Sample Size Determination for Stratified Sampling

I. Quantitative Data

Overall sample size n and allocations n_h's are determined based on considering linear cost function: $C = c_0 + \sum c_h n_h$. The optimum allocation is defined as

$$n_h = n \times \frac{N_h S_h/\sqrt{c_h}}{\sum N_h S_h/\sqrt{c_h}}$$

1. Population Mean

(i). Minimizing B for a fixed cost C.

$$n = \frac{(C - c_0) \sum N_h S_h/\sqrt{c_h}}{\sum N_h S_h \sqrt{c_h}}$$

(ii). Minimizing C for a fixed B.

$$n = \frac{(\sum N_h S_h/\sqrt{c_h})(\sum N_h S_h \sqrt{c_h})}{\sum N_h S_h^2 + \frac{N^2 B^2}{4}}$$

(iii). When $c_h \equiv c$, it reduces to minimizing the the sample size for a fixed B (Neyman allocation).

$$n = \frac{(\sum N_h S_h)^2}{\sum N_h S_h^2 + \frac{N^2 B^2}{4}}$$

$$n_h = n \times \frac{N_h S_h}{\sum N_h S_h}$$

2. Population Total

(i). Minimizing B for a fixed cost C.

$$n = \frac{(C - c_0)\sum N_h S_h/\sqrt{c_h}}{\sum N_h S_h \sqrt{c_h}}$$

(ii). Minimizing C for a fixed B.

$$n = \frac{(\sum N_h S_h/\sqrt{c_h})(\sum N_h S_h \sqrt{c_h})}{\sum N_h S_h^2 + \frac{B^2}{4}}$$

(iii). When $c_h \equiv c$, it reduces to minimizing the the sample size for a fixed B (Neyman allocation).

$$n = \frac{(\sum N_h S_h)^2}{\sum N_h S_h^2 + \frac{B^2}{4}}$$

$$n_h = n \times \frac{N_h S_h}{\sum N_h S_h}$$

II. Qualitative Data

Overall sample size n and allocations n_h's are determined based on considering linear cost function: $C = c_0 + \sum c_h n_h$. The optimum allocation is defined as

$$n_h = n \times \frac{N_h \sqrt{\frac{N_h}{N_h - 1} P_h(1 - P_h)} / \sqrt{c_h}}{\sum N_h \sqrt{\frac{N_h}{N_h - 1} P_h(1 - P_h)} / \sqrt{c_h}}$$

1. Population Proportion

(i). Minimizing B for a fixed cost C.

$$n = \frac{(C - c_0) \sum N_h \sqrt{\frac{N_h}{N_h - 1} P_h(1 - P_h)} / \sqrt{c_h}}{\sum N_h \sqrt{\frac{N_h}{N_h - 1} P_h(1 - P_h)} \sqrt{c_h}}$$

(ii). Minimizing C for a fixed B.

$$n = \frac{\left(\sum N_h \sqrt{\frac{N_h}{N_h - 1} P_h(1 - P_h)} / \sqrt{c_h}\right) \left(\sum N_h \sqrt{\frac{N_h}{N_h - 1} P_h(1 - P_h)} \sqrt{c_h}\right)}{\sum \frac{N_h^2 P_h(1 - P_h)}{N_h - 1} + \frac{N^2 B^2}{4}}$$

(iii). When $c_h \equiv c$, it reduces to minimizing the the sample size for a fixed B (Neyman allocation).

$$n = \frac{\left(\sum N_h \sqrt{\frac{N_h}{N_h - 1} P_h(1 - P_h)}\right)^2}{\sum \frac{N_h^2 P_h(1 - P_h)}{N_h - 1} + \frac{N^2 B^2}{4}}$$

$$n_h = n \times \frac{N_h \sqrt{\frac{N_h}{N_h - 1} P_h(1 - P_h)}}{\sum N_h \sqrt{\frac{N_h}{N_h - 1} P_h(1 - P_h)}}$$

2. Population Total (total number)

(i). Minimizing B for a fixed cost C.

$$n = \frac{(C - c_0) \sum N_h \sqrt{\frac{N_h}{N_h - 1} P_h(1 - P_h)} / \sqrt{c_h}}{\sum N_h \sqrt{\frac{N_h}{N_h - 1} P_h(1 - P_h)} \sqrt{c_h}}$$

(ii). Minimizing C for a fixed B.

$$n = \frac{\left(\sum N_h \sqrt{\frac{N_h}{N_h - 1} P_h(1 - P_h)} / \sqrt{c_h}\right) \left(\sum N_h \sqrt{\frac{N_h}{N_h - 1} P_h(1 - P_h)} \sqrt{c_h}\right)}{\sum \frac{N_h^2 P_h(1 - P_h)}{N_h - 1} + \frac{B^2}{4}}$$

(iii). When $c_h \equiv c$, it reduces to minimizing the the sample size for a fixed B (Neyman allocation).

$$n = \frac{\left(\sum N_h \sqrt{\frac{N_h}{N_h - 1} P_h(1 - P_h)}\right)^2}{\sum \frac{N_h^2 P_h(1 - P_h)}{N_h - 1} + \frac{B^2}{4}}$$

$$n_h = n \times \frac{N_h \sqrt{\frac{N_h}{N_h - 1} P_h(1 - P_h)}}{\sum N_h \sqrt{\frac{N_h}{N_h - 1} P_h(1 - P_h)}}$$