Math 114 **ACTIVITY 8:** Sampling distribution of sample proportion $\hat{p}$

Why

We have looked at the sampling distribution of the sample mean $\bar{x}$ because we want to be able to use the sample mean to give information about the mean of the population from which the sample is drawn. In the same way, we are interested in *proportions* – the proportion of heart attack patients who recover if treated in a certain way, the proportion of voters in a town who will support some initiative, etc. When we are thinking of sampling, this is the p of the binomial distribution, and it is more useful, in most situations, to discuss the *proportion* of successes rather than the *number* of successes .We want to look at the distribution of *sample proportions* (proportion of "Yes" results in samples) for the purpose of tying this to the population proportion.

LEARNING OBJECTIVES

1. Learn the meaning and interpretation of sample proportion in a binomial sampling setting.
2. Be able to calculate the mean and standard deviation of the distribution of sample proportions.
3. Know when the normal distribution is an appropriate model for the distribution of sample proportions
4. Be able to calculate probabilities for sample proportions, based on the normal model.

CITERIA

1. Success in working as a team and in fulfilling the team roles.
2. Success in involving all members of the team in the conversation.
3. Success in completing the exercises.

RESOURCES

1. Your text, section 8.2 (and section 7.5) and the standard normal probability table (Table V) pages A-11 & A-12
2. The team role desk markers (handed out in class for use during the semester)
3. Your class notes from Monday and Wednesday
4. Your calculator
5. 40 minutes

PLAN

1. Select roles, if you have not already done so, and decide how you will carry out steps 2 and 3 (5 minutes)
2. Work through the exercises given here - be sure everyone understands all results & procedures(30 minutes)
3. Assess the team's work and roles performances and prepare the Reflector's and Recorder's reports including team grade (5 minutes).

DISCUSSION

Meaning of $\hat{p}$:
In any population, there is a proportion of individuals who have any characteristic that can be named (proportion of all college students who are tired, proportion of cell phones that do not have picture-taking capability, etc.). We represent this proportion by p. When we take a sample of size n from the population, the *number* of members in the sample that have the characteristic is a binomial variable X (if the sampling is random, sample is a small part of the population) with n trials and probability p on each trial. It is often more useful to express our results in terms of the *sample proportion* $\hat{p} = \frac{X}{n}$ (number of "Yes" results divided by sample size.
For example, if we are studying a stop-smoking program, we may take a sample of 30 people who have been through the program and observe the number who have successfully stopped smoking. If X = number (in the sample) of successful stops, then the sample proportion is the fraction $\hat{p} = \frac{X}{30}$ If we find twelve successful stoppers, we can report that X = 12 or that $\hat{p} = \frac{12}{30} = .40$ (40% of the sample were successful). For many of our sample/population connections, this form is more useful.

Mean and standard deviation of $\hat{p}$:
If we take random samples of the same size n from a large enough population that X can be regarded as binomial (population at least 20 times sample size), then $\mu_{\hat{p}} = \frac{\mu_X}{n} = \frac{np}{n} = p$ and $\sigma_{\hat{p}} = \frac{\sigma_X}{n} = \frac{\sqrt{np(1-p)}}{n} = \sqrt{\frac{p(1-p)}{n}}$.
That is:
$\mu_{\hat{p}} = p$ and $\sigma_{\hat{p}} = \sqrt{\dfrac{p(1-p)}{n}}$

Shape of distribution of $\hat{p}$:
If sample size n is large enough that X is approximately normal (that is, if $np(1-p) \geq 10$) then $\hat{p}$ is approximately normally distributed. however, we do *not* need the continuity correction, because $\hat{p}$ is treated as a continuous variable - we are not converting from discrete to continuous.

Summary:
If we sample at random from a large enough population (at least 20 times sample size) and the sample is large enough that $np(1-p) \geq 10$, then the sample proportion $\hat{p}$ is a random variable which is approximately normal with mean $\mu_{\hat{p}} = p$ and $\sigma_{\hat{p}} = \sqrt{\dfrac{p(1-p)}{n}}$.

MODEL

1. If we take a sample of 80 students and find that 60 love statistics, the our sample proportion of students who love statistics is $\hat{p} = \frac{60}{80} = .75$
2. Approximately 6% of people have O-negative type blood. Consider taking samples of 600 Americans and observing the *proportion* of people in the sample with O-negative blood type. [This is an example we used with the normal approximation to the binomial–but we worked with the *count* of people there]
   (a) The sample size (600) is a small enough part of the population size (millions) that we can consider the variable "number of O-negative people in the sample" to be given by a binomial variable
   (b) Since $np(1-p) = 600(.06)(.94) = 33.84$ is greater than 10, we can consider $\hat{p}$, the proportion of O-negative people in the sample, to be approximately normal
   (c) The mean of $\hat{p}$ is $\mu_{\hat{p}} = .06$ and the standard deviation of $\hat{p}$ is $\sigma_{\hat{p}} = \sqrt{\frac{.06(.94)}{600}} = .0096$
   (d) The probability of getting a sample proportion less than .05 (less than 5% of the sample with O-negative blood type) is $P(\hat{p} < .05) = P(Z < \frac{.05-.06}{.0096}) = P(Z < -1.04) = .1492$. About 15% of samples (of size 600) would give a sample proportion less than .05.
   (e) The probability of getting a sample with sample proportion within 1.5% of the population proportion (remember $p = .06$ – so we're talking about a sample proportion between .045 and .075) is $P(.045 < \hat{p} < .075) = P(\frac{.045-.06}{.0096} < Z < \frac{.075-.06}{.0096}) = P(-1.56 < Z < 1.56) = P(Z < 1.56) - P(Z \leq -1.56) = .9406 - .0594 = .8812$. There is about an 88% chance of getting a sample proportion (with a sample size of 600) which is within .015 of the population proportion.

EXERCISES

1. At Huge U, 42% of the (thousands of) students live off-campus. If we want to consider random samples of 30 students
   (a) Can we consider the *number* of off-campus students in a sample as a *binomial* variable?
   (b) What are the mean and standard deviation of $\hat{p}$ the *proportion* of off-campus students in the sample?
   (c) Is our sample size large enough to allow treating $\hat{p}$ as a *normally distributed* variable?
   (d) How large a sample would we need in order to be justified in treating $\hat{p}$ as a normal variable?
2. A "stop smoking" program claims a 65% success rate – that is, they claim that 65% of their (thousands of) former participants are non-smokers two years later.
   (a) If the claim is true, what are the mean and standard deviation of the proportion ($\hat{p}$) of non-smokers (two years later) in samples of 50 former participants?
   (b) If the claim is true, what is the probability that a sample of 50 of their former participants would include fewer than half who are non-smokers two years later?
3. If the claim made by the program in the previous exercise is true (proportion of non-smokers p really is 65%) and we take a sample of 80 former participants:

(a) What is the probability that the proportion of non-smokers (two years later) in the sample would be between 55% and 70%?

(b) How likely is a sample in which the proportion of non-smokers is less than 50%? Would this be an unusual occurrence (if the claim is true)?

4. Suppose the success rate for the program in the previous exercise is really 55% (which means, among other things, that their claim is not correct).

   (a) What proportion of samples of 80 former participants would give a proportion of non-smokers between 55% and 70%?

   (b) How likely is a sample of 80 former participants in which the proportion of non-smokers is less than 50%?

READING ASSIGNMENT (in preparation for next class)
No new reading assignment – begin reviewing for Test 2 (Chapters 5 – 8) on Wednesday

SKILL EXERCISES:(hand in - individually - at next class meeting) Sullivan p.398 #11–12, 17 – 18, 20 – 22