

IPS9 in R: Sampling distributions (Chapter 5)

Margaret Chien and Nicholas Horton (nhorton@amherst.edu)

November 10, 2018

Introduction and background

These documents are intended to help describe how to undertake analyses introduced as examples in the Ninth Edition of *Introduction to the Practice of Statistics* (2017) by Moore, McCabe, and Craig.

More information about the book can be found [here](#). The data used in these documents can be found under Data Sets in the Student Site. 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/ips9/>.

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>.

Chapter 5: Sampling Distributions

This file replicates the analyses from Chapter 5: Sampling Distributions.

First, load the packages that will be needed for this document:

```
library(mosaic)
library(readr)
```

Section 5.1: Toward Statistical Inference

Section 5.2: The Sampling Distribution of a Sample Mean

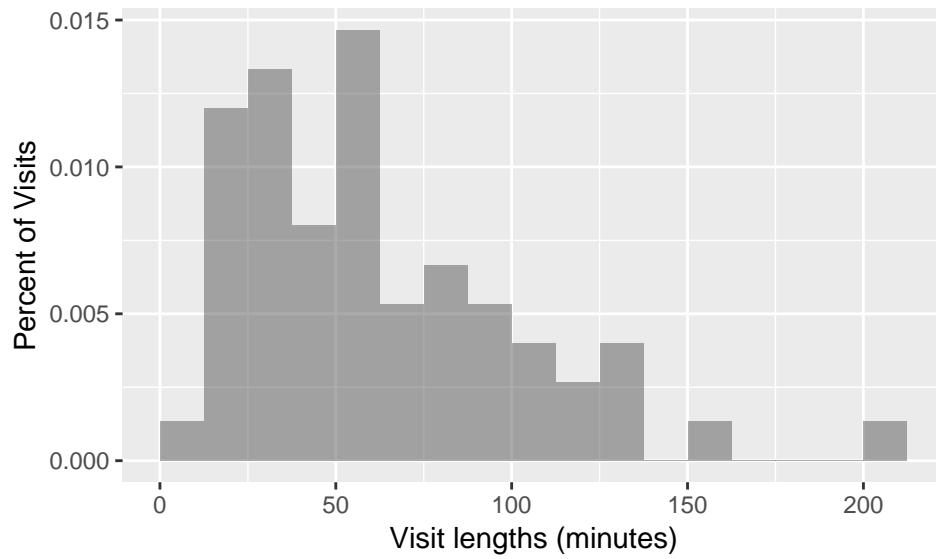
Example 5.5: Sample means are approximately Normal

```
Help <- read_csv("https://nhorton.people.amherst.edu/ips9/data/chapter05/EG05-05HELP60.csv")
```

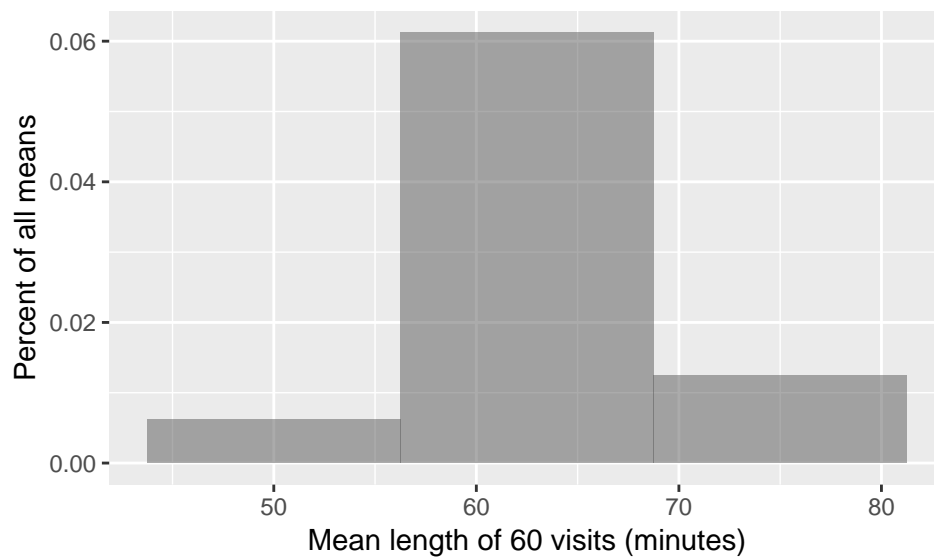
```
## Parsed with column specification:
## cols(
##   Length = col_double()
## )
```

```
# Figure 5.6(a), page 294
```

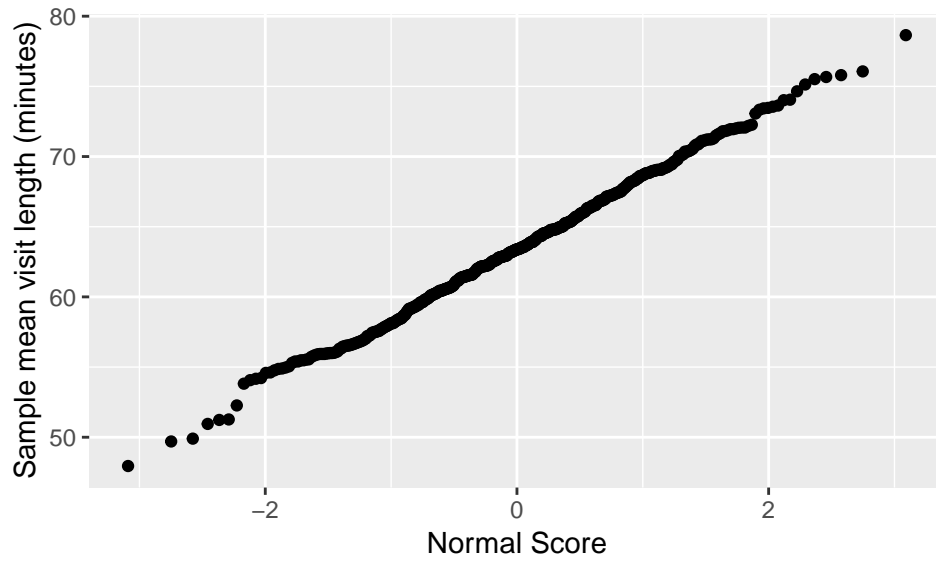
```
gf_dhistogram(~ Length, data = Help, binwidth = 12.5, center = 6.25) %>%
  gf_labs(x = "Visit lengths (minutes)", y = "Percent of Visits")
```



```
# Figure 5.6(b)
set.seed(124)
HelpSamples <- do(500) * mean(~ Length, data = resample(Help, size = 60))
gf_dhistogram(~ mean, data = HelpSamples, binwidth = 12.5) %>%
  gf_labs(x = "Mean length of 60 visits (minutes)", y = "Percent of all means")
```



```
# Figure 5.7
gf_qq(~ mean, data = HelpSamples) %>%
  gf_labs(x = "Normal Score", y = "Sample mean visit length (minutes)")
```



Section 5.3: Sampling Distributions for Counts and Proportions