Stat 581, Problem Set #1 Solutions

(1.) $H_0: \mu_1 = \mu_2$ (The drilling method has no effect on drilling time)

$H_A: \mu_1 \neq \mu_2$ (The drilling method does have an effect on drilling time)

(2.)
- The CRD is simple to implement and analyze.
- The CRD accounts for exp. unit variance, and controls for the type I error probability

(3.)
- The CRD does not account for the type II error probability
- The CRD does not adjust for differences in exp units, or for differences in other factors

(4.) dry: $\bar{y}_1 = 878.83$, $s_1 = 89.47$, $s_p = 88.3$

wet: $\bar{y}_2 = 677.17$, $s_2 = 87.19$

(5.) $t_0 = 5.592$, $t_{.025} = 2.07$, p-value = .000

(6.) The experiments finds that wet drilling leads to reduced drilling times.

(7.) see Box Plot

(8.) The experimental finding is based on a comparison of means. It is not true that all dry drilling times exceed all wet drilling times.

```
> library("readxl")
>
> setwd("C:/Users/aneath/iCloudDrive/Lexar/stat581 fall2021")
> hw1.data = read_excel("handout1data.xlsx")
> str(hw1.data)
Classes ‘tbl_df’, ‘tbl’ and 'data.frame':     24 obs. of  10 variables:
 $ machine : num  1 1 1 1 1 1 1 1 1 1 ...
 $ output  : num  16 16 16.1 16.1 16 ...
 $ flow    : num  125 125 125 125 125 125 200 200 200 200 ...
 $ observed: num  2.7 4.6 2.6 3 3.2 3.8 4.6 3.4 2.9 3.5 ...
 $ 95C     : num  11.18 7.09 8.1 11.74 11.29 ...
 $ 100C    : num  5.26 6.75 7.46 7.01 8.13 ...
 $ modified: num  16.9 16.4 17.2 16.4 16.5 ...
 $ unmod   : num  16.6 16.8 17.4 17.1 17 ...
 $ time    : num  727 965 904 987 847 918 814 750 804 989 ...
 $ method  : chr  "d" "d" "d" "d" ...
>
> time = na.omit(hw1.data$time)
> method = as.factor(na.omit(hw1.data$method))
>
>
> two.sample.test = function(y1,y2,alpha=.05)
+ {
+   n1 = length(y1)
+   n2 = length(y2)
+   ybar1 = mean(y1)
+   ybar2 = mean(y2)
+   s1 = sd(y1)
+   s2 = sd(y2)
+   ybar.diff = ybar1-ybar2
+   s.p = sqrt( ((n1-1)*s1^2+(n2-1)*s2^2) / (n1+n2-2) )
+   SE = s.p*sqrt(1/n1 + 1/n2)
+
+   t.cr = qt(alpha/2,lower.tail = FALSE, df=n1+n2-2)
+   t.0 = ybar.diff / SE
+   p.value = 2*pt(abs(t.0),df=n1+n2-2,lower.tail = FALSE)
+
+   table1 = matrix(c(ybar1,ybar2,s1,s2,s.p),nrow = 1)
+   dimnames(table1) = list(c(""),c("ybar1","ybar2","s1","s2","Sp"))
+   print(table1)
+
+   table2 = matrix(c(t.0,t.cr,p.value),nrow = 1)
+   dimnames(table2) = list(c(""),c("test statistic","critical point","p-value"))
+   print(table2)
+
+ }
>
> two.sample.test(time[method=="d"],time[method=="w"])

    ybar1    ybar2       s1       s2       Sp
 878.8333 677.1667 89.46999 87.18928 88.33699
 test statistic critical point      p-value
          5.592       2.073873 1.272908e-05
>
```

```
> t.test(time~method,var.equal=TRUE)

        Two Sample t-test

data:  time by method
t = 5.592, df = 22, p-value = 1.273e-05
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 126.8757 276.4576
sample estimates:
mean in group d mean in group w
       878.8333        677.1667

>
> boxplot(time[method=="d"],time[method=="w"],names = c("dry drilling","wet drilling"),
```