Notebook PREP for Project 01

Contents

Project 1: Measure

Email before Semester Started

The first project is to collect some data from your peers, family or friends, that we can analyze. Specifically, for 10 persons (yourself included), you will measure: height (standing), the height of your head (chin to top of head), the size of your hand (middle finger to wrist), hand span (pinkie to thumb fully extended), the length from your middle finger to your elbow (measured on the inside), the length from your elbow to your armpit, your wingspan (middle finger of each hand fully stretched out), standing reach (flat footed, how far can you reach on a wall with the tip of your finger), length of foot, sitting length from the floor to your knee pit (like armpit but for your knee), standing length from the floor to your hip bone, standing length (height) from the floor to your armpit. {We need to also collect the person's gender, age, and ethnicity. We also need an identifier for each person: e.g., I would suggest something like MonteS passed through an md5 encryption [http://md5.mshaffer.com/] => 1c2408654ef5a2fe1fc962088312266c .} You can choose how to measure exactly, and the units, although mm/cm may be a bit easier. Due to the current "social distancing" concerns, I am giving you a heads up about this data collection as part of your first assignment. You may need to be a creative problem solver to figure out how to get this done. They could email you the results, and I believe everything can be measured by a person in isolation (e.g,

with a 2 meter soft cloth tape [https://www.amazon.com/uxcell-Measure-Metric-Measuring-Tailor/dp/B07TJT6FG4] although some of it may be tricky. Giving you a heads up early is to give you a chance to figure out how you will measure yourself, and how you will measure 9 others (either remotely or in-person practicing "social distancing"). If you have any questions or concerns, please don't hesitate to ask. We live in extraordinary times, and I appreciate your creativity in context of our current circumstances. Anyway, I will have something formal typed up very soon. When the semester begins, you will have about 2 weeks to collect and submit the data, then another two weeks to do some analyses and submit a writeup on your findings (you will have your data plus the data of the other students to perform the analysis).

Project Outline (September 18, 2020)

Objective: To learn about data collection and exploratory multivariate data analysis.

- Handout, 10 points
- Data Collection (XLS and TXT), total 50 points
- Writeup, was 30 points, **NOW** 60 points
- Video, we will NOT be doing a video

No video

The video presentation is not as important as preparing a first writeup.

Writeup

The purpose of the writeup is to demonstrate that you can build a final work product based on your research questions.

The deadline was supposed to be Monday, 2 November by midnight. For a variety of reasons, I have moved the deadline to Wednesday, 4 November by midnight.

SECRET practice We will build a work product where the data stays in a SECRET or private format. It should not be uploaded to GitHub.

Data cleansing I have provided the data in a compiled format. In the notebook `unit_02_confirmatory_data_analysis\n` I have provided some clues on how to cleanse the data. That task is yours.

I consider changing all of the results to one unit system part of data cleansing. You can choose "inches" (in) or "centimeters" (cm) for your analysis depending on your culture and comfort with a given system. This means, all of the data needs to be converted. Please recall the `distance` work we have done, there is a library `measurements` and a function `conv_unit`.

Data collapsing Some people have data for a person's "left" and "right" side of the body. I have prepared code for you to collapse that data so (we assume) each side of the body is equal. This is one option. You can choose to keep the overall data and address NA's (missing values) if your research question is tied to body symmetry.

In the notebook on correlation in week 6, the section `1.2.3.3 Measure` has some code on to "getOne" measurement from the left or right. If one is NA, it returns the other. If they both are available, it returns the mean.

Data creation There may be a few data features you may want to create. I have the "arm span" and information about the "armpits" which would enable you to compute the internal "chest width" (from armpit to armpit). There may be other data you can create in a similar fashion.

Data proportions It is very likely that for each measure row, you would want to create "scaled variables to that person's height", also known as a proportion.

Alternatively, you could scale everything to a person's head height.

Alternatively, you could review lots of different proportions. I suggested at one point that the foot-size and the "upper arm" (elbow-pit to arm-pit) are the same size (some basic Pythagorean theorem could get you there or close).

There are lots of possibilities, all depends on your interests.

- Some say the unit of length of a "one foot" that we now decompose into 12 inches was a function of the actual length of the King's foot in England, and would change when a new King was crowned.
- Another measure of length, the "cubit" is derived from the Latin word for "elbow"
- Galileo Galilei, the famous Italian polymath, literally sold his body parts when he died (quite the entrepreneur). He had extremely long fingers. In the museum in Firenze, they have on display a few of the fingers recovered. Yes, I have seen them https://www.museogalileo.it/it/. Most people miss this museum because they are too busy admiring David's proportions at the nearby Academia Gallery https://en.wikipedia.org/wiki/David_(Michelangelo).

Data selection Which columns are you going to use in your analysis. The "covariates" will be necessary to describe the sample procedure, but for your research question maybe you just use a few of those, or none of those.

The summary statistics on the sample "covariates" and on the overall data are dependent on which columns you want to research. This depends, or is constrained by your research question. For example, in the "Joireman paper" we did collect a lot of other data, and even showed them a variant of an exercise motivator (Nike Ad):

http://www.mshaffer.com/arizona/videos/exercise/010.mp3

http://www.mshaffer.com/arizona/videos/exercise/101.mp4

http://www.mshaffer.com/arizona/videos/exercise/110.mp4

http://www.mshaffer.com/arizona/videos/exercise/111.mp4

Project Research Question

Have you formulated one primary research question and possibly 2-3 subquestions. Or maybe 2-3 primary questions?

The project was initially intended to be an exploration of the original data (distances), the computed proportions (as a function of head height), and its relationship to correlations. However, you now have some experience with basic clustering techniques, so you could try to use them as well. I would say **focus** on the key research question and don't deviate too far afield into clustering techniques that you don't report on "exploratory findings" that inform your research question.

Preparation for Final Writeup

Use this space to include and run code that gets your data and research question prepared.

Alternatively, you can use another notebook. You will submit a final "ZIP" file that contains the supporting documents and the final "PDF" product. I will build the template for the "PDF" product and help with the tasks new to you in that regard.

The data cleanup, the research question, and the analysis is up to you.

I will be supporting the process of turning the results and data you prepare into a "work product".