Read Any Good Books Lately?

What makes a book easier or more difficult to read? An interesting story line, intriguing characters, and an appealing writing style certainly help. But from an educator's standpoint, the readability of a book or article depends mainly on word length and sentence length. In this Activity, you will learn some methods for evaluating readability. Work with one or two partners.

1. The first step in most standard readability tests is the selection of a random sample of words from the book or article. It would not be very practical to take a census of every word in an entire book! Using the randInt key from your calculator, randomly select a page from the book or article. Your sample will consist of the first 100 words in the randomly selected passage. Continue on to the next page as necessary. (If you have a one-page article, randomly select a paragraph.)
2. For each of the 100 words in your sample, count the number of letters and record your results on a dotplot.
3. Briefly describe the distribution of word length for your sample. Remember your SOCS!

In Chapter 7, we learned when we take a large sample from a population with mean μ and standard deviation σ, the distribution of the sample mean $\bar{x}$ (sampling distribution) is approximately Normal with mean μ and standard deviation $\sigma/\sqrt{n}$. Thus, in about 95% of large samples we would expect the sample mean $\bar{x}$ to be within two standard deviations of the true mean μ (from the 68-95-99.7 rule). This also means that an unknown population mean μ should be within two standard deviations of the sample mean $\bar{x}$ in about 95% of all samples.

4. Calculate the sample mean $\bar{x}$ and the sample standard deviation s_x for your sample of 100 words.
5. Calculate the following interval: $\bar{x} \pm 2\frac{s_x}{\sqrt{n}}$.

   This interval gives a range of plausible values for the population mean μ, the average number of letters per word in the *entire* book or article. Since intervals constructed in this manner will include the true mean about 95% of the time in repeated sampling, they are called 95% confidence intervals.
6. Share your results with the class by drawing your interval on the board and compare the confidence intervals for the different books and articles. Are you convinced that some books (or articles) are more difficult to read than others?

Another method for evaluating readability would be to estimate the *proportion* of "long" words in a book or article.

In Chapter 7, we learned that when we take large samples from a population where the true proportion of success is *p*, the distribution of the sample proportion $\hat{p}$ is approximately Normal with mean *p* and standard deviation $\sqrt{\frac{p(1-p)}{n}}$. Thus, in about 95% of large samples we would expect the sample proportion $\hat{p}$ to be within two standard deviations of the true proportion *p*. This also means that an unknown population proportion p should be within two standard deviations of the sample proportion $\hat{p}$ in about 95% of all samples.

7. How many words in your sample were at least 7 letters long? Calculate the value $\hat{p}$, the sample proportion of "long" words in your sample.
8. Calculate the following interval: $\hat{p} \pm 2\sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}}$.

   This interval gives a range of plausible values for the population proportion *p*, the true proportion of words that are at least 7 letters long in the *entire* book or article.
9. Share your results with the class by drawing your interval on the board and compare the confidence intervals for the different books and articles. Are you convinced that some books (or articles) are more difficult to read than others?