Statistics 480, Winter 2024: sampling project

The requirement for the project is to prepare a class presentation and paper in response to one of two options. These are

1. Design, execute and analyze data from a sample you draw from some real-world population. The reports should describe and justify the sampling design.

2. Consider data from some real-world probability survey of a finite population and discuss (critically appraising if you can) the sampling design and execution.

For both options, if the sampling design differs from a simple random sample, include discussion of the design effect (or of the "relative efficiency" as this is described in the course textbook). Also you should report finite population inferences (with bounds) about at least three quantities of interest, including at least one ratio or regression estimator. All estimates must be based on the same sample (or perhaps on subsamples of one sample). You should explicitly discuss nonresponse and if possible take steps to mitigate the consequences of nonresponse. All aspects of the design and analysis should be explicitly described and justified.

This is a group assignment: teams of three people will be formed in class on April 3, and each team is responsible for independently conducting its own study. Cooperation between teams is allowed, but each team must obtain its own sample dataset independently of every other team and prepare its own report on its work.

The first thing that will happen in class on April 3 is that there will be a vote about what procedure we'll use to form the teams. Teams can form voluntarily, or I can randomly assign people to teams.

At the end of class on April 3 (or by 6pm the same day), each team will give me a paragraph describing in a preliminary way the project the team expects to complete. By 11:59pm on April 6, each team will send me a formal proposal describing the design and data gathering or selection plan in more detail (earlier submission of these proposals is encouraged). I will promptly let each team know whether the plan is approved. The lab sessions of April 5 and 12 will be dedicated to work on the projects.

A written report about the project is due by noon on April 24. Each team will submit a single joint report, and all team members will receive the same grade for the report.

On April 17 or 22, each team will make a presentation to the class about the team's project. The quality of the presentation will factor into the overall grade for the project. The specific date on which each team is to present will be randomly assigned by April 10.

On the final exam, questions will refer to and be based on the work done for the project. For example, you may be asked to justify the sampling design and explain why one or more other designs were not used. You may be asked to discuss, very specifically, the relative efficiency (or design effect) of estimates produced from the sample.

Walter Mebane