Homework Basic Stat Lab

Taylor Vladic

DATA 613, 2023

Contents

1. In Hong Kong, human male height is approximately normally distributed with mean 171.5 cm and standard deviation 5.5 cm. (Use and show R code to produce answers for a – e) You can use any method for part f.

a) What proportion of the Hong Kong population is between 170 cm and 180 cm?

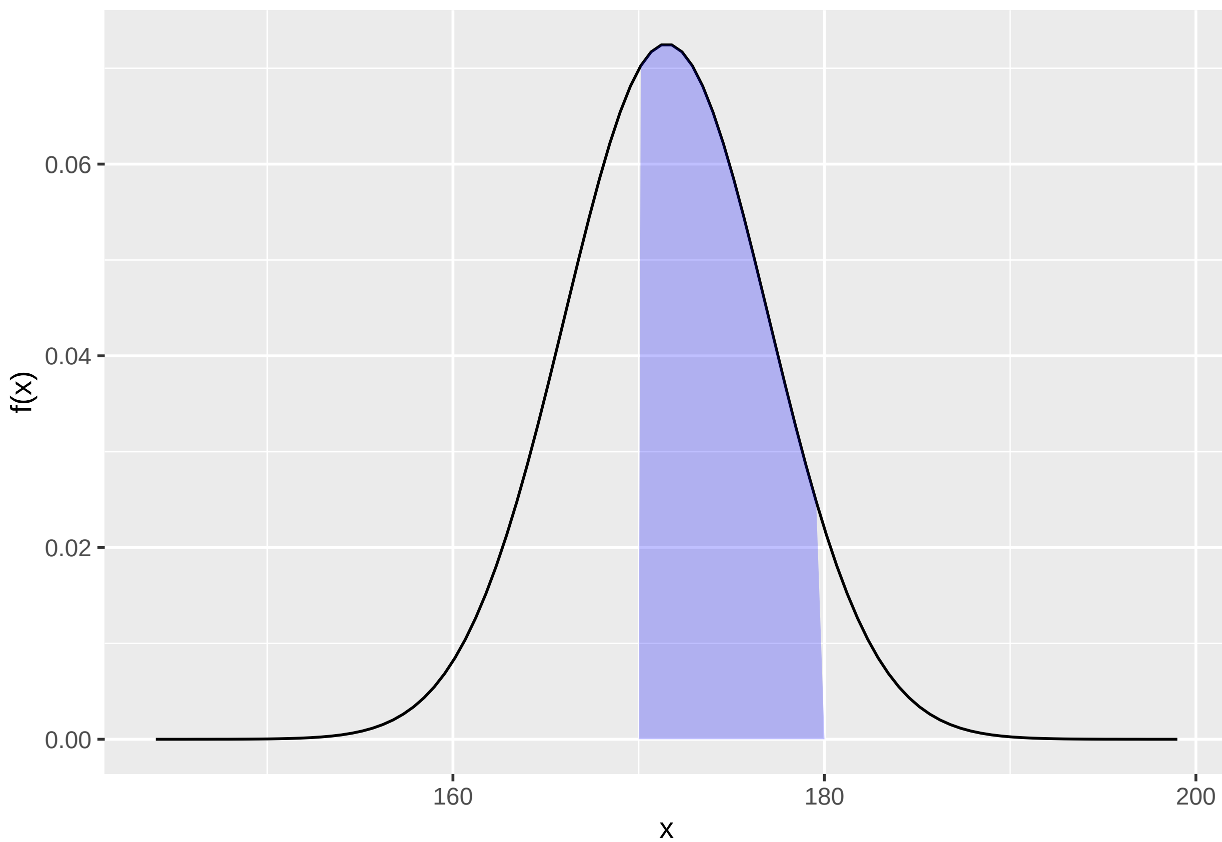
```
pnorm(180, mean = 171.5, sd = 5.5, lower.tail=TRUE) - pnorm(170, mean = 171.5, sd = 5.5, lower.tail=TRU
```

```
## [1] 0.5463504
```

```r
l <- 171.5 - 5*5.5
u <- 171.5 + 5*5.5

x <- seq(l, u, length = 100)
y <- dnorm(x, mean = 171.5, sd = 5.5)

df_1 <- data.frame(x = c(170, x[170 < x & x < 180], 180),
                     y = c(0, y[170 < x & x< 180], 0))
qplot(x, y, geom = "line", ylab = "f(x)") +
  geom_polygon(data = df_1, mapping = aes(x = x, y = y),
               fill = "blue", alpha = 1/4)
```

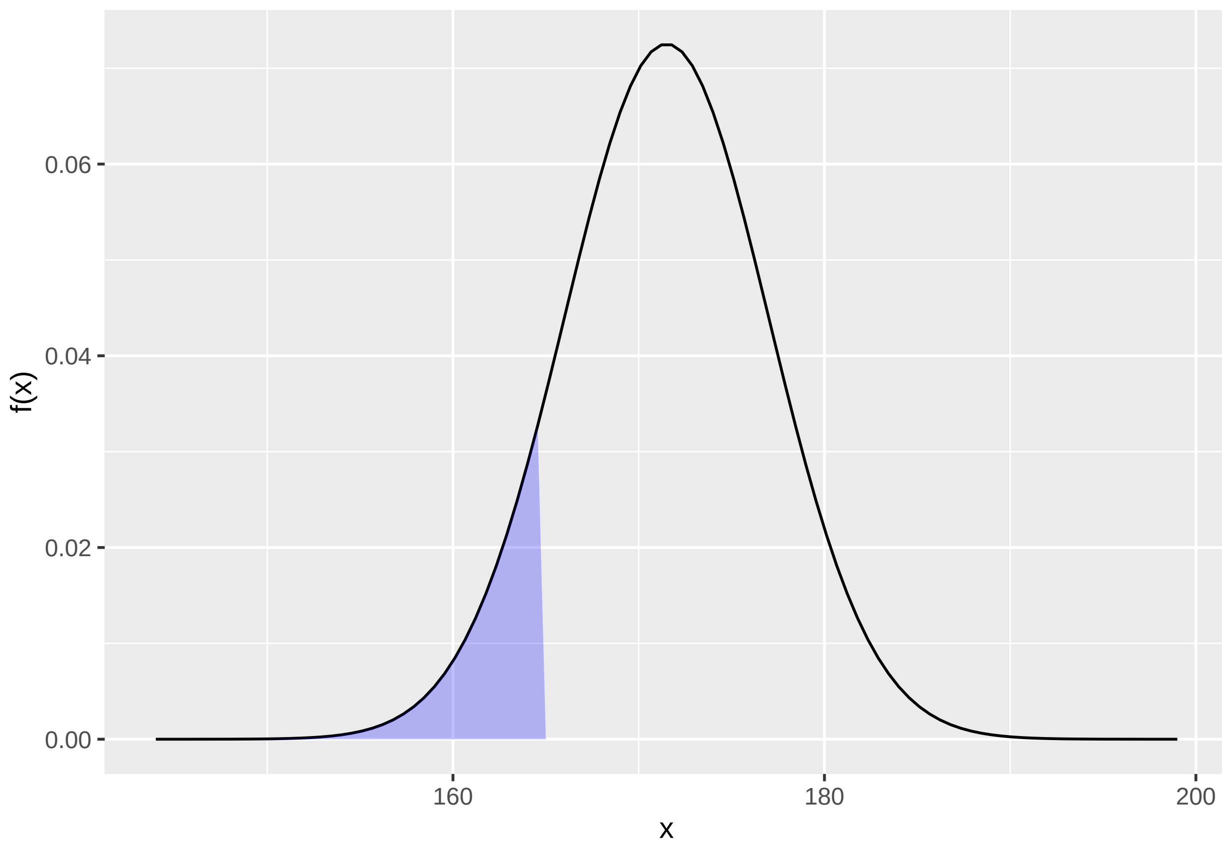


b) What proportion of the Hong Kong population is less that 165 cm ?

```r
pnorm(165, mean = 171.5, sd = 5.5, lower.tail=TRUE)
```

```
## [1] 0.1186389
```

```r
df_2 <- data.frame(x = c(min(x), x[x < 165], 165),
                    y = c(0, y[x < 165], 0))
qplot(x, y, geom = "line", ylab = "f(x)") +
  geom_polygon(data = df_2, mapping = aes(x = x, y = y),
               fill = "blue", alpha = 1/4)
```

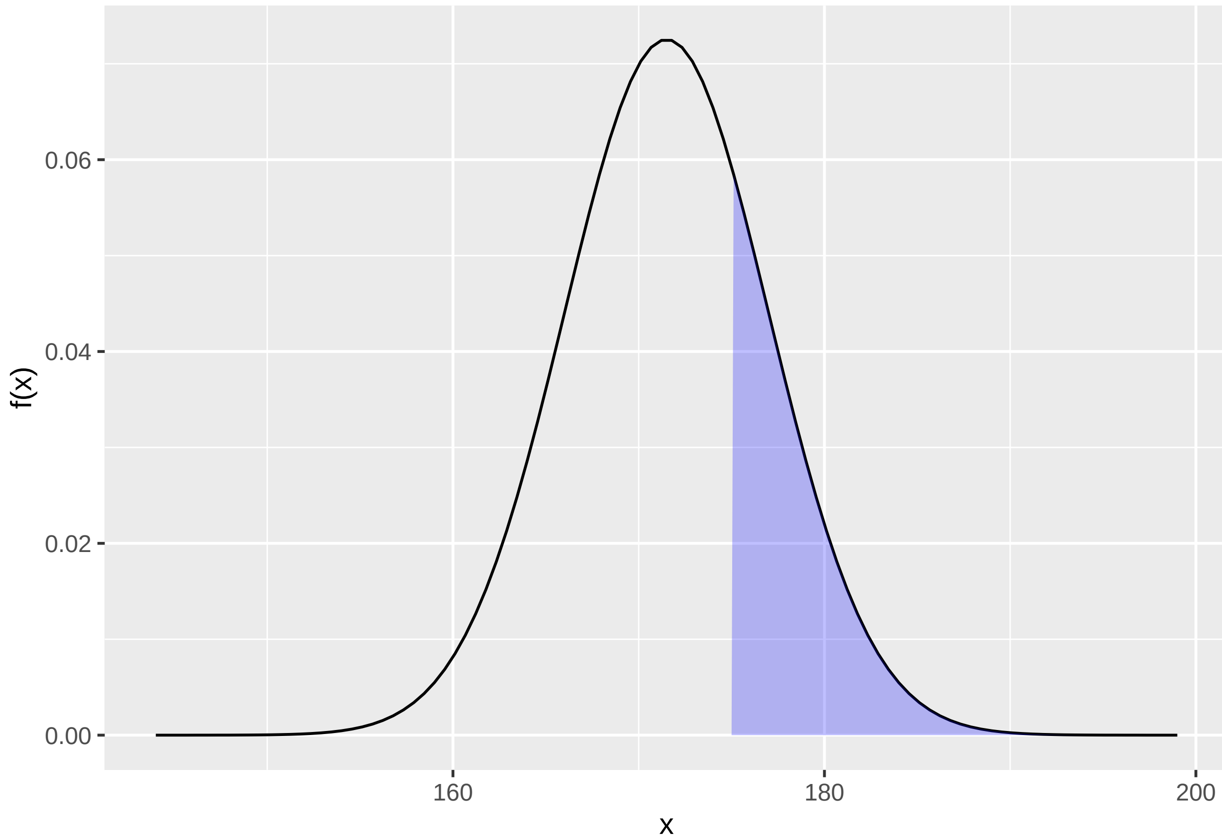


c) What proportion of the Hong Kong population is greater than 175?

```
pnorm(175, mean = 171.5, sd = 5.5, lower.tail=FALSE)
```

```
## [1] 0.2622697
```

```
df_3 <- data.frame(x = c(175, x[x > 175], max(x)),
                   y = c(0, y[x > 175], 0))
qplot(x, y, geom = "line", ylab = "f(x)") +
  geom_polygon(data = df_3, mapping = aes(x = x, y = y),
               fill = "blue", alpha = 1/4)
```

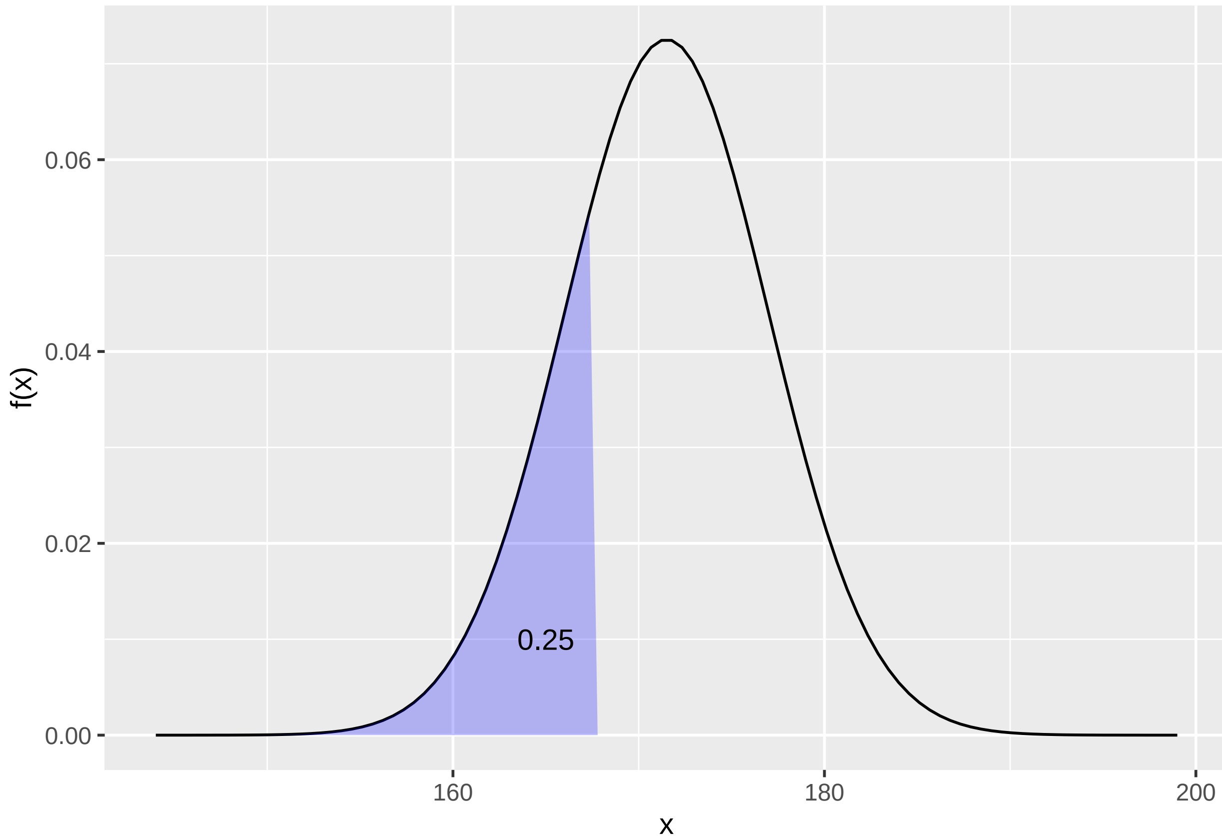


d) The proportion .25 is less than what male height value ?

```
qnorm(p = .25, mean = 171.5, sd = 5.5, lower.tail = TRUE)
```

```
## [1] 167.7903
```

```
df_4 <- data.frame(x = c(min(x), x[x < 167.7903], 167.7903),
                   y = c(0, y[x < 167.7903], 0))
qplot(x, y, geom = "line", ylab = "f(x)") +
  geom_polygon(data = df_4, mapping = aes(x = x, y = y),
               fill = "blue", alpha = 1/4) +
  annotate(geom = "text", x = 165, y = 0.01, label = 0.25, color = "black")
```

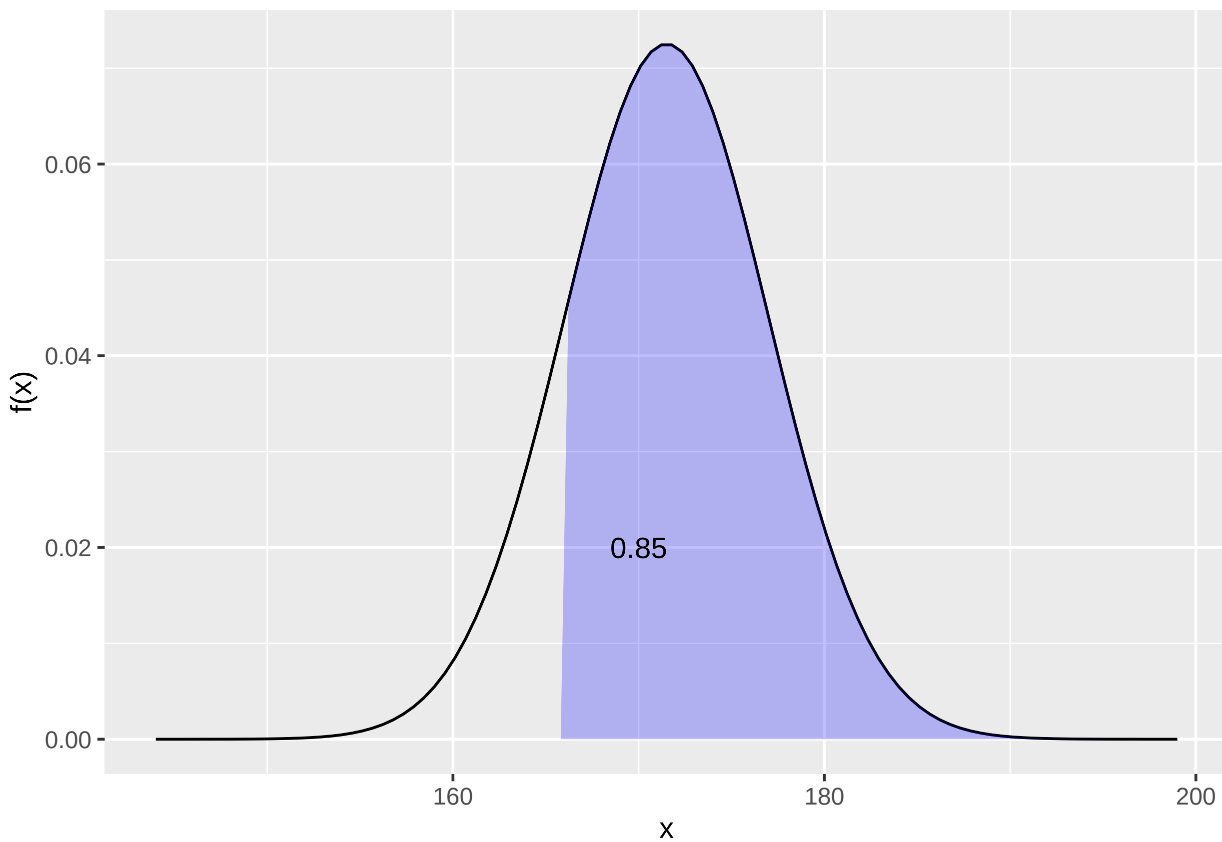


e) The proportion .85 is greater than what male height value ?

```
qnorm(p = .85, mean = 171.5, sd = 5.5, lower.tail = FALSE)
```

```
## [1] 165.7996
```

```
df_5 <- data.frame(x = c(165.7996, x[x > 165.7996], max(x)),
                   y = c(0, y[x > 165.7996], 0))
qplot(x, y, geom = "line", ylab = "f(x)") +
  geom_polygon(data = df_5, mapping = aes(x = x, y = y),
               fill = "blue", alpha = 1/4) +
    annotate(geom = "text", x = 170, y = 0.02, label = 0.85, color = "black")
```



f) The proportion of .68 is between what two male height values ?

```
qnorm(p = 0.16, mean = 171.5, sd = 5.5)
```

```
## [1] 166.0305
```

```
qnorm(p = 0.16, mean = 171.5, sd = 5.5, lower.tail = FALSE)
```

```
## [1] 176.9695
```

2. Use and show R code to determine the height of the standard normal curve at a Z value of 2.5.

```
dnorm(2.5, 0, 1)
```

```
## [1] 0.0175283
```

3. Use and show R code to determine the height of the standard normal curve at a Z value of .4.

```
dnorm(.4, 0, 1)
```

```
## [1] 0.3682701
```

4. Use and show R code to determine the height of a t distribution curve for a t value of 1.5 with 3 degrees of freedom.

```
dt(1.5, 3)
```

```
## [1] 0.1200172
```

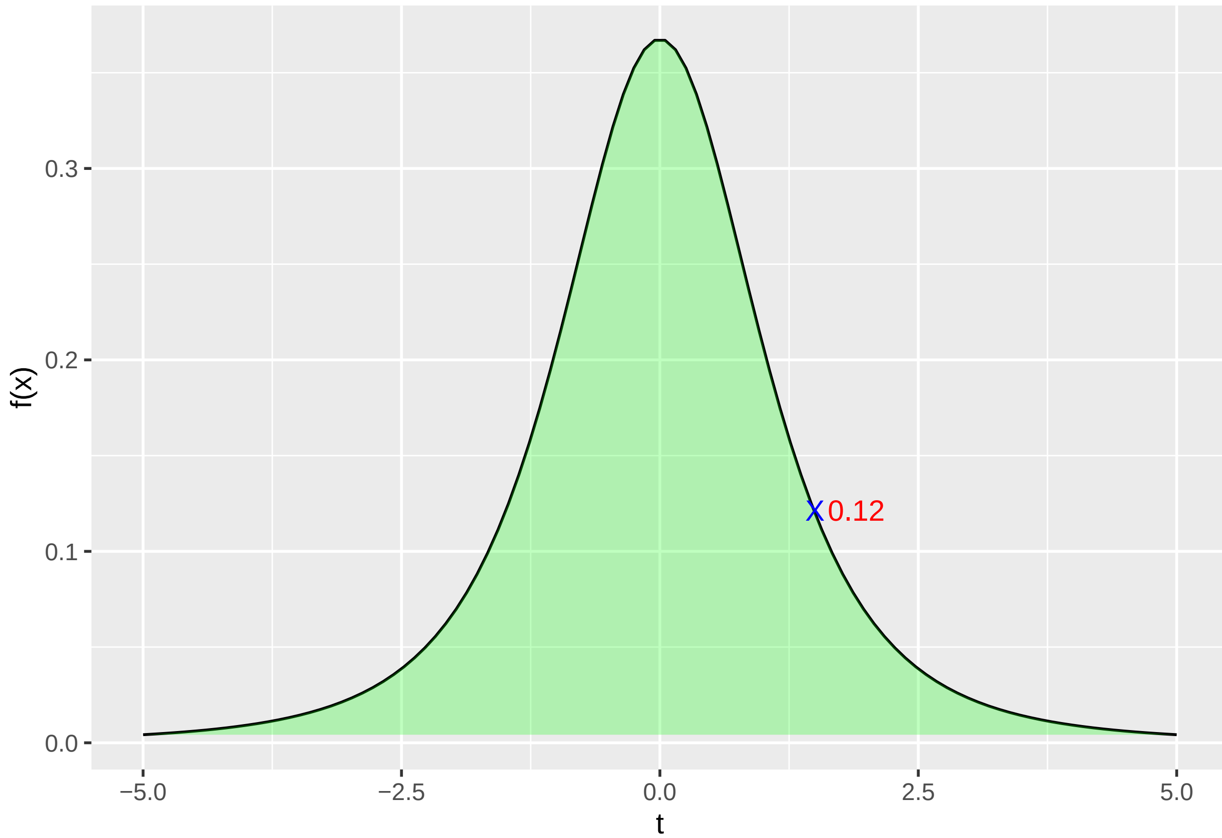
```
t <- seq(-5, 5, length = 100)
y_t <- dt(t, df = 3)
df_t <- tibble(t, y_t)
head(df_t)
```

```
## # A tibble: 6 x 2
##       t     y_t
##   <dbl>   <dbl>
## 1 -5    0.00422
## 2 -4.90 0.00454
## 3 -4.80 0.00489
## 4 -4.70 0.00527
## 5 -4.60 0.00568
## 6 -4.49 0.00614
```

```
dt(1.5, 3)
```

```
## [1] 0.1200172
```

```
 qplot(t, y_t, geom = "line", ylab = "f(x)") +
   geom_polygon(data = df_t, mapping = aes(x = t, y = y_t),
               fill = "green", alpha = 1/4) +
   annotate(geom = "text", x = 1.5, y = .12+0.0015, label = "X", color = "blue") +
   annotate(geom = "text", x = 1.5 + 0.4, y = .12+ 0.0015,
            label = round(.12,2), color = "red")
```



5. Use and show R code to determine the height of a t distribution curve for a t value of -2 with 2 degrees of freedom.

```
dt(-2, 2)
```

```
## [1] 0.06804138
```

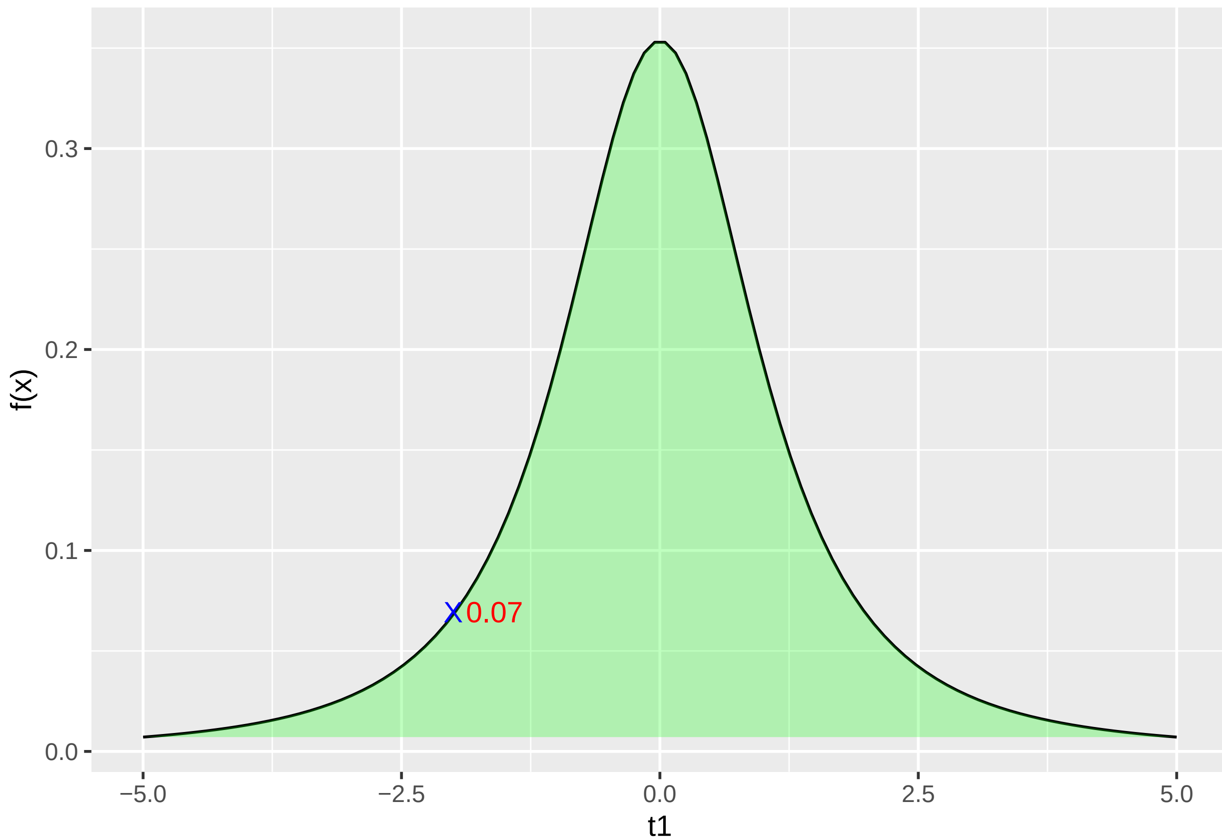
```
t1 <- seq(-5, 5, length = 100)
y_t1 <- dt(t, df = 2)
df_t1 <- tibble(t1, y_t1)
head(df_t1)
```

```
## # A tibble: 6 x 2
##      t1    y_t1
##   <dbl>   <dbl>
## 1 -5    0.00713
## 2 -4.90 0.00754
## 3 -4.80 0.00799
## 4 -4.70 0.00847
## 5 -4.60 0.00899
## 6 -4.49 0.00956
```

```r
dt(-2, 2)
```

```
## [1] 0.06804138
```

```r
qplot(t1, y_t1, geom = "line", ylab = "f(x)") +
  geom_polygon(data = df_t1, mapping = aes(x = t1, y = y_t1),
               fill = "green", alpha = 1/4) +
  annotate(geom = "text", x = -2, y = .068 + 0.0015, label = "X", color = "blue") +
  annotate(geom = "text", x = -2 + 0.4, y = .068 + 0.0015,
           label = round(.068,2), color = "red")
```



6. For a t distribution with 3 degrees of freedom, use and show R code that will find the proportion less than 2.5.

```r
pt(2.5, 3)
```

```
## [1] 0.9561467
```

7. For a t distribution with 1 degree of freedom, use and show R code that will find the proportion that is greater than 1.75.

```
pt(1.75, 1, lower.tail = FALSE)
```

```
## [1] 0.1652493
```

8. For a t distribution with 2 degrees of freedom, use and show R code that will find the value immediately above a proportion of .355.

```
qt(0.355, 2, lower.tail = FALSE)
```

```
## [1] 0.4285376
```

9. An educator believes that new directed reading activities in the classroom will help elementary school pupils improve some aspects of their ability. She arranges for a third -grade class of 23 students to take part in these activities for an eight-week period. A control classroom of 23 third graders follows the same curriculum without the activities. At the end of the eight weeks, all students are given a Degree of Reading Power (DRP) test, which measures the aspects of reading ability that the treatment is designed to improve. The sample data performance results are provided below;

Treatment Group: 24,61,59,46,43,44,52,43,58,67,62,57,71,49,54,43,53,57,49,56,33,74,70

Control Group: 42,33,46,37,43,41,10,42,55,19,17,55,26,54,60,28,62,20,53,48,37,85,42

Design and execute a two sample t test.

```
Treatment_Group <- c(24,61,59,46,43,44,52,43,58,67,62,57,71,49,54,43,53,57,49,56,33,74,70)

Control_Group <- c(42,33,46,37,43,41,10,42,55,19,17,55,26,54,60,28,62,20,53,48,37,85,42)
```

a) State the appropriate null and alternative hypotheses.

Null Hypothesis (H_0): There is no significant difference between the mean scores of the treatment and control groups on the DRP test. (There is no relationship)

Alternative Hypothesis (H_A): There is a significant difference between the mean scores of the treatment and control groups on the DRP test. (There is a relationship)

b) Use and show R code to produce the p value and the confidence interval

```
tout1 <- t.test(Treatment_Group, Control_Group, mu = 0, var.equal = FALSE)
tout1
```

```
##
##  Welch Two Sample t-test
##
## data:  Treatment_Group and Control_Group
## t = 2.6853, df = 39.487, p-value = 0.01054
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   2.900188 20.578072
## sample estimates:
## mean of x mean of y
##  53.26087  41.52174
```

```
tdf1 <- tidy(tout1)
tdf1
```

```
## # A tibble: 1 x 10
##   estim~1 estim~2 estim~3 stati~4 p.value param~5 conf.~6 conf.~7 method alter~8
##     <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl> <chr>  <chr>
## 1    11.7    53.3    41.5    2.69  0.0105    39.5    2.90    20.6 Welch~ two.si~
## # ... with abbreviated variable names 1: estimate, 2: estimate1, 3: estimate2,
## #   4: statistic, 5: parameter, 6: conf.low, 7: conf.high, 8: alternative
```

```
tdf1$p.value
```

```
## [1] 0.01054229
```

```
c(tdf1$conf.low, tdf1$conf.high)
```

```
## [1]  2.900188 20.578072
```

c) Determine if you should reject or fail to reject the null hypothesis using the p value and the confidence interval.

- Looking at the p-value above (0.01054229), we should reject the null hypothesis and accept the alternative that there is a significant difference between the mean scores of the treatment and control groups on the DRP test. There is a relationship between the reading activities and the Degree of Reading Power (DRP) test scores.