Day 07 Quiz on Hypothesis Tests

Carey Kopeikin

4/3/2022

Quiz answered by Chloe Lewis

1. A start-up company is about to market a new computer printer. It decides to gamble by running commercials during the Super Bowl. The company hopes that name recognition will be worth the high cost of the ads. The goal of the company is that over 40% of the public recognize its brand name and associate it with computer equipment. The day after the game, a pollster contacts 420 randomly chosen adults and finds that 181 of them know that this company manufactures printers. Would you recommend that the company continue to advertise during Super Bowls? Explain.

a) Write appropriate hypotheses. Ha - P = 40% Ho - P != 40%

b) What is a type I error in this context? Explain the ramifications of that error. A Type I error would mean that we falsely conclude that the proportion of people who recognize the brand is different from 40% when in reality it is 40%.

c) What is a type II error in this context? Explain the ramifications of that error. In this case, the true proportion is not 40%, but we mistakenly fail to reject the null hypothesis and conclude that it is 40%. This is bad because a company has a missed opportunity to continue advertising.

d) Which is worse for the company in this context? In this context it would be worse for the company to have made a type II error because while a type I error is not ideal, a type II error means a missed opportunity to confirm the success of the companies advertising campaign.

e) Based on your answer to part d) is it better to use an alpha level of .05 or .01. Because a type II error would be considered worse it would make more sense to pick an alpha level of .05

f) Perform the test and find the p-value using the alpha you selected in part e.

```
# number of successes
x <- 181

# sample size
n <- 420

# sample proportion of successes
phat <- x/n

# sample proportion of failures
qhat <- 1-phat

# assumed true proportion of successes
p = .40
```

```r
#assumed true proportion of failures
q = 1-p

#Standard deviation of the sampling distribution of sample proportions
SDphat <- sqrt( (p*q)/n )
```

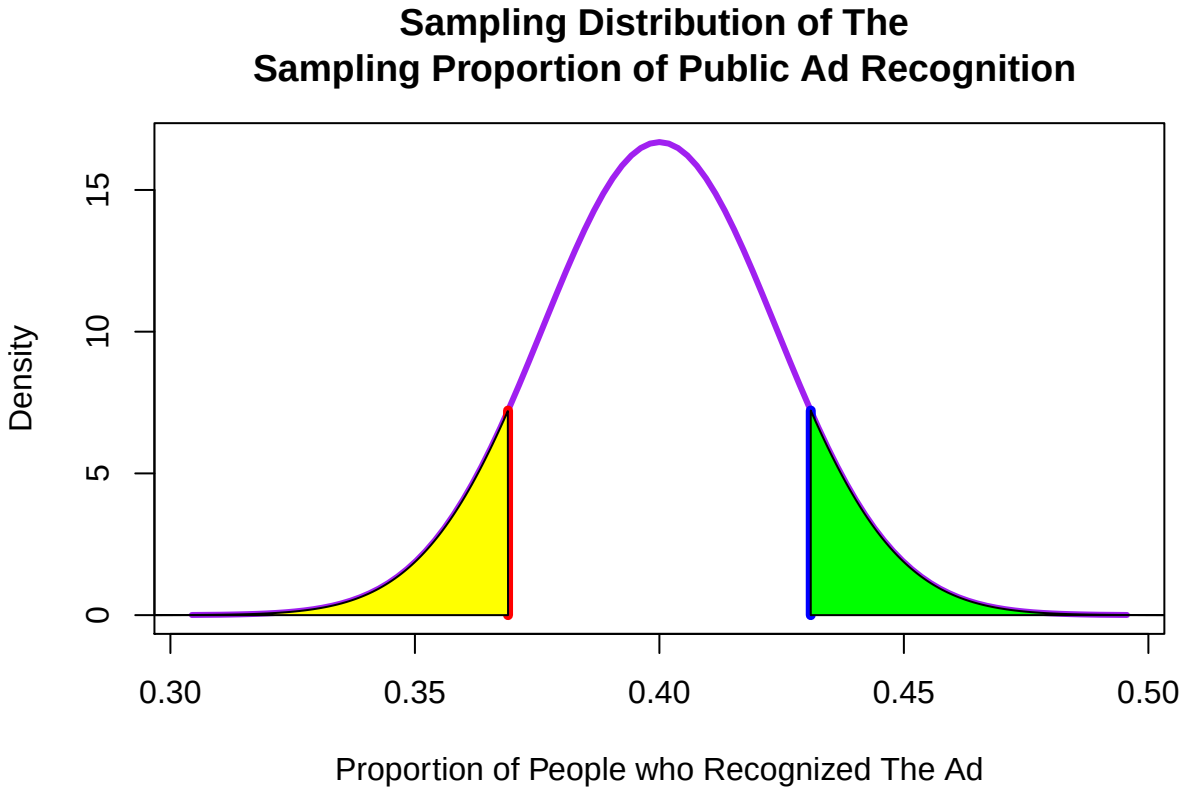
```r
prop.test(x, n, p, alternative = "two.sided", correct = FALSE)
```

```
##
##  1-sample proportions test without continuity correction
##
## data:  x out of n, null probability p
## X-squared = 1.6766, df = 1, p-value = 0.1954
## alternative hypothesis: true p is not equal to 0.4
## 95 percent confidence interval:
##  0.3844291 0.4787273
## sample estimates:
##         p
## 0.4309524
```

g) Make a graph

```r
curve( dnorm( x, p, sqrt(p*q/n)),
       xlim=c(p-4*sqrt(p*q/n), p+4*sqrt(p*q/n)),
       lwd = 3,
       main="Sampling Distribution of The \n Sampling Proportion of Public Ad Recognition",
       xlab="Proportion of People who Recognized The Ad",
       ylab="Density",
       col = "purple",
       )

if (phat < p){
  tail.1 <- phat
  tail.2 <- -1*(phat-p) + p
} else{
    tail.2 <- phat
    tail.1 <- -1*(phat-p) + p
}


segments(tail.1, 0, tail.1, dnorm( tail.1, p, sqrt(p*q/n)),
         col="red",
         lwd=5)

polygon(x=c(0,
            seq(from=-1, to=tail.1, by=.0001),
            tail.1),
        y=c(0,
            dnorm(seq(from=-1, to=tail.1, by=.0001),mean=p, sd=sqrt(p*q/n)),
            0),
        col="yellow")
```

```
segments(tail.2, 0, tail.2, dnorm( tail.2, p, sqrt(p*q/n)),
         col="blue",
         lwd=5)

polygon(x=c(tail.2,
            seq(from=tail.2, to=1, by=.0001),
            1),
        y=c(0,
            dnorm(seq(from=tail.2, to=1, by=.0001),mean=p, sd=sqrt(p*q/n)),
            0),
        col="green")
```



h) State your conclusion.

Because my p-value of 0.4352 is greater than my chosen alpha level of .05, I fail to reject my NULL hypothesis because there is not enough evidence to suggest whether or not the company has an increase, or decrease in add exposure from their goal of over 40%.

2. For over a year, from the spring of 1721 until winter 1722, a smallpox epidemic afflicted the city of Boston. Though tragic, the 1721 epidemic led to a major milestone in the history of vaccination and smallpox eradication. The use of inoculation during this epidemic, and the heated debate that arose surrounding the practice, was one of the first major applications of inoculations in western society, paving the way for Edward Jenner to develop smallpox vaccination by the end of the century.*

The file `smallpox.csv` is a sample of some of the first people in Boston to be inoculated. Perform a hypothesis test to discover if these first inoculations worked. Note that the mortality rate (death rate) of those who were not vaccinated was 14%.

```
smallpox.data <- read.csv("smallpox.csv")
head(smallpox.data)
```

```
##   result inoculated
## 1  lived        yes
## 2  lived        yes
## 3  lived        yes
## 4  lived        yes
## 5  lived        yes
## 6  lived        yes
```

a) Write appropriate hypotheses. Ho- P = .86 Ha- P > .86

b) What is a type I error in this context? Explain the ramifications of that error. A type I error in this context would mean that we wrongly reject the null hypothesis and conclude that the vaccine has a success rate greater than 86%, when it is only 86%.

c) What is a type II error in this context? Explain the ramifications of that error. A type II error in this context would mean that we fail to reject the null and conclude that the vaccines success rate is 86%, when it is actually higher than that.

d) Which is worse for the vaccine creator in this context? In this context, a type I error would be worse because it would lead us to believe that the vaccine is more effective than it actually is.

e) Based on your answer to part d) is it better to use an alpha level of .05 or .01. It would be better to use an alpha level of .01 in this case, because a type I error would be worse.

f) Perform the test and find the p-value using the alpha you selected in part e.

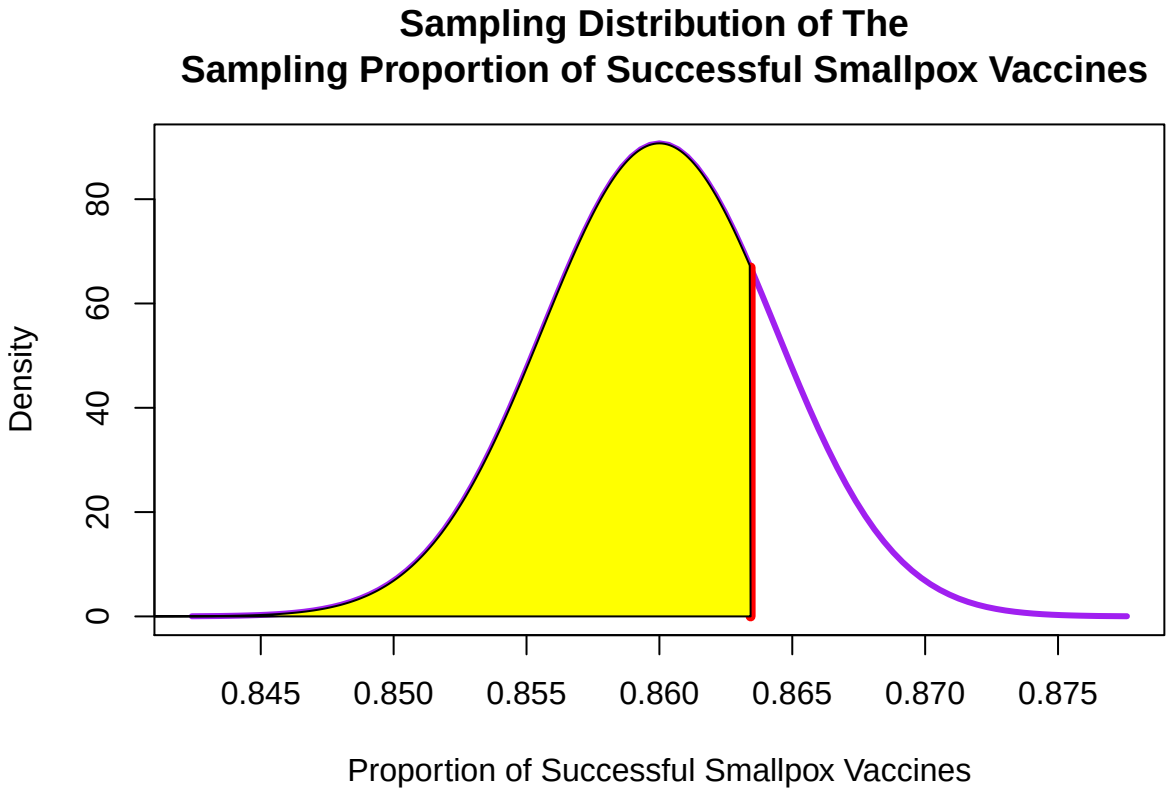
```
# number of successes
x <- sum(smallpox.data$result == "lived")

# sample size
n <- length(smallpox.data$result)

# sample proportion of successes
phat <- x/n

# sample proportion of failures
qhat <- 1-phat

# assumed true proportion of successes
p = .86

#assumed true proportion of failures
q = 1-p

#Standard deviation of the sampling distribution of sample proportions
SDphat <- sqrt( (p*q)/n )
```

```
prop.test( x, n, p, alternative = "greater", correct=FALSE)
```

```
##
##  1-sample proportions test without continuity correction
##
## data:  x out of n, null probability p
## X-squared = 0.60884, df = 1, p-value = 0.2176
## alternative hypothesis: true p is greater than 0.86
## 95 percent confidence interval:
##  0.8561143 1.0000000
## sample estimates:
##         p
## 0.8634319
```

g) Make a graph

```
curve( dnorm( x, p, SDphat),
       xlim=c(p-4*SDphat, p+4*SDphat),
       lwd = 3,
       main="Sampling Distribution of The \n Sampling Proportion of Successful Smallpox Vaccines",
       xlab="Proportion of Successful Smallpox Vaccines",
       ylab="Density",
       col = "purple",
       )


segments(phat, 0, phat, dnorm( phat, p, SDphat),
         col="red",
         lwd=5)

polygon(x=c(0,
            seq(from=-1, to=phat, by=.0001),
            phat),
        y=c(0,
            dnorm(seq(from=-1, to=phat, by=.0001),mean=p, sd=SDphat),
            0),
        col="yellow")
```



h) State your conclusion. Our p value of 0.2176 is greater than our alpha of .01 so we do not have enough evidence to reject the null hypothesis. This means that the data does not have enough evidence to support the claim that the vaccine has a success rate that is greater than 86%.

#i do not remember why this URL is here. I cannot remember if my teacher put it here or if i did but i am going to leave it. * https://sitn.hms.harvard.edu/flash/special-edition-on-infectious-disease/2014/the-fight-over-inoculation-during-the-1721-boston-smallpox-epidemic/#:~:text=The%20Boston%20Epidemic,the%20most%20deadly%