

ST 432 Sampling Theory and Survey Design.

Exam 3 **Modified**

Instructor: Kenneth H. Pollock

Take Home Exam April 15, 2008. Time allowed two weeks. Due Thursday May 1, 2008.

You may use any books or notes and ask me questions of clarification. You may not talk to other students in the class about this exam and evidence of collaboration will result in 0 for both parties. I will be very happy to provide detailed clarification so there is no need for you to even consider collaboration.

This exam is totally my own work.

Name: _____

Signature: _____

-
1. (25 points).
 - (a) First explain briefly using equations and ideas developed in our class how the ideas of double sampling may be used to adjust for nonresponse bias in a mail survey where a telephone followup sample is used.
 - (b) This approach is not always done because of cost. Why would the double sampling method be more costly?
 - (c) There is another problem besides the increased cost with this approach as it is presented in the text. What has been ignored in that analysis?
 2. (25 points) Suppose that a sociologist wants to estimate the total number of retired people in a town based on two stage sampling. At the first stage 4 blocks are sampled at random out of the total of 300 blocks existing in the town. At the second stage 3 households are sampled out of each block. It is also known that there are a total of 3900 households in the town. The data obtained is:

| Blocks | No. Households in Block | No. Households Sampled | No Retired residents |
|--------|-------------------------|------------------------|----------------------|
| 1 | 18 | 3 | 1,0,2 |
| 2 | 14 | 3 | 0,3,0 |
| 3 | 9 | 3 | 1,1,2 |
| 4 | 12 | 3 | 0,1,1 |

Estimate the total number of retired residents in the town and its standard error using the best method available to you. (Note we are not allowing for any nonresponse here).

3. (25 points). **Modified from original.** In class I wrote down the likelihood and then found the MLE for the Lincoln Petersen model. I want you to extend this estimator to the case where there is trap response. Now p_2 is the probability of an unmarked animal being captured in the second sample. Also c_2 is the probability of a marked

animal being recaptured in the second sample. I want you to write down the probability distribution for n_1 and then for m_2 and $(n_2 - m_2)$ conditional on n_1 as before but now modified for trap response making the probabilities of capture slightly different. I want you to show that the MLE of N does not exist in this case when you use $L(N)/L(N-1) = 1$. You will also see that the other MLEs written in

terms of N are $\hat{p}_1 = \frac{n_1}{\hat{N}}$, $\hat{p}_2 = \frac{n_2 - m_2}{\hat{N} - n_1}$, $\hat{c}_2 = \frac{m_2}{n_1}$. The problem is that there are 4

parameters and only three random variables. To get estimates one has to restrict the problem by assuming $p_2 = c_2$ or $p_1 = p_2$. I do not expect you to consider either restriction on the exam though.

4. (25 points) In a pilot study on farm income in a region of third world country an agricultural economist first obtained a random sample of 10 farms out of 90 total farms and ascertained their size in hectares. He then took a subsample of 5 of the 10 farms at random and ascertained their income from a rice crop. The information obtained is presented in the following table and for ease of presentation the 5 farms with income ascertained are presented first.

| Farm No | Size | Income |
|---------|------|--------|
| 1 | 1.2 | 240 |
| 2 | 4.0 | 810 |
| 3 | 3.1 | 625 |
| 4 | 0.5 | 110 |
| 5 | 0.9 | 190 |
| 6 | 2.4 | |
| 7 | 3.4 | |
| 8 | 0.8 | |
| 9 | 1.4 | |
| 10 | 5.1 | |

- (a) First estimate the total farm income in the region first totally ignoring the information about size of farm.
 (b) Next estimate the total farm income based on a ratio estimator where we assume that the total farm size for the region is 180 hectares.
 (c) Next estimate the total farm size based on a more practical ratio estimator where we do not know the total farm size for the region.
 In the interests of time I will not ask for the standard errors.