

# IS5 in R: Testing Hypotheses (Chapter 15)

Margaret Chien and Nicholas Horton ([nhorton@amherst.edu](mailto:nhorton@amherst.edu))

July 11, 2018

## Introduction and background

This document is intended to help describe how to undertake analyses introduced as examples in the Fifth Edition of *Intro Stats* (2018) by De Veaux, Velleman, and Bock. More information about the book can be found at [http://wps.aw.com/aw\\_deveaux\\_stats\\_series](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 <http://nhorton.people.amherst.edu/is5>.

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 15: Testing Hypotheses

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

### Section 15.1: Hypotheses

### Section 15.2: P-Values

### Section 15.3: The Reasoning of Hypothesis Testing

#### Example 15.5: Finding A P-Value

```
n <- 90
x <- 61
p <- .8
phat <- x/n
sdphat <- ((p * (1 - p))/n)^.5
z <- (phat - p)/sdphat
pnorm(z)
```

```
## [1] 0.00187324
```

```
# Or, without calculating the z-score:
pnorm(q = phat, mean = p, sd = sdphat)
```

```
## [1] 0.00187324
```

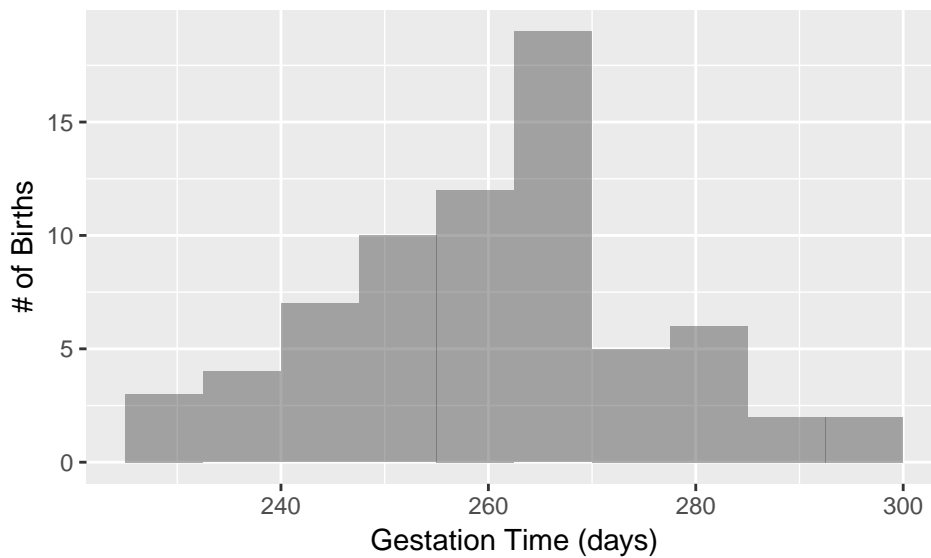
## Section 15.4: A Hypothesis Test for the Mean

```
GestationTime <- read_csv("http://nhorton.people.amherst.edu/is5/data/Nashville.csv")
```

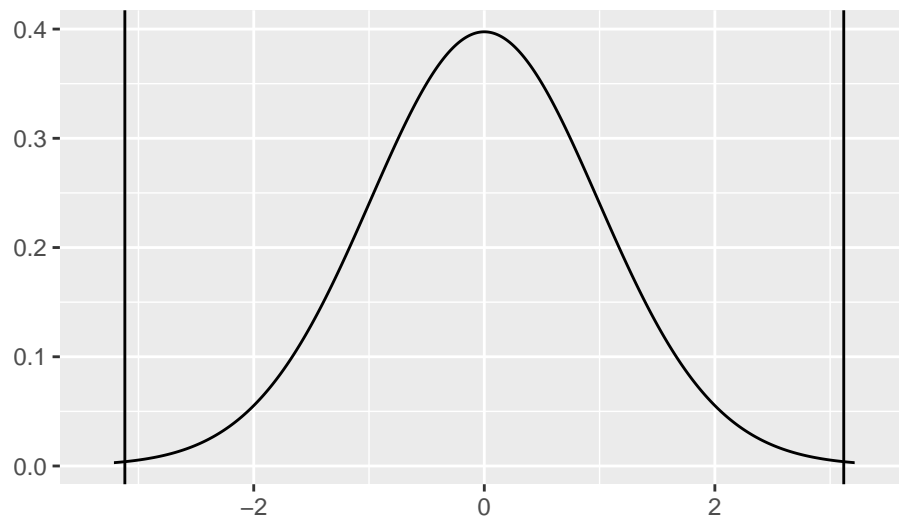
```
## Parsed with column specification:  
## cols(  
##   Gestation = col_integer(),  
##   Time = col_character()  
## )
```

By default, `read_csv()` prints the variable names. These messages can be suppressed using the `message=FALSE` code chunk option to save space and improve readability.

```
# 2. Model (page 482)  
gf_histogram(~ Gestation, data = GestationTime, binwidth = 7.5, center = 3.75) %>%  
  gf_labs(x = "Gestation Time (days)", y = "# of Births")
```



```
# 3. Mechanics  
gf_dist(dist = "t", df = 69) %>%  
  gf_vline(xintercept = -3.118) %>%  
  gf_vline(xintercept = 3.118) %>%  
  gf_labs(x = "", y = "") +  
  xlim(-3.347, 3.347)
```

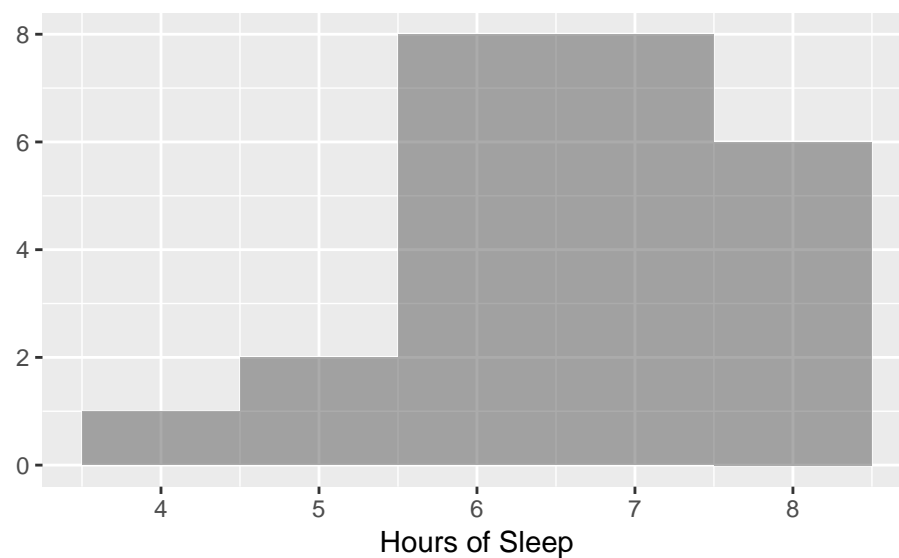


### Step-By-Step Example: A One-Sample $t$ -Test for the Mean

```
Sleep <- read_csv("http://nhorton.people.amherst.edu/is5/data/Sleep.csv")
```

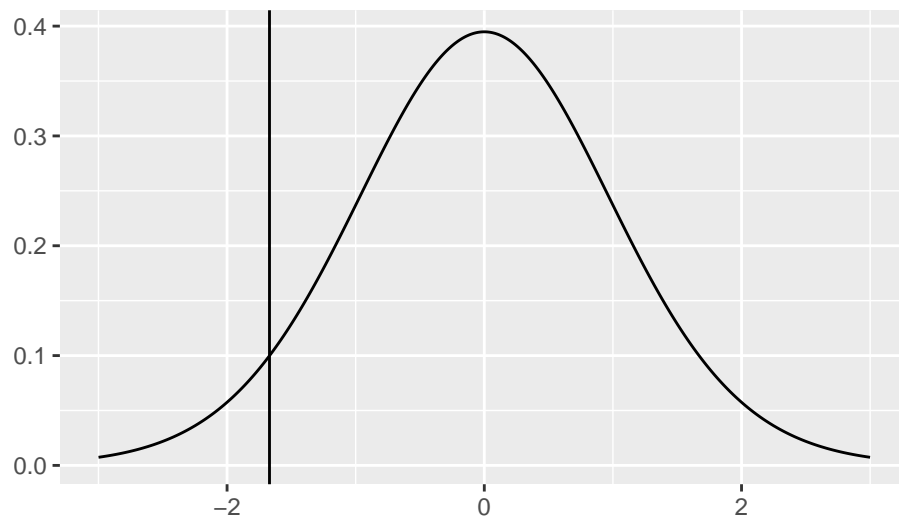
```
## Parsed with column specification:
## cols(
##   Sleep = col_integer()
## )
```

```
# Plan
gf_histogram(~ Sleep, data = Sleep, binwidth = 1) %>%
  gf_labs(x = "Hours of Sleep", y = "")
```



```
gf_dist(dist = "t", df = 24) %>%
  gf_vline(xintercept = -1.67) %>%
  gf_labs(x = "", y = "") +
  xlim(-3, 3)
```

```
## Warning: Removed 674 rows containing missing values (geom_path).
```



```
# Mechanics
n <- 25
mean <- 7.0
df <- 24
y <- 6.64
s <- 1.075
sey <- s/(n^.5)
t <- (y - mean)/sey # t-statistic
pt(q = t, df = df) # p-value

## [1] 0.05351625
```

## Section 15.5: Intervals and Tests

XX NH Can't find temperature data (page 487) - data should be from an exercise (Chapter 14, Exercise 29)

### Step-By-Step Example: Tests and Intervals

*# XX NH There's a baseball 2016 csv, but it doesn't seem to have a variable about home vs away (page 48)*

## Section 15.6: P-Values and Decisions: What to Tell About a Hypothesis Test