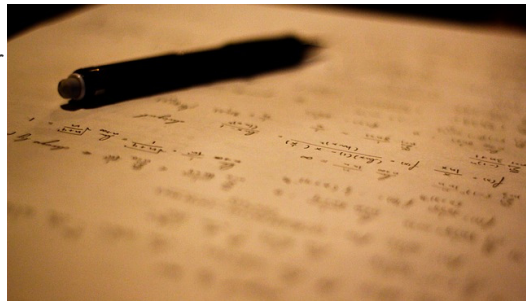
Peer Assessments

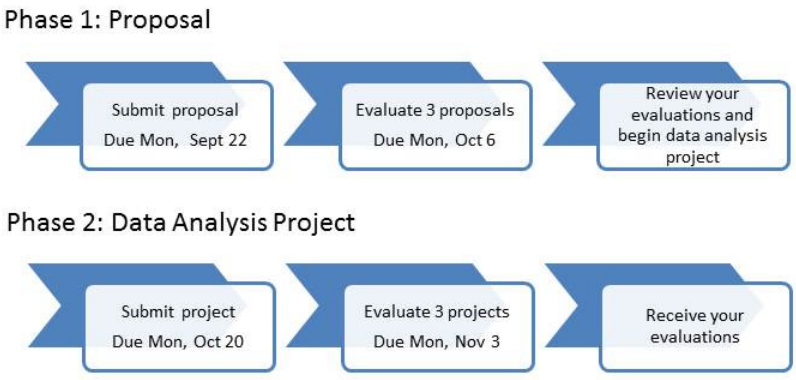
Help



Your independent data analysis project will be done in two phases, each assessed by other students using peer assessment. You can also share your projects via the forums for additional feedback. The project counts as 15% of your grade for a certificate with distinction.

- **Before the deadline,** use the discussion forum to ask specific questions about the assignment. Please don't post your proposal.
- **After the deadline**, we encourage you to post your project submission to the forum to show other students interesting data sets and questions, and to get additional feedback on project (in addition to the three peer assessments you will receive formally that will be used to calculate your score for the project).

Please note: Both phases of the project are open so that you can see what is expected. As we learn how you work, we may make small modifications to the project requirements as the course progresses, so please check again before the due dates. You can resubmit your work up until the due date.



Phase 1: Project Proposal

Identify a research question similar to questions we've talked about in this course. Choose a dataset, and one or two variables from that dataset, with which to answer this question using a hypothesis test or confidence intervals (the dataset used is entirely up to you, it can be one of the datasets listed below

under "Datasets for the Project", or another one of your own choosing). You should pick two variables of interest, and you will be exploring the relationship between them. These variables should be either numerical and categorical, or both categorical (but they cannot be both numerical, see below for examples).

As part of the proposal stage you are also asked to complete a brief exploratory data analysis to determine if your data is appropriate for the project. You should submit a proposal with enough detail so that your peers can give you feedback before you start the full data analysis.

You can begin working on the proposal immediately, following the link at the bottom of this page. Please save your work as you go along. When you're ready to submit your work for evaluation, remember to click the "Submit" button. If the deadline passes and you haven't clicked "Submit" yet, then your saved work will not be evaluated. **Note:** You can re-submit your work for evaluation as many times as you want before the submission deadline March 10th. Only your last submission will be seen and evaluated by your classmates.

After the submission deadline, you will have two weeks to provide feedback to others on their project proposals. Please assess at least 3 proposals before March 24th. This peer assessment will help you prepare for your project and provide you with experience with a variety of data sets and research questions.

Phase 2: Data Analysis Project

Once you receive feedback on your proposal, you will then continue onto the data analysis project. This peer assessment will count towards your grade for a certificate with distinction. You will answer the research question you developed in the proposal phase using methods you've learned in this class, and summarize your findings into a report. More details on what should be included are in the assessment itself, linked below.

This project is due on April 7th, and then you will have two weeks to provide feedback to other students on their projects. You will only receive full credit for your project if you evaluate 3 other projects before the April 21st deadline.

Datasets for the project

We are providing two datasets that you can use for your project. Both of them come from large-scale US surveys, and they have been modified slightly to make them easier to use as part of this course. Even though there are only two datasets, each contains many variables and hopefully you should be able to find some combination of variables that are of interest to you.
(1) General Social Survey (GSS): A sociological survey used to collect data on demographic characteristics and attitudes of residents of the United States. The codebook below lists all variables, the values they take, and the survey questions associated with them. There are a total of 57,061 cases and 114 variables in this dataset. Note that this is a cumulative data file for surveys conducted between 1972 - 2012 and that not all respondents answered all questions in all years.

- Codebook: Review the codebook to view a list of all variables, the values they take, and the original survey questions associated with the variables.
- Use the following code to load the GSS dataset into R:

```
load(url("http://bit.ly/dasi_gss_data"))
```

For access from China try using the following URL in the code above:
https://d396qusza40orc.cloudfront.net/statistics/project/gss.Rdata

The name of the dataset that you load is gss. For example, you can see a list of the variable names using the following command:

```
names(gss)
```

(2) American National Elections Study (ANES): A survey of voters in the United States, conducted before and after every presidential election. The codebook below lists all variables, the values they take, and the survey questions associated with them.There are a total of 5,914 cases and 205 variables in this dataset. Note that not all respondents answered all questions.

- Codebook: Review the codebook to view a list of all variables, the values they take, and the original survey questions associated with the variables.
- Use the following code to load the ANES dataset into R:

```
load(url("http://bit.ly/dasi_anes_data"))
```

For access from China try using the following URL in the code above:
https://d396qusza40orc.cloudfront.net/statistics/project/anes.RData

The name of the dataset that you load is anes. For example, you can see a list of the variable names using

the following command:

```
names(anes)
```

R Markdown template for the project

You will use R Markdown for your project. The template can be downloaded using the following code:

```
download.file(url = "http://bit.ly/dasi_project_template", destfile = "dasi_project_template.R
md")
```

If the above shortened link doesn't work for you, try the following URL

instead: https://d396qusza40orc.cloudfront.net/statistics/project/dasi_project_template.Rmd.

Watch a demo video for using R Markdown for your project here.

Examples of variables and research questions appropriate for the project

Provided below are some examples to give you an idea as to what types of research questions and data are appropriate for the project:

- *One numerical and one categorical:* Is there a relationship between whether the mother worked during the first 5 years of the child's life and the highest level of education the child attains? [Data: Number of years of education of child; Mom's working status - yes, no]
- *Two categorical:* Do racial minority groups in North Carolina have less access to health care coverage? [Data: Ethnicity - various levels; Health coverage - yes, no]

Note that you can work with one numerical and one categorical variable, or two categorical variables (but not two numerical variables, as we'll learn methods associated with those later in the course).

While it may not be clear to you yet which techniques you would use to do statistical inference to answer these questions, you will have learned all the tools you need for the proposal by the end of Week 1, and all the tools you need for the actual project by the end of Week 6.

Questions and Answers about the project

What type of variables should I use for my project?
You need to pick two variables and evaluate the relationship between them. These can be a numerical and a categorical variable or two categorical variables. Do not use two numerical variables.

How will I find a dataset for my project?
You can either find your own dataset on the web, collect your own data, or use one of the datasets we provide specifically to be used in the project (listed above). While finding/collecting your own dataset is an eye-opening experience that we think is immensely valuable, it can be very time consuming. There are a variety of data resources online (some of which we'll point you to throughout the course and some of which you might discover yourself), but the internet is a bottomless well, and so finding a dataset that interests you, that is appropriate for this project, and that can be analyzed using tools from this course can be challenging, so choosing to work with one of the datasets we provide to get a head start on the project quickly can also be a good idea. You're the best judge of how much time you can devote to the project, so it is your choice and responsibility to decide which route you want to take.

Can I use a dataset from the labs?
No, you can to work with just about any dataset you like (as long as it meets the conditions required to apply the inferential methods we learn in this class), but you cannot re-use a dataset from the labs. One of the main objectives of this project is to work with a novel (to you) dataset. You can, however, choose to work with one of the datasets we provide specifically for the project (listed above).

Where can I find a list of R commands that might be useful for the project?
See this document for a list of R commands you encountered in labs as well as a few others that you might find useful for the project. Note that this is not an exhaustive list.

Can I use DataCamp for my project?
No, for the project you will need to use R/RStudio.

Who am I writing for?
Write as if you are explaining your results to whoever would be interested in your research question, whether this is other scholars in your field or peers sharing your interest in the topic. This audience may not have taken statistics. You must be statistically accurate and use correct statistical terminology, but must also explain your conclusions in a way that anyone can understand.

Who will see my work?
Other students who have submitted similar work will be given your work to evaluate. In addition, you will be able to share your work via the forum with the rest of the students to benefit everyone in class. You will benefit by receiving additional ideas about your project.

How will this project be graded?
The projects will be evaluated via peer assessment. Three other students will carefully read your project proposal, and will provide feedback on your project proposal. This part will not count towards your final grade, but will prepare you for the data analysis project, show you other data sets and other ways of approaching questions, and give you practice with peer assessment. For the data analysis project, you will receive a score that is the median of your peer reviewers. You will receive the full score if you perform all 3 of your peer assessments; if you do not complete 3 peer assessments, your final score will

be decremented by 20%. The project score counts as 15% towards your final grade for a certificate with distinction.

What is a peer assessment?
Peer Assessment is when students in a course evaluate each others work. First, each student submits an assignment. Then, the students who have submitted an assignment are given other students assignments to evaluate, according to criteria. Finally, each student receives the grades that the other students assigned to the work. More information is available on the Coursera help site.

What do I have to do when?
There are two phases of the project: a project proposal phase and a data analysis project phase. Each phase contains a submission period and an evaluation period where you will provide feedback on others projects. Although you can access the data analysis project right away, you are more likely to be successful if you wait until you get feedback on your project proposal before submitting a completed data analysis project. See the Due Dates page.

Assignments

Details: OFF

✱ *Timezone:* ***Europe/Dublin***

Project Proposal *(Submitted with authentication)* Go to assignment ›

Results available since Mon 24 Mar 5:00 pm

Data Analysis Project *(Submitted with authentication)* Go to assignment ›

Results available since Mon 21 Apr 6:00 pm