Colm Mooney - 20325583

#Question 2

lol yes, good job

#Rule to go by: If p is low, Reject H0. = null hypothesis

#Two independent samples

tA <- c(-0.15, 8.6, 5, 3.71, 4.29, 7.74, 2.48, 3.25, -1.15, 8.38)

tB <- c(2.55, 12.07, 0.46, 0.35, 2.69, -0.94, 1.73, 0.73, -0.35, -0.37)

par(mfrow = c(2,1))

hist(tA) - show 1/2

hist(tB)

par(mfrow = c(1,1))

median(tA) #4

median(tB) #0.595

#Wilcoxon-Mann Whitney Test

#####WILCOXON-MANN-WHITNEY TEST (TWO-SIDED)

?wilcox.test

wilcox.test(tA,tB, alternative = "two.sided", conf.int = TRUE, conf.level = 0.95) #(,exact = FALSE, correct = FALSE) #p = 0.08921 Concludes the medians of t1 and t2 differ.

#The true location shift is not equal to 0, Both samples are not of similar shapes.

#Wilcoxon-Mann Whitney Test is inappropriate

#Siegel Tukey Test

library(DescTools)

?SiegelTukeyTest

SiegelTukeyTest(tA,tB, conf.int = TRUE, conf.level = 0.95)

#Siegel Tukey test is best used when medians are the same. That is not the case with these samples.

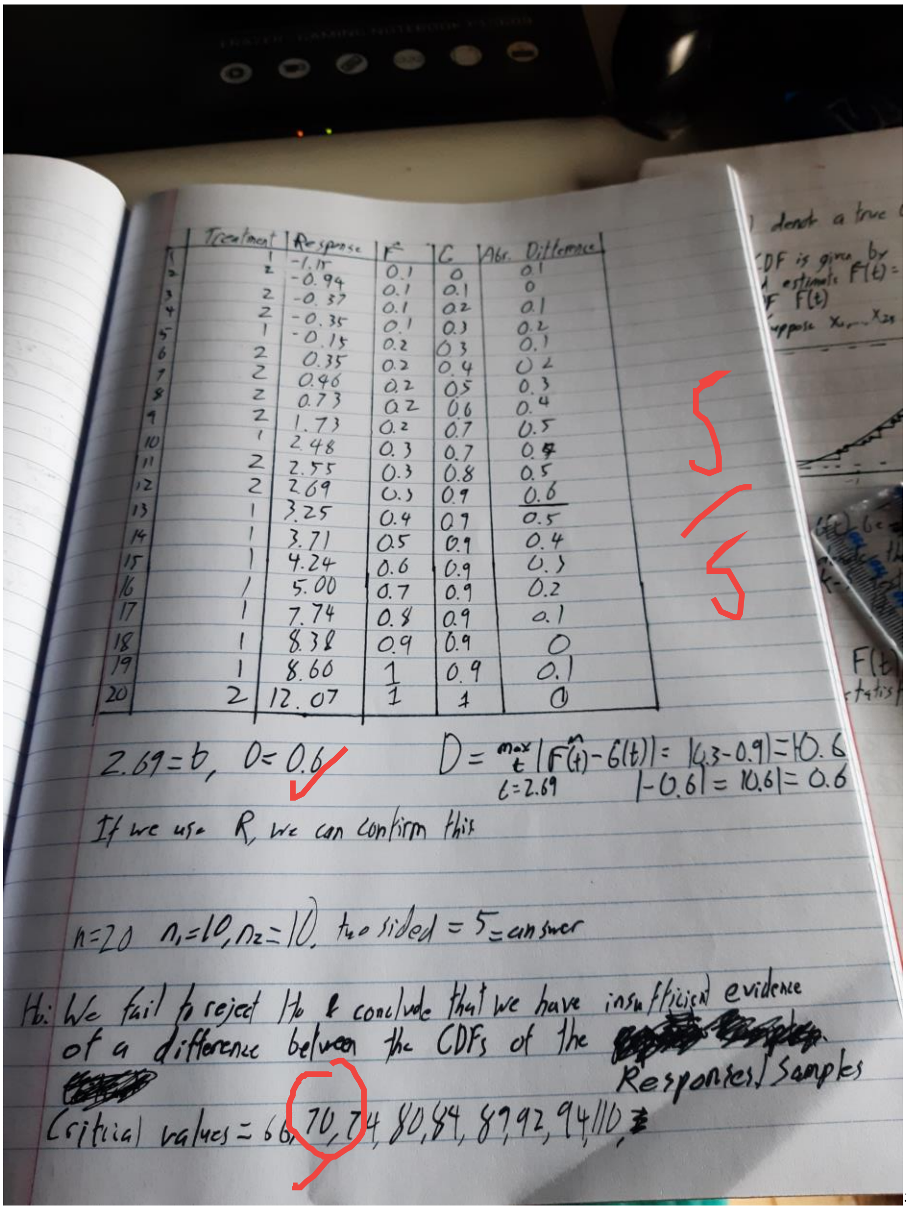
#This means the Siegel Tukey test, while more appropriate than Wilcoxon-Mann Whitney test, is not as appropriate as the Kolmogorov-Smirnov test.

```
#KS doesn't make any assumptons regarding the shape of the underlying distributions, so is the more appropriate test here.

ks.test(x = tA, y = tB, alternative = "two.sided") #p = 0.05245

cdfA <- ecdf(tA)

cdfB <- ecdf(tB)

minMax <- seq(min(tA, tB), max(tA, tB), length.out=length(tA))

x0 <- minMax[which(abs(cdfA(minMax) - cdfB(minMax)) == max(abs(cdfA(minMax) - cdfB(minMax))))]

y0 <- cdfA(x0)

y1 <- cdfB(x0)

y0

y1

plot(cdfA, verticals=TRUE, col.points="blue", col="blue", main="Liquid Beverage Containers ECDFs", ylab="CDF", xlab="t")

lines(ecdf(tB), verticals=TRUE, col.points="green", col="green")

points(c(x0, x0), c(y0, y1), pch=16, col="red")

segments(x0, y0, x0, y1, col="red", lty="dotted")

#Both samples are not a similar shape & have different medians.

#Only slightly so, but we fail to reject the null hypothesis.

#This test is the most appropriate.

#Therefore, we conclude The Kolmogorov-Smirnov is the most appropriate test to do here.
```

Ho + Ha ?

Conclusion?

0/3



	Treatment	Response	$\hat{F}$	$\hat{G}$	Abs. Difference
1	1	-1.15	0.1	0	0.1
2	2	-0.94	0.1	0.1	0
3	2	-0.37	0.1	0.2	0.1
4	2	-0.35	0.1	0.3	0.2
5	1	-0.15	0.2	0.3	0.1
6	2	0.35	0.2	0.4	0.2
7	2	0.46	0.2	0.5	0.3
8	2	0.73	0.2	0.6	0.4
9	2	1.73	0.2	0.7	0.5
10	1	2.48	0.3	0.7	0.4
11	2	2.55	0.3	0.8	0.5
12	2	2.69	0.3	0.9	0.6
13	1	3.25	0.4	0.9	0.5
14	1	3.71	0.5	0.9	0.4
15	1	4.24	0.6	0.9	0.3
16	1	5.00	0.7	0.9	0.2
17	1	7.74	0.8	0.9	0.1
18	1	8.38	0.9	0.9	0
19	1	8.60	1	0.9	0.1
20	2	12.07	1	1	0

$2.69 = t,\ D = 0.6$

$D = \max_{t} |\hat{F}(t) - \hat{G}(t)| = |0.3 - 0.9| = |-0.6| = |0.6| = 0.6$, $t = 2.69$

If we use R, we can confirm this

$n = 20$ $n_1 = 10, n_2 = 10$, two sided $= 5$ = answer

H_0: We fail to reject H_0 & conclude that we have insufficient evidence of a difference between the CDFs of the Responses/Samples

Critical values $= 66, 70, 74, 80, 84, 89, 92, 94, 110$

#Q3

```
tN <- c(74,79,92,65,253)

tS <- c(159,72,248,129,120)


median(tN)

median(tS)


hist(tN)

hist(tS)


#Wilcoxon test is not appropriate
```

```
#Different shape & Median

#Siegel-Tukey test not appropriate

#Medians are not the same


#KS test is best, different shapes and medians are also different.

#KS doesn't make any assumptons regarding the shape of the underlying distributions,

#so is the more appropriate test here.



ks.test(Number, Symbol, alternative = "two.sided") #p-value = 0.3571


cdfN <- ecdf(tN)

cdfS <- ecdf(tS)

minMax <- seq(min(tN, tS), max(tN, tS), length.out=length(tN))

x0 <- minMax[which(abs(cdfN(minMax) - cdfS(minMax)) == max(abs(cdfN(minMax) - cdfS(minMax))))]

y0 <- cdfN(x0)

y1 <- cdfS(x0)


plot(cdfN, verticals=TRUE, col.points="blue", col="blue", main="Liquid Beverage Containers ECDFs", ylab="CDF", xlab="t")

lines(ecdf(tS), verticals=TRUE, col.points="green", col="green")

points(c(x0, x0), c(y0, y1), pch=16, col="red")

segments(x0, y0, x0, y1, col="red", lty="dotted")
```