Administrative

• Office Hours
  • Shifted to 4-5pm Eastern today due to a conflict

• Homework 1 graded
  • Generally well done
  • *Read the questions carefully;* most people lost points for not fully answering the question, not because they didn’t understand how to do it
  • Grades, feedback on Blackboard

• Homework 2
  • Due next Monday, October 3, at 12 noon Eastern

• Project
  • Proposal revisions due 11:59pm Eastern *on Blackboard* today for those who were emailed feedback
  • Deliverables I due on October 31
# Homework 1 Statistics

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Average</strong></td>
<td>97</td>
</tr>
<tr>
<td><strong>Median</strong></td>
<td>100</td>
</tr>
<tr>
<td><strong>Standard Deviation</strong></td>
<td>5</td>
</tr>
<tr>
<td><strong>Max</strong></td>
<td>100</td>
</tr>
<tr>
<td><strong>Min</strong></td>
<td>82</td>
</tr>
</tbody>
</table>
Recitation on AdFisher

September 30, 2016
Motivation for building AdFisher

Web Browsing → Ad Ecosystem → Ads

Control → Experimental

Different Outcome?

Slide Credit: Amit Datta
How AdFisher works

Simulate users having a particular interest by visiting webpages associated with that interest.

Visit a webpage, collects ads, settings provided to the simulated users.

Automatically analyzes the data to determine whether statistically significant differences between groups of users with different interests exist.

Github: https://github.com/tadatitam/info-flow-experiments
Running experiments on AdFisher to determine if behavioral targeting occurs on the web

- **Input:**
  - Behaviors the control / experimental groups are to perform (e.g., visiting websites)
  - Measurements (e.g. ads) to collect afterwards

- **AdFisher:**
  - Takes input runs experiments collect measurements
  - Determines if there is a statistically significant difference between the data collected from the control and experimental groups
Running AdFisher

Technical Details
Installation & Running

• Installation
  • Instructions on Github
  • Tested on MacOS / Linux only
  • Windows users may want to use a VM to run AdFisher

• Running
  • Scripts to run experiments and collect data in AdFisher/examples
Modifying AdFisher

Technical Details
AdFisher/core

- **AdFisher/core/adfisher.py**
  - Runs the entire experiment
- **AdFisher/core/driver**
  - Stages the experiment and sets up the browser units that collects data
  - Stores measurements in a log file
- **AdFisher/core/converter**
  - Reads data written in the log file
  - Converts data to machine readable format for statistical analysis
- **AdFisher/core/web**
  - Codes that collect data (such as ads, news articles) from various websites
  - *Groups that want to collect data from various websites*
  - Familiarity with Selenium needed
- **AdFisher/core/analysis**
  - Determine if the difference between data collected from the control group and data collected from the experimental group is statistically significant
  - *Groups that want to implement different forms of statistical analysis*
Selenium Python

• Web Scraping tool that simulates a user surfing the Internet

• Limitations:
  • Works only on HTML elements
  • Does not work on XML, alert boxes, chrome://settings, firefox preferences page, etc

http://selenium-python.readthedocs.io/
Automated experiments and Site Policies

Technical Details
Creating Multiple Accounts

Exercise judgment when sending email to a large number of recipients, even if the recipients elected to receive emails from you in the past. When Gmail users mark emails as spam, it increases the likelihood that future messages you send will also be classified as spam by our anti-abuse systems.

The Creation and Use of Multiple Gmail Accounts

Don't create or use multiple accounts to abuse Google policies, bypass Gmail account limitations, circumvent filters, or otherwise subvert restrictions placed on your account. (For example, if you've been blocked by another user or have had your Gmail account disabled due to abuse, don't create a replacement account that engages in similar activity.)

You're also not allowed to create Gmail accounts by automated means or buy, sell, trade, or re-sell Gmail accounts to others.
Creating Multiple Accounts

2. Obligations

Here are some promises you make to us in this Agreement:

You're eligible to enter into this Agreement and you are at least our “Minimum Age.”

2.1. Service Eligibility

To use the Services, you agree that: (1) you must be the “Minimum Age” (defined below) or older; (2) you will only have one LinkedIn account (and/or one SlideShare or Pulse account, if applicable), which must be in your real name; and (3) you are not already restricted by LinkedIn from using the Services.

“Minimum Age” means (a) 18 years old for the People's Republic of China, (b) 16 years old for the Netherlands, (c) 14 years old for the United States, Canada, Germany, Spain, Australia and South Korea, and (d) 13 years old for all other countries. However, if law requires that you must be older in order for LinkedIn to lawfully provide the Services to you (including the collection, storage and use of your information) then the Minimum Age is such older age. The Services are not for use by anyone under the age of 13.
Number of requests a website expects
#Requests / unit time
#requests / unit time

http://blog.hubspot.com/marketing/reduce-http-requests#sm.0001t43gs9739cxeq3916bv5w0cmy