CRI 2120H: Data Analysis
Class hours: Tuesdays 10 a.m. to Noon, Winter 2020
Anthony N. Doob

Part of the challenge of social science is to provide descriptions of events, relationships, etc., that both capture the “meaning” of the matter being described but are also efficient and effective. This challenge exists in what is typically referred to as ‘qualitative’ research as well as in ‘quantitative’ research. Whether your task is to characterize the issues that occur in debates in Parliament about criminal justice policy or to describe ordinary Canadians’ victimization experiences, you are necessarily going to be simplifying (summarizing) your observations.

Inevitably, in any description, details are going to be left out. The challenge in both ‘qualitative’ and ‘quantitative’ research is to provide the descriptive information in a manner that allows readers to assess the evidence and come to their own conclusions about the meaning (and meaningfulness) of the observations.

Much of this course, therefore, will be focused on how best to describe criminological phenomena. This course will, however, focus on quantitative data. Although much of the challenge of social science is in conceptualizing and describing the social world, the course will also expose you to the issue of drawing general inferences from the data that you have. I hope that the course will help give you a better understanding of the meaning of ‘statistics’ as they are presented in published criminological research. In addition, however, the course is designed to give you the skills that are necessary to carry out basic and intermediate quantitative analysis of data.

There will, therefore, be two inter-related parts of this course. In the first place, I will help you acquire basic skills in interpreting and handling of data. The course will use what I believe to be the data analysis software most often used in the social sciences: SPSS. Copies have been installed on a few of the Centre’s computers. We will discuss various ways in which data can be described and, working with actual data, we will carry out various inferential tests.

I will not be assuming any prior statistical knowledge, nor will the course assume that you have any special background in mathematics. This is a course in ‘data analysis’ not a statistics course per se. Hence we will discuss what statistical tests do, and what assumptions are important, but I will not be teaching you ‘statistics’ per se. For those of you who have a background in statistics at the undergraduate level, some of the basic concepts will be
familiar. Those of you with some statistical background may find this course to be useful, however, because it will give you an opportunity to apply your knowledge to an actual research project.

In addition, an important part of the course will involve a project where you carry out a study involving the analyzing and interpreting of criminological data. For those of you who are interested in carrying out original research, this will provide you with an opportunity to do so, using real data. For those of you who are interested in criminology, but do not anticipate doing original research as part of your careers, it will give you an opportunity to see, first hand, how research decisions are made. Hence it will help you understand the limits of any research findings – not only in criminology, but in other social sciences as well.

* * * * * * * * * * * * * *

No special ‘statistics’ text is required for the course. However, you may find it useful to supplement the work in class with an account in a textbook. Statistics textbooks present material in quite different – and usually equally useful – manners. Hence if you have a text that you are comfortable with, or, in looking over a number of texts, you find one that seems compatible with what you are interested in, use that one. If you want to look at a ‘statistics’ book, see me. On the other hand, my own experience is that a Google search is likely to be helpful in finding a description that explains, in a manner that is understandable to you, what the issues are. For example, when I was looking for an understandable description of “multicollinearity” (something you probably won’t have to deal with in this course), I easily found this site: http://www.thejuliagroup.com/blog/?p=1405. It answered my questions quite adequately in a manner I could understand. And the author has a sense of humour.

Similarly, although there may be books that teach you SPSS, my experience is that if you don’t know how to do something, you should either use as a starting point something we did in class, or simply Google the problem you are addressing.

SPSS has two approaches to working with the program: using the “pull down menus” and working from syntax (where you write your own commands). I would strongly urge you to work largely from syntax. Aside from anything else, it gives you a record of exactly what you’ve done. But the pull-down menus are also very useful. You will see that I use them all the time typically because I haven’t bothered to learn the exact form of a command. But I usually try to remember to copy that syntax that SPSS creates into my own syntax file so I know what I’ve done. You will see, in the syntax files that I use, that there is almost always a mixture of commands that I wrote and commands that I copied from the SPSS output into my syntax file.

Much of what you will be doing in the course will be focused on your research project. In class we will be using the public access data set from the Statistics Canada General Social Survey on Victimization (GSS – 28) carried out in 2014 (the most recent victimization survey released as of late 2019). The “public use” data were released in August 2016.

You can, if you wish, use some other large data set that you can get access to. Check with me.
first to ensure that it has the characteristics that are necessary for this course. The most important disadvantage in some other data set is that I will not be able to assist you as much as I can with a data set that I, myself, know. The idea, however, is that you work with ‘raw’ data – individual observations rather than aggregate data. The purpose of using a larger data set is that it allows certain kinds of analyses and manipulations that would make little sense with smaller data sets. An advantage, for those of you not used to working with data, of using GSS-28 is that I will be using this data set in class and you can build on those analyses for your own assignments and project.

The research question that forms the basis of your project should be similar to a research question that forms the basis of an article in a criminological journal. In other words, you should ask a question that is of broad criminological interest. For example, you could try to understand the relationship between certain kinds of victimization experiences and people’s views of the criminal justice system. You should plan on working with the public use data set. Some data are not available except in a secure Statistics Canada site (with restrictions related to privacy on what can be released).

Your papers will follow the broad format of a journal article in criminology, though the emphasis should be on your analysis and presentation of data as well as the interpretation of what your findings mean. I would expect that the introduction will, in most cases be relatively short.

You can work together in pairs on the homework assignments and on your final paper if you like. If you do, it makes sense to hand in a single assignment and paper. There is no need to be consistent on this: you can work in pairs on only some of the assignments and/or the final paper.

The School of Graduate Studies asks (requires?) instructors to remind you of the following: "Plagiarism and other academic offences will not be tolerated at the University of Toronto. Academic discipline ranges from a mark of zero on an exam or assignment to dismissal from the University. For important information see UofT’s Academic Integrity website at http://academicintegrity.utoronto.ca/. See also these academic integrity websites: School of Graduate Studies at http://www.sgs.utoronto.ca/facultyandstaff/Pages/Academic-Integrity.aspx and Faculty of Arts and Science at http://www.artsci.utoronto.ca/osai".

Evaluation:

**Five homework assignments.** There is one required homework assignment – due on January 14. In addition, there will be six subsequent opportunities to do homework assignments. Of these six, your top 4 marks will count. I would urge you to do all six assignments since working with data is the best way to learn the skills. Homework must be handed in at the beginning of class when it is due, or must be received by the University of Toronto’s server (if you are not coming to class and/or want to hand in your assignment by email) by 10:15 a.m. on the day it is due. Late homework assignments will not be accepted. Each counts 4%. (20% total).
Research paper: 40%

Tests: Total - 40%. There is an optional ‘practice’ test on 11 February which will count 15% or 0%. The required test on 31 March will count 25% or 40%. The weighting (adding to a total of 40%) will be whichever weighting system works best for you (Counting the practice test 15% if you do better on it than on the final test; counting the practice test 0% if you do worse on it than on the final test). I would, therefore, encourage everyone to take the optional practice test since you cannot be hurt by it.

Tentative class schedule (subject to change, with notice if possible, but possibly without notice). Note that all or most of these classes will involve being taught how to use SPSS.

7 January: Descriptive and inferential statistics – an introduction. Getting to know the data you are working with. Understanding different kinds of measures. An overview of the General Social Survey.

14 January: [Required homework assignment due] Describing data (tables, graphs, etc.)

21 January: Normal curves and hypothesis testing. Samples and populations. Type 1 and Type 2 errors. [Homework due]

28 January: Contingency tables and the Chi Square. Basic ‘controlling for third variables.’

4 February: Working with subsets of the data. Creating new variables. [Homework due]

11 February: Practice test

[18 February: Reading week]

25 February: Regression and Correlation.

3 March: Creating scales (weighting of variables; reliability). Testing the difference between two means. The Logic of analysis of variance [Homework due]

10 March: Factorial designs and the logic of interactions [Homework due].

17 March: An introduction to multiple regression [Homework due]

24 March: Review – identifying the ‘best’ way of presenting findings and ‘testing’ hypotheses concerning associations. [Homework due]

31 March: Final test

Research report due: 5 p.m. Monday 20 April 2020. Penalty for late papers: One third of a letter grade for each two days (or part thereof) that the paper is late (For example, the mark will be reduced from an A- to a B+ if the paper is up to two days late; it will be reduced from A- to B if the paper is more than 2 but less than 4 days late, etc.).
compute personweight = wght_per/889.372298.
weight by personweight.

Frequencies variables = TIP_10  TIP_25.
crosstabs tables = tip_10 by tip_25/cells=count,row/statistics=chisquare.
compute trust = tip_10 + tip_25.
variable labels trust 'Trust in neighbours and strangers, high = high trust'.
fre var = trust.

compute trust2 = trust.
var lab trust2 'Trust in neighbours and strangers'.
recode trust2 (2 thru 5= 1)(6=2)(7=3)(8=4)(9,10=5).
value labels trust2 1 'Low' 2 'Moderate' 3 'Moderate-high' 4 'High' 5 'Very High'.
crosstabs tables = trust by trust2.
fre var = trust2.

compute policeconfidence = cip_10.
compute courtconfidence = pcc_145.
recode policeconfidence, courtconfidence (1,2=1)(3,4=2).
var lab policeconfidence 'Confidence in the police'/ courtconfidence 'Confidence in the criminal courts'.
val lab policeconfidence, courtconfidence 1 'A great deal or some' 2 'Not very much or none'.
cro tab = cip_10 by policeconfidence/pcc_145 by courtconfidence.

cro tab = policeconfidence by courtconfidence/cells=count,row/statistics=chisquare.
compute confpc=0.
var lab confpc 'Confidence in police and courts'.
val lab confpc 1 'Little confidence' 2 'Moderate' 3 'High'.
if (policeconfidence = 2 and courtconfidence = 2) confpc=1.
if ((policeconfidence =1 and courtconfidence = 2) or (policeconfidence =2 and courtconfidence = 1) ) confPc=2.
if (policeconfidence = 1 and courtconfidence = 1) confpc = 3.
fre var = confpc.

missing value confpc (0).
cro tab = policeconfidence, courtconfidence by confpc.

compute confpc2= 0.
Variable label confpc2 'Confidence in police and courts'.
count confpc2 =policeconfidence, courtconfidence (1).
fre confpc, confpc2.

compute confmissing = 0.
if (missing(Policeconfidence)) confmissing = confmissing +1.
if (missing (courtconfidence)) confmissing = confmissing + 1.
var lab confmissing 'Number of missings for confidence ratings'.
fre var = confmissing.
if (confmissing ge 1) confpc2 = 9.
miss val confpc2 (9).
fre confpc, confpc2.

******************************************************************************
cro tab = trust2 by confpc/cel=cou,row/stat=chi.
fre var = totvic.
cro tab = totvic by trust2, confpc/cel=cou,row/stat=chi.
cro tab = trust2 by confpc by totvic/cel=cou,row/stat=chi.
******************************************************************************
** What if you are interested only in women in urban areas in Ontario?.
******************************************************************************
fre var = prv, sex, luc_rst.

USE ALL.
COMPUTE filter_$(SEX=2 and prv = 35 and luc_rst =1).
VARIABLE LABELS filter_$(SEX=2 and prv = 35 and urban =1 (FILTER).
VALUE LABELS filter_$(0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$(
EXECUTE.

fre var = prv, sex, luc_rst.
cro tab = trust2 by confpc/cel=cou,row/stat=chi.

Use all.
fre var = prv, sex, luc_rst.
fre var= trust.

T-TEST GROUPS=sex (1,2)
/MISSING=ANALYSIS
/VARIABLES= trust
/CRITERIA=CI(.95).
fre var = cip_10, pcc_145.

T-TEST PAIRS=cip_10 WITH PCC_145 (PAIRED)
/CRITERIA=CI(.9500)

ONEWAY TRUST BY TOTVIC
/STATISTICS DESCRIPTIVES
/MISSING ANALYSIS.
compute ANYVICTIM = TOTVIC.
recode anyvictim (1 thru 3 = 1).
var lab anyvictim 'Has the respondent been victimized (non-spousal) in the previous year?'.
val lab anyvictim 0 'No' 1 'Yes'.
fre var = anyvictim.
fre var = amb_01.
USE ALL.
COMPUTE filter_$=(AMB_01=1).
VARIABLE LABELS filter_ $ 'AMB_01=1 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_ $ (f1.0).
FILTER BY filter_ $.
EXECUTE.
fre amb_01.
fre var = pcc_145, cip_10.
UNIANOVA pcc_145 BY SEX ANYVICTIM
   /METHOD=SSTYPE(3)
   /INTERCEPT=INCLUDE
   /EMMEANS=TABLES(OVERALL)
   /EMMEANS=TABLES(SEX)
   /EMMEANS=TABLES(ANYVICTIM)
   /EMMEANS=TABLES(SEX*ANYVICTIM)
   /PRINT DESCRIPTIVE
   /CRITERIA=ALPHA(.05)
   /DESIGN=SEX ANYVICTIM SEX*ANYVICTIM.
REGRESSION
   /MISSING LISTWISE
   /STATISTICS COEFF OUTS R ANOVA CHANGE
   /CRITERIA=PIN(.05) POUT(.10)
   /NOORIGIN
   /DEPENDENT pcc_145
   /METHOD=ENTER sex
   /METHOD=ENTER anyvictim.