



STA 3011 Workshop on Data Analysis and Statistical Computing

March, 2018



■ Instructor:

Fan, Xiaodan Room 107 (3943 7930)

Email: xfan@cuhk.edu.hk

SONG, Xinyuan Room 114 (39437929)

Email: xy-song@sta.cuhk.edu.hk

■ Teaching Assistant:

Kang, Kai Room G32 (39438535)

Email: 1155070327@link.cuhk.edu.hk

Project 2:

- Two datasets: 1. China dataset and India dataset, and a summary of the data sets and meaning of the variables are provided; 2. ADNIMERGE and a website for data dictionary search.
- You can use various statistical methods to analyze the data, draw conclusions and report results. The statistical methods include, but are not limited to
 1. Exploratory data analysis
 2. Regression (simple or multiple regression)
 3. Multiple group analysis
 4. Generalized linear model/generalized linear mixed effects model
 5. Mixture model
 6. Missing data problem
 7. Various computer algorithms



Project Description

Main objectives:

1. Identify interesting problems.
2. Explore possible findings from the data.
3. Write a technical report to present and illustrate your findings. The report should be well organized and written in English.
4. Give a clear presentation of your report in English.



Requirements

- You are asked to work in groups to study the given dataset. Each group should give a presentation and write a report on its findings both in English. Each member of a group should participate in the project and presentation.
- Each team has 14 minutes for presentation and 1 minute for QA.
- The report should not exceed 15 pages including all tables and figures.

Academic Honesty and Plagiarism

- Attention is drawn to University policy and regulations on honesty in academic work, and to the disciplinary guidelines and procedures applicable to breaches of such policy and regulations. Details may be found at the CUHK [academic honesty website](#).
- References
 - (1) Cite any resource you use
 - (2) Pick a citation style and use it consistently.
 - (3) Include every resource you cite in a bibliography at the end of your report




Report Structure

- Introduction (A brief introduction of the topic)
- Method (What statistical methods?)
- Results (Tables, graphs, etc.)
- Limitations (Possible extensions?)
- Conclusions (Interpretation?)
- References (A list of the papers, books, websites)
- Appendix (Technical details, codes, questionnaires)



Assessment scheme

- Methodology 30%
creativity, innovation, proper use of statistical methods
- Report 30%
organization, completeness, critical thinking
- Presentation 40%
clearness, time management, presentation

- 
- To assess individual contribution within each group, each individual MUST e-mail to the TA his/her own assessment on the percentage of contribution of each team member on or before Apr 24th , 2018.

Submission: Send both PPT and Report to

1155070327@link.cuhk.edu.hk by 23:59, April 23 2017.

- Must be submitted to Veriguide with a report of similarity.



■ Presentation time list (project 2):

Group 1-11 : 17/04/2017

Session A (group 1-6): 16:30 – 18:15

Session B (group 7-11): 19:00 – 20:15

Venue: Lai Chan Pui Ngong LT




Data Background 1

This data collection is designed to enable cross-national comparison of values and opinions in a wide variety of areas and to monitor changes in values and attitudes of mass publics in 45 societies around the world.

Broad topics covered are

- work,
- the meaning and purpose of life,
- family life,
- contemporary social issues.



Respondents were asked to rate the importance of

- (1) work,
- (2) family,
- (3) friends,
- (4) leisure time,
- (5) politics,
- (6) religion in their lives.

They were also asked

- (1) how satisfied they were with their present lives,
- (2) whether they tended to persuade others close to them,
- (3) whether they discussed political matters,
- (4) how they viewed society.

Questionnaire

V91 Depressed or very unhappy Yes 1 No 2

V96 All things considered, how satisfied are you with your life as a whole these days?

1 2 3 4 5 6 7 8 9 10

Dissatisfied

Satisfied

V116 Overall, how satisfied or dissatisfied are you with your job?

1 2 3 4 5 6 7 8 9 10

Dissatisfied

Satisfied

V132 How satisfied are you with the financial situation of your household?

1 2 3 4 5 6 7 8 9 10

Dissatisfied

Satisfied

V146 Were you brought up religiously at home? Yes 1 No 2

V296 Claiming government benefits which you are not entitled to

1 2 3 4 5 6 7 8 9 10

Never

Always

V297 Avoiding a fare on public transport

1 2 3 4 5 6 7 8 9 10

Never

Always

V298 Cheating on tax if you have the change

1 2 3 4 5 6 7 8 9 10

Never

Always

V299 Buying something you knew was stolen

1 2 3 4 5 6 7 8 9 10

Never

Always

V300 Taking and driving away a car belonging to someone else joyriding

1 2 3 4 5 6 7 8 9 10

Never

Always

V301 Taking the drug marijuana or hashish

1 2 3 4 5 6 7 8 9 10

Never

Always

V302 Keeping money that you have found

1 2 3 4 5 6 7 8 9 10

Never

Always

V303 Lying in your own interest

1 2 3 4 5 6 7 8 9 10

Never

Always

V304 Married men/women having an affair

1 2 3 4 5 6 7 8 9 10

Never

Always

V305 Sex under the legal age of consent

1 2 3 4 5 6 7 8 9 10

Never

Always

V306 Someone accepting a bribe in the course of their duties

1 2 3 4 5 6 7 8 9 10

Never

Always

V307 Homosexuality

1 2 3 4 5 6 7 8 9 10

Never

Always

V308 Prostitution

1 2 3 4 5 6 7 8 9 10

Never

Always

V309 Abortion

1 2 3 4 5 6 7 8 9 10

Never

Always

V310 Divorce

1 2 3 4 5 6 7 8 9 10

Never

Always

V311 Fighting with the police

1 2 3 4 5 6 7 8 9 10

Never

Always

V312 Euthanasia terminating the life of the incurably sick

1 2 3 4 5 6 7 8 9 10

Never

Always

V313 Suicide

1 2 3 4 5 6 7 8 9 10

Never

Always

V314 Failing to report damage you've done accidentally to a parked vehicle

1 2 3 4 5 6 7 8 9 10

Never

Always

V315 Threatening workers who refuse to join a strike

1 2 3 4 5 6 7 8 9 10

Never

Always

V316 Killing in self-defence

1 2 3 4 5 6 7 8 9 10

Never

Always

V317 Political assassinations

1 2 3 4 5 6 7 8 9 10

Never

Always

V318 Throwing away litter in a public place

1 2 3 4 5 6 7 8 9 10

Never

Always

V319 Driving under the influence of alcohol

1 2 3 4 5 6 7 8 9 10

Never

Always

V353 Sex of respondents: 1 Male 2 Female

V355 Age



Data Background 2

The Alzheimer's Disease Neuroimaging Initiative (ADNI) unites researchers with study data as they work to define the progression of Alzheimer's disease.

Data from the North American ADNI's study participants, including Alzheimer's disease patients, mild cognitive impairment subjects and elderly controls, are available from this website: <http://adni.loni.usc.edu/>




ADNI researchers collect, validate and utilize data such as

- magnetic resonance imaging (MRI)
- genetics
- cognitive tests
- cerebrospinal fluid (CSF)

The ADNI dataset also provides tabular information about ADNI participants, such as

- gender, race, ethnicity categories
- age, mean age for scans
- education
- diagnostic categories: Alzheimer's disease (AD), mild cognitive impairment (MCI), cognitive normal (Norma)
- apolipoprotein E- ϵ 4 (APOE4)



All the available variables are included in the dataset
ADNIMERGE.csv. For more information, a website for data
dictionary search is also given here:

<http://adni.loni.usc.edu/data-dictionary-search/>