Confidence intervals for proportions (Chapter 16)

Patrick Frenett, Vickie Ip, and Nicholas Horton (nhorton@amherst.edu)

June 19, 2018

Introduction and background

This document is intended to help describe how to undertake analyses introduced as examples in the Fourth Edition of *Intro Stats* (2013) by De Veaux, Velleman, and Bock. More information about the book can be found at http://wps.aw.com/aw_deveaux_stats_series. This file as well as the associated R Markdown reproducible analysis source file used to create it can be found at https://nhorton.people.amherst.edu/is4.

This work leverages initiatives undertaken by Project MOSAIC (http://www.mosaic-web.org), an NSF-funded effort to improve the teaching of statistics, calculus, science and computing in the undergraduate curriculum. In particular, we utilize the `mosaic` package, which was written to simplify the use of R for introductory statistics courses. A short summary of the R needed to teach introductory statistics can be found in the mosaic package vignettes (http://cran.r-project.org/web/packages/mosaic). A paper describing the mosaic approach was published in the *R Journal*: https://journal.r-project.org/archive/2017/RJ-2017-024.

Note that some of the figures in this document may differ slightly from those in the IS4 book due to small differences in datasets. However in all cases the analysis and techniques in R are accurate.

Chapter 16: Confidence intervals for proportions

Section 16.1: A confidence interval

The Facebook survey of 156 respondents yielded 48 who updated their status every day (or more often).

```
n <- 156
binom.test(48, n)      # exact binomial
```

```
##
##
##
## data:  48 out of n
## number of successes = 48, number of trials = 156, p-value =
## 1.748e-06
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.2363644 0.3864933
## sample estimates:
## probability of success
##              0.3076923
```

Calculation on page 429

```
phat <- 48/n
phat
```

```
## [1] 0.3076923
```

```
sep <- sqrt(phat*(1 - phat)/n)
sep     # matches value on page 429
```

```
## [1] 0.03695265
```

```
interval <- phat + c(-2, 2)*sep
interval
```
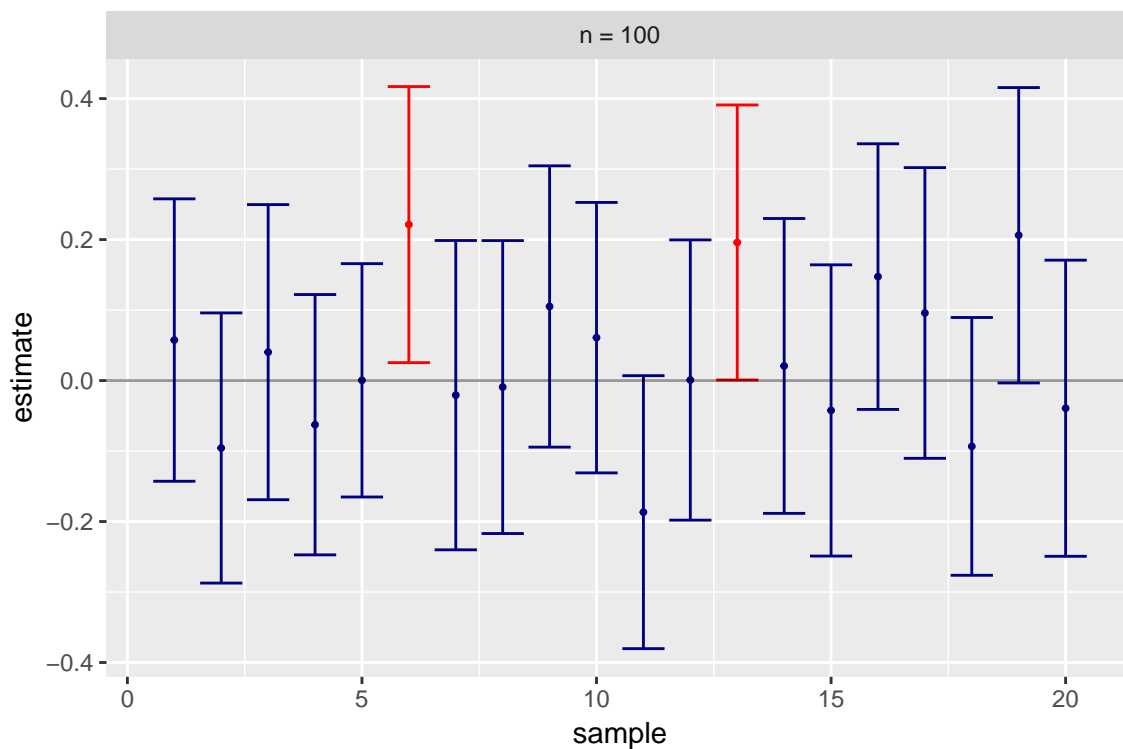
```
## [1] 0.2337870 0.3815976
```

Section 16.2: Interpreting confidence intervals

```
set.seed(1988)
CIsim(n = 100, samples = 20)
```

```
## Interval coverage:
```

```
##        cover
## n      Low Yes High
##   100 0.0 0.9  0.1
```


We would expect 19 out of 20 of the intervals to cover the true (population) value, but here only 18 out of 20 actually covered that value (see Figure at top of page 432).

Section 16.3: Margin of error

We can replicate the calculation for the "For Example: Finding the margin of error Take 1" (page 434)

```
sdp <- sqrt(.5*.5/1010)
sdp    # worst case margin of error (based on p=0.5)
```

```
## [1] 0.01573292
```

```
me <- 2*sdp
me
```

```
## [1] 0.03146584
```

We can replicate the calculation for the "For Example: Finding the margin of error Take 1" (pages 434-435)

```
qnorm(.95, mean = 0, sd = 1)    # z-star for 90% confidence interval
```

```
## [1] 1.644854
```

```
sep <- sqrt(.4*.6/1010)
sep
```

```
## [1] 0.01541505
```

```
me <- 1.6445*sep
me
```

```
## [1] 0.02535005
```

Section 16.4: Assumptions and Condition

We can replicate the calculation for the "For Example: choosing a sample size" (page 438)

```
zstar <- qnorm(.975, mean = 0, sd = 1)
zstar
```

```
## [1] 1.959964
```

```
me <- 0.02 # desired margin of error
p <- 0.40
n <- (zstar*sqrt(p*(1 - p))/me)^2
n
```

```
## [1] 2304.875
```

We will need about 2305 subjects to yield a margin of error of 2% under these assumptions.