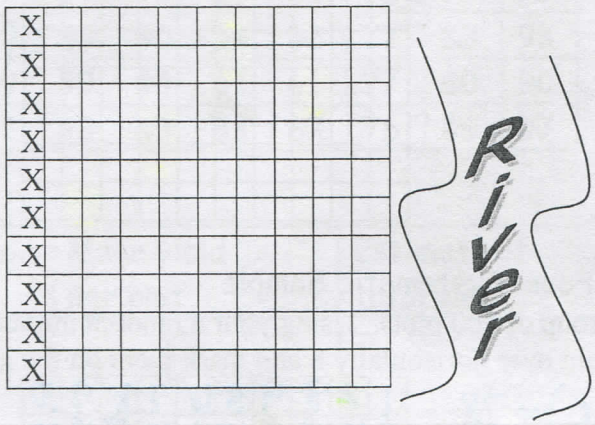
A River Runs Through It **Partners:** Key

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 others. The farmer is not sure that harvesting the field is worth the expense. He has decided to harvest 10 plots and use this information to estimate the total yield. Based on this estimate, he will decide whether to harvest the remaining plots.

Part I.

A. Method Number One: Convenience Sample

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



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

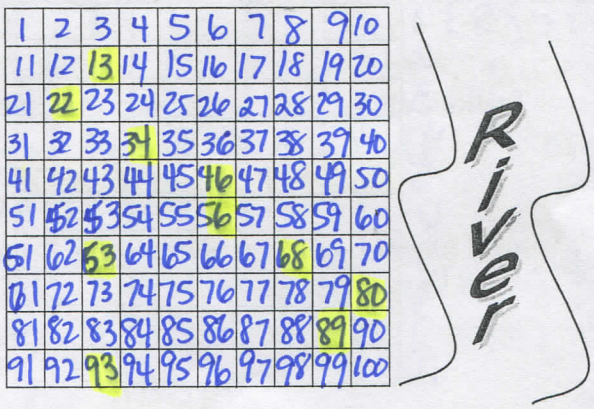
You will still be allowed to pick 10 plots to harvest early. Your job is to determine which of the following methods is the best one to use – and to decide if this is an improvement over the farmer's original plan.

B. Method Number Two: Simple Random Sample

Use the random generator at (https://www.random.org/) to choose 10 plots to harvest. Mark them on the grid below, and describe your method of selection. Note you will do this by labeling the grid below from 1 to 100. Then randomly generate a number and put a X on the plot below that corresponds to that number. If you get repeats, simply generate another random number until you have 10 distinct plots in your sample.

randint(1,100)

{93, 46, 13, 89, 63, 22, 34, 56, 68, 80}

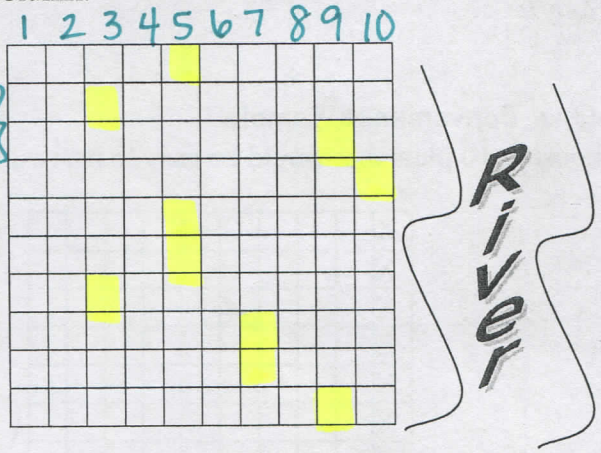


C. Method Number Three: Stratified Sample

Consider the field as grouped in vertical columns (called strata). Using the random generator at (https://www.random.org/), randomly choose one plot from each vertical column and mark on the grid. Note you will do this by labeling plots from1-10 in each vertical column and then select a random number from 1-10 for each column.

randint (1, 10, 10)

{5, 3, 9, 10, 5, 5, 3, 7, 7, 8}

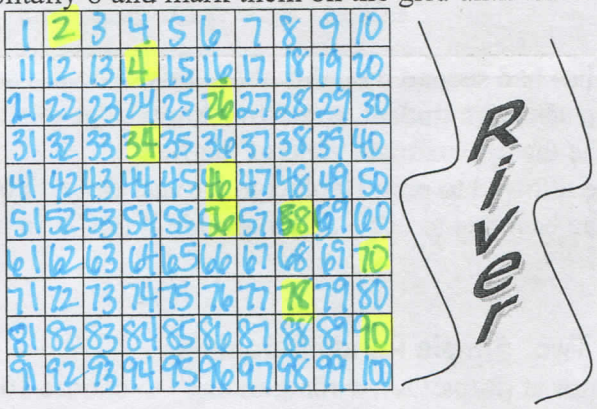


D. Method Number Four: Systematic Sample

Consider the field as group of 100 plots. Using your a random number generator, randomly choose one plot from 100, then count over horizontally 8 and mark them on the grid until ten are marked.

randint (1, 100) = 78

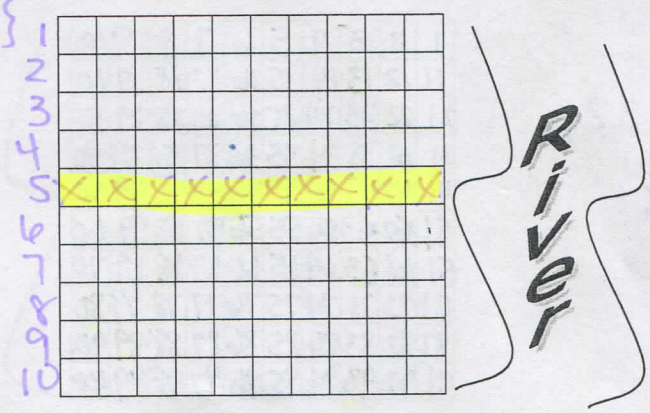
{78, 56, 34, 90, 46, 2, 58, 14, 70, 26}



E. Method Number Four: Cluster Sample

Consider the field as grouped in horizontal rows. Using the random generator at (https://www.random.org/), randomly choose one row to sample.

randint (1, 10) = {5}



OK, the crop is ready. Below is a grid with the yield per plot. Estimate the average yield per plot based on each of the four sampling techniques.

6	17	20	38	47	55	69	76	82	97
7	14	23	34	43	56	63	75	81	92
2	14	28	30	50	50	62	80	85	96
9	15	27	34	43	51	65	72	88	91
4	15	28	32	44	50	64	76	82	97
5	16	27	31	48	59	69	72	86	99
5	18	28	34	50	60	62	75	90	90
8	15	20	38	40	54	62	77	88	93
7	17	29	39	44	53	61	77	80	90
7	19	22	33	49	53	67	76	86	97

6+7+2+9+4+56+8+7+7 / 10

23+14+34+50+59+34+93+80+22 / 10

47+23+85+91+43+44+28+62+61+86 / 10

17+34+50+34+50+59+72+90+77+90

4+15+28+32+44+50+64+76+82+97

Sampling Method	Mean yield per plot	Estimate of total yield
Convenience Sample (farmer's)	6	6 · 100 = 600
Simple Random Sample	409/10 = 40.9	40.9 · 100 = 4090
Vertical Strata	570/10 = 57	57(100) = 5700
Systematic Sample	573/10 = 57.3	57.3(100) = 5730
Cluster Sample	492/10 = 49.2	49.2(100) = 4920

Observations:

You have looked at five different methods of choosing plots. Is there a reason, other than convenience, to choose one method over another?

Depending on if there are any patterns in your population, one sampling method may provide a more reliable estimate.

How did your estimates vary according to the different sampling methods you used?

The convenience sample was much lower than the rest. SRS & cluster were close together, and the Strata & systematic were highest together.

Compare your results to someone else in the class. Were your results similar?

they were the same highest to lowest order but all #'s were off slightly.

Which sampling method should you use? Why do you think this method is best?

Answers may vary. I think I got lucky with systematic. It makes sense to yield better closer to the river, but cluster is the easiest to harvest.

What was the actual yield of the farmer's field?

the actual yield is 4998.

Part II:

The farmer was very impressed with the results of your study and seeks to improve the yield of that part of the field the following year. Believing that irrigation is the answer, a new system was installed. The following year's yield was:

79	81	95	69	65	59	88	65	66	91
80	75	88	80	82	66	76	99	62	61
97	50	92	92	91	84	75	85	63	89
99	71	55	75	65	66	66	86	96	50
57	95	51	79	98	71	70	86	89	76
57	53	90	71	50	76	56	91	85	64
69	95	98	90	93	97	79	95	73	90
58	99	75	51	67	81	55	63	89	74
98	62	73	54	50	76	91	50	90	55
91	59	69	59	71	72	85	85	86	97

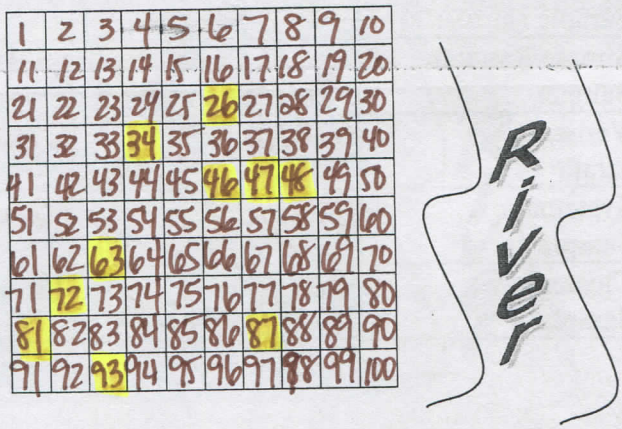
River

Redo your sampling using a SRS, vertical stratification, and cluster. Be certain to mark on the grids the plots you choose.

A. Simple Random Sample:

randint (1,100,10)

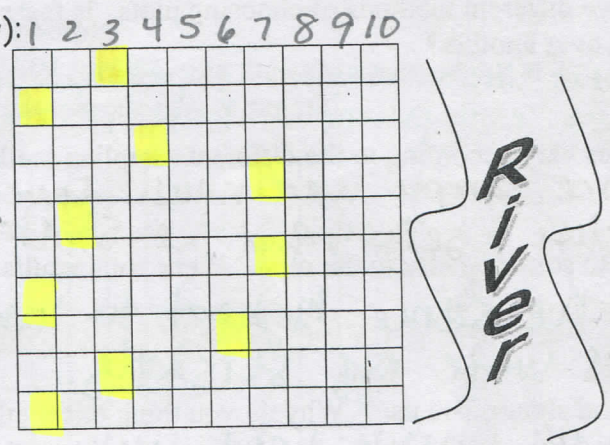
{93, 63, 34, 46, 81, 87, 26, 72, 47, 48}



B. Stratified Sample (vertically):

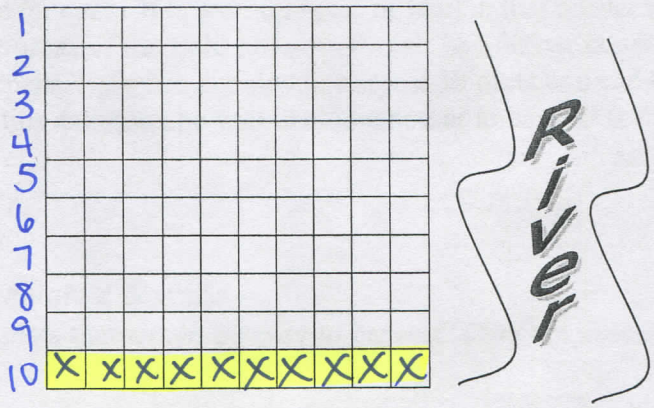
randint (1,10,10)

{3, 1, 4, 7, 2, 7, 1, 6, 3, 1}



C. Cluster Sample (1 horizontal row):

randint (1,10) = {10}



84+75+71+70+86+
98+99+98+69+91
95+80+92+66+95+
56+69+81+73+91
91+59+69+59+71+
72+85+85+86+97

Sampling Method	Mean yield per plot	Estimate of total yield
Simple Random Sample	841/10 = 84.1	84.1(100) = 8410
Vertical Strata	798/10 = 79.8	79.8(100) = 7980
Cluster Sample	774/10 = 77.4	77.4(100) = 7740

actual yield is 7603 in part II.

Observations:

Based on the results of both activities, under what conditions is it more useful to use stratified sampling?

When there is a reasonable expectation or a known pattern in the data.

Based on the results of both activities, under what conditions is it more useful to use a simple random sample?

when there is no known pattern in the data.