

## 7.1 SAMPLING DISTRIBUTIONS: BIAS, VARIABILITY, SHAPE

### A COMPLETE SAMPLING DISTRIBUTION

Suppose one year the major movie studios are attacked by hackers and only 6 movies are released that year. These are B movies and so they don't gross very much. Here are the *total gross* earnings for these 6 movies in millions of dollars: {2, 4, 4, 6, 10, 10}

1. Compute the population mean *total gross*.
2. We are interested in taking samples of size  $n = 2$  from a population of size  $N = 6$ . Combinatorics formulas tell us that there are  ${}_6C_2 = 6!/(4! \times 2!) = 15$  possible samples of size  $n = 2$ . List all samples of size  $n = 2$  below.
3. Next to each sample you listed, write the sample mean  $\bar{x}$ . On the number line below, make a dotplot of all 15 sample means. Carefully label the x-axis. What is the name for this distribution?



4. Find the mean of the sample means (mean of the  $\bar{x}$ s). Does this number look familiar?

When the mean (center) of a sampling distribution is equal to the corresponding parameter, we say that the statistic is an **unbiased estimator** of the population parameter.

5. Is the sample mean a biased or unbiased estimate of the population mean? How can you tell?

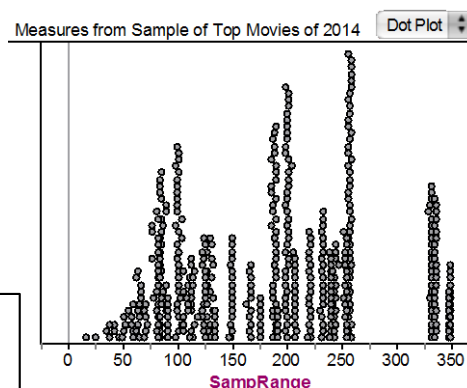
## BIAS AND VARIABILITY AND SHAPE

6. Let's go back to the movies of 2014. What is the population range?

Use Fathom to simulate the sampling distribution of the sample range of *total gross* for samples of size 10. Sample without replacement and collect about 500 different sample ranges. A screenshot of the distribution from one such simulation is shown, along with the mean and standard deviation of these 500 statistics.

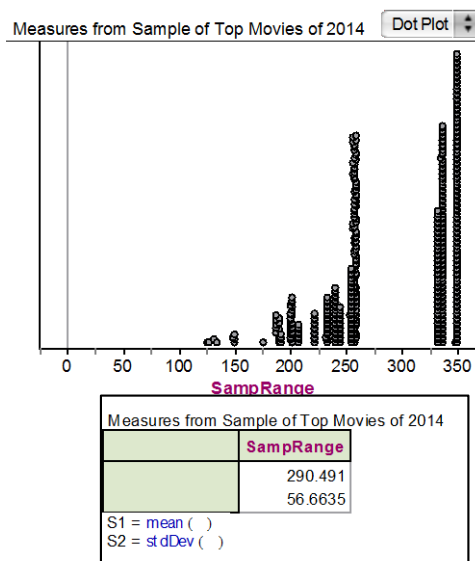
7. From this simulation, do you suspect the sample range is a biased or unbiased estimator of the population range? Explain.

Measures from Sample of Top Movies of 2014	
	SampRange
	187.578
	85.5028
S1 =	mean ( )
S2 =	stdDev ( )



Change the sample size in Fathom to samples of size 40. Again collect 500 different sample ranges, but for samples of size 40. A screenshot of the distribution from one such simulation is shown, along with the mean and standard deviation of these 500 statistics.

8. How does the variability of the sample ranges in this simulation compare to the variability of the sample ranges in the previous simulation? Why does this make sense?



9. Complete this statement (and be specific): As sample size increases ...

**About shape:** We have seen many different shapes for sampling distributions. In general, for quantitative variables the shape of a single sample should look roughly like the shape of the population. The shape of the sampling distribution, however, is much more complex. It does not necessarily look like the shape of the population.