Spring 2014-Specific Notes to Graders for SOS3003

You should be made aware that in Spring, 2014, the textbooks and statistical software taught in SOS3003 at NTNU were changed completely for the first time in approximately 12 – 15 years. The labs for students were based completely on STATA, while none of the students had prior experience with any software other than SPSS; the instructor had only begun to learn to use STATA three months earlier; and, most of the students needed to work with STATA through the computer labs and remote terminal services, as NTNU does not have as inexpensive a group license as it does for SPSS. Needless to say, this caused me a very great amount of concern. The class of spring 2014 did not have the wealth of examples from previous cohorts, SPSS cookbooks, SPSS on their notebooks, or instructors with quite so much experience with the software as earlier cohorts.

Despite this, it seems to me that the transition went much better than I expected. My impression is that the downsides of this transition were not nearly as bad as I had feared; and, the presumed advantages of this transition – 1) that the most quantitative students could pass on to more advanced courses without an introduction to STATA and 2) that being able to do conditional effect plots with confidence intervals easily would help students tackle more complicated models with interactions and curvilinear effects – are serious advantages of this transition.

Nevertheless, I would like to make the following suggestions to graders for this term: 1) Please take into consideration that this cohort had disadvantages relative to earlier cohorts in terms of the issues I mentioned above; 2) Despite the fact that the students have been told to include both unstandardized and standardized coefficients for their “final/best” OLS model/s, and both the logistic regression coefficients and exponential/odds ratio/multiplicative coefficients in logistic regression, I anticipate that more students are likely to forget including both these statistics for either OLS or LR, given that STATA does not produce both without separate commands, as does SPSS (so, perhaps not doing this should not be taken quite so seriously as earlier. Needless to say, many published articles do not include both.); and 3) Although I have provided the students with model syntax statements for the kinds of graphs required, it is difficult for students this term to produce graphs with a high degree of quality without having more instruction and buying an additional book on graphics with STATA, so I would suggest being lenient on details of the production of figures, particularly with regard to graphs displaying influential outliers.

All of the strict requirements remain the same as for earlier years, as described in my rather lengthy instructions to the students. However, there are a couple of areas where my lectures and advising have extended my recommendations to students. First, with STATA, doing conditional effects plots is much easier than with syntax statements with SPSS. For this reason, I have allowed/encouraged students to try more complicated models with more interaction effects and non-linear effects than earlier, stressing the use of conditional effect plots to help illustrate the combined effects of combinations of such effects. So, the inclusion of conditional effect plots is still not required, but has been strongly recommended. I would be happy to hear from you later if this has worked very well.

Second, over the years I have encouraged students to do two-country comparisons, and more recently, two-period comparisons, so that they would be more likely to have to deal with significant interaction effects. My advice in analysis strategy has been to produce matching models for both countries/periods separately, but to include a model for the merged data for both countries/years so that they would have estimates with standard errors and tests of significance for the pair-wise
differences which were most obvious. So, such students should have at least one test of differences based on the merged data. (There were a couple of cases where the cross-country differences were so great that I thought it was OK to present parallel models and interpretations.)

Finally, there is an issue where many student wish to include subjective attitude questions as determinants of other subjective attitude questions, when the causal directions and relationships are very open to debate. In these situations, my advice has been as follows: 1) Be very clear you recognize the issue, 2) Make at least a brief argument in favor of the causal relationships you assume; and 3) Show the model coefficients both for a reduced model based on the obvious demographic determinants and a model including the more debatable attitudinal determinants.
FORMAT AND REQUIREMENTS FOR THE SEMESTER PAPER

The main requirements for this paper have not substantially changed for a number of years, regardless of changes in instructors. However, I have added comments and emphasis regarding aspects of the papers which have influenced the graders for the papers over the earlier terms. Advice on various sections of the paper is much more detailed than in other descriptions of the paper. I have commented on these same points in class lectures. And, I will remind the censors of these points in my grading instructions for the exam.

Note also: There is a distinction between the paper requirements, and my advice about what improves a paper. The list of analyses which must be included and the paper length considerations are requirements. Descriptions of sections, section length, and format are recommendations.

The basic model for this paper is that it should have the same form as a serious research article for an academic journal, under length constraints insisted upon by the editor. The course paper is different from that format in one respect, and that is that the course paper includes a much longer appendix documenting additional data descriptions, analyses, and checks for problems than would be included in a published journal article.

The analysis project upon which the paper is based: The papers are to be based on regression analysis, using either OLS or logistic regression. The analyses should study the effects of a set of independent variables on a single dependent variable. (Some students may study two to four dependent variables within the context of a structural equation model; but they should concentrate their detailed analyses relevant to various requirements listed below on a single dependent variable.) The data set and dependent variable used by each student must be approved by the instructor. Every student in a given term must choose a unique combination of dependent variable(s) and data set. Most students are encouraged to choose variables from the European Social Survey (ESS), because this provides many possible variables, is easily available for the class, and is familiar to the instructors. However, with permission, they may use other data sets instead. The data do need to be based on approximately random, representative samples of cases, of a reasonable size for the use of the regression statistics taught.

If a student took SOS3003 earlier, turned in a paper for that course and either did not complete the course, or received an unsatisfactory grade, they may use the same paper (hopefully improved) as their paper for this term; and they have a priority in using the variable used previously. To get a final grade for the course this term, they also must take the school exam again in addition to turning in a course paper which will be graded with a completely new evaluation.

Structure of the paper

The paper must be 20 pages in length, 1.5 or 2.0 spacing, font of 11 or 12 pts. These 20 pages include the abstract/summary, the bibliography/list of references, and tables and figures included in the text. This limit on page length has been part of the formal published course requirement for some time. I am instructing graders to take it seriously! They are under no obligation to grade on the basis of pages beyond 20, and are instructed to count exceeding page length as negatively as a
paper not being complete enough! A journal editor would not publish your paper at all if you exceeded his/her instructions regarding length.

The paper should contain the following sections:

1. An abstract/summary of 100 – 200 words, summarizing very briefly the data source and the main findings. (This section should be formatted separately, with spacing of 1.0, and looks nice in italics and/or a smaller font, so as to be about 1/3 – 1/2 page.)
2. An introduction of 1/3 - 1 page stating the research question and describing why the research question is interesting and/or valuable.
3. A short discussion of theory relevant to the research question (1 – 3 pages)
4. A short summary/mention of previous research relevant to the question (1 – 2 pages).
5. A description of main hypotheses (1 – 2) pages.
6. A description of the data set, the dependent variable (in some detail) and the independent variables (in much less detail)(1-2 pages).
7. A description of the analysis results based on the basic beginning model, tests for more complicated effects, eliminated variables, and the final model (7 – 13 pages).
8. A short conclusions section, summarizing the most important findings. This should be about one page, longer than the conclusions statements in the abstract, which should be only 1-3 sentences (in the abstract).
9. A bibliography/list of references (1/2-1 page).
10. An appendix, which includes tables, figures (graphs), and a little explanatory text to describe aspects of the data relevant to variable transformations, scale construction, tests for additional variable transformations, tests for violations of statistical assumptions, and the examination of influential outliers. (There are no formal page requirements for the appendix, but in most cases, this should not be longer than about 5 – 15 pages, as appropriate to your analyses.)

**Formal analysis requirements (The results of these will not necessarily be in the text of the paper. A number of them should only be mentioned in the text and documented in the appendix.)**

1. Students should test for the significance of at least one “interaction effect” and interpret its meaning. If no interactions prove statistically significant, you should still describe at least one hypothesized interaction in theoretical terms, and state the test and results which showed it to be insignificant.
2. You should test for the effects of a set of dummy variables representing a theoretical variable involving at least three categories.
3. You should test for non-linear effects by trying one or test for such effects.
4. You should show and interpret at least one “conditional effects plot.”
5. You should carry out, and report the results from, tests for whether the data conform to the assumptions required for the model, and tests for influential outliers.
6. You should end up testing more than one model.
7. Of course, for all models, you need to report and correctly interpret the important and relevant tests of significance, parameters, coefficients of determination, and sometimes confidence intervals.
Additional comments on various sections of the paper

Country-specific discussion and theory:

Students are encouraged to study countries other than Norway, even though they may be less familiar with these countries and previous research in these countries. The main purpose of this course is to learn to use and understand methods of statistical analysis, not to develop deep understandings of other countries. Accordingly, the graders for the course papers will be instructed to give little or no weight to country-specific knowledge, but to grade based on the correctness, complexity, and creativity of the data analyses and description of results.

The theory and previous research section: Everybody asks how important it is and how long it should be. Different instructors may have different opinions on this, and sensors (graders) can also. I think most of us think that this paper is very demanding, and it is really tough to write a good theory section when you forced to take an available variable in competition with all the other students. For this reason, I and I think most of us, do not want to put undue emphasis on the theory section, and it should not be more than 1 – 3 pages. Nevertheless, we are pushed to grade papers in a competitive model, and there is no question that everything else being equal, the quality of the theory section has an influence, especially related to the highest grades. Note that the “theory” does not have to be based mostly on literature and can and should include your own ideas and hypotheses.

The same things can be said about the discussion of previous research. It does not have the highest importance, but is still influential. For those using the latest ESS data, it is a little early for there to be a huge mass of published articles available, but there is a lot, especially if you can find conference papers, and much of it is relevant to many papers published based on the World Values Surveys, the International Social Survey Program (ISSP), and other similar projects. If you can find such papers, they will lead you to theory.

Your hypotheses: Much of the statistical literature approaches issues as if you only had one hypothesis. This is a model for hard (physical and medical – “real-fag”) science research, which is not always so relevant for social science research these days, when we can tear through hypotheses in minutes of analysis based on secondary analysis. You should be guided first by your theoretical hypotheses, while you will often test additional hypotheses quickly, as your data teach you and lead you on. No one will want to read a null and alternative hypothesis for all possible relationships. So, what you should do is to settle on one to four main beginning hypotheses to describe before the analysis section. These may be described in terms of very specific null and alternative hypotheses. If you discover new relationships during analysis, you can describe them there. There are likely to be a number of variables which you think have possible or likely effects on your dependent variable, but are not part of your main theoretical hypotheses. You can label these “control variables,” and comment on the reason for including them, and their likely effects, as briefly as possible.

Tables and figures to include in the text: This depends partly upon how much space you have within the 20 pages, but the test is whether these would be included in the text of a professional article. Such articles often include a table describing all of the variables included in the various models, including their number of valid cases, maximum and minimum values, means, and standard deviations. Generally, you should include such a table, though it can be omitted if you have more important things for the text, but in such cases it should be in the appendix. In general, you do not
need to present figures (graphs) for the univariate distributions of the independent variables, even in the appendix. You might present such a figure or figures for the dependent variable in the appendix, if you think that its distribution was problematic, or involved a transformation, and needed some discussion of how to deal with it.

In the text you definitely should include tables showing the usual statistics for coefficients, and tests of significance for coefficients for your most important models. How many such models are represented in tables in the text depends on your topic, data, and analyses, but usually there should be 2 – 4 such models. Sometimes the coefficients for more than one model can be included in a single table. Other times, separate tables may be best.

The statistics for tests of significance of overall models, contrasts between models, and coefficients of determination or their analogues, can be reported in the text, or as part of these larger tables (they do not require tables of their own). Conditional effect plots might be included in either the text or the appendix. Graphs showing non-linear effects might also be in either the text or the appendix. The results of your analyses testing assumptions, looking for influential cases, testing scaling, or looking for non-linearities should definitely be stated in your text (with references to where these might be included in the appendix), but the actual graphs and tables of statistics for these should usually be in the appendix.

What belongs in the appendix? The appendix is the place for graphs and tables of statistics related to tests of whether the assumptions underlying the models are met, and whether influential cases of outliers are problematic. The appendix also may include statistics for additional models which were tested, and led to the development of the final model. It might also include small matrices or tables related to scale construction. You should number or label your graphs and figures, so that you can refer to them clearly by page and label in the text of your paper. You definitely should NOT thoughtlessly include masses of SPSS output on univariate variable distributions, large bivariate correlation matrices, etc.

Mistakes you should definitely avoid: You definitely should not include passages of general explanations of various statistics in an abstract way unrelated to your specific research analyses. There should not be any sections which are simple restatements or summaries of textbook explanations of statistics. You should show that you understand the statistics by the way you use them, and the way you interpret them relative to your specific research questions and results.

Take great care in your assumptions about the direction of causality between variables, especially subjective attitude variables measured at the same time. If such assumptions are implicit in your model, they should be explicitly stated and justified in your discussion. If such assumptions are questionable, you should mention this. Such considerations might lead you to try different models, including and excluding the relationships with debatable causal relations.

Be sure to describe each of your variables very clearly - the exact meanings of your dummy categories and the omitted reference category, any variable transformations, the wording of variable questions for attitude items, what high versus low values mean, and the items included in any scale. If this takes space, some of these details can be placed in the appendix.
Features of the very best quality papers: The analysis requirements listed above (such as test at least one interaction effect, one set of dummy variables for a variable with at least three categories, on non-linear term, etc.) are minimum requirements for passing grades. Graders are likely to be impressed with more extensive analyses in addition to this minimum, especially if they are related to theory and/or yield substantive findings. Special attention to technical details of data exploration is important, as is a good discussion of the relation between theoretical ideas and your data and analyses. Unusually ambitious choices of data or analysis techniques are appreciated; but, you should be very careful not to “get in over your head,” with analysis problems which might take time and space, and possibly lead you into mistakes, which would end up hurting the final overall paper. It is very difficult in an exercise with these kinds of constraints, but signs of creativity in variable choice and formation, hypotheses, and analyses, are appreciated indeed.

You will receive only one overall grade for the course. Both the paper and the school exam are given equal weight in the graders’ evaluation of the overall grade.
EXAMINATION FOR SOS3003: APPLIED SOCIAL STATISTICS
SPRING 2014

ENGLISH VERSION WITH GRADER’S INSTRUCTIONS

General comments to graders:

1. While approximate weights have been given for the importance of the three sections of the examination, and weights for the individual answers to specific questions should be related to the number of questions in each section, the overall evaluation of the examination, and the weight given the examination in determining the overall grade for the student is intended to take into account the grader’s evaluation of the difficulty of individual questions, and the overall impression of the student’s knowledge and performance. Particularly good answers to difficult questions or particularly poor answers to very easy questions may contribute more to the overall evaluation than indicated by the weight of the section and the number of questions per section.

2. For the relatively long answers expected for some questions in sections II and III, there is room for different correct and articulate arguments and judgments for answers for the same question. I have suggested one way of answering the questions, and the issues I intended students to notice; but grader’s may see and acknowledge additional and different conclusions which are also reasonable, insightful, and articulate.

3. It should be taken into account that the students were not allowed to bring any books, notes, or lists of concepts and formulas into the exam.
4. Of course, it is possible that I have made errors in these instructions or answers. If so, please inform me by e-mail.

5. Finally, students have been instructed that since the course lectures, many labs, and one of the main pensum books were in English, they may use English terms and expressions for statistics and statistical concepts in English, even if their answers and sentences are written in Norwegian.

Academic contact during the examination: Albert Andrew Simkus

Telephone: 99 53 21 74

Examination date and time: 04 May 2014 09:00 – 15:00

Examination duration: 6 hours

Study points: 15

Number of pages in English, excluding appendix/attachments 9 – including this page.

Number of pages in bokmål, excluding cover page and appendix/attachments: 7
Number of pages in nynorsk, excluding cover page and appendix/attachments: 7

Number of pages in appended attachments: 8

Internal Grading date: 22. June 2014

**Materials allowed to be brought into the examination:**

A hand calculator, which should have functions for natural logarithms and anti-logarithms, and may have higher functions, but may not have internet connectivity or memory storage for large amounts of text.

*No books or notes of any kind are allowed for the examination! What formulas as may be needed are provided on the page of general instructions.*
ENGLISH

General Instructions: For each set of questions, the approximate weight given the answers for each section toward the total grade, and the approximate length of good answers for that section are stated before each set of questions. These are guidelines so that you can tell approximately how long an answer is expected for each part, and you should write neither too much nor too little.

The quality of the answers is much more important than the length of the answers. The grading of the exam will be based on the overall quality of the examination, and particularly good answers for some questions may or may not balance out particularly bad answers on others. Weight will also be given to the consideration that – regardless of question length - some questions are intentionally very easy, while others are intentionally difficult. Of course, the purpose of this is to allow graders the best possible test of the range of your knowledge on these topics.

If Norwegian is your native language, it is still wise to also read the English-language version of the questions as well, to be extra certain you understand the question details clearly. The Norwegian version of questions will be sufficient for doing well on the exam; but, given that the lectures and pensum have been in English, it is a good idea to read the questions in their English versions as well as Norwegian versions, so as to help relating them to your lectures and readings.

You must write clearly! Clearly label and number the question you are answering, and keep your answer pages in the order of the questions!

WARNING! Be extremely careful how you divide your time between questions! It would be very easy to spend way too much time on Section II, questions 2 & 3, and Section III, question 2, and run out of time to answer the questions requiring much shorter answers!

Reference formulas which may be required:

\[ Y^* = b_0 + b_1 + b_2 + b_3 + b_4 + b_5 \ldots \]

\[ L^* = b_0 + b_1 + b_2 + b_3 + b_4 + b_5 \ldots \]

\[ L = \ln(O) \]

\[ O = e^L \]
\[ P = \frac{1}{1 + e^{-1}} \]

Section I. Knowledge and understanding of statistical terms. (40%) Be as precise and concise as possible.

1. What are the names of three kinds of univariate “measures of central tendency” or “averages,” and what do they indicate?

The mean, based on the sum of the scores of a variable divided by the number of cases. The median, the middle score when the scores are placed in an array from lowest to highest (the exact middle score when there is an odd number of scores, halfway between the two middle scores if there is an even number of scores). The mode, the score or category of scores including the most cases.

2. What are the names of three kinds of univariate “measures of variation” and what do they indicate.

The range, indicating the difference between the lowest and highest scores. The variance, the mean squared deviation of scores from the mean of the scores (for a population). For a sample, the sum of the squared deviations of the scores from the overall mean score, divided by the number of scores, minus 1. The standard deviation – the square root of the variance. Students may also use the interquartile range, the difference between the value for the first quartile and that for the third quartile as one of the three measures requested.

3. What kind of statistics are used to examine the patterns of correlation among variables in a correlation matrix in order to see which groups of variables are most highly correlated with each
other? What statistic and/or graph is used to decide how many significant groups of variables are in
the correlation matrix studied?

“Factor analysis” (or principal components analysis) is the form of analysis/statistics. The criteria for
significant factors are the “eigenvalues” (which should be at least 1.0) and “scree plots” (additional
factors may be disregarded after the point where the plot begins to “flatten out.”

4. What measure of reliability is used to measure whether a set of attitude questions form a good
unidimensional scale? What is the range of acceptable values for this measure? Is this standard
related to the number of attitude questions included in the scale?

Cronbach’s alpha, which should have a value of .55 - < 1.0 for a very small number of items (3 or 4),
or above .7 or .8 for larger numbers of items. Yes, the more items, the higher the value is likely to be.

5. In OLS, what is the difference between a “b” coefficient and a “Beta” coefficient?

The b coefficient, the unstandardized regression coefficient indicates the average amount the
dependent variable is expected to change with one unit of increase in the independent variable,
controlling on the effects of the other independent variables which may be in the regression. For the
bs, the “units” are the original units of measurement – years of age, years of schooling, etc. The Beta
coefficient, the standardized regression coefficient indicates the same thing, EXCEPT the “units” are
standard deviations of the independent and dependent variables. This is equivalent to the regression
coefficients when the variables are transformed to “z-scores” or “standardized scores” by subtracting
the mean value for each of the individual variables from the actual values and dividing by the
standard deviations of the variables.

6. In logistic regression, what is the difference between a “b” coefficient and an
“exponential”/”multiplicative”/”odds-ratio” coefficient?

The exponential/multiplicative/odds-ratio coefficients are the natural anti-logarithms of the bs (the
logistic regression coefficients). The bs indicate additive effects on the logits, while the exponential
coefficients measure multiplicative effects on the odds.

7. Describe the advantages of measuring an attitude using a scale based on several questionnaire
items, as opposed to a single question.
The range of possible values in increased, the likelihood of the scores involving an interval or ratio level of measurement is increased, and the reliability of the variable as a measure is improved. Test-re-test reliabilities are likely to be higher, and there is likely to be less random measurement error.

8. In both OLS and logistic regression, you may need to use a statistical test of the significance of adding variables to a simpler model. What do we call this statistical test in OLS?

The F-test for the change in the model, also called the F-test for the change in R-square.

9. Referring to the previous question 8 above, what do we call the equivalent test in logistic regression?

The likelihood-ratio, or likelihood Chi-square test for the change in the model.

10. In OLS, name three different statistics which measure the effect of removing an extreme “outlier” from the set of cases included in a regression. Describe how each of these three tells you something different about the effects of removing the outlier.

“Leverage, or “h”, which measures “leverage” (the potential influence of a variable on the regression estimates). Cook’ D, which measures the actual influence of a case on all the regression coefficients together. The DFBETAs, which measure the actual influence of individual cases on specific regression coefficients.

11. If, after running your final model in OLS, you find that your distribution of residuals has a symmetric shape but has very “heavy” or “thick” “tails,” which of the following statistics can you believe the most, and which can you believe the least? The tests of significance, the b coefficients, the standard errors, or the tests of significance?

The “b”s, the regression coefficients will not be biased, but all of the inferential statistics – standard errors, confidence intervals, and tests-of-significance cannot be trusted.

12. What is the name of the test of significance for individual b coefficients in an OLS regression?
The “t” test. This is what we have taught. Sometimes “z” tests are reported, which are equivalent with large samples.

13. How do you calculate a variable to represent the “interaction” effect of the combination of two independent variables, $X_1$ and $X_2$ on the dependent variable?

By computing a variable for the interaction effect by multiplying $X_1$ and $X_2$ together. (Students can be given credit if they answer in terms of a form of STATA syntax, saying by adding the term $X_1##X_2$ to the list of variables in the regression command.)

14. Imagine you are doing an analysis of a survey for Norway, and you create dummy variables for each of the “fylker,” and you run your analysis omitting the dummy for Troms, using Troms as your reference category. Next you run the same analysis, but leave out the dummy variables for Troms, Finnmark, and Nordland. Is this wrong? What happens to the meaning of your coefficients?

This is not necessarily incorrect, if you understand and intend that you have changed your “reference category” to “fylker” in Northern Norway.

15. What does the statistic “tolerance” (a function of VIF) measure? What are worrisome values of tolerance? What are two kinds of independent variables which it is OK to use, even if they have problematic values of tolerance/VIF?

Tolerance is likely to be a problem if it is less than .1 (some texts will be more conservative and say .2). It is OK to have such levels of tolerance involving variables and the interaction or squared/cubed/etc. variables based on them.

16. What is a “conditional effect plot” “conditional” on? What does it tell you that a simple bivariate regression line would not?

It is conditional on controlling the other variables in a multiple regression (OLS or logistic) and setting them to “fixed” (constant) values, such as their means. It shows you the predicted values
“controlling” for the effects of the other independent variables, and “conditional” upon the other independent variables having the values to which they are set.

17. What is the difference between a standard error and a standard deviation?

A “standard error” is the standard deviation of a sampling distribution.

18. What is an “adjusted R-square” “adjusted” for? When should you use it rather than the ordinary R-square?

It is adjusted for the number of independent variables in the regression. You should use it when evaluating whether an improvement in R-square is simply due to adding the random correlation of additional independent variables with the dependent variable.

19. There are various measures of “influence” for outliers in regression. And, for some of the more important measures, you have learned that there are different standards for evaluating whether an influence measure is too high for given cases. Sometimes we differentiate between “absolute” and “conservative” or “adjusted” standards. What is the principal behind the difference between these standards?

The more “conservative” or “adjusted” standards involve “correcting for” or “adjusting” the standards for the fact that larger sample sizes will naturally result in it being less likely that individual outliers or influential cases will influence the regression coefficients. With very large samples, it is very unlikely that individual cases will have values over the absolute standards. The more conservative values will distinguish the “most influential” cases, although these cases may have very little actual effect on the coefficients.

20. Briefly list two reasons why we cannot use OLS regression with a dependent variable which is a binary, dichotomous, variable with two possible values.

1) Dichotomous dependent variables necessarily involve high degrees of heteroscedasticity, which violates the assumptions behind OLS, but logistic regression does not require the assumption of this kind of homoscedasticity.
2) OLS can produce estimates of the predicted probability of the dependent variable having a value of 1 of greater than 1.0 or less than 0.0, which is logically impossible. Logistic regression avoids this problem.

Sections II and III: Interpreting OLS and Logistic Regression Models (approximately 30 % each).
Data definitions for both sections.
The problems for sections II and III are both based on the data from the latest wave of the European Social Survey for Norway and Russia. The variables included in the models are as follows:

Unsafe - The respondents’ response to being asked if they felt safe walking alone in their local area after dark. The possible answers and codes were “1” - “Very Safe,” “2” - “Safe,” “3” - “Unsafe,” and 4 - “Very Unsafe.” In section II, dealing with OLS, this variable will be assumed to reflect an underlying continuous feeling of a lack of safety, with approximately equal intervals between the possible scores of 1 – 4. In other words, it is assumed to constitute an approximately interval level of measurement. This is the dependent variable for the analyses estimated in section II, using OLS.

UnsafeB - This variable is based on recoding Unsafe into a binary/dichotomous variable, in which answers of Unsafe and Very Unsafe were coded as “1” and “Safe” and “Very Safe” were coded as “0.” This is the dependent variable used in the analyses in section III, using logistic regression.

Age – Age in years.

Male - A dummy variable for gender, with males coded as “1”.

Eduyrs - The respondent’s number of years of completed formal education.

Minority - A dichotomous variable coded “1” if the respondent answered that he/she is a member of an ethnic minority group in the country.

Disabled - A dummy variable coded as “1” if the respondent stated that they were seriously disabled by illness, disability, infirmity, or mental problems.

Disabled_Some - A dummy variable coded as “1” if the respondent stated that they were somewhat disabled.

The reference category for the two variables above is respondents who stated that they were not incapacitated.
Poor Health-The respondent’s subjective reported health status, coded as “1”-“Very good,” “2”-“Good,” “3”-“Fair,” “4”-“Bad,” and “5”-“Very bad. This variable is treated as an interval-level 5-point scale.

City-A dummy variable coded “1” if the respondent lived in a “city.”

Suburb-A dummy variable coded “1” if the respondent lived in a suburb of a city.

Town-A dummy variable coded “1” if the respondent lived in a “town.”

The reference category for the variables indicating the size of the respondents place of residence involves persons living in small villages or on isolated farms.

Happy-The respondent’s answer to how happy they were at the time of the interview, on a scale from 0 to 10, with 0 indicating “extremely unhappy” and 10 indicating “extremely happy.”

Age2-Years of age squared.

MalexAge - the interaction effect between Male and Age.

In attached tables, the variable names are sometimes shortened or abbreviated, but they should be obvious.

In Tables 3,4, 7, and 8 interactions between country (Russia) and other variables are named with an “R” before the variable name or abbreviation of the second variable, as in RxAge.

Section II, OLS analyses:

Tables 1 and 3 of the attachments show OLS regression coefficients, tests of significance, standard errors and adjusted R-square values for 14 models, labeled “Model 1 through 14.” The standard errors are in parentheses, and significance levels are indicated by 1 – 3 “*” symbols, signifying the .05, .01, and .001 levels of significance.

In addition, the adjusted R-square for each model is listed at the bottom of each column.
The Beta coefficients for these same OLS models are shown in Tables 2 and 4 of the attachments.

Note that Model 5 is repeated in Tables 1 – 4. Also, the models in Tables 3 and 4 concentrate on tests of interactions between country (Russia compared to Norway) and the other independent variables.

The probability levels indicating the results of significance tests for differences between models are listed below.

<table>
<thead>
<tr>
<th>Models Contrasted</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1</td>
<td></td>
</tr>
<tr>
<td>Model 2 versus Model 1</td>
<td>0.000</td>
</tr>
<tr>
<td>Model 3 versus Model 2</td>
<td>0.000</td>
</tr>
<tr>
<td>Model 4 versus Model 3</td>
<td>0.545</td>
</tr>
<tr>
<td>Model 4 versus Model 1</td>
<td>0.000</td>
</tr>
<tr>
<td>Model 5 versus Model 4</td>
<td>0.000</td>
</tr>
<tr>
<td>Model 6 versus Model 5</td>
<td>0.010</td>
</tr>
<tr>
<td>Model 7 versus Model 5</td>
<td>0.679</td>
</tr>
<tr>
<td>Model 8 versus Model 5</td>
<td>0.000</td>
</tr>
</tbody>
</table>

| Table 3           |         |
| Model 9 versus Model 5 | 0.000 |
| Model 10 versus Model 5 | 0.376 |
| Model 11 versus Model 5 | 0.000 |
| Model 12 versus Model 5 | 0.008 |
| Model 13 versus Model 5 | 0.006 |
| Model 14 versus Model 5 | 0.000 |
Questions for Section II, OLS, based on Tables 1 – 4. (Approximately 30% of the weight of the exam)

1. Based on the models in Tables 1 and 2 only, which model is the “best” model? Why? (Be careful, pay attention to some social scientific theoretical considerations, not just the numbers.)

The simplest, straightforward answer is that Model 6 is the best model. This is based on the fact that adding the term for age-squared does produce a significant improvement over Model 5, and its adjusted R-square is slightly larger than that of Model 5.

Why not Model 8? Its R-square is largest of all, the regression coefficient for “happy” is statistically significant at the .001 level, the improvement of the model is statistically significant at the .001 level, and its Beta is the sixth highest in Table 2. BUT, this is where my warning to think about the social science / theoretical meaning of the models, not just their statistics! It seems extremely unlikely that the respondents’ subjective reports of happiness had a unidirectional causal effect on the dependent variable of whether the respondents were afraid to walk in the vicinity where they lived after dark. At best, an argument for a weak effect of happiness on the feeling of fear might be made, but it is far more likely that living in a situation of such fear could have a significant effect on happiness. If we take into account causal arguments for why fear and happiness are correlated, it is hard to believe that the causal effect is unidirectional and completely in the direction of happiness causing fear of this specific type. Model 8 was included make it most likely that students would judge the models partly on theoretical grounds.

Model 7 is excluded from consideration because adding the interaction between Male and Age was not significant.

On comparing models 5 and 6. I have instructed students that when one has a very large sample size, and there is an effect which makes a trivial contribution to explaining the independent variable and makes the interpretation and further development of the model much more complex, then it is reasonable to accept and present the simpler model. In this case, with nearly 4000 cases, adding
age-squared to the model is significant on only the .01 level, and the improvement in the adjusted R-square is .001 or slightly less. Furthermore, further analyses show that, surprisingly, age, and age x country have very small and generally insignificant effects, particularly after controlling other variables. And, (in other analyses), adding age-square is barely significant in the Russian sample and not in the Norwegian sample. This is a case where I believe the principle of “parsimony” outweighs including trivial, complicating, effects just because they are barely statistically significant.

For this reason, I would argue that among Models 1 – 8, Model 5 is the best model. However, I would not count it against a student for pointing out the significance of age-square in Model 6, and recommending that it be included in more complicated models like those in Models 9 – 14, or a more complicated model.

2. Based on the models in Tables 1 and 2, which variables have the greatest influence on the feeling of safety? Describe the directions and sizes of their effects in very specific terms using the coefficients shown, and, if possible, their standard errors.

Based on the Beta coefficients in Table 2, and based on the improvements in the adjusted R-square when variables or sets of variables are added to the models, Country (Russia) has the strongest effect, followed by the effect of size of the place of residence, closely followed by the effect of gender. Since all three of these variables are dummy variables or dichotomies, their unstandardized regression coefficients can be compared directly. Using the unstandardized regression coefficients (“b”s) in Model 5, the net effect (b) of Russia (vs Norway) is 0.506; in the comparison of sizes of places, the biggest net effect is that between “(big)city” and the reference category (village & farm), for which the b is 0.390; and the b for Male is -0.371. Russians have a higher net average value on “Unsafe” than do Norwegians, residents of large cities have higher values than residents of smaller places, villages in particular. And males have a lower average value on Unsafe than do females. The standard errors for these coefficients are all approximately 0.03, so the confidence intervals for the regression coefficients are only approximately +/- 0.06.

Based on its Beta value, Poor Health has an important and statistically significant “positive” effect (Beta= 0.197, b=0.195), an effect negligibly smaller than that of being male. The only other variable whose effect is statistically significant significant (at the .05 level) in model 5 is years of education, which has a negative effect. Perhaps surprisingly, the effects of age and being a member of an ethnic minority are not statistically significant.

3. The size and significance of the effects of some variables changes as other variables are added in Models 1 – 8. Describe a couple of instances of this and give a brief explanation of why a couple of the coefficients change across models.
The effects of being disabled are statistically significant in Model 2, but not after the variable for poor health is added in Model 3. Model 4, including poor health but omitting the dummy variables for being disabled, has the same adjusted R-square as Model 3, which includes both variables, and the F-test for the difference between models is not significant. Evidently, once the variable for poor health is added, the variables for being disabled add nothing to the model and are omitted from the subsequent models. The measures overlap, and the variable for poor health seems to be a better and more parsimonious predictor.

The effect of age, while relatively small, is statistically significant in Model 1, but becomes insignificant and is reduced to $b<.000$ after the variable for poor health is introduced in Models 3 and 4. Evidently all the effect of age is through its positive association with poorer health.

The positive effect of the country difference, Russia, shrinks from 0.758 and 0.752 in Models 1 and 2 to 0.596 and 0.602 in Models 3 and 4, suggesting that the country effect is partly explained by poorer health and more persons disabled in Russia than in Norway. It further shrinks to 0.506 in Model 5, suggesting that a greater urban concentration in Russia also makes a slight positive contribution to the Russian country difference. These differences are bigger than the coefficients confidence intervals of approximately 0.06.

The effect of being a member of a minority group is small, (perhaps surprisingly) negative, but statistically significant in Models 1 – 4, but becomes insignificant after size of place of residence is introduced in Model 5. It appears that ethnic minority members may be overrepresented in the smaller places of residence, and part of the negative effect of their minority status is explained by this.

4. Tables 3 and 4 show regression coefficients for models including interactions between the dummy for country and the other independent variables. Which of these interactions are statistically significant, and which are most important in the Russian-Norwegian differences? Be specific about the direction and sizes of the interaction effects.

Model 9 shows that the interaction between Russia and Male is statistically significant, positive, and large ($b=0.228$). Males in Russia have higher net average scores on Unsafe than do Norwegians. The negative effect of being male vs female is much stronger in Norway than in Russia, where the effect is still negative, but to a much smaller degree.
In models 12 and 14, we see that the interaction between Russia and Minority is also significant, and slightly larger than the interaction between Russia and Age (for Russia by Minority $b=-0.296$ in Model 12, $-0.277$ in Model 14). In fact, in Models 12 and 14, comparing this interaction effect to the effect of the simple variable for Minority, it seems that the net effect of Minority is negative in Russia and evidently not significant for Norway. This negative interaction between Minority and Russia does not decrease significantly when all of the other country interactions (including those for size of place) are included in Model 14.

The interaction between Russia and Poor Health is also significant, large (as indicated by its Betas of $0.221$ in Model 11 and $0.231$ in Model 14 (Table 4). The unstandardized coefficients in Table 3, Models 11 and 14 indicate that the positive effect of poor health is about twice as great for Russia as for Norway, although this positive effect is significant for Norway as well as Russia.

The interactions between size of place of residence and country were also significant and reasonably large. Larger places of residence had higher net predicted values of Unsafe than did the smallest places; and this difference was larger in Russia than in Norway, although this effect of place was significant in Norway as well.

There was not a significant interaction between country and age.

5. In Tables 3 and 4, the effect of country changes across models. What does this tell you about the Norwegian-Russian differences?

The effect ($b$) of country goes from $0.506$ in Model 5 to $-0.93$ in Model 14! The biggest decreases occur when interactions are added for Russia x Male, Russia x Poor Health, and Russia x Residence. This tells us that not only is a significant part of the overall Russian vs Norwegian difference due to worse intervening factors in Russia (Table 1), but that the “positive” effects of Male, Poor Health, and urban residence, are greatly due to these effects being much greater in Russia than in Norway. The Russian Norwegian overall gross difference is due to both intervening “handicaps” in Russia, and the effects of these handicaps being much larger in Russia than in Norway.

6. If you were to test the best model for the data and model meeting the assumptions behind OLS statistics, describe three different types of graphs you could create to test these assumptions. Be specific about the name of the type of graph and what would be plotted on the X and Y axes.
You should create a graph to test for heteroscedasticity, here the predicted values of the dependent variable would be plotted on the X-axis and the absolute values or squared values of the residuals would be plotted on the Y-axis. A regression line, or better, a Loess regression should be fitted to the resulting scatterplot. A horizontal fit line would be consistent with a lack of heteroscedasticity and a strongly tilted or curved line would indicate a problem of heteroscedasticity.

You should create a histogram with the values of your residuals represented by the horizontal X-axis and the percentage or number of cases with residuals of various values indicated by the Y-axis. A curve for a normal distribution should be superimposed on this histogram, and the difference between the distribution of your residuals and a normal distribution would indicate the degree to which the assumption of “normally distributed residuals” is being violated.

A quantile-normal plot is also helpful in testing the same distribution, here the quantile distribution for your residuals would be plotted by the quantile distribution for a normal distribution. You should see a straight line if the distribution of residuals is normal. Which is plotted on the X versus Y axes varies by statistical software programs and textbooks.

The issue of influential outliers is somewhat different from the other “assumptions,” but the answer could include a graph for the leverage statistics, Cooks D, or the DFBETAS, plotted as either box charts, with the distribution of the measure of influence on the vertical axis, or scatterplots with the predicted values of the dependent variable on the X axis, with outliers identified by case numbers.

Section III, Logistic regression analyses

Tables 5 through 8 show equivalent models to those in Tables 1 – 4, are based on the same ESS data for Norway and Russia and include identical independent variables; however, the dependent variable has been recoded to a dichotomy, where “1” indicates that respondents did not feel safe walking in the dark in the area where they lived, and “0” indicates respondents who reported feeling safe. And, the models are estimated using logistic regression.

Tables 5 and 7 of the attachments show logistic regression coefficients, tests of significance, standard errors and pseudo R-square values for fourteen models, labeled “Model 1 through 14.”
The exponential/odds ratio coefficients for these same models are shown in Tables 6 and 8 of the attachments.

The probability levels indicating the results of significance tests for differences between models are listed below.

<table>
<thead>
<tr>
<th>Models Contrasted</th>
<th>p-value</th>
</tr>
</thead>
</table>

Table 5 (NOTE TO GRADERS! THESE CONTRASTS WERE FOR TABLES 5 AND 7, BUT WERE LABELED AS FOR TABLES 1 AND 3 ON THE EXAM. NO ONE POINTED THIS OUT IN THE EXAM, AND I BELIEVE EVERYONE UNDERSTOOD THAT THEY SHOULD BE FOR TABLES 5 AND 7 BECAUSE THEY WERE IN THE LOGISTIC REGRESSION SECTION OF THE EXAM AND WERE DISCUSSED AS TABLES 5 AND 7 ABOVE. BUT THIS ERROR SHOULD BE NOTED AND PERHAPS TAKEN INTO ACCOUNT.) THE SIGNIFICANCE OF THE CONTRASTS LEADS TO THE SAME CONCLUSIONS, WITH THE EXCEPTION OF THE CONTRAST OF MODELS 11 vs 5 WHICH IS SIGNIFICANT FOR TABLE 3, BUT NOT TABLE 7.

Model 2 versus Model 1 0.000
Model 3 versus Model 2 0.000
Model 4 versus Model 3 0.470
Model 4 versus Model 1 0.000
Model 5 versus Model 4 0.000
Model 6 versus Model 5 0.011
Model 7 versus Model 5 0.985
Model 8 versus Model 5 0.000

Table 7
Model 9 versus Model 5 0.000
Model 10 versus Model 5 0.074
Model 11 versus Model 5 0.166
Model 12 versus Model 5 0.001
Model 13 versus Model 5 0.000
Questions for Section III, Logistic regression, based on Tables 5 – 8.

1. Do the logistic regression analyses in Tables 5 – 8 lead you to the same conclusions as the OLS analysis in Tables 1 – 4?

Yes, they would not necessarily have to, but in this case, both analyses lead to the same conclusions.

The one small difference between the OLS and logistic regression analyses is that the interaction effect between Russia and Poor Health is statistically significant in the OLS analysis but not in the logistic regression analyses.

2. Describe the most important findings from the logistic regression analyses in terms of the coefficients and statistics specific to logistic regression.

This discussion should be largely the same as for the similar question in Section II, except that the effects of the logistic regression coefficients should be in terms of effects on the LOGIT, the exponential/odds ratio effects should be described in terms of multiplicative effects on the ODDS, and there are no coefficients to be discussed as being equivalent to Beta effects in OLS. The differences in pseudo-R-squares should be called that rather than adjusted R-squares; and the sizes of the pseudo R-squares are much smaller than the OLS adjusted R-squares, as are the differences in the pseudo R-squares smaller than the differences in the OLS adjusted R-squares.

The one small difference between the OLS and logistic regression analyses is that the interaction effect between Russia and Poor Health is statistically significant in the OLS analysis but not in the logistic regression analyses.

GRADERS: I expect the answers to this question to be “fleshed out” in more specific detail than I have written here. But, given the similarity of results, I would consider it satisfactory if a student answered this question as tersely as I have.
3. What are the relative advantages and disadvantages of using OLS versus logistic regression for these analyses?

The advantage of logistic regression over OLS is that using a single 4-category variable as the dependent variable may be questioned from the point of view of the level of measurement assumed, whereas no one would question using logistic regression with a dichotomous recode of the categories. On the other hand, one could argue that the OLS coefficients are somewhat easier to interpret, having the Beta coefficients makes it easier to compare the effects of independent variables measured with different scales of values, and having four values for the dependent variable allows taking into account a greater range of variation in the degree to which respondents felt unsafe.

4. Based on Model 9, calculate the predicted probability of a respondent feeling afraid for respondents with the following characteristics.

a. A male in Russia, 30 years old, with 12 years of education, not a member of an ethnic minority, in bad health, living in a large city.

0.507

b. A female in Russia, 30 years old, with 12 years of education, not a member of an ethnic minority, in bad health, living in a large city.

0.637

c. A female in Norway, 30 years old, with 12 years of education, not a member of an ethnic minority, in bad health, living in a large city.
d. A female in Norway, 30 years old, with 12 years of education, not a member of an ethnic minority, in very good health, living in a village.

<table>
<thead>
<tr>
<th></th>
<th>Case 1</th>
<th>Case 2</th>
<th>Case 3</th>
<th>Case 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-3,152</td>
<td>-3,152</td>
<td>-3,152</td>
<td>-3,152</td>
</tr>
<tr>
<td>Russia</td>
<td>0,842</td>
<td>0,842</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Age</td>
<td>0,03</td>
<td>0,03</td>
<td>0,03</td>
<td>0,03</td>
</tr>
<tr>
<td>Male</td>
<td>-2,366</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Eduyrs</td>
<td>-0,336</td>
<td>-0,336</td>
<td>-0,336</td>
<td>-0,336</td>
</tr>
<tr>
<td>Minority</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Phealth</td>
<td>2,16</td>
<td>2,16</td>
<td>2,16</td>
<td>0,54</td>
</tr>
<tr>
<td>City</td>
<td>1,02</td>
<td>1,02</td>
<td>1,02</td>
<td>0</td>
</tr>
<tr>
<td>R*Male</td>
<td>1,829</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Sum/logit</td>
<td>0,027</td>
<td>0,564</td>
<td>-0,278</td>
<td>-2,918</td>
</tr>
<tr>
<td>p</td>
<td>0,50675</td>
<td>0,63738</td>
<td>0,430944</td>
<td>0,051271</td>
</tr>
<tr>
<td></td>
<td>Case 1</td>
<td>Case 2</td>
<td>Case 3</td>
<td>Case 4</td>
</tr>
<tr>
<td>---------</td>
<td>--------------</td>
<td>--------------</td>
<td>--------------</td>
<td>--------------</td>
</tr>
<tr>
<td>Constant</td>
<td>-3,152</td>
<td>-3,152</td>
<td>-3,152</td>
<td>-3,152</td>
</tr>
<tr>
<td>Russia</td>
<td>(0,842*1)</td>
<td>(0,842*1)</td>
<td>(0,842*0)</td>
<td>(0,842*0)</td>
</tr>
<tr>
<td>Age</td>
<td>=0,001*30</td>
<td>=0,001*30</td>
<td>=0,001*30</td>
<td>=0,001*30</td>
</tr>
<tr>
<td>Male</td>
<td>=-2,366*1</td>
<td>=-2,366*0</td>
<td>=-2,366*0</td>
<td>=-2,366*0</td>
</tr>
<tr>
<td>Eduyrs</td>
<td>=-0,028*12</td>
<td>=-0,028*12</td>
<td>=-0,028*12</td>
<td>=-0,028*12</td>
</tr>
<tr>
<td>Minority</td>
<td>=-0,264*0</td>
<td>=-0,264*0</td>
<td>=-0,264*0</td>
<td>=-0,264*0</td>
</tr>
<tr>
<td>Phealth</td>
<td>=0,54*4</td>
<td>=0,54*4</td>
<td>=0,54*4</td>
<td>=0,54*1</td>
</tr>
<tr>
<td>City</td>
<td>=1,02*1</td>
<td>=1,02*1</td>
<td>=1,02*1</td>
<td>=1,02*0</td>
</tr>
<tr>
<td>R*Male</td>
<td>=1,829*1</td>
<td>=1,829*0</td>
<td>=1,829*0</td>
<td>=1,829*0</td>
</tr>
<tr>
<td>Sum/logit</td>
<td>=SUM(F4:F12)</td>
<td>=SUM(H4:H12)</td>
<td>=SUM(J4:J12)</td>
<td>=SUM(L4:L12)</td>
</tr>
<tr>
<td>p</td>
<td>1/(1+EXP(-F13))</td>
<td>1/(1+EXP(-H13))</td>
<td>1/(1+EXP(-J13))</td>
<td>1/(1+EXP(-L13))</td>
</tr>
</tbody>
</table>
5. If you could create three different conditional effect plots which best illustrate the most important relationships revealed by the best model, which variables would you plot on the X axis, and which variables would you use to distinguish lines? And why would you choose these?

In this case, I consider this mostly a subjective matter of judgement, and the rationale for the choice should be as important or more important than the specific choices.

The effect of age is insignificant and the effect of education is very small, as are their interactions with country, so I would not waste a conditional effect plot on them.

My first choice would be a plot where size of the place of residence is on the horizontal axis. This is because 1) this is the only “variable” which would have more than two values on the horizontal axis (it would have 4) or a simple linear line, 2) it involves one of the 3 or 4 biggest effects, and 3) it has one of the biggest interaction effects with country. So, I would let country and place of residence vary, creating a plot with two lines for Russia and Norway, varying across four categories of size of place.

Second, I would make a plot where country and minority status varied, and I would place the two categories of minority status on the horizontal axis with two lines for country. I would do this because this is a very important and interesting interaction effect, and the country differences are easier to understand in such a plot than in the table.

Third would be a toss-up between country and gender or country and health. Two categories of either gender or the scores of the poor health variables on the horizontal axis, with two lines for
Norway and Russia. I would favor these because not only are the simple effects of country, gender, and health the largest effects, but the interactions between country and gender and between country and health are both quite large. Given the choice, maybe I would either leave both out, or include both. Forced to choose just one – gender x country.