STAT 8010 R Lab 10

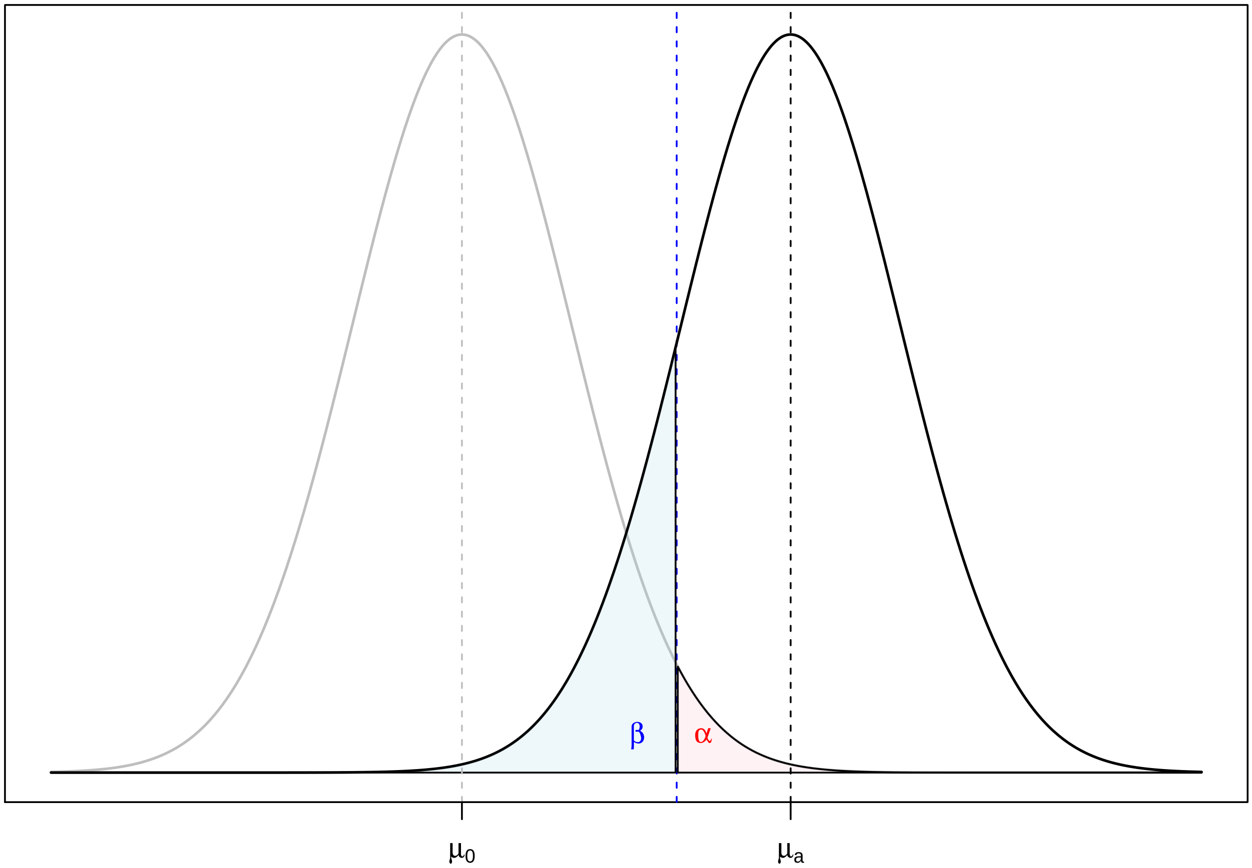
Whitney Huang

October 13, 2020

Contents

Type I and Type II Errors

```r
library(scales)
curve(dnorm(x), from = -3.75, to = 6.75, n = 1001, lwd = 1.5, col = "gray",
      xlab = "",
      ylab = "", las = 1,
      xaxt = "n", yaxt = "n")
abline(v = 1.96, lty = 2, col = "blue")
x_grid <- seq(-3.75, 6.75, 0.01)
y_grid <- dnorm(x_grid)
polygon(c(x_grid[x_grid > 1.96], rev(x_grid[x_grid > 1.96])),
        c(y_grid[x_grid > 1.96], rep(0, length(y_grid[x_grid > 1.96]))), col = alpha("pink", 0.2))
curve(dnorm(x, mean = 3), from = -3.75, to = 6.75, n = 1001, lwd = 1.5, add = T,
      xlab = "",
      ylab = "", las = 1)
x_grid <- seq(-3.75, 6.75, 0.01)
y_grid <- dnorm(x_grid, mean = 3)
polygon(c(x_grid[x_grid < 1.96], rev(x_grid[x_grid < 1.96])),
        c(y_grid[x_grid < 1.96], rep(0, length(y_grid[x_grid < 1.96]))), col = alpha("lightblue", 0.2))
abline(v = 0, lty = 2, col = "gray")
abline(v = 3, lty = 2)
axis(1, at = 0, labels = expression(mu[0]))
axis(1, at = 3, labels = expression(mu[a]))
text(2.2, 0.02, expression(alpha), col = "red")
text(1.6, 0.02, expression(beta), col = "blue")
```



Power analysis

```r
library(asbio)
```

```
## Warning: package 'asbio' was built under R version 3.6.2
```

```
## Loading required package: tcltk
```

```r
power.z.test(sigma = 10, n = 25, power = NULL, alpha = 0.05,
             effect = 4,  test = c("one.tail"))
```

```
## $sigma
## [1] 10
##
## $n
## [1] 25
##
## $power
## [1] 0.63876
##
## $alpha
## [1] 0.05
##
## $effect
## [1] 4
##
## $test
```

```
## [1] "one.tail"
```

```r
power.z.test(sigma = 10, n = NULL, power = 0.8, alpha = 0.05,
             effect = 4,  test = c("one.tail"))
```

```
## $sigma
## [1] 10
##
## $n
## [1] 38.64098
##
## $power
## [1] 0.8
##
## $alpha
## [1] 0.05
##
## $effect
## [1] 4
##
## $test
## [1] "one.tail"
```

```r
power.t.test(n = 25, delta = 4, sd = 10, sig.level = 0.05,
             power = NULL, type = "one.sample", alternative = "one.sided")
```

```
##
##      One-sample t test power calculation
##
##               n = 25
##           delta = 4
##              sd = 10
##       sig.level = 0.05
##           power = 0.617259
##     alternative = one.sided
```

```r
power.t.test(n = NULL, delta = 4, sd = 10, sig.level = 0.05,
             power = 0.8, type = "one.sample", alternative = "one.sided")
```

```
##
##      One-sample t test power calculation
##
##               n = 40.02908
##           delta = 4
##              sd = 10
##       sig.level = 0.05
##           power = 0.8
##     alternative = one.sided
```

Inference for two population means

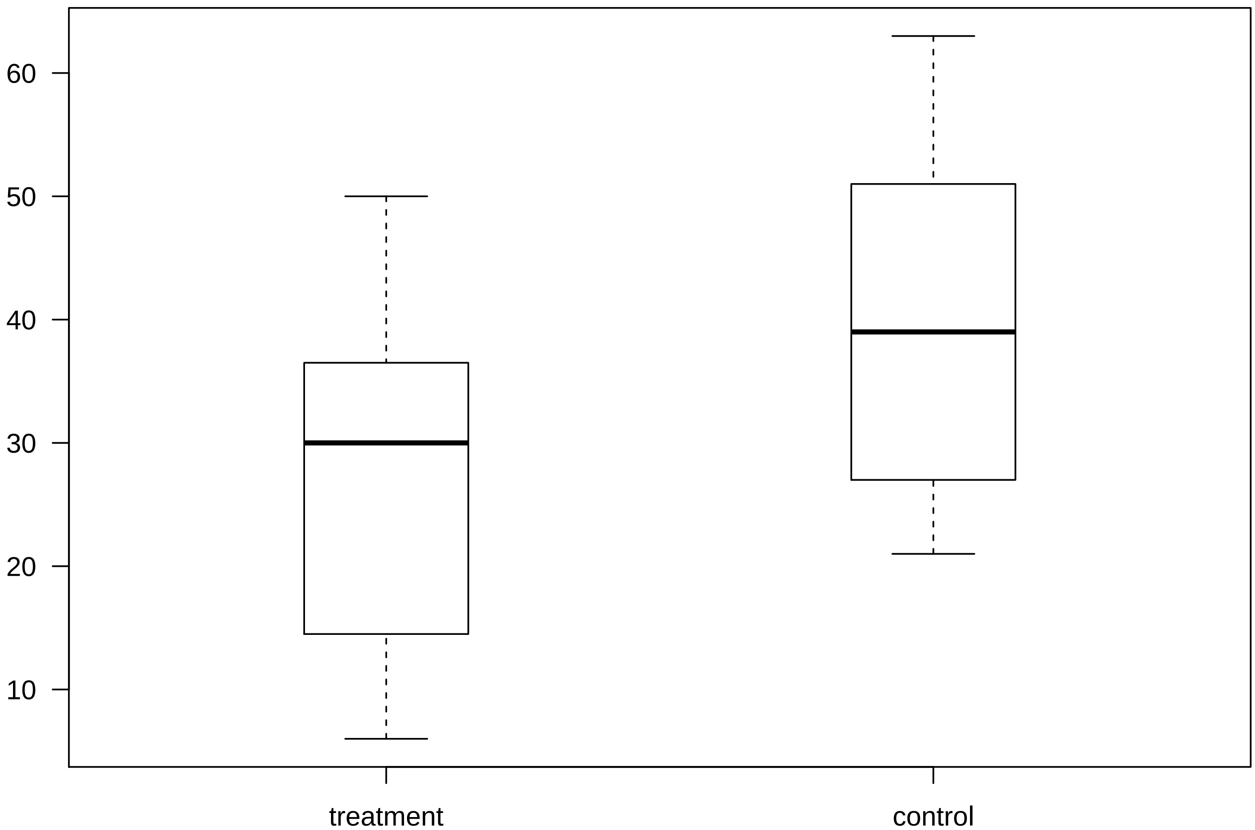
Tapeworm example

```r
treatment <- c(18, 43, 28, 50, 16, 32, 13, 35, 38, 33, 6, 7)
control <- c(40, 54, 26, 63, 21, 37, 39, 23, 48, 58, 28, 39)
dat <- data.frame(cbind(treatment, control))
```

```r
summary(dat)
```

```
##    treatment        control     
##  Min.   : 6.00   Min.   :21.00  
##  1st Qu.:15.25   1st Qu.:27.50  
##  Median :30.00   Median :39.00  
##  Mean   :26.58   Mean   :39.67  
##  3rd Qu.:35.75   3rd Qu.:49.50  
##  Max.   :50.00   Max.   :63.00
```

```r
boxplot(dat, boxwex = 0.3, las = 1)
```




```r
apply(dat, 2, mean)
```

```
## treatment   control 
##  26.58333  39.66667
```

```r
apply(dat, 2, sd)
```

```
## treatment   control 
##  14.36193  13.85859
```

```r
var.test(treatment, control)
```

```
## 
##  F test to compare two variances
## 
## data:  treatment and control
## F = 1.074, num df = 11, denom df = 11, p-value = 0.9079
## alternative hypothesis: true ratio of variances is not equal to 1
```

```
## 95 percent confidence interval:
##  0.3091686 3.7306092
## sample estimates:
## ratio of variances
##           1.073959
```

```r
# Assuming $\sigma_{1}=\sigma_{2}$
t.test(treatment, control, var.equal = T)
```

```
##
##  Two Sample t-test
##
## data:  treatment and control
## t = -2.2709, df = 22, p-value = 0.03329
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -25.031761  -1.134906
## sample estimates:
## mean of x mean of y
##  26.58333  39.66667
```

```r
# Assuming $\sigma_{1}\neq\sigma_{2}$
t.test(treatment, control, var.equal = F)
```

```
##
##  Welch Two Sample t-test
##
## data:  treatment and control
## t = -2.2709, df = 21.972, p-value = 0.03331
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -25.032642  -1.134025
## sample estimates:
## mean of x mean of y
##  26.58333  39.66667
```

```r
# Left-tailed test
t.test(treatment, control, alternative = "less")
```

```
##
##  Welch Two Sample t-test
##
## data:  treatment and control
## t = -2.2709, df = 21.972, p-value = 0.01665
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##       -Inf -3.189613
## sample estimates:
## mean of x mean of y
##  26.58333  39.66667
```

Two sample t test with only sample statistics

```r
t.test.from.summary.data <- function(mean1, sd1, n1, mean2, sd2, n2, ...) {
    data1 <- scale(1:n1)*sd1 + mean1
    data2 <- scale(1:n2)*sd2 + mean2
```

```r
    t.test(data1, data2, ...)
}

t.test.from.summary.data(12.5, 7.63, 10, 27.5, 15.3, 10)
```

```
##
##  Welch Two Sample t-test
##
## data:  data1 and data2
## t = -2.7744, df = 13.216, p-value = 0.01558
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -26.660768  -3.339232
## sample estimates:
## mean of x mean of y
##      12.5      27.5
```

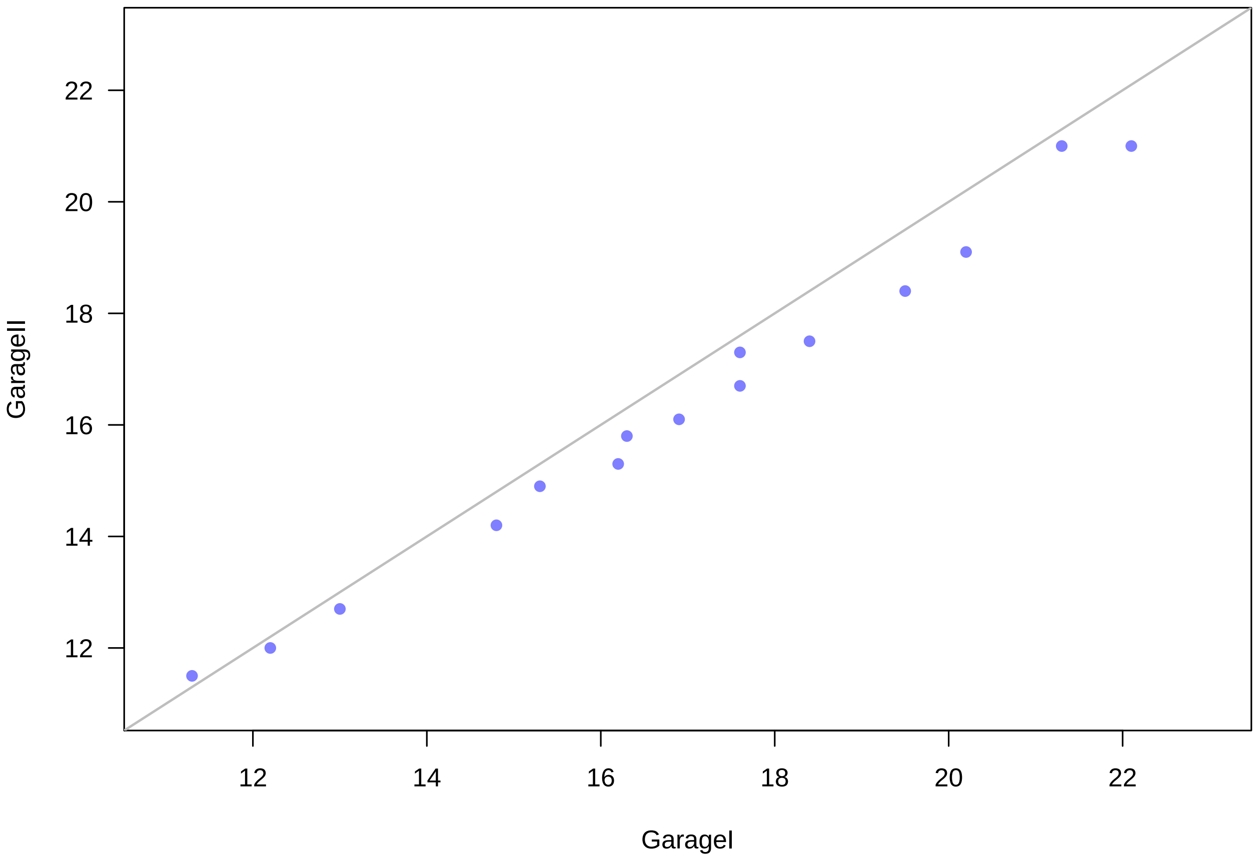
```r
## Check
df = ((4.3^2)/37 + (2.2^2)/31)^2 / (((4.3^2)/37)^2 / 36 +  ((2.2^2)/31)^2 / 30)

se <- sqrt(4.3^2 / 37 + 2.2^2 / 31)

tstat <- (19.45 - 18.2) / se

Pvalue <- 2 * (1 - pt(1.5435, df))
```

Paired T-Test

```r
repair <- c(17.6, 17.3, 20.2, 19.1,
            19.5, 18.4, 11.3, 11.5,
            13.0, 12.7, 16.3, 15.8,
            15.3, 14.9, 16.2, 15.3,
            12.2, 12.0, 14.8, 14.2,
            21.3, 21.0, 22.1, 21.0,
            16.9, 16.1, 17.6, 16.7,
            18.4, 17.5)
GarageI <- repair[seq(1, 29, 2)]
GarageII <- repair[seq(2, 30, 2)]
dat <- cbind(GarageI, GarageII)
apply(dat, 2, mean)
```

```
##  GarageI GarageII
## 16.84667 16.23333
```

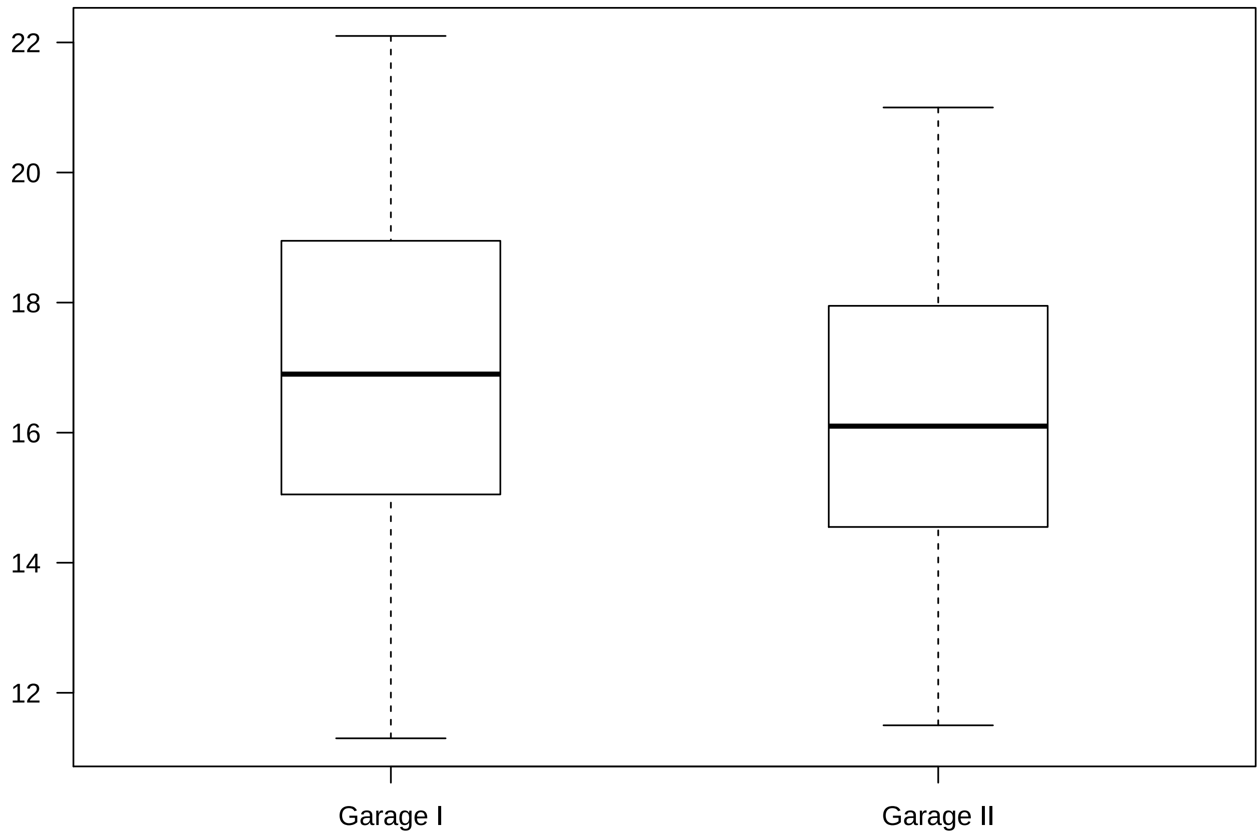
```r
library(scales)
plot(GarageI, GarageII,
     pch = 16, col = alpha("blue", 0.5), las = 1,
     xlim = c(11, 23),
     ylim = c(11, 23))
abline(0, 1, col = "gray", lwd = 1.5)
```




```r
t.test(GarageI, GarageII,
       alternative = c("greater"),
       var.equal = F)
```

```
##
##  Welch Two Sample t-test
##
## data:  GarageI and GarageII
## t = 0.54616, df = 27.797, p-value = 0.2947
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  -1.29749      Inf
## sample estimates:
## mean of x mean of y
##  16.84667  16.23333
```

```r
boxplot(GarageI, GarageII, boxwex = 0.4,
        xaxt = "n", las = 1)
axis(1, at = 1:2, labels = c("Garage I", "Garage II"))
```




```r
boxplot(GarageI - GarageII, boxwex = 0.4,
        xaxt = "n", las = 1)
axis(1, at = 1, labels = "Garage I - Garage II")
abline(h = 0, col = "blue", lty = 2)
```


```
t.test(GarageI, GarageII, alternative = c("greater"), paired = T)
```

```
##
##  Paired t-test
##
## data:  GarageI and GarageII
## t = 6.0234, df = 14, p-value = 1.563e-05
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.4339886       Inf
## sample estimates:
## mean of the differences
##               0.6133333
```