4.1a Time in the Hospital

Name(s):

What you'll need: the Fathom document **HospitalSim.ftm**

In this activity, you'll estimate the average length of stay in a five-bed hospital ward. You'll use a pre-made Fathom simulation of this hospital ward. In the simulation, each bed gets filled with a patient whose length of stay is chosen at random from between 1 and 10 days. You'll estimate the average length from a sample of five patients.

1. Open **HospitalSim.ftm.** You'll see a case table that is similar, but not identical, to the one shown here. Each column (attribute) corresponds to one bed in the ward. The row number indicates the day. Look, for example, at **Bed5Stay.** The first 2 in that column stands for a person who comes to that bed on the 1st day and stays for 2 days. The second 2 in that column represents a continuation of that person's stay—the second day in his or her two-day stay. Continuing down, we have seven 9's representing a person who stays 9 days.

Days

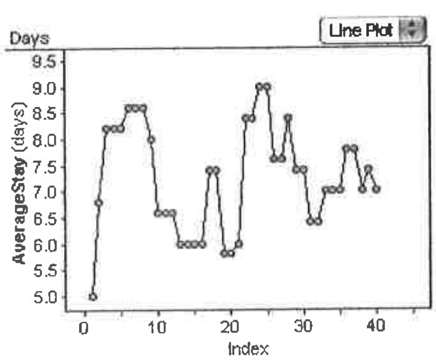
	Bed1Stay	Bed2Stay	Bed3Stay	Bed4Stay	Bed5Stay
1	1 d	5 d	8 d	9 d	2 d
2	10 d	5 d	8 d	9 d	2 d
3	10 d	5 d	8 d	9 d	9 d
4	10 d	5 d	8 d	9 d	9 d
5	10 d	5 d	8 d	9 d	9 d
6	10 d	7 d	8 d	9 d	9 d
7	10 d	7 d	8 d	9 d	9 d
8	10 d	7 d	8 d	9 d	9 d
9	10 d	7 d	5 d	9 d	9 d

One way to attack the problem of estimating the average length of stay is to compute the average length of stay for the five patients in the beds on each day, then choose a random day as typical.

2. In the case table, define a new attribute, **AverageStay,** with the formula

$$\frac{(\textbf{Bed1Stay} + \textbf{Bed2Stay} + \textbf{Bed3Stay} + \textbf{Bed4Stay} + \textbf{Bed5Stay})}{5}$$

3. Make a line plot of **AverageStay.** Make a new graph, drag **AverageStay** to the y-axis, and choose **Line Plot** from the pop-up menu. What does each dot in the plot correspond to?

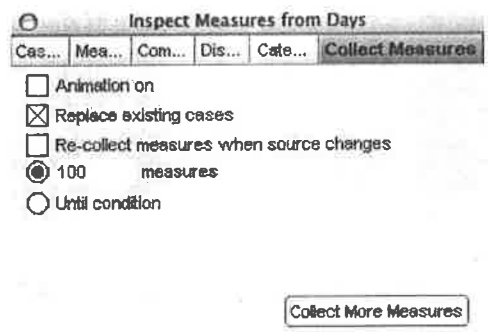


You can also use the keyboard shortcut ⌘ + Y (Macintosh) or Ctrl + Y (Windows) to rerandomize your data.

4. Choose **Rerandomize** from the **Collection** menu. Notice that the case table changes and you get a new plot of **AverageStay.** Do this several times to see how the randomized data change. You probably notice that there is nothing special about any given day. In fact, you are going to treat the 40th day as a typical day and record the value of **AverageStay** over and over again as Fathom rerandomizes.
5. To record the value of **AverageStay** on the 40th day, you need a measure. Double-click the **Days** collection to bring up its inspector. Click the Measures panel and define a new measure, **AnAverageStay,** using the formula **last(AverageStay).** You should see that the value of **AnAverageStay** is, in fact, the last value in the **AverageStay** column of the case table. Close the inspector.

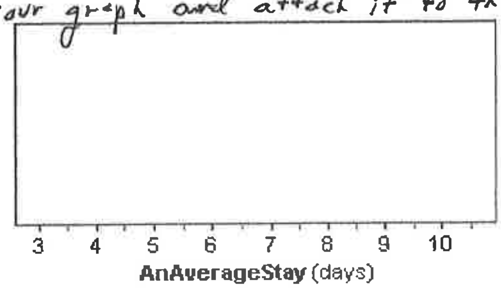

Measures from Days

6. Select the **Days** collection and choose **Collect Measures** from the **Collection** menu. By default, Fathom rerandomizes **Days** five times and places the **AnAverageStay** measures in a new collection named **Measures from Days.**
7. Bring up the inspector for **Measures from Days** by double-clicking the collection. On the Collect Measures panel, change the number of measures to **100.** You may want to uncheck "Animation on" to make it go faster. Click the **Collect More Measures** button.



8. Make a dot plot of the 100 averages. Make a new graph, select the Cases panel of the inspector for **Measures from Days,** and drag the attribute **AnAverageStay** to the *x*-axis of the graph. ~~Sketch the result here.~~ Where is the distribution centered? Is there much variability in your estimates?

Print out your graph and attach it to this handout.



9. If a patient's length of stay is randomly chosen between 1 and 10 days, what is the average stay for the whole population?

10. Compare your results from steps 8 and 9. Are your estimates clearly too low, clearly too high, or about right?

11. You are trying to estimate the average length of stay of a patient. In the foregoing simulation, did every possible length (1–10) have an equal chance of being in the sample? If so, explain. If not, which lengths had the greater chance? Explain.

Suggested Assignments

Classwork		
Essential	**Recommended**	**Optional**
D2–6 P1–5	Activity 4.1a D1, D7, D8 P6	

Homework		
Essential	**Recommended**	**Optional**
E1, E3, E5, E7, E9	E2, E4, E6, E8, E10, E11, E13	E12, E14

Lesson Notes: Why Take Samples, and How Not To

This section is part cautionary tale and part catalog of common sampling pathologies.

Lesson Notes: Bias

Activity 4.1a: Time in the Hospital

For this activity, each group of students will need a deck of cards. You can give your students copies of the blackline master of an empty grid at the end of this section on page 176. This grid will speed the data collection process.

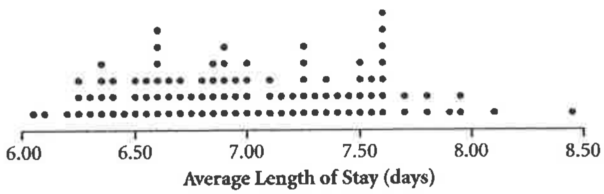
Before students start this activity, be sure they understand that the numbers on the playing cards represent the lengths of stay of 40 different patients. The order in which the patients arrive at the hospital is determined by shuffling the deck. Each patient is put into the first bed available. The goal is to estimate the average number of days a patient stays in the hospital.

This activity illustrates bias in the method of taking a sample. The method of sampling of choosing days at random results in longer stays being chosen more often than shorter stays. This results in an estimate of an average length of stay that is too long.

The bias is in the method, not in the result. A biased sampling method tends to result in a nonrepresentative sample, but not always. The dialogue between the student and the statistician in the student book on pages 222 and 223 reinforces the difference between a biased sampling method and a nonrepresentative sample.

1. Students randomly select five cards. See Display 4.1 on page 221 for an example.
2. See Display 4.2 on page 221 for an example.
3. See Display 4.3 on page 221 for an example.
4. Suppose the day selected is 12.
5. For this example, on day 12, the average stay is $\frac{3+1+7+7+5}{5} = 4.6$ days.

6. This figure shows a typical plot, based on a computer simulation of 100 trials of the activity. The distribution of these sample averages is centered at about 7 days with not much variability ($SD \approx 0.5$).



7. The population average is 5.5:

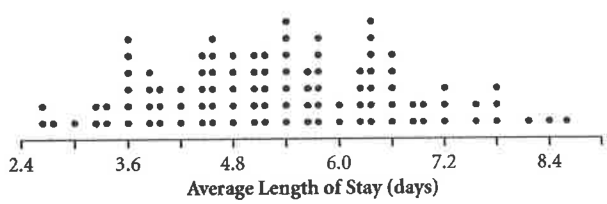
$$\frac{(1 + 1 + 1 + 1 + 2 + 2 + 2 + 2 + \cdots + 10 + 10 + 10 + 10)}{40}$$

$$= \frac{(1 + 2 + \cdots + 10)}{10} = \frac{[(10)(11)/2]}{10} = 5.5 \text{ days}$$

8. The center of the dot plot lies well to the right of the population mean of 5.5, and in general, class estimates should be high. The sampling method is more likely to choose long stays than to choose short stays. If students do not see the bias, ask how many groups had a patient who stayed 1 day, 2 days, . . . , 10 days in their sample. They should begin to see the bias as longer stays are mentioned more often than shorter stays. size bias

9. The units are the individual patients. Each patient did not have an equal chance of being chosen. Patients with longer stays are more likely to be chosen. A stay of, say, 7 days had seven times the chance of being chosen as a stay of only 1 day.

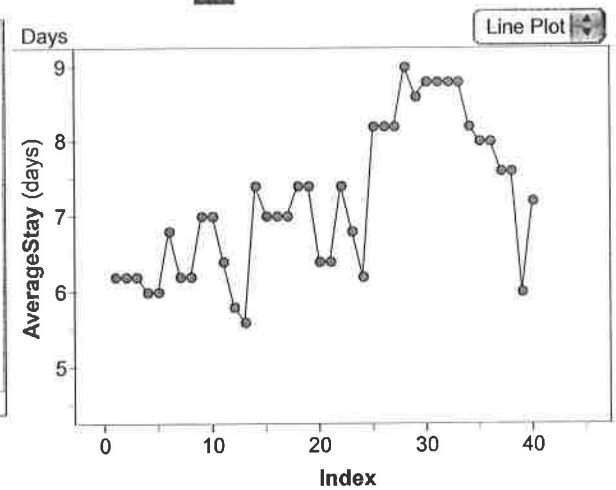
10. A better way to take the sample is to randomly sample five patient records rather than the five patients in the hospital on a randomly selected day. (See the following Optional Extension to Activity 4.1a.) Having each student sample five "records" from the cards set aside in steps 2 and 3 and calculating the average for those five "records" will generate a distribution of the mean length of stay for this sampling method. The dot plot below shows averages for 100 random samples of size 5. This distribution is centered roughly at 5.4 days, very close to the population mean of 5.5. The variability here is larger ($SD \approx 1.26$).

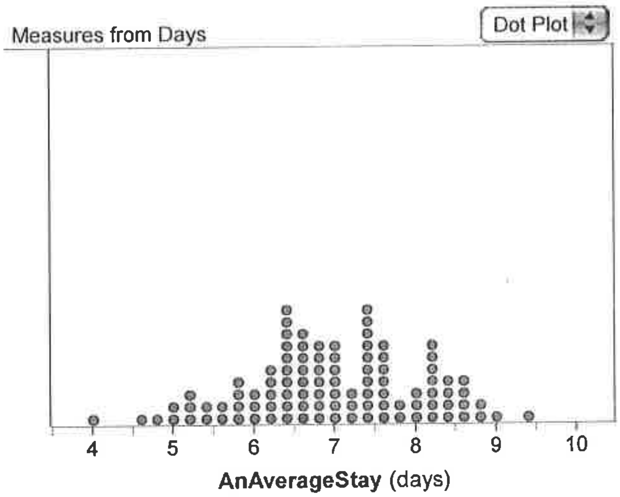


The most important difference between this plot and the one for the size-biased method in step 6 is that the averages for the random samples (described above) tend to cluster near the population average of 5.5, whereas those samples from the biased method tend to be too large. A second difference is that the averages from the random samples are spread out more than those from step 6, because you are likely to get both short and long stays in a random sample.

Days

	Bed1Stay	Bed2Stay	Bed3Stay	Bed4Stay	Bed5Stay	Average...
1	3 d	9 d	10 d	3 d	6 d	6.2 d
2	3 d	9 d	10 d	3 d	6 d	6.2 d
3	3 d	9 d	10 d	3 d	6 d	6.2 d
4	3 d	9 d	10 d	2 d	6 d	6 d
5	3 d	9 d	10 d	2 d	6 d	6 d
6	3 d	9 d	10 d	6 d	6 d	6.8 d
7	4 d	9 d	10 d	6 d	2 d	6.2 d
8	4 d	9 d	10 d	6 d	2 d	6.2 d





Measures from Days

AnAverageStay	6.964 d

S1 = mean ( )