7.1 What is a Sampling Distribution?

Sampling Movies (Again)

Populations are large groups of individuals (people or things) that we would like to understand. In reality, statisticians rarely get a census (data on the whole population). The goal of this unit, however, is to understand the mathematics of random sampling. To understand that, we will start with a population for which we *do* have a census: the top 200 movies of 2014. Get a laptop, go to our course website, download the file "7.1 Top Movies of 2014.ftm," and then open it in Fathom.

1. Who/what are the individuals/cases/observational units?

2. What are the variables? List them.

Taking Samples (Genre)

3. Let's focus on the genre of the movie. Is *genre* a categorical or quantitative variable?

4. Use Fathom to find the proportion of all 200 movies that are comedies. Is this value a parameter or statistic? Record this value with an appropriate symbol.

5. Make an appropriate graphical display of the **population distribution** of *genre*. (To make this simple, you can collapse all of the other genres into one category called "not comedy".)

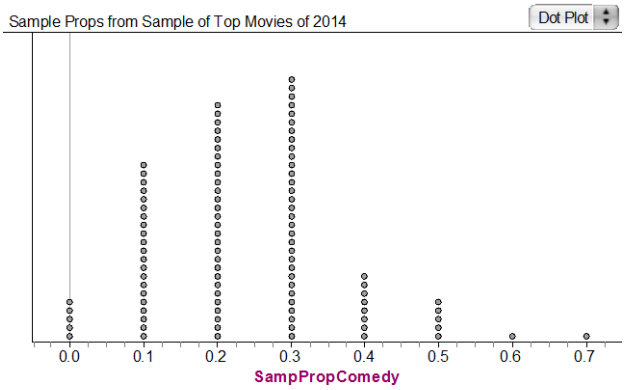
6. Use Fathom to take a random sample of 10 movies from the population, without replacement. Make an appropriate graphical display of the **sample distribution** of *genre*. (To make this simple, you can collapse all of the other genres into one category called "not comedy".)

7. If you take other random samples of 10 movies, would you expect to get the same sample as you did this time? What is the name for this phenomenon?

8. Calculate the proportion of these 10 movies that are comedies. Is this value a parameter or statistic? Record this value with an appropriate symbol.

9. Mr. Tyson will now walk you through the steps to have Fathom calculate the sample proportion of comedies for this sample. Then, take 4 more samples and record the sample proportions below, along with an appropriate symbol. Did you get the same sample proportion each time?

Turn off the animation and collect more sample proportions until you get a total of 100 sample proportions. Make a dotplot of these sample proportions. This dotplot is NOT the **population distribution**, NOR is it the **sample distribution**. This dotplot is the simulated **sampling distribution** of the sample proportion of comedies. One such simulated sampling distribution is show here, along with the mean and standard deviation of the sample proportion of comedies.



Sample Props from Sample of Top Movies of 2014

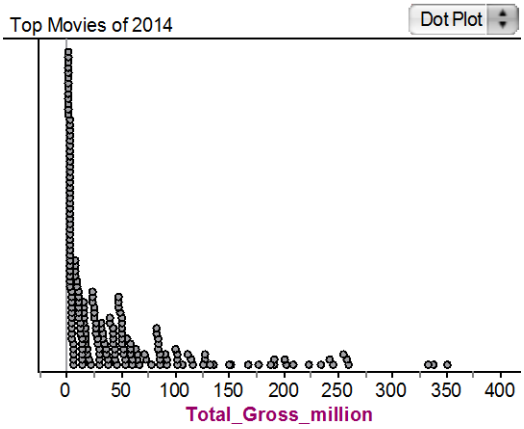
	SampPropComedy
	0.24
	0.132574

S1 = mean ( )
S2 = stdDev ( )

10. There is one dot at 0.6. Explain carefully what that dot represents.

11. Based on the simulated sampling distribution above, would you be surprised to see a sample proportion of comedies of 0.6 or more? Explain why or why not.

TAKING SAMPLES (TOTAL GROSS)

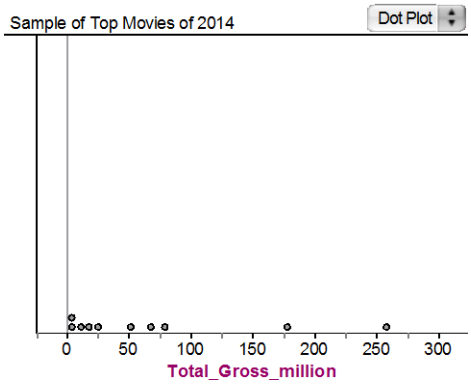


12. Let's focus on the total gross income of the movie. Is *total gross* a categorical or quantitative variable?

13. To the right is a histogram of the **population distribution** of *total gross*. How would you describe its shape?

14. There is one dot at 222.5. Explain carefully what this dot represents.

15. Use Fathom to find the mean, median, and maximum *total gross* of all 200 movies. Are these values parameters or statistics? Record these values with appropriate symbols.



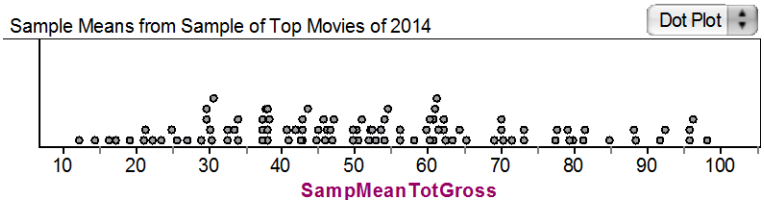
16. Use Fathom to take a random sample of 10 movies from the population. To the right is a dotplot of the **sample distribution** of *total gross* from one such sample. Explain what the dot at 177 represents.

17. Use Fathom to find the mean *total gross* of these 10 movies. Is this value a parameter or statistic? Record this value with an appropriate symbol.

18. Mr. Tyson will now walk you through the steps to have Fathom calculate the sample mean *total gross* for this sample. After he does this, take 4 more samples. Did you get the same sample (and sample mean) each time?

19. How do your sample means for *total gross* compare to the population mean *total gross*?

Turn off the animation and collect more sample means until you get a total of 100 sample means. Make a dotplot of these sample means. This dotplot is NOT the **population distribution**, NOR is it the **sample distribution**. This dotplot is the simulated **sampling distribution** of the sample mean *total gross*. One such simulated sampling distribution is shown to the right, along with the mean and standard deviation of these 100 sample means.



Sample Means from Sample of Top Movies of 2014

	SampMeanTotGross
	51.884
	21.006

S1 = mean ( )
S2 = stdDev ( )

20. There is one dot at 26.9. Explain carefully what that dot represents.

21. Based on this simulated sampling distribution, would you be surprised to see a sample mean *total gross* of $26.9 million or less? Explain why or why not.

22. Go back and make a small change to your work so that you can collect 100 sample median *gross incomes* for samples of 10 movies. Describe the shape, center, and variability of the simulated sampling distribution.

23. Go back and make a small change to your work so that you can collect 100 sample maximum *gross incomes* for samples of 10 movies. Describe the shape, center, and variability of the simulated sampling distribution.

24. What would happen if you were to change the sample size to 20 movies? Make a guess and then try making this change and describe how the simulated sampling distributions of the sample proportion, sample mean, sample median, and sample maximum change.