S Supplement: Additional R-Macros

Estimate population means, proportions, totals, and ratios, and compute the corresponding bound on the error of estimation. Also, determine the necessary sample sizes. The macros to be used are "stratified.survey", "cluster.survey" and "ratio.survey".

Caution: In general, do not feed these macros more information than is necessary. For example, when using the macro "stratified.survey", do not enter the sample standard deviation "stand.dev" or the sample size "n" if the original data set is entered. Also, when using "stratified.survey", do not enter values for both "error.bound.mu" and "error.bound.tau" simultaneously.

S.1 Simple Random Sampling from Chapter 4

Simple random sampling is a special case of stratified sampling, using only one stratum. Hence, we will use "stratified.survey". The exercises begin on p. 106.

Exercise 4.19: Compute $\hat{\mu}$ and B from the data set.

> z = scan2("EXER4_19.DAT", T)

> y = z[, 2] # number of cavities

> stratified.survey(y, N = 1000)

$design

[1] "simple random sample"

$mu.hat

[1] 2

$var.mu.hat

[1] 0.22

$error.bound.mu.based.on.data

[1] 0.9380832

$ci.mu

[1] 1.061917 2.938083

$tau.hat

[1] 2000

$var.tau.hat

[1] 220000

$error.bound.tau.based.on.data

[1] 938.0832

$ci.tau

[1] 1061.917 2938.083

Note: Answer the questions by highlighting the answers (as done above), or write the following:

"The estimate of the population mean is 2, and the bound on the error of estimation is 0.9380832."

Exercise 4.23: Compute $\hat{\mu}$ and B from the summary statistics.

> stratified.survey(2.1, 0.4, 20, N = 200)

"The estimate of the population mean is 2.1 seconds, and the bound on the error of estimation is 0.1697056 seconds."

Exercise 4.24 (with modified numbers): *Sample size determination involving μ.* In Exercise 4.23, how large a sample should be taken in order to estimate μ with a bound of 0.1 seconds on the error of estimation? Use 0.4 seconds as an approximation of the population standard deviation.

> stratified.survey(2.1, 0.4, 20, N = 200, error.bound.mu = 0.1)

"The sample size needed is 48.6692."

Exercise 4.27: Compute $\hat{\tau}$ and B from the summary statistics.

$\hat{\tau} = 37,800$ and $B = 3,379.941$

Exercise 4.28: Sample size determination involving τ.

$n_{\text{new}} = 399.4126$

Exercise 4.25: Compute $\hat{p}$ and B from the summary statistics.

$\hat{p} = 0.1833333$ and $B = 0.0957597$

Exercise 4.41: Compute $\hat{\tau}$, $\hat{\mu}$, (both values of) B, and a 95% confidence interval on μ from the data.

> z = scan2("EXER4_41.DAT", T, T)

> y = z[, 2]

$\hat{\tau} = \$98,550$ where \$19,905.83 is the bound on the error of estimation, $\hat{\mu} = \$197.10$ where \$39.81166 is the bound on the error of estimation, and a 95% confidence interval on μ is (\$157.2883, \$236.9117).

Exercise 4.42: Compute $\hat{p}$, B, and a 95% confidence interval on p from the data.

> z = scan2(“EXER4_41.DAT”, T)

> y1 = as.numeric(z[, 3] == “N”) # Hence, treat the data as zeros and ones.

$\hat{p} = 0.3$, $B = 0.2060148$, and a 95% confidence interval on p is $(0.09398518, 0.50601482)$.

> # Alternatively:

> y2 = sum(z[, 3] == “N”) # 6 accounts are not in compliance.

Exercise 4.50: Compute $\hat{p}$, B, and a 95% confidence interval on p from the summary data, where the population size is **infinite**.

$\hat{p} = 0.25$, $B = 0.07938842$, and a 95% confidence interval on p is $(0.1706116, 0.3293884)$.

S.2 Stratified Random Sampling from Chapter 5

The exercises begin on p. 157.

Exercise 5.6 (a): Compute $\hat{\mu}$ and B from the data set.

> z = scan2("EXER5_6.DAT", T)

> strata = z[, 1]

> y = scan2("EXER5_6.DAT", F, T)

$\hat{\mu} = 59.98862$ and $B = 3.032395$

Exercise 5.8: Sample size determination involving μ based on the data, and proportional allocation.

$n_{\text{new}} = 32.05365$

Exercise 5.9: Sample size determination involving μ based on the data, and Neyman allocation.

$n_{\text{new}} = 31.68249$

Exercise 5.10: Compute $\hat{\tau}$ and B from the data set.

> z = scan2("EXER5_10.DAT", T)

> strata = z[, 1]

> y = scan2("EXER5_10.DAT", F, T)

$\hat{\tau} = 50,505.6$ acres and $B = 8,663.124$ acres

Exercise 5.11: Sample size determination involving τ based on the data, and Neyman allocation.

$n_{\text{new}} = 59.67771$

Exercise 5.19: Compute $\hat{\mu}$, B, and a 95% confidence interval on μ from the summary statistics.

> z = scan2("EXER5_19.DAT", T)

> strata = z[, 1]

> y = scan2("EXER5_19.DAT", F, T)

$\hat{\mu} = 63.88333$ pounds, $B = 0.6280481$ pounds, and a 95% confidence interval on μ is (63.25529 pounds, 64.51138 pounds).

Exercise 5.1: Compute $\hat{p}$ and B from the summary statistics.

$\hat{p} = 0.3$ and $B = 0.1172979$

Exercise 5.13: Sample size determination involving p based on summary statistics, and optimal allocation with unequal costs.

$n_{\text{new}} = 157.2032$

The allocation is (38.99698, 17.28732, 68.70067, 32.21824).

Exercise 5.31 (a): Compute $\hat{p}$ and B from the summary statistics.

```
> yes = c( 913, 136, 860 )
> no = c( 417, 29, 240 )
> n = yes + no
```

$\hat{p} = 0.7383846$ and $B = 0.01722262$

Example from class notes, chapter 5: Compute $\hat{\mu}$, B, and a 95% confidence interval on μ from the summary statistics, using **poststratification**.

Suppose a sampling frame lists all households in an area, and you would like to estimate the average monthly household food bill.

From U.S. Census data, the distribution of household sizes in the region might be known:

Number of persons in Household	Percentage of Households
1	26
2	31
3	18
4	15
5+	10
total	100

However, the sampling frame does not include information on household size. The sampling frame lists only the households. Take a simple random sample of, say, 100 households.

Suppose we obtain the following sample sizes:

$n_1 = 21,\ n_2 = 36,\ n_3 = 19,\ n_4 = 11,\ n_5 = 13$

Suppose the sample mean monthly food bills for the five strata are the following:

$\bar{Y}_1 = \$270,\ \bar{Y}_2 = \$420,\ \bar{Y}_3 = \$560,\ \bar{Y}_4 = \$690,\ \bar{Y}_5 = \$810$

Suppose the sample standard deviation monthly food bills are $s_1 = \$150$, $s_2 = \$270$, $s_3 = \$390$, $s_4 = \$470$, and $s_5 = \$560$.

Suppose that this community has 10,000 households.

$\hat{\mu} = \$485.7$, $B = \$70.91842$, and a 95% confidence interval on μ is $(\$414.7816, \$556.6184)$.

S.3 Ratio, Regression, and Difference Estimation from Chapter 6

The exercises begin on p. 218.

Exercise 6.1: Compute $\hat{\tau}_y$ and B from the data set, using a ratio estimator.

> z = scan2(“EXER6_1.DAT”, T)

> x = z[, 2] ; y = z[, 3]

$\hat{\tau}_y = 1,589.552$ ft^3 and $B = 186.3176$ ft^3

Exercise 6.2: Compute $\hat{\tau}_y$ and B from the data set, using techniques based on simple random sampling.

$\hat{\tau}_y = 2,958.333$ ft^3 and $B = 730.1342$ ft^3

Exercise 6.4: Compute $\hat{\tau}_y$ and B from the data set, using a ratio estimator.

```
> z = scan2( "EXER6_4.DAT", T )
> x = z[ , 2 ] ;  y = z[ , 3 ]
```

$\hat{\tau}_y = 145,943.8$ and $B = 7,353.668$

Exercise 6.5: Compute $\hat{\mu}_y$ and B from the data set, using a ratio estimator.

$\hat{\mu}_y = 1,186.535$ and $B = 59.78592$

Exercise 6.6: Compute $\hat{\mu}_y$ and B from the data set, using a ratio estimator.

```
> z = scan2( "EXER6_6.DAT", T )
> x = z[ , 2 ] ;  y = z[ , 3 ]
```

$\hat{\mu}_y = 17.58918$ seconds and $B = 0.2710168$ seconds

Exercise 6.10: Compute the sample ratio r and B from the data set, using a ratio estimator (and unknown μ_x).

> z = scan2("EXER6_10.DAT", T, T)

> x = z[, 2] ; y = z[, 3]

$\hat{\mu} = 1.038028$ and $B = 0.003623381$

Exercise 6.12: Compute $\hat{\tau}_y$ and B from the data set, using a ratio estimator.

> z = scan2("EXER6_12.DAT", T)

> x = z[, 2] ; y = z[, 3]

$\hat{\tau}_y = 231,611.9$ and $B = 3,073.833$

Exercise 6.13: Sample size determination involving τ_y based on the data set, using a ratio estimator.

$n_{\text{new}} = 13.28976$

Exercise 6.17: Compute the sample ratio r and B from the data set, using a ratio estimator (and unknown μ_x).

> z = scan2("EXER6_17.DAT", T)

> x = z[, 2] ; y = z[, 1] # Note that 'x' and 'y' are reversed.

$\hat{\mu} = 0.8352573$ and $B = 0.01194476$

Exercise 6.23 (a,b,c): Compute $\hat{\tau}_y$ and B from the data set, using ratio, regression, and difference estimators.

> z = scan2("EXER6_23.DAT", T, T)

> x = z[, 1] ; y = z[, 2]

(a) *Ratio estimation:* $\hat{\tau}_y = 899.5837$ billion dollars and $B = 92.25955$ billion dollars

(b) *Regression estimation:* $\hat{\tau}_y = 889.759$ billion dollars and $B = 99.39048$ billion dollars

(c) *Difference estimation:* $\hat{\tau}_y = 933.6667$ billion dollars and $B = 141.6594$ billion dollars

Exercise 6.33: Compute $\hat{\mu}_y$, B, and a 95% confidence interval on μ_y from the data set, using a regression estimator and unknown N.

> z = scan2("EXER6_33.DAT", T)

> x = z[, 2] ; y = z[, 3]

$\hat{\mu}_y$ = 196.9714 pounds, B = 21.60458 pounds, and a 95% confidence interval on μ_y is (175.3668 pounds, 218.5760 pounds).

Exercise 6.34: Compute $\hat{\mu}_y$ and B from the data set, using a regression estimator and unknown N.

> z = scan2("EXER6_34.DAT", T)

> x = z[, 2] ; y = z[, 1] # Note that 'x' and 'y' are reversed.

$\hat{\mu}_y$ = 0.3958196 ml/sec and B = 0.02285520 ml/sec

S.4 Cluster Sampling from Chapter 8

The exercises begin on p. 292.

Exercise 8.2: Compute $\hat{\mu}$ and B, with unknown M.

> z = scan2("EXER8_2.DAT", T)

> m = z[, 2] ; y = z[, 3]

$\hat{\mu} = \$19.73077$ per saw and $B = \$1.780103$ per saw

Exercise 8.3: Compute $\hat{\tau}$ and B, with **unknown** M.

$\hat{\tau} = \$12,312$ and $B = \$3,175.068$

Exercise 8.4: Compute $\hat{\tau}$ and B, with **known** M.

$\hat{\tau} = \$14,008.85$ and $B = \$1,110.785$

Exercise 8.5: Sample size determination involving μ, with **unknown** and **known** M.

(a) Assume M is unknown. Error bound when estimating μ is \$2.

$n_{\text{new}} = 16.56081$

(b) Assume $M = 710$. Error bound when estimating μ is \$2.

$n_{\text{new}} = 13.3146$

(c) *New problem:* Assume $M = 710$. Error bound when estimating τ is \$1420.

$n_{\text{new}} = 13.3146$

(d) *New problem:* Assume M is unknown. Error bound when estimating τ is \$1420.

$n_{\text{new}} = 54.54325$

Exercise 8.8: Compute $\hat{p}$ and B, with **unknown** M.

```
> z = scan2( "EXER8_8.DAT", T )
> m = z[ , 2 ] ;  y = z[ , 3 ]
```

$\hat{p} = 0.7091109$ and $B = 0.04813522$

Exercise 8.9: Sample size determination involving p, with **unknown** M.

$n_{\text{new}} = 6.101613$

Exercise 8.10: Compute $\hat{\mu}$ and B, with **unknown**

M.

> z = scan2("EXER8_10.DAT", T)

> m = z[, 2] ; y = z[, 3]

$\hat{\mu} = \$40.16884$ and $B = \$0.6404408$

Exercise 8.11: Compute $\hat{\tau}$ and B, with **unknown** M.

$\hat{\mu} = \$157,020$ and $B = \$6,927.875$

Exercise 8.12: Sample size determination involving p, with **unknown** M.

$n_{\text{new}} = 29.38633$