

SDM4 in R: The Standard Deviation as a Ruler and the Normal Model (Chapter 5)

Nicholas Horton (nhorton@amherst.edu)

January 31, 2016

Introduction and background

This document is intended to help describe how to undertake analyses introduced as examples in the Fourth Edition of *Stats: Data and Models* (2014) 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>).

Chapter 5: The standard deviation as a ruler and the normal model

Section 5.1: Standardizing with z-scores

```
library(mosaic); library(readr)
options(na.rm=TRUE)
options(digits=3)
(6.54 - 5.91)/0.56 # should be 1.1 sd better, see page 112

## [1] 1.12

Heptathlon <-
read_delim("http://www.amherst.edu/~nhorton/sdm4/data/Womens_Heptathlon_2012.txt",
  delim="\t")
nrow(Heptathlon)

## [1] 38

filter(Heptathlon, LJ >= max(LJ, na.rm=TRUE)) %>% data.frame()

##      Rank      Athlete Total_Points
## 1      3 Chernova, TatyanaTatyana Chernova (RUS)      6628
##      X100_m_hurdle_points X100_m_hurdles HJ_Points  HJ SP_Points  SP
## 1              1053              13.5          978 1.8          805 14.2
##      X200_m_Points X200_m LJ_Points  LJ JT_Points  JT X800_m_Points X800_m
## 1              1013      23.7      1020 6.54          788 46.5          971      130
```

```
favstats(~ LJ, data=Heptathlon)
```

```
## min Q1 median Q3 max mean sd n missing
## 3.7 5.83 6.01 6.19 6.54 5.91 0.564 35 3
```

```
(6.54 - mean(~ LJ, data=Heptathlon))/sd(~ LJ, data=Heptathlon)
```

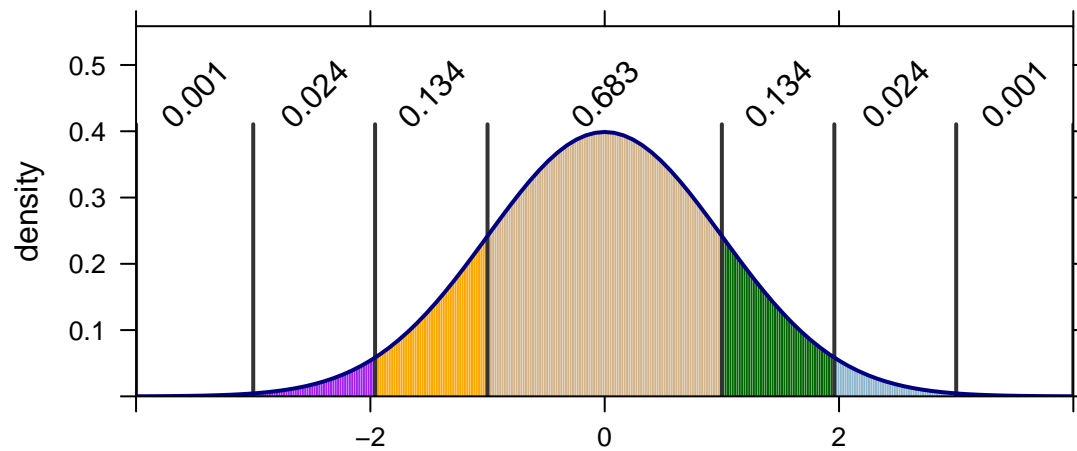
```
## [1] 1.11
```

Section 5.2: Shifting and scaling

Section 5.3: Normal models

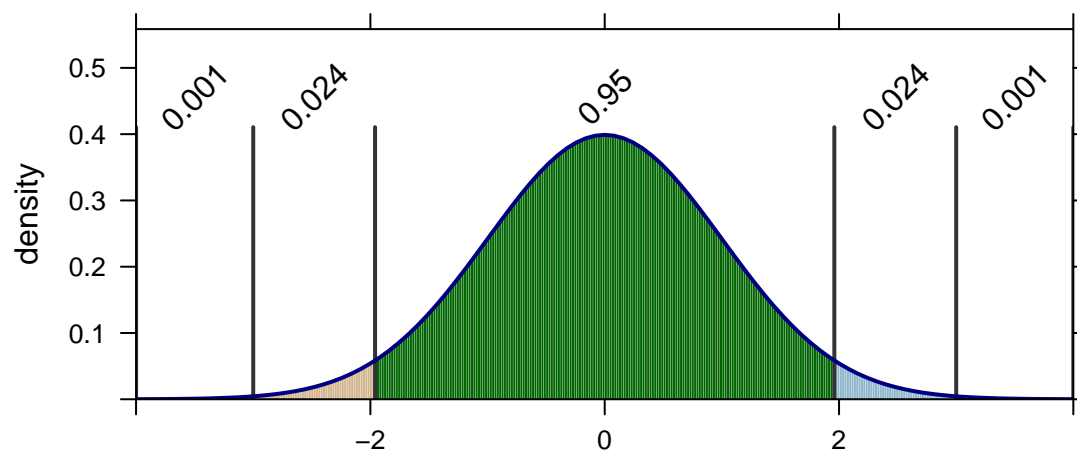
The 68-95-99.7 rule

```
xpnorm(c(-3, -1.96, -1, 1, 1.96, 3), mean=0, sd=1, verbose=FALSE)
```



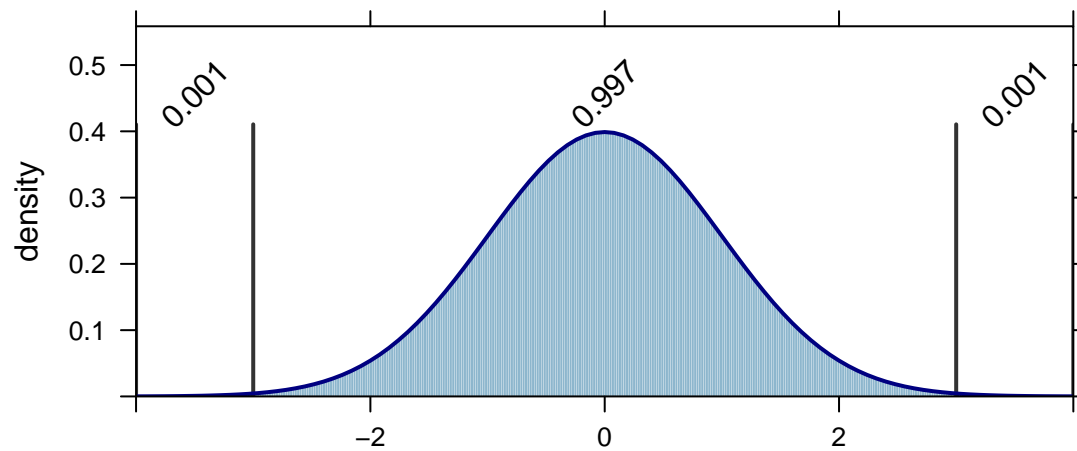
```
## [1] 0.00135 0.02500 0.15866 0.84134 0.97500 0.99865
```

```
xpnorm(c(-3, -1.96, 1.96, 3), mean=0, sd=1, verbose=FALSE)
```



```
## [1] 0.00135 0.02500 0.97500 0.99865
```

```
xpnorm(c(-3, 3), mean=0, sd=1, verbose=FALSE)
```

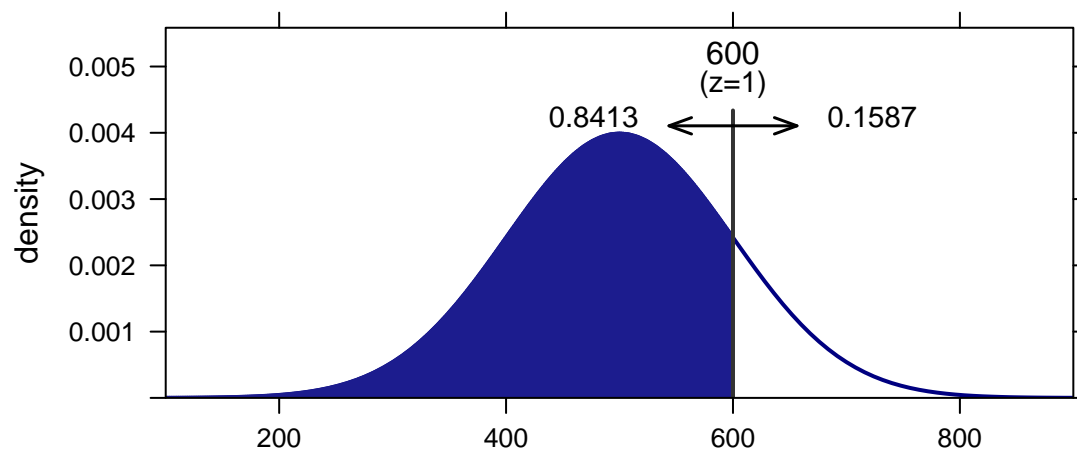


```
## [1] 0.00135 0.99865
```

Step-by-step (page 122)

```
xpnorm(600, mean=500, sd=100)
```

```
##  
## If  $X \sim N(500, 100)$ , then  
##  
##  $P(X \leq 600) = P(Z \leq 1) = 0.841$   
##  $P(X > 600) = P(Z > 1) = 0.159$ 
```



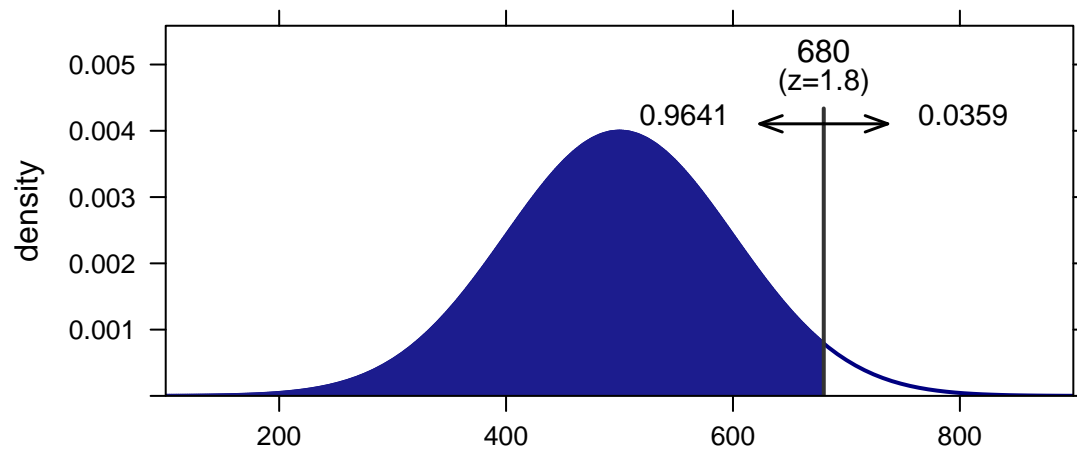
```
## [1] 0.841
```

Section 5.4: Finding normal percentiles

as on page 123

```
xpnorm(680, mean=500, sd=100)
```

```
##  
## If  $X \sim N(500, 100)$ , then  
##  
##  $P(X \leq 680) = P(Z \leq 1.8) = 0.964$   
##  $P(X > 680) = P(Z > 1.8) = 0.0359$ 
```



```
## [1] 0.964
```

```
qnorm(0.964, mean=500, sd=100)  # inverse of pnorm()
```

```
## [1] 680
```

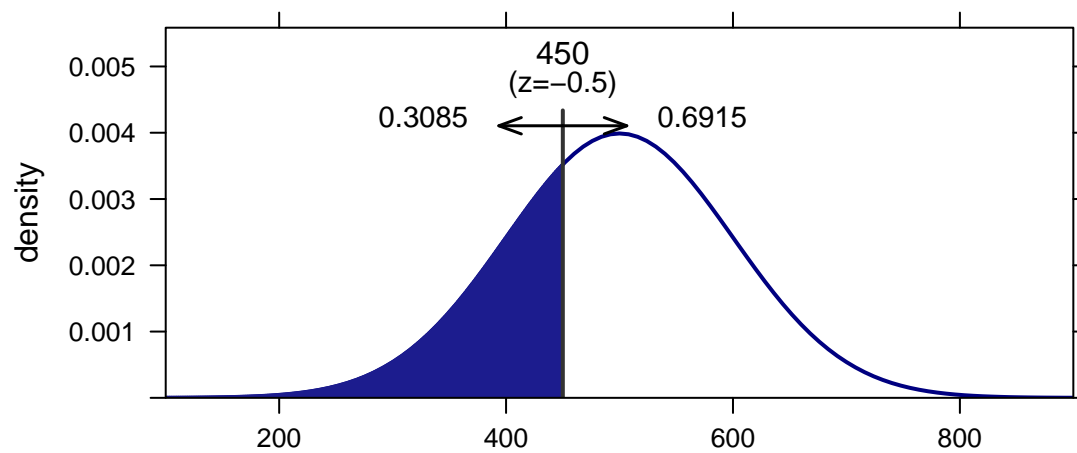
```
qnorm(0.964, mean=0, sd=1)  # what is the z-score?
```

```
## [1] 1.8
```

or on page 124

```
xpnorm(450, mean=500, sd=100)
```

```
##  
## If  $X \sim N(500, 100)$ , then  
##  
##  $P(X \leq 450) = P(Z \leq -0.5) = 0.309$   
##  $P(X > 450) = P(Z > -0.5) = 0.691$ 
```



```
## [1] 0.309
```

and page 125

```
qnorm(.9, mean=500, sd=100)
```

```
## [1] 628
```

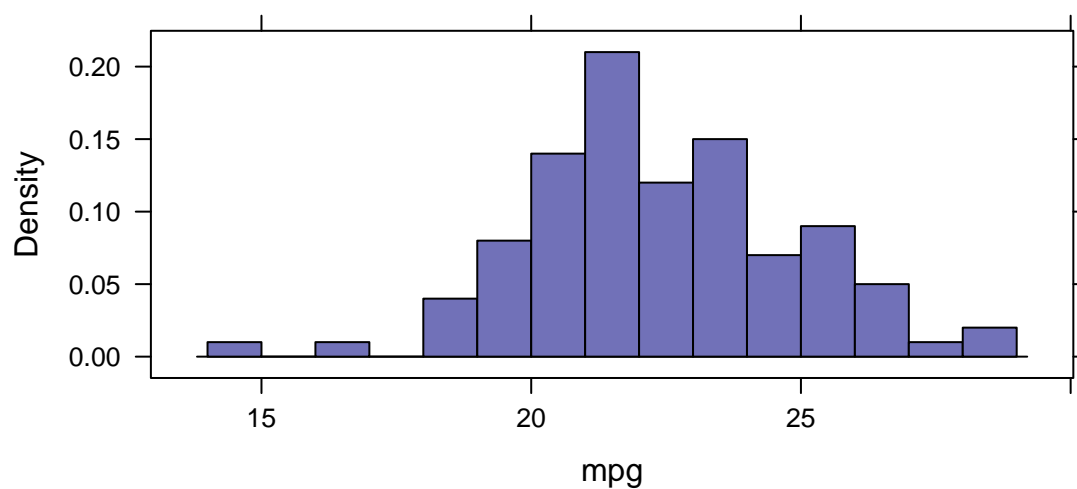
```
qnorm(.9, mean=0, sd=1) # or as a Z-score
```

```
## [1] 1.28
```

Section 5.5: Normal probability plots

See Figure 5.8 on page 129

```
Nissan <-  
read_delim("http://www.amherst.edu/~nhorton/sdm4/data/Nissan.txt",  
  delim="\t")  
histogram(~ mpg, width=1, center=0.5, data=Nissan)
```



```
qqmath(~ mpg, data=Nissan)
```

