Stat 446 **Name:** _____
In class final
12/13/2016

---

1. You have been hired by a Bozeman area appraisal company. Appraisals are an important element in financing houses, as banks will not lend more many than a house is deemed to be worth. Your new employer currrently has a model that works well in the Bozeman area, but is interested in growing into a national company. To do so, they need to develop a model that is reliable at modeling home prices across the country.

   (a) (5 points) Describe the data that you think will be need to be collected to create a national level appraisal model for predicting home prices across the country. Assuming appraisals require an in person assessment of the home, what logistical and financial constraints will need to be addressed for your data collection scheme.

   (b) (5 points) Describe and justify a sampling scheme to collect the necessary data.

(c) (5 points) Given your sampling scheme and the acquired data, how will you estimate housing prices for properties that require an appraisal?

(d) (5 points) You report your results as a confidence interval, but the bank typically receives a single value from appraisals. Justify the use of the interval rather than a point estimate.

Short Answer Questions:

Answer **only five** of the six short answer questions. Be concise with your answer, 3 or 4 sentences should be sufficient.

1. (4 points) How are post-stratification and regression estimation similar and how are they different?

2. (4 points) We have seen how variance estimation for domains can be challenging due to the fact that the sample size in domain h, n_h, is a random variable. Describe how a bootstrap procedure could be used to calculate the variance within the domains.

3. (4 points) Assume unequal probability sampling is conducted where the probability of selection is proportional to the size of the response. Why is it necessary to adjust our standard SRS estimator: $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$?

4. (4 points) For this class, we have taken samples where the response rate is 1 (all units respond) and the responses are without error. Typically real-world data collection will be more problematic. Describe one common problem that plagues sampling schemes, how you would handle it, and the implications it has for the scope of inference.

5. (4 points) How do the principles of efficient stratified random sampling differ from the principles of efficient cluster sampling?

6. (4 points) Describe a situation where you would implement: SRS, stratified random sampling, and cluster sampling. (Include a different situation for each sampling strategy).