Lecture 14: Statistical Significance of 2x2 Contingency Tables, Part 2

Statistics 10

Colin Rundel

March 12, 2012

Remember when...

MythBusters yawning experiment:

	Seeded	Control	Total
Yawn	10	4	14
Not Yawn	24	12	36
Total	34	16	50

$$P(\text{yawn}|\text{seeded}) = \frac{10}{34} = 0.29$$

$$P(\text{yawn}|\text{control}) = \frac{4}{16} = 0.25$$

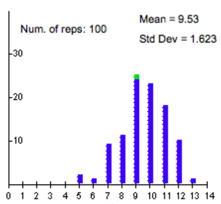
$$P(\text{yawn}|\text{seeded}) - P(\text{yawn}|\text{control}) = 0.0414$$

Possible explanations:

- Yawning is *independent* of seeing someone else yawn; therefore, the difference between the proportions of yawners in the control and seeded groups is *due to chance*. $\rightarrow$ *nothing is going on*
- Yawning is *dependent* on seeing someone else yawn; therefore, the difference between the proportions of yawners in the control and seeded groups is *real*. $\rightarrow$ *something is going on*

With a slightly different terminology

- We started with the assumption that yawning is independent of seeing someone else yawn. $\rightarrow$ *null hypothesis*
- We then investigated how the results would look if we simulated the experiment many times assuming the null hypothesis is true. $\rightarrow$ *testing*



- Since the simulation results were similar to the actual data (on average roughly 10 people yawning in the seeded group), we decided not to reject the null hypothesis in favor of the *alternative hypothesis*.

A trial as a hypothesis test

- Hypothesis testing is very much like a court trial.
  - In a trial, the burden of proof is on the prosecution.
  - In a hypothesis test, the burden of proof is on the unusual claim.
- H_0 : Defendant is innocent
  H_A : Defendant is guilty
- Collect data - The null hypothesis is the ordinary state of affairs (the status quo), so it's the alternative hypothesis that we consider unusual (and for which we must gather evidence).
- Then we judge the evidence - "Could these data plausibly have happened by chance if the null hypothesis were true?"
  - If they were very unlikely to have occurred, then the evidence raises more than a reasonable doubt in our minds about the null hypothesis.
- Ultimately we must make a decision. How unlikely is unlikely?

A trial as a hypothesis test (cont.)

- If the evidence is not strong enough to reject the assumption of innocence, the jury returns with a verdict of "not guilty".
  - The jury does not say that the defendant is innocent, just that there is not enough evidence to convict.
  - The defendant may, in fact, be innocent, but the jury has no way of being sure.
- Said statistically, we fail to reject the null hypothesis.
  - We never declare the null hypothesis to be true, because we simply do not know whether it's true or not.
  - Therefore we never "accept the null hypothesis".

Recap: hypothesis testing framework

- We start with a *null hypothesis (H_0)* that represents the status quo.
- We also have an *alternative hypothesis (H_A)* that represents our research question, i.e. what we're testing for.
- We conduct a hypothesis test under the assumption that the null hypothesis is true, either via simulation or theoretical methods (coming soon).
- If the test results suggest that the data do not provide convincing evidence for the alternative hypothesis, we stick with the null hypothesis. If they do, then we reject the null hypothesis in favor of the alternative.

Back to MythBusters

We are interested in testing if yawning and seeding are independent.

- We start with the assumption that they are independent

$$H_0 : P(\text{Yawn}|\text{Seeded}) = P(\text{Yawn}|\text{Control})$$

- We test the claim that the conditional probabilities are different

$$H_A : P(\text{Yawn}|\text{Seeded}) \neq P(\text{Yawn}|\text{Control})$$

We perform the test using a randomization test (simulations) and make a decision whether or not we reject the null hypothesis.

p-values

- The *p-value* is the probability of observing data at least as favorable to the alternative hypothesis as our current data set, if the null hypothesis was true.
- If the p-value is *low* we say that it would be very unlikely to observe the data if the null hypothesis were true, and hence *reject H_0*.
- If the p-value is *high* we say that it is likely to observe the data even if the null hypothesis were true, and hence *do not reject H_0*.
- We never accept H_0 since we're not in the business of trying to prove it. We simply want to know if the data provide convincing evidence to support H_A.
- How low is low enough to *reject H_0*? We usually use a cutoff we call the significance level designated by α, which is usually taken at 5%
- The choice of $\alpha = 0.05$ is arbitrary, we could just as easily use $\alpha = 0.01$ or $\alpha = 0.10$

Calculating the p-value

p-value: probability of observing a more extreme difference between $P(\text{Yawn}|\text{Seeded})$ and $P(\text{Yawn}|\text{Control})$ than our observed value if we assume H_0 is true.

We can rewrite the hypothesis as follows:

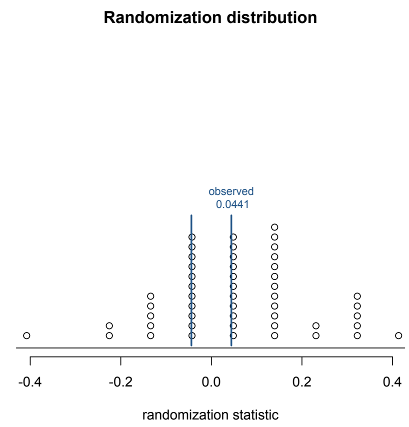
$$H_0 : P(\text{Yawn}|\text{Seeded}) - P(\text{Yawn}|\text{Control}) = 0$$

$$H_A : P(\text{Yawn}|\text{Seeded}) - P(\text{Yawn}|\text{Control}) \neq 0$$

and run the simulations where each iteration calculated the difference between these two probabilities (instead of counts which we saw in Lab 6).

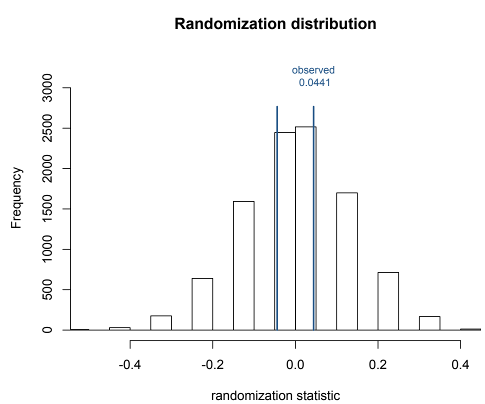
We can then estimate the p-value by the percentage of simulations where the absolute difference is larger than the observed difference.

Calculating the p-value, cont.



p-value $= 28/50 = 0.56$

Calculating the p-value, cont.



p-value $= 0.5184$

Making a decision

- p-value $= 0.5184$
  - If the true difference between $P(\text{Yawn}|\text{Seeded})$ and $P(\text{Yawn}|\text{Control})$ is 0, then there is a 51.84% chance of observing a sample of 50 subjects, with 34 seeds and 16 controls, where there is at least a difference of 0.0414.
  - This is a pretty high probability that our observed difference could have happened simply by chance.
- Since p-value is *high* (greater than 5%) we *fail to reject H_0* at the $\alpha = 0.05$ significance level.
- These data do not provide convincing evidence yawning is contagious.
- We have not shown that yawning is not contagious (we never accept H_0), only that there is not evidence that it is contagious

Gender Discrimination

		Promotion		
		Promoted	Not Promoted	Total
Gender	Male	21	3	24
	Female	14	10	24
	Total	35	13	48

$$P(\text{Promoted}|\text{Male}) = 21/24 = 0.875$$

$$P(\text{Promoted}|\text{Female}) = 14/24 = 0.583$$

$$P(\text{Promoted}|\text{Male}) - P(\text{Promoted}|\text{Female}) = 0.292$$

Gender Discrimination Hypotheses

This experiment was undertaken because the researchers believed that females were being discriminated against during promotions, therefore it is our alternative hypothesis that a male employee is more likely to be selected for promotion than a female employee.

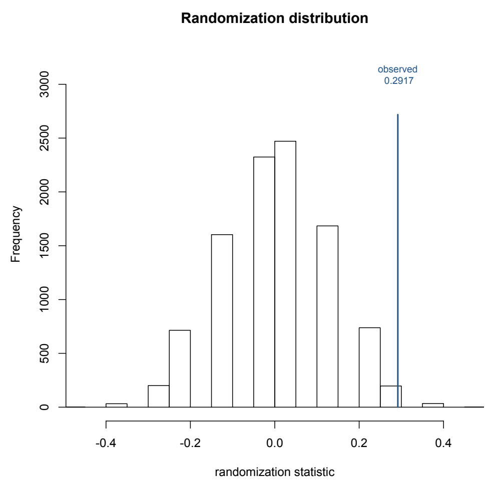
$$H_0 : P(\text{Promoted}|\text{Male}) = P(\text{Promoted}|\text{Female})$$

$$H_A : P(\text{Promoted}|\text{Male}) > P(\text{Promoted}|\text{Female})$$

Consequently, the calculated p-value will be based on the number of simulation results where

$$P(\text{Promoted}|\text{Male}) - P(\text{Promoted}|\text{Female}) \geq 0.292$$

Gender p-value.



p-value = 0.0035

Making a decision

- p-value = 0.0035
  - If the true difference between $P(\text{Promoted}|\text{Male})$ and $P(\text{Promoted}|\text{Female})$ is 0, then there is a 0.35% chance of observing a sample of 48 subjects, with 24 men and 24 women, where there is at least a difference between conditional probabilities of 0.292.
  - This is a pretty low probability that our observed difference could have happened simply by chance.
- Since p-value is *low* (less than 5%) we *reject H_0* at the $\alpha = 0.05$ significance level.
- These data do provide convincing evidence males where more likely to be promoted than females.