MIDTERM INSTRUCTIONS: MUST READ

Academic Integrity Statement: MUST SIGN TO RECEIVE CREDIT FOR EXAM

The Data: Are We Learning Statistics?

Problem 1 (13 points)

Problem 2 (15 points)

Problem 3 (8 points)

Problem 4 (10 points)

Problem 5 (17 points)

Problem 6 (12 points)

Problem 7 (13 points)

Problem 8 (12 points)

Problem 9 (3pts, BUT ONLY REQUIRED FOR 36-749 STUDENTS; OPTIONAL BONUS QUESTION FOR 36-309 STUDENTS)

36-309 / 36-749 Midterm 1

Code ▾

Due Friday, October 14, 8pm (via Gradescope)

Sanaz Saadatifar

MIDTERM INSTRUCTIONS: MUST READ

- The format of the midterm is identical to that of the homework: You do everything in RStudio, and you submit a PDF on Gradescope.
- During the midterm (October 13-14), you are not allowed to talk with anyone about 36-309/36-749 material. You will be asked to sign an academic integrity statement below - you must sign this statement to receive any credit for the midterm.
- Although you can't talk with anyone else, the midterm is still "open everything." Remember that you can always refer back to previous homework/lab solutions, lectures, R demos, and the course textbook.
- **Throughout the midterm, include all the R code you used to arrive at your answers.** For example, if a question asks you to run a statistical analysis or make a graph, you should include the code that runs that analysis or makes that graph. As a result, your midterm should include the code, analyses, and graphs you used to arrive at your answers. This is equivalent to "showing your work" in other classes - if you don't show your work, we won't know how you arrived at your answers.

Academic Integrity Statement: MUST SIGN TO RECEIVE CREDIT FOR EXAM

The Data: Are We Learning Statistics?

Professor Miyawaki is interested in understanding if it's beneficial for student learning to have an instructor help students with assignments, or if learning would be just as good with a computer-based instructor or even nothing at all.

This exam focuses on data from an education experiment. First, here is the dataset:

Hide

```
educationData = read.csv("https://raw.githubusercontent.com/zjbranson/stat309fall2022/main/educationExpData.csv")
educationData$treatment = factor(educationData$treatment)
educationData$pastStats = factor(educationData$pastStats)
```

This dataset comes from a randomized experiment (conducted by Professor Miyawaki (https://en.wikipedia.org/wiki/Mitski)) to see what kind of instruction can help students learn statistics. First, Professor Miyawaki recruits undergraduate students at her university to participate in the experiment. In the experiment, each student goes to a room with a computer, where they are asked to answer statistics questions to the best of their ability. However, before the student starts answering questions, they are randomized to one of three treatment groups:

1. *Control.* The student simply answer questions on their own.
2. *Computer-assisted tutor.* While working on the statistics questions, for any question, the student can ask the computer for help on the question, e.g. a hint. This is similar to Clippy (https://en.wikipedia.org/wiki/Office_Assistant).
3. *Human tutor.* While working on the statistics questions, for any question, the student can raise their hand, and a human instructor will come into the room to help.

At the end of the study, the percentage of questions the student answered correctly is recorded.

The dataset contains the following variables:

- `perCorrect` : The percentage of statistics questions the student answered correctly during the experiment. This is the outcome of the experiment.
- `treatment` : The treatment group that the student was assigned. This is either "nothing", "computer", or "human", denoting the three above treatment groups.
- `prevMath` : The grade (in percentage) that the student received in their last math class before the experiment.
- `pastStats` : Whether the student had taken a statistics class prior to the experiment. This is either "yes" or "no".

Professor Miyawaki's main goal for this experiment is to understand if students' test performance is affected by a computer-assisted tutor or human tutor. In what follows, you'll complete questions below to help achieve this goal.

Problem 1 (13 points)

Before considering the full `treatment` variable, Professor Miyawaki first wants to understand how *any* treatment (human or computer) compares to no treatment. They thus make the following variable in the dataset:

Hide

```
educationData$anyTreatment = ifelse(educationData$treatment == "nothing", "no", "yes")
```

The above code creates a new variable, `anyTreatment`, that is equal to "no" if someone's `treatment` is "nothing" and is equal to "yes" if someone's `treatment` is "computer" or "human".

As an initial analysis, Professor Miyawaki runs the following t-test and simple linear regression.

Hide

```
#t-test
t.test(perCorrect ~ anyTreatment, var.equal = TRUE, data = educationData)
```

```
##
##  Two Sample t-test
##
## data:  perCorrect by anyTreatment
## t = -4.4609, df = 148, p-value = 1.605e-05
## alternative hypothesis: true difference in means between group no and group yes is not
equal to 0
## 95 percent confidence interval:
##  -6.586653 -2.542547
## sample estimates:
##  mean in group no mean in group yes
##           77.4030           81.9676
```

Hide

```
#simple linear regression
summary(lm(perCorrect ~ anyTreatment, data = educationData))
```

```
## 
## Call:
## lm(formula = perCorrect ~ anyTreatment, data = educationData)
## 
## Residuals:
##      Min      1Q   Median      3Q      Max
## -15.1630  -3.9501   0.3474   3.6710  15.6924
## 
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       77.4030     0.8355  92.646  < 2e-16 ***
## anyTreatmentyes    4.5646     1.0232   4.461  1.6e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 5.908 on 148 degrees of freedom
## Multiple R-squared:  0.1185, Adjusted R-squared:  0.1126
## F-statistic:  19.9 on 1 and 148 DF,  p-value: 1.605e-05
```

Hide

```
81.9676-77.4030
```

```
## [1] 4.5646
```

Using the output from the above code, answer the following three questions.

- (4pts) State the scientific conclusion you would make based on the t-test output. In your conclusion, be sure to mention a p-value, a point estimate, and a null hypothesis.

[H0 :μ(perCorrect of yes) = μ(perCorrect of no.)The null hypothesis says that there is no difference between mean of the percentage of statistics questions the student answered correctly during the experiment for two groups of YES (students who received a type of treatment [human or computer]) and NO (no treatment received at all). Since the P_value is less than 0.05 (1.605e-05), we reject the null hypothesis. The point estimate is mean in group yes - mean in group no = 81.9676 - 77.4030 = 4.5646. and with the 95% confidence, the true populations mean difference would fall within this range of (2.542547, 6.586653), (or in other words population of "Yes" would be higher than "NO" in this range).]

- (5pts) Using mathematical notation, write down two **statistical models**: One for the t-test analysis run above, and one for the simple linear regression analysis run above. Your models should include **parameters**; be sure to clarify what your parameters represent.

[A mathematical way to write the statistical model for simple regression is: perCorrect iid~N($\beta 0$ + βy · $Iy,\sigma 2$) : perCorrect = 77.4030 + 4.5646 Iy Here, $\beta 0$ is the intercept parameter. Iy denotes the indicator variable for treatment, where Iy = 1 if someone received the treatment and Iy = 0 if they did not receive the treatment. βy is the change in the intercept when the student is in a treatment group or in other words it is an additional effect of Treatment condition on the intercept (in reference to "treatment NO group"), and $\sigma 2$ is the variance parameter. Meanwhile, the notation N($\mu,\sigma 2$) denotes a Normal distribution with mean μ and variance $\sigma 2$. Also note that "iid" stands for independent and identically distributed. Here's a word-based way to state the statistical model: We assume that the

outcome variable (time) is Normally distributed, and outcome measurements are independent of each other. We assume that the mean perCorrect is equal to an intercept (which can change depending on whether the student is in Yes group or NO group) and that the variance is equal to a constant value (usually denoted as "sigma-squared").]

[A mathematical way to write the statistical model for t-test is: Y iid ~ N($\mu 1$, $\sigma 2$) : perCorrect iid ~ N(μno, $\sigma 2$) if assigned to Treatment "NO" Y iid ~ N($\mu 2$, $\sigma 2$) : perCorrect iid ~ N(μyes, $\sigma 2$) if assigned to Treatment "YES" The statistical model of the two-sample t-test for this example is that perCorrect, the percentage of statistics questions the student answered correctly during the experiment, denoted by Y, has a Normal distribution with mean μno and variance $\sigma 2$ when assigned to the "No" group. Meanwhile, perCorrect has a Normal distribution with mean μyes and variance $\sigma 2$ when assigned to the "YES" group. All measures of perCorrect are independent and identically distributed.]

- (4pts) In 1-3 sentences, discuss to what extent the parameters in the t-test statistical model are related to the parameters in the simple linear regression statistical model. For example: Do any of the parameters represent the same quantity, or can any of the parameters in one model be represented by parameters in another model?

[Both statistical models have same outcomes as perCorrect which is considered to be independent and identically distributed with variance equal to $\sigma 2$ in both models. On the other hand, regression model's intercept equals to the mean μno (when students have no treatment which is equal to 77.4030). moreover for the time when students have treatment (Yes group), the mean of that equals to the intercept of regression model plus the effect of treatment condition on the intercept (mean in group yes = 81.9676 = 77.4030 + 4.5646)]

This is the only question that uses the `anyTreatment` variable. We'll use the original variables in the dataset for the remainder of this exam.

Problem 2 (15 points)

Before running more analyses on the data, Professor Miyawaki wants to do a bit more exploration of the data. Answer the following four questions.

- (5pts) Professor Miyawaki asks, "I would like to know the relationship between `treatment` and `perCorrect` ." Given this, what kind of **graphical EDA** is most appropriate here? Explain in 1-2 sentences. Then, make the appropriate graphical EDA involving `treatment` and `perCorrect` . Then, write a 1-2 sentence interpretation explaining the main takeaways of your EDA.

[Boxplot would be the best EDA because the outcome variable, perCorrect, is a quantitative outcome, while the explanatory variable, treatment, is a categorial variable.]

Hide

```
boxplot(perCorrect~treatment, data = educationData)
```
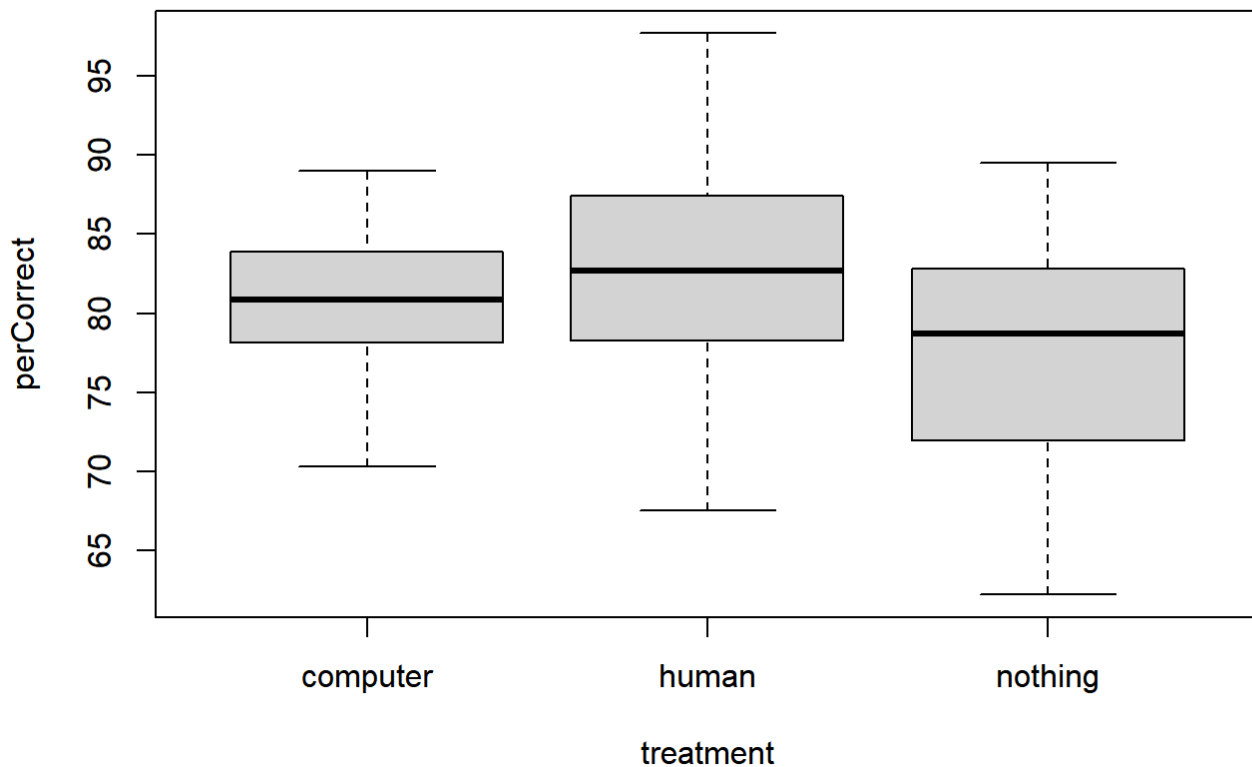
[It is showing that median perCorrect of human group is higher than computer group, and computer group is higher than nothing group. No outliers is seen. Moreover, this plot can be used to assed equal variance assumption. It might be a bit violating for "nothing" and "compute"r groups, but generally it appears safe to assume that the variances are equal to each other - the IQR of each group (i.e., the height of each box in the above plot) appear similar, in the sense that one IQR isn't at least twice as big as another]
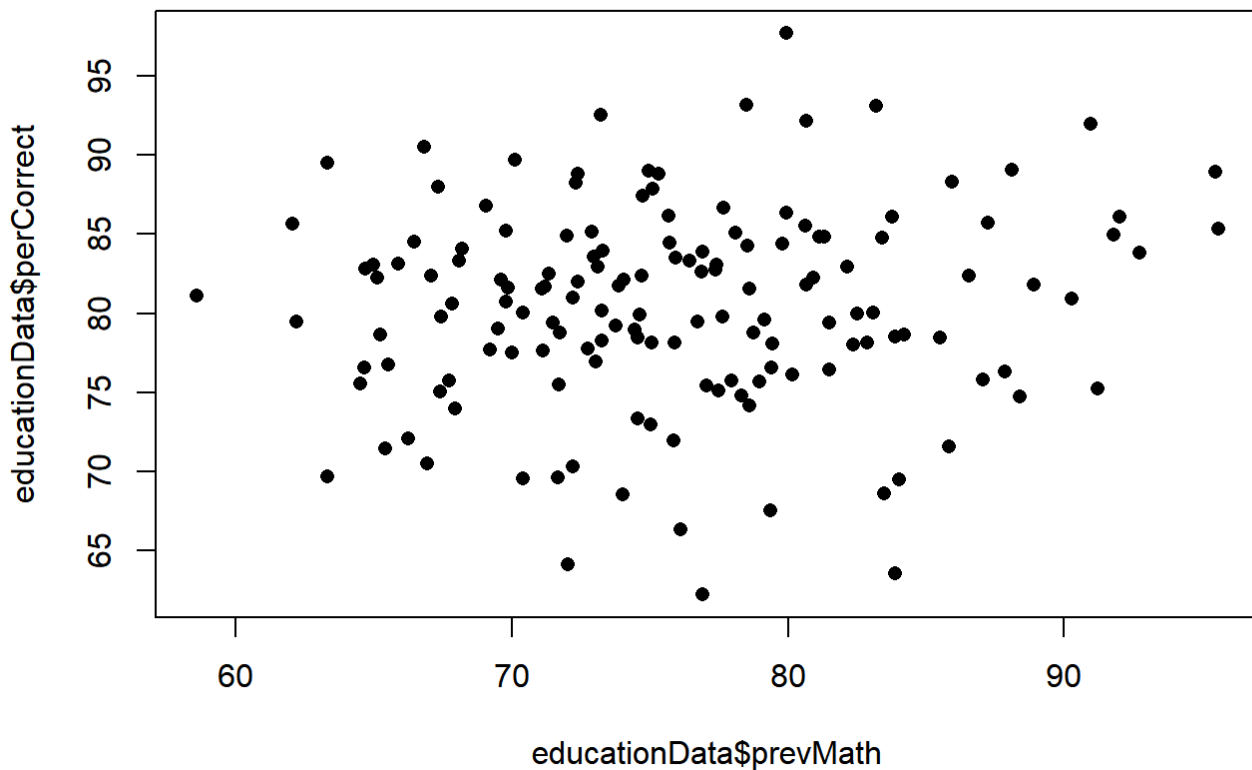
- (3pts) Professor Miyawaki asks, "I would like to know the relationship between `prevMath` and `perCorrect` ." Given this, what kind of **graphical EDA** is most appropriate here? Explain in 1-2 sentences. Then, make the appropriate graphical EDA involving `prevMath` and `perCorrect` . (For this part, you don't need to interpret the EDA.)

[scattorplot would be the best EDA because the outcome variable, perCorrect, is a quantitative outcome, while the explanatory variable, prevMath, is also a quantitative variable.]

Hide

```
plot(educationData$prevMath, educationData$perCorrect,
main = "Scatterplot of prevMath vs perCorrect",
pch = 16)
```

Scatterplot of prevMath vs perCorrect

- (3pts) Professor Miyawaki asks, "I would like to know the relationship between `treatment` and `pastStats` ." Given this, what kind of **non-graphical EDA** is most appropriate here? Explain in 1-2 sentences. Then, make the appropriate non-graphical EDA involving `treatment` and `pastStats` . (For this part, you don't need to interpret the EDA.)

[A contingency table would be appropriate since both are explanatory variables and both are categorial variables. So contingency table would be better to use to understand How many subjects were assigned to each combination of treatment and pastStats.]

Hide

```
table(educationData$treatment, educationData$pastStats)
```

```
##
##           no yes
##   computer 29  21
##   human    33  17
##   nothing  33  17
```

- (4pts) In the above, you should have made three forms of EDA. Which of the EDA, if any, could be used to assess the *internal validity* of this study? Explain in 1-3 sentences.

[Internal validity is tested to see whether there is big differences in explanatory variables between treatment groups that might violate the internal validity. The graphical EDAs above, were to see the relation between the outcome variable and an explanatory variable. Only the last contingency table

was used to compare two explanatory variables in terms of number of subjects assigned to each. So only the last one can be used to see whether there is a major difference between number of subjects in each group.]

Problem 3 (8 points)

Now Professor Miyawaki would like to assess if the average `perCorrect` differs across levels of `treatment`. What is the most appropriate statistical analysis in this case? Explain in 1-2 sentences. Then, perform the appropriate statistical analysis, and state your scientific conclusions in 1-2 sentences. After doing so, compare the scientific conclusion you made here to the scientific conclusion you made from the t-test in Problem 1. In particular: Which scientific conclusion do you think provides more information about how the average outcome differs across the three different treatment levels? Explain in 1-2 sentences.

Hide

```
summary(aov(perCorrect ~ treatment, data = educationData))
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## treatment     2    806   403.0   11.72 1.89e-05 ***
## Residuals   147   5054    34.4
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

[Outcome is quantitative, however the explanatory variable is a categorial variable with more than 2 levels (3 levels), hence the ANOVA should be used. Since the P-value is less than 0.05 (1.89e-05), the we can reject the null hypothesis that preCorrect is same between three groups of computer, human, and nothing (H0: μhuman = μcomputer = μnothing). I think the t-test provides more information, because it not only says whether there is a difference between 2 groups but also says which one is higher than the other one and how high or less is. Also 95% confidence interval is also given in t-test. However, the ANOVA, only says whether there is a difference between 3 groups here. It does not say exactly which groups are different than the other ones, and ho different they are. Also no confidence interval information is mentioned in ANOVA.]

Problem 4 (10 points)

Now we will consider analyses that use `treatment` and `prevMath` as explanatory variables for the outcome `perCorrect`. Professor Miyawaki decides to run two different *interactive regression models* involving `treatment` and `prevMath`. They are presented below.

Hide

```
#the two regression models
intReg1 = lm(perCorrect ~ treatment*prevMath, data = educationData)
educationData$treatment = relevel(educationData$treatment, ref = "nothing")
intReg2 = lm(perCorrect ~ treatment*prevMath, data = educationData)
#the corresponding summary output
summary(intReg1)
```

```
## 
## Call:
## lm(formula = perCorrect ~ treatment * prevMath, data = educationData)
## 
## Residuals:
##      Min      1Q  Median      3Q     Max
## -16.0867  -3.1117   0.6922   3.3865  13.8608
## 
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 79.82444    8.42365   9.476   <2e-16 ***
## treatmenthuman              -19.01779   11.87302  -1.602   0.1114
## treatmentnothing              7.10375   11.86709   0.599   0.5504
## prevMath                      0.01458    0.11241   0.130   0.8970
## treatmenthuman:prevMath       0.27301    0.15575   1.753   0.0818 .
## treatmentnothing:prevMath    -0.14025    0.15711  -0.893   0.3735
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 5.758 on 144 degrees of freedom
## Multiple R-squared:  0.1853, Adjusted R-squared:  0.157
## F-statistic: 6.551 on 5 and 144 DF,  p-value: 1.604e-05
```

Hide

```
summary(intReg2)
```

```
## 
## Call:
## lm(formula = perCorrect ~ treatment * prevMath, data = educationData)
## 
## Residuals:
##      Min      1Q  Median      3Q     Max
## -16.0867  -3.1117   0.6922   3.3865  13.8608
## 
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  86.9282     8.3588  10.400  < 2e-16 ***
## treatmentcomputer            -7.1038    11.8671  -0.599  0.55037
## treatmenthuman              -26.1215    11.8271  -2.209  0.02878 *
## prevMath                     -0.1257     0.1098  -1.145  0.25412
## treatmentcomputer:prevMath    0.1403     0.1571   0.893  0.37351
## treatmenthuman:prevMath       0.4133     0.1538   2.686  0.00808 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 5.758 on 144 degrees of freedom
## Multiple R-squared:  0.1853, Adjusted R-squared:  0.157
## F-statistic: 6.551 on 5 and 144 DF,  p-value: 1.604e-05
```

Answer the following questions about the above output.

- (4pts) Looking at the above output, Professor Miyawaki says, "There are more significant p-values in the summary output for `intReg2` compared to `intReg1`. Thus, I think the `intReg2` model fits the data better than the `intReg1` model." Do you Agree or Disagree with this claim? State Agree or Disagree, and then explain in 1-2 sentences.

[I disagree, because these to models are same, just the reference level has been changed. Another sign for this is that both models have the same R-squared and adjusted R-squared.]

- (6pts) Looking at the above output, Professor Miyawaki makes two observations:

1. "It looks like, for `intReg1`, the estimated coefficient for `treatmentnothing:prevMath` is exactly equal to the estimated coefficient for `treatmentcomputer:prevMath` in `intReg2`, but multiplied by -1. Also, the p-values for these two coefficients are exactly the same."

2. "The estimated coefficient for `treatmenthuman:prevMath` is completely different between the two models. Also, this estimated coefficient is only significant for `intReg2`."

Given the above, Professor Miyawaki is a bit confused. To help ease Professor Miyawaki's confusion, explain in 1-3 sentences why the phenomena Professor Miyawaki noted above occur. In other words: Why does #1 and #2 above occur?

[In model 1, the reference level is computer, but in model 2 the reference level is nothing, and inference from regression is always in reference to the baseline category (reference). Therefore, it is understandable if treatmenthuman:prevMath is different in both models, because in first model, human is compared to the computer (results show that the interaction between these two (0.27301) is not significant (p-value above 0.05 (0.0818))), but in the second model, human is compared to nothing (results show that the interaction between these two (0.4133) is significant(p-value less than 0.05 (0.00808)))]

[For the first observation as well, it is understandable to see same coefficients and P-values because in the first model, nothing is compare to the computer, and in the second model, computer is compared to the nothing. These two are same comparisons so they would have same coefficients and P-values. If inclusion of computer adds to the nothing (model 2 where the coefficient is positive), the inclusion of nothing to computer would has negative effect at the same amount so it is understandable that coefficient is multiplied by -1 .]

(**Hint**: In #1, the numbers between `intReg1` and `intReg2` really are exactly the same. The very minor discrepancy between the two sets of output is due to very slight rounding error in R.)

Problem 5 (17 points)

For this part, let's continue to focus on the `linReg2` model from Problem 4. For reference, here is the summary output for `linReg2`:

Hide

```
summary(intReg2)
```

```
##
## Call:
## lm(formula = perCorrect ~ treatment * prevMath, data = educationData)
##
## Residuals:
##      Min      1Q  Median      3Q     Max
## -16.0867  -3.1117   0.6922   3.3865  13.8608
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   86.9282     8.3588  10.400  < 2e-16 ***
## treatmentcomputer             -7.1038    11.8671  -0.599  0.55037
## treatmenthuman               -26.1215    11.8271  -2.209  0.02878 *
## prevMath                      -0.1257     0.1098  -1.145  0.25412
## treatmentcomputer:prevMath     0.1403     0.1571   0.893  0.37351
## treatmenthuman:prevMath        0.4133     0.1538   2.686  0.00808 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.758 on 144 degrees of freedom
## Multiple R-squared:  0.1853, Adjusted R-squared:  0.157
## F-statistic: 6.551 on 5 and 144 DF,  p-value: 1.604e-05
```

Answer the following questions about the above output.

- (5pts) Write the **prediction equation** according to the above output. Your prediction equation should communicate what the estimated mean `perCorrect` is for any given value of `prevMath` and `treatment`, according to `intReg2`. Be sure to define any additional notation you need to write out your prediction equation. (**Hint**: For this question, you should write out just *one* prediction equation; to do so, you indeed may have to introduce some additional notation, which you must define.)

[perCorrect = ß0 + ßp* prevMath + ßc * Icomputer+ ßh * Ihuman + ßpc * prevMath * Icomputer + ßph * prevMath * Ihuman]

[perCorrect = 86.9282 - 0.1257* prevMath - 7.1038 * Icomputer - 26.1215* Ihuman + 0.1403 * prevMath * Icomputer + 0.4133 * prevMath * Ihuman]

[ß0: intercept. ßp: slope of prevMath in nothing group (Estimated mean change in perCorrect associated with a one-unit increase in prevMath when othe variables are constant). Icomputer: indicator variable for computer (when data is in computer group, this is 1, other wise is 0). Ihuman: indicator variable for human (when data is in human group, this is 1, otherwise is 0). ßc: change in the intercept if the data is in the computer group. ßh: change in the intercept if the data is in the human group. ßpc: Estimated mean change in perCorrect associated with a one-unit increase in prevMath for computer group and others are constant. ßph: Estimated mean change in perCorrect associated with a one-unit increase in prevMath for human group and others are constant]

- (8pts) Professor Miyawaki is particularly interested in the difference in estimated mean `perCorrect` when someone is assigned to the human treatment ("human") versus no treatment ("nothing"). For this problem, call this "the treatment effect" (i.e., the estimated mean `perCorrect` when assigned to "human" *minus* the estimated mean `perCorrect` when assigned to "nothing"). Using your prediction equation above, what is the estimated treatment effect (between "human" and "nothing") for someone whose `prevMath = 80`? Show your work, and report a single number. After doing so: Is your estimate

of the treatment effect in this case considered an *interpolation* or an *extrapolation*? Explain in one sentence, and provide any additional code/output you need to answer this question.

```
TRColor = ifelse(educationData$treatment == "nothing", "black",
ifelse(educationData$treatment == "human", "red", "blue"))

#make a scatterplot:
plot(educationData$prevMath, educationData$perCorrect,
     col = TRColor,
     main = "Scatterplot of prevMath vs perCorrect",
     pch = 16)

#make a legend in the top-left of the plot:
legend("topleft",
       legend = c("nothing", "human", "computer"),
       col = c("black", "red", "blue"),
       pch = 16)
```


Scatterplot of prevMath vs perCorrect
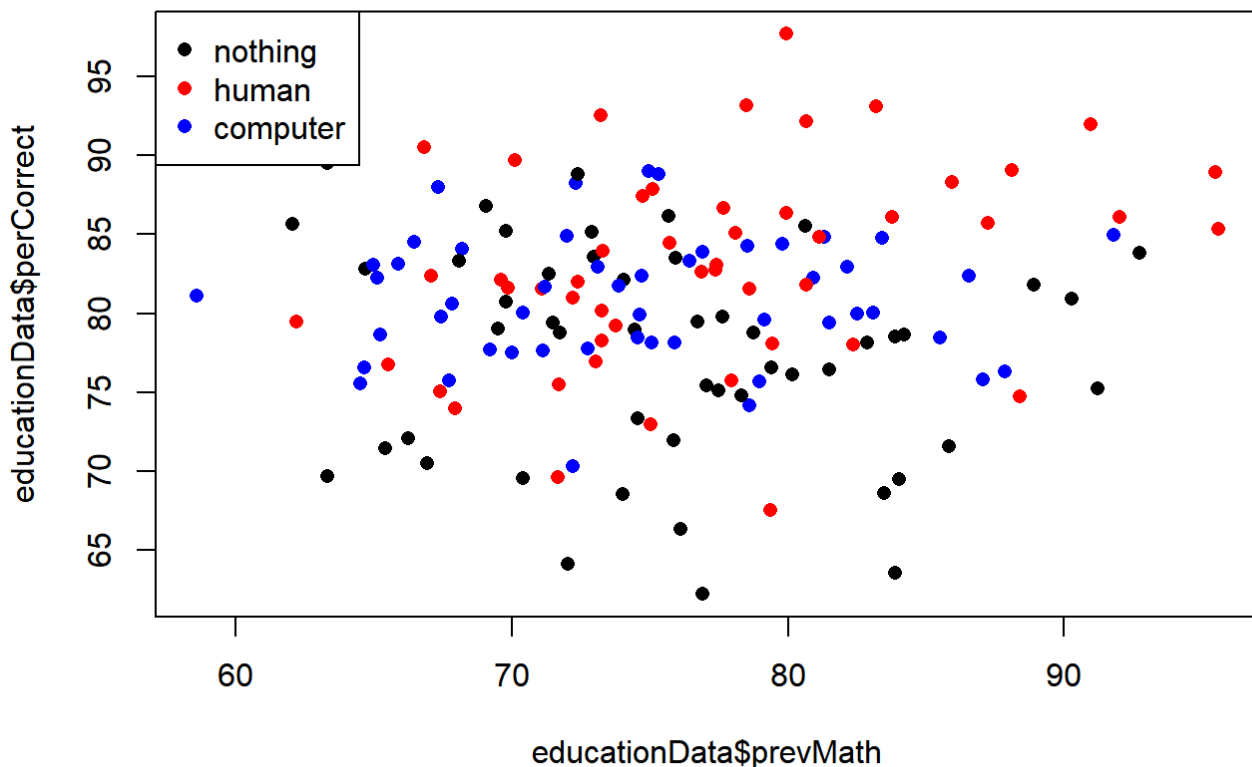
[human group: perCorrect = 86.9282 - 0.1257* prevMath - 26.1215 + 0.4133 * prevMath = 60.8067 + 0.2876* prevMath = 60.8067 + 0.2876* 80 = 83.8147]

[nothing group: perCorrect = 86.9282 - 0.1257* prevMath = 86.9282 - 0.1257* 80 = 76.8722]

[treatment effect = 83.8147 - 76.8722 = 6.9425]

[it is interpolation, because based on the scatterplot shown in problem 2 and the scatterplot here, there are already many data in the range of prevmath = 80. So it is interpolation.]

- (4pts) Looking at the above `intReg2` output, Professor Miyawaki states: "I see that the estimated coefficient for `treatmenthuman` is -26.12, and the p-value is (approximately) 0.03, denoting statistical significance. Thus, I conclude that this is evidence that being assigned to the"human" group decreases the `perCorrect` , on average, compared to being assigned to the "nothing" group." Do you Agree or Disagree with this claim? State Agree or Disagree, and then explain in 1-3 sentences.

[I disagree, we just calculated an example above which showed that human group's perCorrect was higher than nothing group's. the reason is that since we used the interactive model, there is an interaction between these two factors. So being assigned to the human group not only affects the intercept but also affects the slop of the prevmath. Which in this case it has a positive effect on slop so overall increases the perCorrect.]

Problem 6 (12 points)

Now Professor Miyawaki wants to run a simple linear regression, where `perCorrect` is the outcome and `prevMath` is the explanatory variable. Answer the following questions.

- (7pts) First, run the aforementioned simple linear regression, where `perCorrect` is the outcome and `prevMath` is the explanatory variable. Be sure to include the `summary()` output from this regression. After doing so, you should see output for `prevMath` . Using this output, write your interpretation of the **point estimate** of `prevMath` . Then, contrast this interpretation with how you would interpret the point estimate of `prevMath` for the `inReg2` output in the previous problem (Problem 5).

Hide

```
summary(lm(perCorrect~prevMath, data = educationData))
```

```
##
## Call:
## lm(formula = perCorrect ~ prevMath, data = educationData)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -18.289  -3.204   0.769   3.709  16.888
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 74.37851    5.20843  14.280   <2e-16 ***
## prevMath     0.07997    0.06831   1.171    0.244
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.263 on 148 degrees of freedom
## Multiple R-squared:  0.009174,   Adjusted R-squared:  0.002479
## F-statistic:  1.37 on 1 and 148 DF,  p-value: 0.2436
```

[For this simple regression, we see that the p-value for the slope is above than 0.05 (0.244), and our point estimate is positive (0.07997); thus, we fail to reject the null hypothesis that the slope parameter is zero. However considering only this model for this given sample, the point estimate is positive. This suggests that, as change in prevMath increases, this is associated with a higher change in perCorrect. In terms of contrasting with the interaction model, since the interaction between multiple factors are considered in that model, the slop for prevMath is different. For that model, we see that the p-value for the slope is above than 0.05 (0.25412), and our point estimate is negative (-0.1257); thus, we fail to reject the null hypothesis that the slope parameter is zero. However considering only this model for this given sample, the point estimate is negative. This suggests that, as change in prevMath increases, this is associated with a lower change in perCorrect.]

- (5pts) Now write your interpretation of the intercept coefficient in this simple linear regression model. After writing your interpretation, answer the following: Does the intercept have a scientifically meaningful interpretation, within the context of this dataset? Explain in 1-2 sentences.

[For this regression model, we see that intercept's P-value is less than 0.05 (<2e-16), which reject the null hypothesis that intercept is zero. It is not zero and it is 74.37851. which means that in case, prevMath = 0, the perCorrect would be 74.37851.for the second part, I think Yes, it has a scientifically meaningful interpretation, within the context of this dataset. Because prevMath is the grade of students in math class. Although our dataset does not include a data where prevMath = 0, but there could be such a data, as some students can get zero for a course grade.]

Problem 7 (13 points)

We'll continue working with the simple linear regression model in Problem 6. But now, we'll consider various assumptions of that model. Answer the following questions.

- (7pts) Using your simple linear regression model from Problem 6, make the appropriate graphical EDA to assess the *Equal Variance assumption*. Using your graphical EDA, state whether you think the Equal Variance assumption is Definitely Plausible, Somewhat Plausible, Somewhat Not Plausible, or Definitely Not Plausible, and explain in 1-2 sentences. Then, answer the following: According to the simple linear regression `summary()` output from Problem 6, what is your estimate of the "variance" in the Equal Variance assumption? Explain in 1-2 sentences.

Hide

```
#making the residuals
res = residuals(lm(perCorrect~prevMath, data = educationData))
#making the fitted values:
fits = fitted(lm(perCorrect~prevMath, data = educationData))
#making the residual-vs-fit plot:
plot(fits, res)
abline(h=0)
```

fits
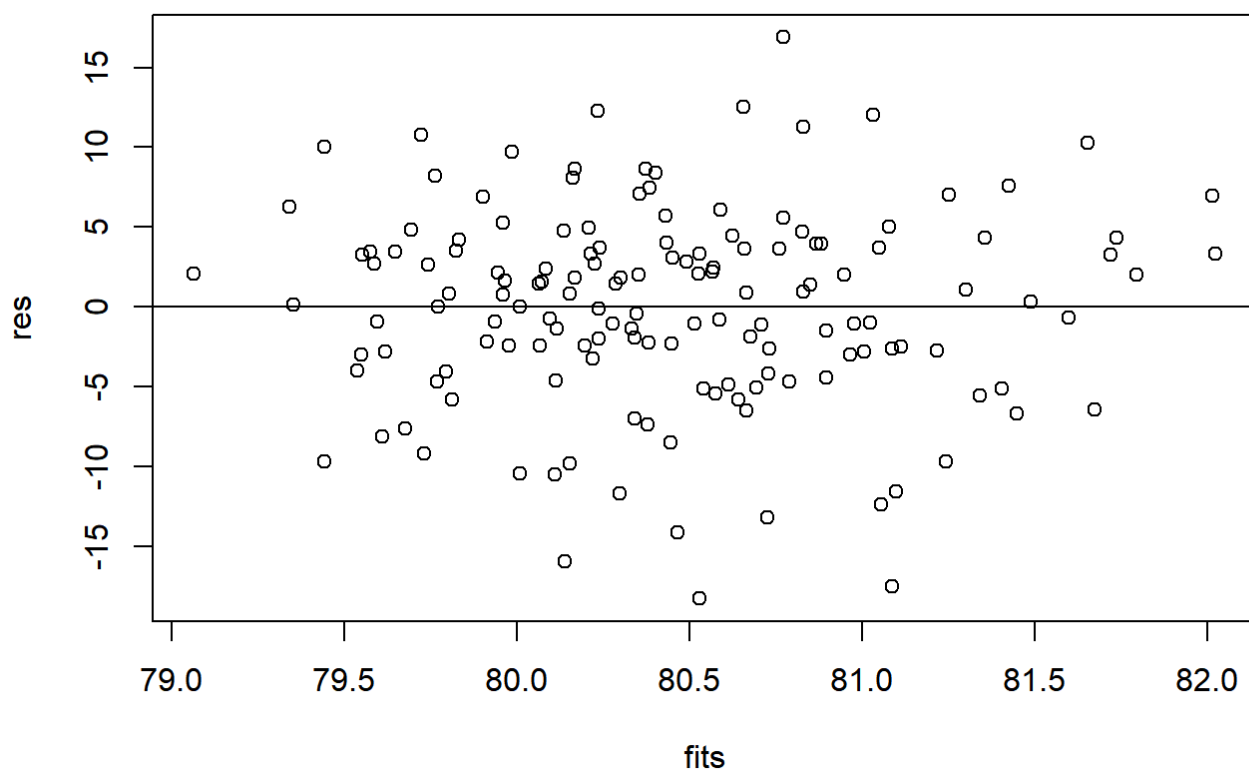
[The equal variance assumption is Somewhat Plausible There does not appear to be a strong violation of the equal variance assumption: The points appear to be equally vertical spread from left-to-right. Variance is σ 2 which is s 2. From the model in part 6, Residual Standard Error is s, not s 2 ! So, our estimate of σ 2 is 6.263^2 = 39.22.]

- (3pts) Thinking more about the simple linear regression model, Professor Miyawaki decides to make the following histogram:

Hide

```
hist(educationData$prevMath,
     breaks = 20,
     main = "Histogram of Previous Math Grade",
     xlab = "Previous Math Grade", col = "forestgreen")
```

Histogram of Previous Math Grade
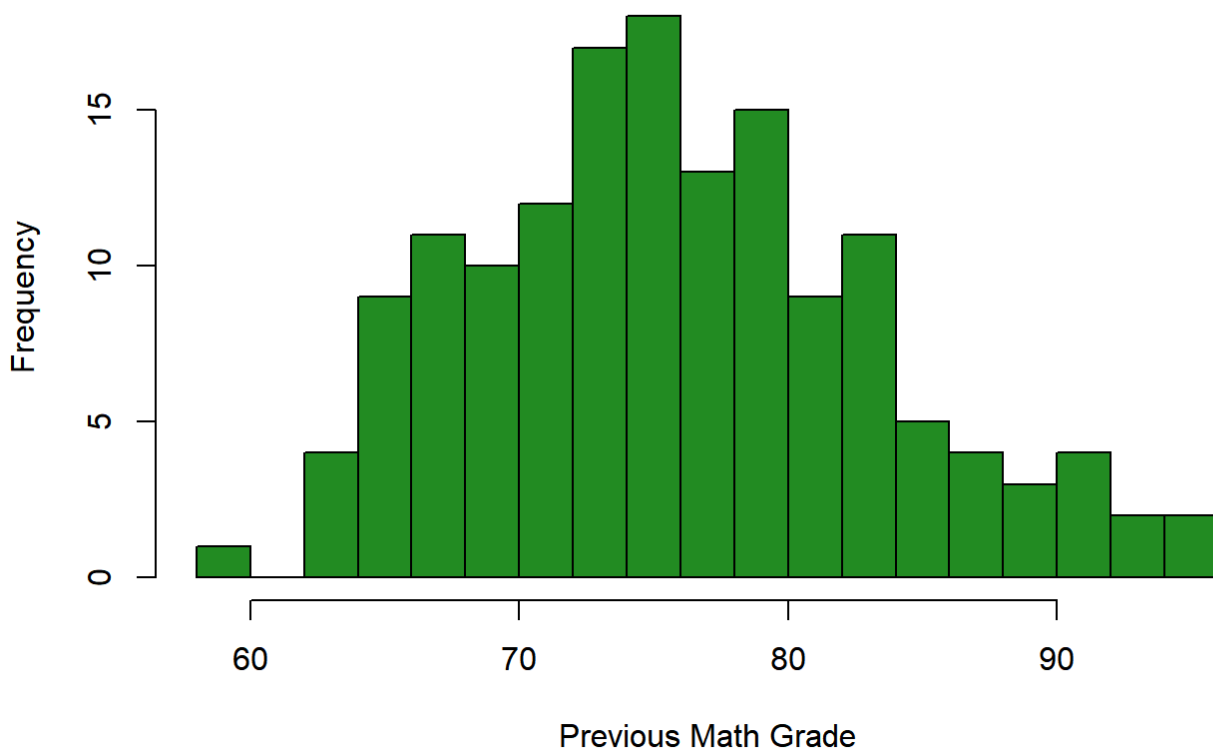
Professor Miyawaki correctly notes that this variable appears to be slightly right-skewed, such that it appears to be slightly non-Normally distributed. Because of this, Professor Miyawaki says, "This histogram serves as evidence that the Normality assumption for the simple linear regression model may be violated." Do you Agree or Disagree with this claim? State Agree or Disagree, and explain in 1-3 sentences.

[I disagree, because to check the normality assumption, the distribution of residuals should be checked not the distribution of prevMath. Since no data so far is provided about the distribution of residuals, e cannot make any conclusion about the normality assumption of this simple linear model.]

- (3pts) Professor Miyawaki then says: "Wait, I just realized something about this experiment. For students assigned to the"human" group, they are collaborating with another person—the human tutor. Thus, the Independence assumption for this model probably doesn't hold." Do you Agree or Disagree with this claim? State Agree or Disagree, and then explain in 1-2 sentences.

[I disagree, because they are independently collaborating with the human tutor not as a group. Also they are not collaborating with each other, they are collaborating with the same human tutor which it should be, otherwise the internal validity would be at risk.]

Problem 8 (12 points)

For this last problem, Professor Miyawaki would like to understand how `perCorrect` may depend on `pastStats` and `treatment`. Because these are two categorical explanatory variables, we should use two-way ANOVA. For this part, answer the following questions:

- (3pts) Professor Miyawaki knows that there is *additive* two-way ANOVA and *interactive* two-way ANOVA. For this part, write code that appropriately runs both of these two-way ANOVA models. Be sure that `summary()` output is displayed, which is all you have to do for this part.

Hide

```
summary(aov(perCorrect ~ pastStats*treatment, data = educationData))
```

```
##                      Df Sum Sq Mean Sq F value    Pr(>F)
## pastStats             1      0     0.0   0.000     0.995
## treatment             2    806   403.0  11.838  1.74e-05 ***
## pastStats:treatment   2    151    75.5   2.218     0.113
## Residuals           144   4903    34.0
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Hide

```
summary(aov(perCorrect ~ pastStats+treatment, data = educationData))
```

```
##             Df Sum Sq Mean Sq F value    Pr(>F)
## pastStats    1      0     0.0    0.00     0.995
## treatment    2    806   403.0   11.64  2.03e-05 ***
## Residuals  146   5054    34.6
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- (4pts) Now, using the above output, state your scientific conclusions, based on the appropriate two-way ANOVA analysis. In your answer, be sure to explain which two-way ANOVA model you're using to make scientific conclusions (the additive model or the interactive model).

[The null hypothesis is that there is not an interaction between pastStats and treatment. Since the P-value for the pastStats:treatment is above 0.05 (0.113), then We fail to reject the null hypothesis, meaning that we conclude that there is not an interaction between pastStats and treatment. Thus, we will be using the additive model to make scientific conclusions.]

- (5pts) Now make an *interaction plot* to complement the two-way ANOVA analyses you ran above. After making your plot, answer the following: What additional information, if any, does the interaction plot provide, that isn't already provided by the two-way ANOVA analyses above? Discuss in 1-3 sentences.

Hide

```
interaction.plot(x.factor = educationData$pastStats, trace.factor = educationData$treatment,
response = educationData$perCorrect)
```

mean of educationData$perCorrect vs educationData$pastStats, with legend educationData$treatme: human (solid), computer (dashed), nothing (dotted).
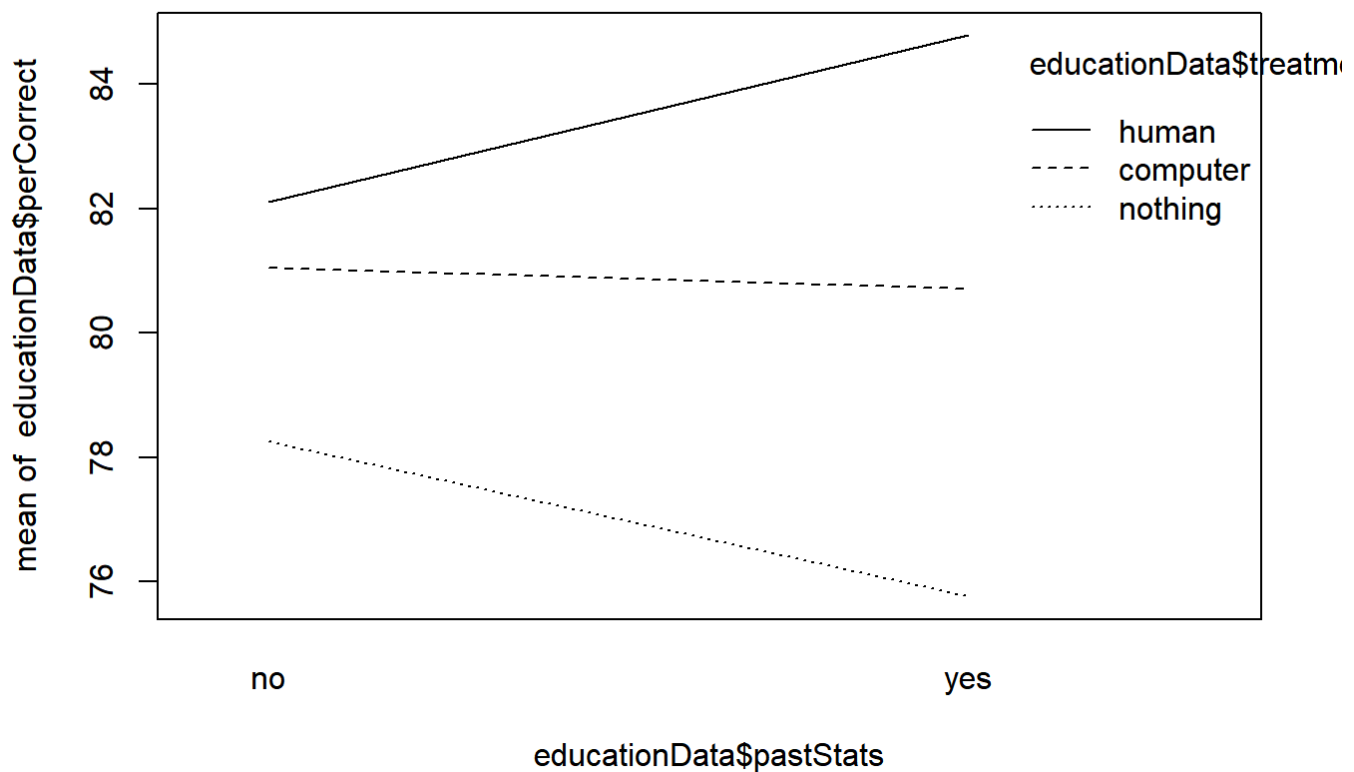
[Interaction plot shows the sample means. In terms of the information that is not provided in ANOVA, This plot mentions that sample mean of perCorrect is generally highest in the human, which is higher than computer, and computer is higher than the nothing group's mean of perCorrect. Moreover the plot shows that having taken a stat course previously for the human group resulted in higher mean of perCorrect. This is opposite for the nothing group. having taken a stat course previously for the nothing group resulted in lower mean of perCorrect. The plot does not show a significant effect of the previous stats course being taken on the computer group's mean of perCorrect (however it very slightly decreases).]

Problem 9 (3pts, BUT ONLY REQUIRED FOR 36-749 STUDENTS; OPTIONAL BONUS QUESTION FOR 36-309 STUDENTS)

Throughout this exam, you've used many different statistical methods to analyses this dataset. Now we'll take a step back and think about situations that would bring doubt to these analyses. For this problem, do the following.

- In 1-3 sentences, provide an example of additional information that would bring into doubt the *external validity* of this study. In your discussion, be sure to explain why your example would make a researcher doubt the external validity of this study.

[Professor Miyawaki is interested in understanding if it's beneficial for student learning to have an instructor help students with assignments, or if learning would be just as good with a computer-based instructor or even nothing at all. So he will generalize his study results to all students, however he only used undergraduate students, also other factors can affect the results such as the years of experience students have working with mathematics., which his not considered in the data set. So we might not generalize the results to the general student learning concept. Imagine comparing a PhD students learning with undergrad students learning, no PhD student was involved in the experiment, and a main difference between these two is that PHD students might have more experience in mathematics which might affect the results and they would prefer to learn by themselves rather than using a tutor or a computer.]

- In 1-3 sentences, provide an example of additional information that would bring into doubt the *construct validity* of this study. In your discussion, be sure to explain why your example would make a researcher doubt the construct validity of this study.

[In terms of current variables in data set they are directly measured, however for example for the test scores (perCorrect: The percentage of statistics questions the student answered correctly during the experiment. This is the outcome of the experiment.) we did not consider factors or variables that can affect the perCorrect but cannot directly measured (like wellness of students, distraction vs focus levels, or their stress level at the time of taking the test). These factors not being considered can be concerning in terms of construct validity.]

Hint: For both of these above, the "additional information" is hypothetical. In other words: You can come up with any example of "additional information" you'd like, as long as it's clearly related to external validity or construct validity, as well as the study discussed in this exam.