Question 1: Collaborating with Two-Way ANOVA (60 points)

Question 2: Built Two (https://en.wikipedia.org/wiki/Built_to_Spill) -Way ANOVA (40 points)

Code

36-309 / 36-749 Homework 5: Two-Way ANOVA

Due Wednesday, October 12, 11:59pm

Sanaz Saadatifar

Question 1: Collaborating with Two-Way ANOVA (60 points)

The dataset in this problem is based on the paper “Allocator-Recipient Similarity and the Equitable Division of Rewards” (https://www.jstor.org/stable/3033586?casa_token=HkWmc_dJ8OAAAAAA%3ANptMttbleK09oA-G2XuLtQ-UpfRsfVjhY-zWerqibGUuWcRwsOsmFHREAYP76sXDQcBp5czuiYhf2_l9nDa-VmGzjVjt9Cc37pMY9JavC2bBuJi03wBM&seq=1#metadata_info_tab_contents) by Jerald Greenberg (1978). You do not need to read the paper.

This paper was an early study on collaboration and “reward sharing” behavior, which has become a classical topic in psychology. In this study, 90 undergraduates (45 were male, and 45 were female) were first given an attitude survey to measure their views on controversial issues. Then, each subject performed a 5-minute “digit coding” task. Before the task began, each subject was (falsely) told that they are collaborating with an unknown student and that they should be as accurate as possible, because they and their collaborator would be rewarded for completing the task accurately. One week later, each subject was told that they performed “better than average” and that their collaborator performed “worse than average.” Two thirds of the subjects were told that they might like to know more about their collaborator, so they were shown their own attitude survey results as well as (fake) survey results from their collaborator. The fake results were arranged to randomly have either “similar” or “dissimilar” answers in reference to whatever the subject wrote on their own survey. A third group, called “unknown,” was a control group who were not shown any survey results about their fictional collaborator.

At the end of the experiment, each subject was asked how much of the reward they think they earned, with the remainder of the reward going to the collaborator. The response was recorded as 5% increments, i.e., 0%, 5%, 10%, …, 100%. (0% corresponds to, “I didn’t earn any of the reward, and my collaborator should get all of it,” and 100% corresponds to, “I earned all of the reward, and my collaborator shouldn’t get any of it.”)

Here is the dataset from this study:

Hide

```
shareAllocation = read.csv("https://raw.githubusercontent.com/zjbranson/stat309fall2022/ma
in/shareAllocation.csv")
#make sure the categorical variables are factors
shareAllocation$gender = factor(shareAllocation$gender)
shareAllocation$treatment = factor(shareAllocation$treatment)
```

There are three variables in this dataset:

- `gender` : The gender of each subject (male or female). (The paper was from 1978, so they took a binary view of gender at the time.)
- `treatment` : Either “similar,” “dissimilar,” or “unknown,” representing the three treatment groups subjects were randomly assigned.
- `allocPct` : The share allocation percentage (this is the outcome variable).

a. (20pts) Because this is a randomized experiment involving a quantitative outcome and two categorical explanatory variables, bells should be ringing that we’ll use two-way ANOVA to analyze this experiment. However, before we dive immediately into statistical modeling, let’s first do some preliminary exploration of this dataset. For this part, answer the following questions.

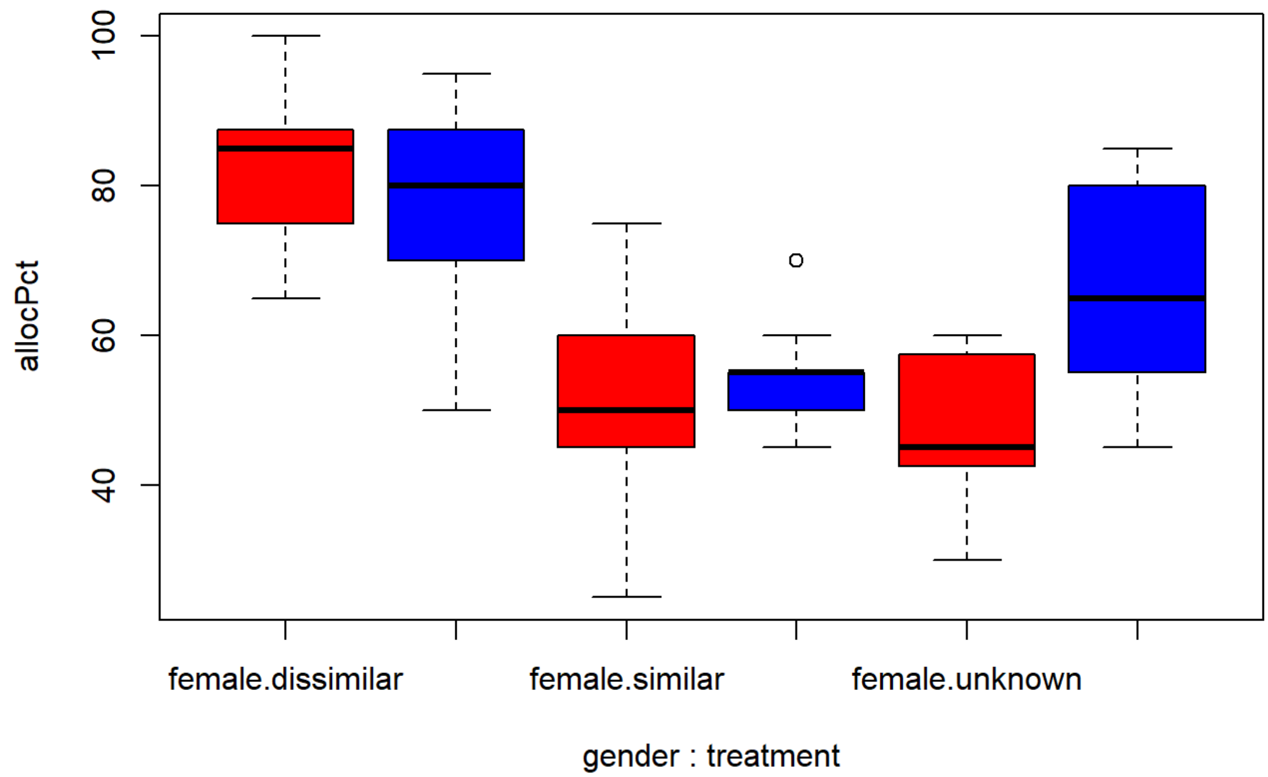
- (4pts) First, consider an alternate experiment, where instead each subject had been tested in **all three** treatment conditions (“similar”, “dissimilar”, and “unknown”). Thus, the resulting dataset would have three rows for each subject (corresponding to their outcomes under the three different treatment conditions). In that case, we’d still have a quantitative outcome and two categorical explanatory variables. However, in thi scase, which model assumption violation would prevent us from validly using two-way ANOVA? Explain your answer in 1-3 sentences. (**Hint**: The assumptions behind two-way ANOVA are exactly the same as those behind one-way ANOVA.)

[The independency of each data could be violate because all of the results for these three categories could be dependent to each other because they all are results of one single subject, so they can be related.]

- (6pts) Now create a side-by-side boxplot that let’s you evaluate the *equal variance assumption*. Does that assumption seem plausible for this dataset? Explain your answer in 1-3 sentences. (**Regardless of what you conclude in this question, please disregard this part when interpreting the upcoming two-way ANOVA analysis.**)

Hide

```
boxplot(allocPct ~ gender + treatment, data = shareAllocation,
col = c("red", "blue", "red", "blue", "red", "blue"))
```

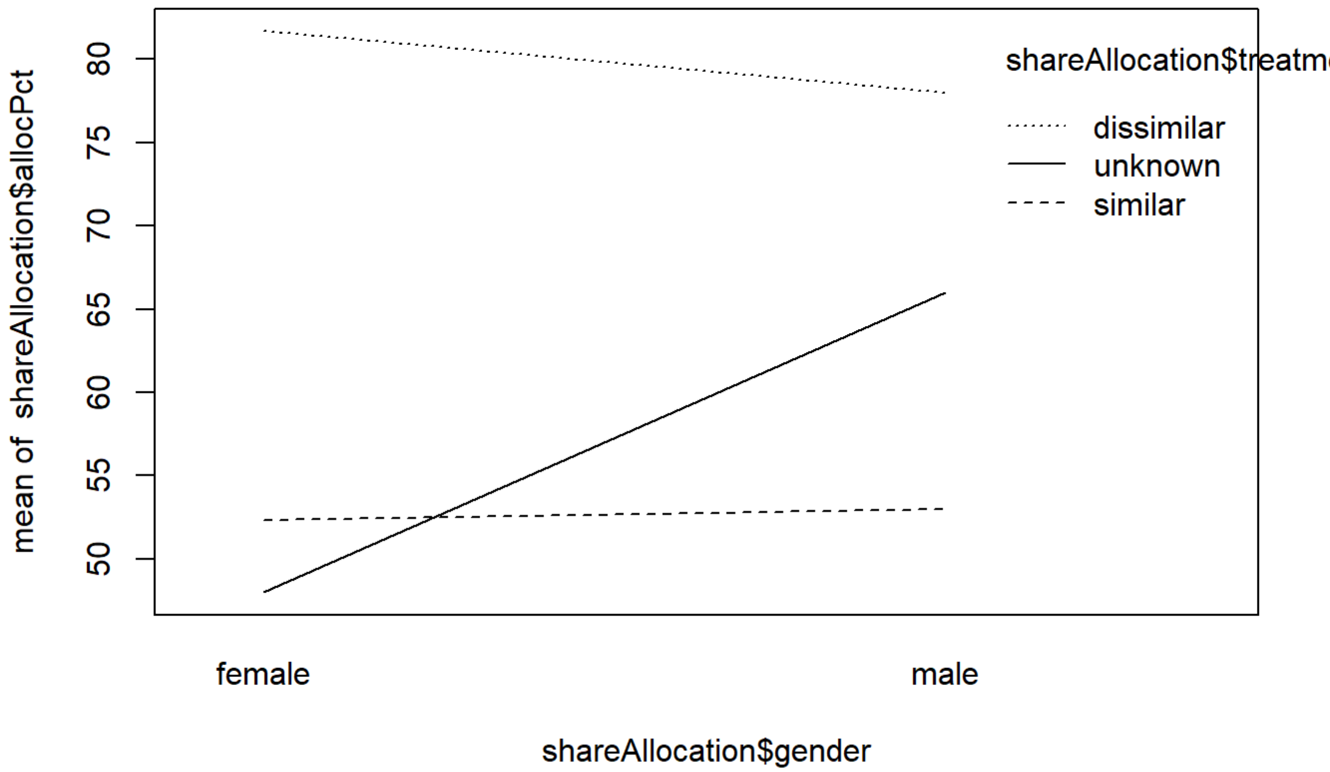


[Equal variance assumption does not seem very plausible because we see some of these 6 groups have significant differences in terms of variance (more than twice of IQR especially male.similar vs male and female dissimilar groups)]

- (6pts) Now turn in an **interaction plot** of these data. (Your interaction plot *does not* need to display confidence intervals.) Interpret the interaction plot in 1-3 sentences. In your interpretation, be sure to discuss the effects of each factor (`treatment` and `gender`), as well as the possibility of an interaction between the two factors.

Hide

```
interaction.plot(x.factor = shareAllocation$gender, trace.factor = shareAllocation$treatme
nt, response = shareAllocation$allocPct)
```



[The plot shows that share allocation percentage for unknown groups is much higher in male compared to female. Also the plot does not show a significant difference between malae and female of similar group for share allocation percentages. Moreover, in terms of dissimilar group, the plot shows that share allocation percentages of males is less than female.
Also this plot shows that share allocation percentage is generally higher for dissimilar group compared to other two groups, and similar group is higher than unknown group for females, and lower for males. In terms of interactions, of course formal statistical analysis would give a better and more reliable answer, but this plot shows that unknown group's line intersects with similar group which seems to be an interaction. Also unknown and dissimilar groups' lines are not parallel which might be an interaction as well. For similar and dissimilar they both seems slightly parallel, hence there does not seem to be an interaction between these two]

- (4pts) At this point, you've made a side-by-side boxplot and an interaction plot. Consider the information displayed in each of these plots. In what ways do these two plots display similar information, and in what ways do they display different information? Discuss in 1-3 sentences. (**Hint**: There is indeed at least one way these two plots are similar, and at least one way they are different. Thus, you should discuss at least one similarity and one difference.)

[Similarity: both two plots show that dissimilar group (female and male), has higher mean of share allocation percentage compared to other groups. Dissimilarity: the interaction plot shows that there is no big difference between the male vs female of similar group in terms of mean share allocation percentages. However, the box plot shows that the variance of female.similar group is higher than male.similar group. And it is possible to see a major difference between female.similar and male.similar.]

b. (8pts) Now perform an *interactive* two-way ANOVA model (being sure to include your code and ANOVA output). Using this two-way ANOVA model, what is your scientific conclusion for this study? In particular, what can we conclude about the effects of gender and treatment on share allocation?

Hide

```
summary(aov(allocPct ~ gender * treatment, data = shareAllocation))
```

```
##                  Df Sum Sq Mean Sq F value   Pr(>F)    
## gender            1    563     563   4.312 0.040891 *  
## treatment         2  12782    6391  48.996  7.9e-15 ***
## gender:treatment  2   1972     986   7.558 0.000959 ***
## Residuals        84  10957     130                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

[first, we reject the null hypothesis that the mean allocPct is equal across the two gender groups, because the p-value is below 0.05 (0.0408). Furthermore, we reject the null hypothesis that the mean allocPct is equal across the three treatment groups, because the p-value is below 0.05 (7.9e-15). Also in terms of the interaction between gender and treatment, The null hypothesis is that there is not an interaction between treatment and gender group. We can also reject the null hypothesis, meaning that there are interactions, one factor's effects depends on the values of other factors.]

c. (10pts) For each of the 6 combinations of gender and treatment, there is a true population mean for share allocation. We can estimate each of these 6 true population means using the sample mean outcome in each of the 6 groups. In turn, we could also compute a confidence interval for each of these 6 sample means. I won't ask you to compute the confidence intervals here, but note that each confidence interval will be a function of three things: (1) the sample mean outcome in each group, (2) the number of subjects assigned to that group, and (3) our estimate of sigma-squared.

Thus, for this question, answer the following three questions.

- (4pts) Write code that displays the sample mean `allocPct` for each of the 6 combinations of `gender` and `treatment` . For the sake of this question, there's no need to restate each mean - just write the code that displays the sample means. (**Hint**: You can do this in one line of code using the `aggregate()` function, where you include `gender` *and* `treatment` on the right-hand side of the tilde within `aggregate()` .)

Hide

```
aggregate(allocPct ~ treatment + gender, data = shareAllocation, FUN = mean)
```

```
##    treatment gender allocPct
## 1 dissimilar female 81.66667
## 2    similar female 52.33333
## 3    unknown female 48.00000
## 4 dissimilar   male 78.00000
## 5    similar   male 53.00000
## 6    unknown   male 66.00000
```

- (3pts) What is the number of subjects in each of the 6 groups? Write one line of code that answers this question, and then state what the number of subjects is for each group outside of your code.

Hide

```
table(shareAllocation$treatment, shareAllocation$gender)
```

```
##
##              female male
##   dissimilar     15   15
##   similar        15   15
##   unknown        15   15
```

[The number of subjects is 15 for each group]

- (3pts) What is our estimate of sigma-squared for the two-way ANOVA model you ran in Part B? Explain your answer in 1-2 sentences.

[Sigma square is the mean sq of residuals (or in other words within-MS) which equals to 130.]

d. (8pts) For this part, we will more closely examine your ANOVA table from Part B. Please do the following:

- (4pts) In the ANOVA table, you should see that there is a p-value for `gender`, which can be used to determine if we should reject or fail to reject a null hypothesis. What is that null hypothesis? State your answer using mathematical notation, and be sure to define what your notation means. **[H0 :μ(allocPct of female) = μ(allocPct of male) Null hypothesis indicates that mean share allocation percentage of male group equals to female group's (there is no difference between male and female share allocation percentages)]**
- (4pts) In addition, compute the sample mean outcome for females and for males. After writing code that outputs the sample means, answer the following: Given the sample means and your p-value for `gender`, what would you conclude from this particular row of the ANOVA table (**without looking at the rest of the ANOVA table**)? (**Hint**: This isn't meant to be a trick question; when stating your conclusion, you can indeed ignore the rest of the ANOVA table for the sake of this question. Just focus on the `gender` row specifically.)

Hide

```
aggregate(allocPct ~ gender, data = shareAllocation, FUN = mean)
```

```
##   gender allocPct
## 1 female 60.66667
## 2   male 65.66667
```

[Since the p-value is less than 0.05 (0.040891), then we can reject the null hypothesis. So the mean allocPct varies across two gender groups. but we do not know whether the mean of male group is higher or female group. For this, the sample means that we calculated in this question shows that the male group's mean of allocPct is higher than the female groups mean of allocPct.]

e. (14pts) Note that it would be inappropriate to report the conclusions you wrote in Part D when you consider the `gender` and `treatment` rows within the context of the rest of the ANOVA table. (If you’re unsure why, you may want to revisit your ANOVA table in Part B, and make sure that you ran the interactive two-way ANOVA correctly.) We’ll explore that “inappropriateness” a bit further in this problem.

Let’s consider estimating the `gender` effect within each `treatment` subgroup. Here are the datasets for each `treatment` subgroup:

Hide

```
#dataset of just subjects assigned to "dissimilar"
shareAllocation.dissimilar = subset(shareAllocation, treatment == "dissimilar")
#dataset of just subjects assigned to "similar"
shareAllocation.similar = subset(shareAllocation, treatment == "similar")
#dataset of just subjects assigned to "unknown"
shareAllocation.unknown = subset(shareAllocation, treatment == "unknown")
```

For this problem, do the following:

- (8pts) For each of the three subgroups, run the relevant t-test, where you compare the mean `allocPct` between the two gender groups. For each t-test, state your scientific conclusion based on that t-test.

Hide

```
t.test(allocPct ~ gender, data = shareAllocation.dissimilar)
```

```
##
##  Welch Two Sample t-test
##
## data:  allocPct by gender
## t = 0.8767, df = 25.273, p-value = 0.3889
## alternative hypothesis: true difference in means between group female and group male is
not equal to 0
## 95 percent confidence interval:
##  -4.942334 12.275668
## sample estimates:
## mean in group female   mean in group male
##             81.66667             78.00000
```

Hide

```
t.test(allocPct ~ gender, data = shareAllocation.similar)
```

```
##
##  Welch Two Sample t-test
##
## data:  allocPct by gender
## t = -0.1699, df = 19.953, p-value = 0.8668
## alternative hypothesis: true difference in means between group female and group male is
not equal to 0
## 95 percent confidence interval:
##  -8.852968  7.519635
## sample estimates:
## mean in group female   mean in group male
##             52.33333             53.00000
```

Hide

```
t.test(allocPct ~ gender, data = shareAllocation.unknown)
```

```
##
##  Welch Two Sample t-test
##
## data:  allocPct by gender
## t = -4.0988, df = 24.741, p-value = 0.0003907
## alternative hypothesis: true difference in means between group female and group male is
not equal to 0
## 95 percent confidence interval:
##  -27.049378  -8.950622
## sample estimates:
## mean in group female   mean in group male
##                   48                   66
```

[For all data sets the null hypothesis is H0 :μ(allocPct of female) = μ(allocPct of male) for each data set. For shareAllocation.dissimilar dataset: we fail to reject the null hypothesis that the mean allocPct is equal across the two gender groups, because the p-value is above 0.05 (0.3889). For shareAllocation.similar dataset: we fail to reject the null hypothesis that the mean allocPct is equal across the two gender groups, because the p-value is above 0.05 (0.8668). For shareAllocation.unknown dataset: we reject the null hypothesis that the mean allocPct is equal across the two gender groups, because the p-value is less than 0.05 (0.0003907).]

- (6pts) Which t-tests seem to contradict the `gender` main effect conclusion that you made in Part D, and which one(s) do not seem to contradict that conclusion? Explain in 1-3 sentences.

[In part D we generally concluded that allocPct varies across the two gender groups, however two t-tests for the shareAllocation.dissimilar dataset and shareAllocation.similar dataset contradicts with our general conclusion because we failed to reject the null hypothesis for these two data sets, and allocPct does not necessarily vary across female and male groups for these two data sets. However, the last dataset (shareAllocation.unknown) does not contradict with the general conclusion. This datasets t-test results showed that with 95% confidence, the male groups would have higher allocPct than female group CI (8.950622 , 27.049378)]

[As an aside: To better understand why there are or are not contradictions popping up in this problem, it may be helpful to look back at your interaction plot in Part A. However, for the sake of this problem, you do not have to explain *why* contradictions occur or do not occur.]

Question 2: Built Two (https://en.wikipedia.org/wiki/Built_to_Spill)-Way ANOVA (40 points)

This problem consists of fake data loosely based on a number of recent studies. The experimental units are groups of 4 subjects who are given a box of (uncooked) spaghetti and a bag of marshmallows and told to build the highest structure they can. The groups are either all kindergartners, all 5th graders, or all graduate students. Furthermore, the groups are either all male, all female, or 2 of each (mixed gender).

Here's the dataset we'll be working with:

Hide

```
build = read.csv("https://raw.githubusercontent.com/zjbranson/stat309fall2022/main/build.c
sv")
#make sure the categorical variables are factors
build$age = factor(build$age)
build$gender = factor(build$gender)
```

We see that the two explanatory variables are `age` and `gender`. Meanwhile, the outcome variable is `cmHigh`, denoting the height (in centimeters) of each structure.

a. (14pts) First, let's do some initial exploratory data analysis. For this part, answer the following questions.

- (3pts) How many treatment combinations are there in this experiment? Explain your answer in 1-2 sentences.

[There are 9 treatment combinations because both of categorial variables have 3 levels. So 3*3 = 9]

- (3pts) An experiment is considered a "balanced design" if the same number of subjects are assigned to each treatment group. Given this definition, did this experiment have a balanced design? Explain your answer in 1-2 sentences. Please include any non-graphical EDA you used to answer this question.

Hide

```
table(build$gender, build$age)
```

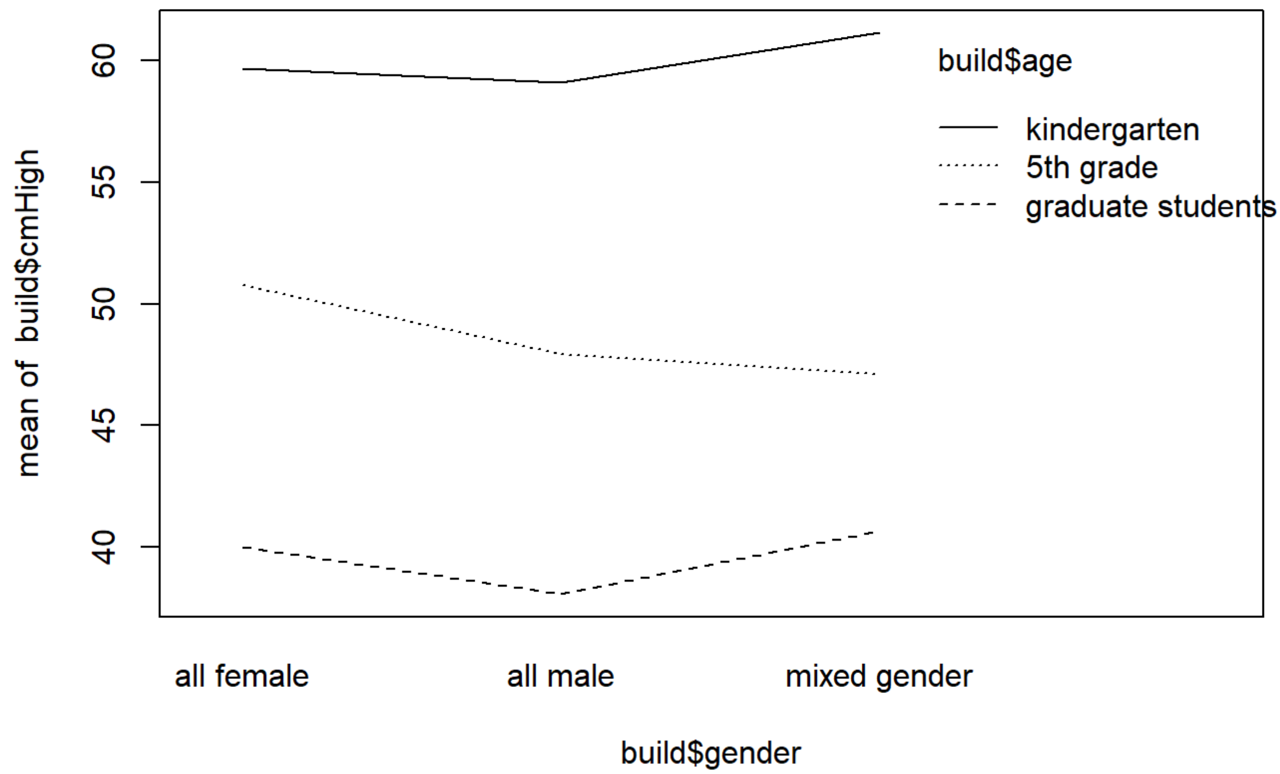
```
##
##                 5th grade graduate students kindergarten
##   all female            6                 7            7
##   all male              7                 6            7
##   mixed gender          7                 7            6
```

[Some groups have 7 subjects but some other have 6 subjects. So I cannot say there are exactly similar and there is a super balanced design, but generally I think it is a balanced design because there is no such a big difference between number of subjects in different groups.]

- (8pts) Now turn in an **interaction plot** for these data. Then, interpret the interaction plot in 1-3 sentences.

Hide

```
interaction.plot(x.factor = build$gender, trace.factor = build$age, response = build$cmHig
h)
```



[We see that younger age is associated with a higher cmHigh height (in centimeters) of each structure. Also all male group seems to have lower cmHigh compared to all female group, however no such a conclusion can be inferred for mixed gender. The lines seems parallel for (all female -all male), and partially not parallel for (all male – mixed gender) which might be a sign of the interaction for not parallel parts and not interaction for parallel parts. overall formal statistical analysis is needed to clearly understand whether there is an interaction or not.]

b. (10pts) For this part, perform two different two-way ANOVA analyses:

1. An interactive two-way ANOVA model involving gender and age.
2. An additive two-way ANOVA model involving gender and age.

In your answer, please include the output from both models. Then, answer the following: Which of the two models should be used for inference and making scientific conclusions? Explain your answer in 1-3 sentences.

Hide

```
summary(aov(cmHigh ~ gender * age, data = build))
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)    
## gender       2     18     9.0   0.242    0.786    
## age          2   4138  2068.8  55.719 1.48e-13 ***
## gender:age   4     54    13.4   0.361    0.836    
## Residuals   51   1894    37.1                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Hide

```
summary(aov(cmHigh ~ gender + age, data = build))
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)    
## gender       2     18     9.0   0.254    0.776    
## age          2   4138  2068.8  58.437 2.46e-14 ***
## Residuals   55   1947    35.4                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

[The null hypothesis is that there is not an interaction between age and gender. We fail to reject the null hypothesis, meaning that we conclude that there is not an interaction between age and gender because the p-value of gender:age interaction is above 0.05 (0.836). so one factor's effect does not depend on the other factor. hence the additive model (the second model) should be used.]

c. (16pts) Using the two-way ANOVA model you chose for inference, please do the following:

- (5pts) State your scientific conclusion for the `age` row of the ANOVA table. Be sure to report your conclusion in terms of a hypothesis test (including the corresponding null hypothesis for this test).

[H0 :μ(cmHigh of 5th grade) = μ(cmHigh of graduate students)= μ(cmHigh of kindergarten) The null hypothesis says that cmHigh does not vary across different age groups. we reject the null hypothesis Since the P value is less than 0.05 (2.46e-14). Hence, cmHigh Varies across different age groups.]

- (5pts) State your scientific conclusion for the `gender` row of the ANOVA table. Be sure to report your conclusion in terms of a hypothesis test (including the corresponding null hypothesis for this test).

[H0 :μ(cmHigh of 5th mixed gender) = μ(cmHigh of all male)= μ(cmHigh of all female) The null hypothesis says that cmHigh does not vary across different gender groups. we fail to reject the null hypothesis Since the P value is above the 0.05 (0.776). Hence, we do not have enough evidence to say that cmHigh Varies across different age groups.]

- (6pts) Finally, give an example of a question that would be a good follow up to your conclusions for `age` and `gender`, but for which we do not have (in this homework) a p-value to answer it. After stating your question, explain why it would be a natural follow-up question to the conclusions you made for `age` and `gender`. (**Hint**: At least one of the conclusions you made should be limited in some way. Thus, a natural follow-up question would be one that addresses those limitations.)

[For the age related conclusion, I stated that cmHigh Varies across different age groups. Hence, the follow up question would be "which of these three age groups are exactly different from each other in terms of mean cmHigh, And how?" So the conclusion I made for the age P value, it only stated that the cmHigh varies across different age groups but no further information is provided in terms of which groups are exactly different. Are all three groups different, or just two of them are different. if the two are different what are those two groups. also we do not have information about how they are different in terms of having higher or lower cmHigh. for example if specific two groups are different, with 95% confidence, which one has higher cmHigh than the other one. So all these questions are still un-answered, And no P value associated for these questions is presented in the additive model mentioned above. therefore further statical analysis and hypothesis testing are required to answer all these questions.]