Sample Term Project

Lauren Cappiello

December 2020

Introduction

For the past seven years, my husband and I have been getting together with our closest friends for a "Friendsgiving" celebration. This generally occurs on a Saturday in November and is primarily an excuse to drink wine and eat way too much with friends. Surely such a delightful tradition is not to be missed... or is it? I set out to find out about people's participation in Friendsgiving celebrations. I decided to examine whether a majority of people attend Friendsgiving celebrations. I also thought these celebrations might be more popular among a younger crowd.

Methods

Data was collected using SurveyMonkey, a website which allows users to create and administer surveys[1]. The target population was US adults. The survey was advertised on social media and ultimately received 1058 responses. After cleaning up the data, there were 947 responses. The questions asked were

1. Do you celebrate Thanksgiving? (Yes/No)
2. Have you ever attended a Friendsgiving? (Yes/No)
3. What is your age?
   (a) 18-29
   (b) 30-44
   (c) 45-59

[1]Since I'm the professor and this is just a sample project, I did not personally collect these data.

(d) 60+

An online survey was chosen to maximize number of responses and to increase the geographic reach of the survey. I grouped the data into ages 18-44 and 45+. I chose this cutoff to be reasonably close to grouping the data into Gen Z/Millennials and Gen X/Baby Boomers[2].

Originally, I hoped to examine whether people who celebrate Thanksgiving are more likely to celebrate Friendsgiving. However, no respondents answered "no" to Question 1. Therefore, this analysis considers only Friendsgiving attendance among US adults who celebrate Thanksgiving.

Data was analyzed using a confidence interval for the proportion of US adults who have attended a Friendsgiving celebration. I also used a one-sample hypothesis test for a proportion, since an examination of the p-value was of additional interest.

Next, I created a confidence interval for the difference in proportions between US adults who have attended a Friendsgiving celebration and are in the 18-44 group and those in the 45+ group. I also conducting a two-sample hypothesis test for a difference of proportions, again to examine the p-value.

Results

Of the 951 respondents, the mode response to Question 2 was "No": only respondents 268 had ever attended a Friendsgiving celebration. For the age groups, the mode was 45-59 with a frequency of 269. When the ages were regrouped into 18-44 and 45+, the mode was 45+ with a frequency of 527.

		Friendsgiving		
		Yes	No	Total
Age Group	18-44	162	258	420
	45+	105	422	527
	Total	267	680	947

		Friendsgiving		
		Yes	No	Total
Age Group	18-44	0.17	0.27	0.44
	45+	0.11	0.45	0.56
	Total	0.28	0.72	1

Table 1: Frequency and relative frequency tables for US adults, broken down by Friendsgiving celebration attendance and age group.

The first test was of overall Friendsgiving attendance. I wanted to examine whether a majority of US adults had attended a Friendsgiving celebration or not.

[2]The cutoff is inexact because the original age groups were determined by whoever originally collected these data.

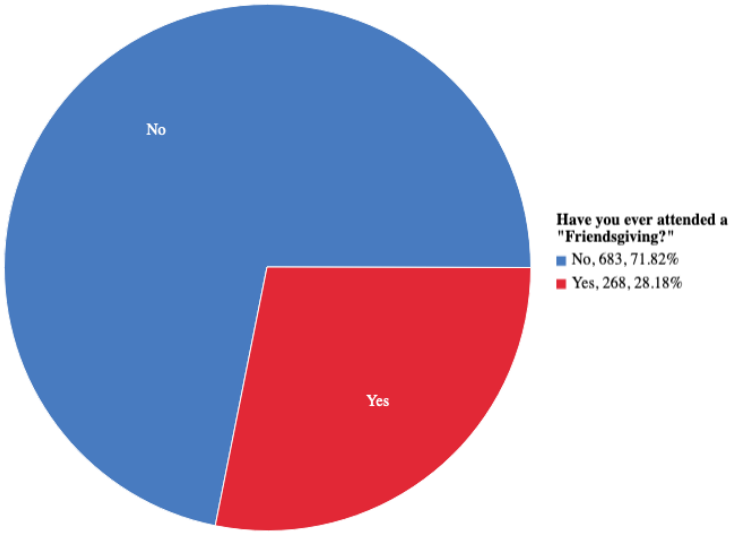

Figure 1: Pie chart of responses to Question 2. 71.82% of survey respondents have not attended a Friendsgiving.

The sample proportion of US adults who have attended a Friendsgiving celebration was 267/947 ($\hat{p} = 0.28$).

$$\begin{aligned} \mathrm{H}_0 &: p = 0.5 \\ \mathrm{H}_A &: p \neq 0.5 \end{aligned}$$

To confirm this test was appropriate, I checked the success-failure condition for the smaller of $\hat{p}$ and $1 - \hat{p}$:

$$947 \times 0.28 = 265.16 \geq 10$$

and concluded that the Central Limit Theorem was satisfied and a test using the normal distribution was appropriate.

Based on these data, a 95% confidence interval for p is (0.253,0.311). It is reasonable to conclude that the true proportion of US adults who have participated in a Friendsgiving celebration is between 25.3% and 31.3%. The corresponding hypothesis test has test statistic $z = -13.42$ with corresponding p-value 4.57×10^{-41} ($\alpha = 0.05$). There is approximately 0% chance that, given the present data, the true proportion of US adults who have attended a Friendsgiving is 50%. In fact, these data suggest that the true proportion is less than 50%.

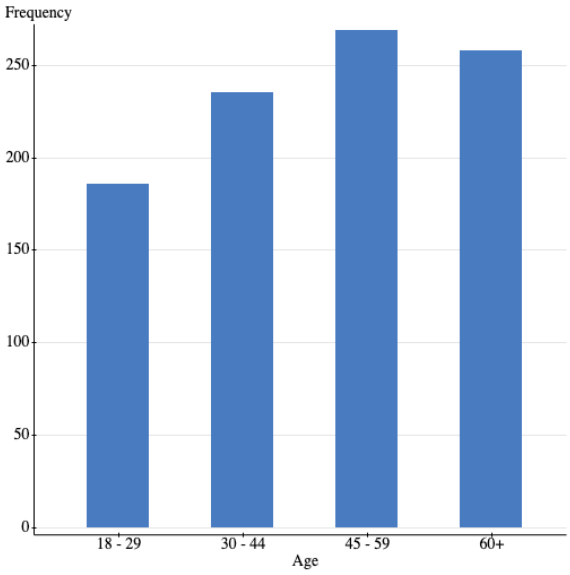

Figure 2: Bar plot of the original age groups used in the survey.

The second test was designed to examine whether younger US adults are more or less likely to attend Friendsgiving celebrations. Of those in the 18-44 group, 162/420 had attended a Friendsgiving celebration ($\hat{p}_1 = 0.386$); in the 45+ group, 105/527 had attended a Friendsgiving ($\hat{p}_2 = 0.199$).

$$\begin{aligned} \mathrm{H}_0 &: p_1 = p_2 \\ \mathrm{H}_A &: p_1 \neq p_2 \end{aligned}$$

To confirm this test was appropriate, I checked the success-failure condition for the smaller of $\hat{p}_1$ and $1 - \hat{p}_1$:

$$420 \times 0.386 = 162.12 \geq 10$$

and the smaller of $\hat{p}_2$ and $1 - \hat{p}_2$:

$$527 \times 0.199 = 104.87 \geq 10$$

and concluded that the Central Limit Theorem was satisfied and a test using the normal distribution was appropriate.

A 95% confidence interval for the difference in the proportion of Friendsgiving-goers between the two age groups is (0.129, 0.244). We can be 95% confident that the difference in proportions between ages 18-44 and ages 45+ is between 21.9% and

24.4%. The test statistic for the difference of these proportions is $z = 6.34$ with a corresponding p-value of 2.36×10^{-10} ($\alpha = 0.05$). There is essentially a 0% chance that the proportion of Friendsgiving-goers in the 18-44 group is the same as that of the 45+ group. Based on this test and the 95% confidence interval, we can conclude at the 0.05 level of significance that the younger US adults (those in the 18-44 group) are more likely to have attended a Friendsgiving celebration.

Discussion

It is reasonable to assume that most people do not attend Friendsgiving celebrations. There are several possible explanations for this. First, some people may celebrate Thanksgiving with friends and do not consider that a dedicated "Friendsgiving". Friendsgiving is also a significant effort for those who host, especially if they also host a Thanksgiving meal.

I also found that it is reasonable to assume that Friendsgiving is more prevalent among younger US adults, specifically among Millennials and Generation Z. It may be that Friendsgiving is a more modern idea. Younger generations may have fewer resources to travel to visit extended family and so may choose to spend time locally with friends. Younger generations may also be less likely to have children, which may impact their holiday decisions.

It is worth noting that I was not able to answer my initial questions directly, nor was I able to adequately address my target population. All survey respondents celebrated Thanksgiving, so these data only tell us about the population of US adults who celebrate Thanksgiving. It does not capture the behaviors of those who do not celebrate Thanksgiving. I also wanted to answer questions about those who do and do not celebrate Friendsgiving. However, the question posed (Question 2) was "Have you ever attended a Friendsgiving?". This question should capture everyone who celebrates regularly, but would also include those who attended once and chose not to continue with Friendsgiving celebrations.

Conclusions

Based on the results and limitations of this study, if I were to do this again I would make several changes. First, I would try to identify a significant population of US adults who do not celebrate Thanksgiving. I would also change the questions to ask whether respondents celebrate Friendsgiving instead of whether they have ever attended a Friendsgiving. This would give me a better picture of those who

actively celebrate. Finally, I would change the age groups to break things down along generation lines, separating Millenials and Generation Z from older generations[3].

Since I found that most people do not celebrate Friendsgiving and that younger generations are more likely to celebrate, a follow up study might examine the attitudes of Generation Z and Millennials toward Thanksgiving and Friendsgiving. This could help to clarify some of the speculation in the previous section. Specifically, follow up questions could help identify why people choose to celebrate - or not celebrate - Friendsgiving.

It might also be of interest to examine attitudes toward Thanksgiving and how these impact attitudes toward Friendsgiving. Some possible follow up questions are: As people become increasingly aware of the history behind Thanksgiving, how will that impact attitudes toward Friendsgiving celebrations? Is Friendsgiving a viable replacement for those who might not want to celebrate Thanksgiving, or is it similar in ways that might make it fall out of favor with those who prefer not to celebrate Thanksgiving?

Appendix

The data for this sample project may be found here:
https://www.statcrunch.com/app/index.php?dataid=2833634# All graphs were created in StatCrunch. The analysis was done in `R`. The code is shown below[4]:

```
tday <- read.csv("~/Downloads/Thanksgiving_2015_Poll_Data.csv")
summary(tday[,1])
fg <- tday$Have.you.ever.attended.a.Friendsgiving.
age <- tday$Age
dat <- data.frame(fg,age)
dat <- dat[-which(dat$fg=="" | dat$age==""),]
dat$age <- as.factor(ifelse(dat$age=="18 - 29" |
                              dat$age=="30 - 44", "18 - 44", "45+"))

sum(dat$fg=="Yes" & dat$age=="18 - 44")
sum(dat$fg=="Yes" & dat$age=="45+")
sum(dat$fg=="No" & dat$age=="18 - 44")
```

[3]Generation Z is born 1997 and earlier; Millennials 1981-1996; Generation X 1965-1980, Baby Boomers 1946-1966; and the Silent Generation 1925-1945

[4]You are certainly not expected to use `R` for this project, but you are encouraged to include any supplementary materials such as handwritten work, StatCrunch output, and the data itself.

```
sum(dat$fg=="No" & dat$age=="45+")

summary(dat$age)
summary(dat$fg)

freq <- matrix(data=c(162,258,420,
                      105,422,527,
                      680,267,947),
               nrow=3, byrow=TRUE)
round(freq/947,3)

p <- sum(dat$fg=="Yes")/947
se_p <- sqrt((p*(1-p))/947)
p - c(-1,1)*qnorm(0.025)*se_p
se_p2 <- sqrt((.5^2)/947)
(p-0.5)/se_p2
2*pnorm((p-0.5)/se_p2)

p1 <- 162/420
p2 <- 105/527
se_d <- sqrt((p1*(1-p1))/420 + (p2*(1-p2))/527)
(p1-p2) - c(-1,1)*qnorm(0.025)*se_d
p_pool <- 267/947
se_pool <- sqrt((p_pool*(1-p_pool))/420 + (p_pool*(1-p_pool))/527)
2*pnorm(-(p1-p2)/se_pool)
```