

Chapter 4: Methods for Two Categorical Variables

Example 4.1: MythBusters and the Yawning Experiment

MythBusters, a popular television program on the Discovery Channel, once conducted an experiment to investigate whether or not yawning is contagious. The premise of the experiment was to invite a stranger to sit in a booth for an extended period of time. Fifty subjects were said to be tested in total, of which 34 were “seeded” with a yawn by the person conducting the experiment. The other 16 were not given a yawn seed. Using a two-way mirror and a hidden camera, the experimenters observed and recorded the results which are given below.

	Yawned	Did not Yawn	Total
Seeded with a yawn	10	24	34
Not seeded with a yawn	4	12	16

Research Question: Do these data provide statistical evidence that those “seeded” with a yawn are more likely to actually yawn than those who are not seeded?

When we analyze data on two variables, our first step is to distinguish between the *response variable* and the *explanatory* (or *predictor*) *variable*.

Note

- **Response variable:** The outcome variable on which comparisons are made.
- **Explanatory (or predictor) variable:** This defines the groups to be compared.

1. What are the variables in the MythBusters Yawning experiment? Are they categorical or numerical?
2. Which is the response variable? Which is the explanatory variable?

i Descriptive Methods for Two Categorical Variables:

- A **contingency table** shows the joint frequencies of two categorical variables. The rows of the table denote the categories of the first variable, and the columns denote the categories of the second variable.
- A **mosaic plot** gives a visual representation of the relationship between two categorical variables. A mosaic plot graphically presents the information given in the contingency table.

SHOW THESE HERE!!!

1. Find the proportion that yawn in the Seeded group.
2. Find the proportion that yawn in the Not Seeded group.
3. Find the difference in the proportion that yawn between these two groups. Do these proportions differ in the direction conjectured by the researchers?
4. Even if the seeding of a yawn had absolutely no effect on whether or not a subject actually yawned, is it possible to have obtained a difference such as this by random chance alone? Explain.

Statistical Inference: Conducting a Simulation Study

The above descriptive analysis tells us what we have learned about the 50 subjects in the study. Can we make any inferences beyond what happened in the study (i.e., general statements about the population)? Does the higher proportion of yawners in the seeded group provide convincing evidence that being seeded with a yawn actually makes a person more likely to yawn? Note that it is possible that random chance alone could have led to this large of a difference. That is, while it is possible that the yawn seeding had no effect and the MythBusters happened to observe more yawners in the seeded group just by chance, the key question is whether it is probable.

We will answer this question by replicating the experiment over and over again, but in a situation where we know that yawn seeding has no effect (the null model). We'll start with 14 yawners and 36 non-yawners, and we'll randomly assign 34 of these 50 subjects to the seeded group and the remaining 16 to the non-seeded group.

Note that we could use playing cards to replicate this experiment: let 14 red cards represent the yawners, and let 36 black cards represent the non-yawners. Shuffle the cards well, and randomly deal out 34 to be the seeded group. Then, construct the contingency table to show the number of yawners and non-yawners in each group.

	Yawned	Did not Yawn	Total
Seeded with a yawn			34
Not seeded with a yawn			16
Total	14	36	50

Next, note that if you know the number of yawners in the Seeded group, then you can fill in the rest of the cells in the contingency table. So, we need only focus on the number of yawners in the Seeded group. Record the number of yawners in the Seeded group from your first simulated experiment in the table below. Then, repeat this randomization process nine more times, recording your results in the table.

Simulated Ex- peri- ment	1	2	3	4	5	6	7	8	9	10
Number of Yawn- ers in Seeded Group										

Next, create a dotplot of the results:

GRAPHIC HERE

1. How many randomizations were performed by the class as a whole? What proportion gave results at least as extreme as the actual study (10 or more yawners in the seeded group)?
2. Note that the random process used in the simulation study models the situation where the yawn seeding has no effect on whether the subject actually yawns—we simply assume there are 14 people who will yawn no matter what group they are in, and they are assigned to the two groups at random. Based on this simulation study, does it appear that random assignment of yawners to groups will result in 10 or more yawners in the seeded group just by chance? Explain.

3. Recall that the MythBusters obtained 10 yawners in the Seeded group. Given your answer to Question 2, would you say that the data provide convincing statistical evidence to support the research question? Explain.

LARGE SCALE SIMULATION WITH TECHNOLOGY

4. What does each dot on the above plot represent?
5. What value(s) occurred most often by chance under the null model? Explain why this makes sense.
6. How often did we see results at least as extreme as the observed data (10 or more yawners in the seeded group) under the null model? Calculate the proportion of simulated results in which we observed 10 or more yawners in the seeded group. Note that this is an approximate p-value!
7. The MythBusters reported the following results: 25% yawned of those not given a yawn seed, and 29% yawned of those given a yawn seed. Then, they cited the “large sample size” and the 4% difference in the proportion that yawned between the seeded and non-seeded group to confidently conclude that yawn seed had a significant effect on the subjects. Therefore, they concluded that the yawn is decisively contagious. Do you agree or disagree with their answer? Justify your reasoning.

Statistical Inference: Using The Chi-Square Test

Example 4.2: Vested Interest and Task Performance

Example 4.2: Vested Interest and Task Performance

This example is from *Investigating Statistical Concepts, Applications, and Methods* by Beth Chance and Allan Rossman. 2006. Thomson-Brooks/Cole.

“A study published in the *Journal of Personality and Social Psychology* (Butler and Baumeister, 1998) investigated a conjecture that having an observer with a vested interest would decrease subjects’ performance on a skill-based task. Subjects were given time to practice playing a video game that required them to navigate an obstacle course as quickly as possible. They were then told to play the game one final time with an observer present. Subjects were randomly assigned to one of two groups. One group (A) was told that the participant and observer would each win \$3 if the participant beat a certain threshold time, and the other group (B) was told only that the participant would win the prize if the threshold were beaten. The threshold was chosen to be a time that they beat in 30% of their practice turns. The following results are very similar to those found in the experiment: 3 of the 12 subjects in group A beat the threshold, and 8 of 12 subjects in group B achieved success.”

Research Question: Does having an observer with a vested interest decrease performance on a skill-based task?

1. What are the variables in the study? Are they categorical or numerical?
2. Which is the response variable? Which is the explanatory variable?

Example 4.3 High-Salt Diets and Cardiovascular Disease

Suppose a retrospective study is carried out among men aged 50-54 in a specific county who died over a one-month period. The investigators try to include approximately an equal number of men who died from CVD (the cases) and men who died from other causes (the controls). Of 15 people who died from CVD, 4 were on a high-salt diet before they died. In contrast, of 15 people who died from other causes, only 1 was on a high-salt diet. The data are shown in the following contingency table:

Research Question: Is the proportion of the CVD group which has a high-salt diet greater than the proportion of the Non-CVD group which has a high-salt diet?

Observational Studies vs. Designed Experiments

Reconsider the “Vested Interest and Task Performance” example. Fisher’s exact test provided evidence that the proportion of successes was in fact smaller for the vested interest group ($p\text{-value} = .0498$). Now, the question is this: can we conclude that having a vested interest really is the cause of the decreased performance? Similarly, consider the “High-Salt Diet and Cardiovascular Disease” example. If this would have yielded a statistically significant result, would it have been fair to conclude that high-salt diets cause cardiovascular disease based on this study? The answer to these questions lies in how the data were collected; i.e., we must first determine whether the study was a designed experiment or an observational study.

Note

- An **observational study** involves collecting and analyzing data without randomly assigning treatments to experimental units.
- On the other hand, in a **designed experiment**, a treatment is randomly imposed on individual subjects in order to observe whether the treatment causes a change in the response.

Key statistical idea:

Observational studies establish only that an association exists between the variables under study. With observational studies, it is always possible that there are other lurking variables not controlled for in the study that may be impacting the response. Since we can’t be certain these other factors are balanced out between treatment groups, it is

possible that these other factors could explain the difference between treatment groups. With designed experiments, there still may be lurking variables not specifically controlled for in the study; however, the random assignment of treatments used by researchers in a designed experiment should balance out between the treatment groups any other factors that might be related to the response variable. In a sense, we control for these other lurking variables by balancing their effects between the two groups via random assignment. Therefore, designed experiments can be used to establish a cause-and-effect relationship if the results are significant.

Note that the “Vested Interest and Task Performance” study is an example of a designed experiment since participants were randomly assigned to the two groups. We were trying to show that having a vested interest caused a decreased task performance. The small p-value rules out observing the decreased performance in the vested interest group simply by chance, and the randomization of subjects to treatment groups should have balanced out any other factors that might explain the difference. So, the most likely explanation left is that having a vested interest really does decrease task performance. In the “High-salt and Cardiovascular

Disease” study, however, there could be other factors that explain the outcome. For example, it’s reasonable to assume that those with high-salt diets are less health conscious and tend to exercise less. Maybe the proportion with CVD is higher in this group because they exercise less!

Example 4.4: Alcoholism and Depression

Past research has suggested a high rate of alcoholism in families among patients with primary unipolar depression. A study of 210 families of females with primary unipolar depression found that 89 families had alcoholism present. A set of 299 control families found 94 present.

Research Question: Does the proportion of families afflicted by alcoholism differ between those families in which the female has primary unipolar depression and the control group? That is, is the proportion of the Depression group with Alcoholism in the family different from the proportion of the Control group with Alcoholism in the family?

1. Identify the proportion with Alcoholism in both groups.
2. Is there evidence that the proportion of the Depression group with Alcoholism is different from the proportion of the Control group with Alcoholism? Use the JMP output to answer this question.
3. Can we say that having unipolar depression causes alcoholism? Explain your reasoning.

Chi-Square Test (Alternative to Fisher's Exact Test)

Fisher's exact test from the previous section can be used for either upper-tailed, lower-tailed, or two-tailed hypothesis tests for differences in two proportions. The Chi-square test is another procedure we can use to test for differences between two proportions. However, this procedure can be used for ONLY TWO-SIDED TESTS! Consider the data from Example 4.4.

H_o : The proportion of the Depression group with Alcoholism is the same as the proportion of the Control group with Alcoholism.

H_a : The proportion of the Depression group with Alcoholism is different from the proportion of the Control group with Alcoholism.

With the chi-square test, the evidence from our sample will consist of a test statistic, which for this test is given as follows:

$$\text{Test Statistic} = \sum \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}}$$

1. Where do we get the expected counts?

EXPECTED COUNTS HERE

We can also use JMP to calculate the expected counts, deviances, and cell contributions to find the test statistic:

2. What does it mean when the test statistic is "large"?
3. At what point does the test statistic provide evidence to support our research question?

Let's find the p-value!

Recall that the Chi-Square test uses a distribution known as the chi-square (2) distribution. The chi-square distribution takes on only positive numbers and is a continuous distribution. In addition, this distribution is indexed by its degrees of freedom (or df). For this test, this is given by $df = (r - 1)(c - 1)$. When the null hypothesis is true, the test-statistic follows the chi-square distribution with $df = (r - 1)(c - 1)$. .

The following graph shows the chi-square distribution with $df = 1$. The p-value is found by plotting the chi-square test statistic on the x-axis and calculating the area under the curve above the test statistic.

i Assumptions behind the Chi-Square Test:

The chi-square test for independence may be inappropriate for tables with very small expected cell frequencies. One rule of thumb suggests that most of the expected cell frequencies in the table should be 5 or more; otherwise, the chi-square approximation

may not be reliable. JMP and most other statistical software packages will warn you when the results of the chi-square test are suspect. Also, each observation in the study can be classified into only one cell of the contingency table, and the observations must be independent.

Example 4.5: Mice and Tobacco Smoke

Ten mice (6–8 weeks old) were randomly assigned to one of two groups; five were exposed to simulated environmental tobacco smoke for 6 h/day, 5 days/week for 5 months. The other 5 mice were kept in clean air during this time period. Then, all of the mice were allowed to recover for a further 4 months in filtered air before being killed for analysis of lung tumor incidence. The results are shown below.

Research question: Does the proportion of mice that develop a lung tumor differ between those exposed to tobacco smoke and the control group?

1. Convert the Research Question into H_o and H_a .

H_o : The proportion of mice that develop a lung tumor when exposed to tobacco does not differ from the proportion that develop a lung tumor when not exposed to tobacco.

H_a : The proportion of mice that develop a lung tumor differs between those exposed to tobacco smoke and those not exposed.

Equivalently, we can state the hypotheses as follows:

H_o :

H_a :

Warning

Keep the assumptions behind the chi-square test in mind:

The chi-square test for independence may be inappropriate for tables with very small expected cell frequencies. One rule of thumb suggests that most of the expected cell frequencies in the table should be 5 or more; otherwise, the chi-square approximation may not be reliable.

Also, each observation in the study can be classified into only one cell of the contingency table, and the observations must be independent.

3. Do the assumptions behind the chi-square test appear to be met? Explain.
4. Find the p-value for the data
5. Write a conclusion in terms of the original research question.