# Sampling From Population With Unknown & Known Parameters

**Description:** Each bag comprises a population of values. Each bag contains 10 pieces of data (i.e., wood chips) from a population with unknown parameters. Do not mix chips from different bags.

**For Instructor:** There are six bags. There are two bags with a mean of 20, two with mean of 25 and two bags with a mean of 30. Each bag comprises a normally distributed data set with a standard deviation of 5.

**Setup Time:** 1 to 5 minutes.

**Learning Goals:** Improved understanding of and ability to explain the process of sampling, creation of sampling distribution of means, central limit theorem, sampling error and comparison of populations being estimated by sample statistics. Improved understanding of confidence intervals—how to calculate and interpret.

----------------------------------------------------

## Suggested Activities

**ACTIVITY 1: Sampling Distribution Creation and Discussion, Central Limit Theorem (10 – 15 minutes)**

1. Put students in groups of two or three. Give a sampling bag to each group.
2. Have the students draw a horizontal line across their note page.
3. Give a bag to each group. Tell the students that they are going to randomly select 10 samples from their bag. **NOTE:** Groups can all perform sampling using same *n* or different *n* can be assigned to different groups to allow for expanded discussion of sampling error.
4. After drawing their first sample, have all groups calculate their mean. Round to one decimal place. Have students write their first mean value at the midpoint of their horizontal line. **NOTE:** If you’ve covered it already, this would be a good place for students to calculate the est. standard error of the mean.
5. On the horizontal line, have students add 10 “tick”marks (equally spaced) to the left of the midpoint and 10 “tick” marks to the right. The distance between tick marks will be a single unit on the measurement scale.
6. After each sample draw, students must put the data chips back in the bag, shake the bag and then redraw a new sample until all 10 samples have been drawn and means calculated for each of the 10 samples. Students should mark the location (with an “*M*”) of each sample mean, on the horizontal line.

DISCUSSION: After all groups have finished, have them discuss their findings as a class or within groups. The following might be good prompts for encouraging discussion.

* What do we call this chart that we’ve created? Why is it called that? How is it used?
* What is your best guess for your population mean? Can you even know if your guess is close? How?
* What if you sampled using a smaller sample size? What about a larger sample size? How would that change the way your chart looks?
* If you calculated standard error of the mean, does that seem to agree with your completed chart?
* How does the central limit theorem figure into this activity?

**ACTIVITY 2: Comparing Known Population Mean With Sample Mean (15 - 20 minutes)**

This could be an effective precursor for one-sample *t* testing.

1. Put students in groups of two or three. Give a sampling bag to each group.
2. Have the students draw a horizontal line across their note page.
3. At the midpoint of this line, have students write the **population mean value of 25**.
4. On the horizontal line, have students add 15 “tick”marks (equally spaced) to the left of the midpoint and 15 “tick” marks to the right. The distance between tick marks will be a single unit on the measurement scale, which means our number line has a range of 10 to 40.
5. Give a bag to each group. Tell the students that they are going to randomly select 10 samples (with replacement) from their bag. Have all groups use the same sample size of n = 3.
6. Have the students draw their first sample, calculate the mean (round to one decimal place) and then mark the location (with an “*M*”) on the horizontal line.

DISCUSSION: Ask the students what they think about the difference between the population mean and their first sample mean? Can they draw any conclusions at this point? Are they really different? What is it that they’re really comparing? Means? Estimates of means? Distributions?

1. Students should calculate the **estimated standard error of the mean**, based on the mean, sample size and standard deviation of their first sample. If desired, have students label the horizontal line with standard error units.
2. Continue sampling. Remind the students that after each sample draw, they must put the data chips back in the bag, shake the bag and then redraw a new sample--until all 10 samples have been drawn, means calculated for each of the 10 samples and the location of each mean placed on the chart.
3. Have students draw two normal curves—one curve centering on the population mean of 25 and the other curve centering on the mean of the sampling distribution of *M’*s.

DISCUSSION: After all groups have finished, have them discuss their findings as a class or within groups. The following might be good prompts for encouraging discussion.

* Now look at your chart again. How might your initial conclusions, based only on locations of population and sample means, change?
* What is it that we’re really comparing when we are analyzing the mean differences for statistical significance?
* What does the height of the curves represent?
* Which of these curves represents the null distribution and which represents the alternative distribution? What do “null” and “alternative” mean within this context?
* If we conclude there is or is not a big difference between our means, is there a chance that our conclusion will be in error? Based on your chart, would that be a large or small chance? Why? How would I even know if my conclusion was in error?
* How was our sampling activity like data collection and analysis in a real research study? How was our sampling activity different?

**Activity 3: Confidence Interval Of The Mean (15 – 20 minutes)**

Use as a supplement to sampling-distribution activities and for discussing unknown population parameters.

1. Put students into groups of two or three and give a sampling bag to each group.
2. Have students take their first sample (with replacement) from the bag. Use a sample size of n = 5.
3. Have students calculate the mean and standard deviation from their first sample.
4. Have students calculate a 95% confidence interval of the mean for their first sample.
5. Have students write down their first confidence interval.

Discussion: Ask the students about the difference between samples and populations. Discuss what a confidence interval actually tells us. How does sample size affect our estimates. Ask the students how confidence intervals are used to supplement significance testing.

1. Have students put the data chips back in the bag. Shake the bag and randomly select their next sample, calculate the mean, standard deviation and confidence interval. Ideally, this process would be repeated another nine times (so, 10 total sampling rounds).
2. Students need to write down all of their confidence intervals. Best to just put them in a list.
3. Now, students will calculate the population mean. After completing all of their sampling rounds and getting 10 confidence intervals of the mean, have the students pull out all 10 data chips from the bag and calculate the mean of the population.
4. Next, have students check to see how many of their confidence intervals contain the population mean. We would anticipate that 95% of these confidence intervals would contain the true population mean. Do they?

Discussion: Ask students how a confidence interval of the mean can be used to supplement significance testing.

**ACTIVITY 4: Creating A Sampling Distribution Of Mean Differences (15 - 20 minutes)**

This could be an effective precursor for independent-groups *t* testing. For this one, pretend like you don’t know the population means and standard deviations.

1. Put students in groups of two or three. Give a bag to each group of students.
2. Have the students draw a horizontal line across their note page.
3. At the midpoint of this line, have students write the **population mean difference value of 0**.
4. On the horizontal line, have students add 10 “tick”marks (equally spaced) to the left of the midpoint and 10 “tick” marks to the right. The distance between tick marks will be a single unit on the measurement scale, which means our number line has a range of -10 to 10.
5. Give a bag to each group. Tell the students that they are going to randomly select 10 pairs of samples (with replacement) from their bag. Have all groups use a sample size of n = 3.
6. Have the students draw their first two samples, calculate the means, calculate the difference between the means (round to one decimal place), and then mark the location (with an “*M - M*”) on the horizontal line. This will be the center of the H1 distribution.
7. **IMPORTANT!** Always subtract the mean of the second sample from the mean of the first sample. This way you will get a representative sample of both positive and negative differences.

DISCUSSION: Ask the students what they think about the difference between the first two sample means? Can they draw any conclusions at this point? Are they really different? What is it that they’re really comparing? Means? Estimates of mean differences? Distributions?

1. Students should calculate the **estimated standard error of mean differences** of the mean, based on the data from their first two samples. If desired, have students label the horizontal line with standard-error-of-the-difference units.
2. Continue sampling. Remind the students that after each sample draw, they must put the data chips back in the bag, shake the bag and then redraw a new sample--until all 10 pairs of samples have been drawn, mean differences calculated for each of the 10 sample pairs and the location of each mean difference placed on the chart.
3. Have students draw two normal curves—one curve (H0) centering on the population mean difference of 0 and the other curve (H1) centering on the middle of the sampling distribution of mean differences.

DISCUSSION: After all groups have finished, have them discuss their findings as a class or within groups. The following might be good prompts for encouraging discussion.

* Now look at your chart again. What do you think of your previously calculated standard error of mean differences? Does it seem about right.
* What is it that we’re really comparing when we are analyzing the mean differences for statistical significance?
* Which of these curves represents the null distribution and which represents the alternative distribution? What do “null” and “alternative” mean within this context?
* If we conclude there is or is not a big difference between our means, is there a chance that our conclusion will be in error? Based on your chart, would that be a large or small chance? Why? How would I even know if my conclusion was in error?
* How was our sampling activity like data collection and analysis in a real research study? How was our sampling activity different?