C2M3: Peer Reviewed Assignment

Outline:

The objectives for this assignment:

1. Motivate the use of two-way ANOVA through real data analysis examples.
2. Interpret the two-way ANOVA model, with and without interaction terms.
3. Construct and interpret interaction plots to visually assess the importance of an interaction term.
4. Conduct hypothesis tests to decide whether a two-way ANOVA interaction term is statistically significant.
5. Use the two-way ANOVA and ANCOVA models to answer research questions using real data.

General tips:

1. Read the questions carefully to understand what is being asked.
2. This work will be reviewed by another human, so make sure that you are clear and concise in what your explanations and answers.

Problem 1: e-reader data

In this assignment, we learn to answer our two-way ANOVA research questions through the analysis of real data. We will use the ereader data. The study that generated these data can be found here: [P.-C. Chang, S.-Y. Chou, K.-K. Shieh (2013). "Reading Performance and Visual Fatigue When Using Electronic Displays in Long-Duration Reading Tasks Under Various Lighting Conditions," Displays, Vol. 34, pp. 208-214. (http://users.stat.ufl.edu/~winner/data/ereader1.txt)](http://users.stat.ufl.edu/~winner/data/ereader1.txt))

Electronic paper display devices, such as the Amazon Kindle have changed the way that people read. But has it changed it for the better? In a 2013 study, titled "Reading Performance and Visual Fatigue When Using Electronic Displays in Long-Duration Reading Tasks Under Various Lighting Conditions", researchers set out to ask whether reading speed, a continuous variable, differed across different electronic paper displays. But, in addition, they were also interested in whether different lighting conditions impacted reading speed. As such, this experiment had one response with two different factors:

1. Device type: three different types.
   A. Sony PRS-700 with a 6-in. display, 800 × 600 resolution;
   B. Amazon Kindle DX with a 9.7-in. display, 1200 × 824 resolution; and
   C. iRex 1000S with a 10.2-in. display, 1024 × 1280 resolution.

1. Lighting Condition: four different conditions (200Lx,500Lx, 1000Lx, 1500Lx), Lx = lux, one lumen per square meter

1. Reading Time: measured in seconds.

With these data, we might ask the following **research questions**:

1. Are the effects of device type significant? That is, is there any evidence that suggests that individuals read faster or slower based on the device that they are using?

1. Are the effects of lighting conditions significant? That is, is there any evidence that suggests that individuals read faster or slower based on the reading lighting conditions?

1. Do device type and lighting conditions *interact*? For example, Suppose that, on average, people can read for longer on device A than on device B, in low light. Is that trend the same in medium light, or bright light? If not, for example, if B is better than A in bright light, then type and lighting interact.

Through this entire analysis, let's set $\alpha = 0.05$.

First, let's read in the data, and store the appropriate variables as factors.

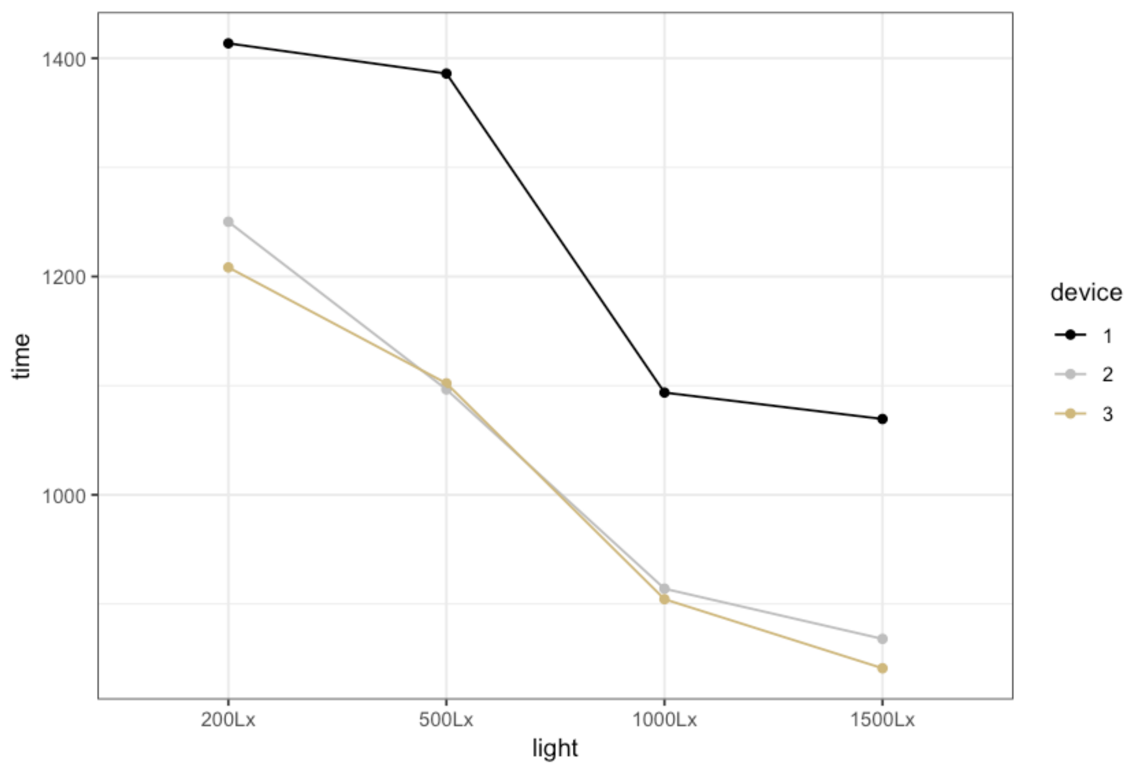
In [13]:
```
library(tidyverse)
#read = read.table("https://stat.ufl.edu/~winner/data/ereader1.da
t", sep = "")
library(RCurl) #a package that includes the function getURL(), whic
h allows for reading data from github.
library(ggplot2) #a package for nice plots!
library(dplyr)

#getURL is a nice way of reading in data from the web
#data from https://dasl.datadescription.com/datafile/graduate-earni
ngs/
url = getURL(paste0("https://raw.githubusercontent.com/bzaharatos/-
Statistical-Modeling-for-Data-Science-Applications/master/ANOVA%20a
nd%20Experimental%20Design/Datasets/ereader.txt"))
#stores the data in the dataframe amazon
read = read.csv(text = url, sep = "\t")

names(read) = c("device", "light", "time")
read$device = as_factor(read$device)
read$light = as.factor(read$light)
read$light = recode(read$light, "1" = "200Lx", "2" = "500Lx", "3" =
"1000Lx", "4" = "1500Lx")


summary(read)
```

```
 device    light          time       
 1:19   200Lx :14   Min.   : 543.8  
 2:20   500Lx :15   1st Qu.: 861.4  
 3:20   1000Lx:15   Median :1105.4  
        1500Lx:15   Mean   :1090.2  
                    3rd Qu.:1300.0  
                    Max.   :1797.2  
```

1.(a) Construct interaction plots, and visually assess and comment on whether interactions are present.

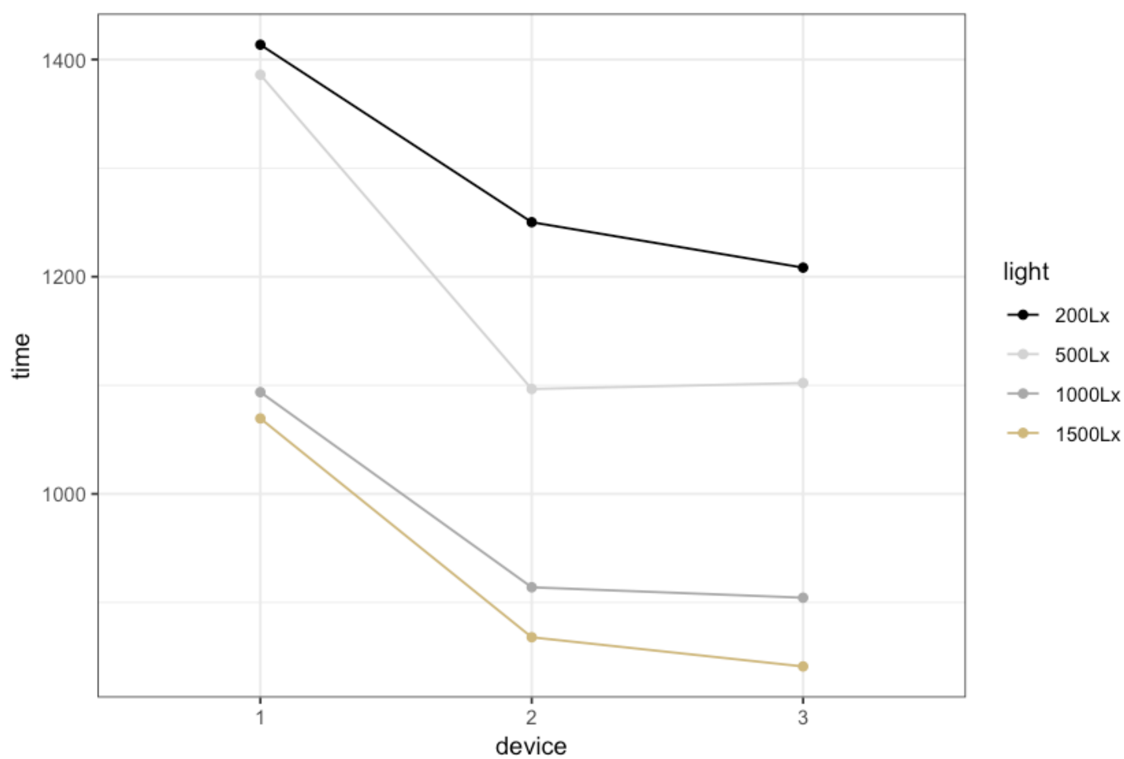
In [14]:

```
read %>%
  ggplot() +
  aes(x = light, color = device, group = device, y = time) +
  stat_summary(fun.y = mean, geom = "point") +
  stat_summary(fun.y = mean, geom = "line") +
  scale_color_manual(values=c("black", "grey","#CFB87C")) +
  theme_bw() +
  coord_fixed(ratio=1/200)
```



In this plot, we see a few pieces of visual evidence that there are interactions, but that those interactions may or may not be statistically significant. For example, we see that, at 200Lx, reading time is longer for device two than device three, but that for 500Lx, reading time is longer for device three than two. However, the difference between the mean reading time for device types two and three seem small, and so it may just be an artifact of the sample data, and not a reflection of the population.

In [15]:
```
read %>%
  ggplot() +
  aes(x = device, color = light, group = light, y = time) +
  stat_summary(fun.y = mean, geom = "point") +
  stat_summary(fun.y = mean, geom = "line") +
  scale_color_manual(values=c("black", "lightgrey", "darkgrey","#CF
B87C")) +
  theme_bw() +
  coord_fixed(ratio=1/250)
```



This plot does not provide evidence of interactions between factors.

1.(b) Now, let's formally test for an interaction. Fit a model with an interaction, and one without, and conduct an F-test. State the appropriate decision for the test.

In [16]:
```
twoway_interact = lm(time ~ device + light + device:light, data = r
ead)
twoway = lm(time ~ device + light, data = read)
anova(twoway, twoway_interact)
```

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
53	3628970	NA	NA	NA	NA
47	3603108	6	25861.55	0.05622427	0.9992146

Here, the p-value, 0.99, is larger than α, and thus, we see no evidence of an interaction. So, we've answered our **research question #3**! Thus, we can proceed with model without interaction term.

In [17]:
```
summary(twoway)
```

```
Call:
lm(formula = time ~ device + light, data = read)

Residuals:
   Min     1Q Median     3Q    Max 
-500.0 -194.6  -24.8  204.9  460.5 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1438.25      87.22  16.489  < 2e-16 ***
device2      -209.73      83.89  -2.500 0.015547 *  
device3      -227.93      83.89  -2.717 0.008879 ** 
light500Lx    -97.46      97.30  -1.002 0.321052    
light1000Lx  -321.66      97.30  -3.306 0.001704 ** 
light1500Lx  -366.16      97.30  -3.763 0.000421 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 261.7 on 53 degrees of freedom
Multiple R-squared:  0.3455,	Adjusted R-squared:  0.2838 
F-statistic: 5.596 on 5 and 53 DF,  p-value: 0.0003268
```

1.(c) Before we interpret this model with respect to research question #1 above (just below the data description), let's decide whether the differences that the model reports are statistically significant.

Investigate this question using Bonferroni post hoc comparisons. That is, conduct all pairwise post hoc comparisons for device type using a Bonferroni correction and an overall type I error rate of $\alpha = 0.05$. Comment on the results.

```
In [18]: library(lsmeans)
         pairs(lsmeans(twoway, "device"), adjust = "bonferroni")
```

```
 contrast estimate   SE df t.ratio p.value
 1 - 2       209.7 83.9 53 2.500   0.0466
 1 - 3       227.9 83.9 53 2.717   0.0266
 2 - 3        18.2 82.7 53 0.220   1.0000

Results are averaged over the levels of: light
P value adjustment: bonferroni method for 3 tests
```

Note that the p-values given in this table have been adjusted, so we can still use the threshold $\alpha = 0.05$. We see that, adjusting for `light` , the differences between device 1 and device 2, and device 1 and device 3 are statistically significant. The difference between device 2 and device 3 is not statistically significant.

1.(d) Using the post hoc comparisons from above, let's focus on research question #1 from above: Are the effects of device type significant? That is, is there any evidence that suggests that individuals read faster or slower based on the device that they are using

Recall that the device type variable included three different types:

```
1. Sony PRS-700 with a 6-in. display, 800 × 600 resolution;
2. Amazon Kindle DX with a 9.7-in. display, 1200 × 824 resolution; and
3. iRex 1000S with a 10.2-in. display, 1024 × 1280 resolution.
```

1. The difference with respect to mean reading time between the Sony PRS 700 and the Amazon Kindle is statistically significant, even after the Bonferroni correction ($p \approx 0.03$). So, we can conclude that there is a difference between these two devices, and that, on average, holding lighting conditions constant, one can read for approximately 210 seconds longer with the Sony device. Whether or not this amount of time is practically significant will depend on the views of the stakeholders involved (e.g., the researchers, technology companies, and readers).
2. The difference with respect to tmean reading time between the Sony PRS 700 and the iRex 1000S is statistically significant ($p = 0.047$). So, we can conclude that there is a difference between these two devices, and that, on average, holding lighting conditions constant, one can read for approximately 228 seconds longer with the Sony device. Again, whether or not this amount of time is practically significant will depend on the views of the stakeholders involved (e.g., the researchers, technology companies, and readers).
3. The difference with respect to mean reading time between the Amazon Kindle and the iRex 1000S is not statistically significant. So, we cannot conclude that mean reading time is better for one of these devices.