Perspectives on Computational Research: Term Paper Re-write

Laurence Warner

lpwarner@uchicago.edu

Research Proposal

Topic: Demographics and Smartphone Usage

Background

Smartphones are a technological device that has rapidly changed society in a unique way.

It is projected that by 2022, 90% of all mobile subscriptions will be for Internet-enabled "smartphones," which are already in the majority (Ericsson, 2016). In the United States, more time is spent on digital activity on smartphones than on computers (ComScore, 2015).

However, the distribution of smartphones across society is not uniform. I propose to investigate this.

Avenue 1: Static Analysis

Key Dataset

Pew Research Center's Internet Project Core Trends Survey, 2018. It can be accessed freely online here: http://www.pewinternet.org/datasets/

Interviews with a nationally representative sample of 2,002 adults were conducted between January 3-10 2018. The target population for the study is non-institutionalized persons age 18 and over, living in the US. Most of the interviews were conducted using cellphones (n = 1,502) with the remainder conducted using landlines (n=500); both groups were included in the final sample. According to the Pew Research Center, the landline sample was collected using a proportional sample based on listed telephone households. The cellphone sample was selected systematically from dedicated wireless numbers. This dataset contains questions about social media use in 2018 and attitudes toward the internet and whether Americans think

it's good or bad for society. Random digit dialing was used to collect survey responses and the final sample was weighted to represent the American adult population. The sample response rate was 11%.

Hypothesis 1

I propose that demographic variables greatly determine smartphone usage. The demographic variables that I will consider are: gender, age, education, race, income.

Hypothesis 1a

In particular, I posit that age and income are the primary determinants.

Literature

The primary theoretical and empirical inspiration is Tsetsi et al. (2017): "Smartphone Internet access and use: Extending the digital divide and usage gap".

They introduce the concept of the usage gap. This is an adaptation of the sociological theory of the knowledge gap, which posits that that knowledge, like other forms of wealth, is differentially distributed throughout a social system. So it states that technology usage is also distrubted differentially.

They also use the Pew Internet survey - but from 2012. So I will be interested to compare results to see how demographic influences have changed in this fast-changing technological landscape.

Marler (2018): "Mobile phones and inequality: Findings, trends, and future directions" finds that socioeconomic effects are prevalent. But interestingly, low socioeconomic groups are often prone to having smartphones, but no other internet-devices - leading them to be "smartphone dependent". It would be interesting to compare the income elasticity of smartphones to other consumer products.

Andone et al. (2016) "How age and gender affect smartphone usage" use an alternative dataset: their own app "Menthal" to analyze smartphone usage. This data is not available - but I feel that survey data is superior because it represents the entire nation. Non-smartphone users are not captured by their smartphone app.

Methods

I have done some initial EDA in the following Jupyter notebook: 'proposal.ipynb'. Here is a heatmap of the variables:
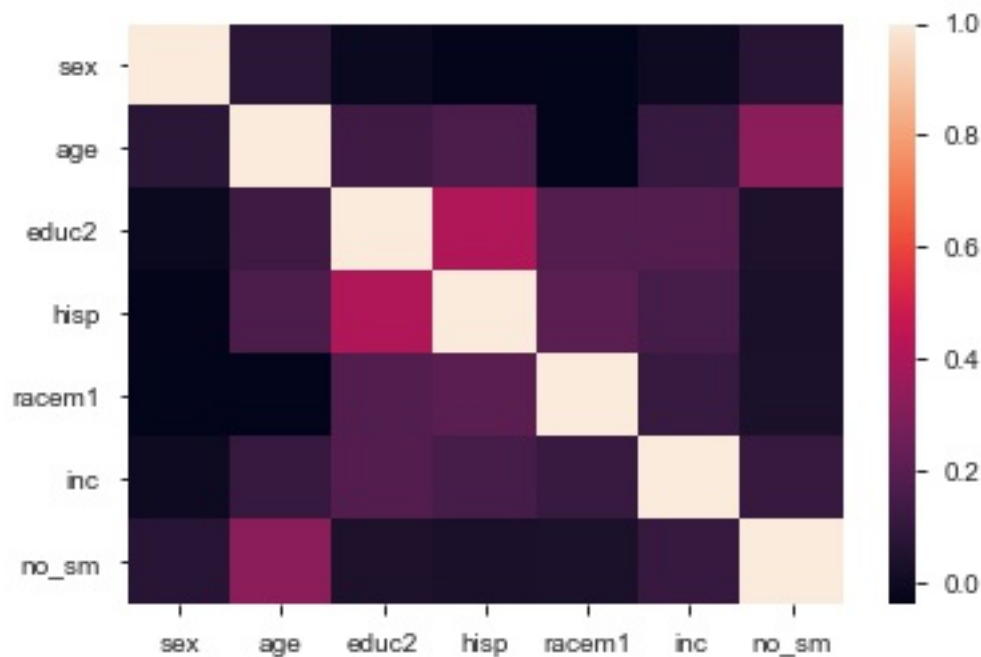
As can be seen, age is the most highly correlated, whilst income level is second most highly correlated.

I will further investigate the relationships using logistic regression models.

Hypothesis 2

I propose that we can use a machine learning model, for example decision trees, to predict whether an individual owns a smartphone based upon demographic variables with a high degree of precision.

The use of these sophisticated methods, and treating the problem as a prediction problem, will constitute a novel approach in the literature.

Avenue 2: Longitudinal analysis of trends

Key Dataset

I propose to use previous iterations of Pew Research Center's Internet Project Core Trends Survey. Unfortunately, they don't appear to be currently available online. I have in contact with Pew to try to obtain them. If this is not possible, I will focus on avenue 1, which contains ample material for interesting research.

Hypothesis 1

Related to Ericsson's claim that 90% of phones will be smartphones by 2022, I will make my own predictions based upon trends from the latest data.

Hypothesis 1a

I could also analyze how smartphone uptake has changed among specific demographic groups over time.

Literature

Andersen (2016): "Internet use and Problematic Internet Use: A systematic review of longitudinal research trends in adolescence and emergent adulthood" gives some data on internet usage amongst young adults specifically. It finds that internet usage is particularly high among young adults. It will be interesting to see if this result pans out for smartphone usage specifically.

Methods

I hope to fit trend lines to the time series data, using different curves to find the least squares fit.