Stat 581 - Problem Set #3a Solutions

(1.) (a) CI for $\delta = [127, 276]$, $\delta = \mu_1 - \mu_2$, $\hat{\delta} = 202$

Based on the observed data, we estimate that the difference in mean drilling time, (dry - wet) is between 127 and 276 $\frac{1}{100}$ minutes

(b) A hypothesis test looks to determine if an effect exists. A confidence interval looks to determine the size of the effect. (Also, a CI provides a measure of evidence strength)

(2.) (a) See output for confidence interval display

(b) If 0 is contained in the CI for $\mu_i - \mu_j$, the decision is in favor of $H_0^{(ij)}: \mu_i = \mu_j$.

If 0 is not contained in the CI for $\mu_i - \mu_j$, the decision is in favor of $H_A^{(ij)}: \mu_i \neq \mu_j$.

(c) When multiple intervals are computed, then

$$P(\text{at least 1 type I error}) > \alpha,$$

where $P(\text{type I error } (i,j)) = \alpha$.

```
> library("readxl")
> library("multcomp")
>
> setwd("C:/Users/aneath/iCloudDrive/Lexar/stat581 fall2021")
> hw3a.data = read_excel("handout1data.xlsx")
> str(hw3a.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(hw3a.data$time)
> method = as.factor(na.omit(hw3a.data$method))
>
> two.sample.interval = 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.mult = qt(alpha/2,lower.tail = FALSE, df=n1+n2-2)
+   lower.est = ybar.diff - t.mult*SE
+   upper.est = ybar.diff + t.mult*SE
+
+   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(ybar.diff,lower.est,upper.est),nrow = 1)
+   dimnames(table2) = list(c(""),c("estimated difference","lower limit","upp
er limit"))
+   print(table2,digits = 3)
+
+ }
> two.sample.interval(time[method=="d"],time[method=="w"])
    ybar1    ybar2       s1       s2       Sp
 878.8333 677.1667 89.46999 87.18928 88.33699
 estimated difference lower limit upper limit
                  202         127         276
>
> t.test(time~method,var.equal=TRUE)

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
```

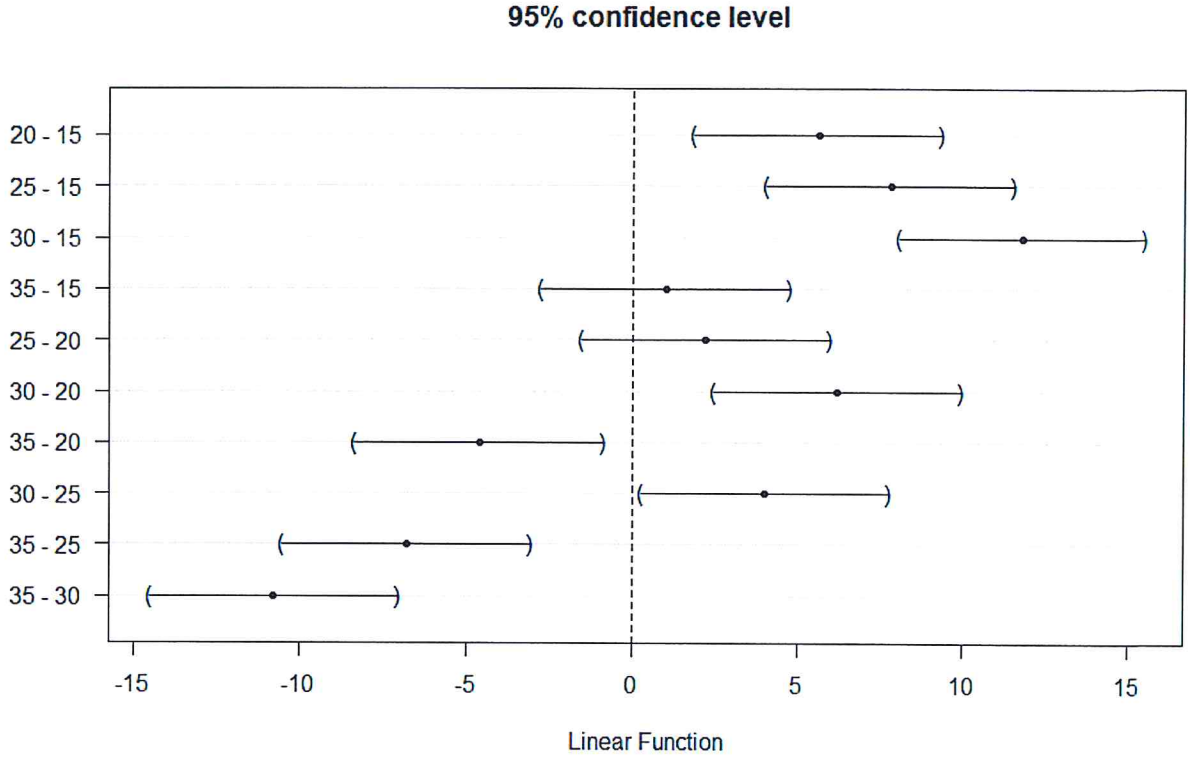
```
> hw3b.data = read_excel("handout2data.xlsx")
> str(hw3b.data)
Classes ‘tbl_df’, ‘tbl’ and 'data.frame':    25 obs. of  11 variables:
 $ strength: num  7 7 15 11 9 12 17 12 18 18 ...
 $ percent : num  15 15 15 15 15 20 20 20 20 20 ...
 $ 20g     : num  24 28 37 30 NA NA NA NA NA NA ...
 $ 30g     : num  37 44 31 35 NA NA NA NA NA NA ...
 $ 40g     : num  42 47 52 38 NA NA NA NA NA NA ...
 $ life    : num  17.6 18.9 16.3 17.4 20.1 21.6 16.9 15.3 18.6 17.1 ...
 $ fluid   : num  1 1 1 1 1 1 2 2 2 2 ...
 $ rate    : num  575 542 530 539 570 565 593 590 579 610 ...
 $ rf power: num  160 160 160 160 160 180 180 180 180 180 ...
 $ brand   : chr  "acme" "acme" "acme" "acme" ...
 $ wear    : num  2.1 2.4 2.5 2.3 2.2 2 1.9 2.1 2.2 2.4 ...
>
> percent = as.factor(na.omit(hw3b.data$percent))
> strength = na.omit(hw3b.data$strength)
>
> hw3.model = aov(strength~percent)
>
> comps = glht(hw3.model,linfct = mcp(percent="Tukey"))
> ci_lsd = confint(comps,calpha = univariate_calpha())
>
> plot(ci_lsd)
> ci_lsd

Quantile = 2.086
95% confidence level


Linear Hypotheses:
               Estimate lwr      upr
20 - 15 == 0   5.6000   1.8545   9.3455
25 - 15 == 0   7.8000   4.0545  11.5455
30 - 15 == 0  11.8000   8.0545  15.5455
35 - 15 == 0   1.0000  -2.7455   4.7455
25 - 20 == 0   2.2000  -1.5455   5.9455
30 - 20 == 0   6.2000   2.4545   9.9455
35 - 20 == 0  -4.6000  -8.3455  -0.8545
30 - 25 == 0   4.0000   0.2545   7.7455
35 - 25 == 0  -6.8000 -10.5455  -3.0545
35 - 30 == 0 -10.8000 -14.5455  -7.0545

> summary(ci_lsd)


Linear Hypotheses:
               Estimate Std. Error t value Pr(>|t|)
20 - 15 == 0    5.600      1.796   3.119  0.03851 *
25 - 15 == 0    7.800      1.796   4.344  0.00261 **
30 - 15 == 0   11.800      1.796   6.572  < 0.001 ***
35 - 15 == 0    1.000      1.796   0.557  0.97977
25 - 20 == 0    2.200      1.796   1.225  0.73727
30 - 20 == 0    6.200      1.796   3.453  0.01885 *
35 - 20 == 0   -4.600      1.796  -2.562  0.11631
30 - 25 == 0    4.000      1.796   2.228  0.21016
35 - 25 == 0   -6.800      1.796  -3.787  0.00901 **
35 - 30 == 0  -10.800      1.796  -6.015  < 0.001 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
(Adjusted p values reported -- single-step method)
```


95% confidence level
20 - 15
25 - 15
30 - 15
35 - 15
25 - 20
30 - 20
35 - 20
30 - 25
35 - 25
35 - 30
-15
-10
-5
0
5
10
15
Linear Function