Author: Chris Franklin

Situation: Binomial Mean and Standard Deviation

Prompt

In a university statistics course designed for secondary mathematics majors, students were asked to consider the following scenario. [Source: Agresti, A. and Franklin, C. (2009). *Statistics: The Art and Science of learning from Data* (p.300). Upper Saddle River, NJ: Pearson/Prentice Hall.]

An exit poll is taken of 2789 voters in a statewide election. Let X denote the number who voted in favor of a special proposition designed to lower property taxes and raise the sales tax. Let's suppose that in the population, exactly 67% voted for it.

(a) Find the mean and standard deviation of the probability distribution of X.

(b) Now, suppose that the exit poll had x = 1774; i.e, 1774 indicated being in favor of the special proposition. Using the results from part (a), what do the results of this exit poll suggest about assuming the actual population percentage is 67%?

A student (a pre-service secondary teacher) approached answering part (a) as follows:

X is a binomial random variable with n = 2789 and the probability of a success p = 0.67. The mean of a binomial is given by np and the standard deviation is given by SQRT[np(1-p)]. Therefore the mean is 2789(.67) = 1868.63 = 1869. The standard deviation is given by SQRT[2789(.67)(.33)] = 24.83 = 25.

The student then proceeded to answer part (b) as follows:

If the population percentage is actually 67%, we would expect the mean of the counts from repeated random sampling of this population to be 1869 with the standard deviation being 25. The exit poll result of 1774 is almost 4 standard deviations below the expected mean of 1869; therefore, it is doubtful that this exit poll was taken from a population with the actual percentage being 67%. We would expect a count from the randomly selected exit poll to be within 2 or 3 standard deviations of the mean 1869 if the actual population percentage is 67%.

Commentary

The student recognized that the random variable X follows a binomial distribution and uses the correct formulas for finding the mean and standard deviation. However, he/she did not justify why the random variable is binomial. It is important to note that a binomial random variable is discrete, not continuous. Thus, when calculating the mean and standard, the student mistakenly rounded both the mean and standard deviation to a whole number. The student provides a nicely written statistical justification in part (b) for why he/she questions the population percentage being 67%. One detail missing from part (b) is why we would expect the expect the exit poll count to be within 2 or 3 standard deviations.

Focus 1

This focus will describe the conditions for a binomial distribution, applied to the above scenario. One condition, the *n* trials are independent, will lead to a discussion in Focus 2 of sampling with replacement versus sampling without replacement.

Focus 2

This focus will describe sampling with replacement versus sampling without replacement and why even though the exit poll used sampling without replacement, it is still reasonable to use the binomial distribution.

Focus 3

This focus will describe why the mean and standard deviation values should not be rounded to a whole number.

Focus 4

This focus will describe why the counts from the exit poll are expected to be within 2 or 3 standards deviations of the mean if the population percentage is actually 67%. This is because the specific binomial distribution in this scenario can be approximated by the normal distribution.

Focus 5 (or may prefer as Post-commentary)

This focus (or post-commentary) describes how the formulas for finding the mean and standard deviation of a binomial distribution are derived mathematically using the binomial theorem.