# PSY 221A - Homework 4

*Cristopher Garduno Luna*

*10/25/2017*

## Some Useful Functions

The function **znorm** takes in either a raw score or a z score, along with the mean and standard deviation, and returns the z score or the raw score (specify type = "z" for z score and type = "x" for raw score) using the following equation:

$$z = \frac{x - \mu}{\sigma}$$

```
znorm = function(s, m, sd, type) {
  if (type == "z") {
    x = s
    return ((x-m)/sd)
  }
  if (type == "x") {
    z = s
    return (m+(z*sd))
  }
}
```

The function **zshade** shades in the area under the curve between 2 z scores, or to either side of a z score.

```
zshade = function(z, shade = "left") {
  # If more than 2 z scores are given
  if (length(z) > 2) {
    stop("Error: Too many z scores given!")
  }

  # If two z scores are given
  if (length(z) > 1) {
    z1 = min(z)
    z2 = max(z)
    cord.x = c(z1, seq(z1, z2, 0.01), z2)
    cord.y = c(0, dnorm(seq(z1, z2, 0.01)), 0)
    area = paste("Area Between z =", as.character(z1), "and z =", as.character(z2), ": ",
                 as.character(round(pnorm(z2)-pnorm(z1), digits=3)))
    curve(dnorm(x, 0, 1), xlim = c(-4, 4), main = area,
          ylab = "", xlab = "")
    polygon(cord.x, cord.y, col = "skyblue")
  }

  if (length(z)==1) {
    # If a single z score is given
    if (shade == "left") {
      z1 = -4
      z2 = z
      cord.x = c(z1, seq(z1, z2, 0.01), z2)
      cord.y = c(0, dnorm(seq(z1, z2, 0.01)), 0)
```

```
      area = paste("Area to the left of z =", as.character(z2), ": ",
                   as.character(round(pnorm(z2), digits=3)))
      curve(dnorm(x, 0, 1), xlim = c(-4, 4), main = area,
            ylab = "", xlab = "")
      polygon(cord.x, cord.y, col = "skyblue")
    }
    if (shade == "right") {
      z1 = z
      z2 = 4
      cord.x = c(z1, seq(z1, z2, 0.01), z2)
      cord.y = c(0, dnorm(seq(z1, z2, 0.01)), 0)
      area = paste("Area to the right of z =", as.character(z1), ": ",
                   as.character(round(1-pnorm(z1), digits=3)))
      curve(dnorm(x, 0, 1), xlim = c(-4, 4), main = area,
            ylab = "", xlab = "")
      polygon(cord.x, cord.y, col = "skyblue")
    }
  }
}
```

# Chapter 4

When conducting your z score problems, please sketch a normal curve.

## A1

If you convert each score in a set of scores to a z score, which of the following will be true about the resulting set of z scores?

**The variance will equal 1.**

## A2

The distribution of body weights for adults is somewhat positively skewed—there is much more room for people to be above average than below. If you take the mean weights for random groups of 10 adults each and form a new distribution, how will this new distribution compare to the distribution of individuals?

**The new distribution will more closely resemble the normal distribution.**

## A3

Assume that the mean height for adult women ($\mu$) is 65 inches, and that the standard deviation ($\sigma$) is 3 inches.

```
m = 65; sd = 3
```

a. What is the z score for a woman who is exactly 5 feet tall? Who is 5 feet 5 inches tall?

```
cat("Z score for woman that is 5 feet tall          : ", znorm(5*12, m, sd, "z"))
```

```
## Z score for woman that is 5 feet tall          :  -1.666667
```

```r
cat("Z score for woman that is 5 feet 5 inches tall: ", znorm(5*12+5, m, sd, "z"))
```

```
## Z score for woman that is 5 feet 5 inches tall:  0
```

b. What is the z score for a woman who is 70 inches tall? Who is 75 inches tall? Who is 64 inches tall?

```r
cat("Z score for woman that is 70 inches tall: ", znorm(70, m, sd, "z"))
```

```
## Z score for woman that is 70 inches tall:  1.666667
```

```r
cat("Z score for woman that is 75 inches tall: ", znorm(75, m, sd, "z"))
```

```
## Z score for woman that is 75 inches tall:  3.333333
```

```r
cat("Z score for woman that is 64 inches tall: ", znorm(64, m, sd, "z"))
```

```
## Z score for woman that is 64 inches tall:  -0.3333333
```

c. How tall is a woman whose z score for height is -3? -1.33? -0.3? -2.1?

```r
cat("Height for woman with z score of -3.00: ", znorm(-3, m, sd, "x"), "inches")
```

```
## Height for woman with z score of -3.00:  56 inches
```

```r
cat("Height for woman with z score of -1.33: ", znorm(-1.33, m, sd, "x"), "inches")
```

```
## Height for woman with z score of -1.33:  61.01 inches
```

```r
cat("Height for woman with z score of -0.33: ", znorm(-0.3, m, sd, "x"), "inches")
```

```
## Height for woman with z score of -0.33:  64.1 inches
```

```r
cat("Height for woman with z score of -2.10: ", znorm(-2.1, m, sd, "x"), "inches")
```

```
## Height for woman with z score of -2.10:  58.7 inches
```

d. How tall is a woman whose z score for height is +3? +2.33? +1.7? +.9?

```r
cat("Height for woman with z score of 3.00: ", znorm(3.0,  m, sd, "x"), "inches")
```

```
## Height for woman with z score of 3.00:  74 inches
```

```r
cat("Height for woman with z score of 2.33: ", znorm(2.33, m, sd, "x"), "inches")
```

```
## Height for woman with z score of 2.33:  71.99 inches
```

```r
cat("Height for woman with z score of 1.70: ", znorm(1.7,  m, sd, "x"), "inches")
```

```
## Height for woman with z score of 1.70:  70.1 inches
```

```r
cat("Height for woman with z score of 0.90: ", znorm(0.9,  m, sd, "x"), "inches")
```

```
## Height for woman with z score of 0.90:  67.7 inches
```

### A9

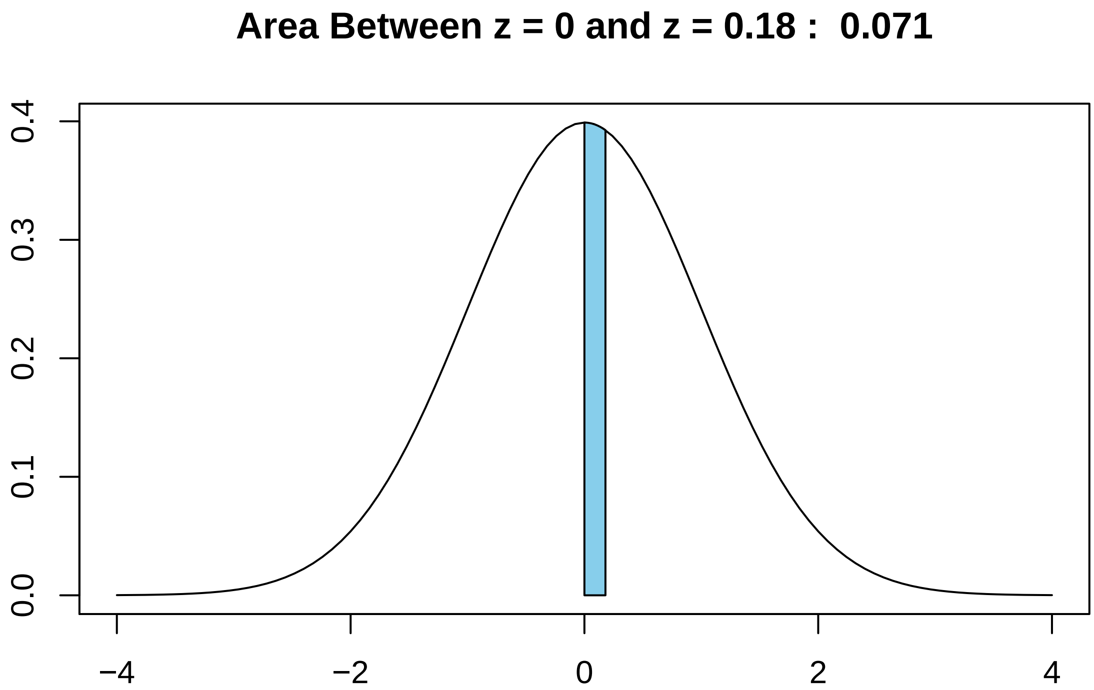
Use Table A.1 to find the area of the normal distribution between the mean and z, when z equals

a. 0.18

```r
cat("Area under the curve between mean and z = 0.18: ", pnorm(0.18)-0.5)
```

```
## Area under the curve between mean and z = 0.18:  0.07142372
```

```r
zshade(c(0.18, 0))
```

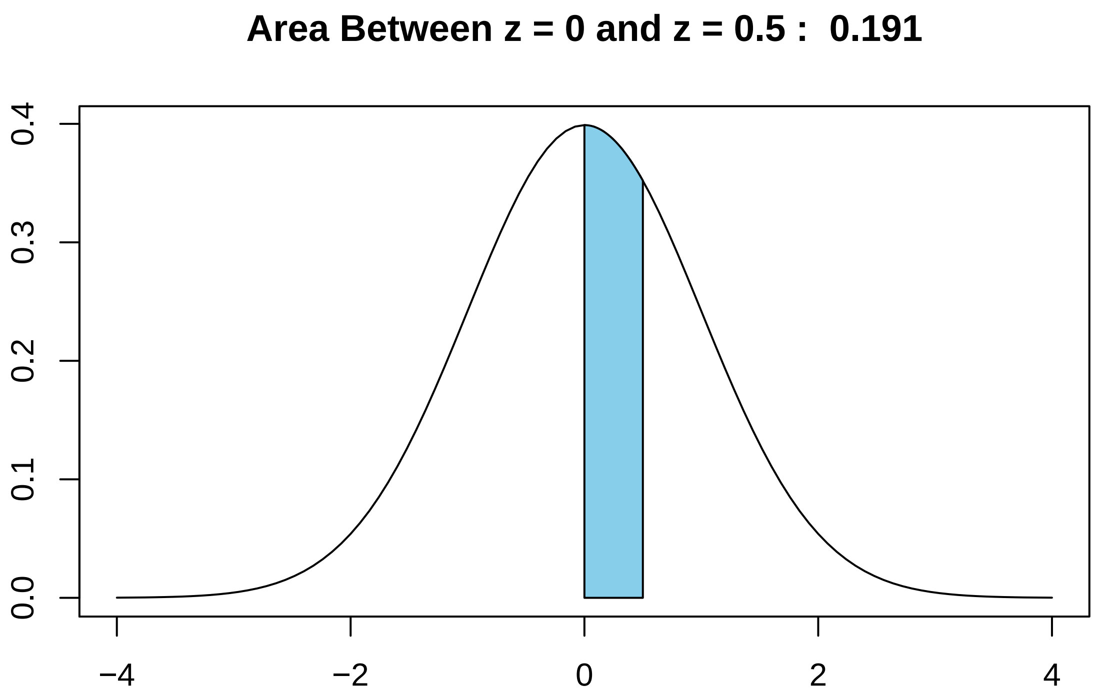



b. 0.50

```r
cat("Area under the curve between mean and z = 0.50: ", pnorm(0.50)-0.5)
```

```
## Area under the curve between mean and z = 0.50:  0.1914625
```

```r
zshade(c(.5, 0))
```

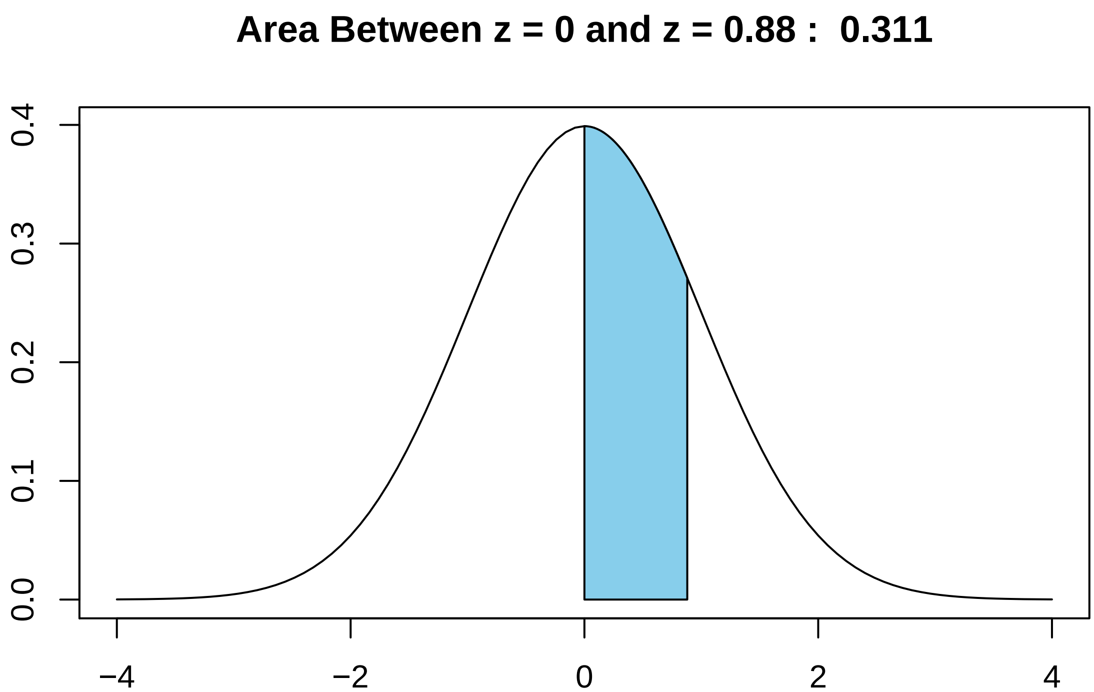



c. 0.88

```r
cat("Area under the curve between mean and z = 0.88: ", pnorm(0.88)-0.5)
```

```
## Area under the curve between mean and z = 0.88:  0.3105703
```

```r
zshade(c(.88, 0))
```

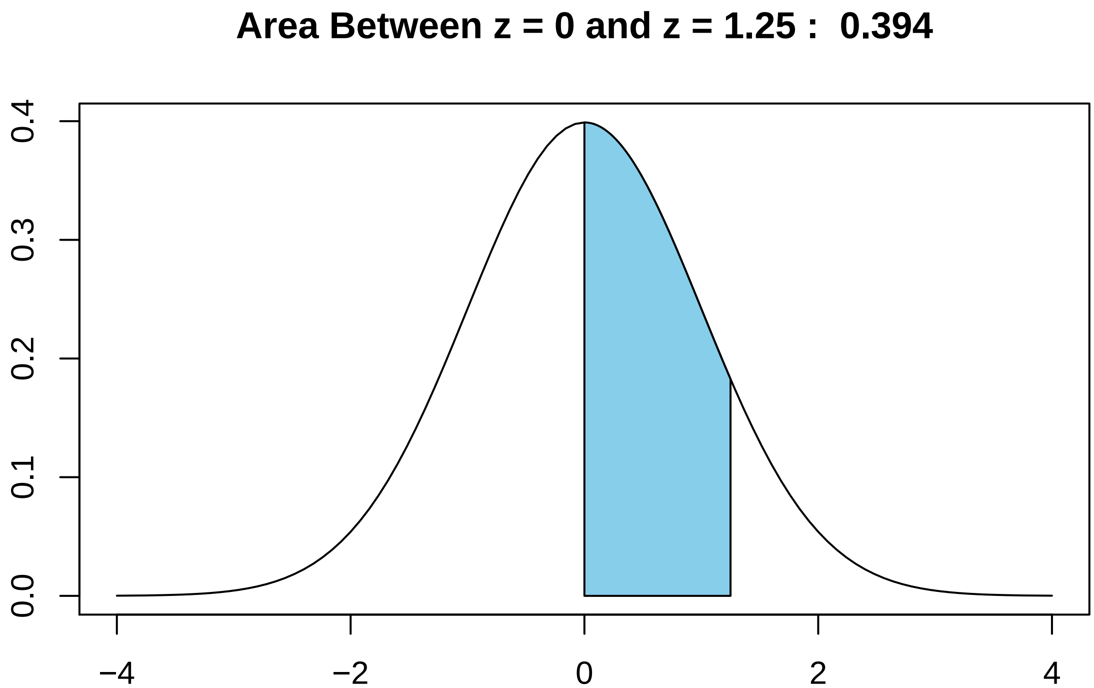



d. 1.25

```r
cat("Area under the curve between mean and z = 1.25: ", pnorm(1.25)-0.5)
```

```
## Area under the curve between mean and z = 1.25:  0.3943502
```

```r
zshade(c(1.25, 0))
```

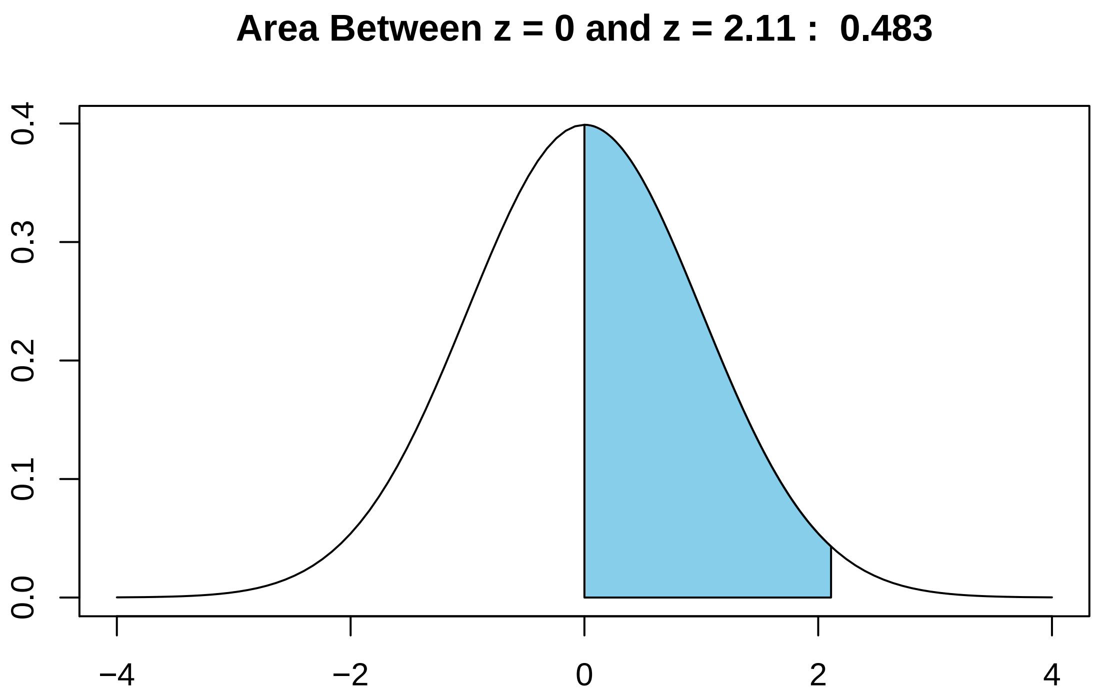



e. 2.11

```
cat("Area under the curve between mean and z = 2.11: ", pnorm(2.11)-0.5)
```

```
## Area under the curve between mean and z = 2.11:  0.4825708
```

```
zshade(c(2.11, 0))
```




### A10

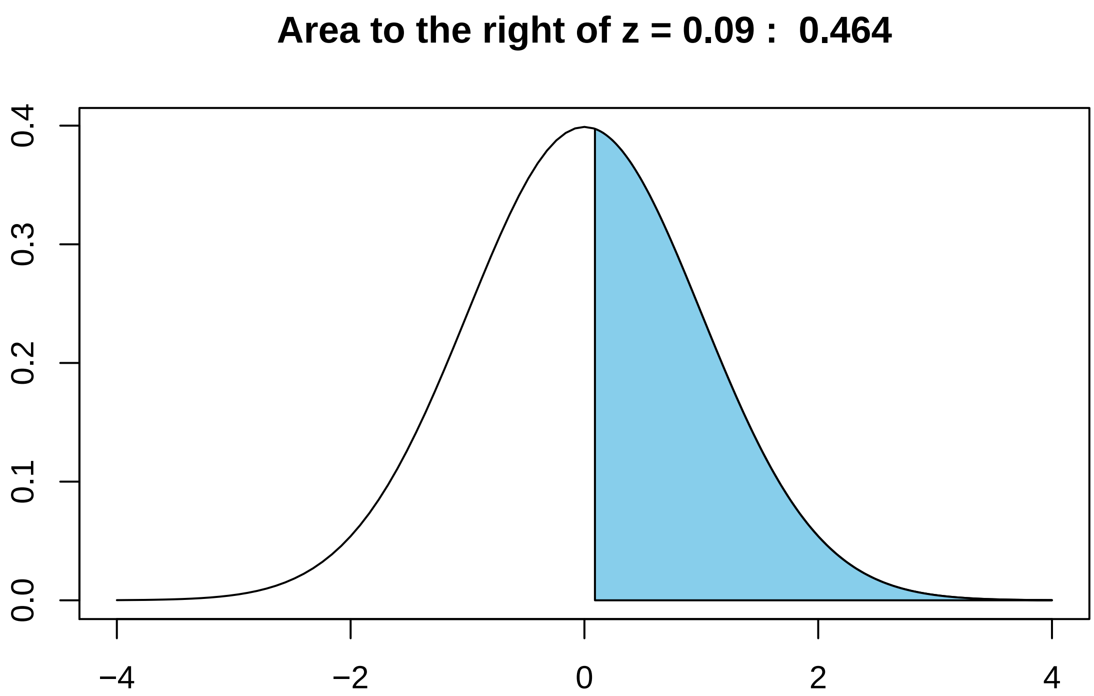
Use Table A.1 to find the area of the normal distribution beyond z, when z equals

a. 0.09

```r
cat("Area under the curve beyond z = 0.09: ", 1-pnorm(0.09))
```

```
## Area under the curve beyond z = 0.09:  0.4641436
```

```r
zshade(0.09, "right")
```

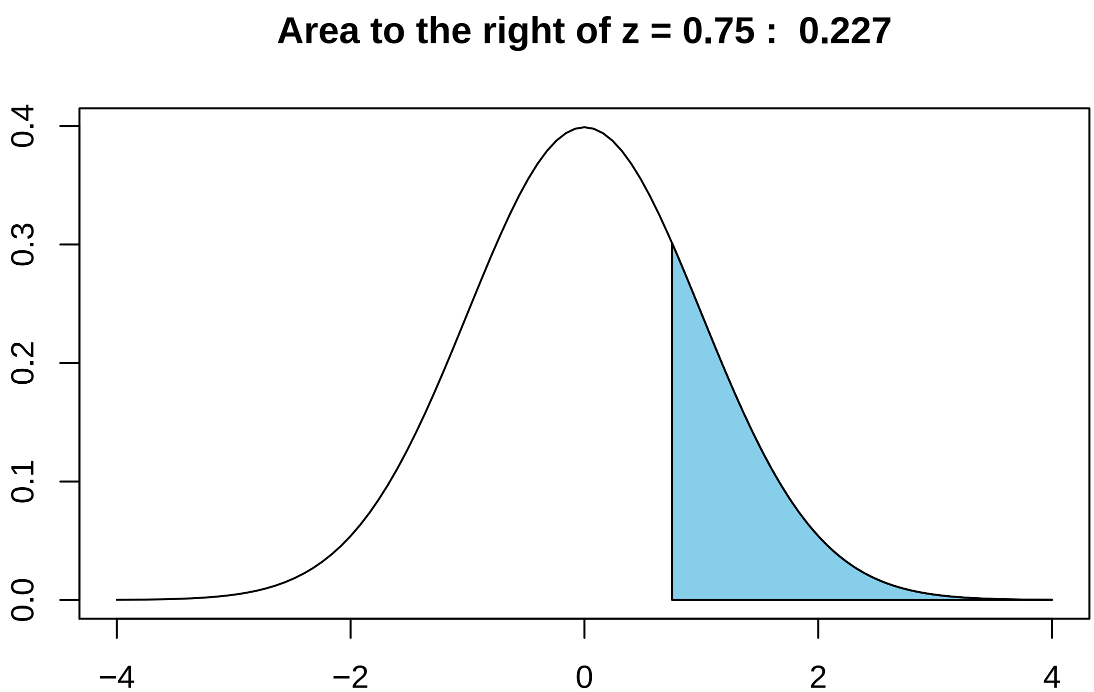



b. 0.75

```r
cat("Area under the curve beyond z = 0.75: ", 1-pnorm(0.75))
```

```
## Area under the curve beyond z = 0.75:  0.2266274
```

```r
zshade(.75, "right")
```

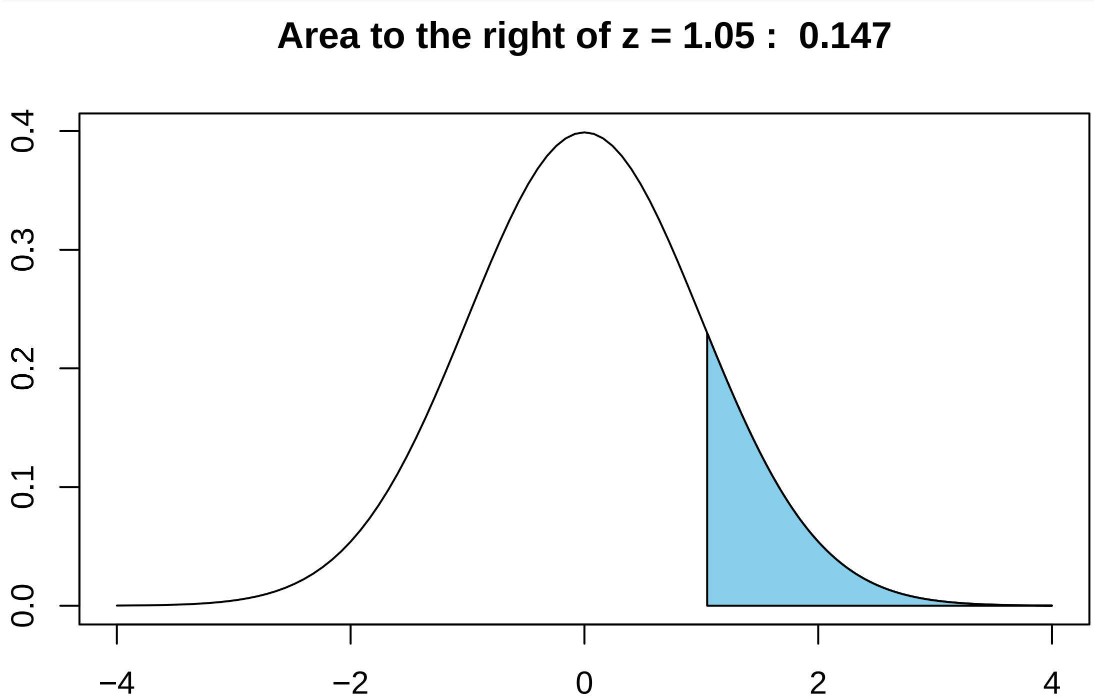



c. 1.05

```r
cat("Area under the curve beyond z = 1.05: ", 1-pnorm(1.05))
```

```
## Area under the curve beyond z = 1.05:  0.1468591
```

```r
zshade(1.05, "right")
```



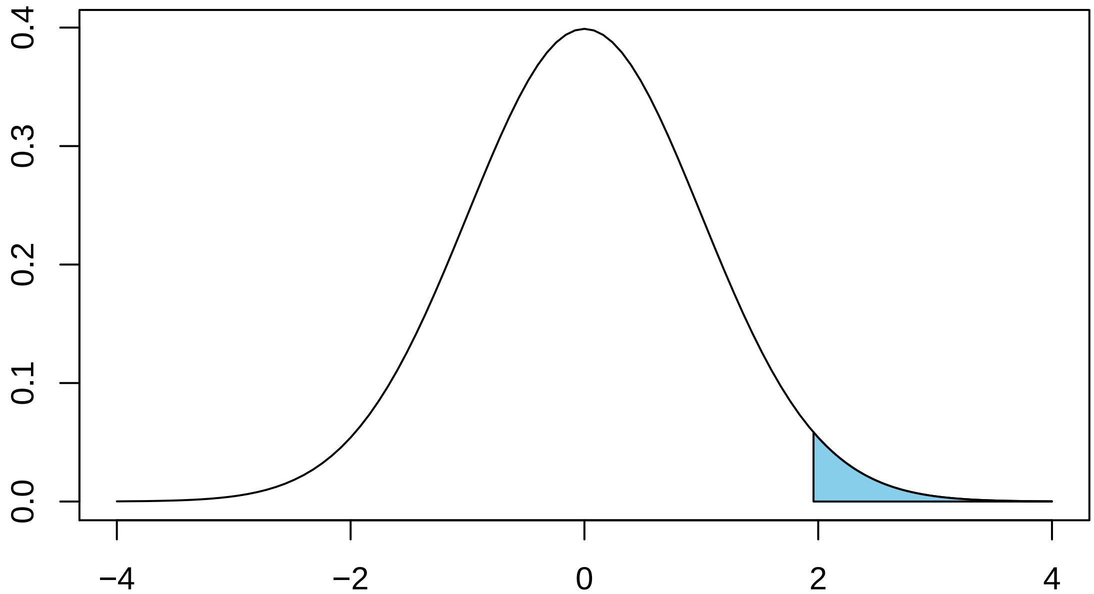

d. 1.96

```r
cat("Area under the curve beyond z = 1.96: ", 1-pnorm(1.96))
```

```
## Area under the curve beyond z = 1.96:  0.0249979
```

```r
zshade(1.96, "right")
```

**Area to the right of z = 1.96 :  0.025**



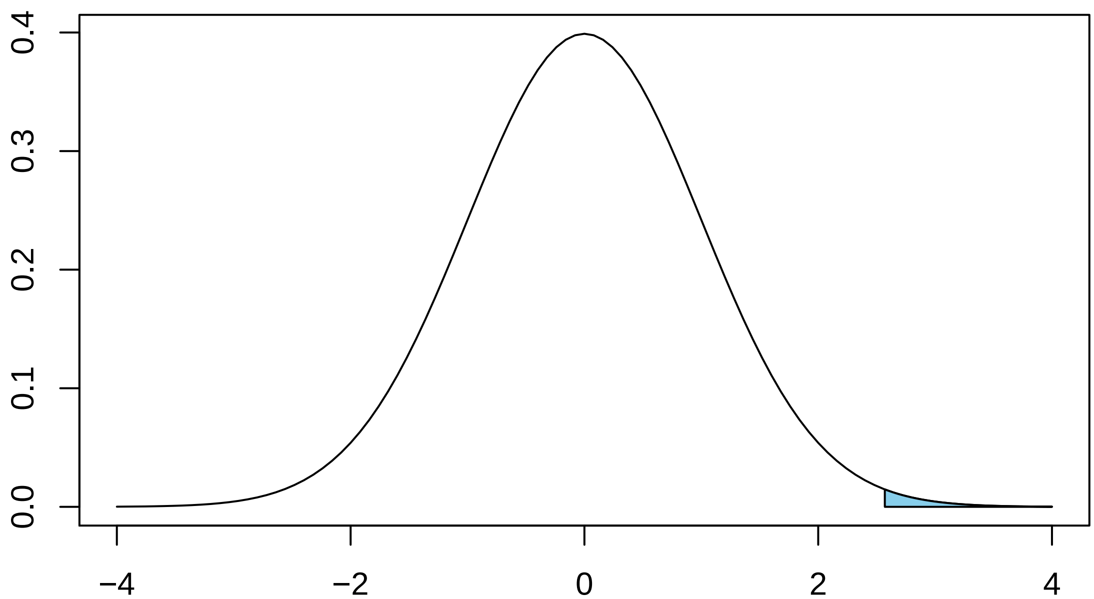

e. 2.57

```r
cat("Area under the curve beyond z = 2.57: ", 1-pnorm(2.57))
```

```
## Area under the curve beyond z = 2.57:  0.005084926
```

```r
zshade(2.57, "right")
```

**Area to the right of z = 2.57 :  0.005**



### B1

Suppose that a large Introduction to Psychology class has taken a midterm exam, and the scores are normally distributed (approximately) with $\mu = 75$ and $\sigma = 9$. What is the percentile rank (PR) for a student

```
m = 75; sd = 9
```

a. Who scores 90?

```
cat("Percentile Rank for score 90: ", 100*pnorm(znorm(90, m, sd, "z")), "%")
```

```
## Percentile Rank for score 90:  95.22096 %
```

b. Who scores 70?

```
cat("Percentile Rank for score 70: ", 100*pnorm(znorm(70, m, sd, "z")), "%")
```

```
## Percentile Rank for score 70:  28.92574 %
```

c. Who scores 60?

```
cat("Percentile Rank for score 60: ", 100*pnorm(znorm(60, m, sd, "z")), "%")
```

```
## Percentile Rank for score 60:  4.779035 %
```

d. Who scores 94?

```
cat("Percentile Rank for score 94: ", 100*pnorm(znorm(94, m, sd, "z")), "%")
```

```
## Percentile Rank for score 94:  98.26186 %
```

### B2

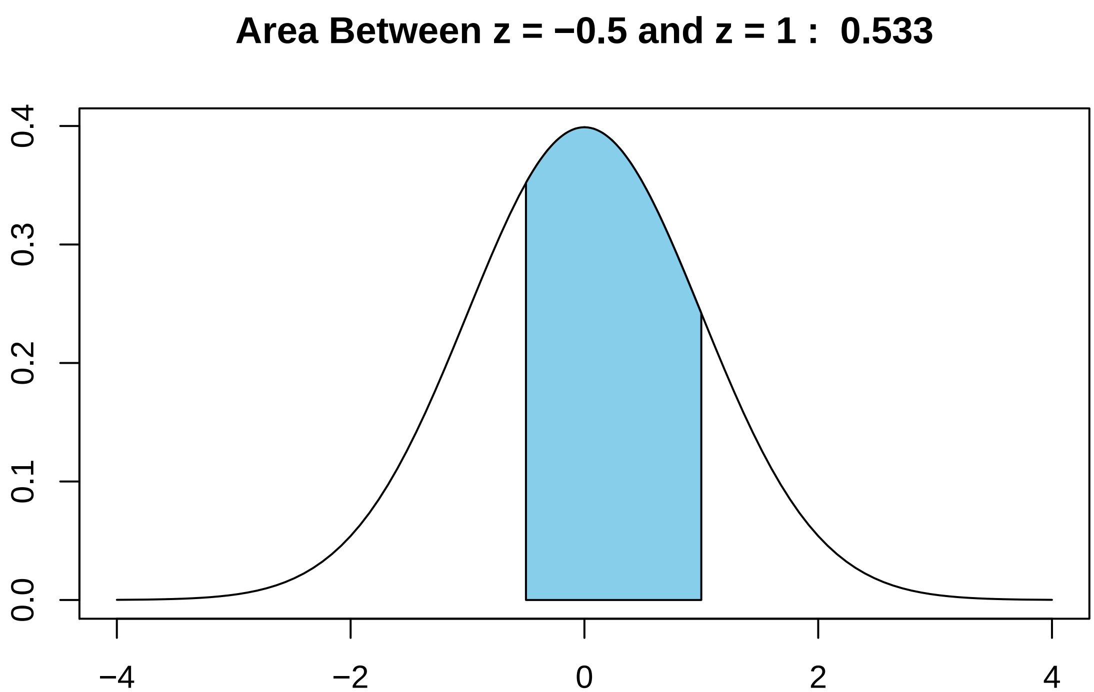
Find the area between

a. $z = -0.5$ and $z = +1.0$

```
cat("Area between two given z scores: ", pnorm(1)-pnorm(0.5))
```

```
## Area between two given z scores:  0.1498823
```

```
zshade(c(-.5, 1))
```

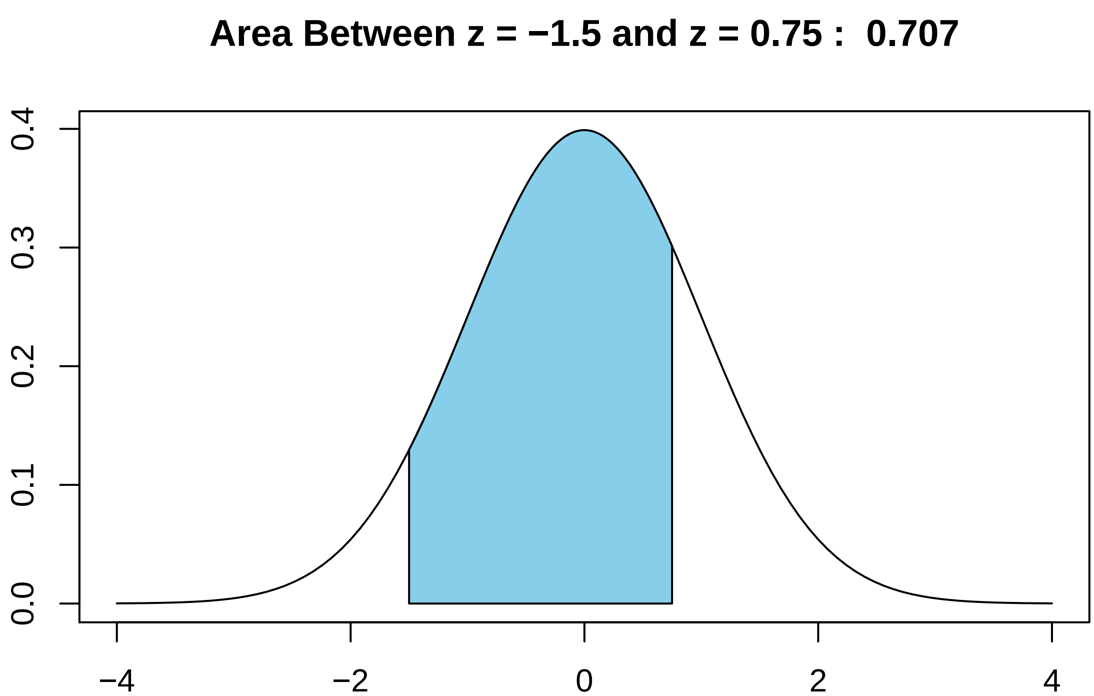



b. $z = -1.5$ and $z = +0.75$

```r
cat("Area between two given z scores: ", pnorm(0.75)-pnorm(-1.5))
```

```
## Area between two given z scores:  0.7065654
```

```r
zshade(c(-1.5, .75))
```

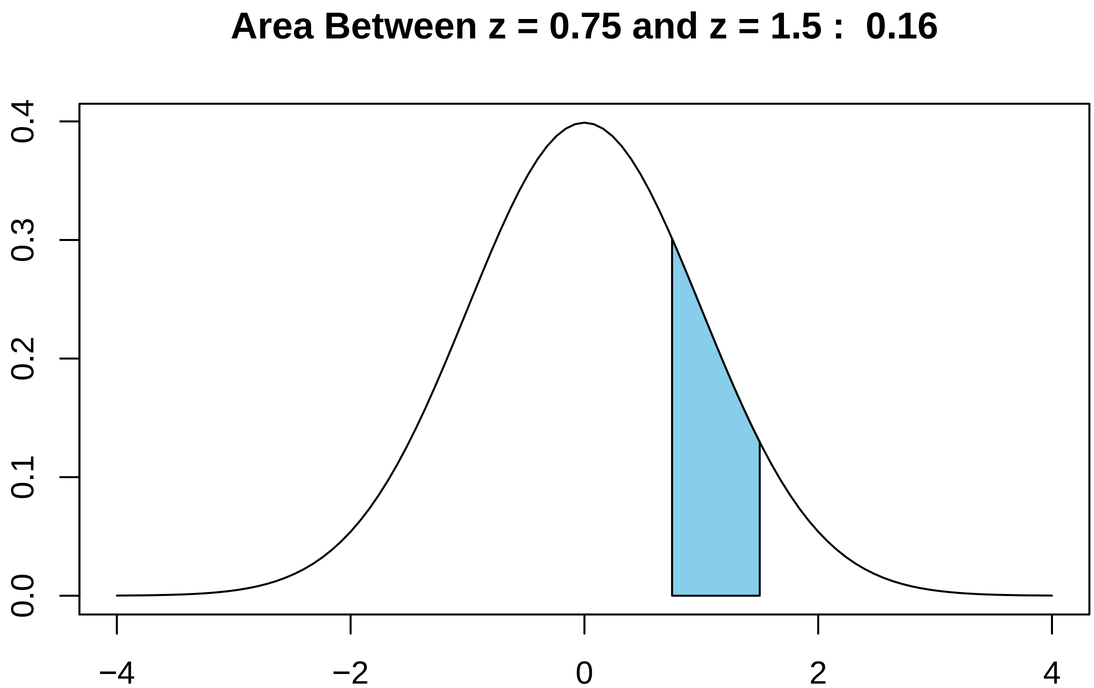



c. $z = +0.75$ and $z = +1.5$

```r
cat("Area between two given z scores: ", pnorm(1.5)-pnorm(0.75))
```

```
## Area between two given z scores:  0.1598202
```

```r
zshade(c(0.75, 1.5))
```

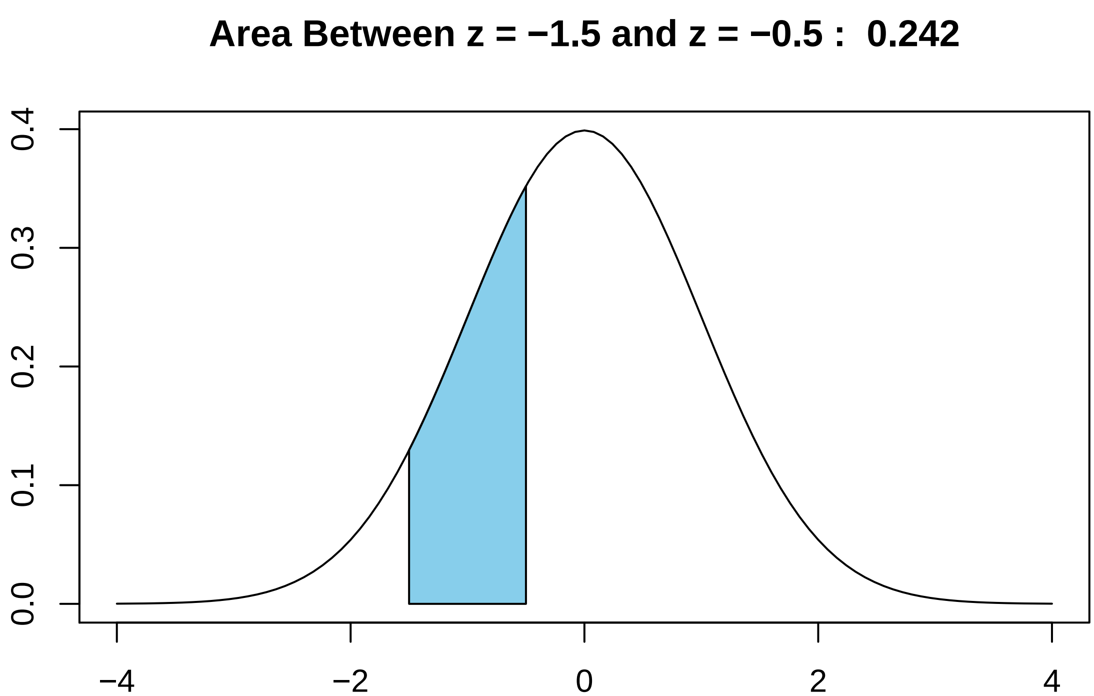



d. $z = -0.5$ and $z = -1.5$

```r
cat("Area between two given z scores: ", pnorm(-.5)-pnorm(-1.5))
```

```
## Area between two given z scores:  0.2417303
```

```r
zshade(c(-.5, -1.5))
```



### B6

A teacher thinks her class has an unusually high IQ, because her 36 students have an average IQ (X) of 108. If the population mean is 100 and $\sigma = 15$,

a. What is the z score for this class?

b. What percentage of classes (n = 36, randomly selected) would be even higher on IQ?

### B9

Suppose that the average person sleeps 8 hours each night and that $\sigma = 0.7$ hour(s).

a. If a group of 50 joggers is found to sleep an average of 7.6 hours per night, what is the z score for this group?

b. If a group of 200 joggers also has a mean of 7.6, what is the z score for this larger group?

c. Comparing your answers to parts a and b, can you determine the mathematical relation between sample size and z (when X remains constant)?

### C1

## Additional Problems

```
library(haven)
```

```
## Warning: package 'haven' was built under R version 3.2.5
```

```
path = "~/Desktop/UCSB/fall2017/psych221a/hw/ihno.sav"
data = as.data.frame(read_sav(path))
```

1) Combine Mathquiz and Statquiz. Normalize variables before combining.

a) Report mean, standard deviation, lowest/highest observed score

```
mathquiz = as.vector(data$Mathquiz); statquiz = as.vector(data$Statquiz)
summary(mathquiz)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    9.00   22.00   30.00   29.07   35.00   49.00      15
```

```
sd(mathquiz, na.rm = TRUE)
```

```
## [1] 9.480291
```

```
summary(statquiz)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1.00    6.00    7.00    6.86    8.00   10.00
```

```
sd(statquiz, na.rm = TRUE)
```

```
## [1] 1.699792
```

b) Compute Pearson Correlation coefficient between mathquiz and statquiz.

```
# Compute correlation of complete cases
cor(mathquiz, statquiz, "complete.obs", "pearson")
```

```
## [1] 0.5078723
```

c) Normalize mathquiz and statquiz. Call the new variables zmathquiz and zstatquiz. Code below normalizes vectors but does not produce output.

```
zmathquiz = c()
zstatquiz = c()
for (i in 1:length(mathquiz)) {
  # Calculate means and standard deviations
  zmathmean = mean(mathquiz, na.rm = TRUE)
  zmathsd   = sd(mathquiz, na.rm = TRUE)
  zstatmean = mean(statquiz, na.rm = TRUE)
  zstatsd   = sd(statquiz, na.rm = TRUE)

  # normalize and add to z vector
  zmathquiz[i] = znorm(mathquiz[i], zmathmean, zmathsd, "z")
  zstatquiz[i] = znorm(statquiz[i], zstatmean, zstatsd, "z")
}
```

d) Compute mean of 2 standardized variables. Call it 'mperform'.

```
zmathquiz = zmathquiz[is.na(zmathquiz)==FALSE]
zstatquiz = zstatquiz[is.na(zstatquiz)==FALSE]
mperform = mean(c(zmathquiz, zstatquiz))
mperform
```

```
## [1] -9.467206e-17
```

e) Compute mean and standard deviation. Explain.

```
mean(c(zmathquiz, zstatquiz))
```

```
## [1] -9.467206e-17
```

```
sd(c(zmathquiz, zstatquiz))
```

```
## [1] 0.9972789
```

The mean is approximately 0 because

$$\sum z = \sum \frac{z-\mu_z}{\sigma_z} = \frac{1}{\sigma_z}\sum (z-\mu_z) = \frac{1}{\sigma_z}((\sum z) - N\mu_z) = \frac{1}{\sigma_z}(N\mu_z - N\mu_z) = 0$$

If $\sum z = 0$ then $\bar{z} = \frac{0}{N} = 0$.

The standard deviation is approximately 1 because

$$\sigma_z = \sqrt{\frac{\sum(z-\mu_z)^2}{N}} = \sqrt{\frac{\sum z^2}{N}} \implies \sigma_z^2 = \frac{\sum z^2}{N} \implies N\sigma_z^2 = \sum z^2$$

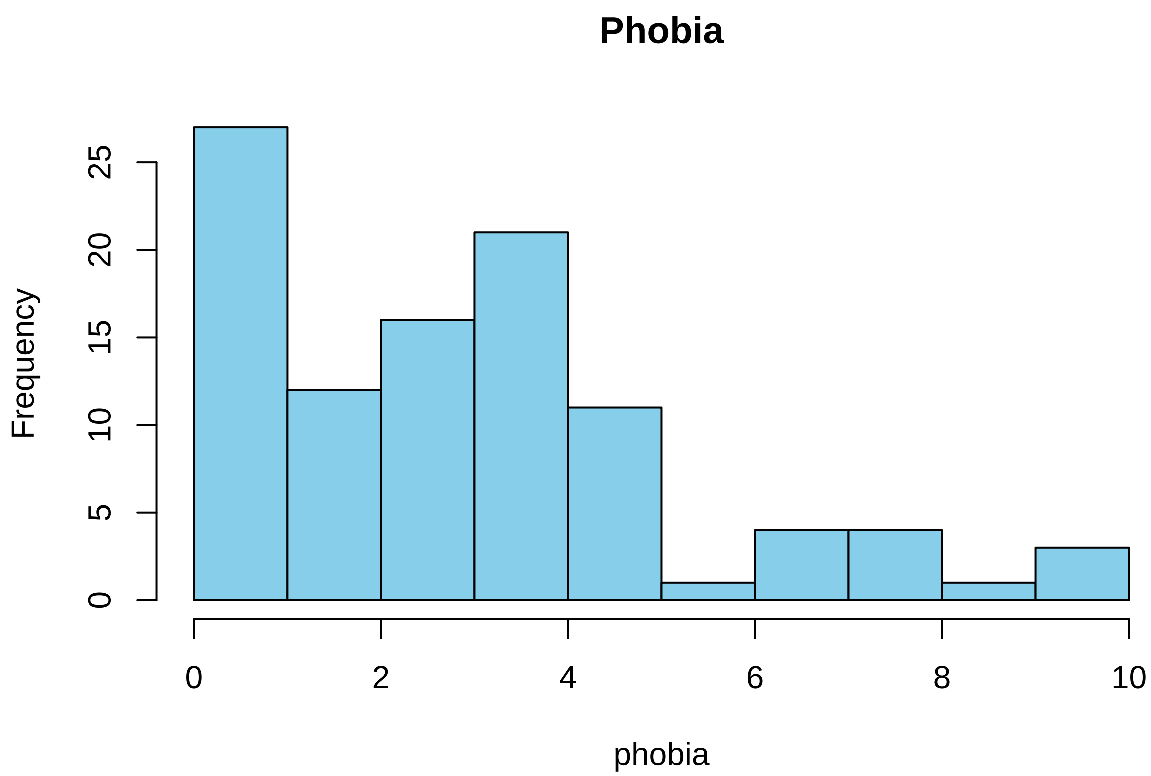
$$\sum z^2 = \sum \frac{(z-\mu_z)^2}{\sigma_z^2} = \frac{1}{\sigma_z^2}\sum z^2 = \frac{1}{\sigma_z^2}(N\sigma_z^2) = N$$

So

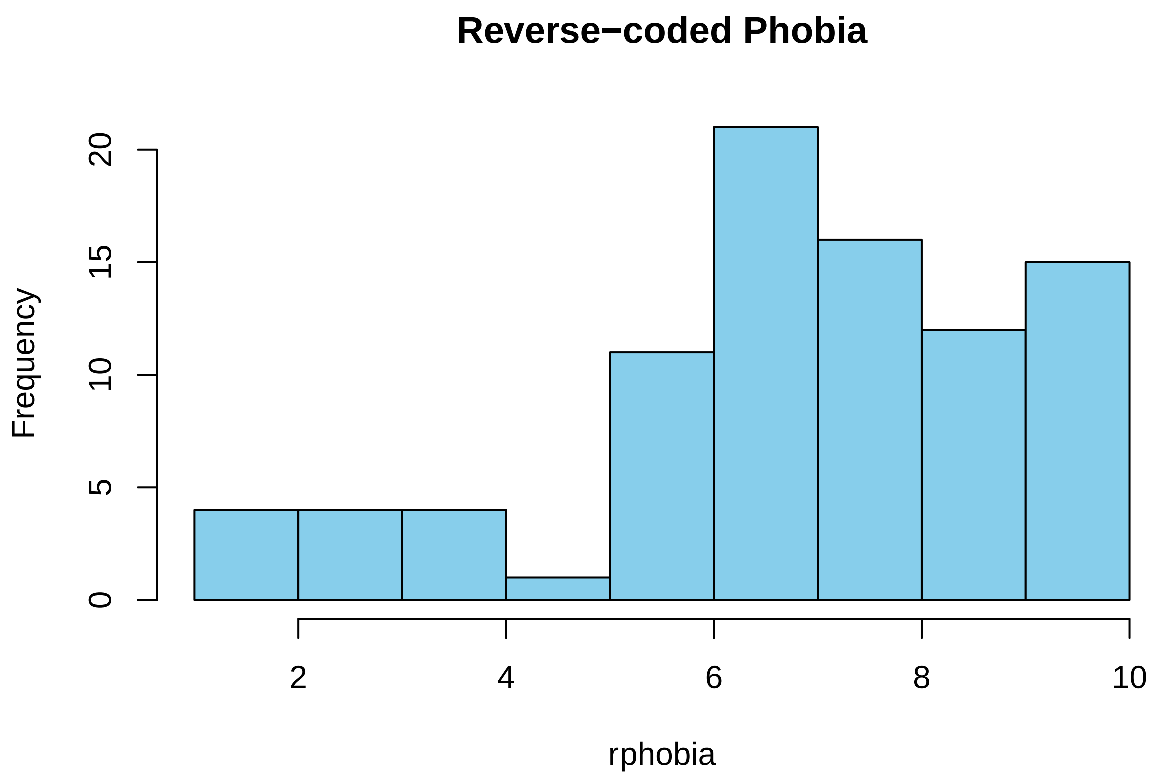
$$\sigma_z = \sqrt{\frac{\sum (z^2) - \frac{(\sum z)^2}{N}}{N}} = \sqrt{\frac{N - \frac{0}{N}}{N}} = \sqrt{\frac{N}{N}} = \pm 1$$

2) Reverse-code math phobia into variable 'rphobia' and generate frequency distribution for phobia and rphobia.

```r
phobia = as.vector(data$Phobia)
rphobia = c()
j = 10
for (i in 1:10) {
  rphobia[phobia==i] = j
  j = j - 1
}
hist(phobia, main = "Phobia", col = "skyblue")
```

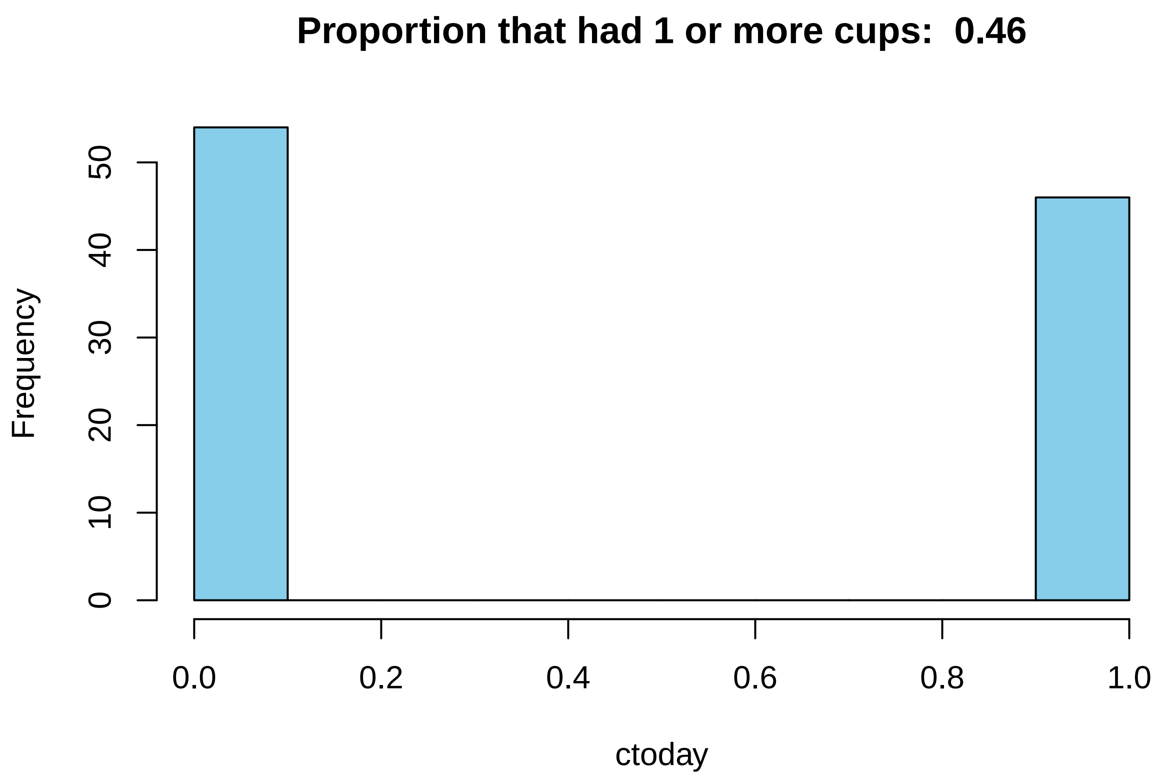



```r
hist(rphobia, main = "Reverse-coded Phobia", col = "skyblue")
```




3) Create variable ‘ctoday’ using if statements to score 0 for no cups that day and 1 for one or more cups.

```
numcups = data$Num_cups
ctoday = c()
for (i in 1:length(as.vector(numcups))) {
  if (numcups[i] == 0) {
    ctoday[i] = 0
  }
  else {
    ctoday[i] = 1
  }
}
oneOrMore = table(ctoday)[[2]]
title = paste("Proportion that had 1 or more cups: ", as.character(oneOrMore/length(numcups)))
hist(ctoday, main = title, col = "skyblue", xlim = c(0, 1))
```




4)

a) Create variable that equals sum of the 3 anxiety measures.

```
sumAnx = as.vector(data$Anx_base) + as.vector(data$Anx_pre) + as.vector(data$Anx_post)
```

b) Create variable that equals average of 3 heart rate measures

```
meanHR = cbind(as.vector(data$Hr_base), as.vector(data$Hr_pre), as.vector(data$Hr_post))
```

c) Create new variable by dividing baseline HR by 60 and give label showing it is beats per second.

```
hr1bps = data$Hr_base / 60
```

d) For each variable, report mean, sd, and standard error.

```
# sumAnx
meAnx = mean(sumAnx, na.rm = TRUE)
sdAnx = sd(sumAnx, na.rm = TRUE)
seAnx = sd(sumAnx, na.rm = TRUE)/sqrt(length(sumAnx))
meAnx
```

```
## [1] 57.41
```

```
sdAnx
```

```
## [1] 13.68631
```

```
seAnx
```

```
## [1] 1.368631
```

```
# meanHR
meHR = mean(meanHR, na.rm = TRUE)
sdHR = sd(meanHR, na.rm = TRUE)
```

```r
seHR = sd(meanHR, na.rm = TRUE)/sqrt(length(meanHR))
meHR
```

```
## [1] 72.97333
```

```r
sdHR
```

```
## [1] 4.473552
```

```r
seHR
```

```
## [1] 0.2582806
```

```r
# hr1bps
meBps = mean(hr1bps, na.rm = TRUE)
sdBps = sd(hr1bps, na.rm = TRUE)
seBps = sd(hr1bps, na.rm = TRUE)/sqrt(length(hr1bps))
meBps
```

```
## [1] 1.2045
```

```r
sdBps
```

```
## [1] 0.05359443
```

```r
seBps
```

```
## [1] 0.005359443
```