Sampling: HW6

Leslie Gains-Germain

1. 6(b) page 103

	true % practicing energy conservation	Total in Population	# Sampled
House Dwellers	45%	35000	504
Apartment Dwellers	25%	45000	324
Condiminium Residents	3%	10000	72

The true proportion of households that practice energy conservation is 0.45(35000/90000) + 0.25(45000/90000) + 0.03(10000/90000) = 0.303. If we take a SRS of size $n = 900$, then the variance of $\hat{p}$ is as follows:

$$V_{SRS}(\hat{p}_{SRS}) = \left(\frac{N-n}{N}\right)\left(\frac{p(1-p)}{n}\right) = \left(\frac{90000-900}{90000}\right)\left(\frac{0.303 * 0.697}{900}\right)$$
$$= 0.000232$$

If we take a stratified sample of the size given in the table, the variance of the estimator $\hat{p}$ is given as follows:

$$\begin{aligned} V(\hat{p}_{STR}) &= \sum_{h=1}^{h=3}\left(\frac{N_h}{N}\right)^2\left(\frac{N_h-n_h}{N_h-1}\right)\left(\frac{p_h(1-p_h)}{n_h}\right) \\ &= \left(\frac{35000}{90000}\right)^2\left(\frac{35000-504}{35000-1}\right)\left(\frac{0.45*0.55}{504}\right) + \\ &\left(\frac{45000}{90000}\right)^2\left(\frac{45000-324}{45000-1}\right)\left(\frac{0.25*0.75}{324}\right) + \\ &\left(\frac{10000}{90000}\right)^2\left(\frac{10000-72}{10000-1}\right)\left(\frac{0.03*0.97}{72}\right) \\ &= 0.000222 \end{aligned}$$

The gain in variance of $\hat{p}$ with the stratified sample compared to the SRS is 0.000222/0.000232 = 0.957

2. 7(c) page 104

Department	Total in Stratum	# in Sample	# refereed publications in sample
Biological Sciences	102	7	1
Physical Sciences	310	19	10
Social Sciences	217	13	9
Humanities	178	11	8
Total	807	50	

The estimated proportion of all faculty with no refereed publications is 0.567. The standard error of $\hat{p}$ is 0.0658. My work is shown in the code below.

$$\hat{V}(\hat{p}_{STR}) = \sum_{h=1}^{h=3} \left(\frac{N_h}{N}\right)^2 \left(\frac{N_h - n_h}{N_h}\right) \left(\frac{p_h(1-p_h)}{n_h - 1}\right)$$

```
102/807*1/7+310/807*10/19+217/807*9/13+178/807*8/11
#0.567
sqrt((102/807)^2*(102-7)/102*(1/7*6/7)/6+
(310/807)^2*(310-19)/310*(10/19*9/19)/18+
(217/807)^2*(217-13)/217*(9/13*4/13)/12+
  (178/807)^2*(178-11)/178*(8/11*3/11)/10)
#0.0658
```

3. The first table shows the data for the Empathy-building exercises method, and the second table shows the data for the Gestalt exercises method.

Stratum	Stratum Total	# in Sample	empathy building as 'exp learning'
General nursing tutors (GT)	150	109	54
Psychiatric nursing tutors (PT)	34	26	20
General nursing students (GS)	2680	222	89
Psychiatric nursing students (PS)	570	40	25

The overall estimate for the proportion of students and tutors who identify the empathy-building exercises as 'experiential learning' is 0.4459. The standard error for this estimate is 0.02761. My work is shown below.

```
34/3434*20/26+150/3434*54/109+2680/3434*89/222+570/3434*25/40
#0.44587
sqrt((34/3434)^2*(34-26)/34*(20/26*6/26)/25+
(150/3434)^2*(150-109)/150*(54/109*55/109)/108+
(2680/3434)^2*(2680-222)/2680*(89/222*133/222)/221+
  (570/3434)^2*(570-40)/570*(25/40*15/40)/39)
#0.02761
```

Stratum	Stratum Total	# in Sample	Gestalt ex. as 'exp learning'
General nursing tutors (GT)	150	109	12
Psychiatric nursing tutors (PT)	34	26	5
General nursing students (GS)	2680	222	24
Psychiatric nursing students (PS)	570	40	4

The overall estimate for the proportion of students and tutors who identify the Gestalt exercises as 'experiential learning' is 0.1077. The standard error for this estimate is 0.0174. My work is shown below.

```
34/3434*5/26+150/3434*12/109+2680/3434*24/222+570/3434*4/40
#0.1077
sqrt((34/3434)^2*(34-26)/34*(5/26*21/26)/25+
(150/3434)^2*(150-109)/150*(12/109*97/109)/108+
(2680/3434)^2*(2680-222)/2680*(24/222*198/222)/221+
  (570/3434)^2*(570-40)/570*(4/40*36/40)/39)
#0.0174
```

4. (a) The optimal allocation is as follows:

$$n_h = \frac{nN_h s_h}{\sum_{h=1}^{2} N_h s_h}$$

First, I solve for s_1 and s_2. I assume that 1/N is approximately equal to 0 because the population is really large (Google says that there are about 1.5 million residents of

Milwaukee).

$$s_1^2 = \frac{N_1 p_1(1-p_1)}{N_1 - 1} = \frac{N_1/N p_1(1-p_1)}{N_1/N - 1/N} \approx \frac{N_1/N p_1(1-p_1)}{N_1/N} = 0.1 * 0.9 = 0.090$$
$$s_2^2 = \frac{N_2 p_2(1-p_2)}{N_2 - 1} = \frac{N_2/N p_2(1-p_2)}{N_2/N - 2/N} \approx \frac{N_2/N p_2(1-p_2)}{N_2/N} = 0.03 * 0.97 = 0.029$$

Then, I find the optimal allocation given different variances. I assume that the total sample size, n, is 2000. I find that the optimal allocation is to sample 1079 residents from Stratum 1 and 921 residents from Stratum 2.

$$\begin{aligned} n_1 &= \frac{nN_1 s_1}{N_1 s_1 + N_2 s_2} = \frac{2000 N_1/N s_1}{N_1/N s_1 + N_2/N s_2} \\ &= \frac{2000 * 0.4\sqrt{(0.09)}}{0.4 * \sqrt{(0.09)} + 0.6 * \sqrt{(0.029)}} \approx 1079 \\ n_2 &= \frac{nN_2 s_2}{N_1 s_1 + N_2 s_2} = \frac{2000 N_2/N s_2}{N_1/N s_1 + N_2/N s_2} \\ &= \frac{2000 * 0.6\sqrt{(0.029)}}{0.4 * \sqrt{(0.09)} + 0.6 * \sqrt{(0.029)}} \approx 921 \end{aligned}$$

(b) First, I find the variance of $\hat{p}$ under optimal allocation. I assume that $1/N \approx 0$, and I assume that n_1/N is ≈ 0 because the sample size is so large.

$$\begin{aligned} V(\hat{p}_{STR}) &= \left(\frac{N_1}{N}\right)^2 \left(\frac{N_1 - n_1}{N_1 - 1}\right) \left(\frac{p_1(1-p_1)}{n_1}\right) + \left(\frac{N_2}{N}\right)^2 \left(\frac{N_2 - n_2}{N_2 - 1}\right) \left(\frac{p_2(1-p_2)}{n_2}\right) \\ &= \left(\frac{N_1}{N}\right)^2 \left(\frac{N_1/N - n_1/N}{N_1/N + 1/N}\right) \left(\frac{p_1(1-p_1)}{n_1}\right) + \left(\frac{N_2}{N}\right)^2 \left(\frac{N_2/N - n_2/N}{N_2/N + 1/N}\right) \left(\frac{p_2(1-p_2)}{n_2}\right) \\ &\approx \left(\frac{N_1}{N}\right)^2 \left(\frac{p_1(1-p_1)}{n_1}\right) + \left(\frac{N_2}{N}\right)^2 \left(\frac{p_2(1-p_2)}{n_2}\right) \\ &= 0.4^2 \left(\frac{0.1(1-0.1)}{1079}\right) + 0.6^2 \left(\frac{0.03(1-0.03)}{921}\right) = 0.0000247 \end{aligned}$$

Under proportional allocation, $n_1 = N_1/N = 2000 * 0.4 = 800$ and $n_2 = N_2/N =$

$2000 * 0.6 = 1200$. Then the variance of of $\hat{p}$, using the same formula as above, is:

$$V(\hat{p}_{STR}) = 0.4^2 \left(\frac{0.1(1-0.1)}{800}\right) + 0.6^2 \left(\frac{0.03(1-0.03)}{1200}\right) = 0.0000267$$

For a simple random sample of 2000 from the population, the variance of $\hat{p}$ is shown below. I first find the true p in the population, $p = 0.4 * 0.1 + 0.6 * 0.03 = 0.058$. Again, we assume that N is sufficiently large that the finite population correction is not necessary.

$$V(\hat{p}_{SRS}) = \left(\frac{N-n}{N-1}\right)\left(\frac{p(1-p)}{n}\right) \approx \frac{p(1-p)}{n} = 0.0000273$$

We noticed that the variance of the estimator is smallest for the optimal allocation sampling scheme and largest for the simple random sample from entire population.