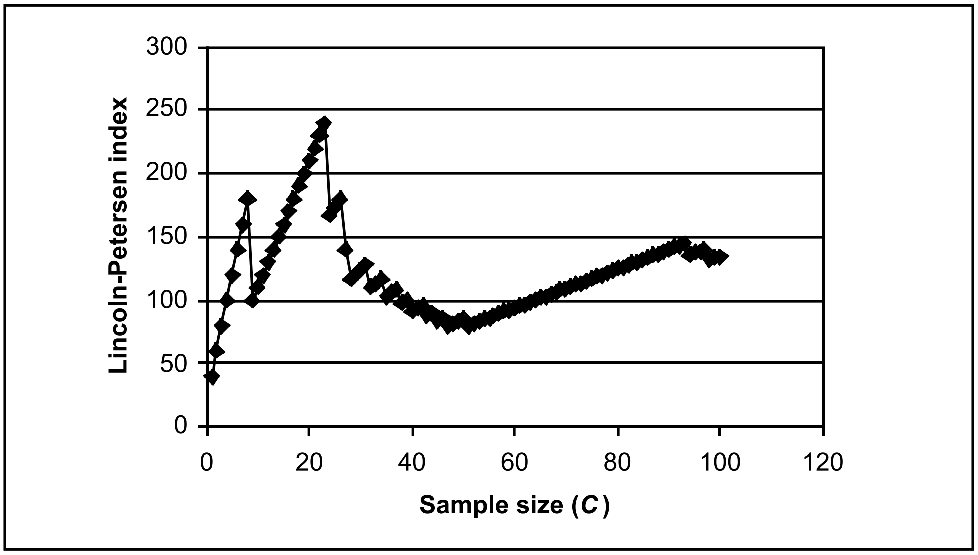
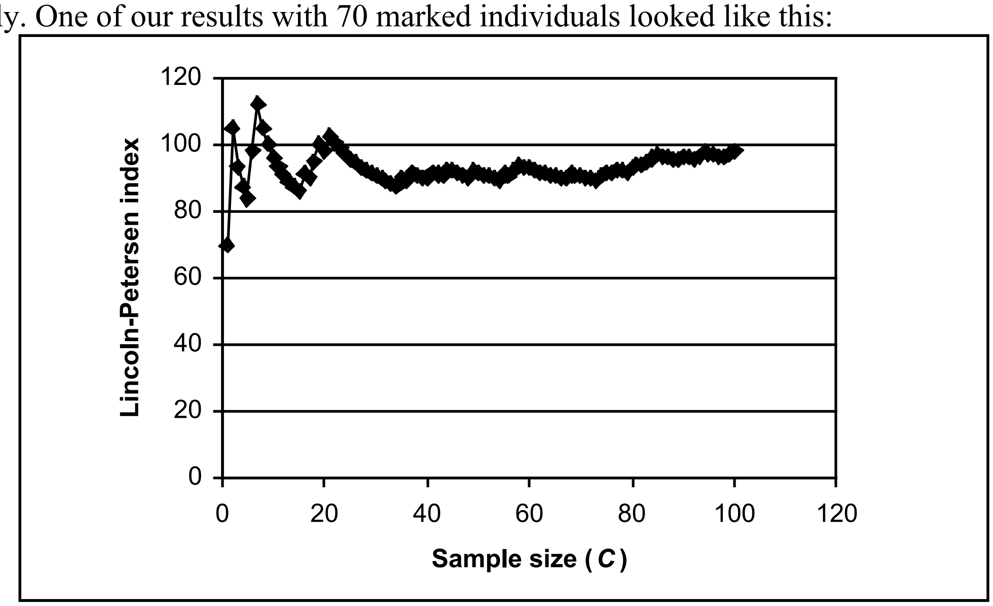
Answers to Exercise 18

Population Estimation and Mark-Recapture Techniques

1. If $M = 20$ and C is low, when you press F9 several times you should see that the Lincoln-Petersen index is poorly estimated and quite variable. Often (but not always) the index becomes somewhat more stable as C increases to 30 or 40. In some cases, the Lincoln-Petersen index will come close to estimating the true population size of 100. In other cases, as shown in the graph below, the index will be off, even after 100 individuals are recaptured. This question should impress on you the need to compute confidence intervals.



2. You should see that as the proportion of individuals in the population that are marked increases, the Lincoln-Petersen index tends to stabilize around the true population size more quickly. One of our results with 70 marked individuals looked like this:



When 100 individuals are marked, the Lincoln-Petersen index is 100, the true population size, because all individuals in the population are marked. The index is 100 no matter how many individuals are recaptured in the second sampling bout.

3. The relationship between *C* and the Lincoln-Petersen index, as developed so far, is sequentially autocorrelated—a value for a specific sample size is correlated with its values in other sample sizes. In other words, sample 99 takes advantage of the sampling up through sample 98; sample 98 takes advantage of the sampling up through sample 97, and so on. Because of this autocorrelation, this method does not reveal the true relationship between *C* and the Lincoln-Petersen index. Hence, we need to push ahead and determine the relationship when the Lincoln-Petersen index is calculated independently for a variety of levels of *C*.

4. Note first that the Lincoln-Petersen index returns only a handful of different estimates under these circumstances; if we mark 20 individuals and recapture 20 individuals, 4 of which are marked, the estimate will always be 84, regardless of the total population size. If 3 marked individuals are recaptured, the estimate will be 105. Under these circumstances, there is no way to get an estimate between 84 and 105, so if you require more sensitive measurements than this (i.e., if you are expecting small fluctuations in population size), you will need to arrange to mark and recapture more individuals when you design the experiment.

   Even if you are anticipating large fluctuations in population size, the Monte Carlo simulation shows that if the population is indeed around 100, the Lincoln Petersen estimate will regulary return estimates from 46.7 to 210 without *any* change in the actual population size. Most likely you will want more reliable estimates than this.

   M and *C* do not seem to have equal influence on the range of estimates returned. When we tried $M = 50$ and $C = 20$, we found that 95% of the estimates fell between 70 and 150, but with $M = 20$ and $C = 50$, the range was still 60 to 204—something to bear in mind during the design phase of your study. Getting meaningful results in this case may actually involve marking *and* resampling more than 50% of the population.

5. Suprisingly, the violations do not appear to drastically alter the range of estimates returned 95% of the time when $M = 50$ and $C = 30$. The mean Lincoln-Petersen index did not appear to differ significantly either (100.47 versus 99.27 by our results). This lack of difference can be explained by the fact that marked and unmarked individuals are equally likely to leave the population (or evade capture), so the ratio of marked to unmarked individuals remaining does not change much. However, the Monte Carlo simulations of the violations appeared to

increase the total range of Lincoln-Petersen estimates and produced a skewed distribution. Our results are shown on the following page; yours will be a bit different. Under different levels of C and M, the violations may be more severe.

	M	N	O	P	Q
32	*M = 50, C = 30, no violations*			*M = 50, C = 30, violations*	
33					
34	Mean	100.4789		Mean	99.26785
35	Standard Error	0.614051		Standard Error	0.607237
36	Median	96.875		Median	96.875
37	Mode	103.3333		Mode	103.3333
38	Standard Deviation	19.418		Standard Deviation	19.20251
39	Sample Variance	377.0586		Sample Variance	368.7364
40	Kurtosis	2.572796		Kurtosis	7.518532
41	Skewness	1.171355		Skewness	1.753205
42	Range	156.8452		Range	193.75
43	Minimum	64.58333		Minimum	64.58333
44	Maximum	221.4286		Maximum	258.3333
45	Sum	100478.9		Sum	99267.85
46	Count	1000		Count	1000
47	Largest(25)	140.9091		Largest(25)	140.9091
48	Smallest(25)	70.45455		Smallest(25)	70.45455

6. There are many ways to modify the model; here is one suggestion using nested functions. Set cell E6 to 1 since the population is closed. Set cell E7 to 1 since this cell represents the probability that both marked and unmarked individuals become trap shy. Add a new probability in cell E8, which will represent the probability that only marked individuals become trap shy. Insert new cells in column E (shift the current cells to the right). Use a nested **IF**, **AND**, and **RAND** formulae in new cell E11:

   =IF(AND(D11="m",RAND()<E8),"m",IF(D11="u","u","."))

   The first section, **IF(AND(D11="m",RAND()<E8),"m"**, tells the spreadsheet to evaluate the two conditions in the **AND** formula: Condition 1 is that cell D11 is **"m"** (the individual sampled is marked), and condition 2 is that that a random number is less than the value in cell E8 (which indicates that the individual could be recaptured). If both of these conditions are met, the program returns the letter m, otherwise it walks through the second IF statement, **IF(D11="u","u",".")**. This statement returns the letter "u" if the individual in cell D11 is unmarked, and returns a missing value for any other condition. The missing value (.) will be generated for marked individuals that have become trap-shy.