Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ *Source: NCSSM Statistics Leadership Institute, 2000*

**Rolling Down the River: An Exercise in Sampling**

A farmer has just cleared a new field for corn. It is a unique plot of land in that a river runs along one side. The corn looks good in some areas of the field but not in others. The farmer is not sure that harvesting the field is worth the expense. Out of the 100 plots on the field, he has decided to harvest 10 plots and use this information to estimate the total yield of the whole field. Based on this estimate, he will decide whether to harvest the remaining plots.

**The Question:** Which 10 plots should the farmer choose to harvest?

a) Method Number 1: **Convenience Sample**

The farmer began by choosing the 10 plots that would be easiest to harvest. They are marked on the grid:



Since then, the farmer has had second thoughts about his selection and has decided to come to you (knowing that you are an Statistics student, somewhat knowledgeable, but far cheaper than a professional statistician) to determine the approximate total yield of the field.

You will still be allowed to pick 10 plots to harvest early. Your job is to determine which of the following methods is the best one to use – and to decide if this is an improvement over the farmer’s original plan.

b) Method Number 2: **Simple Random Sample**

Number the plots from 1 to 100. Use your calculator to randomly choose 10 plots to harvest. Mark them on the grid below.



c) Method Number 3: **Systematic Random Sample**

Number the plots from 1 to 100. Use your calculator to randomly choose a number from 1 to 10. Harvest this plot and every 10th plot after it – for example, if you randomly choose the number 7, then harvest plots 7, 17, 27, etc. Mark your harvested plots on the grid below.



d) Method Number 4: **Stratified Sample**

Consider the field as grouped in vertical columns (called **strata**). Number the plots in each column from 1 to 10. Use your calculator to randomly choose one plot from each vertical column and mark these on the grid.



e) Method Number 5: **Stratified Sample**

Consider the field as grouped in horizontal rows (also called **strata**). Number the plots in each row from 1 to 10. Use your calculator to randomly choose one plot from each horizontal row and mark these on the grid.



f) Method Number 6: **Cluster Sample**

Consider the field as grouped in 10 regions as shown. Randomly (select a random # between 1-10) select on region and survey each plots on the grid below.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |

Okay, the crop is ready. Below is a grid with the yield for each of the 100 plots. For each sampling method, find the 10 plots that are harvested early, and find the mean crop yield per plot for this sample. Use the mean yield to estimate the total yield for the entire field.



|  |  |  |
| --- | --- | --- |
| **Sampling Method** | **Mean Yield Per Plot** | **Estimate of Total Yield** |
| 1. Convenience Sample |  |  |
| 2. Simple Random Sample |  |  |
| 3. Systematic Random Sample |  |  |
| 4. Stratified Sample: Vertical |  |  |
| 5. Stratified Sample: Horizontal |  |  |
| 6. Cluster Sample |  |  |

Observations:

1. Is there a reason to choose one sampling method over another?

2. Compare the class boxplots for estimates of total yield. What do you notice? Which sampling method is best?

Suppose the farmer installs an irrigation system in the field to improve the crop yield in plots that are far away from the river. With the irrigation system in place, the plot yields are:



Method Number: **Simple Random Sample**

Number the plots from 1 to 100. Use your calculator to randomly choose 10 plots to harvest. Mark them on the grid below.



e) Method Number 5: **Stratified Sample**

Consider the field as grouped in horizontal rows (also called **strata**). Number the plots in each row from 1 to 10. Use your calculator to randomly choose one plot from each horizontal row and mark these on the grid.



f) Method Number 6: **Cluster Sample**

Consider the field as grouped in 10 regions as shown. Randomly (select a random # between 1-10) select on region and survey each plots on the grid below.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |

|  |  |  |
| --- | --- | --- |
| **Sampling Method** | **Mean Yield Per Plot** | **Estimate of Total Yield** |
| Simple Random Sample |  |  |
| Stratified Sample: Horizontal |  |  |
| Cluster Sample |  |  |

In this scenario, what is your preferred sampling method? Why or why not?

When is stratified sampling more useful?

When is it more useful to use methods like a simple random sample, which doesn’t split subjects into groups first?