

## OPTION #2: Coil Data Analysis Project Using Your Own Data

Students from the Chita Institute of Baikal State University of Economics and Law (RU) in a 4th year class on [Surveying the American Economic System](#) are interested in meeting and discussing issues and data with American students. Irene has been working with Dr. Elena Pischerskaya for 5 years and have come to enjoy working with her and her students who know some English and are eager to have conversations with Americans.

### Process for completing this COIL project:

This option involves statistical analysis of a set of data to answer a research question generated by you and your Russian student partners that involves an association between two variables. If you choose this option, you may find that you need some extra guidance from the Irene Duranczyk, so it is best to **start early** if you choose to embark on this project. There is a meeting virtual or in person on Jan 29 to talk about our engagement with Russian Students. See the [Specifics of the COIL project](#) below.

#### Step 1:

**Either:** Your first step will be to find a set of raw data. In choosing your data, keep in mind the following points:

- In a data file, each column is a variable and each row is a case. Make sure your data set consists of raw data where each line is a case (similar to the data sets provided on Moodle), and not of a summary report or summary table.
- Be sure your data is from a **sample**, and not an entire population. This project will focus on statistical inference, which is pointless when you already know everything about the population!
- There are no strict sample size requirements, since the necessary sample size will depend on the method you want to do. A sample size of 30 is the absolute minimum, but for most methods you will need more than that.

**OR:** You will design a google forms with 6-10 questions and then distribute the survey to 40-50 US or RU students.

- In a data file from the survey will have each column as variable and each row as a case.
- Your data will be from a **sample**, and not an entire population. This project will focus on statistical inference, which is pointless when you already know everything about the population!
- There are no strict sample size requirements, since the necessary sample size will depend on the method you want to do. A sample size of 40-50 will work.

**Please e-mail or talk to Irene Duranczyk to make sure your dataset is appropriate before completing Part 1.** It is highly recommended that you either e-mail your data or show Irene Duranczyk your data in person no later than **1 week before (February 14<sup>th</sup>) Part 1 is due** (but

the earlier the better!). It would be best if you also shared an idea of what variables you would like to explore when you share the data with Irene Duranczyk, but that is not absolutely required.

**Part 1 (10 points) (Due February 21<sup>st</sup>):** E-mail Irene Duranczyk your data file, (unless data privacy/IRB rules do not allow you to do so), the Google Survey form if you collected your own data and a MS Word or PDF document containing the following information:

1. Where did you obtain your data? (e.g., if you found it online, please provide the link to the website. If you obtained it from a professor, what department is the professor from?) If you collected it from the Google Form, include a link to the form. **(0.5 pts.)**
2. Describe your sample. How large is it? How were the data collected? **(1 pt.)**
3. Choose TWO variables of interest to explore. (If you really want to go above and beyond, you could choose more than two, but you would have to learn methods on your own that are beyond the scope of this class.) You will have one of *three* possibilities here:
  - a. Two categorical variables.
  - b. Two quantitative variables.
  - c. One categorical, one quantitative variable. *Note that if you choose this option, the categorical variable must have no more than TWO categories, because this class will not cover methods for comparing more than two groups on a quantitative variable (i.e., ANOVA). If you're interested in a categorical variable with more than two categories, you may need to re-code it into two, such as putting all cases who said "Strongly Agree" or "Agree" into a single category that says "Agree."*

Describe each of your chosen variables as completely as possible (e.g., definition; type of variable, categorical or quantitative; explanatory/response variables, if applicable, etc.) **(2 pts.)**

4. Pose a research question. **(1 pt.)** Here are some guidelines to help you out:
  - 1) If you have *two categorical variables* (one variable that has multiple groups or no variable with distinct groups that you want to compare) or *two quantitative variables* that you wish going to compare, you would say:
    - "Is there a relationship between variable 1 and variable 2?"
  - 2) If you have one categorical variable (with two groups) and one quantitative variable (a difference in means situation) or two categorical variables (one variable that has distinct groups that you want to compare with regards to the second variable...a difference in proportions situation), you would say:
    - "Is there a difference in (parameter for the quantitative/categorical variable) between group 1 and group 2?"
  - 3) If you have paired data (e.g. quantitative data collected on the same subjects at two different time points), you would say something like:
    - "Is there a difference in (parameter for the quantitative variable) from time 1 to time 2? (Or a similar question, depending on the nature of your data.)
5. Provide appropriate descriptive statistics and graphical summaries for each of your variables of interest. **(2 pts.)** (You do not need to write anything for this part, just provide

output from your statistical software of choice.) Here are some guidelines for what type of output to provide:

If you have two categorical variables, see chapter 2.1 and the “Describing Two Categorical Variables” part of your technology reference guide on how to examine (descriptively and graphically) categorical variables.

If you have two quantitative variables, examine each variable individually (e.g. histogram, dotplot, and/or boxplot, and summary statistics)

If you have one categorical variable and one quantitative variable, examine the quantitative variable for each group (e.g. histogram, dotplot, and/or boxplot, and summary statistics, but separately for each group). (See the “Comparing Groups on a Quantitative Variable” part of your technology reference guide.)

6. Describe in as much detail as you can the distribution for each of your variables of interest. **(3 pts.)**
- If you are examining any quantitative variables, be sure to address shape, center, and variability for each variable.
  - If you are comparing two groups on a quantitative outcome, be sure to compare the two distributions on their shape, center, and variability.
  - If you are examining two categorical variables, comment on the frequencies of each variable individually and also comment on any tables and graphs you made in question #2.

Make sure you write in a way that someone who doesn’t know your data (or statistics) understands.

7. Propose a type of analysis. **(0.5 pt.)** I don’t want you to actually conduct the analysis, just state what type of analysis that you would do, based off of your research question.

The table below will help you decide which type of analysis would be appropriate to answer your research question.

<b>Types of Data</b>	<b>Possible Analysis</b>
2 categorical variables (one with two groupings and one response variable)	Hypothesis test for a difference in proportions Confidence interval for a difference in proportions
2 categorical variables where you want to investigate if there is a relationship between the two	Chi-Square Test
1 categorical variable (two groups) and one quantitative variable	Hypothesis test for a difference in means Confidence interval for a difference in means
2 quantitative variables where you want to investigate if there is a relationship	Correlation

between the two	
Paired data	Hypothesis test for a mean difference (paired data) Confidence interval for a mean difference (paired data)

As previously mentioned, please e-mail your data file (if allowable) and your Word/PDF document addressing the above points to Elizabeth Fry. If you worked in a group, be sure to CC your group member(s) on the e-mail and put all of your names on the document. The instructor will then provide feedback to help you move on to Part 2.

**Part 2 (20 points) (Draft Due March 12 to share with Russian students, but final will be due Apr. 30) :** Create a report or presentation (this may be either a Word/PDF document or a PowerPoint presentation) to share with the class on the discussion forum *Data Analysis Projects*. It is up to you what type of format you'd like to use (a written report versus a PowerPoint presentation), but your document must include the following components:

1. Title of your project and group member names, if applicable **(0.5 pt.)**
2. Introduction **(3 pts.)**
  - i. Motivation for studying the topic
  - ii. Background information about what you investigated / what the variables are
  - iii. Your research question(s)
3. Descriptive statistics including
  - i. Graphs of your data, with clear labels and titles (note that these should be presentable, i.e. no screenshots of your entire computer desktop or labels that include "R-speak") **(2 pts)**
  - ii. Numerical summaries (only ones that are helpful & relevant; don't include a long list of summaries if they don't enhance the story of the data!) **(2 pts)**
  - iii. Sample size(s) (also, if you had missing data please include how many cases were missing) **(0.5 pt)**
4. Method used to analyze the data **(2 pts)**
  - i. If you did a test for means or proportions: Randomization or Traditional? (Why, if applicable)
  - ii. H-test, CI, or both?
5. Results **(3 pts.)**
  - i. This should include your confidence interval and/or  $p$ -value, depending on the approach you took. Also - If you did a hypothesis test and found the result to be statistically significant, do you think it is practically significant? Why or why not?
6. Conclusions **(6 pts.)**
  - i. Answer in context, (e.g., "Reject  $H_0$ " alone is not appropriate).
  - ii. Address what generalizations and causal conclusions can or cannot be made
  - iii. Be sure that the research question is clearly answered and results are communicated to an audience that may not be well-versed in statistics.
  - iv. Follow up research questions, if appropriate

7. Citations (APA format preferred, but if your department uses another format, you may use that style) (1 pt.)
  - i. Credit the data sources
  - ii. Note other related studies that you may have compared your results to, if applicable
  - iii. Cite any other research you used to complete your project), if applicable

Suggested length for Part 2: There is no required length, but if you do a PowerPoint presentation, try to limit it to no more than 15 slides. If you have a document, try to limit it to no more than 4 pages single-spaced (including graphs/output). It can certainly be shorter, but please do not write a novel!

Instructions for posting your project:

- On the course website, find the module labeled *Create Your Own Assignment*.
- Click on the *Option #2: Data Analysis Project (using your own data)* forum and then click on the “Reply” button.
- At the top of the text box, type the title of your project. (You may choose to click on the button that says “Paragraph” and choose the style “Header 2” to make the title stand out more.)
- Below the title, provide the link to your video and the names of all group members (if you worked in a group).
- Click on the *Attach* button at the bottom of the text box.
- Click on *Browse...* and find your file to upload, then complete your post.

## Specifics of the COIL project

The Chita students have class assignments related to one of the topics below. These topics are broad and can be defined and narrowed by Chita and Minneapolis students negotiating the space and their specific interests. Their research will include data about the US on one of the topics within these broad areas and may also include data about the same topic from Chita or Russia for comparative purposes. This is an opportunity for you to share data to support their studies and also to learn about these same topics in Russia from a Russian student.

Broad topics are:

- Social policies affecting healthcare, social security, temporary assistance for needy families, food stamps, education, unemployment insurance;
- Income inequality including data, reasons, history, protest movements;
- Foreign trade including data, major trade partners, trade deficits, most exported/imported goods, trade restrictions;
- The New Deal developments in the U.S. during the 1930s and 1940s;
- Government shutdowns;

- U.S. monetary and fiscal policy; and
- the concept of “the American Dream” and development of “Manifest Destiny”, its history, and one’s present-day understanding.

Chita students are very interested in linking up with one or more students from the University of Minnesota by social media including video conferencing to discuss with American students their insights and research into these topics and identify resources that can help round out their ideas on the U.S. Economy by getting specific feedback on their research topic from an American student. As their semester started on January 9th and ends on Mar 13th your involvement with Russia students will only be from Jan. 29th through March 13. During this time you will share your articles and data even though your write up of your analysis will not occur until April. The student in 4th year Russian students will be in an internship program for the last 10 weeks of the semester, however they will have an opportunity to present their work with you at a year end conference so they may keep in contact with you so that your paper may become a part of their final presentation at the conference and you will be credited with this work.

- Our first video conference for all students wishing to participate in this "version" of the project will be on January 29th between 6:30 pm and 8:00 pm (You will be traveling space and time as the Chita, RU students will be experiencing January 30 at 10:30 in the morning.). I will have a video conference room available through WebEx: <https://umn.webex.com/join/duran026>.
- You can join me in this virtual room or you can meet me in a video conference room on the St. Paul campus Building: **Peters Hall** Room #: **145**. The Chita, RU students will also be in the virtual room.
- Prior to that date I would like you to create an introductory statement about yourself so I can compile a document for the Russian students introducing UMN students. I will need your name, a picture of yourself, social media contact information that you would like to share (e.g. email, Twitter, Facebook, WeChat, WhatsApp, Instagram, etc) and a bit about yourself and your interests (2-3 sentences). The Chita students will be doing the same for you.

**NOTE:** Last year, the Chita Institute of Baikal State University course-based economics project involved all the students from the U.S. Economy (18 students) course plus the Minnesota participants for a total of 28 members interacting with unique 15 papers. Some of the economic projects were subsequently submitted to an annual undergraduate student research project contest in Chita. They were highly ranked above their peer researchers and undergraduate institutions. The project gave students the necessary first-hand information and personal interaction experience that is critical for better global understanding and further practical use of the theoretical knowledge they learn in classes.

Last year Chita students conducted research in groups of 1, 2, or 3 with 1, 2, or 3 UMN students using the following process:

- They shared drafts of their research with UMN students to ask for their personal view/experience/ideas about major economic and social aspects of life in the U.S.
- UMN students answered questions posed by Chita students about their topics.
- UMN students found documents, resources, and created personal statements about the research topics to share with Chita students.

- UMN students read papers written by the Chita students and provided comments and clarifications.
- Russian students analyzed the feedback in their papers and presented the results in class and at an annual youth conference.

This year as we are in a statistics class and data supporting your stance will be required. For this assignment, you will find a data set to support your international topic OR you will use Google Form to create a short (6-10) question survey to collect data from 40-50 US or RU students on your chosen topic. The follow follows the guidelines for Option #2 Data Analysis Project Using your Own Data, however, you have the twist of the International Focus and some sped up deadlines for submission.