Sampling: HW4
Leslie Gains-Germain

1. (a) $\bar{y}_{1U} = 3.5$ and $\bar{y}_{2U} = 7.75$

   (b) $S_1^2 = 3.0$ and $S_2^2 = 7.5833$

   (c) $V(\hat{t}_1) = 4(4-3)3.0/3 = 4$ and $V(\hat{t}_1) = 4(4-3)7.5833/3 = 10.1111$

   (d) $V(\hat{t}_{STR}) = 14.1111$ and $V(\bar{y}_{U_{STR}}) = \frac{14.1111}{8^2} = 0.2205$

   (e) The possible simple random samples in each stratum are shown below. The units selected (not the y-values) are displayed.

Stratum 1	$\hat{t}_1$	Stratum 2	$\hat{t}_2$
1, 2, 3	12	5, 6, 7	26.6667
1, 3, 4	13.3333	5, 7, 8	33.3333
1, 2, 4	13.3333	5, 6, 8	29.3333
2, 3, 4	17.3333	6, 7, 8	34.6667

The values of $\hat{t}$ for each stratified sample are shown in the table below.

Stratum 1	Stratum 2	$\hat{t}$
1, 2, 3	5, 6, 7	38.6667
1, 2, 3	5, 7, 8	45.3333
1, 2, 3	5, 6, 8	41.3333
1, 2, 3	6, 7, 8	46.6667
1, 3, 4	5, 6, 7	40
1, 3, 4	5, 7, 8	46.6666
1, 3, 4	5, 6, 8	42.6666
1, 3, 4	6, 7, 8	48
1, 2, 4	5, 6, 7	40
1, 2, 4	5, 6, 8	46.6666
1, 2, 4	6, 7, 8	42.6666
1, 2, 4	6, 7, 8	48
2, 3, 4	5, 6, 7	44
2, 3, 4	5, 7, 8	50.6666
2, 3, 4	5, 6, 8	46.6666
2, 3, 4	6, 7, 8	52

   (f) The sampling distribution of $\hat{t}$ is shown in the table below.

$\hat{t}$	Probability
38.6667	1/16
45.3333	1/16
41.3333	1/16
46.6667	1/4
40	1/8
42.6667	1/8
48	1/8
44	1/16
50.6667	1/16
52	1/16

(g) The expected value of $\hat{t}$ is $E[\hat{t}] = 38.67(1/16) + 45.33(1/16) + 41.33(1/16) + 46.67(1/4) + 40(1/8) + 42.67(1/8) + 48(1/8) + 44(1/16) + 50.67(1/16) + 52(1/16) = 45$. The true population total is 45, so $\hat{t}$ is unbiased.

2. I would sample 504 houses, 324 apartments, and 72 condiminiums. My work is shown in the R code below.

```
35000+45000/2+10000/2
#62500
900*22500/62500
#324
900*5000/62500
#72
900*35000/62500
#504
```

3. The total number of breathing holes found in the sample are shown in the following table for each zone. Using the formula for $\hat{t}$, we find that $\hat{t} = 68/17(30) + 84/12(53) + 48/11(116) = 997.18$. Since holes is a whole number, the total number of holes in the study region is estimated to be 997.

	zone	sum(holes)	var(holes)
1	1	30	3.32
2	2	53	11.54
3	3	116	46.07

The variance of the $\hat{t}$ is estimated to be $\widehat{V(\hat{t})} = 68(68-17)3.3162/17+84(84-12)11.5379/12+48(48-11)46.0727/11 = 13930.25$, and the standard error of $\hat{t}$ is $\sqrt{13930.25} = 118.03$. A 95% confidence interval for the true population total, t, is $[758, 1237]$. We are 95% confident that the true total number of breathing holes in the study region is between 758 and 1237.

4. I would sample 8 counties from the Northeast region, 69 counties from the Northcentral region, 122 counties from the South, and 101 counties from the West. My work is shown below.

```
sqrt(7647472708)
sqrt(29618183543)
sqrt(53587487856)
sqrt(396185950266)
220*87449.83+1054*172099.3+1382*231489.7+422*629433
#786171116
300*220*87449.83/786171116
#7.34
300*1054*172099.3/786171116
#69.22
300*1382*231489.7/786171116
#122.08
300*422*629433/786171116
#101.36
```

5. The total number of otter dens found in the sample are shown for each terrain type in the following table. Using the formula for $\hat{t}$, we find that $\hat{t} = 89/19*33+61/20*35+40/22*292+47/21*86 = 984.71$. We round up, and the estimate for the total number of otter dens along the coastline of Shetland, UK is 985.

	terrain	sum(dens)	var(dens)
1	1	33	5.43
2	2	35	6.83
3	3	292	58.78
4	4	86	15.59

The variance of the $\hat{t}$ is estimated to be $\widehat{V(\hat{t})} = 89*(89-19)*5.4269/19+61*(61-20)*6.8289/20+40*(40-22)*58.7792/22+47*(47-21)*15.5905/21 = 5464.31$, and the standard error of $\hat{t}$ is $\sqrt{5464.31} = 73.92$. A 95% confidence interval for the true population total, t, is

$[837, 1132]$. We are 95% confident that the true total number of otter dens along the 1400 km coastline of Shetland, UK is between 837 and 1132.

6. The population I came up with is given below:

Stratum 1 y-value	Stratum 2 y-value
1	1.5
2	2.5
3	3.5

$\bar{y}_{1U} = 2$, $\bar{y}_{2U} = 2.5$, and $\bar{y}_U = 2.25$. Then $SSB = 3(0.25^2 + 0.25^2) = 0.375$ and $\sum_{h=1}^{H}(1 - N_H/N)s_h^2 = .5 * 1 + .5 * 1 = 1$. Since $0.375 < 1$, this is an example where $V(\hat{t}_{STR})$ is larger using proportional allocation than what it would be from a SRS with the same number of observations. I think the main message here is that there is no benefit of proportional allocation if the between strata variance is small and the strata are similar to eachother.