

December 2017

# Effectively Comparing Differences in Proportions

Lonnie Turpin Jr.

*McNeese State University, lturpin@mcneese.edu*

 Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

---

## Recommended Citation

Turpin, L. (2017). Effectively Comparing Differences in Proportions. *Journal of Modern Applied Statistical Methods*, 16(2), 186-199.  
doi: 10.22237/jmasm/1509495000

# Effectively Comparing Differences in Proportions

**Lonnie Turpin**  
McNeese State University  
Lake Charles, LA

---

A single framework of developing and implementing tests about proportions is outlined. It avoids some of the pitfalls of methods commonly put forward in an introductory data analysis course.

*Keywords:* Binary data analysis, proportions, significance, inference

---

## Introduction

Proportions derived from binary variables are simple predominantly due to the nature of the variables involved. Because of this, the logic behind the methods are able to be grasped, as opposed to resorting to memorizing formulas. However, confusion arises when making a connection within the equations between a Bernoulli random variable  $X$  and the associated estimator  $\hat{p}$  of the sample proportion  $p$ . A way to mitigate this confusion will be shown where only a basic knowledge of descriptive/inferential statistics and linear combinations are required. With all necessary formulations included, this study is essentially self-contained and aimed at analysts (practitioners and teachers focusing on applications who need a quick guide for analyzing proportions).

A subject can represent any object of analysis (people, products, etc.) and the method refers to the two examples used to illustrate proportions, not the statistical technique used in analysis. A single framework of developing and implementing tests about proportions will be outlined, which avoids some of the pitfalls of methods commonly put forward in introductory classes. The methods will entail using the simple two-way probability table to easily calculate confidence intervals and  $Z$ -scores while accounting for the correlation between the proportions. The key contributions are: 1) to make use of this simple two-way table as an easy way to

---

*Dr. Lonnie Turpin is an Assistant Professor of Operations Management and Business Statistics. Email him at: [lturpin@mcneese.edu](mailto:lturpin@mcneese.edu).*

compute the correlation (which is often ignored in intro level treatments); and 2) to motivate the methods in a way that makes use of the student's intuition about how sample means relate to population means. This study was motivated by experience teaching graduate-level applied business statistics courses. Common points of confusion are highlighted as remarks. For theoretical expositions on the topics presented, Brownlee (1965) and Bickel and Doksum (2015) are recommended.

## Comparing Proportions

Begin with an example that involves comparing the effectiveness of two different methods of testing a product in order to provide a single method of comparing two proportions. An illustrative example concerns comparing the effectiveness of two methods. The methods are denoted as  $M1$  (a chosen method 1) and  $M2$  (a chosen method 2 that is different from method 1). It is a loose assumption that there exists an initial method of comparison ( $M0$ ) with a known statistical measure of effectiveness. However, it is not necessary for the comparison of the test methods ( $M1$  and  $M2$ ), and this is therefore omitted for the remainder of the paper. Both of these methods will be given to a group of  $n$  test subjects. For a classic example of comparing two proportions from an identical survey, see Scott and Seber (1983) and Wild and Seber (1993).

### Procedure

**Assumption 1:** To control for the potential individual-specific factors, each subject in the test group is given both methods. Practical outlets for this assumption are in clinical trials where each subject receives both the treatment and the control (Senn, 2002).

For each subject  $i$ , define  $X_i^M = 1$  if method  $M$  is successful and 0 otherwise. Prior to beginning the experiment, it is important to note that  $X_i^M$  is a random variable, since we do not know in advance whether the method will work for any given subject.

**Assumption 2:** Outcomes for different subjects are independent and identically distributed (i.i.d.).

Given this assumption, and based on the nature of the variables, we let the distribution of each variable expressed in terms of  $p$  as

## EFFECTIVELY COMPARING DIFFERENCES IN PROPORTIONS

$$X_i^M \sim \text{Bernoulli}(p^M) \text{ i.i.d.} \quad (1)$$

where  $p^M$  is the familiar probability  $\Pr(X_i^M) = 1$  representing the success of a given method for the  $i^{\text{th}}$  subject. Let  $\hat{p}^M$  be the fraction of subjects for which method  $M$  is successful. For dummy variables, let

$$\hat{p}^M = \frac{1}{n} \sum_{i=1}^n X_i^M \quad (2)$$

be the estimator for  $p^M$ , representing the sample proportion of ones contained in the data set.

**Assumption 3:** The sample size  $n$  is large enough to follow the Central Limit Theorem, thus define the sampling distribution of  $\hat{p}^M$  as

$$\hat{p}^M \sim N\left(p^M, \frac{p^M(1-p^M)}{n}\right) \quad (3)$$

For effective techniques in working with smaller data sets (especially in teaching the logic behind these techniques), see Agresti and Caffo (2000).

Notice that in Assumption 3, the sampling distribution of  $\hat{p}^M$  depends on the unknown parameter  $p^M$ . With a large  $n$ , by equation (2), rewrite equation (3) in the form

$$\hat{p}^M \approx N\left(\hat{p}^M, \frac{\hat{p}^M(1-\hat{p}^M)}{n}\right) \quad (4)$$

Replace  $p$  with  $\hat{p}$  in the equations to follow with the understanding that  $\hat{p} \approx p$  by Assumption 3 and equation (4). Then, define the mean and variance of  $X_i^M$  in terms of  $p^M$  and  $\hat{p}^M$  as

$$\begin{aligned} E[X_i^M] &= p^M \\ &\approx \hat{p}^M \end{aligned} \quad (5)$$

and

$$\begin{aligned} V[X_i^M] &= p^M(1-p^M) \\ &\approx \hat{p}^M(1-\hat{p}^M) \end{aligned} \tag{6}$$

Now, figure out which of the two methods is more effective. That is, the goal is to find and explain the difference  $p^{M1} - p^{M2}$ . To do this, think about the two methods together. Considering the potential unique nature of the subjects highlighted by Assumption 1, the following assumption is introduced:

**Assumption 4:** Although it was assumed the outcomes are i.i.d. across different subjects, do not assume the two outcomes  $X_i^{M1}$  and  $X_i^{M2}$  are independent for the same subject. Because there is no assumption that the two methods are independent, the question arises how strongly they are related. Use the familiar covariance formula

$$\begin{aligned} \text{Cov}[X_i^{M1}, X_i^{M2}] &= \sum_{X_i^{M1}, X_i^{M2} \in \{0,1\}} \Pr(X_i^{M1}, X_i^{M2})(X_i^{M1} - E[X_i^{M1}])(X_i^{M2} - E[X_i^{M2}]) \end{aligned} \tag{7}$$

where the weights  $\Pr(X_i^{M1}, X_i^{M2})$  are the joint probabilities easily derived using the two-way Table 1. In Appendix B, the numerical example is used to show a connection between covariance and the assumption of independence via Bayes' Rule.

**Table 1.** Two-way joint probability table

		$X_i^{M1}$		$\Pr(X_i^{M2})$
		0	1	
$X_i^{M2}$	0	$\Pr(0, 0)$	$\Pr(1, 0)$	$1 - p^{M2}$
	1	$\Pr(0, 1)$	$\Pr(1, 1)$	$p^{M2}$
$\Pr(X_i^{M1})$		$1 - p^{M1}$	$p^{M1}$	

## EFFECTIVELY COMPARING DIFFERENCES IN PROPORTIONS

**Table 2.** Two-way joint probability table with  $\hat{p}^M$  estimators

		$X_i^{M1}$		$\Pr(X_i^{M2})$
		0	1	
$X_i^{M2}$	0	Pr(0,0)	Pr(1,0)	$1 - \hat{p}^{M2}$
	1	Pr(0,1)	Pr(1,1)	$\hat{p}^{M2}$
$\Pr(X_i^{M1})$		$1 - \hat{p}^{M1}$	$\hat{p}^{M1}$	

Replace the unknown  $p^{M1}$ ,  $p^{M2}$ , and  $\Pr(X_i^{M1}, X_i^{M2})$  Table 1 with the estimators  $\hat{p}^{M1}$ ,  $\hat{p}^{M2}$ , and  $\Pr(X_i^{M1}, X_i^{M2})$  in the modified Table 2.

As with equation (5) and equation (6), rewrite equation (7) in terms of  $p$  as

$$\begin{aligned}
 & \text{Cov}[X_i^{M1}, X_i^{M2}] \\
 &= \sum_{X_i^{M1}, X_i^{M2} \in \{0,1\}} \Pr(X_i^{M1}, X_i^{M2}) (X_i^{M1} - E[X_i^{M1}]) (X_i^{M2} - E[X_i^{M2}]) \\
 &= \sum_{X_i^{M1}, X_i^{M2} \in \{0,1\}} \Pr(X_i^{M1}, X_i^{M2}) (X_i^{M1} - p^{M1}) (X_i^{M2} - p^{M2}) \\
 &\approx \sum_{X_i^{M1}, X_i^{M2} \in \{0,1\}} \Pr(X_i^{M1}, X_i^{M2}) (X_i^{M1} - \hat{p}^{M1}) (X_i^{M2} - \hat{p}^{M2})
 \end{aligned} \tag{8}$$

where  $\Pr(X_i^{M1}, X_i^{M2})$  and  $\text{Cov}[X_i^{M1}, X_i^{M2}]$  are estimates for  $\Pr(X_i^{M1}, X_i^{M2})$  and  $\text{Cov}[X_i^{M1}, X_i^{M2}]$ , respectively.

**Remark 1:** The covariance of the methods are represented in the familiar probabilistic form as  $\text{Cov}[M1, M2] = \sigma_{M1, M2}$ . Let  $\text{Cov}[X_i^{M1}, X_i^{M2}] = \sigma_{M1, M2}$ . Like  $p^{M1}$  and  $p^{M2}$ , treat  $\sigma_{M1, M2}$  as an unknown parameter.

Following Remark 1, for a particular subject  $i$ , define the mean and variance of  $X_i^{M1}$  and  $X_i^{M2}$  in two steps just as  $\text{Cov}[X_i^{M1}, X_i^{M2}]$  in equation (7) and equation (8). First, use the linear formulas to write  $E[X_i^{M1} - X_i^{M2}]$  and

$V[X_i^{M1} - X_i^{M2}]$  in terms of the means, variances, and the covariance of  $X_i^{M1}$  and  $X_i^{M2}$  as

$$E[X_i^{M1} - X_i^{M2}] = E[X_i^{M1}] - E[X_i^{M2}] \quad (9)$$

and

$$V[X_i^{M1} - X_i^{M2}] = V[X_i^{M1}] + V[X_i^{M2}] - 2(\text{Cov}[X_i^{M1}, X_i^{M2}]) \quad (10)$$

Then, figure out what each mean, variance, and covariance is in terms of  $p^{M1}$ ,  $p^{M2}$ , and  $\sigma_{M1,M2}$  as

$$\begin{aligned} E[X_i^{M1} - X_i^{M2}] &= E[X_i^{M1}] - E[X_i^{M2}] \\ &= p^{M1} - p^{M2} \\ &\approx \hat{p}^{M1} - \hat{p}^{M2} \end{aligned} \quad (11)$$

and

$$\begin{aligned} V[X_i^{M1} - X_i^{M2}] &= V[X_i^{M1}] + V[X_i^{M2}] - 2(\text{Cov}[X_i^{M1}, X_i^{M2}]) \\ &= p^{M1}(1 - p^{M1}) + p^{M2}(1 - p^{M2}) - 2\sigma_{M1,M2} \\ &\approx \hat{p}^{M1}(1 - \hat{p}^{M1}) + \hat{p}^{M2}(1 - \hat{p}^{M2}) - 2\sigma_{M1,M2} \end{aligned} \quad (12)$$

This leads to estimating the mean and variance of  $\hat{p}^{M1} - \hat{p}^{M2}$ . The mean tends to cause less confusion than the variance since a very simple connection can be made in the form of the average difference

$$\hat{p}^{M1} - \hat{p}^{M2} = \frac{1}{n} \sum_{i=1}^n (X_i^{M1} - X_i^{M2}) \quad (13)$$

**Remark 2:** The benefit of equation (13) is that it follows from a bit of basic algebra, so a formal proof is not needed. The analyst only needs to recall that the expected difference in two variables is the difference in expected values. For the

## EFFECTIVELY COMPARING DIFFERENCES IN PROPORTIONS

variance, it is important to recall the connection between the variance of averages and individual values.

To achieve Remark 2, recognize that the average simply equals  $1/n$  times the sum. Therefore,  $1/n$  acts as a constant. When a random variable is multiplied by a constant, the mean gets multiplied by the same constant and the variance gets multiplied by that constant squared.

$$\begin{aligned} \mathbf{V}[\hat{p}^{M1} - \hat{p}^{M2}] &= \mathbf{V}\left[\frac{1}{n} \sum_{i=1}^n (X_i^{M1} - X_i^{M2})\right] \\ &= \left(\frac{1}{n}\right)^2 n \mathbf{V}[X_i^{M1} - X_i^{M2}] \\ &= \frac{\mathbf{V}[X_i^{M1} - X_i^{M2}]}{n} \end{aligned} \quad (14)$$

Then, the difference in means can be estimated as

$$\begin{aligned} \mathbf{E}[\hat{p}^{M1} - \hat{p}^{M2}] &= \mathbf{E}[\hat{p}^{M1}] - \mathbf{E}[\hat{p}^{M2}] \\ &= p^{M1} - p^{M2} \\ &\approx \hat{p}^{M1} - \hat{p}^{M2} \end{aligned} \quad (15)$$

and the variance as

$$\begin{aligned} \mathbf{V}[\hat{p}^{M1} - \hat{p}^{M2}] &= \frac{\mathbf{V}[X_i^{M1} - X_i^{M2}]}{n} \\ &\approx \frac{\hat{p}^{M1}(1 - \hat{p}^{M1}) + \hat{p}^{M2}(1 - \hat{p}^{M2}) - 2\sigma_{M1,M2}}{n} \end{aligned} \quad (16)$$

**Remark 3:**  $p^{M1}$ ,  $p^{M2}$ , or  $\text{Cov}[X_i^{M1}, X_i^{M2}]$  is not known, and the values of  $\hat{p}^{M1}$ ,  $\hat{p}^{M2}$ , and  $\text{Cov}[X_i^{M1}, X_i^{M2}]$  are just approximations even though the sample size  $n$  is such that  $\hat{p}^{M1} \approx p^{M1}$  and  $\hat{p}^{M2} \approx p^{M2}$  by Assumption 3.



### Testing

Suppose the intent is test the null hypothesis  $H_0: p^{M1} = p^{M2}$ , which says the two methods are equally effective. To clarify this point, another way to write the null hypothesis is  $H_0: p^{M1} - p^{M2} = 0$ . However, the consequence of Remark 3 is there is no guarantee of the equivalency  $\hat{p}^{M1} - \hat{p}^{M2} = p^{M1} - p^{M2}$ . By all preceding logic, at best  $\hat{p}^{M1} - \hat{p}^{M2} \approx p^{M1} - p^{M2}$ , which still leaves unknown values for  $p^{M1}$  and  $p^{M2}$ . Therefore, construct a 95% confidence interval for the difference  $p^{M1} - p^{M2}$  as

$$(\hat{p}^{M1} - \hat{p}^{M2}) \pm 1.96 \sqrt{V[\hat{p}^{M1} - \hat{p}^{M2}]} \quad (17)$$

which follows from considerations about the weight in the tails of the standard normal distribution. Two may be used instead of the usual 1.96 in equation (17). If the interval does not contain 0, the null hypothesis is rejected at the usual 5% level (Bickel & Doksum, 2015).

The hypothesis could also be tested using the Z-score

$$Z = \frac{\hat{p}^{M1} - \hat{p}^{M2}}{\sqrt{V[\hat{p}^{M1} - \hat{p}^{M2}]}} \quad (18)$$

where  $Z$  represents the number of standard errors,  $\sqrt{V[\hat{p}^{M1} - \hat{p}^{M2}]}$ , the estimate,  $\hat{p}^{M1} - \hat{p}^{M2}$ , and the null, 0, are from each other. In the event that  $|Z| > 1.96$ , then based on the data, there is evidence that one method is more effective than the other. For an example of the power of the traditional Z-test in comparing Bernoulli proportions, see Suissa and Shuster (1984). An advantage of these tests is that the confidence interval is always consistent with the hypothesis test decision for a two-tailed test. A formal connection is provided in Appendix A.

**Remark 4:** Although the method of examining the overlap between two confidence intervals is a recognized technique, avoid using this method in formal significance testing; for justification, see Schenker and Gentleman (2001). This is highlighted the following example:

## EFFECTIVELY COMPARING DIFFERENCES IN PROPORTIONS

### Example

Suppose the success of two types of nail polish removal methods are tested on  $n = 100$  people. Each participant has the exact type of nail polish on each testable finger. For each subject  $i$ , define  $X_i^{M1} = 1$  if method  $M$  is successful (that is, reducing the target amount of polish on the applied nail within a given timeframe) and 0 otherwise. As the aim of this article is for statistical clarity more than method content, for brevity assume this can be nail polish remover of any type (acetone-based vs. acetone-free, remover strips vs. soaking, etc.). The uniqueness of people helps make clear Assumptions 1 and 4. To satisfy Assumption 1, we can apply Method 1 ( $M1$ ) to any right-hand finger and Method 2 ( $M2$ ) to the corresponding finger on the left hand simultaneously. The data on each of the 100 participants, including Excel functions, is shown in the supplementary material. Initially, use the logic of Table 2 to complete Table 3.

The mean values given in equation (5) for each variable are shown in the table and, by equation (1), note  $X_i^{M1} \sim \text{Bernoulli}(0.65)$  and  $X_i^{M2} \sim \text{Bernoulli}(0.49)$ . The corresponding variances given in equation (6) are derived as  $\hat{p}^{M1}(1-\hat{p}^{M1}) = 0.2275$  and  $\hat{p}^{M2}(1-\hat{p}^{M2}) = 0.2499$ . Notice this is multiplying the two marginal probabilities  $\Pr(X_i^M = 1) \times \Pr(X_i^M = 0)$  for each method  $M$ . By equation (8), calculate the covariance as

$$\begin{aligned}
 \text{Cov}[X_i^{M1}, X_i^{M2}] &= \Pr(0,0)(0-0.65)(0-0.49) + \Pr(1,0)(1-0.65)(0-0.49) \\
 &\quad + \Pr(0,1)(0-0.65)(1-0.49) + \Pr(1,1)(1-0.65)(1-0.49) \\
 &= 0.0615
 \end{aligned} \tag{19}$$

**Table 3.** Two-way joint probability table for the persuasion example

		$X_i^{M1}$		$\Pr(X_i^{M2})$
		0	1	
$X_i^{M2}$	0	$\Pr(0,0) = 0.24$	$\Pr(1,0) = 0.27$	$1 - \hat{p}^{M2} = 0.51$
	1	$\Pr(0,1) = 0.11$	$\Pr(1,1) = 0.38$	$\hat{p}^{M2} = 0.49$
$\Pr(X_i^{M1})$		$1 - \hat{p}^{M1} = 0.35$	$\hat{p}^{M1} = 0.65$	

Then, calculate the variance of the difference by equation (15) as

$$\begin{aligned} V[\hat{p}^{M1} - \hat{p}^{M2}] &= \frac{0.65(0.35) + 0.49(0.51) - 2(0.0615)}{100} \\ &= 0.0035 \end{aligned} \quad (20)$$

Using equation (20), the interval (17) for the difference  $\hat{p}^{M1} - \hat{p}^{M2}$  is calculated to be (0.0409, 0.2791). Notice it does not contain 0, reject the null, and conclude the proportions are significantly different. In this case,  $M1$  is more effective.

To highlight Remark 4, consider the interval

$$\hat{p}^M \pm 1.96 \sqrt{\frac{\hat{p}^M (1 - \hat{p}^M)}{n}} \quad (21)$$

representing the 95% confidence interval for a single sample  $\hat{p}^M$ . The interval in (21) is the familiar nominal 95% confidence interval shown in Brownlee (1965). Should Assumption 4 be violated, the covariance would be 0, thereby eliminating the covariance term in equation (16). This condition (via the assumption of independence) mirrors the calculation for  $V[\hat{p}^{M1} - \hat{p}^{M2}]$  in Brownlee, and is used to compare proportions discussed in Schenker and Gentleman (2001). Applying the interval (21) to the proportions  $\hat{p}^{M1}$  and  $\hat{p}^{M2}$  results in (0.5546, 0.7454) and (0.3900, 0.5899), respectively. By the overlap method, conclude that the proportions are not significantly different resulting in a contradiction with our preceding analysis.

## Conclusion

Proportions are a key part of applied statistics and merit the attention of useful guides in clarifying common techniques. A classic technique was discussed for comparing proportions with a target audience of students and practitioners (and to a degree, teachers) dealing with statistics of everyday life. A step-by-step procedure was presented for analysis to mitigate the confusion of the audience when making a connection within the equations between a Bernoulli random variable  $X$  and the associated proportion  $p$  (as well as its respective estimator  $\hat{p}$ ). This procedure culminated with a basic example where we showed the connection between

## EFFECTIVELY COMPARING DIFFERENCES IN PROPORTIONS

confidence intervals and hypothesis testing and highlighted the deficiencies of relying on overlapping confidence intervals as a means of inference. This article should serve as a good secondary reference analysts who needs to not only apply statistical procedures to their research, but also to appreciate the basic connections within them.

### Acknowledgements

The authors would like to thank the anonymous reviewers for their detailed and constructive comments. We would also like to thank Dr. William Galose and Morgan Turpin for their discussions involving the logical framework and examples presented in this research.

### References

- Agresti, A., & Caffo, B. (2000). Simple and effective confidence intervals for proportions and differences of proportions result from adding two successes and two failures. *The American Statistician*, 54(4), 280-288. doi: 10.2307/2685779
- Bickel, P. J., & Doksum, K. A. (2015). *Mathematical statistics: Basic ideas and selected topics* (2nd ed.) (Vol. 2). Boca Raton, FL: CRC Press.
- Brownlee, K. A. (1965). *Statistical theory and methodology: In science and engineering* (2nd ed.). New York: John Wiley & Sons.
- Schenker, N., & Gentleman, J. F. (2001). On judging the significance of differences by examining the overlap between confidence intervals. *The American Statistician*, 55(3), 182-186. doi: 10.1198/000313001317097960
- Scott, A. J., & Seber, G. A. (1983). Difference of proportions from the same survey. *The American Statistician*, 37(4), 319-320. doi: 10.2307/2682774
- Senn, S. (2002). *Cross-over trials in clinical research* (2nd ed.). Chichester, UK: John Wiley & Sons. doi: 10.1002/0470854596
- Suissa, S., & Shuster, J. J. (1984). Are uniformly most powerful unbiased tests really best? *The American Statistician*, 38(3), 204-206. doi: 10.2307/2683654
- Wild, C. J., & Seber, G. A. F. (1993). Comparing two proportions from the same survey. *The American Statistician*, 47(3), 178-181. doi: 10.2307/2684972

## Appendices

In the following appendices, we first present a formal connection of confidence intervals and hypothesis tests; then, via Bayes' Rule, we show a connection between covariance and the assumption of independence.

### Appendix A

The confidence interval and hypothesis test are two ways of saying what we think about the true value of the unknown difference  $p^{M1} - p^{M2}$ . To make clear this notion, recall the formula for the 95% confidence interval in (17) expressed in its equivalency as

$$\left(\hat{p}^{M1} - \hat{p}^{M2}\right) \pm 1.969 \text{ SE} \quad (\text{A22})$$

Now consider the formula for the test statistic in equation (18) in its comparative form

$$Z = \frac{\hat{p}^{M1} - \hat{p}^{M2}}{\text{SE}} \quad (\text{A23})$$

allowing us to compare two values for  $p^{M1} - p^{M2}$ . These two values are the value we guessed, 0 (from  $H_0: p^{M1} - p^{M2} = 0$ ), and the value we actually estimated from our data,  $\hat{p}^{M1} - \hat{p}^{M2}$ . The difference between the two values gets divided by the standard error  $\sqrt{V[\hat{p}^{M1} - \hat{p}^{M2}]}$ , labeled simply as SE.

Just like with the single sample hypothesis test, we want to calculate the number of standard errors away from the null hypothesis value our estimate actually is. So if  $\hat{p}^{M1} - \hat{p}^{M2}$  and 0 are more than 1.96 standard errors apart, we will get a Z-score greater than 1.96 and will reject the null at the 5% level. Now, recall that the 95% confidence interval contains all the values within 1.96 standard errors of  $\hat{p}^{M1} - \hat{p}^{M2}$ . If our guess, 0, lies outside the 95% confidence interval, we will reject the null.

## Appendix B

There is a positive relationship between the two methods by the sign of the covariance 0.0615 calculated in equation (19). We can verify this by Bayes' Rule

$$\Pr(X_i^{M2} | X_i^{M1}) = \frac{\Pr(X_i^{M2}, X_i^{M1})}{\Pr(X_i^{M1})} \quad (\text{B24})$$

where  $\Pr(X_i^{M2} | X_i^{M1})$  represents the conditional distribution and  $\Pr(X_i^{M2}, X_i^{M1})$  and  $\Pr(X_i^{M1})$  each represent the joint and marginal distributions discussed previously. We could also structure equation (B1) with respect to  $X_i^{M2}$  by solving for  $\Pr(X_i^{M2} | X_i^{M1})$ . Using the data from the supplemental material, we estimate

$$\begin{aligned} \Pr(X_i^{M2} = 1 | X_i^{M1} = 1) &= \frac{\Pr(X_i^{M2} = 1, X_i^{M1} = 1)}{\Pr(X_i^{M1} = 1)} \\ &= \frac{0.38}{0.65} \\ &= 0.5846 \end{aligned} \quad (\text{B25})$$

With equation (B2) yielding  $0.5846 \approx 0.58$ , we can now compare the marginal distribution  $X_i^{M2} \sim \text{Bernoulli}(0.49)$  to the conditional distribution  $X_i^{M2} | X_i^{M1} \sim \text{Bernoulli}(0.58)$ . Notice that we didn't really need to calculate  $\Pr(X_i^{M2} = 0 | X_i^{M1} = 1)$  since

$$\Pr(X_i^{M2} = 1 | X_i^{M1} = 1) + \Pr(X_i^{M2} = 0 | X_i^{M1} = 1) = 1$$

by the definition of a distribution. Thus, the marginal distribution  $\Pr(X_i^{M2})$  and the conditional distribution  $\Pr(X_i^{M2} | X_i^{M1} = 1)$  are not the same. The distribution of  $X_i^{M2}$  depends on what we observe for  $X_i^{M1}$ . Therefore, they are not independent, validating Assumption 4.

Given

## LONNIE TURPIN

$$\Pr(X_i^{M2}) \neq \Pr(X_i^{M2} | X_i^{M1} = 1)$$

it is easily inferred that

$$\Pr(X_i^{M2}) \neq \Pr(X_i^{M2} | X_i^{M1} = 0)$$

Thus, the conditional distributions  $\Pr(X_i^{M2} | X_i^{M1} = 1)$  and  $\Pr(X_i^{M2} | X_i^{M1} = 0)$  are not the same. They each depend on what we observe for  $X_i^{M1}$ .

To help see the how  $X_i^{M1}$  and  $X_i^{M2}$  are positively related, we compare the conditional distributions  $\Pr(X_i^{M2} = 1 | X_i^{M1} = 1)$  and  $\Pr(X_i^{M2} = 1 | X_i^{M1} = 0)$  to  $\Pr(X_i^{M2} = 0 | X_i^{M1} = 1)$  and  $\Pr(X_i^{M2} = 0 | X_i^{M1} = 0)$ . Applying equation (A2), we get the following calculations:

$$\Pr(X_i^{M2} = 0 | X_i^{M1} = 0) = 0.6857 \approx 0.69$$

$$\Pr(X_i^{M2} = 1 | X_i^{M1} = 0) = 0.3143 \approx 0.31$$

$$\Pr(X_i^{M2} = 0 | X_i^{M1} = 1) = 0.4154 \approx 0.42$$

$$\Pr(X_i^{M2} = 1 | X_i^{M1} = 1) = 0.5846 \approx 0.58$$

Notice that conditional on  $X_i^{M1}$  being small (large), the probabilities get larger for  $X_i^{M2}$  when it is also small (large). Therefore, they are positively related.