Kyle Beitz

STAT 4155 Homework 4

1. SMOG 8.24

Department	Number of Items	Range
1	12	1 to 12
2	9	13 to 21
3	27	22 to 48
4	40	49 to 88
5	35	89 to 123
6	15	124 to 138
7	18	139 to 156
8	10	157 to 166
9	31	167 to 198
10	26	199 to 224
11	22	225 to 246
12	19	247 to 265
13	16	266 to 281
14	33	282 to 314
15	6	315 to 319
Total -->	319	

Random Numbers		
Cluster 1	202	<-- Dept 10
Cluster 2	33	<-- Dept 3
Cluster 3	62	<-- Dept 4

2. SMOG 8.25

$\underline{Y_1}$**= 2,** $\underline{Y_2}$**= 2,** $\underline{Y_3}$**= 2** $\rightarrow \frac{2}{26}, \frac{2}{40}, \frac{2}{27} = 0.0769, 0.05, 0.0741$

= $\frac{1}{3}(0.0769 + 0.05 + 0.0741) = 0.067$

= $\frac{1}{3(2)}[(0.0769 - 0.067)^2 + (0.05 - 0.067)^2 + (0.0741 - 0.067)^2] = .00007$

Thus, the bound on the error of estimation is:

$$2\sqrt{.00007} = 0.01673$$

The estimate the average number of improperly identified items is: 0.067 ± 0.0167

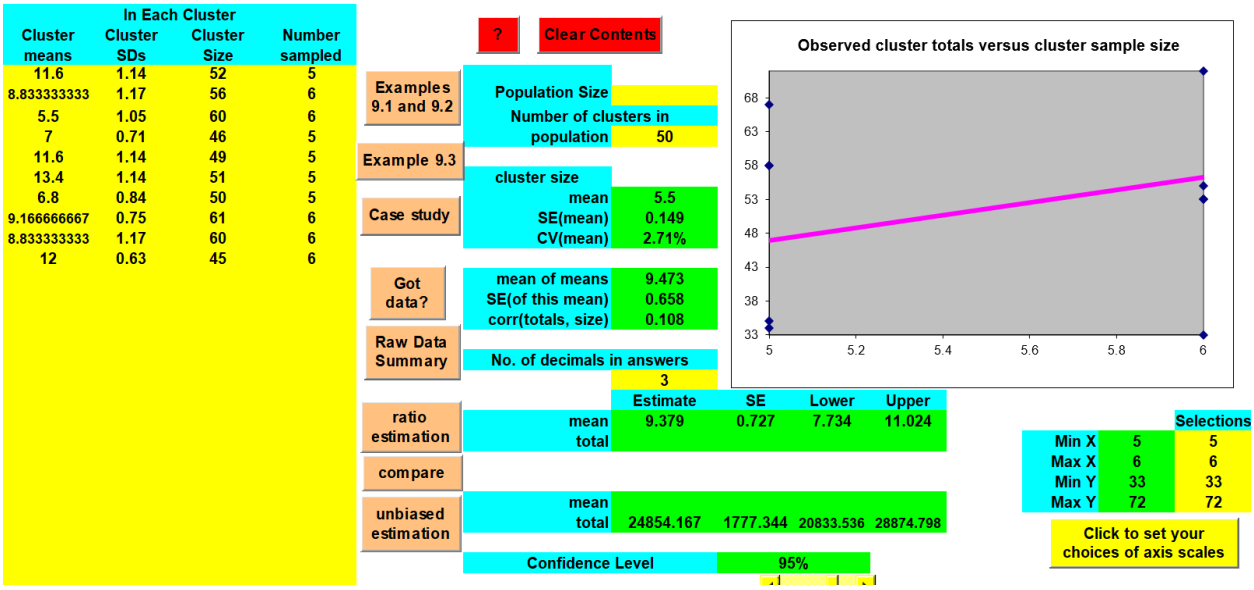
The estimation of total ($\hat{T}_{y,ratio} = 0.067 * 319 = 21.373$ **(where 319 is the number of elements)**

3. SMOG 8.33 (a)

There are a few reasons cluster sampling would be more beneficial than stratified random sampling in estimating the proportion of students favoring a certain candidate in a student government election. For one, stratifying by collections of housing units (dorms, fraternity houses, sorority houses, etc.) seems to be not a very optimal idea because the variation on favorite candidate within each strata would have to be pretty large because students living in

each strata (picture a dormitory) will likely have very different opinions on who they want to be elected. Thus, the main situation to use stratified random sampling, when strata have very similar elements within but strata are meaningfully different from each other, is not being met. Another reason clustering by household seems to be a more useful sampling method is because this will decrease time and cost of the study. Each housing unit can be sampled one after another because these are close in proximity to one another, making it more time and effort efficient than in the StRS case of having to travel to a dorm, sample from it, travel to a fraternity, sample from it, travel to a sorority, sample from it, and so on.

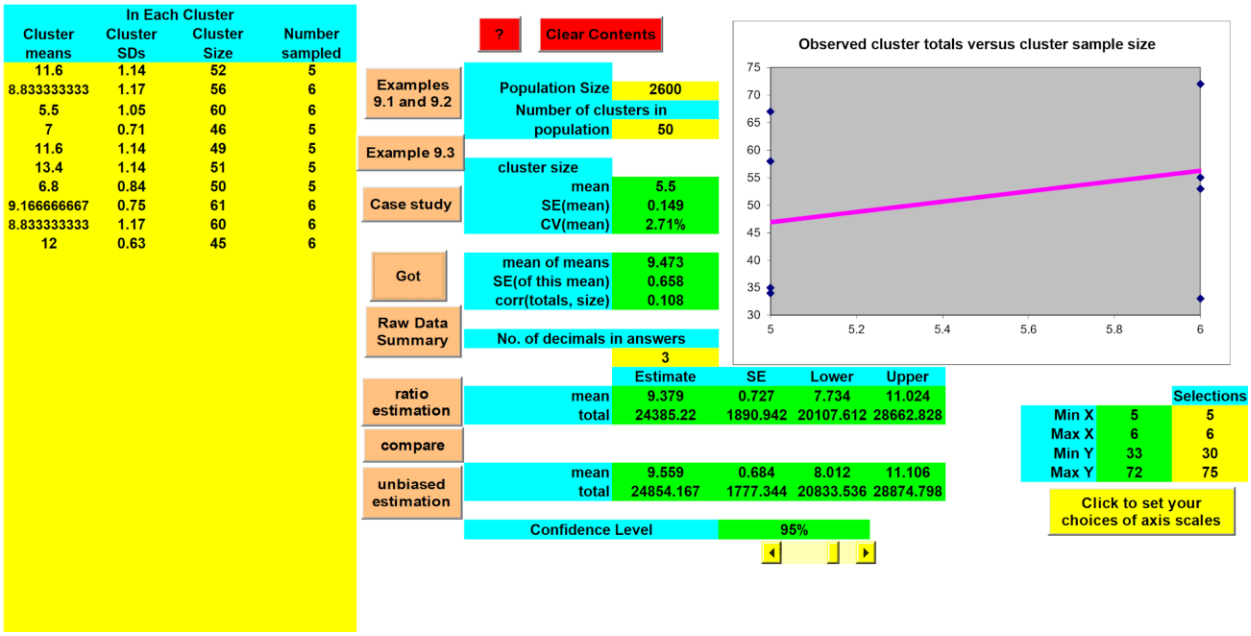
4. SMOG 9.2



The average height of seedlings is 9.379 with a bound on error of 1.645.

5. SMOG 9.3

Cluster						
1	12	11	12	10	13	
2	10	9	7	9	8	10
3	6	5	7	5	6	4
4	7	8	7	7	6	
5	10	11	13	12	12	
6	14	15	13	12	13	
7	6	7	6	8	7	
8	9	10	8	9	9	10
9	7	10	8	9	9	10
10	12	11	12	13	12	12



I used the raw data function of this tool (screenshot of raw data entry above).

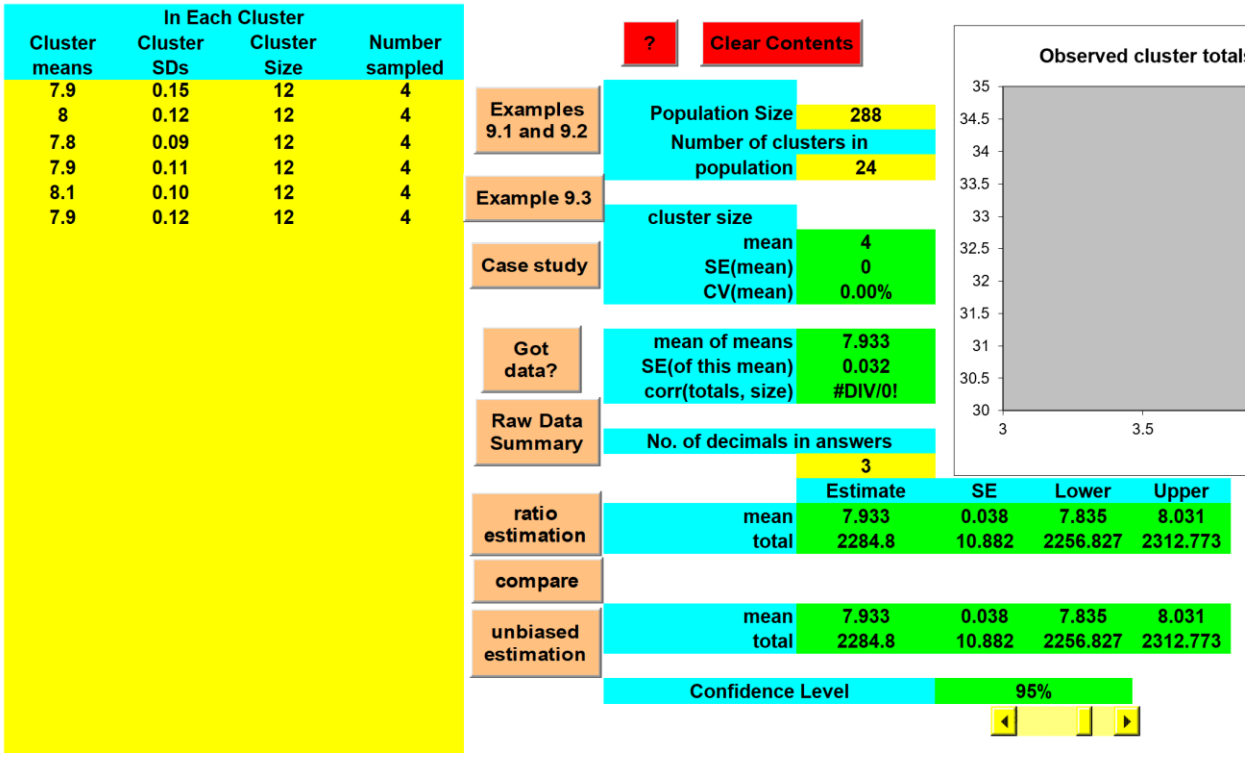
Estimate of the Average Height With a Bound on The Error of Estimation:

$$\hat{\mu} = 9.559$$

$$B = 1.547$$

$$9.559 \pm 1.547$$

6. SMOG 9.9 (next page)



Estimate of the Average Number of Ounces Per Bottle With a Bound on The Error of Estimation:

$$\hat{\mu} = 7.933$$
$$B = 0.098$$
$$7.933 \pm 0.098$$

7.

5 Statistical Sampling Points/Concepts We Understand Better After This Course:

1. The following quote describes a lot of what a statistician must keep in mind when determining sampling methods. WIlliams writes, "Accuracy must be balanced against effort and expense, and very likely a study that is as accurate as is desirable will not be practicable, but if the results are expected to be too inaccurate, then the utility of the study should be questioned" (Williams 2). The reason there are so many sampling methods and time must be taken to determine the optimal one is because each method has its own pros and cons. A method that works perfectly in one study might be a terrible choice for another. It is all about using a method that is accurate, but also practical and cost-efficient as well.
2. "Accordingly, as described in the textbooks cited above, a finite population correction factor should be included in the variance estimates, unless the fraction of the sampling universe contained in the samples is small" (Williams 4). We have talked about the fpc all semester, and this seems to be a good summary of when to use it.
3. The following sentence is so relevant to what we have learned in this class. Williams writes, "Precision is determined partly by the variation in the thing being sampled, which the investigator cannot control, but also by the sample size and the sampling design" (Williams 5). The precision of estimates is determined by many things. The variation is one such thing, and that cannot be controlled in any way. There are, however, a few aspects of the sampling design that can play huge roles in precision that *can* be controlled. As you increase sample size, you increase precision and accuracy of estimates because your sample is getting closer and closer to the real population. While it is sometimes hard to get a large sample, the investigator can also determine sampling methods that will increase accuracy as much as possible, optimizing their predictions based on patterns in the data and cost.
4. "Sampling is ordinary in science and is a well-developed part of statistics, but has received surprisingly little attention in environmental flow assessments" (Williams 2). This sentence hits home for me because before this class I thought sampling was really one dimensional, meaning I thought there was really only one straightforward way of doing it. I have now learned that depending on a large list of variables including cost, data type, time, relationships to name a few that there are many different sampling methods to pick from.
5. "Stratified and multi-stage sampling are common approaches to reducing the effort and cost of the sampling required to achieve a given level of accuracy" (Williams 5). I feel that I can relate to this sentence now that I have taken this class. A Stratified random sample is one obtained by separating the population elements into non-overlapping groups, called strata and then selecting a simple random sample from each strata. The

cost per observation may be reduced by stratification of the population elements into stratas. Stratification may produce a smaller bound on error of estimation than what would be produced by a simple random sample of the same size especially if the stratas are homogeneous.

2 Statistical Sampling Points/Concepts That Aren't Clear:

1. The article mentions sampling units for DFA (Demonstration Flow Assessments). What would sampling units for this look like?
2. On page 6 of the article, Williams explains that spatial stratification could be more useful than systematic sampling. He briefly explains spatial stratification but it wasn't really talked about in depth, most likely because he still didn't think it was the optimal method to use. What is the general procedure of spatial stratification?

1 Fish/Fish-Science-Related Point We Want You to Explain:

1. What is a pool-dwelling and riffle-dwelling fish?