Stat 141XP – Statistics homework three
Professor Esfandiari
Week four – Session Two - Question two will be part of statistics homework three

Question one.

Refer to ChatGPT and type the following question and discuss the answer with your neighbor.

Can I run logistic regression with random factors. Please give me the R codes.

Question two

Open the report on Dr Rootman's project with rare events, go to the part on random intercept models, and go through the output for random intercept model and the plot of odds they created and discuss their findings.

Question three. In a study with Jules Stein, we wanted to examine the effect of light on the perception of beauty by individuals from two different countries (India and USA). Six different lights were used including:

1) 90° overhead box light,
2) ring light,
3) 45° superior box light,
4) built-in camera flash,
5) 2 straight on box lights, each 45° from midline, and
6) natural light. Participants were instructed to maintain a neutral

Participants were instructed to maintain a neutral expression, were placed in front of a standardized blue-gray background, and were photographed during a single session. Photographs were imported into an online survey platform (Qualtrics 2020) and displayed in random order. Volunteer survey respondents were instructed to rate the subject's attractiveness on a scale of 0 to 10.

This data set is given to you as lighting.csv in the summary four homework folder on week five.

Model one:

1. Treating country and light as fixed factors, conduct multiple linear regression **"lm";** using average rating as outcome with Country and light as fixed predictors.

```
> m1<-lm(average_rating~Country+light)
> summary(m1)
```

2. Interpret the findings within context.
3. Create the boxplot of average ratings by light and country. Interpret the resulting boxplots.
4. What is MSE for m1?
5. Draw the plot of means for average rating of beauty for different by country and light; using the following commands
   library(car)

library(effects)
library(lattice)
>plot(allEffects(m1),ask=FALSE)

`

Model two – m2

1. Treating country and light as fixed factors and rater as random effect, conduct **"lmer"** with library **"lme4"** ; using average rating as outcome with Country and light as fixed predictors and "Rater" as random factor
2. Compare the outputs resulting from models one and two and comment on similarities and differences.
3. Create the relevant intercepts and explain what the output shows.
4. Do you recommend using model one or mode two and why?

Model three

1. Treating country as fixed factor and rater and light as random effects, conduct **"lmer"** with library **"lme4"** .
2. Compare the outputs resulting from models three and two and comment on similarities and differences.
3. Create the relevant intercepts and explain why and how they differ from intercepts create for model two.
4. Under what condition do you recommend model two and under what conditions do you recommend model three.

Model four – m4

1. Run a linear model with an interaction between country (treating it as between subject factor and light as within subject factor. Then create an interaction plot showing the mean of rating under the six lights for two countries.

   Use the following commands to create the plot:

   Library(car)
   Library(effects)
   Library(lattice)
   Plot(allEffects(m4), aks=FALSE)
   Interpret the output and the plot

Model five – m5

1. Run an lmer model with an interaction between country and light and using "Rater" as random effect. Then create the interaction plot.

Question three. Create a fictitious scenario with two fixed effects, one repeated factor, and the nesting of subjects within one of the fixed factors. Assume the following:

a) Factor A fixed factor has two levels. What is your factor A (say gender)
b) Factor B fixed factor has three levels. What is your factor B.
c) All levels of factor A cross with factor B.
d) Your repeated factor has four levels. What is your repeated factor.
e) You have a total of 60 subjects participating in the study.
f) Draw a table showing the levels of the fixed and repeated factors as well as number of subjects in each cell of the table. Make fixed factors rows and repeated factor the column.
g) Draw the shape of a few rows of the data that you need to run the analysis with R. As an example look at the lighting data.
h) How many questions does this study help you answer? Please write these questions within context.

Question four.

a) If you were a student who just finished Stat 10, how would you explain confounding factors and whether you have any strategies to deal with them.

b) Now that you are a senior in data science, how would you explain confounding factors (covariates and random factors) and what strategies/statistical methods do you recommend for dealing with them?

Question five

This data was published in a medical journal and below are given the abstract.

The Effect of Lighting and Photograph Exposure on Perceived Attractiveness

Kelsey A. Roelofs, Mahtash Esfandiari, Stefania B. Diniz, Liza M. Cohen, Samuel Baugh, Justin N. Karlin, Robert A. Goldberg, Daniel B. Rootman

Purpose: To assess the effect of various lighting conditions and photograph exposures on perceived attractiveness. Methods: In the first experiment, 5 variably exposed photographs were taken of 10 subjects using a consistent lighting condition (45° superior box light). In the second experiment, 10 subjects were photographed under variable lighting conditions with consistent exposure: 1) 90° overhead box light, 2) ring light, 3) 45° superior box light, 4) built-in camera flash, 5) 2 straight on box lights, each 45° from midline, and 6) natural light.

Participants were instructed to maintain a neutral expression, were placed in front of a standardized blue-gray background, and were photographed during a single session. Photographs were imported into an online survey platform (Qualtrics 2020) and displayed in random order. Volunteer survey respondents were instructed to rate the subject's attractiveness on a scale of 0 to 10. Between the two experiments, a total of 22,000 scored photographs were included in the analysis.

Statistical analysis: Mixed ANOVA and pairwise comparisons with Bonferroni correction were used to compare between- and within-subject ratings. Results: Lighting condition had a significant impact on perceived attractiveness ($p < 0.001$), with the 90° overhead box light achieving lower scores and the 45° superior box light yielding greater scores of attractiveness relative to the other conditions. Photograph exposure did not have a significant impact on subjective attractiveness ($p = 1.000$).

Conclusions: Our findings suggest that perceived attractiveness is enhanced when a 45° superior box light is used for illumination, and attractiveness is reduced when 90° overhead exposure is utilized. Exposure did not play a prominent role in perceived attractiveness.

Question to be answered:

a) What is the outcome variable
b) What are the fixed effects
c) What are the random effect
d) What are the research questions asked. Write them within context.