Unit 7 Family Support Materials

Statistical Inferences



In this unit, your student will use a small *sample* of data to estimate information about a larger group called a *population* and use simulation to determine a range of values for the estimate. A population is the entire set of subjects of interest for a question, and a sample is a smaller group within that population.

For example, we may want to figure out the *mean* (or average) amount that families in the United States spend on food every month. The population includes all families in the United States, but collecting information from everyone would be very difficult and would cost a lot of money, so we might begin gathering data with a sample of 50 families.

An important question to consider when beginning to collect information from a sample is how the sample is to be selected. The data you collect might be very different if you ask families who are shopping at a local grocery store compared to families outside of a fancy restaurant. Similarly, the amount spent on food in San Francisco is likely very different from the amount spent in rural Iowa. There may even be some spending habits that are hidden in ways we haven't thought about yet. So, how do you make sure that your sample is representative of families in the United States, without using too many families from groups that are not typical in their spending?

The solution is to use *randomness*. We can select 50 families using a random process, such as having a computer select the families from a database at random without considering other factors. This should reduce the *bias* that might be introduced by humans trying to get information about people, and it will likely include more accurate proportions of the different types of families in the United States. While randomness may not entirely eliminate bias from the sample selection, it will significantly reduce the bias present when compared to selection without randomness.

Researchers have done studies like these to answer the question about spending. A report says that the mean amount spent on food each month is \$1,355 with a *margin of error* (MoE) of \$250. The margin of error is used to say that our estimate from the sample is probably not exact. We might expect the actual average amount of money spent on food each month to be somewhere between \$1,105 ($1,355 - 250 = 1,105$) and \$1,605 ($1,355 + 250 = 1,605$).

The margin of error is important to look for in statistical results. It is irresponsible to discuss statistics without providing a margin of error to describe how much the value is expected to vary. Many graphs included in news reports will report the margin of error in small print on the graphic. Look for something like $\pm 3\%$ when there is a graphic about approval rating for an official or about polls during the next election. This means that the percentages shown in the graphic could

actually be up to 3% lower or higher than the number shown.

Here is a task to try with your student:

A town has an upcoming vote about whether to raise corporate income taxes by 2% to increase funding for public schools. The local news shows an image that indicates that 52% of the voting population are in favor of the tax increase, and in the corner it shows "margin of error $\pm 3.5\%$." The reporter sounds confident that the corporate taxes will be increased because anything greater than 50% of the vote in favor of the taxes will pass the law.

1. The reporter got the result of 52% by driving to 4 of the 20 different neighborhoods around town and asking residents their opinion. Is there anything wrong with how that was done? Can you think of a better way to collect data?
2. What does the margin of error mean in this image?
3. Should you be confident that the taxes will be increased? Explain your reasoning.

Solutions:

1. Going to only 4 neighborhoods in the town might leave out the opinions of many voters from other neighborhoods in which the reporter did not go. A better way to collect information could be to randomly select several households in the town to survey about their opinions. The random selection is more likely to avoid any biases the reporter has about which neighborhoods to visit.
2. The margin of error means that the actual percentage in favor of increasing taxes could be 3.5% higher or 3.5% lower than the 52% reported from the sample. The actual percentage is most likely between 48.5% and 55.5%.
3. Sample responses:
    - I think there is still a good chance the taxes will be increased. Although the actual percentage could be as low as 48.5% because of the margin of error, it could also be as high as 55.5%. Most of the possible percentages are above 50%, so I think the increase will happen.
    - I think it is still unclear whether the increase will happen. The actual percentage could be as low as 48.5% because of the margin of error, which would cause the increase not to happen. I'm also not sure about the reporter's methods for collecting a sample in this report, so the report may not be very accurate.