Lesson 19: Sampling Variability in the Sample Mean

Classwork

This lesson uses simulation to approximate the sampling distribution of the sample mean for random samples from a population, explores how the simulated sampling distribution provides insight into the anticipated estimation error when using a sample mean to estimate a population mean, and covers how sample size affects the distribution of the sample mean.

Exercises 1–6: SAT scores

1. SAT test scores vary a lot. The table displays the 506 scores for students in one New York school district for a given year.

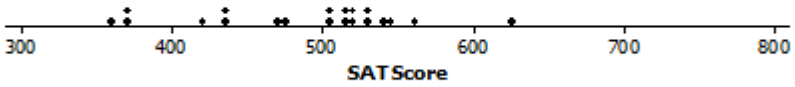
Table 1: SAT scores for district students

441	395	369	350	521	691	648	521	498	413	486
440	415	481	392	800	448	603	503	486	476	500
391	359	447	550	432	158	379	394	495	442	507
395	504	399	424	456	729	356	392	514	388	518
445	436	386	493	467	493	440	387	512	431	467
499	412	457	389	323	319	550	450	517	405	506
486	519	369	373	348	532	496	488	504	444	
396	473	319	367	679	472	613	561	522	408	
451	427	369	560	602	520	567	495	473	424	
362	391	371	407	436	366	582	528	533	463	
328	613	357	438	436	713	603	525	553	446	
414	466	382	362	777	259	557	508	495	466	
409	486	627	589	749	410	639	516	520	632	
526	334	608	374	634	443	556	506	506	526	
391	497	378	358	566	442	496	568	544	546	
529	392	387	373	198	555	499	476	525	529	
529	426	470	378	345	431	613	490	548	455	
574	379	380	561	712	197	556	547	543	431	
363	382	370	379	504	254	596	489	474	386	
486	434	365	530	685	372	580	506	529	434	
418	722	674	504	645	501	605	511	566	362	
527	437	388	525	509	662	445	489	487	426	
441	395	377	561	448	503	602	523	510	404	
467	463	427	519	491	448	638	530	518	493	
387	433	446	525	352	662	570	507	515	515	
503	371	394	569	779	158	558	504	516	407	
350	392	368	484	689	691	535	522	505	409	
583	416	406	416	513	729	623	503	536	422	
370	370	350	446	624	493	465	524	547	612	
499	422	344	420	465	319	460	523	528	486	
399	532	347	446	504	532	375	524	527	394	
374	545	377	462	390	472	540	501	523	424	
372	427	391	528	576	520	564	482	540	393	
559	371	339	533	756	366	547	502	480	420	
330	390	404	543	451	713	568	503	516	415	
567	529	377	460	505	259	588	439	501	394	
371	341	469	391	540	410	502	474	452	473	
503	356	417	623	436	443	510	477	507	531	
327	351	356	587	298	442	589	458	486	469	
528	377	370	528	449	555	537	494	500	453	
447	404	355	356	352	431	410	447	507	442	
572	369	364	523	574	197	330	517	518	509	
379	396	383	404	518	460	500	457	467	435	
456	396	400	505	682	623	531	471	506	427	
406	535	404	512	474	587	509	541	509	489	
420	388	375	514	629	528	571	513	597	480	
395	370	398	516	656	523	527	441	509	516	
355	417	376	498	539	505	457	489	567	501	
423	419	451	460	553	514	552	498	509	452	
438	348	369	541	400	629	561	538	597	507	

a. Looking at the table above, how would you describe the population of SAT scores?

b. Jason used technology to draw a random sample of size 20 from all of the scores and found a sample mean of 487. What does this value represent in terms of the graph below?

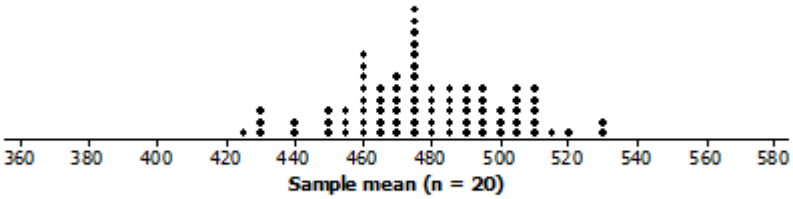
Random sample from District SAT scores



2. If you were to take many different random samples of 20 from this population, describe what you think the sampling distribution of these sample means would look like.

3. Everyone in Jason's class drew several random samples of size 20 and found the mean SAT score. The plot below displays the distribution of the mean SAT scores for their samples.

Random sample from District SAT scores



a. How does the simulated sampling distribution compare to your conjecture in Exercise 2? Explain any differences.

b. Use technology to generate many more samples of size 20, and plot the means of those samples. Describe the shape of the simulated distribution of sample mean SAT scores.

c. How did the simulated distribution using more samples compare to the one you generated in Exercise 3?

d. What are the mean and standard deviation of the simulated distribution of the sample mean SAT scores you found in part (b)? (Use technology and your simulated distribution of the sample means to find the values.)

e. Write a sentence describing the distribution of sample means that uses the mean and standard deviation you calculated in part (d).

4. Reflect on some of the simulated sampling distributions you have considered in previous lessons.

   a. Make a conjecture about how you think the size of the sample might affect the distribution of the sample SAT means.

   b. To test the conjecture, investigate the following sample sizes: $5, 10, 40$, and 50 as well as the simulated distribution of sample means from Exercise 3. Divide the sample sizes among your group members, and use technology to simulate sampling distributions of mean SAT scores for samples of the different sizes. Find the mean and standard deviation of each simulated sampling distribution.

   c. How does the sample size seem to affect the simulated distributions of the sample SAT mean scores? Include the simulated distribution from part (b) of Exercise 3 in your response. Why do you think this is true?

5.

   a. For each of the sample sizes, consider how the standard deviation seems to be related to the range of the sample means in the simulated distributions of the sample SAT means you found in Exercise 4.

   b. How do your answers to part (a) compare to the answers from other groups?

6.

a. Make a graph of the distribution of the population consisting of the SAT scores for all of the students.

b. Find the mean of the distribution of SAT scores. How does it compare to the mean of the sampling distributions you have been simulating?

Lesson 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.
- Sample 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 two standard deviations of the mean of the simulated sampling distribution.

Problem Set

1. Which of the following will have the smallest standard deviation? Explain your reasoning.

   A sampling distribution of sample means for samples of size:

   a. 15 b. 25 c. 100

2. In light of the distributions of sample means you have investigated in the lesson, comment on the statements below for random samples of size 20 chosen from the District SAT scores.
   a. Josh claimed he took a random sample of size 20 and had a sample mean score of 320.
   b. Sarfina stated she took a random sample of size 20 and had a sample mean of 520.
   c. Ana announced that it would be pretty rare for the mean SAT score in a random sample to be more than three standard deviations from the mean SAT score of 475.

3. Refer to your answers for Exercise 4, and then comment on each of the following:
   a. A random sample of size 50 produced a mean SAT score of 400.
   b. A random sample of size 10 produced a mean SAT score of 400.
   c. For what sample sizes was a sample mean SAT score of 420 plausible? Explain your thinking.

4. Explain the difference between the sample mean and the mean of the sampling distribution.