

**Statistics 6510**  
**Survey Sampling Methods**  
**Autumn 2014**  
**Project Guidelines**

To solidify the concepts covered in class, as part of this course you (as part of a small group) will design and conduct a high quality survey, analyze the resulting data in a statistically sound manner, and present your results via a virtual poster and a written report.

To keep you on track in your project development, beginning in the sixth week of the course, you must submit progress reports by 5:00 pm on the Monday of that week. Reports should be electronically submitted to the corresponding Carmen drop-box in .pdf format. If you need help creating .pdf documents, please contact the instructor. Only one report should be submitted per group. Each progress report should include the names of all the project participants at the top of the first page. The information I expect to be in these reports (and resulting expectations regarding the overall project) is described below. Each report will be evaluated. Because each of these steps must be completed for a successful project, you will be asked to **resubmit** any unacceptable progress reports. The interim reports will make up 10% of your final grade. The poster presentation in week 14 will make up 5% of your final grade. The final report will be worth 20% of your final grade.

You are free to divvy up the project work in any way your group would like to, with the exception that **all** members must be involved in developing the project sampling plan and analyzing the data, and **all** members must read and approve the final report before submission. That is, all members of the group must be involved in the project in some way from start to finish. If you would like assistance with any aspect of your project, please contact the instructor via email to set up a meeting.

**Week 6 Progress Report: Propose a Research Question, due Oct. 3**

Propose a research question. Because the time we have for this class prohibits the lengthy approval process required for human or animal subjects research, **your survey may not involve humans or other animals in any way**. This includes passively recording human or animal behavior. Some broad populations for research might include library books, web sites, parking spaces, trees, or groceries. You may not propose an idea for which any member of the group has already received course credit or payment (i.e., no RA projects). Your question should be specific enough to provide direction, yet broad enough to provide an *interesting* and rich research question. Specifically, at a *bare minimum* answering your question must involve recording one binary variable that can be used to group sampled units and a total of at least four variables for each sampled unit **that are not known before sampling**.

Examples of past projects:

- What factors account for the differences in prices of used cars, specifically 2005 Ford Mustangs with V-6 engines?
- How much and what kind of graffiti is in OSU Columbus campus' public restroom facilities?

You may use these for inspiration, but you may not repeat these exact projects. Note that it is difficult to devise a good and simple sampling plan for cars on campus; I advise you to choose a different topic.

Submit a report that includes:

1. Your question of interest
2. A description of an observation unit
3. The variables you will collect for that observation unit (at least 4 total, at least one of which is categorical)

### **Week 7 Progress Report: Sampling Plan, due Oct 10**

For this week, you should plan your actual sampling.

Submit a report that includes:

1. Your target population and population to be surveyed. Be specific about how these differ, if at all.
2. Your sampling frame. If this is a “virtual” frame, explain what that frame is.
3. How you plan to sample units. Be *specific* enough in your description that someone could carry out your sampling plan based on this description without consulting you.

Note that this includes a method for random selection. A statement that your group will “randomly select cars” is not enough for someone in the field to actually identify the sample of cars. Instead write something like: “We will use SAS to randomly generate 100 numbers between 1 and 100 [without replacement], which correspond to parking spaces on the first floor of the Northwest Garage. We will go to each space in the order they were randomly generated until we observe 5 cars parked in the corresponding spaces.”

Think about what might happen in your sampling, and plan for the worst case scenario. For example, what if there aren’t 5 cars parked there? What if one car parked 1/2 way between two spaces?

4. A justification for your sample size. If you have chosen a SRS or stratified random sampling design, use the strategies for sample size estimation we have talked about in class. If you have chosen a cluster or complex design, we have not yet learned about how to estimate sample sizes. Thus, it is OK for the purposes of this project to select a sample size based on the SRS formula, knowing that this will likely under-estimate the sample size you would ideally like. Regardless of the sampling design you use, **I expect that you will adjust your sample sizes to limit the time you will spend collecting data (the total “cost” of collecting data)**. My expectation is that you will use a sample size that is large enough to calculate reasonable estimates, but not so large that you will be able to find “statistically significant” results with a 95% confidence level.

**Week 8 Progress Report: Plan for Data Collection and Initial Plan for Analysis, due Oct. 17**

For this week, you should make final preparations for data collection, and connect your sampling plan to what you intend to do with the information once you collect it. That is, how do you intend to answer your question with the data?

Submit a report that includes:

1. A detailed plan for data collection, including who will be collecting what data where and how (e.g., will you have a chart to fill in using pen and paper?), and who and how you will digitizing your data.
2. a list of the design-based analyses you intend to complete. You should be specific as to the variables you will use and the general methodology (SRS vs. stratified vs. cluster-based analysis). You do not need to explicitly write down the formulas or SAS commands you will use.

**Weeks 9-10 Progress Report: Collect Data, due Oct 31**

Carry out your data collection plan. Digitize your data.

Submit:

1. a file containing your data
2. a brief report recounting any difficulties you encountered in data collection. If you did not encounter any difficulties, submit a note that says so.

**Week 11 Progress Report: Exploratory Data Analysis, due Nov. 7**

Perform exploratory data analysis for your data. Make appropriate tables, charts and graphs. Note any unusual observations. (Are these perhaps typos? Clean your data.)

Submit a report that includes:

1. a short summary of your exploratory data analysis (include the interesting tables, charts, and graphs.
2. information about any data cleaning you performed. If you did not need to “fix” any data, note that. If you did, do not include any pre-cleaning analyses, as you then believe these are not valid.

**Weeks 12-13 Progress Report: Design-Based Data Analysis, due Nov 21**

Use the sampling-based techniques we learned in class to analyze your data in such a way as to answer (parts of) your original research question.

Submit a description of your analyses and results, and the conclusions you draw.

**Week 14, Virtual Poster, due Nov 25**

Create a virtual scientific poster that summarizes your project so far (up to and including a design-based analysis) and that also includes some preliminary discussion of your results. That is, the poster should contain the same information as for items 1-5

of the final report described below, with the exception that no model-based analysis is required. Information about how to prepare a scientific poster is available from many sources online; here is one site that includes both tips for good poster preparation and links to software tutorials: [http://www.waspacegrant.org/for\\_students/student\\_internships/wsgc\\_internships/posterdesign.html](http://www.waspacegrant.org/for_students/student_internships/wsgc_internships/posterdesign.html).

Upload your poster as part of a new thread in the appropriate discussion forum on Carmen by 5:00pm.

### **Week 15a Virtual Poster Session, Nov 25 - Dec 5**

This week the class will be participating in a virtual poster session, where all students should review all the other project posters and provide meaningful feedback to the other groups. At a minimum, each student should comment on two pre-assigned posters by December 3 via the Carmen discussion system. (Assignments will be emailed to each student by November 25.) Participation in the virtual poster session will be worth 2 points out of the participation portion of the course grade.

### **Week 15b Progress Report: Model-Based Data Analysis, due Dec 5**

Use the model-based techniques we learned in class to analyze your data in such a way as to answer (parts of) your original research question.

Submit a description of your analyses and results, and the conclusions you draw.

### **Week 16: Final Report, due Dec 12**

Submit a *polished* final report of your research. While there are no maximum or minimum number of pages, I expect the final reports to be approximately eight 1½-spaced pages. This report should include the following sections:

1. Introduction (10%): What is your question? Why is it interesting? If you have background information, include it.
2. Data (20%): The information from Week 7, your chosen sample size and justification for that sample size (and any updates to this justification that you might make now that we know more about sample sizes in complex surveys), and a description of any problems you encountered in data collection that might affect your conclusions.
3. Methods (20%): What statistical methods did you use to analyze your data? Be specific enough that anyone with access to the course textbook or notes could repeat your analyses. You do not need to actually reproduce the formulas from the textbook. You should combine the design- and model-based methods in this section.
4. Results (20%): Present your analyses and conclusions based on those analyses, including tables and graphs as appropriate. You should include both the design- and model-based results in this section.
5. Discussion (15%): What is the answer to your research question? How does this relate to the world (e.g., do you recommend some action based on your results)?

What are the limitations of your study? What would you do differently next time? Are there more questions your research implies would be interesting?

6. Appendix (required for any grade): Include your computer code as an appendix. You may also include other appendices if you feel they would add to the understanding of your project.
7. Data (required for any grade): Upload your dataset to Carmen in a file that can be easily loaded into R.

This report makes up the bulk of the project portion of your grade. Make sure it is both complete and easy to read. That is, do *not* just copy and paste from previous reports. Make sure *all* members of the group read the final report from start to finish before submitting it. If you include figures or tables, be sure they are numbered and referred to in the text. Do not include any raw computer output. Refer to any major research journal for ideas about what kinds of formatting would be good. Presentation comprises 15% of the grade for the final report.