Question 1)

Book Problem 1.3

Recognition and the statement of the problem: We want to compare the growth of a flower and determine the effects of certain conditions. These conditions are sunlight, water, fertilizer, and soil. The overall problem is to see if flowers can grow quicker under a certain condition. The next step is the selection of the response variable. In my opinion the most effective response variable would be the height between all of the flowers to see a difference between all of them. You could make an average height and have a variability between heights to make this experiment more useful. In order to control an experiment, you need to have factors involved. There should be control factors that way we can differentiate between the data points. I would say making a held-constant factor would help that way we can see what condition is the most effective, if that be water, sunlight, fertilizer, or soil.

Question 2)

In an observational study we find if something is correlated between a certain treatment group or not. In an observational study correlation does not imply causation. We cannot conclude that because there might be other treatment factors affecting the study. For example, if we want to say tanning beds cause cancer, we cannot say that we can conclude because what if some people have other factors like, some people are outside more often, or maybe if someone uses tanning products as well. These additional treatment factors cause standard errors. The data may be inconsistent as well.

Question 3)

Book problem 1.9

Randomization if a very important process to do in an experiment because you eliminate biasness. In an experiment this makes all possible treatments to experiment units all equally likely to appear in the experiment. This will cause the same probability for all outcomes.

Question 4)

randomvalues = runif(12)

> randomvalues

[1] 0.36873331 0.57263878 0.37880448 0.65767323 0.72201315 0.01712232 0.90862773 0.19878683 0.43592257 0.81776413

[11] 0.81873946 0.65881120

names = c("A","B","C","D","E","F","G","H","I","J","K","L")

> names

[1] "A" "B" "C" "D" "E" "F" "G" "H" "I" "J" "K" "L"

RNs = runif(12)

> RNs

[1] 0.9718382 0.3032138 0.6793978 0.6151122 0.6267247 0.9677971 0.2671738 0.1596083 0.1365499 0.6218375 0.2446019

[12] 0.4811720

names[order(RNs)]

[1] "I" "H" "K" "G" "B" "L" "D" "J" "E" "C" "F" "A"

> sample(names)

[1] "K" "D" "B" "F" "I" "H" "C" "L" "E" "G" "A" "J"

my.plan = data.frame(treatment = rep(c("trt1","trt2", "trt3"), 4), subj = sample(names))

> my.plan

treatment subj

1 trt1 G

2 trt2 L

3 trt3 K

4 trt1 A

5 trt2 J

6 trt3 B

7 trt1 E

8 trt2 F

9 trt3 H

10 trt1 C

11 trt2 D

12 trt3 I

my.plan = my.plan[sample(1:12), ]

> my.plan = cbind(run=1:12, my.plan)

> my.plan

run treatment subj

1 1 trt1 G

4 2 trt1 A

8 3 trt2 F

11 4 trt2 D

12 5 trt3 I

9 6 trt3 H

10 7 trt1 C

3 8 trt3 K

7 9 trt1 E

6 10 trt3 B

2 11 trt2 L

5 12 trt2 J

Question 5)

The three basic principles of experimental design are, randomization, replication, and blocking.

Question 6)

A) We fail to reject that the variances are equal based on the p-value of 0.9744. We can see that the variances are equal based in this situation because we fail to reject the null hypothesis.

F test to compare two variances

data: prob0228$Type1 and prob0228$Type2

F = 0.97822, num df = 9, denom df = 9, p-value = 0.9744

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

0.2429752 3.9382952

sample estimates:

ratio of variances

0.9782168

B) Yes, from part A in question six we can assume that we have equal variances and use a pooled t-test. 95% Confidence interval is (-8.552441, 8.952441).

Two Sample t-test

data: prob0228$Type1 and prob0228$Type2

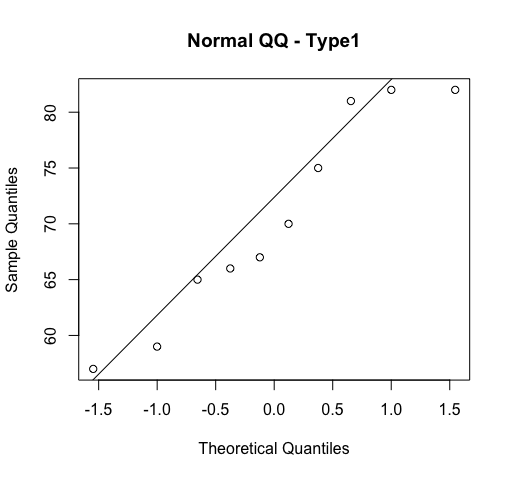
t = 0.048008, df = 18, p-value = 0.9622

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

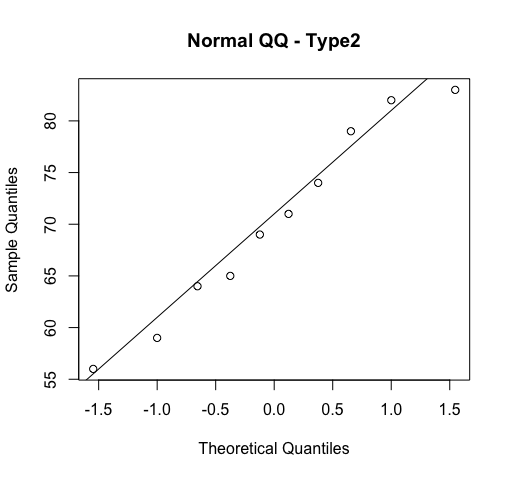
-8.552441 8.952441

sample estimates:

mean of x mean of y

70.4 70.2

C)



Yes, we can assume normality based on these two plots due to the fact that the line on both of the Normal QQ plot is straight and this implies normality. We can also assume equal variance based off of the slope and how the ratio is close to 1 which implies equal variance.

D)

Welch Two Sample t-test

data: prob0228$Type1 and prob0228$Type2

t = 0.048008, df = 17.998, p-value = 0.9622

alternative hypothesis: true difference in means is not equal to 0

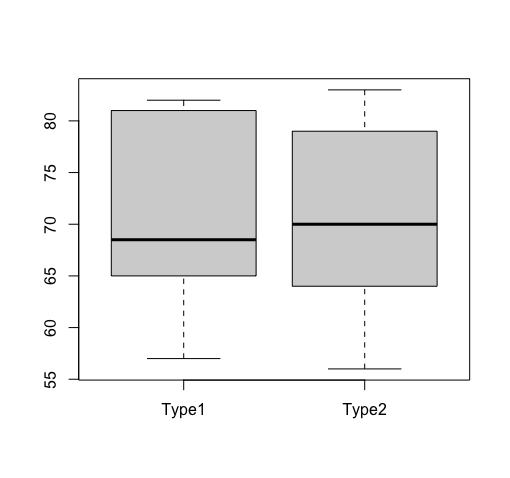
95 percent confidence interval:

-8.552517 8.952517

sample estimates:

mean of x mean of y

70.4 70.2

 E)

We can see from the boxplots that their medians are very similar to each other. Along with the variability being relatively equal as well. In conclusion we can see that the variances are the same and this is why we picked the pooled t-test so we could assume independence, normality, and equal variance. With our significance level alpha being at 0.05 and our p-values from the pooled t-test we fail to reject the null hypothesis, the data suggest that type 1 and type 2 do not have enough significance between one another.

CODE

randomvalues = runif(12)

randomvalues

sort(randomvalues)

names = c("A","B","C","D","E","F","G","H","I","J","K","L")

names

RNs = runif(12)

RNs

names[order(RNs)]

sample(names)

my.plan = data.frame(treatment = rep(c("trt1","trt2", "trt3"), 4), subj = sample(names))

my.plan

my.plan = my.plan[sample(1:12), ]

my.plan = cbind(run=1:12, my.plan)

my.plan

prob0228 <- read.table("https://www.stat.uiowa.edu/~ernli/DOEdata/problem0228.txt",

header = TRUE)

prob0228

var.test(prob0228$Type1, prob0228$Type2, ratio = 1)

t.test(prob0228$Type1, prob0228$Type2, var.equal = TRUE)

qqnorm(prob0228[ ,"Type1"], main = "Normal QQ - Type1");

qqline(prob0228[ ,"Type1"]);

qqnorm(prob0228[ , "Type2"], main = "Normal QQ - Type2");

qqline(prob0228[ , "Type2"]);

t.test(prob0228$Type1, prob0228$Type2, var.equal = FALSE)

boxplot(prob0228)