SOCI 502 Research Design and Techniques (Quantitative)

Course Description

This course focuses on first principles for analyzing quantitative survey data. It addresses epistemological and theoretical orientations, causality, research designs, descriptive statistics, statistical inference, latent constructs and missing data. A practical outcome of the course is a research proposal based upon multivariate analysis of secondary survey data that can be implemented in SOCI 514 Analyzing Quantitative Data in Sociology.

Instructor

Gerry Veenstra, ANSO 1321, (604) 822-4351, gerry.veenstra@ubc.ca

Textbook

Agresti, Alan. 2018. *Statistical Methods for the Social Sciences. Fifth Edition*. Prentice Hall

Evaluation

Assignments (20%)
Four assignments based upon analyzing survey data

Video presentation (10%)
A narrated slide-based presentation of the research proposal

Research proposal (35%)
Proposal for a project of journal article length that comprises secondary analysis of survey data

Participation (5%)
Based on attendance with one percentage point deducted for each missed class

Final exam (30%)
Drawn from the supplementary Agresti exercises

Statistical Software

We will use the statistical software packages Stata and jamovi. Stata is available in the ANSO Computing Lab (Room 150), Buchanan B101, B121, B123, B125 and B126 and Koerner Library Room 218A and can be purchased at www.stata.com. jamovi is open-source software which can be downloaded for free at www.jamovi.org. We will apply Stata and jamovi to two datasets throughout the semester: a dataset comprised of characteristics of 1,500 adult Canadians (Statistics Canada's General Social Survey on Time Use) and a dataset comprised of characteristics of 154 countries.

Canvas

The following materials are available in Canvas (www.canvas.ubc.ca):

- Lecture slides
- GSS and Country datasets
- Descriptions of variables in the GSS and Country datasets
- Agresti exercises

Schedule

Topic 1: Introduction (Sep 9)

- *Topics*: Introduction to the course; tutorial on Stata and jamovi
- *Supplementary videos*: Crash Course Statistics Preview, #1, #2

Topic 2: Epistemology and causality (Sep 16)

- *Topics*: Epistemological standpoints; conditions for causality; research designs (experimental, longitudinal, cross-sectional); multivariate causal relationships
- *Readings*: Michalski, J.H. 2016. The epistemological diversity of Canadian sociology. *Canadian Journal of Sociology* 41, 4, 525-556. McCall, L. 2005. The complexity of intersectionality. *Signs* 30, 3, 1771-1800.

Topic 3: Variables and sampling (Sep 23)

- *Topics*: Descriptive and inferential statistics; populations and samples; parameters and statistics; variables; levels of measurement; sampling error; types of bias; sampling strategies; survey weights
- *Readings*: Agresti Chapters 1 & 2
- *Supplementary exercises*: Agresti 1.1-1.5, 1.8-1.9, 1.14-1.17, 2.1-2.10, 2.12-2.14, 2.16-2.21, 2.25-2.40

Topic 4: Descriptive statistics (summarizing variables) (Sep 30)

- *Topics*: Frequency tables; pie charts and bar charts; central tendency (mean, median, mode); dispersion (range, standard deviation, interquartile range); shape (histograms, stemplots, boxplots); outliers; recoding and transforming variables
- *Readings*: Agresti 3.1-3.4, 3.6
- *Supplementary videos*: Crash Course Statistics #3, #4, #5, #6
- *Supplementary exercises*: Agresti 3.3, 3.10, 3.13, 3.15, 3.17, 3.18, 3.21, 3.23, 3.24, 3.26, 3.29-3.32, 3.36-3.38, 3.40, 3.42, 3.45, 3.69, 3.70, 3.75

Topic 5: Descriptive statistics (bivariate associations) (Oct 7)

- *Topics*: Cross-tabulations; Cramer's V; Kendall's tau-b; scatterplots; Pearson's *r*; OLS regression; Spearman's rho; comparing central tendencies, dispersions and shapes; introduction to multiple regression
- *Readings*: Agresti 3.5
- *Supplementary video*: Crash Course Statistics #8
- *Supplementary exercises*: Agresti 3.47-3.50

No class Oct 14 (Thanksgiving)

Topic 6: Probability distributions and confidence intervals (Oct 21)

- *Topics*: Rules of probability; random variables; probability distributions for discrete and continuous variables; normal distributions; z-scores; sampling distributions; standard errors; Central Limit Theorem; confidence intervals for means and proportions
- *Readings*: Agresti Chapter 4, 5.1, 5.2
- *Supplementary video*: Crash Course Statistics #20
- *Supplementary exercises*: Agresti 4.1, 4.2, 4.4, 4.23, 4.30, 4.32, 4.37, 4.43, 4.50, 5.1, 5.4-5.10, 5.12, 5.15, 5.17, 5.18, 5.20

Topic 7: Tests of significance (Oct 28)

- *Topics*: Null and alternative hypotheses; test statistics; p-values; alpha levels; Type I and Type II errors; Chi-squared and oneway ANOVA tests of significance; statistical significance versus practical significance; multiplicity
- *Readings*: Agresti 6.1, 6.4, 6.5. Simmons, J.P., Nelson, L.D. & Simonsohn, U. 2011. False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science* 22, 11, 1359-1366.
- *Supplementary video*: Crash Course Statistics #21
- *Supplementary exercises*: Agresti 6.26, 6.44, 6.45, 6.47-6.51, 6.59, 6.60

Topic 8: Scales and indices (Nov 4)

- *Topics*: Latent factors; indices and scales; Cronbach's alpha
- *Readings*: Neuman, W.L. & Robson, K. 2015. *Basics of Social Research: Qualitative and Quantitative Approaches*. Third Canadian Edition. Toronto: Pearson, pp. 104-132.

No class Nov 11 (Remembrance Day)

Topic 9: Missing data (Nov 18)

- *Topics*: Types of missingness; strategies for dealing with missing data

Topic 10: Computer lab (Nov 25)

Assignments

Assignment 1: Summarizing variables
Select a categorical variable from the GSS dataset (you choose the variable). What is the variable measuring and what is its level of measurement? Describe and summarize the distribution of values for the variable. Can you provide any insights regarding the distributions of values for this variable? Feel free to copy charts and graphs from your statistical software package into your assignment. Now repeat the process with a quantitative variable of your choice. Present your analyses and insights in sentence and paragraph form accompanied by graphs and/or tables. Due Oct 7 in class.

Assignment 2: Bivariate associations
Investigate the bivariate association between two categorical variables in the GSS dataset (a different pair of variables for each student).

1. Theoretical rationale: How would you expect the two variables to be related (if at all)? Does it make sense to designate one as independent and the other as dependent?
2. Examining each variable individually: What do the variables measure or assess? What are their levels of measurement? What do their distributions look like in this dataset?
3. Examining the relationship between the variables: Describe and summarize the relationship between the two variables. You may have to recode or transform a variable to make the relationship more intelligible.
4. Interpretation: Provide some interpretive insights regarding the relationship between the variables. Were your theoretical expectations met?

Now repeat the process for a relationship between two quantitative variables and again for a relationship between a categorical variable and a quantitative variable (you choose the variables). Present your analyses and insights in sentence and paragraph form accompanied by graphs and/or tables. Due Oct 21 in class.

Assignment 3: Scales and indices
Consider the variables TCS_Q130, TCS_Q140, TCS_Q150, TCS_Q160, TCS_Q170, TCS_Q180, TCS_Q190 and TCS_Q200 in the GSS dataset. Do they form an internally coherent scale? What does an index or scale produced from these variables look like in this dataset? Prepare a short report documenting your investigation into whether and how to produce an index or scale from these variables. Due Nov 18 in class.

Assignment 4: Missing data
Consider the variable incmhsd (household income) in the GSS dataset. How much missing data is there? What can you conclude about the nature of the missingness? Prepare a short report documenting the results of your investigation into missing data for this variable. Due Nov 25 in class.

Video Presentation

The video presentation of the research proposal should be approximately 15 minutes in length. Ensure that the storyline is clearly told (especially the research question, operationalization of key concepts and proposed line of analytical inquiry) and issues germane to your project are addressed (such as creating indices or dealing with missing data). Prepare attractive and professional slides for the presentation and record your narration of the slides (this can be easily done in PowerPoint). The video presentation is due @ 9:00 am on Monday, Dec 2 (submit by email or in Canvas).

Research Proposal

The research proposal should be written for a general audience. The text of the proposal should be between 15 and 25 double-spaced pages in length (Times New Roman 12-point font) excluding tables, references and appendices. The proposal should address the following issues:

Introduction: State the objectives of the proposed research and give a brief overview of the topic. Provide an overview of theories and concepts that are relevant to the topic and research problem and (briefly) review past research.

Research problem: Clearly describe the main research problem(s).

Research design and unit of analysis: Identify the research design of the study (experimental, longitudinal, cross-sectional). Identify the unit of analysis for the study (individuals, businesses, neighbourhoods, etc.).

Sampling: Describe the population being studied and the details of the sampling strategy. Briefly describe the data collection technique. What is the response rate, if applicable? What is the size of the working sample?

Operationalization of variables: Identify the key variables for addressing the research problem. How are they coded? Describe composite variables (indices or scales) if applicable. Describe control variables if applicable. Describe the distributions of these variables in the dataset.

Missing data: Describe the nature and extent of missing data in the survey sample. Outline a plan for analytically dealing with these missing data.

Proposed analysis: Describe your plans for analyzing the data (e.g., specified sequences of regression models).

Limitations: Describe some of the more important limitations of the proposed research. Suggest how future research might potentially overcome these limitations.

The research proposal is due @ 9:00 am on Thurs Dec 19 (by email is fine).