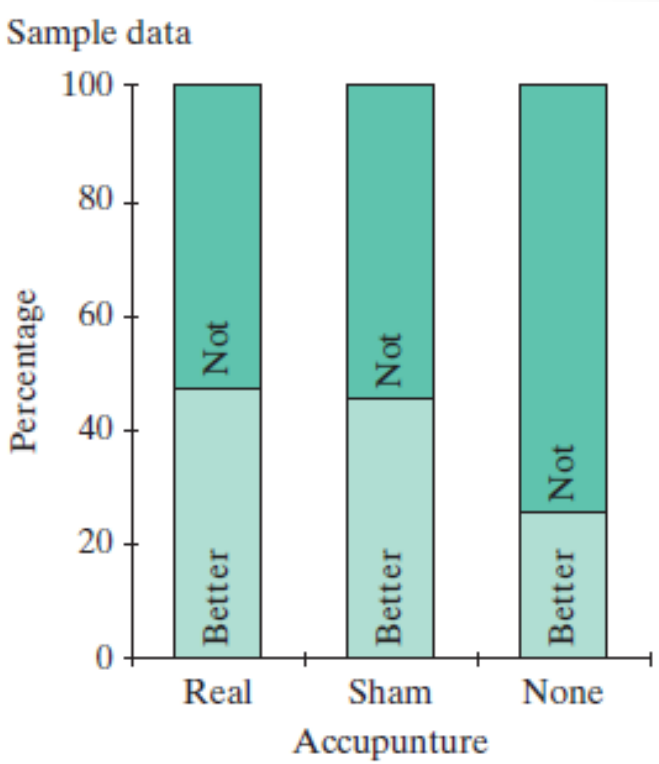
Null hypothesis - no association between type of treatment received and reduction in back pain.

Alternative hypothesis - there is an association between type of treatment and reduction in back pain.

H_0: $\pi_{real} = \pi_{Sham} = \pi_{None}$

H_a: Not all the probabilities of back pain reduction are the same (at least one is different).

Results



	Real acupuncture	Sham acupuncture	Nonacupuncture	Total
Substantial reduction in pain	184 (0.475)	171 (0.442)	106 (0.273)	461
Not a substantial reduction in pain	203 (0.525)	216 (0.558)	282 (0.727)	701
Total	387	387	388	1,162

MAD Statistic

- Remember that the MAD statistic is the mean of the absolute value of differences in proportions for all 3 groups:

  Real vs. Sham: 0.476-0.442=0.034

  Real vs. None:0.476-0.274= 0.202

  Sham vs. None: 0.442-0.274= 0.168

- The statistic (MAD) is (0.034+0.202+0.168)/3=0.135
- Larger MAD statistics give more evidence against the null

Statistic

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

- Another statistic commonly used is the chi-square statistic.
- Its distribution is easier to fit with a theory-based distribution than a MAD statistic distribution.

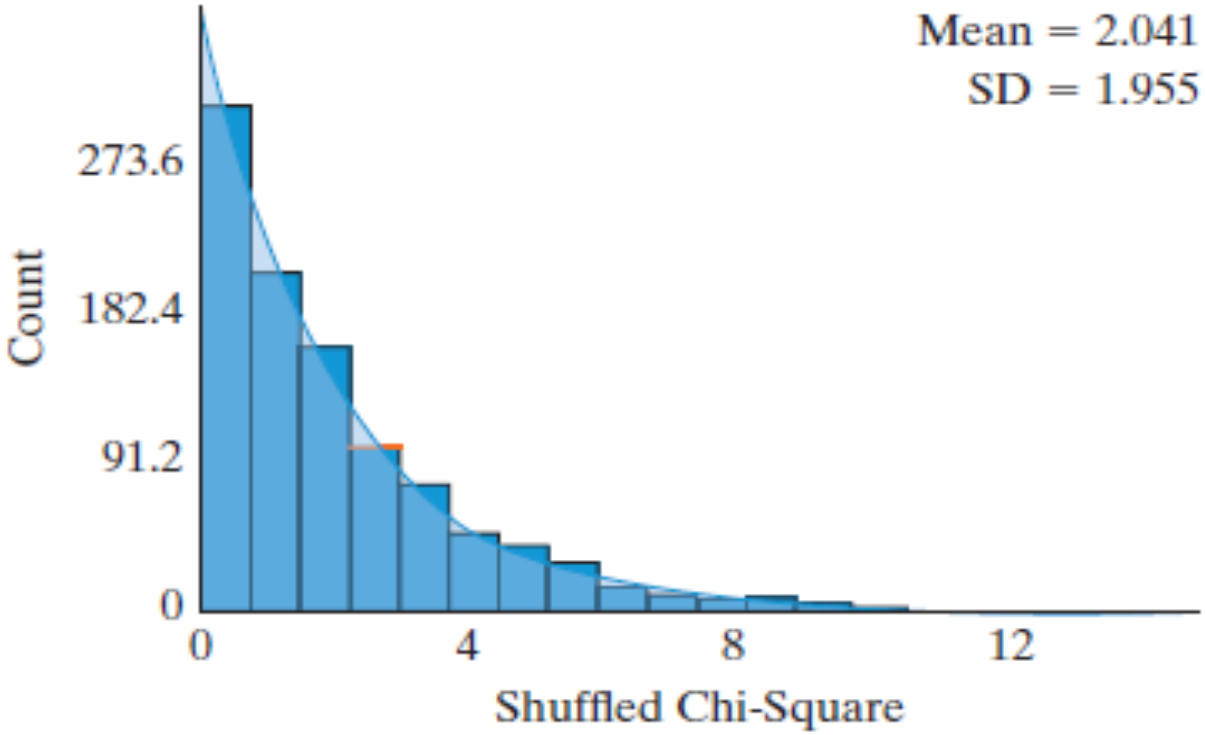
$$\chi^2 = \left(\frac{1}{\hat{p}(1-\hat{p})} \right) \sum n_i (\hat{p}_i - \hat{p})^2$$

- $\hat{p}_i$ is the proportion of successes in category *i*
- $\hat{p}$ is the overall proportion of successes in the dataset.
- n_i is the sample size of category *i*

Sham Acupuncture

- The chi-square statistic is another way to measure how far apart the proportions are from each other.
- Just like the MAD statistic, the larger the chi-square statistic the more evidence there is against the null, and a chi-square of 0 means proportions are all the same.
- We can test this using **both MAD and chi-square** simulation.

Theory-based p-value



Count Samples | Greater Than | 38.05 | Count

Count = 0/1000 (0.0000)

☑ Overlay Chi-square distribution

theory-based p-value = 0.0000

Sham Acupuncture

Strength of evidence:

- In both cases, we got p-values of 0.
- Nothing as large as a MAD statistic of 0.135 or larger ever occurred in this simulation.
- Likewise nothing as large as a chi-square statistic of 38.05 or larger ever occurred in this simulation.
- Hence we have very strong evidence against the null and in support of the type of acupuncture used is associated with pain reduction.

Sham Acupuncture

- The theory based method, namely the chi square test, only works well when each cell in the 2-way table has at least 10 observations.
- This is easily met here. The smallest count was 106.

Follow up

- What if you find evidence of an association?
- How do we determine exactly where the association is?
  - One can use pairwise confidence intervals for the difference in proportions to determine exactly where the association lies.

Follow up

- 95% confidence intervals on the difference in improvement percentages comparing:
  - Real to sham (-0.0366, 0.1038)
  - Real to **none** (0.1356,0.2689)*
  - Sham to **none** (0.1022, 0.2351)*
- There is evidence that the probability of pain reduction is significantly lower for no acupuncture treatment than the other two.
- There is no significant difference between real and sham acupuncture however.