**Part A.**  **One-sided, Two-Sample z-Test of Population Proportions.** (See LN10.)

In Part A, get your data from the U.S. Census.

A1. *Develop* your own research hypothesis to use in a one-sided two-sample z-test for population proportions. One way to start is to imagine the potential social impacts of a major historical event, such as the Civil War, Industrialization, or the Great Migration. Then, think how such an impact might be reflected in a quantitative form in the Census. What can be measured depends on what questions were asked in Census used at that time. Search on "index of census questions" to find a breakdown of which questions were asked and when.

Finding a good topic requires working back-and-forth between history and data availability. For example, consider the Pike's Peak Gold Rush, occurring 1858-1861 in parts of what became Kansas and Nebraska. In principle, comparing data from the 1850 and 1860 Census could be used to study the impact of that event. However, Kansas and Nebraska would not appear in either of the 1850 or 1860 Census because they had not yet achieved statehood. So, there is some needle-threading that is required.

A2. Explainto a novice how your population proportions would in principle be computed.

A3. Gather data from U.S. Census using original handwritten records. Those are readily accessed using FamilySearch. For each of the two samples, I recommend working from all persons shown on a single sheet of the census, which is usually between 30 and 40 people per sheet. If you are restricting yourself to a smaller groups, say, school-age children, then your working sample sizes will of course be smaller.

Describ*e* what you are doing as you present and process the data. Show the reader where and how you got the data using cropped screenshots. Those images will not be self-explanatory; they must be accompanied with text. Write as if you are interested in the subject and the people, and are addressing someone else who is interested, too.

A4. Walk the reader through the steps of the hypothesis test in the context of your data. As you go, explain how the test progressively answers the question, “How far is far?”

A5. Graph the test, labeling all relevant portions.

A6. Explain how the computation of your particular p-value is connected to the null hypothesis.

A7. As we have stressed, taking a single page from the U.S. Census does not give a "random" sample. What specific problems do you see in the use of a single page? How would a random sample of the same size help overcome those issues? Is it guaranteed to be better?

**Part B. One-sided, Two-Sample t-Test of Population Averages** (see LN9)

In Part B, get your data from Ellis Island or other immigration records. The data site I recommend is presented in Section 4 of LN10.

B1. Develop a research hypothesis to use in a one-sided two-sample test of population averages concerning age, height, or family size. To do so, consider how an event such as a World War, famine, or the shifts in immigration policy in 1920's could produce a change in who immigrates and when. As always, write as if you are interested, and are addressing someone else who is interested, too.

(Note: if you want to use family size, please see me during office hours, soon, as you will need a somewhat different approach.)

B2. Create your own data set consisting of two different samples drawn from immigration data. Use sample sizes between 5 and 70. Show details of the sourcing and context. Show and use images of the manifests of the ships. Explain what you are doing as you go.

B3. State the null hypotheses in words that apply to the particular topic you are addressing. Define the populations you are referencing, their approximate sizes, and how the population averages would (in principle) be computed.

B4. Perform a one-sided, two-sample t-test. Explain what the software is doing on each line of the table of output. In particular, name the sample statistics that are computed and explain how those are connected with population parameters.

B5. Show the graph of the test with all features shown and labeled.

B6. State the conclusion of the test and the grounds. Explain the *reasoning* behind the conclusion.

B7. All serious studies are preceded by a small trial run to look for problems, including lack of clarity in definitions, problems in acquiring data, or a mismatch in what is measured. Thinking of your study as a trial run for a larger study, what problems did you encounter and what changes would you make?

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*