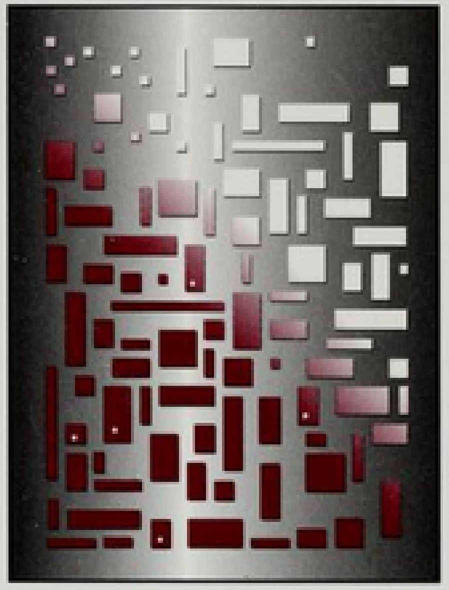
Student Solutions Manual
for Scheaffer, Mendenhall, and Ott's



Elementary Survey Sampling

SIXTH EDITION

Richard L. Scheaffer

DUXBURY ADVANCED SERIES

Duxbury

Student Solutions Manual

for

Scheaffer, Mendenhall, and Ott's

Elementary Survey Sampling

Sixth Edition

Richard L. Sheaffer
University of Florida, Emeritus

THOMSON
BROOKS/COLE

Australia • Canada • Mexico • Singapore • Spain • United Kingdom • United States

Printed in the United States of America
1 2 3 4 5 6 7 09 08 07 06 05

Printer: Malloy Incorporated

ISBN: 0-495-01017-0

For more information about our products, contact us at:
Thomson Learning Academic Resource Center
1-800-423-0563

For permission to use material from this text or product, submit a request online at **http://www.thomsonrights.com.**
Any additional questions about permissions can be submitted by email to **thomsonrights@thomson.com.**

Thomson Higher Education
10 Davis Drive
Belmont, CA 94002-3098
USA

Asia (including India)
Thomson Learning
5 Shenton Way
#01-01 UIC Building
Singapore 068808

Australia/New Zealand
Thomson Learning Australia
102 Dodds Street
Southbank, Victoria 3006
Australia

Canada
Thomson Nelson
1120 Birchmount Road
Toronto, Ontario M1K 5G4
Canada

UK/Europe/Middle East/Africa
Thomson Learning
High Holborn House
50–51 Bedford Road
London WC1R 4LR
United Kingdom

Latin America
Thomson Learning
Seneca, 53
Colonia Polanco
11560 Mexico
D.F. Mexico

Spain (including Portugal)
Thomson Paraninfo
Calle Magallanes, 25
28015 Madrid, Spain

Table of Contents

CHAPTER 2

ELEMENTS OF THE SAMPLING PROBLEM

2.1 An adequate frame listing individuals in a city is difficult to obtain. For that reason, and because data is desired on a family basis, it would be better to sample dwelling units. An adequate frame for dwelling units is also difficult to obtain, so a cluster sampling approach could be used by sampling city blocks and then measuring water consumption for the families living in the sampled blocks.

2.3 The sampling design depends on a careful definition of the population of interest. As it would be almost impossible to get a listing of all cars owned by residents of a city, a better option would be to restrict the population of cars to something like "cars that use city parking lots on a working day" or "cars that belong to people visiting the malls on a weekend." Then, a listing of parking lots or sections of parking lots could serve as frames for collections of cars.

2.5 An area as large as a state is generally broken up into smaller areas, such as counties and farms within counties, for sampling. Each county may contain a number of farms, so there are various sampling options. Counties could be viewed as strata, with farms being sampled from each. If there are many counties, one might sample counties as clusters of farms and then sample farms from each sampled county. In either of these scenarios a list of farms by county would be needed as a frame.

2.7 **(a)** A telephone survey would be the only way to cover the country with a well designed sampling plan in a reasonable time.

(b) If the population is defined as subscribers to the paper, then a mailed questionnaire or interviews could be used. If the population is less well defined to include all readers or potential readers, than a telephone survey with random digit dialing may have to be used.

(c) Homeowners are a well-defined group, and a sample could be contacted through either mailed questionnaires or personal interviews, although the latter would be time consuming. Telephone interviews could also be used, and random digit dialing would not be necessary.

(d) Assuming dogs are registered, it should be relatively easy to sample from the list of registered owners and obtain the survey information by either telephone

5.37 (a) This is a stratified random sample with three strata, all having the same sample size (340). Because the populations are large and approximately equal in size we can ignore the finite population corrections and assume that the stratum relative sizes, N_i/N, are all approximately 1/3. The estimate of the population proportion then becomes:

$$\hat{p}_{st} = \frac{1}{N}(N_1\hat{p}_1 + N_2\hat{p}_2 + N_3\hat{p}_3) = \frac{1}{3}\left(\frac{214}{340} + \frac{249}{340} + \frac{261}{340}\right) =$$

$$\frac{1}{3}\left(\frac{724}{340}\right) = \frac{724}{1020} = .725$$

The estimated variance becomes:

$$V(\hat{p}_{st}) = \frac{1}{N^2}[N_1^2\,V(\hat{p}_1) + N_2^2\,V(\hat{p}_2) + N_3^2\,V(\hat{p}_3)] =$$

$$\frac{1}{3^2}\left[\frac{.629(.371)}{339} + \frac{.732(.268)}{339} + \frac{.768(.232)}{339}\right] = \frac{.001793}{3^2}$$

The margin of error for this estimate is:

$$2\sqrt{V(\hat{p}_{st})} = 2\sqrt{\frac{.001793}{9}} = 2(.014) = .028$$

The plausible values for the true proportion of residents that recycled over the past month are in the interval .725±.028 or (.697, .753).

(b) This question calls for a confidence interval on a difference based on two independent proportions. The interval estimate of the population difference is:

$$(.768 - .629) \pm 2\sqrt{\frac{.768(.232)}{339} + \frac{.629(.371)}{339}} \text{ or}$$

$$.139 \pm 2(.035) \text{ or } .139 \pm .070$$

Because this interval does not include zero, there is evidence to say that there is a significant increase in the proportion who recycle as we move for the stratum with low educational effort to the one with high educational effort.

(c) Again, the question calls for an estimate of the difference based on two independent proportions. The interval estimate of the population difference is:

$$(.768 - .732) \pm 2\sqrt{\frac{.768(.232)}{339} + \frac{.732(.268)}{339}} \text{ or}$$

$$.036 \pm 2(.033) \text{ or } .036 \pm .066$$

Because this interval does include zero, there is no evidence of a real difference between the recycling proportions for stratum 2 and 3. We cannot reject a claim that the medium educational effort does as well as the high educational effort.

(d) Using the same methodology as in part (a), the estimate of the proportion, the estimate of variance, and margin of error turn out to be as shown below:

$$\hat{p}_{st} = \frac{1}{N}(N_1\hat{p}_1 + N_2\hat{p}_2 + N_3\hat{p}_3) = \frac{1}{3}\left(\frac{211}{340} + \frac{225}{340} + \frac{255}{340}\right) =$$

$$\frac{1}{3}\left(\frac{691}{340}\right) = \frac{691}{1020} = .677$$

$$V(\hat{p}_{st}) = \frac{1}{N^2}[N_1^2\,V(\hat{p}_1) + N_2^2\,V(\hat{p}_2) + N_3^2\,V(\hat{p}_3)] =$$

$$\frac{1}{3^2}\left[\frac{.62(.38)}{339} + \frac{.662(.338)}{339} + \frac{.75(.25)}{339}\right] = \frac{.001908}{3^2}$$

$$2\sqrt{V(\hat{p}_{st})} = 2\sqrt{\frac{.001908}{9}} = 2(.015) = .030$$

The plausible values for the true proportion of residents who find recycling at least somewhat convenient are in the interval .677±.030 or (.647, .707).

(e) This calls for the estimate of a difference based on two independent proportions. The interval estimate of the difference in population proportions is:

$$(.750 - .620) \pm 2\sqrt{\frac{.75(.25)}{339} + \frac{.62(.38)}{339}} \text{ or}$$

$$.130 \pm 2(.035) \text{ or } .130 \pm .070$$

The plausible values for the difference in population proportions lie in the interval (.06, .20). The proportion who think it is somewhat or very convenient to recycle is higher in the high education stratum by at least .06.

(f) This question calls for an estimate of a difference in population proportions based on dependent sample proportions. (They are dependent because the both come from the same sample, the sample of stratum 1.) The interval estimate is:

(d) For Exercise 7.20 (a),

Successive Differences (d_i)

465 161 498 29 587 468 149 397

$$\sum d_i^2 = 1234{,}394$$

$$\hat{V}_d(\hat{\tau}_{sy}) = N^2\hat{V}_d(\bar{y}_{sy}) = N^2 \frac{N-n}{nN}\frac{1}{2(n-1)}\sum d_i^2$$

$$= (41)^2 \frac{41-9}{9(41)}\frac{1}{2(8)}(1234394) = 11246701$$

$$\hat{V}(\hat{\tau}_{sy}) = N^2 \frac{s^2}{n}\left(\frac{N-n}{N}\right) = (41)^2 \frac{115959}{9}\left(\frac{41-9}{41}\right) = 16904245$$

For Exercise 7.20 (b),

Successive Differences (d_i)

.9 1.3 4.3 1.0 3.8 1.3 .1 .9

$$\sum d_i^2 = 38.94$$

$$\hat{V}_d(\bar{y}_{sy}) = \frac{N-n}{nN}\frac{1}{2(n-1)}\sum d_i^2 = \frac{41-9}{9(41)}\frac{1}{2(8)}(38.94) = .21$$

$$\hat{V}(\bar{y}_{sy}) = \frac{s^2}{n}\left(\frac{N-n}{N}\right) = \frac{16.0461}{9}\left(\frac{41-9}{41}\right) = 1.39$$

(e) For Exercise 7.21 ,

Successive Differences (d_i)

8 16 86 229 328 153 10 15

$$\sum d_i^2 = 191475$$

$$\hat{V}_d(\hat{\tau}_{sy}) = N^2\left(\frac{N-n}{nN}\right)\frac{1}{2(n-1)}\sum d_i^2$$

$$= (41)^2\left(\frac{41-9}{9(41)}\right)\frac{1}{2(8)}(191475) = 1744550$$

$$\hat{V}(\hat{\tau}_{sy}) = N^2 \frac{s^2}{n}\left(\frac{N-n}{N}\right) = (41)^2 \frac{140009}{9}\left(\frac{41-9}{41}\right) = 2.04102 \times 10^7$$

For the first four situations the difference method gives approximately the same estimated variance as the standard simple random sampling result; there are no pronounced trends in these data. In the last situation the divorce rates have a very

strong trend and the estimated variance from the difference method is much smaller (and probably more realistic) than the one from simple random sampling.

7.29 In such situations systematic sampling is often used merely for convenience. If, however, there is a trend in number of unemployed adults per dwelling along the street (perhaps the street runs from affluent neighborhoods to those of lower economic status), then systematic sampling can actually improve the precision of the results.

The right hand column of the table shows the 15 estimates of the mean count per cell. Because all of these are equally likely in random sampling, their expected value is simply their average, which is 1.5, the population mean count per cell.

CHAPTER 11

SUPPLEMENTAL TOPICS

11.1 $k = 8$

$\bar{y}_1 = 322.6$ $\quad\bar{y}_5 = 404.6$
$\bar{y}_2 = 345.8$ $\quad\bar{y}_6 = 593.8$
$\bar{y}_3 = 493.8$ $\quad\bar{y}_7 = 584.8$
$\bar{y}_4 = 224.0$ $\quad\bar{y}_8 = 287.6$

$\sum \bar{y}_i = 3257.0$

The estimated mean $\bar{y}$ is, given in Equation (11.2),

$$\bar{y} = \frac{1}{k}\sum \bar{y}_i = \frac{3257.0}{8} = 407.125$$

The estimated variance of $\bar{y}$ is, given in Equation (11.3), then becomes

$$\hat{V}(\bar{y}) = \left(\frac{N-n}{N}\right)\frac{s_k^2}{k}$$

$$= \left(\frac{545-40}{545}\right)\frac{137.7^2}{8}$$

with $B = 2\sqrt{\hat{V}(\bar{y})} = 93.7$

11.3 $N = 95,\ n = 20,\ n_1 = 16$

$\sum y_{1j} = 377.78$

The estimator of the population mean is $\bar{y}_1$, given by Equation (11.5), which yields an estimate of

$$\bar{y}_1 = \frac{1}{n_1}\sum y_{1j} = \frac{377.78}{16} = 23.61$$

The quantity $(N_1 - n_1)/N_1$ must be estimated by $(N-n)/N$, since N_1 is unknown. The estimate of $\bar{y}_1$, given in Equation (11.6), then becomes