Mon, Wed and next Mon: Chapter 9
Understanding Sampling Distributions: Statistics as Random Variables

Friday: Section 7.7 and additional material on probability and intuition
(Guest lecturer: Thao Duong, TA for Sections 3 and 4.)

ANNOUNCEMENTS:
Given in class.

HOMEWORK:
Chapter 9: #25, 37

Introduction to statistical inference: On chalk board.

Five situations we will cover for the rest of this quarter:

Table 9.1 Population Parameters and Sample Statistics for the Big Five Scenarios

Parameter Name and Description	Symbol for the Population Parameter	Symbol for the Sample Statistic
For Categorical Variables:		
One population proportion (or probability)	p	$\hat{p}$
Difference in two population proportions	$p_1 - p_2$	$\hat{p}_1 - \hat{p}_2$
For Quantitative Variables:		
One population mean	μ	$\overline{x}$
Population mean of paired differences (dependent)	μ_d	$\overline{d}$
Difference in two population means (independent)	$\mu_1 - \mu_2$	$\overline{x}_1 - \overline{x}_2$

For each parameter will we:

- Learn how to find a confidence interval for its true value
- Test hypotheses about its true value

Examples of each of these on chalk board.

GOAL: Estimate and test *parameters* based on *statistics*.
(Get confidence intervals and do hypothesis tests)

Intermediate step: Find out what *statistics* to expect, *given* the value of a *parameter*.

SOME LOGICAL NOTES:

1. Assuming the sample is representative of the population, the *sample statistic* should represent the *population parameter* fairly well.

2. But, the sample statistic will have some error associated with it, i.e. it won't equal the parameter exactly.

3. If repeated samples are taken from the same population and the sample statistic is computed each time, these sample statistics will *vary* but in a *predictable way*, i.e. they will have a *distribution*. It is a *pdf* for the statistic. It is called a **sampling distribution** for the statistic.

The Sampling Distribution for a Sample Proportion $\hat{p}$

1. Physical situation: binomial.
   Actual population with fixed proportion w/trait or opinion (e.g. polls, TV ratings, etc.)
   OR
   A repeatable situation with fixed probability of a certain outcome (e.g. birth is a boy, probability of heart attach if one takes aspirin, etc.)

2. The Experiment
   A random sample of size n from the population, $\hat{p}$ = proportion w/trait
   OR
   Repeat the situation n times, $\hat{p}$ = proportion with specified outcome

3. Size requirement: In either case, must have np and n(1- p) at least 5, prefer at least 10.

THE NORMAL CURVE APPROXIMATION RULE FOR SAMPLE PROPORTIONS: Assuming the above conditions are met,

the distribution of *possible* values of $\hat{p}$ is approximately normal with

mean μ = p and standard deviation $\sigma = \sqrt{\frac{p(1-p)}{n}}$

The resulting normal distribution is called the ***sampling distribution of*** $\hat{p}$

NOTATION:

s.d.($\hat{p}$) = standard deviation of $\hat{p}$ $= \sqrt{\frac{p(1-p)}{n}}$

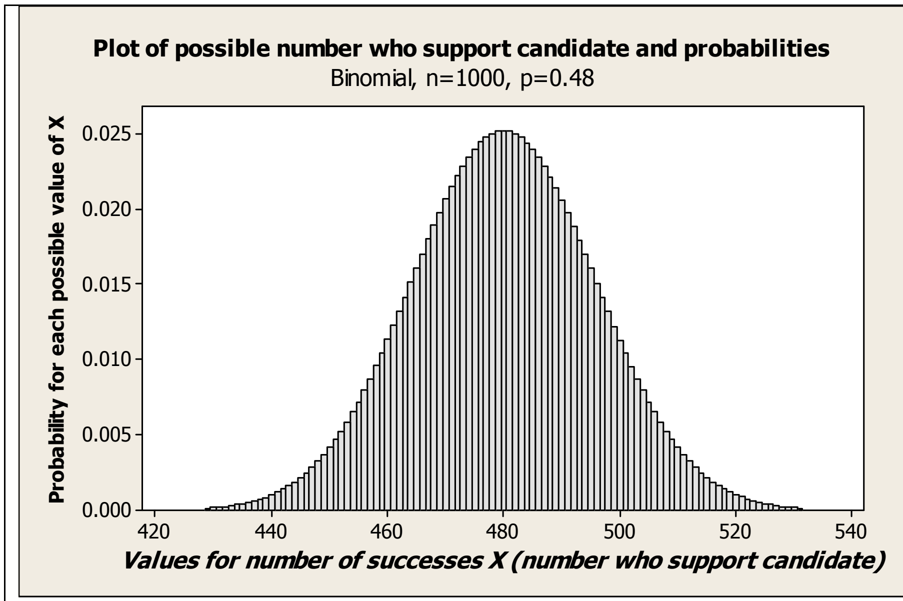
s.e.($\hat{p}$) = standard error of $\hat{p}$ $= \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

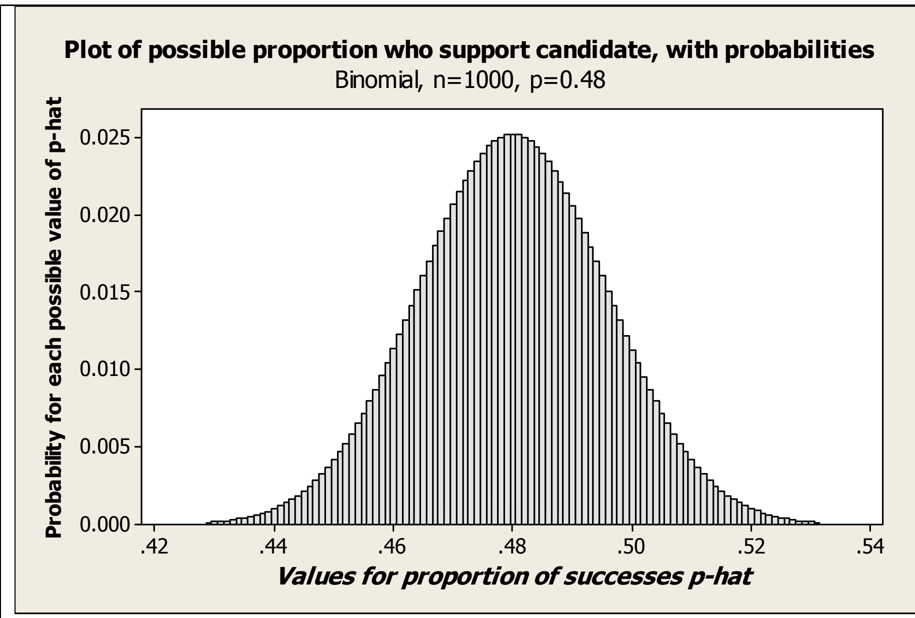
= *estimate* of s.d.($\hat{p}$)

(Note that if p is unknown, s.d.($\hat{p}$) can't be computed but s.e.($\hat{p}$) can.)

Familiar example with picture: Suppose 48% of a population supports a candidate. In a poll of 1000 randomly selected people, what do we expect to get for the *sample proportion* $\hat{p}$?

$\hat{p} = \frac{X}{n}$ where X is a binomial random variable. We have seen picture of possible values of X. Divide all values by n to get picture for possible $\hat{p}$.





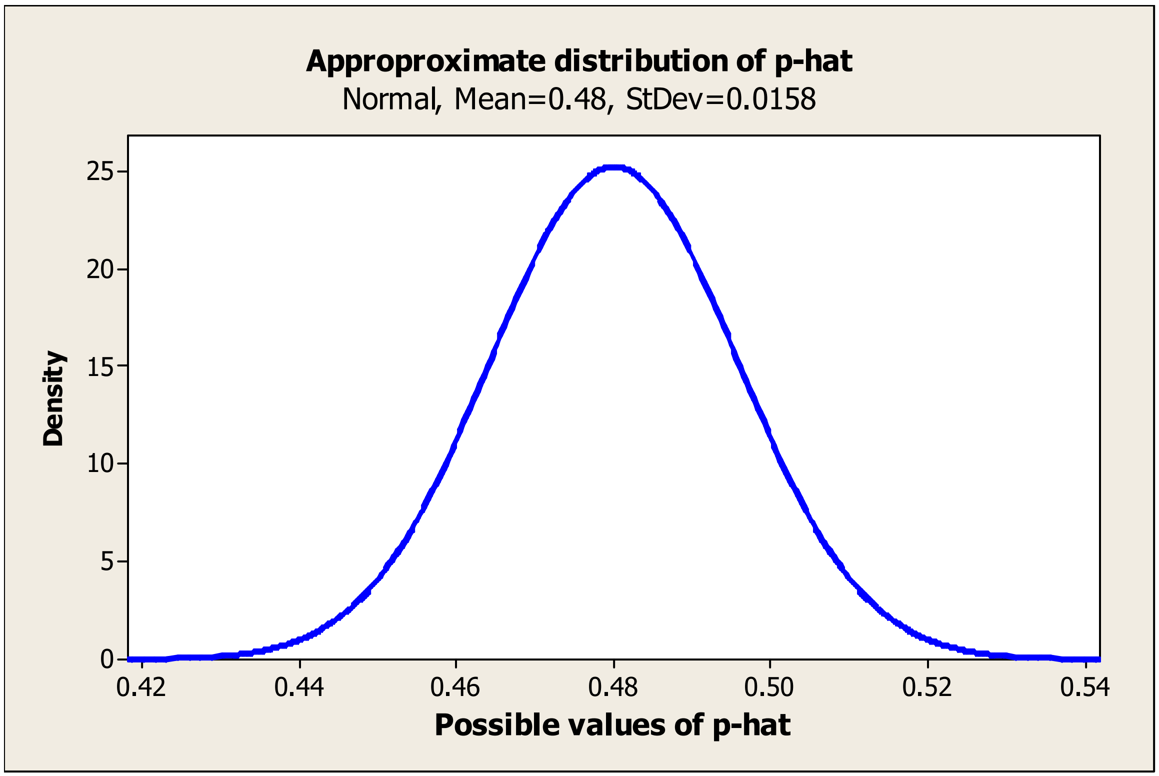
What's different and what's the same about these two pictures?

The normal curve approximation rule for sample proportions for this example:

Poll of n = 1000 people, where the *true population proportion* p = .48.

The distribution of *possible* values of $\hat{p}$ is approximately normal with

$$\mu = p = .48 \quad \text{and} \quad \sigma = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{.48(1-.48)}{1000}} = .0158$$



Examples (details on chalk board):

1. Nielsen ratings.

2. Proportion of young adults who often text while driving. Marist poll taken October 2009. n = 200 aged 18 to 29 “How often do you text while driving?”