STAT 8020 R Lab 13: Inference for Proportions

Whitney

November 09, 2020

Contents

Cancer treatment survival rate

Researchers in the development of new treatments for cancer patients often evaluate the effectiveness of new therapies by reporting the *proportion* of patients who survive for a specified period of time after completion of the treatment. A new genetic treatment of 870 patients with a particular type of cancer resulted in 330 patients surviving at least 5 years after treatment. *Estimate* the proportion of all patients with the specified type of cancer who would survive at least 5 years after being administered this treatment.

```
n = 870; x = 330
# point estimate
phat <- x / n
phat
```

```
## [1] 0.3793103
```

```
# 95% CI for p
library(fastR)
wald.ci(x, n, conf.level = 0.95)
```

```
## [1] 0.3470683 0.4115524
## attr(,"conf.level")
## [1] 0.95
```

```
# 99% CI for p
wald.ci(x, n, conf.level = 0.9)
```

```
## [1] 0.3522519 0.4063688
## attr(,"conf.level")
## [1] 0.9
```

Bird flu example

Among 900 randomly selected registered voters nationwide, 63% of them are somewhat or very concerned about the spread of bird flu in the United States.

- What is the point estimate for p, the proportion of U.S. voters who are concerned about the spread of bird flu?
- Construct a 95% CI for p

```
n = 900; x = 900 * .63
# point estimate
```

```r
phat <- x / n
phat
```

```
## [1] 0.63
```

```r
# 95% CI for p
wald.ci(x, n, conf.level = 0.95)
```

```
## [1] 0.5984574 0.6615426
## attr(,"conf.level")
## [1] 0.95
```

Sample size calculation example

A researcher wants to estimate the proportion of voters who will vote for candidate A. She wants to estimate to within 0.05 with 90% confidence.

```r
# True proportion is .9
p = 0.9; alpha = 1 - 0.9
n = p * (1 - p) * (qnorm(1 - alpha / 2) / 0.05)^2
n
```

```
## [1] 97.39956
```

```r
# True proportion is .6
p = 0.6
n = p * (1 - p) * (qnorm(1 - alpha / 2) / 0.05)^2
n
```

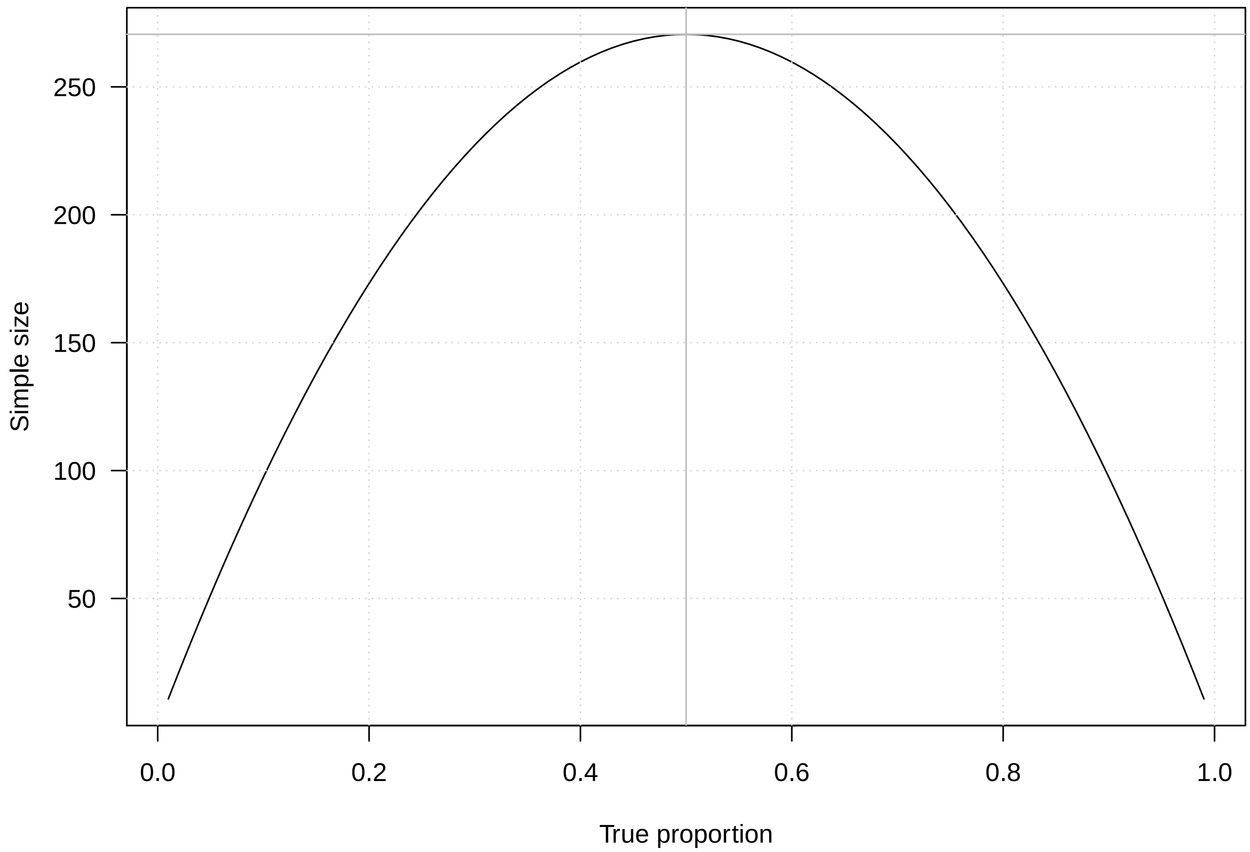
```
## [1] 259.7322
```

```r
# True proportion is .5
p = 0.5
n = p * (1 - p) * (qnorm(1 - alpha / 2) / 0.05)^2
n
```

```
## [1] 270.5543
```

```r
# Just for fun
p <- seq(0.01, 0.99, 0.01)
n = p * (1 - p) * (qnorm(1 - alpha / 2) / 0.05)^2

plot(p, n, type = "l",
     xlab = "True proportion",
     ylab = "Simple size",
     las = 1)
abline(v = 0.5, col = "gray")
abline(h = max(n), col = "gray")
grid()
```



Hypothesis testing: Bird flu example

Among 900 randomly selected registered voters nationwide, 63% of them are somewhat or very concerned about the spread of bird flu in the United States. Conduct a hypothesis test at .01 level to assess the research hypothesis: $p > .6$.

```
n = 900; x = .63 * 900
p_null = .6; alpha = 0.01

prop.test(x, n, p = p_null,
          conf.level = 1 - alpha,
          alternative = "greater",
          correct = FALSE)
```

```
##
##  1-sample proportions test without continuity correction
##
## data:  x out of n
## X-squared = 3.375, df = 1, p-value = 0.0331
## alternative hypothesis: true p is greater than 0.6
## 99 percent confidence interval:
##  0.5918879 1.0000000
## sample estimates:
##    p
## 0.63
```

```
# With Yates' continuity correction
prop.test(x, n, p = p_null,
```

```r
          conf.level = 1 - alpha,
          alternative = "greater",
          correct = TRUE)
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  x out of n
## X-squared = 3.2512, df = 1, p-value = 0.03569
## alternative hypothesis: true p is greater than 0.6
## 99 percent confidence interval:
##  0.5913242 1.0000000
## sample estimates:
##    p
## 0.63
```

Proportion of CU vegetarian

```r
n = 25; x = 0
wald.ci(x, n)
```

```
## [1] 0 0
## attr(,"conf.level")
## [1] 0.95
```

```r
## Wilson score CI
library(PropCIs)
scoreci(x, n, conf.level = .95)
```

```
##
##
##
## data:
##
## 95 percent confidence interval:
##  0.0000 0.1332
```

```r
## Check
prop.test(x, n, conf.level = 0.95,
          correct = F)
```

```
##
##  1-sample proportions test without continuity correction
##
## data:  x out of n
## X-squared = 25, df = 1, p-value = 5.733e-07
## alternative hypothesis: true p is not equal to 0.5
## 95 percent confidence interval:
##  0.0000000 0.1331923
## sample estimates:
## p
## 0
```

Here the rule of three provides a quick and reasonable approximation

Inference for $p_1 - p_2$

A Simple Random Simple of 100 CU graduate students is taken and it is found that 79 strongly agree that they would recommend their current graduate program. A Simple Random Simple of 85 USC graduate students is taken and it is found that 52 strongly agree that they would recommend their current graduate program. At 5 % level, can we conclude that the proportion of strongly agree is higher at CU?

```r
x <- c(79, 52); n <- c(100, 85)
prop.test(x, n, alternative = "greater", correct = F)
```

```
##
##  2-sample test for equality of proportions without continuity
##  correction
##
## data:  x out of n
## X-squared = 7.0618, df = 1, p-value = 0.003937
## alternative hypothesis: greater
## 95 percent confidence interval:
##  0.06847019 1.00000000
## sample estimates:
##    prop 1    prop 2
## 0.7900000 0.6117647
```

```r
# With Yates' continuity correction
prop.test(x, n, alternative = "greater", correct = F)
```

```
##
##  2-sample test for equality of proportions without continuity
##  correction
##
## data:  x out of n
## X-squared = 7.0618, df = 1, p-value = 0.003937
## alternative hypothesis: greater
## 95 percent confidence interval:
##  0.06847019 1.00000000
## sample estimates:
##    prop 1    prop 2
## 0.7900000 0.6117647
```