Stat 549

Fall 2014

Homework 5/Final Exam: due Friday, Dec. 12 by 5 pm.

This is to be done individually using only class notes and the textbook. You are welcome to consult with me, but no one else. Show your work on all calculations.

1. Consider a strip adaptive sampling scheme to estimate the number of objects in the 10x10 grid below. Label the vertical strips 1 to 10 from left to right. The initial sample will be a systematic sample of two vertical strips with a random starting point from 1 to 5; e.g., strips 1 and 6 would be one possible sample and strips 4 and 9 would be another. The neighborhood of a plot is the plot itself plus the four plots above, below, and to the right and left of the plot. If any selected plot has at least one object, then its neighborhood is included in the sample.

					1				
					14	1			
				1	15	1			
2									
26	1					3	1	1	
3						19	4	25	2

(a) Calculate the inclusion probabilities for all the non-zero networks. Remember to treat this as a systematic sample of size 1 and not as an SRS of two transects.

(b) Derive the sampling distribution of the Horvitz-Thompson estimator of τ, the total number of objects. Show that the estimator is unbiased (τ=120) and compute its standard deviation.

(c) Derive also the distribution of v across all possible initial samples where v is the total number of distinct plots which must be surveyed (including edge units). Compute its expected value.

(d) Calculate the standard deviation of the unbiased estimator of the total for an SRS of E(v) plots (with no adaptive component).

(e) Compare the standard deviation of the H-T estimator in part (b) to the standard deviation of the SRS estimator in part (d). The assumption is that the expected cost of the adaptive plan of part (b) is the same as the cost of the SRS plan of part (d) so that this is a fair comparison of their efficiencies. Is that a valid assumption?

2. It was desired to estimate the proportion of a large tract of land that could be classified as wetland. The area was first stratified based on aerial photographs into three terrain types and then the proportion of wetland estimated within each. In the first stratum, comprising 50% of the area, a random sample of 100 points was chosen and the points classified: 30 were classified as wetland. In the second stratum, comprising 30% of the area, a systematic sample of transects with random starting point was used; the estimated proportion of wetland in this stratum was .42 with SE = .08. In the third stratum, comprising the remaining 20% of the area, a two-stage sample with transects as the primary units and points along the transects as secondary units was used; the estimated proportion of wetland was .10 with SE = .04. Estimate the proportion of the entire area that is wetland and compute an SE for this estimate.

3. The National Marine Fisheries Service places observers on commercial fishing vessels. One of their jobs is to sample the hauls of fish for numbers and species of fish. One method of sampling is basket sampling where baskets of fish are taken from the haul as it is unloaded onto the ship. Suppose one particular haul of fish weighs 10,000 kg. Six baskets of fish are selected randomly from the haul. The number and total weight of the fish in each basket are recorded.

(a) How would you estimate the total number of fish in the whole haul and a standard error from the given information? Give the formulas, being careful to specify precisely what each variable represents.

(b) The number of salmon in each basket is also recorded. How would you estimate the proportion and SE of all the fish in the haul that are salmon? Again, give the formulas, being careful to specify precisely what each variable represents.

4. MATH ONLY: The linearization approximation for the variance of the product of two random variables is $\text{Var}(XY) \approx \mu_Y^2\sigma_X^2 + \mu_X^2\sigma_Y^2 + 2\mu_X\mu_Y\rho\sigma_X\sigma_Y$. Suppose X and Y are independent. Derive an exact expression for the variance of XY in terms of the means and variances of X and Y. Note that it is not the same as the linearization approximation with $\rho = 0$.

5. Suppose I'm interested in the total number of ducklings produced in a large breeding region in North Dakota. I have an estimate, say P, of the total number of ponds in the entire region based on aerial photos. I also have a standard error, SE(P) for this estimate. In addition, based on ground surveys independent of the aerial photo survey, I have an estimate of the

average number of breeding pairs, say B, per pond, along with SE(B). Finally, from a different study reported in the literature, I have an estimate of the average number of ducklings per pair, say D, along with SE(D). Estimate the total number of ducklings in the area, along with a standard error for the following data: P = 23890, SE(P) = 1122; B =2.42, SE(B) = 0.43; D = 5.65, SE(D) = 0.51 (Hint: use the linearization approximation given in the previous problem twice. You could also use the exact expression derived in the previous problem since the random variables here are independent.)

6. MATH ONLY: From Chapter 16, derive the expression for the true variance of $\widehat{\tau}$ for simple random sampling with estimated detectability (equation (16.9) on p. 221 of Thompson; 2nd ed: eq. (9) on p. 191). Thompson gives some hints on p. 223 (2nd ed: p. 194).