MH4511 Sampling & Survey

Tutorial 10 Solution AY2025/26 Semester 1

Problem 10.1 (Solution)

In this problem:

$n_1 = 500$ is the size of the first sample,

$n_2 = 300$ is the size of the second sample, and

$m = 120$ is the number of marked fish caught in the second sample.

a) Using the ratio estimator,

$$\hat{N} = \frac{n_1 n_2}{m} = \frac{500 \times 300}{120} = 1250$$

and, ignoring fpc,

$$\widehat{Var}(\hat{N}) \approx \frac{n_1^2\, n_2\, (n_2 - m)}{m^3}$$

$$= \frac{500^2 \times 300 \times (300 - 120)}{120^3} = 7812.5$$

A 95% CI for N can be,

$$1250 \pm 1.96 \times \sqrt{7812.5} = 1250 \pm 173 = (1077,\ 1423)$$

Alternatively, if we include fpc,

$$\widehat{Var}(\hat{N}) \approx \left(1 - \frac{m}{n_1}\right) \frac{n_1^2\, n_2\, (n_2 - m)}{m^3}$$

$$= \left(1 - \frac{120}{500}\right) \times 7812.5 = 5937.5$$

A 95% CI for N can be,

$$1250 \pm 1.96 \times \sqrt{5937.5} = 1250 \pm 151 = (1099,\ 1401)$$

b) Using the contingency table approach,

		In Sample 2?		
		Yes	No	Total
In Sample 1?	Yes	$120(=m)$	380	$500(=n_1)$
	No	180	x_{22}^*	x_{2+}^*
	Total	$300(=n_2)$	x_{+2}^*	$x_{++}^*(=N)$

$$\widehat{N} = \frac{n_1 n_2}{m} = \frac{500 \times 300}{120} = 1250$$

A 95% CI = (1116, 1420), with the computation below.

Using chi-square test, with $u = 436$, the expected frequencies are

		In Sample 2?		
		Yes	No	Total
In Sample 1?	Yes	$134.41(120)$	$365.59(380)$	500
	No	$165.59(180)$	$450.41(436)$	616
	Total	300	816	1116

The chi-square statistic is

$$\chi^2 = \sum_{all\ cells} \frac{(observed - expected)^2}{expected}$$

$$= \frac{(120 - 134.41)^2}{134.41} + \cdots + \frac{(436 - 450.41)^2}{450.41}$$

$$= 1.545 + 0.568 + 1.254 + 0.461 = 3.827$$

With p-value for $\Pr(\chi_1^2 > 3.827) = 0.0504$

Similarly, with $u = 740$, the expected frequencies are

		In Sample 2?		
		Yes	No	Total
In Sample 1?	Yes	$105.63(120)$	$394.37(380)$	500
	No	$194.37(180)$	$725.63(740)$	920
	Total	300	1120	1420

The chi-square statistic is

$$\chi^2 = \sum_{all\ cells} \frac{(observed - expected)^2}{expected}$$

$$= \frac{(120 - 105.63)^2}{105.63} + \cdots + \frac{(740 - 725.63)^2}{725.63}$$

$$= 1.954 + 0.523 + 1.062 + 0.284 = 3.823$$

with p-value for $\Pr(\chi_1^2 > 3.823) = 0.0505$

Problem 10.2 (Solution)

a) $n_1 = 311$ is the number of unique genotypes in the 2001 sample,
$n_2 = 239$ is the number of unique genotypes in the 2002 sample, and
$m = 165$ is the number of unique genotypes in both samples
Hence,

$$\hat{N} = \frac{n_1 n_2}{m} = \frac{311 \times 239}{165} \approx 450$$

And, with fpc,

$$\widehat{Var}(\hat{N}) \approx \left(1 - \frac{m}{n_1}\right) \times \frac{n_1^2\, n_2\, (n_2 - m)}{m^3}$$

$$= \left(1 - \frac{165}{311}\right) \times \frac{311^2 \times 239 \times (239 - 165)}{165^3} = 178.8$$

A 95% CI for N can be

$$450 \pm 1.96 \times \sqrt{178.8} = 450 \pm 26 \approx (424,\ 476)$$

We would need the following assumptions:

- The population is closed –no bears enter or leave the area between the samples. That is, N is the same for each sample.
- Each sample of bears' feces is an SRS from the population. That is, each bear's feces is equally likely to be collected in a sample.
- The two samples are independent. The bears from the first sample become re-mixed in the population, so that the marking status of a bear is unrelated to the probability that the bear is selected in the second sample.

b) $n_1 = 56$ is the number of bears with radio transmitter in 2001,
$n_2 = 311$ is the number of unique genotypes in the 2001 sample, and
$m = 36$ is the number of unique genotypes with radio transmitter
Hence,

$$\hat{N} = \frac{n_1 n_2}{m} = \frac{56 \times 311}{36} \approx 484$$

And, with fpc,

$$\widehat{Var}(\hat{N}) \approx \left(1 - \frac{m}{n_1}\right) \times \frac{n_1^2\, n_2\, (n_2 - m)}{m^3}$$

$$= \left(1 - \frac{36}{56}\right) \times \frac{56^2 \times 311 \times (311 - 36)}{36^3} = 2053.1$$

A 95% CI for N can be

$$484 \pm 1.96 \times \sqrt{2053.1} = 484 \pm 88.8 \approx (395,\ 572)$$

c) $n_1 = 57$ is the number of bears with radio transmitter in 2002,
$n_2 = 239$ is the number of unique genotypes in the 2002 sample, and
$m = 28$ is the number of unique genotypes with radio transmitter
Hence,

$$\hat{N} = \frac{n_1 n_2}{m} = \frac{57 \times 239}{28} \approx 487$$

And, with fpc,

$$\widehat{Var}(\hat{N}) \approx \left(1 - \frac{m}{n_1}\right) \times \frac{n_1^2\, n_2\, (n_2 - m)}{m^3}$$

$$= \left(1 - \frac{28}{57}\right) \times \frac{57^2 \times 239 \times (239 - 28)}{28^3} = 3797.3$$

A 95% CI for N can be

$$487 \pm 1.96 \times \sqrt{3797.3} = 487 \pm 120 \approx (367,\ 607)$$

Problem 10.3 (Solution)

Note that $L(N)$ follows a hypergeometric distribution.

Consider

$$L(N) \geq L(N-1)$$

$$=> \frac{\binom{n_1}{m}\binom{N-n_1}{n_2-m}}{\binom{N}{n_2}} \geq \frac{\binom{n_1}{m}\binom{N-1-n_1}{n_2-m}}{\binom{N-1}{n_2}}$$

$$=> \frac{\frac{(N-n_1)!}{(n_2-m)!\,(N-n_1-n_2+m)!}}{\frac{(N)!}{(n_2)!\,(N-n_2)!}} \geq \frac{\frac{(N-1-n_1)!}{(n_2-m)!\,(N-1-n_1-n_2+m)!}}{\frac{(N-1)!}{(n_2)!\,(N-1-n_2)!}}$$

$$=> \frac{\frac{(N-n_1)}{(N-n_1-n_2+m)}}{\frac{(N)}{(N-n_2)}} \geq 1$$

$$=> \frac{(N-n_1)}{(N-n_1-n_2+m)} \geq \frac{(N)}{(N-n_2)}$$

$$=> (N-n_1)(N-n_2) \geq (N)(N-n_1-n_2+m)$$

$$=> n_1 n_2 \geq Nm$$

$$=> N \leq \frac{n_1 n_2}{m}$$

This shows that for integers $N > 0$, $L(N)$ is increasing (non-decreasing) until $N > \frac{n_1 n_2}{m}$. So, the maximum is attained at (or around) $\frac{n_1 n_2}{m}$.