Serik Sagitov, Chalmers and GU, January 19, 2018

Solutions chapter 7

Problem 7.1

We consider sampling with replacement. For an answer in the case of sampling without replacement consult the book page A36.

Population distribution

Values	1	2	4	8
Probab.	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{1}{5}$	$\frac{1}{5}$

Population mean and variance: $\mu = 3.4$, $\sigma^2 = 6.24$. The list of $\bar{X}$ values and their frequencies for $n = 2$ observations taken with replacement:

	1	2	4	8	Total prob.
1	1.0 (1/25)	1.5 (2/25)	2.5 (1/25)	4.5 (1/25)	1/5
2	1.5 (2/25)	2.0 (4/25)	3.0 (2/25)	5.0 (2/25)	2/5
4	2.5 (1/25)	3.0 (2/25)	4.0 (1/25)	6.0 (1/25)	1/5
8	4.5 (1/25)	5.0 (2/25)	6.0 (1/25)	8.0 (1/25)	1/5
Tot. prob.	1/5	2/5	1/5	1/5	1

The sampling distribution of $\bar{X}$:

Values	1	1.5	2	2.5	3	4	4.5	5	6	8
Probab.	$\frac{1}{25}$	$\frac{4}{25}$	$\frac{4}{25}$	$\frac{2}{25}$	$\frac{4}{25}$	$\frac{1}{25}$	$\frac{2}{25}$	$\frac{4}{25}$	$\frac{2}{25}$	$\frac{1}{25}$

$$\mathrm{E}(\bar{X}) = 3.4 = \mu, \quad \mathrm{E}(\bar{X}^2) = 14.68, \quad \mathrm{Var}(\bar{X}) = 3.12 = \frac{\sigma^2}{n}.$$

Problem 7.9

Dichotomous data

$$n = 1500, \quad \hat{p} = 0.55, \quad 1 - \hat{p} = 0.45.$$

Population margin of victory

$$v = p - (1 - p) = 2p - 1.$$

Estimated margin of victory

$$\hat{v} = \hat{p} - (1 - \hat{p}) = 2\hat{p} - 1 = 0.1.$$

(a) Standard error of $\hat{v}$

$$s_{\hat{v}} = 2s_{\hat{p}} = 0.026,$$

where

$$s_{\hat{p}}^2 = \frac{\hat{p}(1-\hat{p})}{n-1} = \frac{0.55 \times 0.45}{1499} = 0.013^2.$$

(b) Approximate 95% CI for v is

$$\hat{v} \pm 1.96 s_{\hat{v}} = 0.10 \pm 0.05.$$

Problem 7.19

Normal approximation: $\frac{\bar{X}-\mu}{s_{\bar{X}}}$ is asymptotically N(0,1)-distributed. Approximate 95% one-sided CI for μ is obtained from

$$0.95 = \mathrm{P}(\frac{\bar{X}-\mu}{s_{\bar{X}}} < 1.645) = \mathrm{P}(\bar{X} - 1.645 s_{\bar{X}} < \mu < \infty).$$

Problem 7.35

Simple random sample:

$$N = 2000, \quad n = 25, \quad \sum X_i = 2451, \quad \sum X_i^2 = 243505.$$

(a) Unbiased estimate of μ is

$$\bar{X} = \frac{2451}{25} = 98.04.$$

(b) Unbiased estimate of σ^2 is

$$\frac{N-1}{N} s^2 = \frac{1999}{2000} 133.71 = 133.64,$$

where

$$s^2 = \frac{n}{n-1}(\overline{X^2} - \bar{X}^2) = \frac{25}{24}(\frac{243505}{25} - 98.04^2) = 133.71.$$

Unbiased estimate of $\mathrm{Var}(\bar{X})$ is

$$s_{\bar{X}}^2 = \frac{s^2}{n}(1 - \frac{n}{N}) = 5.28.$$

(b) Approximate 95% CI for μ is

$$\bar{X} \pm 1.96 s_{\bar{X}} = 98.04 \pm 1.96\sqrt{5.28} = 98.04 \pm 4.50.$$

Approximate 95% CI for the total $N\mu$ is

$$N\bar{X} \pm 1.96 N s_{\bar{X}} = 196080 \pm 9008.$$

Problem 7.36

Simple random sample. Take $\bar{X}^2$ as a point estimate of μ^2. (Note that this is a method of moments estimate.) Compute the bias of this estimate.

Solution. The bias is

$$\mathrm{E}(\bar{X}^2) - \mu^2 = \mathrm{E}(\bar{X}^2) - (\mathrm{E}\bar{X})^2 = \mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}(1 - \frac{n-1}{N-1}).$$

For large n, the bias is small.

Problem 7.28

Randomized response method. Estimate the proportion q of illegal drug users among prison inmates. We are interested in the population as a whole not in punishing particular individuals.

Randomly chosen n inmates responded to a randomized statement:

"I use heroin" (with probability 5/6)
"I do not use heroin" (with probability 1/6).

Consider

$$X = \text{ number of "yes" responses for } n \text{ inmates}, \quad R = X/n.$$

Then X has Bin (n, p) distribution, where

$$p = \mathrm{P}(\text{a "yes" answer}) = \frac{5}{6} \cdot q + \frac{1}{6} \cdot (1 - q) = \frac{1 + 4q}{6}.$$

Method of moments estimate $\tilde{q}$ for the population proportion q is found from

$$\mathrm{E}(X) = np, \quad \mathrm{E}(R) = \frac{1 + 4q}{6}.$$

Setting

$$R = \frac{1 + 4\tilde{q}}{6},$$

we find

$$\tilde{q} = \frac{6R - 1}{4}.$$

The estimate is unbiased

$$\mathrm{E}(\tilde{q}) = \frac{6p - 1}{4} = q.$$

Its variance equals

$$\mathrm{Var}(\tilde{q}) = \frac{9}{4} \cdot \mathrm{Var}(R) = \frac{9}{4} \cdot \frac{p(1 - p)}{n} = \frac{(1 + 4q)(5 - 4q)}{16n}.$$

Take for example $n = 40$, $X = 8$. Then $R = 0.2$ and $\tilde{q} = \frac{6R-1}{4} = 0.05$. The estimated standard error

$$s_{\tilde{q}} = \sqrt{\frac{(1 + 4\tilde{q})(5 - 4\tilde{q})}{16n}} = 0.095.$$

The estimate is unreliable. We have to increase the sample size.

Problem 7.53

μ_l	5.4	16.3	24.3	34.5	42.1	50.1	63.8
W_l	0.20	0.23	0.19	0.17	0.08	0.06	0.07
σ_l	8.3	13.3	15.1	19.8	24.5	26.0	35.2
Optimal allocation $n\frac{W_l\sigma_l}{\bar{\sigma}_l}$	10	18	17	19	12	9	15
Proportional allocation nW_l	20	23	19	17	8	6	7

(a) $N = 2010$, $L = 7$, $n = 100$, $\bar{\sigma} = 17.04$, , $\overline{\sigma^2} = 347.40$, $\sigma^2 = 347.40 + 275.33 = 622.73$.

(b) $\text{Var}(\bar{X}_{so}) = \frac{\bar{\sigma}^2}{n} = 2.90$, $\text{Var}(\bar{X}_{sp}) = \frac{\overline{\sigma^2}}{n} = 3.44$, $\text{Var}(\bar{X}) = \frac{\sigma^2}{n} = 6.21$.

(c) $\mu = 26.49$, $\sigma = 24.95$.

(d) If $n_1 = \ldots = n_7 = 10$ and $n = 70$, then $\text{Var}(\bar{X}_s) = 4.45$. Find sample size x such that $\text{Var}(\bar{X}) = \frac{\sigma^2}{x} = 4.45$. The solution gives the sample size of 140.

(e) If $n = 70$, then $\text{Var}(\bar{X}_{sp}) = 4.92$. Find sample size x such that $\text{Var}(\bar{X}) = \frac{\sigma^2}{x} = 4.92$. The solution gives the sample size of 127.

Problem 7.57

Stratified population with

$$N = 5, \quad L = 2, \quad W_1 = 0.6, \quad W_2 = 0.4, \quad \mu_1 = 1.67, \quad \mu_2 = 6, \quad \sigma_1^2 = 0.21, \quad \sigma_2^2 = 4.$$

Given $n_1 = n_2 = 1$ and $n = 2$, the sampling distribution of the stratified sample mean $\bar{X}_s = 0.6X_1 + 0.4X_2$ is

	1	2	Total prob.
4	2.2 (1/6)	2.8 (2/6)	1/2
8	3.8 (1/6)	4.4 (2/6)	1/2
Tot. prob.	1/3	2/3	1

$$\text{E}(\bar{X}_s) = 3.4 = \mu, \quad \text{E}(\bar{X}_s)^2 = 12.28, \quad \text{Var}(\bar{X}_s) = 0.72 = 0.36\sigma_1^2 + 0.16\sigma_2^2.$$