W200 Python for Data Science: Project 2 Proposal

James Kajdasz, Kyle Eschen

July 24, 2018

Background

Our project proposal aims to explore two findings related to human judgment and decision making by analyzing a large data set of online survey data.

The first finding relates to the "wisdom of the crowd". When trying to answer some question of uncertainty, it's been found that if you take a large number of individual estimates and then average them, the average is often very accurate. This phenomenon was popularized in James Surowiecki's book The Wisdom of Crowds.

A second finding is what's known as the overconfidence bias. People tend to be confident more than they are correct. Such overconfidence often contributes to decision aide neglect (Sieck & Arkes, 2005), making it a pivotal topic for decision scientists who may find themselves data modeling to make decision aides.

We look to demonstrate these principles and perhaps advance the field by examining a data set of individuals attempting to guess whether a person smiling is being genuine or not.

Data Description

In the course of his research, Dr Paul Ekman created videos of people smiling. Sometimes the smile was a genuine Duchenne smile that occurred spontaneously while participants watched video clips of baby animals playing. Sometimes the participants were simply asked to smile. People can test themselves on whether they can detect the difference on the following survey:

https://www.surveymonkey.com/r/SmileRead

In the survey, a few demographic questions are asked and then the participant views 20 videos of people smiling. Data collected includes:

Start Time/ End Time: (date and time survey was started and exited) IP Address: For example: 68.102.240.36 Confidence: Participant asked how many (of 20 videos) will you get correct? Demographic data: Gender, Age, Height, # children Job experience: Years worked in a job where you had to 'read' people

To date almost 100,000 people have taken the survey. The data is available in a .csv file (subset in Attachment 1). We'd like to work with the full data-set of 100k if possible, but if unanticipated problems dealing with "big data" make this problematic, we'd like to work with a subset of several hundred responses.

Data Preparation

General data cleaning

• Eliminate rows where the subject (Ss) did not complete the full survey. (~65% of total)

• Score responses: Recode Ss answer "Fake/Genuine" to "Correct/ Incorrect" variable

Data Preparation for The Wisdom of crowds

• Calculate individual mean accuracy for each of the 20 smile videos for all data.

• Create function that randomly creates 'virtual groups' of x size by sampling from all data.

• Calculate the accuracy rate of each virtual group, and store this performance rate.

Overconfidence Bias

• Calculate the total correct for each individual.

• Compare to the number they thought they would get correct (confidence)

• Split data into comparison groups of interest: men vs. women, young versus old, etc.

Data Analysis

Wisdom of Crowds

Past research demonstrated and replicated the principles of the "Wisdom of the Crowd" with the smile task data. When performance across all smiles was evaluated, individuals achieved 68% accuracy (13.6/20 smiles). When theoretical groups were created randomly, it was found that overall performance increased with group size (Kajdasz, 2014). This analysis was accomplished largely manually, with much less data, utilizing SPSS.

Figure 1 represents the total average performance on all 20 videos with theoretical groups of various sizes. However, this line only represents a single task with, in this case, an average individual performance rate of 68%. For tasks of higher or lower difficulty, one can imagine additional lines fitting on this graph, representing tasks of various difficulty levels, such as in the hypothetical graph in Figure 2.

If one could obtain sample data of this sort for various task difficulties, it should be possible to then model predicted performance using group size and task difficulty. This model could then be used to predict the optimum group size for tasks of various difficulty. We do not plan on creating this regression model during our Python project, but we do hope to process the observed data to make such a future analysis possible. We will do this by treating each smile video as an individual task. Each smile video has its own individual performance rate. Some videos are obvious fakes/obviously genuine. Some are very difficult to determine. We can use each of the 20 videos to represent a task of varying difficulty.

Overconfidence

Before the survey starts, participants are asked how many smile videos out of 20 they think they will get correct. Here are some of the questions we can answer with out anlysis regarding the overconfidence bias:

• Are people overconfident on this task? Compare predicted number correct to actual correct.

• Are men more overconfident than women? Some research shows men tend to be more overconfident (Barber & Odean, 2001).

• Are younger people more overconfident than older people?

• Do subjects who have had a job "reading people" think they'll do better on the smile test? Do they actually do better? Some research has shown experts tend to be overconfident more often than novices (Bradley, 1981).

• Assuming individuals are overconfident, how big does a group have to get before performance matches predicted performance?

Works Cited

Barber, B. M., & Odean, T. (2001). Boys will be boys: Gender, overconfidence, and common stock investment. Quarterly Journal of Economics, 261-292.

Bradley, J. V. (1981). Overconfidence in ignorant experts. Bulletin of the Psychonomic Society, 17(2), 82-84.

Kajdasz, J. E. (2014). A demonstration of the benefits of aggregation in an analytic task. International Journal of Intelligence and Counterintelligence, 27(4), 752-763.

Surowiecki, J. (2005). The Wisdom of Crowds. New York: Anchor Books.

Figures

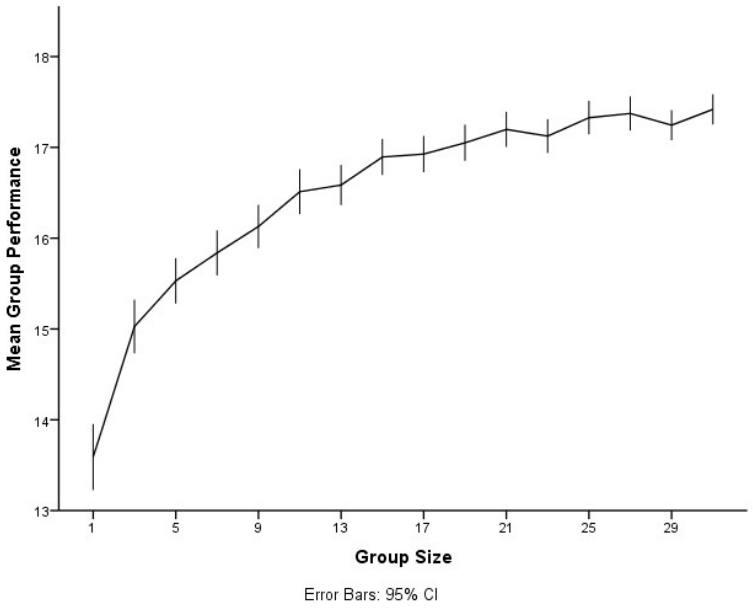

Figure 1: from Kajdasz(2014) Performance for groups of various sizes on SmileVideo Task

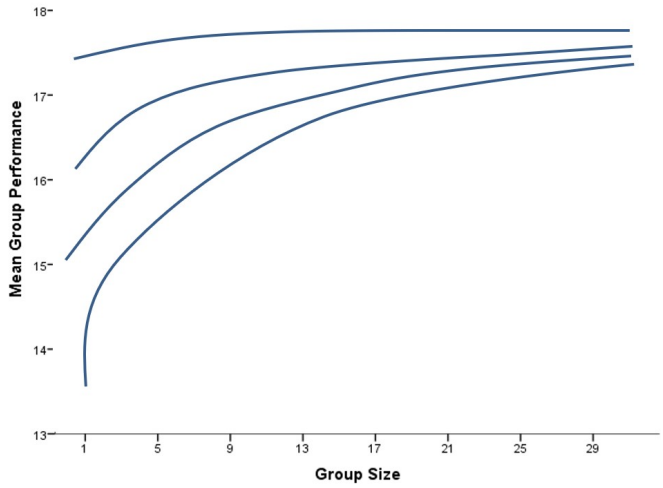

Figure 2: Hypothetical group performance for tasks of varying difficulty