Math 236: Dealing with QUALITATIVE data---using the Sampling Distribution of a Proportion (Section 7.4)

Example: Thirty-three percent of physicians in the United States are women. (Source: Kaiser Family Foundation http://kff.org/other/state-indicator/physicians-by-gender/)

Suppose in doing a study, you drew a sample of 100 U.S. physicians and the sample had 40 female doctors. Would that be a pretty unusual number of women to show up in your sample?

How could we answer this question?

The Sampling Distribution of a Proportion
Suppose we have a population (like all the doctors in the U.S.) with proportion \( p \) of some characteristic (such as being female).

\[
p = \text{the percentage (proportion) of the population having some characteristic.}
\]

\[
q = 1 - p = \text{the percentage (proportion) of the population who don’t have that characteristic.}
\]

In a simple random sample of \( n \) individuals, suppose \( x \) of them have that characteristic. We can use this information to construct the sample proportion, \( \hat{p} \) (read this as “p-hat”).

\[
\hat{p} = \frac{x}{n} \quad \text{For our example} \quad \hat{p} = \frac{\text{number of women in the sample}}{\text{sample size}}
\]

Remember the “p-words” that all measure the same thing: percentage, proportion, probability. The language of probability is often used in this context. (In particular, Minitab uses these words.)

\[
n = \text{the number of “trials” (where a trial is to select a physician). This is just the sample size.}
\]

\[
x = \text{the number of “events” (so in the example above, the “event” is selecting a woman). This is just the number of women physicians who were in your sample.}
\]

Sampling Distribution of the Proportion: Now suppose you draw samples of size \( n \) from the population and find \( \hat{p} \) for all of these samples. You’ve taken your population data and created a whole new set of data from it, just as we did with the mean in the previous sections.
That whole new set of data (consisting of all the proportions from all the samples) is called “the sampling distribution of \( \hat{p} \)” and it has some important characteristics:

<table>
<thead>
<tr>
<th>1. The average of all those sample proportions is the true population proportion.</th>
<th>2. The standard deviation of all those ( \hat{p} )’s is given by the formula ( \sigma_{\hat{p}} = \sqrt{\frac{pq}{n}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu_{\hat{p}} = p )</td>
<td></td>
</tr>
</tbody>
</table>

And most importantly…

3. **The Central Limit Theorem.** As long as the sample size, \( n \), is large enough then all those \( \hat{p} \)’s will be approximately *normally distributed*.

Now we can answer the original question. The proportion of U.S. physicians who are women is 33%. Suppose a sample of 100 U.S. physicians had 40 female doctors. How unusual is this result?

1. Parking Lot!
2. Check that the sample size is large enough.
3. Write the question as a probability problem.
4. Sketch a normal curve, showing \( p, \hat{p} \) and shading in the relevant area.
5. Find \( \sigma_{\hat{p}} = \sqrt{\frac{pq}{n}} \)
6. Find the z-score for \( \hat{p} \)
7. Find the shaded area, using the Normal Distribution Calculator.
8. Write a conclusion.
Example: How Many Jobs Do You Have? The Bureau of Labor Statistics reported in a recent year that 5% of employed adults in the U.S. worked more than 1 job. A random sample of 15 employed adults is chosen and it’s found that 6.5% of the individuals in the sample had more than 1 job. Is it appropriate to use a normal approximation to find the probability that more than 6.5% of the individuals in the sample have multiple jobs? Why or why not?

(b) A new sample of 350 employed adults is chosen. Find the probability that between 3.5% and 6.5% of the individuals in the sample have multiple jobs.

(c) A new sample of 350 employed adults is chosen and 6.5% of this sample has more than 1 job. Find the probability that 6.5% or more of the individuals in this sample would have multiple jobs, if the true population percentage is 5%.