Inferences about Means (Chapter 18)

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

June 16, 2016

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 http://www.amherst.edu/~nhorton/sdm4.

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).

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.

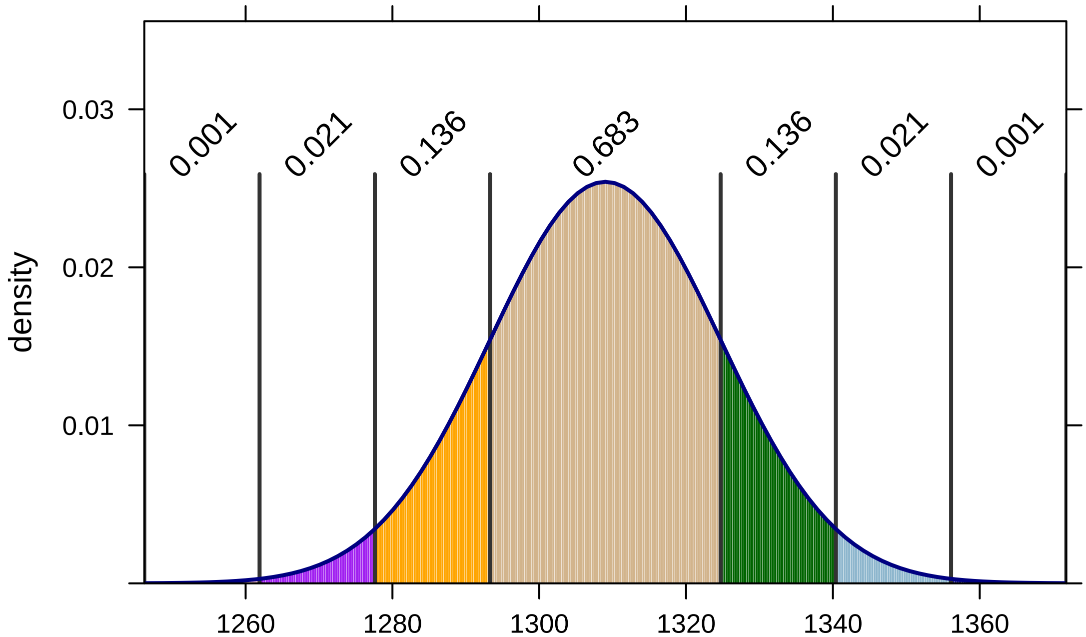
Chapter 18: Inferences about Means

Section 18.1: The Central Limit Theorem

Let's begin by reproducing the figure on the bottom of page 475.

```
mu <- 1309
sd <- 15.7
xpnorm(c(mu-3*sd, mu-2*sd, mu-sd, mu+sd, mu+2*sd, mu+3*sd), mean=mu, sd=sd)
```

```
##
## If X ~ N(1309, 15.7), then
##
##  P(X <= 1261.9) = P(Z <= -3) = 0.001349898
##      P(X <= 1277.6) = P(Z <= -2) = 0.022750132
##      P(X <= 1293.3) = P(Z <= -1) = 0.158655254
##      P(X <= 1324.7) = P(Z <=  1) = 0.841344746
##      P(X <= 1340.4) = P(Z <=  2) = 0.977249868
##      P(X <= 1356.1) = P(Z <=  3) = 0.998650102
##  P(X >  1261.9) = P(Z >  -3) = 0.998650102
##      P(X >  1277.6) = P(Z >  -2) = 0.977249868
##      P(X >  1293.3) = P(Z >  -1) = 0.841344746
##      P(X >  1324.7) = P(Z >   1) = 0.158655254
##      P(X >  1340.4) = P(Z >   2) = 0.022750132
##      P(X >  1356.1) = P(Z >   3) = 0.001349898
```

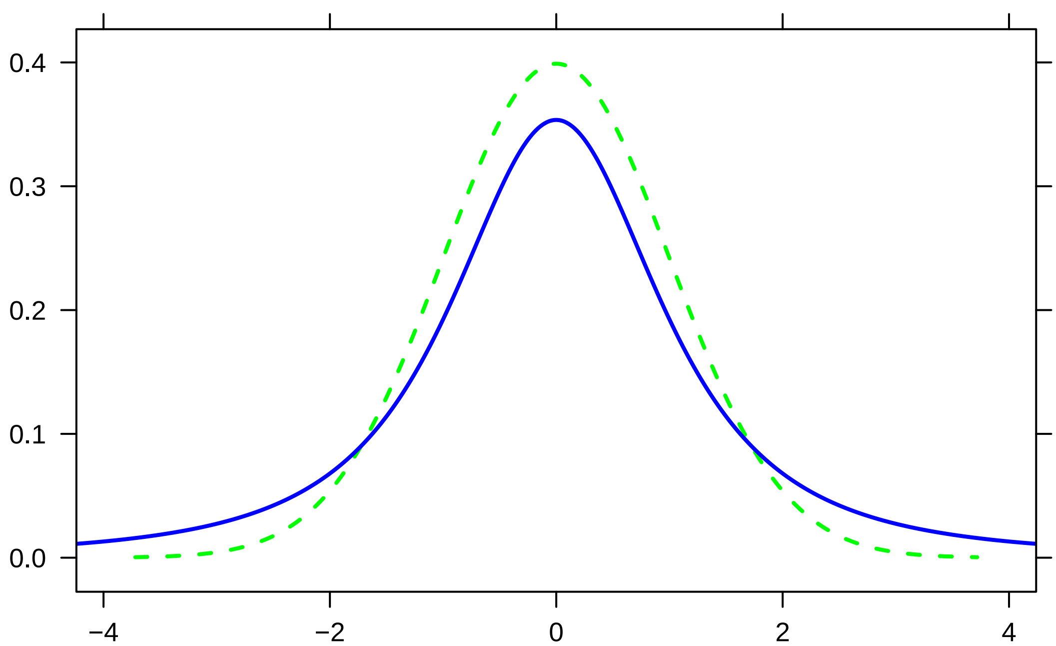



```
## [1] 0.001349898 0.022750132 0.158655254 0.841344746 0.977249868 0.998650102
```

Section 18.2: Gosset's t

Figure 18.1 (page 477) displays a normal curve (dashed green curve) and a t-model with 2 degrees of freedom (solid blue curve).

```
plotDist("norm", lty=2, col="green", lwd=2)
plotDist("t", params=2, lty=1, lwd=2, col="blue", add=TRUE)
```



We can reproduce the calculations for the Farmed salmon example (pages 479-480) using summary statistics:

```
n <- 150; ybar <- 0.0913; s = 0.0495
tstar <- qt(0.975, df=n-1); tstar
```

```
## [1] 1.976013
```

```
ybar + c(-tstar, tstar)*s/sqrt(n)
```

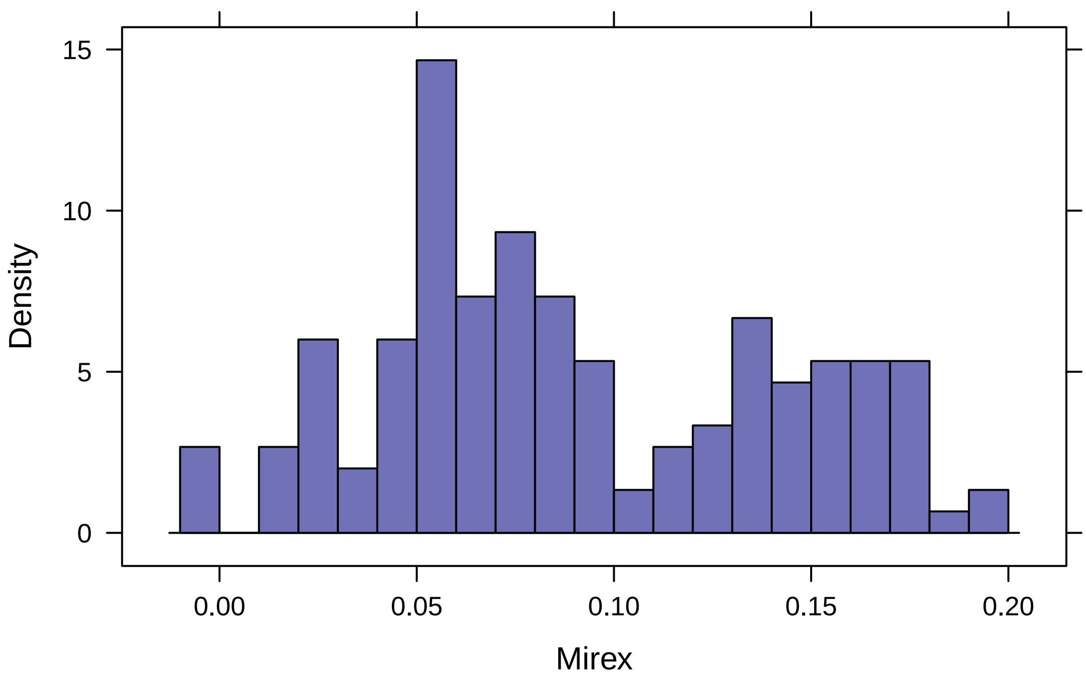
```
## [1] 0.08331363 0.09928637
```

or directly:

```
Salmon <- read.csv("http://www.amherst.edu/~nhorton/sdm4/data/Farmed_Salmon.csv")
favstats(~ Mirex, data=Salmon)
```

```
##   min    Q1 median      Q3   max    mean         sd   n missing
##     0 0.056  0.079 0.13475 0.194 0.09134 0.04952388 150       3
```

```
histogram(~ Mirex, width=0.01, center=0.01/2, data=Salmon)
```




```
t.test(~ Mirex, data=Salmon)
```

```
##
##  One Sample t-test
##
## data:  Salmon$Mirex
## t = 22.589, df = 149, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
```

```
##  0.08334978 0.09933022
## sample estimates:
## mean of x
##   0.09134
```

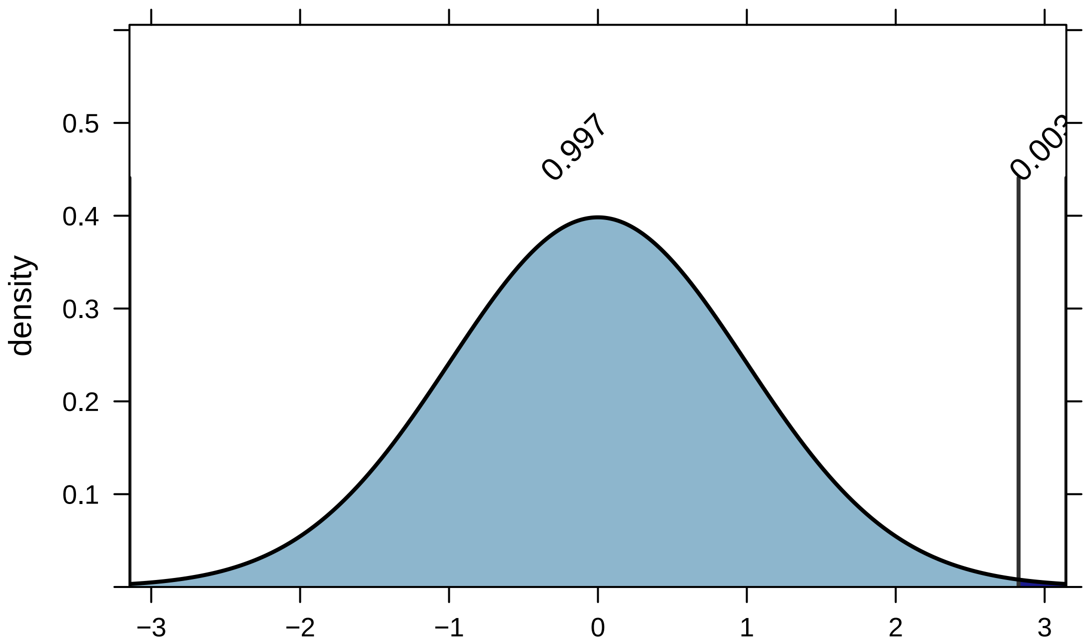
We note that the distribution of measurements is not particularly normal.

Section 18.4: A hypothesis test for the mean We can carry out the one-sided test outlined on page 486:

```
tval <- (.0913-0.08)/0.0040; tval
```

```
## [1] 2.825
```

```
1-xpt(tval, df=149)
```




```
## [1] 0.002688148
```