Stat 576
10-19-17

Question 3 was: How large should n be?

Assume that our total cost is fixed ①

$$C = C_0 + \sum_{h=1}^{H} n_h C_h$$

$$= C_0 + n \sum_{h=1}^{H} \frac{n_h}{n} C_h$$

$$n = \frac{C - C_0}{\sum_{h=1}^{H} \frac{n_h}{n} C_h}$$

For optimal allocation, ②

$$\frac{n_h}{n} = \frac{\frac{N_h S_h}{\sqrt{C_h}}}{\sum_{h=1}^{H} \frac{N_h S_h}{\sqrt{C_h}}}$$

So $$n = \frac{C - C_0}{\sum_{h=1}^{H} \left(\frac{N_h S_h / \sqrt{C_h}}{\sum_{h=1}^{H} N_h S_h / \sqrt{C_h}} \right) C_h}$$

(3)

$$n = (C - c_0)\frac{\sum_{h=1}^{H} N_h S_h / \sqrt{c_h}}{\sum_{h=1}^{H} N_h S_h \sqrt{c_h}}$$

For Neyman allocation, $c_h \equiv c$

$$n = (C - c_0)\frac{\frac{1}{\sqrt{c}}\sum_{h=1}^{H} N_h S_h}{\sqrt{c}\sum_{h=1}^{H} N_h S_h} = \frac{C - c_0}{c}$$

(4)

This remains true for proportional & equal allocation

Question 2 was about demarcation

Goal: Minimize

$$V(\bar{y}_{str}) = \sum_{h=1}^{H} \frac{N_h^2}{N^2} \frac{S_h^2}{n_h}\left(1 - \frac{n_h}{N_h}\right)$$

Approximate by ignoring the fpcs

$$V(\bar{y}_{str}) \approx \sum_{h=1}^{H} W_h^2 \frac{S_h^2}{n_h}, \text{ where } W_h = \frac{N_h}{N} \quad (5)$$

Assume that we will be using Neyman allocation

$$\text{So } n_h = \frac{N_h S_h}{\sum_{h=1}^{H} N_h S_h} \cdot n$$

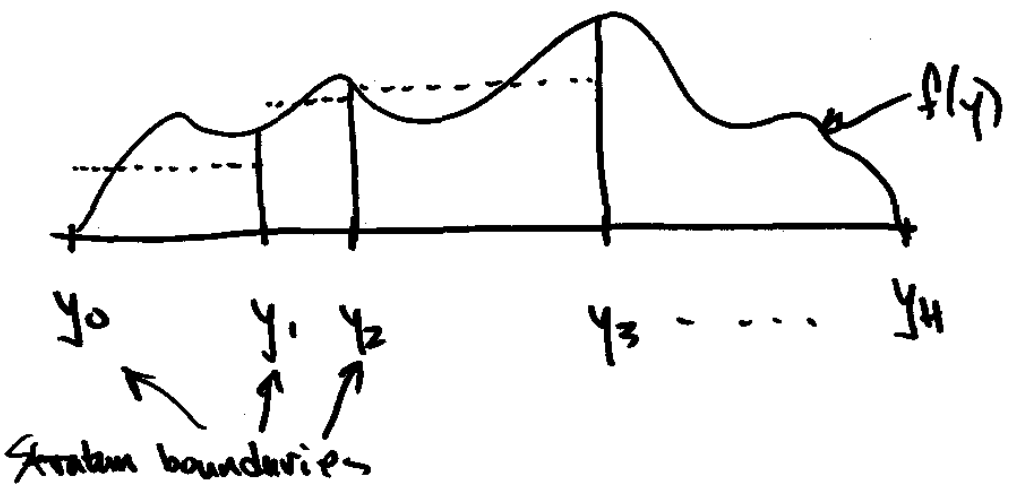
$$V(\bar{y}_{str}) \approx \sum_{h=1}^{H} W_h^2 \frac{S_h^2}{\left[\frac{N_h S_h}{\sum_{h=1}^{H} N_h S_h}\right] n}$$

$$V(\bar{y}_{str}) \approx \left(\sum_{h=1}^{H} N_h S_h\right) \frac{1}{n} \sum_{h=1}^{H} \frac{W_h^2 S_h}{N_h} \quad (6)$$

$$\approx \left(\sum_{h=1}^{H} \frac{N_h}{N} S_h\right) \frac{1}{n} \sum_{h=1}^{H} \frac{W_h^2 S_h}{N_h/N}$$

$$\approx \frac{1}{n}\left(\sum_{h=1}^{H} W_h S_h\right)^2$$

Minimizing $\sum_{h=1}^{H} W_h S_h$ will (approx) minimize $V(\bar{y}_{str})$



Assume $f(y)$ can be approximated by a uniform density within each stratum

Now $W_h = \frac{N_h}{N}$ = proportion of population in stratum h

$$W_h = \int_{y_{h-1}}^{y_h} f(y)\,dy \approx \int_{y_{h-1}}^{y_h} f_h\,dy$$

$$= f_h(y_h - y_{h-1})$$

Also, S_h^2 = population variance in stratum h

$$\approx \frac{(y_h - y_{h-1})^2}{12}$$

(variance of a uniform distribution)

$$\text{So } \sum_{h=1}^{H} W_h S_h \approx \sum_{h=1}^{H} f_h (y_h - y_{h-1}) \sqrt{\frac{(y_h - y_{h-1})^2}{12}}$$

$$= \frac{1}{\sqrt{12}} \sum_{h=1}^{H} f_h (y_h - y_{h-1})^2$$

$$\text{Let } a_h = \sqrt{f_h}(y_h - y_{h-1}) \quad \& \quad b_h = \frac{1}{\sqrt{H}}$$

$$\text{Then } \sum_{h=1}^{H} a_h^2 \sum_{h=1}^{H} b_H^2 = \sum_{h=1}^{H} f_h (y_h - y_{h-1})^2 \cdot 1$$

Remember, Cauchy-Schwarz says

$|a|^2 |b|^2 \geq (a \cdot b)^2$, with equality

when $\vec{a} = k\vec{b}$

That is, $V(\bar{y}_{str})$ is (approx) minimized

when $\sqrt{f_h}(y_h - y_{h-1})$ is constant

This is the Dalenius-Hodges Rule

Example:

We want 2 strata

Annual income	freq	$\sqrt{f}$
20 to 30	11	3.32
30 to 40	14	3.74
40 to 50	9	3
50 to 60	4	2
60 to 70	5	2.24
70 to 80	8	2.83
80 to 90	3	1.73
90 to 100	2	1.41
		20.27

First guess: Use the Cumulative square root frequency (CSRF) rule — divide the total square-root freq. by the desired # of strata

$\frac{20.27}{2} = 10.135$

Check your results: $\sqrt{f_h}(y_h - y_{h-1})$

$\sqrt{34}(50-20) \overset{?}{\approx} \sqrt{22}(100-50)$

175 235

Trial & error: move each boundary up 1 space + down 1 space + see if you get an improvement

$$\sqrt{38}(60-20) \overset{?}{\approx} \sqrt{18}(100-60)$$

$$246 \qquad 109$$

Suppose that we can come close to Dalenius-Hodges and that we are using Neyman allocation

$$n_h = \frac{N_h S_h}{\sum_{h=1}^{H} N_h S_h} \cdot n = \frac{N_h/N \, S_h}{\sum_{h=1}^{H} \frac{N_h}{N} S_h} \cdot n$$

$$= \frac{W_h S_h}{\sum_{h=1}^{H} W_h S_h} \cdot n$$

But $W_h S_h = f_h (y_h - y_{h-1}) \sqrt{\frac{(y_h - y_{h-1})^2}{12}}$

$$= \frac{1}{\sqrt{12}} f_h (y_h - y_{h-1})^2 = \frac{1}{\sqrt{12}} Q_h^2$$

Dalenius-Hodges solution was to make Q_h constant

$$\therefore n_h = \frac{n}{H}$$, which is equal allocation

Summary: To minimize $V(\bar{y}_{st})$, use D-H rule to find the boundaries, & then apply equal allocation !!

7 In Exercise 6 of Chapter 2, data on numbers of publications were given for an SRS of 50 faculty members. Not all departments were represented, however, in the SRS. The SRS contained several faculty members from psychology and from chemistry, but none from foreign languages. The following data are from a stratified sample of faculty, using the areas biological sciences, physical sciences, social sciences, and humanities as the strata.

Stratum	Number of Faculty Members in Stratum	Number of Faculty Members in Sample
Biological Sciences	102	7
Physical Sciences	310	19
Social Sciences	217	13
Humanities	178	11
Total	807	50

The frequency table for number of publications in the strata is given below.

Number of Refereed Publications	Number of Faculty Members			
	Biological	Physical	Social	Humanities
0	1	10	9	8
1	2	2	0	2
2	0	0	1	0
3	1	1	0	1
4	0	2	2	0
5	2	1	0	0
6	0	1	1	0
7	1	0	0	0
8	0	2	0	0

a Estimate the total number of refereed publications by faculty members in the college, and give the standard error.

b How does your result from (a) compare with the result from the SRS in Exercise 6 of Chapter 2?

c Estimate the proportion of faculty with no refereed publications, and give the standard error.

d Did stratification increase precision in this example? Explain why you think it did or did not.

8 A public opinion researcher has a budget of \$20,000 for taking a survey. She knows that 90% of all households have telephones. Telephone interviews cost \$10 per household; in-person interviews cost \$30 each if all interviews are conducted in person, and \$40 each if only nonphone households are interviewed in person (because there will be extra travel costs). Assume that the variances in the phone and nonphone groups are similar, and that the fixed costs are $c_0 = \$5000$. How many households should be interviewed in each group if

a all households are interviewed in person

b households with a phone are contacted by telephone and households without a phone are contacted in person.