Learning Goal: Identify the sampling plan for a study. Recognize the implications and limitations of the plan.

Learning Objectives:

- Define statistical bias.
- Explain how random selection eliminates bias.
- Identify sampling plans that will tend to give the most accurate samples.

Introduction: One of the most important ideas in college-level statistics is that we can learn a lot about a large group (a population) by studying a small piece of it (a sample.)

On the next page you will see a large group of 100 rectangles. We will view this as a population. We want to estimate the average area of the rectangles in this population by taking a sample. In this activity we will use different methods of collecting samples and see which gives us the most reliable estimate of the mean area for the population of all rectangles.

Your initial estimate: Glance through the population of rectangles. What do you think is the average area?

Sampling method #1: Select 5 rectangles that look representative of the rectangles. In other words, choose 5 with an area that you think looks pretty typical. For each rectangle that you chose, write down the rectangle number and its area.

Rectangle number					
Area of the rectangle					

Now find *the mean of the areas of the 5 rectangles that you chose.*
Round your answer to one decimal place (e.g. 8.25 = 8.3) _____

If you have time, select another sample of 5 rectangles and repeat what you just did.

Rectangle number					
Area of the rectangle					

Now find *the mean of the areas of the 5 rectangles that you chose.*
Round your answer to one decimal place (e.g. 8.254 = 8.3) _____

Sampling method #2: We will now select 5 rectangles at random using a random number generator.

Go to *www.rossmanchance.com/applets*. Under Probability, click on *Random Number Generator.* Enter the following settings:
Number of replications to 1 (We will select one sample.)
Numbers per replication to 5 (We will select 5 rectangles for each sample.)
Number range: From 1 to 100. (Each rectangle has a number.)
With replacement? No (Once a rectangle is chosen, it is removed from the list.)
Sort Results? Yes (Seeing an ordered list will make it easier to find the areas.)
Click Generate.

Record the numbers of the random rectangles from the Random Number Generator in the table. Then find the areas in the diagram on the next page.

Rectangle number					
Area of the rectangle					

Now calculate the *mean of the areas of the 5 randomly chosen rectangles*
Round your answer to one decimal place (e.g. 8.25 = 8.3) _____

If you have time, repeat this. Select another random sample of 5 rectangles, etc.

Rectangle number					
Area of the rectangle					

Calculate *the mean of the areas of the 5 randomly chosen rectangles*
Round your answer to one decimal place (e.g. 8.25 = 8.3) _____

Sampling method #3: We will now select 10 rectangles at random. Set **Numbers per replication** to 10. As before, record the number of the chosen rectangles in the table. Then look up the area of each rectangle in the diagram on the next page.

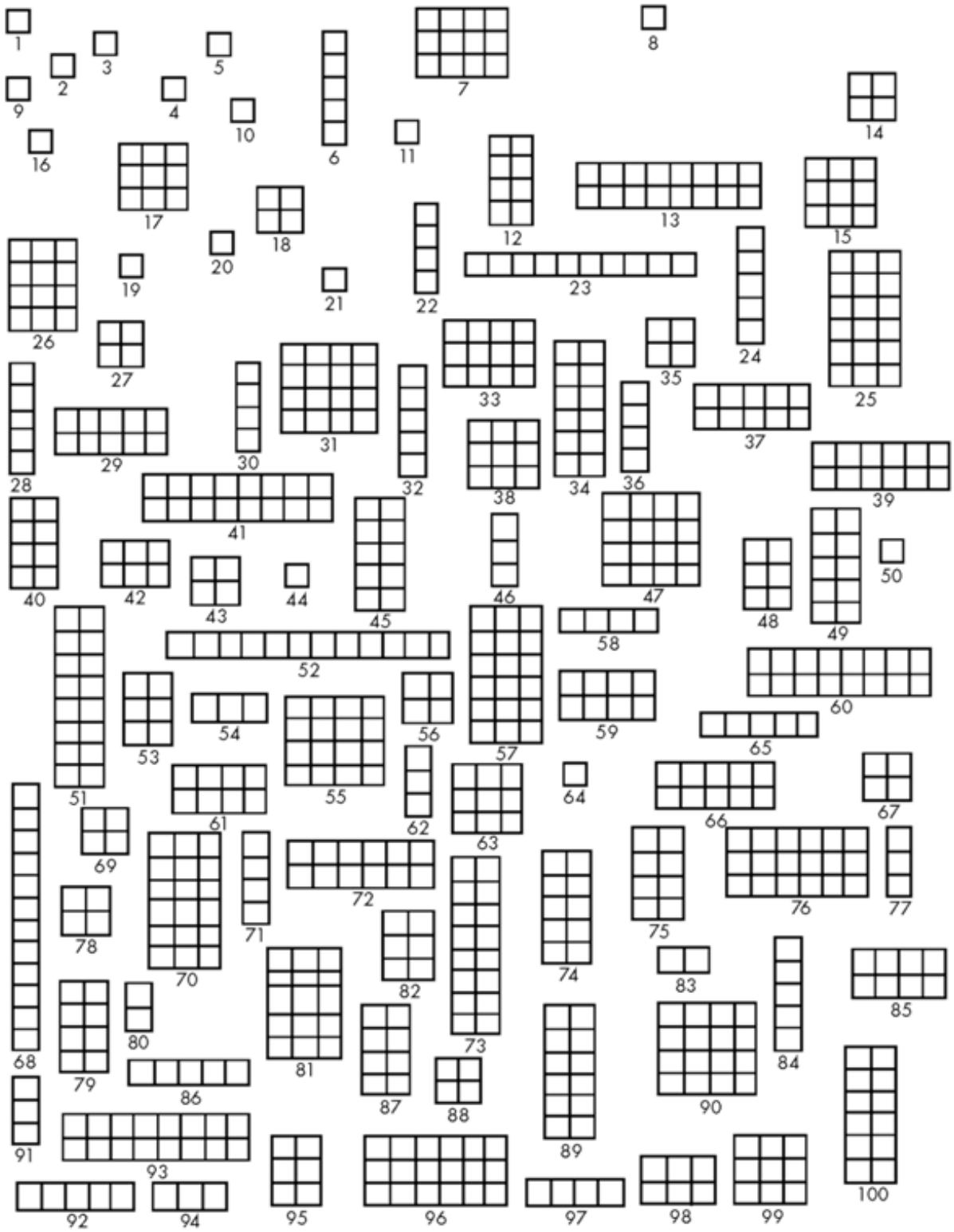
Rectangle number										
Area of the rectangle										

Now calculate the *mean of the areas of the 10 randomly chosen rectangles*
Round your answer to one decimal place (e.g. 8.25 = 8.3) _____

If you have time, repeat this. Select another random sample of 10 rectangles, etc.

Rectangle number										
Area of the rectangle										

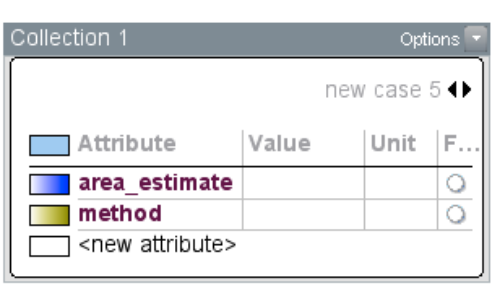
Calculate the *mean of the areas of the 10 randomly chosen rectangles*
Round your answer to one decimal place (e.g. 8.25 = 8.3) _____

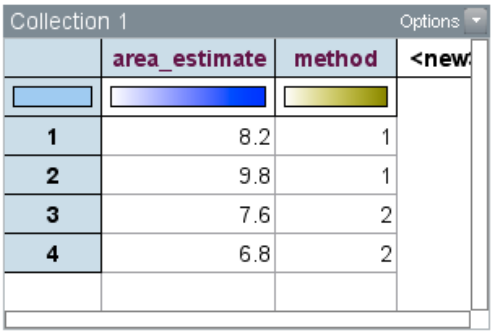

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

We will do this part of the activity together. After everyone is done collecting their samples and finding the mean of each sample, we will compare the results from our three methods. We will use Tinkerplots.

Open a new Tinkerplots file. Click and drag Cards to the work area. Define 2 attributes: area estimate and method.

Now click and drag a table to the work area. You should see two column headings: area estimate and method. We will record each student's results. Each student should report 6 results. Each result will have an associated method.





Now make a dot plot for each sampling method and answer the questions below.

1) For each of the three dot plots describe what a dot represents.

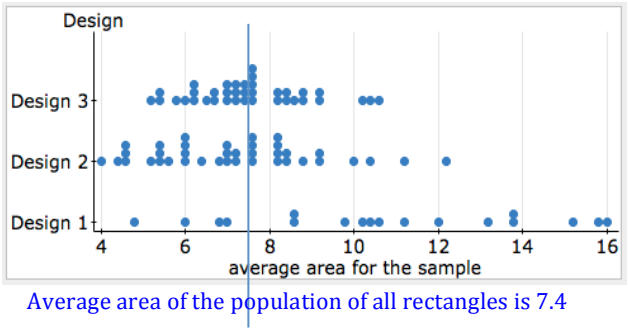
2) The mean area of population of all rectangles is 7.42. Compare our subjective samples of 5 rectangles from Method 1 to the random samples of 5 rectangles from Method 2. Which sampling method produces samples that more accurately estimate the mean of the population? Use the dot plots to support your answer.

3) Now compare the random samples of 5 rectangles from Method 2 to the random samples of 10 rectangles from Method 3. Which sampling method produces samples that more accurately estimate the mean of the population? Use the dot plots to support your answer.

4) After hearing the class discussion of this activity, what do you think are the big ideas about sampling that are demonstrated by this activity?

5) Here are the results from a previous class that did this exact activity with sampling rectangles. They used the same three methods to select samples.

Recall the definition of statistical bias: A sampling method is biased if it systematically favors some outcomes over others.



Average area of the population of all rectangles is 7.4

a) Which sampling method is biased? How can you tell by looking at the graphs?

b) How do these graphs illustrate that random samples give unbiased results?

c) How do these graphs illustrate that larger random samples give more precise estimates of the population mean than smaller random samples?

6) Did I Get This?

Suppose that we want to estimate the mean number of text messages sent by LMC students each day.

Which sampling design is the best for producing this estimate?

- Select 50 students at random from the list of student InSite email addresses. (All LMC students have InSite addresses.)
- Select 100 students at random from the list of student InSite email addresses.
- Select the first 200 students who you see texting in the Quad.
- Select 200 students at random who follow LMC on twitter.

Jot a few notes to explain your choice.