HW 1~~3~~4-1

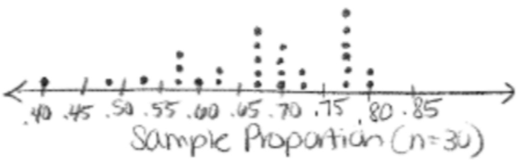
1. a. ~ symmetric centered around 0.50

   b. 0.51 the mean of sampling distribution

   c. SD will decrease

2. Histogram A b/c it has less variability than Histogram B.

3. a.



   b. Skewed Left

   c. $\bar{x}$ = 0.67 S_x = 0.10

   d. Likely 0.7 since it's near the center of the dotplot. Unlikely prop. of all juniors/seniors is 0.5 since there are very few samples with sample prop. of 0.5 or less.

   e. Sampling distr. would be mound shaped with ~ same mean as sampling distr. size 30, but SD of sampling distr. based on size 60 would be smaller than one based on samples of size 30.

Name *Key* Algebra 2 Homework 15-1

1. A group of eleventh graders wanted to estimate the population proportion of students in their high school who drink at least one soda per day. Each student selected a different random sample of 30 students from the high school and calculated the proportion that drink at least one soda per day. The dot plot below shows the sampling distribution. This distribution has a mean of 0.51 and a standard deviation of 0.09.

   a. Describe the shape of the distribution.

   ~ symmetric centered around 0.50

   b. What is your estimate for the proportion of all students who would report that they drink at least one soda per day? *.51 → mean of sampling distr.*

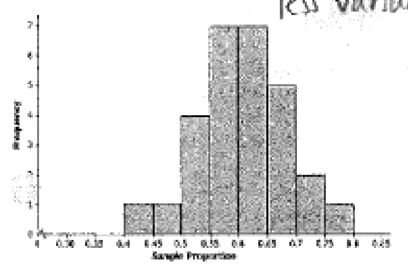
   c. If, instead of taking random samples of 30 students in the high school, the eleventh graders randomly selected samples of size 60, describe what will happen to the standard deviation of the sampling distribution of the sample proportions.

   SD will decrease

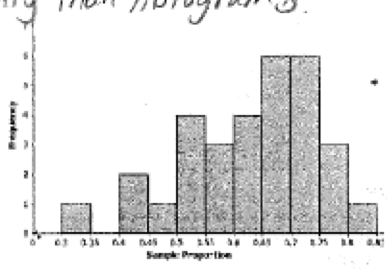
2. A group of eleventh graders wanted to estimate the proportion of all students at their high school who suffer from allergies. Each student in one group of eleventh graders took a random sample of 20 students, while another group of eleventh graders each took a random sample of 40 students. Below are the two sampling distributions (shown as histograms) of the sample proportions of high school students who said they suffer from allergies. Which histogram is based on random samples of size 40? Explain.

Histogram A b/c it has less variability than Histogram B.

Histogram A

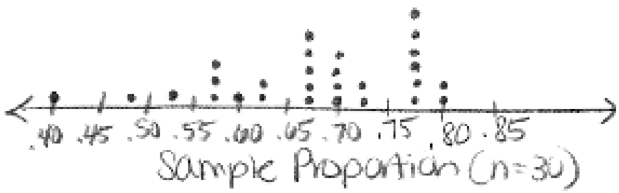


Histogram B



3. A class of 28 eleventh graders wanted to estimate the proportion of all juniors and seniors at their high school with part-time jobs after school. Each eleventh grader took a random sample of 30 juniors and seniors and then calculated the proportion with part-time jobs. Following are the 28 sample proportions.

0.7, 0.8, 0.57, 0.63, 0.7, 0.47, 0.67, 0.67, 0.8, 0.77, 0.4, 0.73, 0.63, 0.67, 0.6, 0.77, 0.77, 0.77, 0.53, 0.57, 0.73, 0.7, 0.67, 0.7, 0.77, 0.57, 0.77, 0.67



a. Construct a dot plot of the sample proportions.

b. Describe the shape of the distribution. Skewed Left

Stat-Calc-1

c. Using technology, find the mean and standard deviation of the sample proportions.

$\bar{X} = .67$ $S_x = .10$

d. Do you think that the proportion of all juniors and seniors at the school with part-time jobs could be 0.7? Do you think it could be 0.5? Justify your answers based on your dot plot.

Likely .7 since it's near the center of the dotplot
Unlikely prop. of all juniors/seniors is 0.5 since there are very few samples with sample prop. of .5 or less.

e. Suppose the eleventh graders had taken random samples of size 60. How would the distribution of sample proportions based on samples of size 60 differ from the distribution for samples of size 30?

Sampling distr. would be mound shaped w/ ≈ same mean as sampling distr. size 30, but SD of sampling distr. based on size 60 would be smaller than one based on samples of size 30.

1. Name your favorite movie of all time.

2. Take a look at the list of the 200 top-grossing movies of 2018 and select 10 that you saw (or wanted to see) in theaters. For purposes of this activity, we will consider these 200 movies as a small population. In practice, populations are often much larger than 200 individuals, sometimes reaching the hundreds of millions or more. Write the titles of those movies in the table below, along with the amount they grossed in 2018. Notice that the listing of movies gives the gross income rounded to the nearest tenth of a million, so that a movie listed as earning $270.3 really grossed $270,300,000. The order in which you write the movies in the table below does not matter.

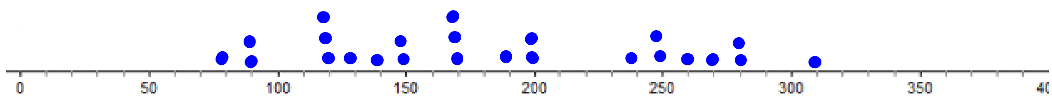
Movie Title	Gross Income (Millions)

3. Compute and record the mean gross income for the 10 movies you selected. This number is called the **sample mean**.

4. Is your sample mean the same as all the other sample means computed by the other students in your class?

It's probably not surprising to you that your sample mean differs from those of other students because you have most likely chosen different samples. The fact that different samples yield different statistics (in this case different sample means) is called **sampling variability**.

5. Combine your results with those of your classmates by creating a dotplot of sample means on the board. Then, record this dotplot on the number line below, carefully labeling the axis.



6. Based on the previous dotplot, without any calculations, what do you suppose the mean gross income for the population of all 200 movies might be? $185,000,000 (center of dist.)

The method you used to take samples from the population is based on your experience and interest in movies. It turns out that this is not a particularly good way to sample if you wish to generate samples that are representative (good images) of the population. Instead of using human experience, judgment, or interest to choose samples, statisticians use chance to select samples from large populations. Samples selected by a chance process are called **random samples**.

7. Your instructor will give you instructions on how to use random chance to select 10 movies from the population of 200 movies (using technology) to generate 10 random numbers from 1-200. Find the ID numbers in the table below that match your randomly chosen numbers. Record the ID number, title of the movie, and gross income in the following table.

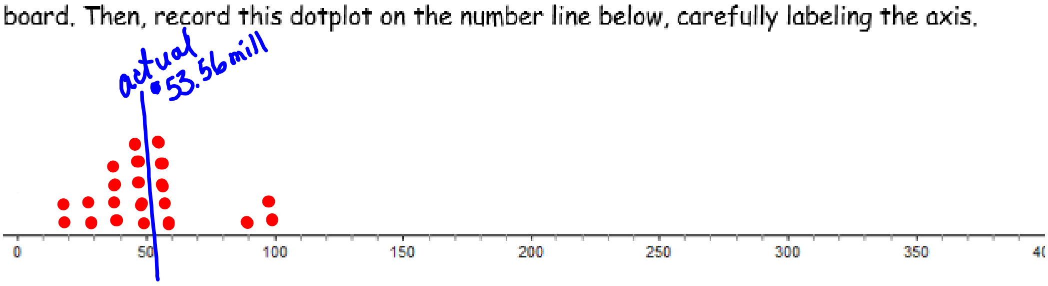
Math-Prob
→5: RandInt
(1,200,10)

ID Number	Movie Title	Gross Income (Millions)

8. Compute and record the mean gross income for the 10 movies you selected. This number is called the **sample mean**, but this time the sample mean has been generated by a **random sample.** Is your sample mean the same as all the other sample means computed by the other students in your class?

   No (but more similar than before)

9. Combine your results with those of your classmates by creating a dotplot of sample means on the board. Then, record this dotplot on the number line below, carefully labeling the axis.



10. Based on the previous dotplot, without any calculations, what do you suppose the mean gross income for the population of all 200 movies might be?

    $ 55,000,000

11. Is your guess for the mean gross income from a random sample somewhat different from your guess when you chose your own sample?

    Yes!

    (185,000,000 vs. 55,000,000)

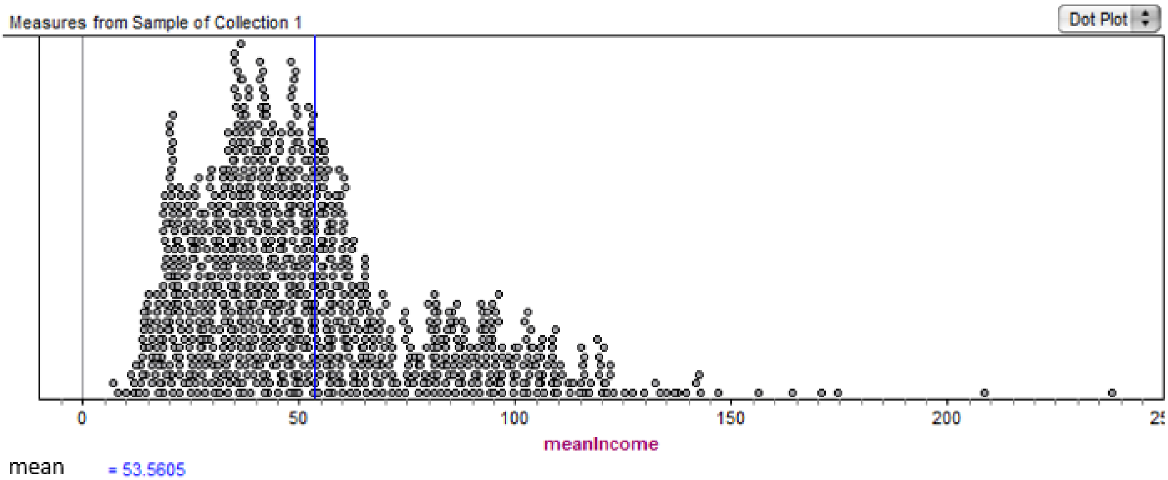
12. The population's mean gross income for the population of all 200 movies is $53.56 million. Go back to your dotplots in #5 and #9 and draw a vertical line on your number line at 53.56. Did the sample means from you and your classmates do a pretty good job of estimating the population mean when you chose your sample by thinking of movies you saw or wished to see? How about when you obtained your samples through chance? What have you learned about the use of random samples from populations?

No - based on personal pref (choice). Chose one's ~~[illegible]~~ that were popular (Higher gross)

yes!

Random are much more accurate at representing the entire population.

13. With enough time, you and your classmates could continue drawing *random samples* of 10 movies, computing the sample mean for each sample, and building the dotplot in #9 above. In order to save time, here is a simulated sampling distribution using 1000 samples of size 10. Notice that the peak is around where we indicated, but the mean of the sample means is right around $53.56 million, as we would expect.



It should be clear that rather than producing haphazard results, random sampling actually follows regular patterns that can be predicted with advanced mathematics and statistics. These patterns can then be used to make inferences (conclusions) about the population from random samples.

Summary:

For a given sample, you can find the sample mean.

- There is __variability__ in the sample mean. The value of the sample mean __varies__ from one random sample to another.
- A graph of the distribution of sample means from many different random samples is a __simulated__ __sampling__ __distribution__.
- Samples means from random samples tend to cluster around the value of the __population__ __mean__. That is, the simulated sampling distribution of the sample mean will be centered close to the value of the __population__ __mean__.
- The __variability__ in the sample mean __decreases__ as the sample size increases.
- Most sample means are within __2__ standard deviations of the mean of the simulated sampling distributions.

~95% of all data is within 2 SD

1. The following segment lengths were selected in four different random samples of size 10.

Stat calc 1Var stats

Lengths Sample A	Lengths Sample B	Lengths Sample C	Lengths Sample D
L_1 1	1	1	2
2	3	5	2
1	1	1	7
5	2	3	2
3	1	4	5
1	5	2	2
2	3	2	3
2	4	4	5
3	3	3	5
1	3	4	4

$\bar{x}$ S_x

a. Find the mean segments length of each sample and the standard deviation of each sample.

A: $\bar{x} = 2.1$ $S_x = 1.29$ B: $\bar{x} = 2.6$ $S_x = 1.35$ C: $\bar{x} = 2.9$ $S_x = 1.37$ D: $\bar{x} = 3.7$ $S_x = 1.77$

b. Find the mean and standard deviation of the four sample means and standard deviations.

Put $\bar{x}$ in L_1 stat-calc-1

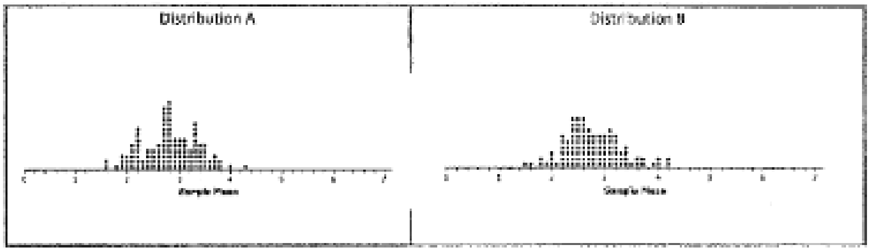
Mean of sample means = 2.825

SD of sample means = .67

c. Interpret your answer to part (b) in terms of the variability in the sampling process.

A typical distance of a sample mean from the mean of the 4 samples (2.825) is .67.

2. Two simulated sampling distributions of the mean segment lengths from random samples of size 10 are displayed below.



a. Compare the distributions with respect to shape, center, and spread.

Both distributions ~ symmetric w/ a center a bit below 3, about 2.8.

The max mean segment length in both is about 4.2 units & min around 1.5 or 1.6. Most of sample means in both distr. are betw. 2 and 4.