36-303: Sampling Surveys and Society
Stratified Sampling Examples

Brian Junker
132E Baker Hall
brian@stat.cmu.edu
February 23, 2010

Contents

1 Key Ideas

1.1 Why stratify?

Strata are just subgroups of the target population that have some feature in common (gender, major, region, income, ...) Reasons to stratify include:

- We need to make a separate inference for each stratum (e.g. we want to estimate mens and womens incomes separately);
- Different sampling schemes would be used in each stratum (PA voters in PA [telephone survey?], vs PA voters in Iraq [mail or email?]);
- Population is geographically diverse (Minnesota, Illinois, Ohio, Pennsylvania, ...);
- Reduce variance of estimates (and reduce sample size) by exploiting similarity among members of the same stratum.

1.2 Notation and Basic Facts

We assume there are H strata. Within each stratum, we will use the following notation:

- N_h = *population size in each stratum*
- n_h = *sample size in each stratum*
- $f_h = n_h/N_h$ = *sampling fraction, each stratum*
- $W_h = N_h/N$ = *stratum weights, each stratum*
- $N = \sum_{h=1}^{H} N_h$ = *total population size*
- $n = \sum_{h=1}^{H} n_h$ = *total sample size*

We also assume throughout that within each stratum we are taking a simple random sample (SRS) without replacement, of size n_h from the stratum population of size N_h.

Let y be some attribute of interest. y could be numerical, like a person's income, or it could be Bernoulli, like $y = 1$ if the person rides PAT buses, and $y = 0$ otherwise.

1.2.1 Means

The population average for y is

$$\overline{y}_{pop} = \frac{1}{N}\sum_{i=1}^{N} y_i = \frac{1}{N}\sum_{h=1}^{H}\sum_{i=1}^{N_h} y_{hi} = \sum_{h=1}^{H}\frac{N_h}{N}\frac{1}{N_h}\sum_{i=1}^{N_h} y_{hi} = \sum_{h=1}^{H}\frac{N_h}{N}\overline{y}_{h,pop} = \sum_{h=1}^{H} W_h\overline{y}_{h,pop} \qquad (1)$$

where $\overline{y}_{h,pop} = \frac{1}{N_h}\sum_{i=1}^{N_h} y_i$ and

$$W_h = N_n/N. \qquad (2)$$

So we can see that $\overline{y}_{pop}$ is a weighted sum of the population averages in each stratum, $\overline{y}_{h,pop}$, where the weights W_h in (2) are proportional to the size of the stratum.

To make an estimate from stratified sample data we can mimic the idea in (1) with sample averages:

$$\overline{y}_{st} = \frac{1}{n}\sum_{i=1}^{n} y_i = \sum_{h=1}^{H} \frac{N_h}{N}\overline{y}_h = \sum_{h=1}^{H} W_h \overline{y}_h \tag{3}$$

where $\overline{y}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} y_{hi}$. and the W_h are as in (2).

It is important to realize that the stratified sample average is unbiased for the population average:

$$\begin{aligned}
E[\overline{y}_{st}] &= E\left[\sum_{h=1}^{H} W_h \overline{y}_h\right] && \text{(by (3))} \\
&= \sum_{h=1}^{H} W_h E[\overline{y}_h] \\
&= \sum_{h=1}^{H} W_h \overline{y}_{h,pop} && (E[\overline{y}_{h,srs}] = \overline{y}_{h,pop} \text{ in each stratum } h) \\
&= \overline{y}_{pop} && \text{(by (1)).}
\end{aligned}$$

1.2.2 Variances

Within each stratum, since we are doing SRS without replacement, we have

$$Var(\overline{y}_h) = (1 - f_h)\frac{s_h^2}{n_h} \qquad \text{where} \qquad s_h^2 = \frac{1}{n_h - 1}\sum_{i=1}^{n_h}(y_{hi} - \overline{y}_h)^2 \tag{4}$$

as usual. Note that for Bernoulli data, there is a simpler expression for the within-cluster standard error:

$$s_h^2/n_h = \frac{p_h(1-p_h)}{n_h - 1} \tag{5}$$

To obtain an overall variance estimate for $\overline{y}_{st}$ we have:

$$\begin{aligned}
Var(\overline{y}_{st}) &= Var\left(\sum_{h=1}^{H} W_h \overline{y}_h\right) \\
&= \sum_{h=1}^{H} Var(W_h \overline{y}_h) \;=\; \sum_{h=1}^{H} W_h^2 Var(\overline{y}_h) \\
&= \sum_{h=1}^{H} W_h^2 (1 - f_h)\frac{s_h^2}{n_h} \qquad (6)
\end{aligned}$$

2 Self-weighting Samples, and Proportionate Sampling

2.1 Self-weighting samples

In general,

$$\overline{y}_{st} = \sum_{h=1}^{H} W_h \overline{y}_h$$

need not be the same as

$$\overline{y}_{srs} = \frac{1}{n}\sum_{i=1}^{n} y_i.$$

When we are lucky enough to have $\overline{y}_{srs} = \overline{y}_{st}$ the stratification is said to be "self-weighting".

2.2 Proportionate sampling

One case of a self-weighing stratification is called "proportionate sampling". We get proportionate sampling when we force the sampling fraction f_h to be the same in every stratum,

$$f_h \equiv f, \ \forall h \tag{7}$$

Since $f_h = n_h/N_h$ it is easy to see that if every $n_h/N_h \equiv f$ then also $n/N = f$, and so

$$n_h/n = (n_h/N_h)(N_h/N)(N/n) = f(N_h/N)(1/f) = N_h/N$$

Therefore,

$$\overline{y}_{st} = \sum_{h=1}^{H} W_h \overline{y}_h = \sum_{h=1}^{H}(N_h/N)\overline{y}_h = \sum_{h=1}^{H}(n_h/n)\overline{y}_h = \sum_{h=1}^{H}(n_h/n)\frac{1}{n_h}\sum_{i=1}^{n_h} y_{hi} = \frac{1}{n}\sum_{i=1}^{n} y_i = \overline{y}_{srs} \tag{8}$$

2.3 Design Effect

The nice part about self-weighting samples is that anybody, with no statistical training, can calculate $\overline{y}_{srs}$ and it will be the right number for $\overline{y}_{st}$ as well. The only difference is that instead of

$$Var(\overline{y}_{srs}) = (1-f)\frac{s^2}{n} \tag{9}$$

we will have

$$Var(\overline{y}_{st}) = \sum_{h=1}^{H} W_h^2(1-f_h)\frac{s_h^2}{n_h} \tag{10}$$

The "art" of stratified sampling is making sure that $Var(\overline{y}_{st})$ is smaller than $Var(\overline{y}_{sr})$. The *design effect* is defined to be

$$d^2 = \frac{Var(\overline{y}_{st})}{Var(\overline{y}_{srs})} = \frac{\sum_{h=1}^{H} W_h^2(1-f_h)\frac{s_h^2}{n_h}}{(1-f)\frac{s^2}{n}} \tag{11}$$

and we have succeeded in the art of stratified sampling when $d^2 < 1$. Cochran (1961) did some experiments with stratification and showed that $d^2 < 1$ when

- 2–6 strata are used (more strata tend to reduce $Var(\overline{y}_{st})$ further, but there are diminishing returns as H grows);
- Elements are more similar to each other within strata than between (e.g., substantively meaningful strata);
- Proportionate sampling is used.

3 Examples

3.1 Confidence Intervals for SRS and Proportionate Stratified Samples

3.1.1 Proportion p_{pop}

Consider the three urns we played with in class:

- Urn 1: $N_1 = 100$ balls; 10 yellow, 90 blue, $p_1 = 0.1$
- Urn 2: $N_2 = 100$ balls; 20 yellow, 80 blue, $p_2 = 0.2$
- Urn 3: $N_3 = 100$ balls; 30 yellow, 70 blue, $p_3 = 0.3$

- Combined urns: $N = 300$, 60 yellow, 240 blue $p = 0.2$.

We want to compare $n = 30$ SRS without replacement from the combined urns, with a stratified sample of $n_h = 10$ balls each from the three urns.

***n* = 30 SRS:** Using formula (9) (and the adjustment in equation (5)),

$$Var(\hat{p}_{srs}) = (1-f)\frac{s^2}{n} = (1-f)\frac{p(1-p)}{n-1} = (1-30/300)\frac{0.2(1-0.2)}{30-1} = 0.00497$$

so, if we draw 30 balls and 8 of them are yellow, then $\hat{p} = 8/30 = 0.27$ and we get a 95% CI that is

$$0.27 \pm (1.96)\sqrt{0.00497} = 0.27 \pm 0.1381$$

n_h = 10 stratified SRS's from each urn: Using formula (10), and noting that $f_h = 10/100 = 0.1$ for all strata (proportionate sampling) and $W_h = 100/300 = 1/3$ for all strata,

$$Var(\hat{p}_{st}) = \sum_{h=1}^{H} W_h^2(1-f_h)\frac{s_h^2}{n_h} = \sum_{h=1}^{H} W_h^2(1-f_h)\frac{p_h(1-p_h)}{n_h-1} = \sum_{h=1}^{H}(1/3)^2(1-0.1)\frac{p_h(1-p_h)}{10-1}$$
$$= \frac{(1/3)^2(0.9)}{9}\left[(0.1)(0.9)+(0.2)(0.8)+(0.3)(0.7)\right] = 0.00511$$

If 2 of the 10 balls from Urn 1 are yellow, 2 from Urn 2 are yellow, and 3 from Urn 3 are yellow, we can use the fact that we are doing *proportionate sampling* to obtain $\hat{p} = 8/30 = 0.27$ (otherwise we have to use formula (3)). We get a 95% CI

$$0.27 \pm (1.96)\sqrt{0.00511} = 0.27 \pm 0.1401$$

Note that in this case, *stratification was not as good as a plain old SRS!*, although the difference wasn't enormous. The problem is that the three strata are a little too much like the combined urn. It is better when the strata have much less internal variability than the combined population would have. For example:

Setup: Same setup but now suppose the three urns have proportions of yellow $p_1 = 0.1$, $p_2 = 0.5$ and $p_3 = 0.9$. The combined population will now have a proportion $p = 0.5$ of yellow balls. We'll continue to have $N_h = 100$, $n_h = 10$.

SRS CI: The margin of error for a 95% CI is going to be

$$(1.96)\times\sqrt{(1-f)\frac{p(1-p)}{n-1}} = (1.96)\times\sqrt{(1-30/300)\frac{(0.5)(1-0.5)}{30-1}} = 0.1726$$

Stratified CI: The margin of error for a 95% CI is going to be

$$(1.96)\times\sqrt{\sum_{h=1}^{H} W_h^2(1-f_h)\frac{p_h(1-p_h)}{n_h-1}}$$
$$= (1.96)\times\sqrt{\frac{(1/3)^2(1-10/100)}{9}\left[(0.1)(0.9)+(0.5)(0.5)+(0.9)(0.1)\right]} = 0.1355$$

Now the stratified CI wins, because stratum 1 and 3 have so much less internal variability than the combined population does. In this case also, the design effect is

$$d^2 = \frac{Var(\hat{p}_{st})}{Var(\hat{p}_{srs})} = \frac{(0.1355/1.96)^2}{(0.1726/1.96)^2} = 0.616$$

so the stratified variance for $\overline{y}$ is only 61% of the SRS variance!

Taxpayer No.	Actual Income ($ thousands)	Reported Income
1	60	50
2	72	56
3	68	66
4	94	76
5	90	90
6	102	100
7	116	112
8	130	110
9	200	175

Table 1: Sampling frame for Williams (1978) example.

3.1.2 Numerical $\overline{y}_{pop}$

This example is a little contrived, in order to make all of the parts "visible". But the underlying principle, that stratification reduces variance, still holds.

Williams (1978) gives a simple illustration of gains from stratification for a small population of measurements on the income (in thousands of dollars) for a population of $N = 9$ taxpayers, shown in Table 1.

SRS: For the complete population we have $N = 9$, and for the actual income $\overline{y}_{pop} = 103.56$ and $s^2_{pop} = 1621.14$. The variance for an SRS of size $n = 3$ is

$$Var(\overline{y}_{srs}) == (1 - f)\frac{s^2}{n} = (1 - 3/9)\frac{1621.14}{3} = 360.25$$

(if we didn't know s^2_{pop} to plug in for s^2 we would plug in a sample variance instead), so the 95% margin of error for estimating $\overline{y}_{pop}$ would be

$$(1.96) \times \sqrt{360.25} = 37.20$$

Stratified: Now suppose we use the values of the reported income in order to stratify the population into two subpopulations consisting of the first eight observations and the ninth observation:

$$N_1 = 8, \overline{y}_{1,pop} = 91.50 \text{ and } s^2_{1,pop} = 515.75; N_2 = 1, \overline{y}_{2,pop} = 200 \text{ and } s^2_{2,pop} = 0.$$

We now take a sample of size $n = 3$, with $n_1 = 2$ from stratum 1 and $n_2 = N_2 = 1$ from stratum 2 (note that this is not a proportionate sample!). The variance for this stratified sample is

$$Var(\overline{y}_{st}) = \sum_{h=1}^{H} W_h^2(1 - f_h)\frac{s_h^2}{n_h} = (8/9)(1-3/8)\frac{515.75}{8} + (1/9)(1-1/9)\frac{0}{1} = 35.82 + 0 = 35.82$$

and the 95% margin of error for estimating $\overline{y}_{pop}$ will be

$$(1.96) \times \sqrt{35.82} = 11.73$$

The design effect in this case is

$$d^2 = \frac{Var(\overline{y}_{st})}{Var(\overline{y}_{srs})} = \frac{35.82}{360.32} = 0.0994$$

so in this case the variance for the stratified sample is *less than 10%* of the variance from an SRS.

Note especially in the Variance calcualtion above the expression "35.82 + 0". The zero comes because there is no variability in stratum 2, just the one observation. Low within stratum variability causes low variance, which causes low sample size (see next subsection!). This is a good thing.

3.2 Sample Size Calculations for SRS and Proportionate Stratified Samples

3.2.1 Proportion p_{pop}

We will illustrate calculating a sample size to get a 0.10 margin of error for a 95% CI. We will assume a population of 300 units.

SRS: We've done this a few times before. We first calculate the SRS with replacement sample size

$$n_0 \geq \frac{(Z)^2(SD)^2}{(ME)^2} = \frac{(1.96)^2(0.5)^2}{(0.10)^2} = 96.04$$

(where we take $SD = 0.5$ as a worst case), and then apply the correction for SRS without replacement.

$$n \geq \frac{Nn_0}{N + n_0} = \frac{(300)(96.04)}{300 + 96.04} = 72.75$$

Stratified: Let us suppose that our population breaks into three strata, with $N_1 = 50$, $N_2 = 100$, and $N_3 = 150$ units per stratum. Here are two possible approaches to determining sample sizes for the three strata:

- *Trial and error with proportionate sampling.* One might imagine a table like this:

f	$n_h = f \times N_h$	p_h	ME = 1.96 × $\sqrt{\textit{Eq. (10)}}$	$n = n_1 + n_2 + n_3$
0.10	(5, 10, 15)	(0.5, 0.5, 0.5)	0.18	30
0.20	(10, 20, 30)	(0.5, 0.5, 0.5)	0.12	60
0.30	(15, 30, 45)	(0.5, 0.5, 0.5)	0.09	90
0.25	(13, 25, 37)	(0.5, 0.5, 0.5)	0.10	75

  Note that here we have used just about the worst possible assumption, namely that $p_1 = p_2 = p_3$, so we get no advantage (and actually probably a slight disdadvantage) for stratifying.

  On the other hand if we knew something about p_1, p_2, p_3 (perhaps from a pilot study?), then things look better:

f	$n_h = f \times N_h$	$\hat{p}_{h,pilot}$	ME = 1.96 × $\sqrt{\textit{Eq. (10)}}$	$n = n_1 + n_2 + n_3$
0.100	(5, 10, 15)	(0.1, 0.5, 0.9)	0.14	30
0.200	(10, 20, 30)	(0.1, 0.5, 0.9)	0.09	60
0.150	(8, 15, 22)	(0.1, 0.5, 0.9)	0.11	45
0.175	(9, 17, 26)	(0.1, 0.5, 0.9)	0.10	53

  $n = 53$ is more efficient than $n = 72$ or 73 that would be needed for SRS. Note that we have used the additional information in the $\hat{p}_{pilot}$'s to get this better sample size!

- *Optimal Allocation.* It turns out that we can achieve greater gains from stratification by choosing our sampling fractions to be proportional to the standard deviations in each stratum. The idea is

$$n_h = cs_h$$

  where c is a constant, and s_h is the (population) SD for stratum h. For fixed overall sample size n, we also need

$$\sum_{h=1}^{H} n_h = n, \quad \text{i.e.} \sum_{h=1}^{H} cs_h = n$$

  and so, clearly,

$$c = \frac{n}{\sum_{h=1}^{H} s_h}$$

  and therefore

$$n_h = \frac{s_h}{\sum_{k=1}^{H} s_k} \times n$$

  We can apply this to our sample size problem for a margin of error of 0.10 with another 'trial and error table', this time for the overall sample size n:

n	$n_h = s_h / \sum_k s_k$	$f_h = n_h/N_h$	ME = $1.96 \times \sqrt{\textit{Eq. (10)}}$
33	(11, 11, 11)	(0.22, 0.11, 0.07)	0.18
66	(22, 22, 22)	(0.44, 0.22, 0.15)	0.12
99	(33, 33, 33)	(0.66, 0.33, 0.22)	0.09
90	(30, 30, 30)	(0.60, 0.30, 0.20)	0.10

Note that this is *not* a proportionate allocation scheme; the f_h's are not all equal.

Note also that we have taken $p = 0.5$ again, so that every s_h is 0.5, and so we are really just doing 1/3-1/3-1/3 allocation. This is probably why the result is not so great ($n = 90$ suggested here, vs. the best so-far of $n = 53$ above.

As before we can improve things by using pilot estimates for p. Now if we use our fictional pilot estimates $\hat{p}_{pilot} = (0.1, 0.5, 0.9)$, we discover that

$$s_1 = \sqrt{(0.1)(0.9)} = 0.3; \; s_2 = \sqrt{(0.5)(0.5)} = 0.5; \; s_3 = \sqrt{(0.9)(0.1)} = 0.3$$

and so the proportions of n allocated to each stratum would be

$$n_1 = s_1/\textstyle\sum_h s_h = (0.27)n; \; n_2 = s_2/\sum_h s_h = (0.45)n; \; n_3 = s_3/\sum_h s_h = (0.27)n;$$

Note that most of the sample is being spent on the stratum with the largest variance; that is why this method works. Our new trial and error table is as follows:

n	$n_h = s_h / \sum_k s_k$	$f_h = n_h/N_h$	ME = $1.96 \times \sqrt{\textit{Eq. (10)}}$
33	(9, 15, 9)	(0.18, 0.15, 0.06)	0.13
66	(18, 30, 18)	(0.36, 0.30, 0.12)	0.09
45	(12, 21, 13)	(0.25, 0.20, 0.08)	0.11
54	(14, 25, 15)	(0.30, 0.25, 0.10)	0.10

Once again this is not a proportionate allocation scheme. It is also worth noting that the N_h's don't seem to matter much: smaller samples are taken in the first and third strata, because that's where the variance, and hence the uncertainty, is the lowest – even though strata 1 and 3 are the smallest and largest strata.

Our answer here is comparable to the answer we got with proportionate sampling; indeed $n = 53$ also produces a 10% margin of error here.

3.2.2 Numerical $\overline{y}_{pop}$

To illustrate sample size calculations for a numerical variable, I will consider the problem of estimating average income for a population of size 300, that can be broken up into three strata:

Stratum (h)	N_h	$\overline{y}_{h,pop}$	$s^2_{h,pop}$
1	50	140	2000
2	100	85	415
3	150	50	200

Note: $\overline{y}_{h,pop}$ is the stratum mean income, in 1,000's of dollars. $s^2_{h,pop}$ is the stratum variance in income.

From this table one can easily calculate that

$$s^2_{pop} = \frac{(50)(2000) + (100)(415) + (150)(200)}{300} = 571.67$$

so that the population SD is $\sqrt{572.67} = 23.91$ (if we didn't know this we could estimate it from pilot data, for example).

We want to know the sample size needed to estimate the population mean income with a 95% margin of error of ± 5 thousand dollars.

SRS This is our old familiar friend. It proceeds pretty much as before. We first calculate the SRS with replacement sample size

$$n_0 \geq \frac{(Z)^2(SD)^2}{(ME)^2} = \frac{(1.96)^2(23.91)^2}{(5)^2} = 87.85$$

and then apply the correction for SRS without replacement.

$$n \geq \frac{Nn_0}{N + n_0} = \frac{(300)(87.85)}{300 + 87.85} = 67.95$$

Stratified We try the same two approaches that we tried for proportions:

- *Trial and error with proportionate sampling.* Here is our trial and error table:

f	$n_h = f \times N_h$	s_h	ME = $1.96 \times \sqrt{\textit{Eq. (10)}}$	$n = n_1 + n_2 + n_3$
0.10	(5, 10, 15)	(44.72, 12.04, 14.14)	7.45	30
0.20	(10, 20, 30)	(44.72, 12.04, 14.14)	4.96	60

  So, in this case, stratification with proportionate allocation requires only a sample of size $n = 60$, which is already beating SRS's sample size requirement of $n = 68$.

- *Optimal Allocation.* Reacalling from above that our standard deviations are $s_1 = 44.72$, $s_2 = 12.04$, and $s_3 = 14.14$ As in the previous subsection, the idea is to take

$$n_1 = s_1/\sum_h s_h = (0.63)n;\ n_2 = s_2/\sum_h s_h = (0.17)n;\ n_3 = s_3/\sum_h s_h = (0.20)n;$$

  Again we build a trial-and-error table for various sample sizes n:

n	$n_h = s_h / \sum_k s_k$	$f_h = n_h/N_h$	ME = $1.96 \times \sqrt{\textit{Eq. (10)}}$
30	(19, 5, 6)	(0.37, 0.05, 0.04)	7.03
60	(38, 10, 12)	(0.76, 0.10, 0.08)	4.64
45	(28, 8, 9)	(0.57, 0.08, 0.06)	5.55
54	(34, 9, 11)	(0.68, 0.09, 0.07)	4.97

Once again this is *not* a proportionate allocation scheme; the f_h's are not all equal. For this problem, the "optimal" scheme produces a somewhat better sample size requirement, $n = 54$ than the proportionate allocation scheme, $n = 60$.

4 Proportionate vs. Non-Proportionate Stratified Samples

Proportionate sampling schemes are easy to explain: the formula for $\overline{y}$ is the same under proportionate stratified sampling as it is under SRS (because proportionate sampling is self-weighting); moreover it is easy to explain to someone that you sampled more from the bigger strata, and that is what proportionate sampling does.

In some ways taking a bigger sample from the bigger stratum is looking at the problem wrong. The things that are hard to estimate are not things in big strata, but rather things that vary a lot within their stratum. *Optimal sampling* takes account of this by allocating sample size to each stratum, proportionate to that stratum's SD, rather than proportionate to that stratum's size. As we saw above, optimal sampling can produce lower sample sizes than proportionate sampling. You pay for this greater efficiency by not being able to use $\overline{y}_{srs}$ as your estimator anymore (instead, use the weighted $\overline{y}_{st}$ from equation (3)).

Proportionate sampling gets you most of the benefits of stratified sampling, in getting you a reduced sample size. Optimal sampling reduces the sample size a little bit more, but sometimes not much more.

If you are only interested in estimating overall population quantities, then once you have assigned sample sizes to each stratum by one of these methods, you are done. However, you may want to estimate some quantities separately in a stratum as well.

Example:

If you were doing a survey of satisfaction with sports facilities at Carnegie Mellon, your strata might be: staff, faculty, student-athletes, and other students. You might be epsecially interested in the proportion of student athletes who are satisfied. But that is a small stratum, so proportionate sampling will allocate only a few observations to that stratum.

You could do a separate, plain-vanilla SRS calculation in that one stratum, to see how many student athletes you need to just estimate their satisfaction. This will probably lead to a larger sample size in that stratum than proportionate sampling suggests. If so, go ahead and increase the sample size for the student-athletes stratum. It will allow you to make a better estimate for student-athletes, and it will not wreck anything else in the survey (in fact it should help your overall estimates just a little bit).

5 Pro's and Con's of Stratified Samples

Pro's:

- Can make your sample "more representative" of the population by making sure you get some observations from each stratum.
- Can "oversample" some strata because you want really good estimates for those strata.
- Can generally design a set of strata so that the margin of error is smaller for stratified samples than for nonstratified.

Con's:

- Figuring out what strata to use, largely a trial-and-error game.
- Figuring out the sample size for each stratum, largely a trial-and-error game.
- Need to know
  - The population sizes of the strata, N_h
  - The overall population size of your sampling frame, N
  - Rough guesses about the variances $\hat{s}^2_{h,pilot}$ and/or the proportions $\hat{p}_{h,pilot}$, often from a pilot study or other background information about your strata.

  in order to make stratification work for you to reduce sample size.
- Once you have committed to stratified sampling, now you have to make the extra effort necessary to reach your target sample size *in each stratum*. It will make for a better survey, but it is also more work for you.