Introduction to the Practice of Statistics using R: Chapter 5

Nicholas J. Horton[*]   Ben Baumer

April 8, 2013

Contents

Introduction

This document is intended to help describe how to undertake analyses introduced as examples in the Sixth Edition of *Introduction to the Practice of Statistics* (2009) by David Moore, George McCabe and Bruce Craig. More information about the book can be found at `http://bcs.whfreeman.com/ips6e/`. This file as well as the associated `knitr` reproducible analysis source file can be found at `http://www.math.smith.edu/~nhorton/ips6e`.

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 vignette (`http://cran.r-project.org/web/packages/mosaic/vignettes/MinimalR.pdf`).

To use a package within R, it must be installed (one time), and loaded (each session). The package can be installed using the following command:

```
> install.packages('mosaic')                 # note the quotation marks
```

The `#` character is a comment in R, and all text after that on the current line is ignored.
Once the package is installed (one time only), it can be loaded by running the command:

```
> require(mosaic)
```

This needs to be done once per session.
We also set some options to improve legibility of graphs and output.

---

[*]Department of Mathematics and Statistics, Smith College, nhorton@smith.edu

```
> trellis.par.set(theme=col.mosaic())  # get a better color scheme for lattice
> options(digits=3)
```

The specific goal of this document is to demonstrate how to replicate the analysis described in Chapter 5: Sampling Distributions.

1   Sampling distributions for counts and proportions

Calculations with the binomial distribution can be undertaken using the `pbinom()` and `dbinom()` functions. For example, the results from Figure 5.1 (page 317) can be reproduced.

```
> dbinom(10, size=150, prob=0.08)

[1] 0.107

> pbinom(10, size=150, prob=0.08)

[1] 0.338
```

The table Figure 5.2 (page 318) can be reproduced using the following command:

```
> cbind(0:9, dbinom(0:9, size=15, p=0.08))

        [,1]      [,2]
 [1,]      0 2.86e-01
 [2,]      1 3.73e-01
 [3,]      2 2.27e-01
 [4,]      3 8.57e-02
 [5,]      4 2.23e-02
 [6,]      5 4.27e-03
 [7,]      6 6.19e-04
 [8,]      7 6.93e-05
 [9,]      8 6.02e-06
[10,]      9 4.07e-07
```

And the calculation on page 318 using the command:

```
> sum(dbinom(0:1, size=15, prob=0.08))

[1] 0.66
```

or using `pbinom()`:

```
> pbinom(1, size=15, prob=0.08)

[1] 0.66
```

Example 5.9 (pages 318-319) can be calculated:

```
> 1 - pbinom(4, size=12, prob=0.25)

[1] 0.158
```

Using R, there is little need for the normal approximation to the binomial. As an example, the probability of interest in Example 5.11 (page 321) can be calculated using the command:

```
> 1 - pbinom(1449, size=2500, prob=0.6)

[1] 0.98
```

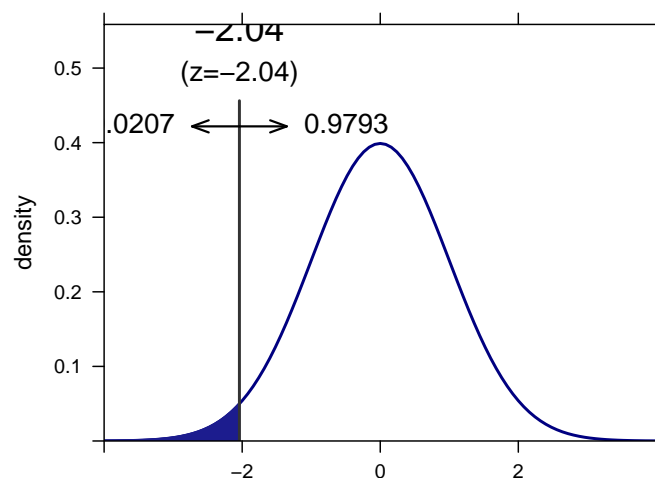Example 5.13 (page 324) uses the normal approximation nonetheless:

```
> xpnorm(-2.04, mean=0, sd=1)


If X ~ N(0,1), then


P(X <= -2.04) = P(Z <= -2.04) = 0.0207
P(X >  -2.04) = P(Z >  -2.04) = 0.9793
[1] 0.0207
```
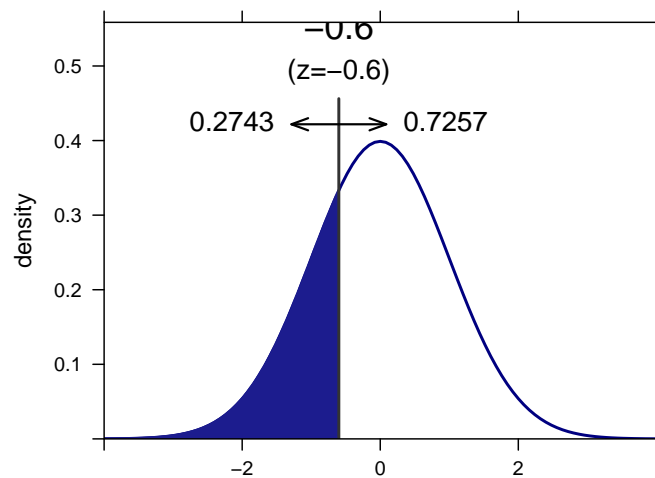

A similar calculation is done in Example 5.14 (page 325):

```
> xpnorm(-0.60, mean=0, sd=1)


If X ~ N(0,1), then
```

Introduction to the Practice of Statistics using R: Chapter 5

```
P(X <= -0.6) = P(Z <= -0.6) = 0.2743
P(X >  -0.6) = P(Z >  -0.6) = 0.7257
[1] 0.274
```


We can compare this to the exact calculation:

```
> pbinom(10, size=150, prob=0.08)

[1] 0.338
```

The approximation isn't great (which is a good reason not to use it).

The binomial probability formula (Example 5.16, page 329) can be used to calculate probabilities, or `dbinom()` and `pbinom()` can do the trick:

```
> dbinom(0, size=15, prob=0.08)

[1] 0.286

> dbinom(1, size=15, prob=0.08)

[1] 0.373

> dbinom(0, size=15, prob=0.08) + dbinom(1, size=15, prob=0.08)

[1] 0.66

> pbinom(1, size=15, prob=0.08)

[1] 0.66
```

2 Sampling distributions for a sample mean

Similar approaches are used for sampling distributions for a sample mean, using `xpnorm()`. For instance, Example 5.24 (page 343) can be found using:
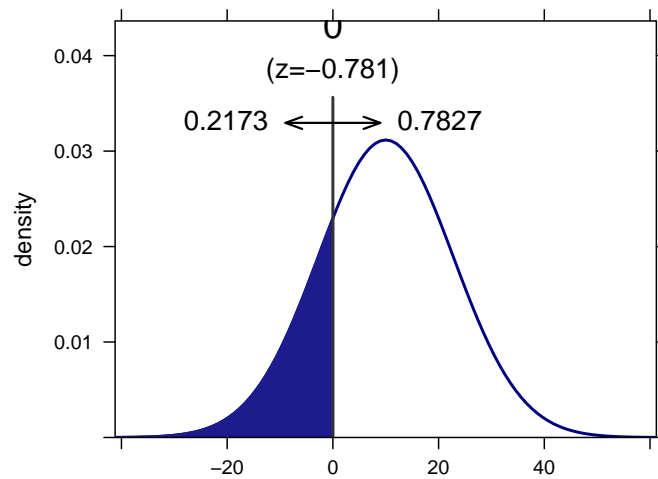
```
> xpnorm(0, mean=10, sd=12.8)


If X ~ N(10,12.8), then

P(X <= 0) = P(Z <= -0.781) = 0.2173
P(X >  0) = P(Z >  -0.781) = 0.7827
[1] 0.217
```


or

```
> xpnorm(-0.78, mean=0, sd=1)


If X ~ N(0,1), then

P(X <= -0.78) = P(Z <= -0.78) = 0.2177
P(X >  -0.78) = P(Z >  -0.78) = 0.7823
[1] 0.218
```