S.1

Teaching Introductory Statistics Courses With a Focus on Developing Statistical Reasoning and Thinking

Beyond the Formula Conference
Rochester, NY
August 2002

Joan Garfield
Dept. of Educational Psychology
University of Minnesota
jbg@umn.edu

Activities to develop statistical reasoning

Reasoning about data and distributions:

- **Matching graphs to descriptions of graphs**
- **Histogram sorting**
- **Creating graphs from descriptions**
- **Matching histograms to summary statistics**
- **Matching histograms to boxplots**

Reasoning about variability

- **Paired graph activity: which has the larger standard deviation?**

Reasoning about covariation

- **Matching scatterplots to correlations and/or descriptions**
- **Guessing correlations**

Reasoning about sampling:

- **Random rectangles**

Reasoning about sampling distributions:

- **Sampling SIM activity**

References

Tools for Teaching and Assessing Statistical Reasoning Website

http://www.gen.umn.edu/faculty_staff/delmas/stat_tools/

STAR Library of Activities

http://starlibrary.net/

***Workshop Statistics*, by Allan Rossman and Beth Chance**
Springer Publishers

***Activity-Based Statistics*, by Richard Scheaffer et al., Springer Publishers.**

Journal of Statistics Education

http://amstat.org/publications/jse/

***Teaching Statistics: Resources for Undergraduate Instructors.* Edited by Tom Moore.**
MAA Notes @52. 2000.

Stay tuned for:

ASA Consortium for the Advancement of Undergradaute Statistics Education

Web Assessment Resource Project
(NSF funded project, Joan Garfield, Bob delMas, and Beth Chance)

***The Challenge of Developing Statistical Reasoning, Literacy, and Thinking.* Edited by**
D. Ben-Zvi and J. Garfield. Kluwer Publishers, 2003.

Innovation in Teaching Statistics
Edited by J. Garfield
MAA Notes series

26

Session S.3.3

Assessing Student Learning Outcomes

Beyond the Formula Conference

Rochester, NY

August, 2002

Joan B. Garfield

University of Minnesota

Jbg@tc.umn.edu

Assessing Statistical Literacy

From "Assessing Statistical Thinking Using the Media" by Jane M. Watson, in Gal and Garfield (1997) *The Assessment Challenge in Statistics Education,* IOS Press.

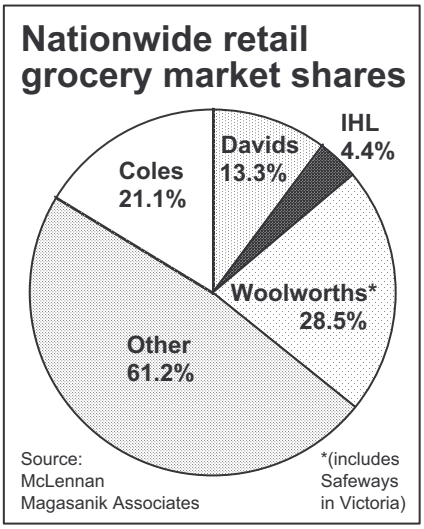
Tier 1: Basic understanding of terminology and representations

Tier 2: Embedding of language and concepts in a wider context

Tier 3: Questioning of claims

Items:

1.



Write an interpretation of this pie graph. Is there anything confusing about this graph?

From the Chance Project

Newspaper critique

Read the attached article and answer the following questions:

1. What do you think is the purpose of the research study described in this article?

2. What method or methods were used to answer the research question?

3. What questions would you like to ask the investigators in order to better understand the study?

4. Are there any aspects of the study that might make you question the conclusions presented in the article?

Assessing Statistical Reasoning

From: Garfield, J. & Gal, I. (1999). Teaching and Assessing Statistical Reasoning. In L. Stiff (ed.) *Developing Mathematical Reasoning in Grades K-12: National Council Teachers of Mathematics 1999 Yearbook*, 207-219.

1. Reasoning about data
2. Reasoning about representations of data
3. Reasoning about statistical measures
4. Reasoning about uncertainty
5. Reasoning about samples
6. Reasoning about association
7. Reasoning about inference

Items:

1. The following message is printed on a bottle of prescription medication:

> **WARNING**: For applications to skin areas there is a 15% chance of developing a rash. If a rash develops, consult your physician.

Which of the following is the best interpretation of this warning?

_____ a. Don't use the medication on your skin — there's a good chance of developing a rash.
_____ b. For application to the skin, apply only 15% of the recommended dose.
_____ c. If a rash develops, it will probably involve only 15% of the skin.
_____ d. About 15 of 100 people who use this medication develop a rash.
_____ e. There is hardly a chance of getting a rash using this medication.

2. Half of all newborns are girls and half are boys. Hospital A records an average of 50 births a day. Hospital B records an average of 10 births a day. On a particular day, which hospital is more likely to record 80% or more female births?

_____ a. Hospital A (with 50 births a day)

_____ b. Hospital B (with 10 births a day)

_____ c. The two hospitals are equally likely to record such an event.

Assessing Statistical Thinking

From "Components of Statistical Thinking and Implications for Instruction and Assessment" by Beth Chance, Paper presented at AERA, 2000.

1. Start from the beginning
2. Understand the statistical process as a whole
3. Always be skeptical
4. Think about the variables involved.
5. Always relate the data to the context
6. Understand (and believe) the relevance of statistics
7. Think beyond the textbook

Items:

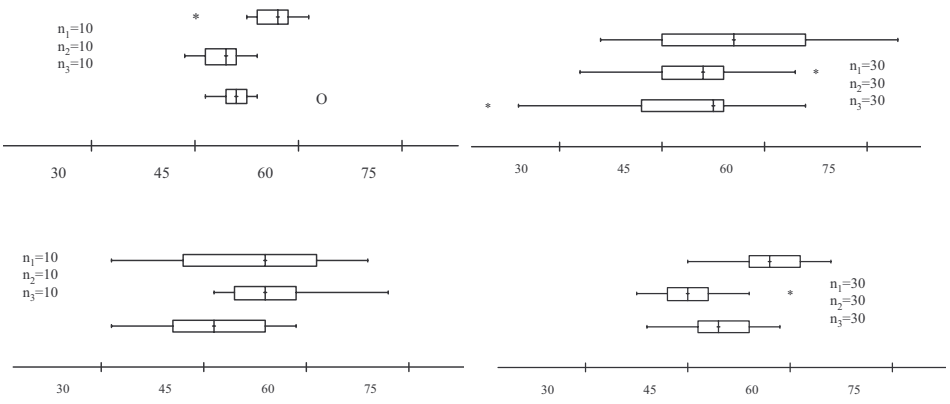
1. The underlying principle of all statistical inference is that one uses sample statistics to learn something (i.e. to infer something) about the population parameters. Convince me that you understand this statement by writing a short paragraph describing a situation in which you might use a sample statistic to infer something about a population parameter. Clearly identify the sample, population, statistic, and parameter in your example. Be as specific as possible, and do not use any example which we have discussed in class

This problem requires students to demonstrate their understanding of the overall statistical process, at least from the point of data collection forward. Students are required to extract a general approach from the isolated methods learned in the course. The focus is on the big picture rather than a specific technique. They also have to demonstrate their ability to apply their statistical knowledge to answer a question of interest (an individual assessment to complement the group project).

2. Given data on calories for several Chinese foods, students are asked to produce a histogram (using technology) and then (b) Do you think it is reasonable to use these data to rank the foods from least to most in terms of calorie content? Explain how else you might look at the data if you were interested counting calories.

In question (b), I'm hoping students will consider the issue of serving size. This serves as a follow-up question to the small soda costs at baseball games examined in class. This approach should be aided by their graph in which egg rolls and soup, the two appetizers, stand out as low outliers. Thus, students are expected to think beyond the statistical method, utilizing context and behavior of the data in their answer.

3. A researcher is examining the time for 3 different medicines to register in the blood system (minutes). She wants to test the null hypothesis that the mean times are all the same: H_0: $\mu_1=\mu_2=\mu_3$. For the following four sets of boxplots, order them by smallest p-value to largest p-value and explain your choices. Your grade will be based mostly on your explanation (inspired by Cobb).



This problem does not focus on application of a particular technique but rather asks students to consider issues of sample size and variation in determining statistical significance. Also notice the emphasis on communication for full credit. Thus, students need to understand the purpose and be able to explain the results of the statistical methods. This is similar to the "explain this result to someone who has not taken statistics" question that can be added to the end of a statistical analysis question.

Putting it all together: An Assessment Framework:

- A listing of desired learning outcomes for students, misconceptions that students should not develop, and prerequisite concepts and skills.

- A pretest which measures prerequisite knowledge and intuitions that may affect students' interactions with learning the new material. The results of the pretest are used as the basis for discussions with students to try to clear up misconceptions that might interfere with the learning activity.

- Assessment items that are imbedded in learning activities/discussions, to help reveal how students are understanding and using the new material.

- A posttest of desired outcomes that assess aspects of statistical literacy, reasoning and thinking.

- A delayed post test which consists of items that could be included in an end of course final exam to measure long-term retention.

An Example: Sampling Distributions of Means

Desired student outcomes

What students should understand about sampling distributions:

- That a sampling distribution for means (based on quantitative data) is a distribution of sample means (statistics) of a given sample size, randomly sampled from a population with mean μ and standard deviation σ. It is a probability distribution for the sample mean.
- The sampling distribution for means has the same mean as the population (parameter)
- As n gets larger, variability of the sample means gets smaller (a statement, a visual recognition, and predicting what will happen or how the next picture will differ)
- The standard error of the mean is a measure of variability of sample statistic values
- How to interpret area/apply areas under curve as probability statements about sample statistics
- The building block of a sampling distribution is a sample statistic
- Some values of statistics are more or less likely than others to be drawn from a particular population
- When it is reasonable to use a normal approximation
- Different sample sizes lead to different probabilities for the same value (know how sample size affects the probability of different outcomes for a statistic)
- Sampling distributions tend to look more normal than like the population, even for small samples
- As sample sizes get very large, all sampling distributions for means look alike (i.e., have the same shape), regardless of the population from which they are drawn
- Averages are more normal and less variable than individual observations
- Distinguish between a distribution of observations in one sample and a distribution of statistics (sample means) from many samples (n greater than 1) that have been randomly selected.

What students should be able to DO with this knowledge:

- Describe what a sampling distribution would look like for different populations and sample sizes (in terms of shape, center and spread, and where the majority of values would be found). What values of the sample mean are likely, and which are less likely.
- Describe the size of the standard error of the mean
- Describe the likelihood of different values of the sample mean
- Describe the mean of the sample means for different shaped populations

***Some common misconceptions students should NOT have*:**

- Sampling distribution should look like the population (for n > 1)
- Sampling distributions for small and large sample sizes have the same variability
- Sampling distributions for large samples have more variability
- Don't understand that a sampling distribution is a distribution of sample statistics
- Confuse one sample (real data) with all possible samples (in distribution) or potential samples
- representing a population (better for large samples) is confused with a sampling distribution for large samples, better representing a population (have more area)
- students pay attention to the wrong things: e.g., heights of bars
- the mean of a positive skewed distribution will be greater than the mean of the sampling distribution for samples taken from this population

Prerequisite knowledge for learning sampling distributions

Students should understand:

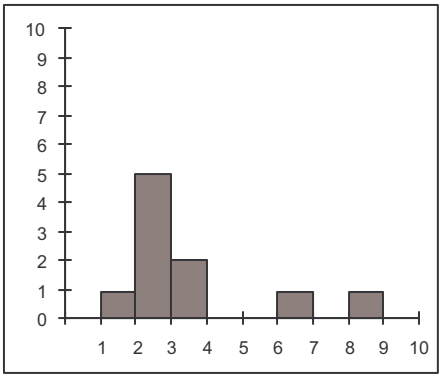
- the idea of a distribution
- what variability means (spread vs. smoothness)
- the idea of the center of a distribution
- a random sample
- a sample statistic vs. a population parameter
- common shapes of distributions: normal, skewed, uniform, bimodal
- being able to see between the data, to recognize overall shapes of distributions
- the idea of area under a curve and how it represents likelihood of outcomes
- properties of the normal distribution
- normal distributions can look different due to different variability (shape vs. variance)
- how to read and interpret histograms

the idea of sampling variability

Sampling Distributions Pretest Items
(L = literacy, R = reasoning, T = thinking)

1. What does the word “distribution” mean to you? **L**

2. Listed below is histogram for a set of test scores.



a. What do the numbers on the horizontal axis represent? **L**

b. What do the numbers on the vertical axis represent? **L**

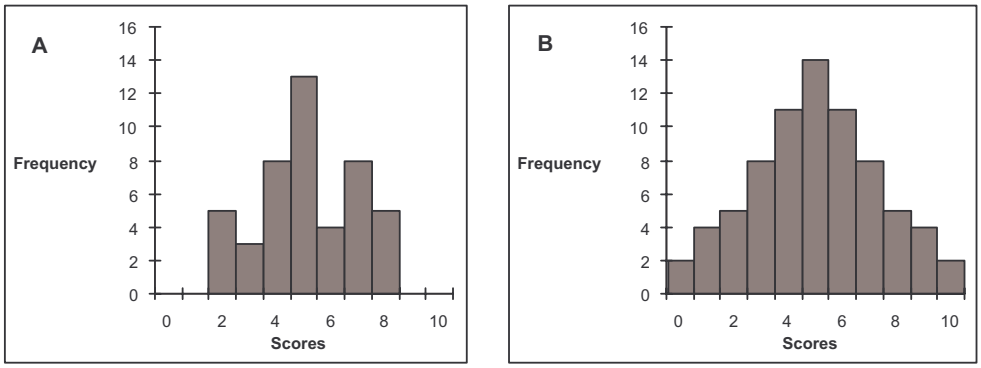
c. How many people received scores of 4 or higher? **L**

d. How many people took the test and have scores represented in the graph? **L**

e. What percent of students received scores of 4 or higher? **L**

2. Which of the following distributions shows MORE variability? Check one of the choices: **L**

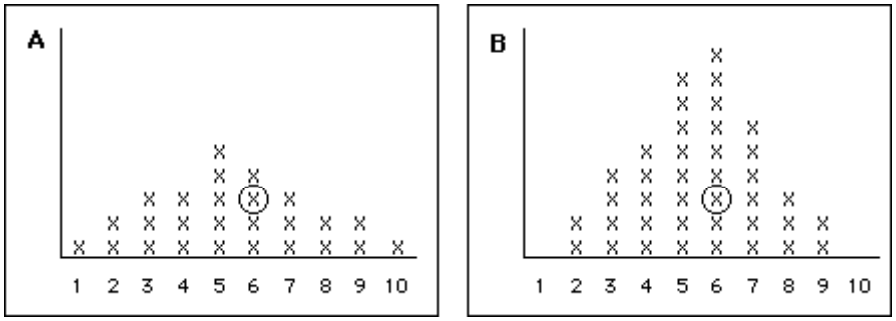
A has more variability _____ B has more variability_____



Check the statement or statements that led you to select your answer above.

___a. Because it's bumpier
___b. Because it's more spread out
___c. Because it has a larger number of different scores
___d. Because the values differ more from the center
___e. Because it has more data values (scores)
___f. Because it has a larger standard deviation
___g. Other (please explain)

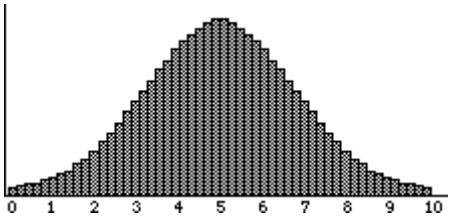
3. Figure A represents a sample of 26 weights and Figure B represents a sampling distribution of mean weights for samples of size 3. One value is circled in each distribution.



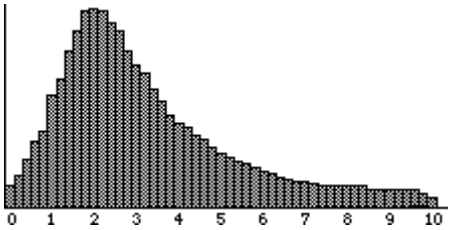
Is there a difference between what is REPRESENTED by the X circled in A and the X circled in B? **R**

a. No
b. Yes (please explain what you see as the difference)

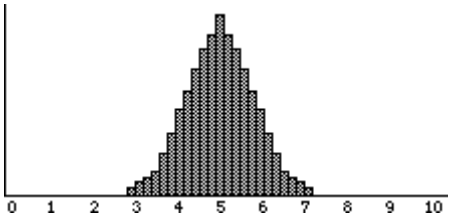
4. Shown below are 4 distributions of test scores. For each one, circle the one descriptor that best represents the shape of the distribution. **L**



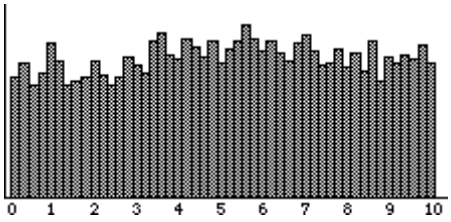
a. normal skewed bimodal uniform even other can't tell



b. a. normal skewed bimodal uniform even other can't tell

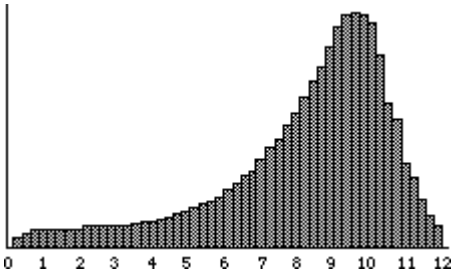


c. normal skewed bimodal uniform even other can't tell



d. normal skewed bimodal uniform even other can't tell

The distribution for a population of measurements is presented below. Suppose that ten values are going to be sampled from this population and the sample mean calculated. Some possible values for this sample mean are 1, 6, 8, and 10.



5. Looking at the graph above, what would you guess to be the value of μ? **R**

6. Which of the four possible sample mean values is MOST likely to be calculated? (circle only one) **R**

   a. 1

   b. 6

   c. 8

   d. 10

7. Which of the four possible sample mean values is LEAST likely to be calculated? (circle only one) **R**

   a. 1

   b. 6

   c. 8

   d. 10

Some Sampling Distributions Posttest Items

1. American males must register at a local post office when they turn 18. In addition to other information, the height of each male is obtained. The national average height for 18-year-old males is 69 inches (5 ft. 9 in.). At the small, local post office, about 5 men register each day. At the large, city post office, about 50 men register each day. At the end of each day, the clerk at each post office computes the average height of the men who registered there that day

   Which of the following predictions would you make regarding the number of days on which the average height for the day was more than 71 inches (5 Ft. 11 in)? **T**

   __a. The number of days with average heights over 71 inches would be greater for the small post office than for the large post office.

   __b.The number of days with average heights over 71 inches would be greater for the large post office than for the small post office.

   __c. There is no basis for predicting which post office would have the greater number of days.

2. In a geology course, students were asked to determine the weight of rock samples. One instructor asked her students to weigh a rock several times on the same scale. This rock is known to weigh exactly 1000 grams. However, the scale is not completely accurate and sometimes it is off in either direction by 25 grams or less. After a lot of practice, one student weighed the rock 20 times, then computed and recorded the average of the 20 weighings. After a lot of practice, a second student weighed the rock 5 times, then computed and recorded the average of the five weighings. **T**

   How would you expect the accuracy of the average weight recorded by the first and second student to compare?

   __a. The student who weighed the rock 20 times would be more likely to have a more accurate average.

   __b. The student who weighed the rock 5 times would be more likely to have a more accurate average.

   __c. Both averages would be equally accurate.

   __d. It is impossible to predict which average would be more accurate.

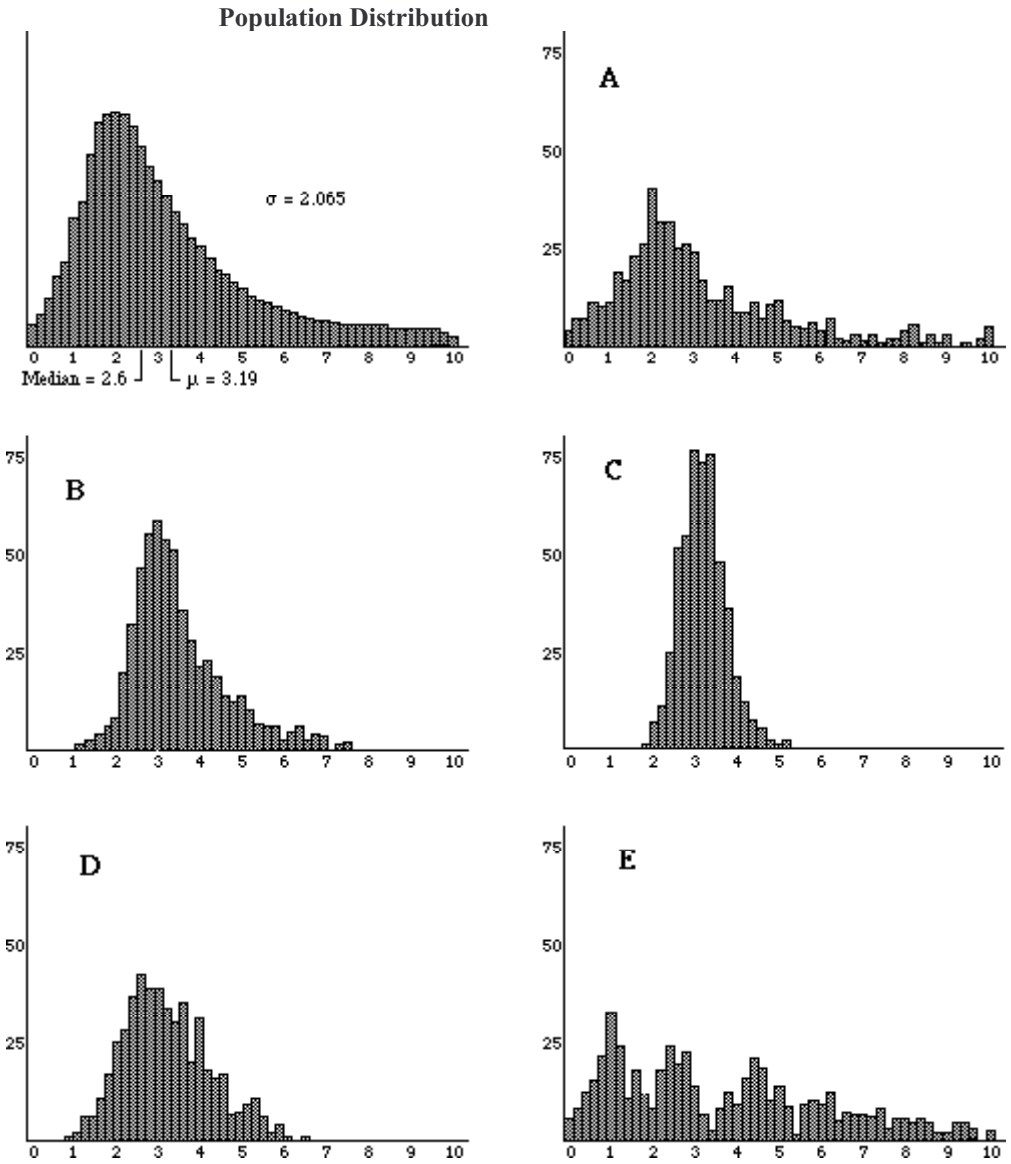
3. A small town has 15,000 families. The average number of children per family is 2.5, with a standard deviation of .60. children. The distribution is not normal since 30% of the families have no children at all. A researcher wants to draw a sample of 100 families and record the number of children in each family. Her inferences would be based on the sampling distribution of sample means for all samples of size 100. Sketch what you think this distribution would look like. Label both axes and indicate the center and spread of the distribution**. T**

On the pages that follow you will find 5 different situations. For each situation, a population distribution is presented in the upper left of the page. For each population you are asked to identify the distribution of sample means that might result when 500 samples are drawn at random for each of two different sample sizes. Each situation presents five graphs labeled A through E. The two problems for each situation are presented on the page that is opposite of the page with the graphs. To answer each problem, circle the letter that corresponds to the graph you believe best represents the distribution of sample means that might result for the given sample size.

Pay careful attention to the sample sizes because they change from question to question.

Then, read through and answer the questions that follow each problem by selecting one, and only one, answer for each question.

6. The distribution for a population of test scores is displayed below on the left. Each of the other five graphs labeled A to E represent possible distributions of sample means for random samples drawn from the population**. R**



Which graph do you think represents a distribution of sample means for 500 samples of size 4? (circle one)

A B C D E

Answer each of the following questions regarding the sampling distribution you chose for question 1.

a) Which do you expect for the shape of the sampling distribution? (check only one)

- O Shaped more like a NORMAL DISTRIBUTION.
- O Shaped more like the POPULATION.

Circle the word between the two vertical lines that comes closest to completing the following sentence:

b) I expect the sampling distribution to have | less / the same / more | VARIABILITY than the population

Which graph do you think represents a distribution of sample means for 500 samples of size 16? (circle one)

A B C D E

Answer each of the following questions regarding the sampling distribution you chose for question 2.

a) Which do you expect for the shape of the sampling distribution? (check only one)

- O Shaped more like a NORMAL DISTRIBUTION.
- O Shaped more like the POPULATION.

Circle the word between the two vertical lines that comes closest to completing each of the following sentences.

b) I expect the sampling distribution to have | less / the same / more | VARIABILITY than the population

c) I expect the sampling distribution I chose for question 2 to have | less / the same / more | VARIABILITY than / as the distribution I chose above for question 1.

Session 10 is not available