CCMR Educational Programs

Title:	**Should We Count the Beans…one at a time?**
Date Created:	August 6, 2006
Author(s):	Joan Erickson
Appropriate Level:	Introductory statistics or AP Statistics. Pre or co-requisite with general chemistry
Abstract:	We use hands-on activities to demonstrate that sampling with the appropriate sample size can be a very effective way to make inferences of a population when it is impossible to measure the entire the population. We use real-life questions to invoke the students' thoughts on how and why we use statistics. We also learn to apply statistical analyses using the correct vocabulary such as statistics and sampling distribution, parameters in population, spread, sampling deviation, etc.
Time Requirement:	2 lecture periods, depending on your intended course coverage.
NY Standards Met:	- Standard 2 – Information Systems: Students will access, generate, process, and transfer information using appropriate technologies. - Standard 3 – Mathematics: Students will understand mathematics and become mathematically confident by communicating and reasoning mathematically, by applying mathematics in real-world settings, and by solving problems through the integrated study of number systems, geometry, algebra, data analysis, probability, and trigonometry. - Standard 4 – Science: Students will understand and apply scientific concepts, principles, and theories pertaining to the physical setting and living environment - Standard 7 – Interdisciplinary Problem Solving: Students will apply the knowledge and thinking skills of mathematics, science, and technology to address real-life problems and make informed decisions.
Equipment/Materials List	- A see-through jar, colored beads (any color combination as long as you know the number of beads for each color) A paddle with 40 holes. The jar & paddle demo is available through many commercial teaching resources. Similar jar & paddle lesson plans are readily available on the internet. - Graph Paper or poster board for frequency distribution chart
Resources/Credits Web links	- Cartoon clips http://www.fotosearch.com/illustration/bean-counting.html - Similar lesson concept http://www.amstat.org/publications/jse/v5n1/schwarz.supp/samplingbowl.html - Resources on New York Learning Standards http://www.emsc.nysed.gov/ciai/mst/math.html

	• National Council of Teachers of Mathematics Learning Standards (NCTM) http://standards.nctm.org/document/chapter7/data.htm • MALDI-TOF mass spectrometer http://www.pslc.ws/mactest/maldi.htm

Statistics Lesson Plan
Random Sampling in Action

SHOULD WE COUNT
THE BEANS...one at a time?

Lesson Plan:

Notes to Teachers:

- This lesson plan is designed to provide ideas for hands-on demonstrations. The lecture outline is laid out for the teacher. It is your choice to interject the lecture with certain activities (See Demos). I use an example from another discipline (chemistry) to probe how well the students understand the difference between a population and a sampling process, how to interpret the various shapes of the density curves, and the very fundamental concept of CLT (central Limit Theorem).
- A printable activity worksheet is included at the end of the lesson plan.
- Keep in mind that it takes a while to build up the frequency table when demonstrate CLT. I do random sampling and sample mean distribution for CLT in lab, where the students can do random sampling on Minitab.
- I usually spend another 3-4 lecture periods on CLT after this lesson and get heavily into the $\frac{\sigma}{\sqrt{n}}$ calculations.

Recapping Previous Knowledge:

- By now the students have learned words such as population, mean (μ), variance, and sigma (σ). They also know the meaning of observation, frequency, spread, center, etc.

New concepts and vocabulary Needed for This Lesson:

- **Randomness** and the need to watch out for **biases**.
- Sampling processes and **sample size**.
- How to relate sampling outcome to the population.
- **Sample mean distribution** and **Central Limit Theorem**. (If time allows)

Lecture Outline:

- What happens when we can't capture EVERY object in a population?
- Introduce the concept of random sampling and necessary terminology.
- If we don't know much about the population, what can we learn by taking random samples? (Making inferences based on sound information)
- Introduce Central Limit Theorem. I talk about the conditions necessary for CLT to work even though the population distribution may not be Normal.
- Explain that CLT is all about sampling the population and how we can use CLT to help us make estimations about an entire population. Examples may vary (see Demo ideas).

Demo #1(Hook):

- Coin-toss. Let heads $= 0$ and tails $= 1$. If the coin is fair, ask the students to toss the coin 20 times and get an average of the 20 outcomes. What they think the most probable outcome should be? (They should know it is about 0.5)
- Verify it by building a frequency table. The peak should be near 0.5.
- (TI-83/84 has a coin toss simulator under APPS. But the degree of fairness is predetermined by the user.)

## Demo #2 (Hands-on, see Activity Worksheet):

- Rolling a die. If the die is fair, what is the most probable average outcome? (I use a weighted die so the students can't just guess the average is 3.5)
- Go to the activity worksheet.

## Demo #3 (Hands-on, see Activity worksheet):

- Show the jar of beads in class. Don't tell the students that the orange beads make up X % of the content. DO tell the students that there are Y beads total in this jar.
- Can we find out how many orange beads there are <u>without</u> having to count the beads one-by-one?
- Find the most probable average % in the sampling distribution. Call it X %. $X \cdot Y = \#$ of orange beads.

Question (Probe):

- We know that in chemistry, a polymer is a long chain of a certain molecule. For example, polystyrene is a polymer that is made of many styrene molecules linked together. Think of the polystyrene molecules in terms of $n \cdot (C_8 H_8)$ where n $=$ integer. Depending on the value of n, the polystyrene molecules can have various weights. A mass spectrometer can show us the molecular weight distribution of polystyrene molecules.
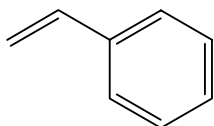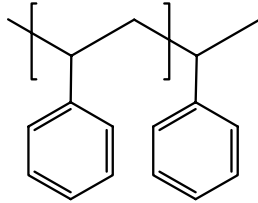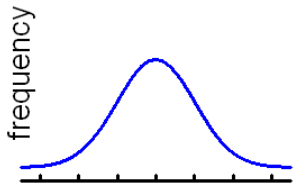

Single Styrene n $= 1$
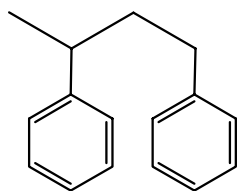
Styrene dimer n $= 2$
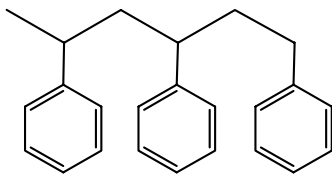
Polystyrene n= 3
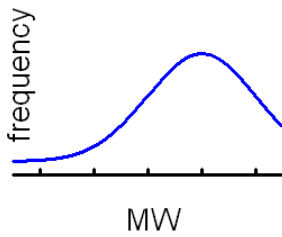
Polystyrene $(C_8H_8)n$

- Think of the population of ALL the polystyrenes where $0 \leq n < \infty$. Is there a most probable value of n? In other words, we know polystyrene likes to grow in length, but is there a finite length for the polymer chain? What do you think the density curve of polystyrenes looks like?
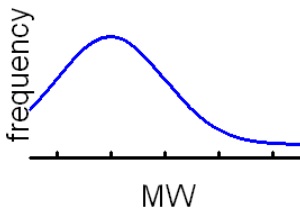
frequency

Is it a symmetric curve? The peak of the curve indicates the most probable weight of the polystyrene molecule?

Or

frequency

MW
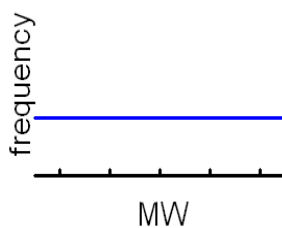
Is it a skew- curve where the most probable value of n is on the high side but there are a few short chains (low molecular weights)?

Or

frequency

MW

Is it a skew+ curve where the value of n is on the low side with a few long chains?

Or

frequency

MW

The polystyrene molecules of various weights are equally likely to be present?

- MALDI-TOF mass spectrometer uses a very simple physics principle F=ma to determine the masses of the molecules. Namely, the lighter the molecule, the "faster" it flies across the fixed distance. MALDI gives a frequency distribution of the various weights of the polymer.
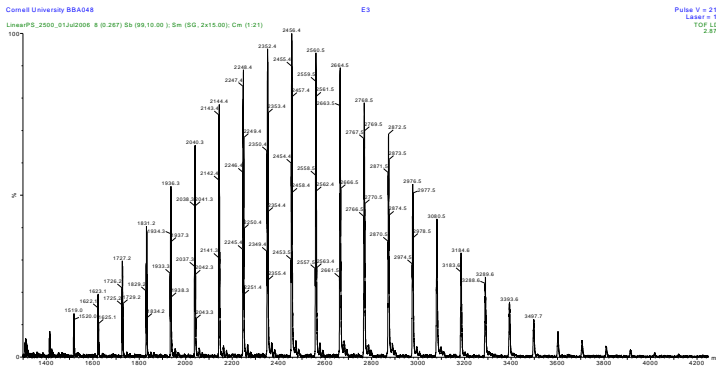
Although the actual MALDI spectrometer plot is Normalized by energy focusing, it is a good example of a sampling distribution when the population is large.

Conclusion:

- Wrap up the activity. Discuss the noticeable trends in the sampling distribution and draw inferences about the population.
- Discuss the possible biases in the activity. Discuss the concept of "randomness".
- Where else can we apply CLT to study quantitative data in real life? Ask the students to give examples and specific ways of measuring the statistics.

Follow-up:

- Assign homework problems. Assign real-life quantitative survey project if there is enough time in the semester

Random Sampling in Action

SHOULD WE COUNT
THE BEANS?

Is the Die Fair? Activity---How to Build a Frequency Table

Part One: Concept Check

a.) You have rolled a die before. You know that there are _____ possible outcomes each time you roll a die.

b.) If the die is fair, the probability of getting a "4" is the same as getting any other number, which is _____ % chance.

c.) If you rolled the die twice and got a "2" and "5", the mean (average) of the two outcomes would be _____.

d.) Assume the die is fair, if you roll it six times and get a different number each time, the mean of these outcomes will be _____

e.) What do you suppose the frequency table of the outcome mean values will look like? Draw a rough sketch.
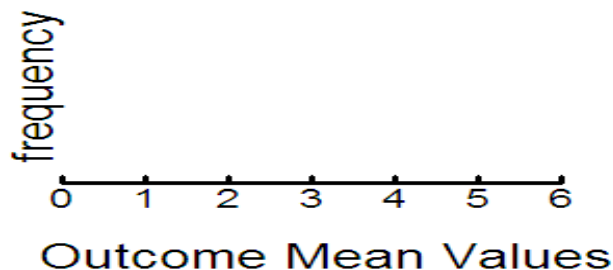
Part Two: Let's Roll It!

    a.) Roll the die 10 times, record the number you get each time, then calculate the mean.
    b.) Repeat the process 3 more times.
    c.) Now go up to the chart on the wall, mark an X for each of the four average outcome values.
    d.) Does the chart on the wall look like the one you predicted? How or how not so?

_____

Beads & Paddle Activity---Random Sampling in Action

Part One: Vocabulary/Concept Review

We are trying to find how many percent of the beads in the jar are orange. Once we have an idea how what percent it is for the orange beads, with the fact there is a total number of _____ beads in the jar (ask me), we can estimate how many orange beads there are.
Before you start sampling the beads, think about how the statistical terms are applied in the beads & paddle activity.

    a.) In the case of the beads & paddle activity, what is the population? _____

    b.) The population size = _____

    c.) What is a "sample" in this activity? _____

    d.) The sample size n =_____

    e.) Briefly explain the sampling process in this activity

        _____

        _____

    f.) We conduct the sampling process because we want to know

        _____

    g.) Each time you sample the jar, what is the lowest possible number of the orange beads that can be included in the sample? _____, which is equivalent to _____ %

    h.) Similarly, what is the highest possible number of the orange beads you can bring up in each sample? _____, which is equivalent to _____ %

    i.) What you think the sampling distribution of the beads & paddle activity should look like? Draw a rough sketch.
    j.) Use any method to make a guess what percent of the orange beads there are in the jar. _____%. Compare your guessed value with the value obtained in Part Three a).

Part Two: Building the Sample Mean Distribution

a.) Each person will "dip" into the jar 3 times. Record the number of the orange beads you see in your sample. Return the beads back in the jar after sampling each time.

	Dip #1	Dip #2	Dip #3	Mean	%
# of orange beads					
# of orange beads					
# of orange beads					
# of orange beads					
# of orange beads					

b.) The mean (average # of orange beads) of the 10 samples is _____

c.) Convert it to % by dividing the mean by 40, then multiply by 100 = _____ %

d.) Repeat steps a). through c). 4 more times. Write down the 4 average percents.

Now go up to the chart on the wall, mark an X for each of the five average % values.

If you have more time, sample the beads as many times as you wish, mark more X's on our chart. Remember, the more samples averages are taken, the better!

Part Three: Estimate the number of white beads in the jar.

a.) Based on the frequency distribution chart, the peak is at _____ %

b.) Therefore, we estimate the total number of orange beads is about _____

c.) So, to answer the question asked in the title of this worksheet "Should we count the beans...one at a time?" Let's think about what possible factors may prevent us from counting every object in the population? How does taking samples help us understand the whole population? Justify your reasons.

_____

_____