

# The “A” of Statistics II (and III)

## Mini Workshop

Bryan Mui - UID 506021334

January 23, 2025

### 0.1 Introduction

We assume you have already viewed a “client interview” (or, client discovery) video featuring Professor Esfandiari and Dr. Rootman.

This assignment is an individual submission which counts for two individual assignments. For this part, please read Ruland’s First and Second Questions in the book “Guide for the New Statistical Consultant”.

### 0.2 Your Work

Make sure to include your name on your submission. Your student ID is not necessary.

Please submit answers as PDF output (include the scripts that produced them - Quarto or LaTeX preferred, you can modify the files that produced this document) to the following prompts:

#### 0.2.1 Imagine yourself as Dr. Rootman. Formulate responses to the following three questions.

1. What does one, complete, single observation look like?
  - One, complete observation would have the picture’s flash position and lighting effect, along with the rating of that person’s attractiveness
2. If you measure this again, will you get the same answer?
  - No, because every person’s attractiveness is subjective, so the results of the analysis might be different if the experiment was repeated

3. If we can answer your research questions, what will you do next?

- We would change the way we take photos for before/after to improve desirability for plastic surgery, so people would support our business

## 0.2.2 Analysis

You have been given access to the original Rootman data (it's the Excel file found in the link on the assignment page). You have also been given a modified form of the original data called "lighting\_tall.tsv". Please use any of the formats or your own modified form of the data to produce two results (whatever language you wish to use, e.g., Python)

1. One data visualization

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
v dplyr      1.1.4      v readr      2.1.5
v forcats    1.0.0      v stringr    1.5.1
v ggplot2    3.5.1      v tibble     3.2.1
v lubridate  1.9.4      v tidyr      1.3.1
v purrr      1.0.2
-- Conflicts ----- tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()     masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become
```

```
library(readxl)
```

```
read_xlsx("./data/20201228perceptionOfAttractivenessRawData.xlsx")
```

```
# A tibble: 200 x 77
```

	ResponseId	StartDate	EndDate	Status	IPAddress	Progress
	<chr>	<dtm>	<dtm>	<chr>	<chr>	<dbl>
1	R_2QVUTvHQ~	2020-08-10 20:57:02	2020-08-10 21:02:27	IP Ad~	100.14.1~	100
2	R_sNAOY7Wi~	2020-08-10 14:26:38	2020-08-10 14:29:55	IP Ad~	100.7.25~	100
3	R_3mjH4mWb~	2020-08-10 20:06:48	2020-08-10 20:14:34	IP Ad~	103.104.~	100
4	R_1lsGctLd~	2020-08-10 18:18:03	2020-08-10 18:24:40	IP Ad~	103.110.~	100
5	R_3HztxfUf~	2020-08-11 02:16:52	2020-08-11 02:21:06	IP Ad~	103.114.~	100
6	R_1LcJ6mHu~	2020-08-10 23:13:41	2020-08-10 23:19:53	IP Ad~	103.195.~	100

```

7 R_24AA1Mkw~ 2020-08-10 20:16:42 2020-08-10 20:24:14 IP Ad~ 103.217.~      100
8 R_1n1s10G3~ 2020-08-10 22:48:25 2020-08-10 22:56:11 IP Ad~ 103.40.1~      100
9 R_YRBYV14t~ 2020-08-10 21:28:30 2020-08-10 21:34:29 IP Ad~ 103.59.2~      100
10 R_268Xe6r6~ 2020-08-10 23:37:16 2020-08-10 23:45:27 IP Ad~ 103.93.1~      100
# i 190 more rows
# i 71 more variables: `Duration (in seconds)` <dbl>, Finished <chr>,
#   RecordedDate <dtm>, LocationLatitude <dbl>, LocationLongitude <dbl>,
#   DistributionChannel <chr>, UserLanguage <chr>, Age <chr>, Race <chr>,
#   Race_other <chr>, Sex <chr>, A1 <dbl>, A2 <dbl>, A3 <dbl>, A4 <dbl>,
#   A5 <dbl>, A6 <dbl>, B1 <dbl>, B2 <dbl>, B3 <dbl>, B4 <dbl>, B5 <dbl>,
#   B6 <dbl>, D1 <dbl>, D2 <dbl>, D3 <dbl>, D4 <dbl>, D5 <dbl>, D6 <dbl>, ...

```

```
# Idea, we could create a plot that compares the age, race, sex to the average beauty score
```

2. One table of some sort (could be simple counts could be more complex like model output)

```
# Idea: we could fit a model to see which of these features(age, race sex) actually have an
```

These should fit into the context of the problem that Dr. Rootman was seeking to solve. You could look at the graduate lab example that can be found with the Ruland reading as an encouragement.

3. Please record yourself (like your UCLA Story video) telling us a story about either your visualization OR your table (what is this result communicating to you and please communicate it to your viewer). Not both, choose one. The story might be really short (e.g., “this is a histogram of score”)

- **Video attached with submission\***

### 0.3 What to turn in

1. A single PDF with your name somewhere answering the questions and providing either a visualization or a table. The PDF help you tell the story you want to tell (writing things out will be helpful before having to talk about the result on video).
2. A video recording (.mp4 or .mov work best)
3. The associated code that produced the PDF and the analysis (visualization code or table code). Quarto or  $\text{\LaTeX}$  or RMarkdown is preferred, but it is OK if you used something else, just include it as part of the submission (like a link).
4. This is due before the next meeting when we will see the team’s answer.

## 0.4 Reminders

1. It's the story that is most important, the graphic and the table only lend support.
2. Your PDF does not need to be very long, a page or two is about all you need.
3. Your recording does not need to be very long, less than minute, but no penalties if it runs long.
4. It's OK to turn this in late, just upload it as a comment but remember, once I finish grading, it's a zero because reopening a completed assessment is a headache...