Sampling Corn Activity

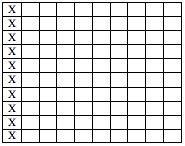
Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

A farmer has just cleared a new field for corn. The plot lies along a river on one side. The corn looks good in some areas of the field, but not in others. It will only be worthwhile for the farmer to harvest the corn if the total yield of the field is high enough. **She wants to accurately estimate the total yield of the corn in the field**. To do this she will harvest 10 of the 100 plots and use the average yield per plot in her sample to estimate the average overall yield of the field. This will determine whether to harvest or not.

We will test different ways the farmer could choose her 10 plots of land by using the applet here: <https://vank-stats.shinyapps.io/riversampling/>

**Method #1: Convenience Sampling**

The farmer’s house is on the left, so her first plan is to select the 10 easiest plots to harvest (seen below). Enter the numbers corresponding to these plots in the applet.



River

What is the total yield of corn for these 10 plots? \_\_\_\_\_\_\_\_\_\_

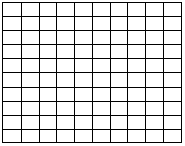
Based on these 10 plots, what is the best estimate of the field’s total yield? \_\_\_\_\_\_\_\_\_\_

Comment on the farmer’s selection. Do you think it will provide a representative sample of the entire field? Why or why not?

The farmer knows that you are taking a statistics class and that you know a thing or two about sampling. She wants your help determining which 10 plots she should harvest to get a good estimate of the overall yield of the field.

**Method #2: Simple Random Sampling (SRS)**

Use StatCrunch (or another method) to *randomly* choose 10 plots to harvest using simple random sampling. In other words, randomly select 10 numbers between 1 and 100. Mark which plots you chose with an X on the field below.



River

What is the total yield of corn for these 10 plots? \_\_\_\_\_\_\_\_\_\_

Based on these 10 plots, what is the best estimate of the field’s total yield? \_\_\_\_\_\_\_\_\_\_

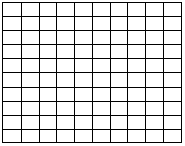
How do your chosen plots look different than in your previous method? Are they spread out differently? Are they grouped together in certain areas? Etc.

Do you think that this will be a representative sample of the entire field? Why or why not?

**Method #3: Stratified Sample (by columns)**

Consider the **columns** to be ten strata (or non-overlapping groups). Use StatCrunch (or another method) to randomly choose one plot from **each** column. You will need to think about how to do this. For example, in column 1 (on the far left) you could randomly choose a number between 1 (top plot) and 10 (bottom plot). Then repeat this for columns 2 through 10. Mark which plots you randomly selected with an X on the field below.

1 2 3 4 5 6 7 8 9 10



River

What is the total yield of corn for these 10 plots? \_\_\_\_\_\_\_\_\_\_

Based on these 10 plots, what is the best estimate of the field’s total yield? \_\_\_\_\_\_\_\_\_\_

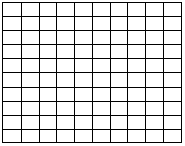
How does the location of your chosen plots look different than in your previous methods?

What was guaranteed to happen with your sample using this method?

Do you think that this will be a representative sample of the entire field? Why or why not?

**Method #4: Stratified Sample (by rows)**

Consider the **horizontal rows** to be ten strata (or non-overlapping groups). Use StatCrunch (or another method) to randomly choose one plot from **each** horizontal row. This will be similar to method 3 but you will select from each row instead of each column. Mark which ones you chose with an X on the field below.



River

What is the total yield of corn for these 10 plots? \_\_\_\_\_\_\_\_\_\_

Based on these 10 plots, what is the best estimate of the field’s total yield? \_\_\_\_\_\_\_\_\_\_

How does the location of your chosen plots look different than in your previous methods?

What was guaranteed to happen with your sample using this method?

Do you think that this will be a representative sample of the entire field? Why or why not?

You have looked at four methods for choosing plots. Give a reason other than convenience for why one method may be preferable over others. Explain your reasoning.

The farmer’s goal was to choose a method that gives her an accurate estimate of the overall yield. If she underestimates the yield she may harvest too late and if she overestimates it she may harvest too early. To see the actual yield of the field (in bushels), type the word **census** in where it asks for a code to reveal the whole field.

The actual overall yield of the farmer’s land was \_\_\_\_\_\_\_\_\_\_\_\_ bushels

Fill in the table below with your estimates based on each of the methods.

|  |  |  |
| --- | --- | --- |
| Sampling Method | Estimate of Total Yield | Difference between estimate and actual yield  (Estimate – 5004) |
| Convenience  Sample | 60 x 10 = 600 | 600 – 5004 = -4404 |
| Simple Random  Sample |  |  |
| Stratified Sample  (by columns) |  |  |
| Stratified Sample  (by rows) |  |  |

**We will compile everyone’s estimates to compare the different methods**

How did your *estimates* from the different sampling methods compare to the actual overall yield and to one another? Would you recommend (or not recommend) certain methods for the farmer to use in the future?

Wait to compare results for the whole class to answer: In hindsight which sampling method would be the best one for the farmer to use? Why?