Ch. 18 – Sampling Distribution Models (Day 1 – Sample Proportions)

Survey Results

- During the first week of the Obama presidency, the Gallup Poll released the new president's first approval ratings. Their survey of 1,591 adults found that 68% approved of how he had performed so far.

- If Gallup could have surveyed all U.S. adults, they would have obtained the true (population) proportion who approved of Obama's job performance

- Obviously this is not practical, so they depended on their sample proportion instead

- p = population proportion (parameter)

- $\hat{p}$ = sample proportion (statistic)

Sampling Variability

- If the Gallup Poll had conducted a second poll asking the same question, they probably would have obtained slightly different results (perhaps 67% or 69%)

- Because each sample will contain different, randomly selected individuals, the results will be a little different each time

- Remember that this is called <u>sampling error</u> or <u>sampling variability</u>

Sampling Distribution

- Suppose Gallup repeated the sampling procedure many times and recorded the proportion who approve of the president in each sample:
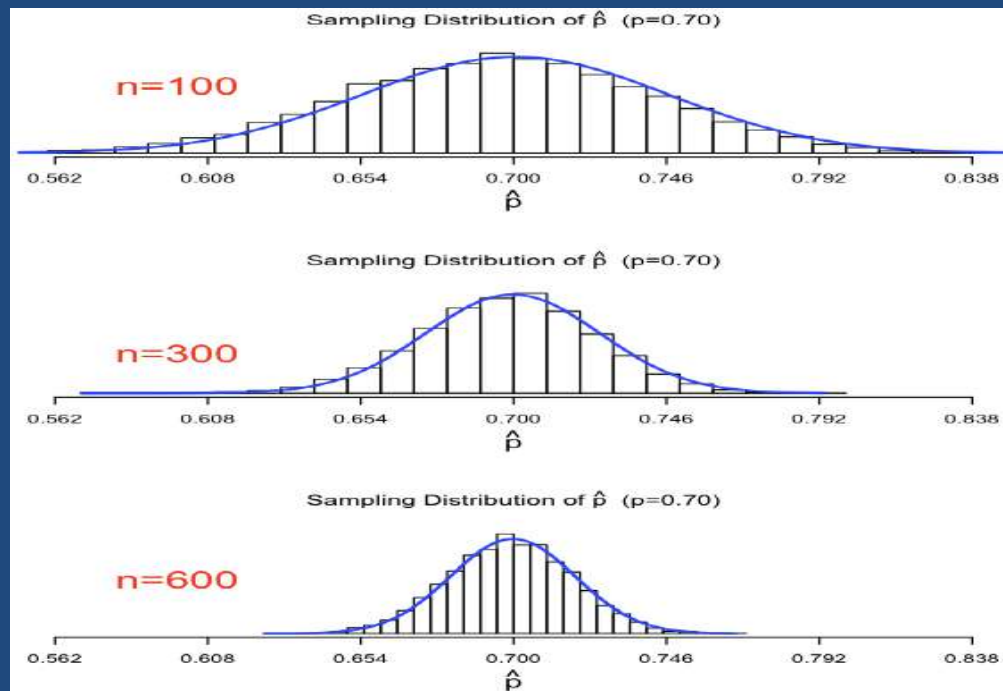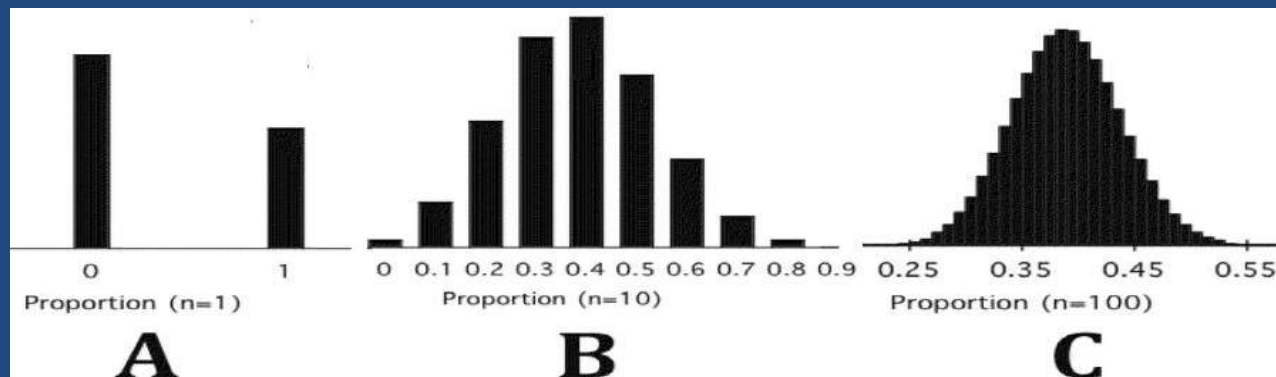
Sampling Distribution

- In real life, we would only conduct the sample once, but we can *imagine* what the distribution of all possible samples would look like

- This is called the sampling distribution

- Sampling Distribution Model: A model representing the values taken on by a statistic over all possible samples of size n

Characteristics of the Sampling Distribution of Sample Proportions

- Suppose that in reality, the true proportion of Americans who approve of Obama's performance is 40%.

- If we took many, many samples, we would expect:

- Does this shape look familiar?

Characteristics of the Sampling Distribution of Sample Proportions

- The model follows a Normal distribution

- Remember that we describe Normal models using the notation N(μ, σ)

- The mean (μ) is the center of the distribution

- In this case, we would expect that to be the true proportion, p

- For the sampling distribution of a proportion:

$$\mu_p =$$

How does this help us?

- We can make calculations involving sample proportions using these ideas

- Ex Suppose the true presidential approval rating is 67%. What is the probability that a random sample of 1000 adults would have less than 650 who approve of the president's performance?

Using your formula sheet

- Remember that we have been using the formula

$$z = \frac{x - \mu}{\sigma}$$

- But now, we are going to use this more general form (on your formula sheet)

$$z = \frac{\text{statistic} - \text{parameter}}{\text{standard deviation}}$$

Using your formula sheet
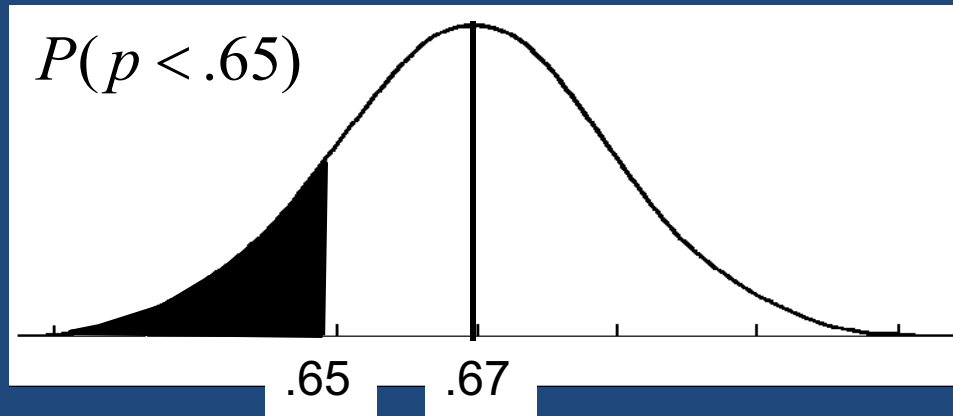
- For sample proportions, this means

$$z = \frac{p - p}{\sqrt{\dfrac{p(1-p)}{n}}}$$

Back to the problem…

$p = .67$

$n = 1000$

$p =$

$P(p < .65)$

.65 .67

$z = \dfrac{ - }{\rule{3cm}{0.4pt}}$

$P(p < .65) =$

Assumptions & Conditions

- Like most statistical procedures, sampling distribution models only work when certain assumptions are met

- The assumptions are:

  1. The sampled values must be independent of each other

  2. The sample size, n, must be large enough (the larger the sample, the better the Normal model fits)

Assumptions & Conditions

- In order for those assumptions to be reasonable, we need to check these conditions:

  1. Randomization: The sample should be a SRS, or at least a non-biased, representative sample

  2. 10% Condition: The sample should be less than 10% of the population

  3. Success/Failure Condition: We should expect at least 10 successes & 10 failures

$$np \geq 10 \ \text{ and } \ n(1-p) \geq 10$$

Checking Conditions

- Do the necessary assumptions seem reasonable for the Gallup Poll example we just did?
  - Random Sample? *Yes, given in problem*
  - Sample < 10% of population?
    Yes, since there are more than 10,000 Americans (1000 is less than 10% of all Americans)

  - n large enough?
    $$np \geq 10 \text{ and } n(1-p) \geq 10$$
    $$1000(.67) \geq 10 \text{ and } 1000(.33) \geq 10$$
    $$670 \geq 10 \text{ and } 330 \geq 10$$

Putting it all together…

- A national study found that 44% of college students engage in binge drinking. After reading about this, a professor conducts a random survey of 244 of the students at his college.

- Use the 68-95-99.7 rule to describe the sampling distribution for this situation

$\mu_p =$

$\sigma_p =$

.35   .38   .41   .44   .47   .50   .53
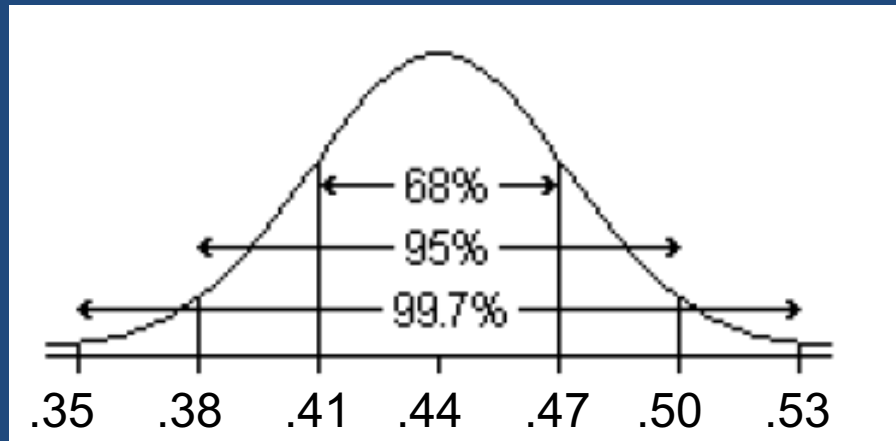
Don't forget assumptions!

- Do you think the appropriate conditions are met to use this model?
  - Random Sample? *Yes, given in problem*
  - Sample < 10% of population?
    We have to assume there are more than 2440 students in the college (244 is less than 10% of all students in the college)
  - n large enough?

$$np \geq 10 \quad \text{and} \quad n(1-p) \geq 10$$

$$244(.44) \geq 10 \quad \text{and} \quad 244(.56) \geq 10$$

$$107.4 \geq 10 \quad \text{and} \quad 136.6 \geq 10$$

Using the Model

- When he conducted his poll, the professor found that 96 of the 244 students he surveyed admitted having been engaged in binge drinking. Should he be surprised?

$$p = \frac{96}{244} = 0.393$$

.35   .38   .41   .44   .47   .50   .53
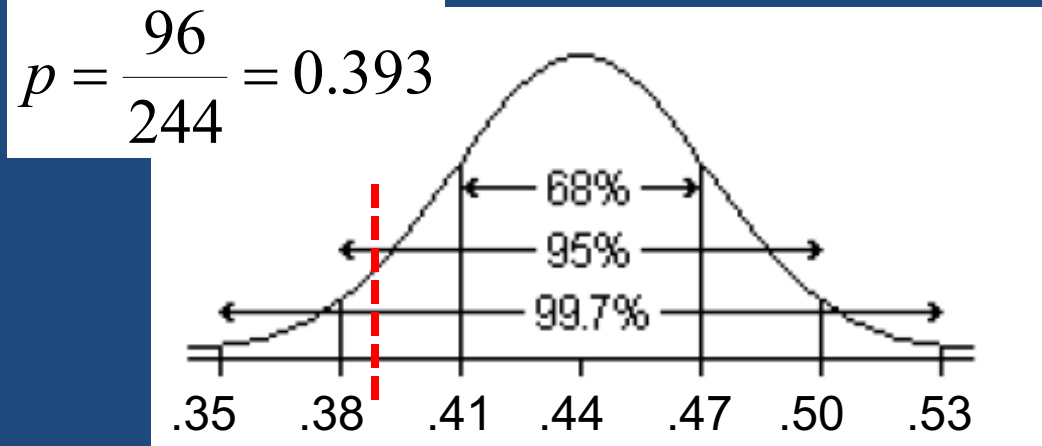
He should not be surprised – this is less than 2 standard deviations away from the mean, so it's not that unusual.
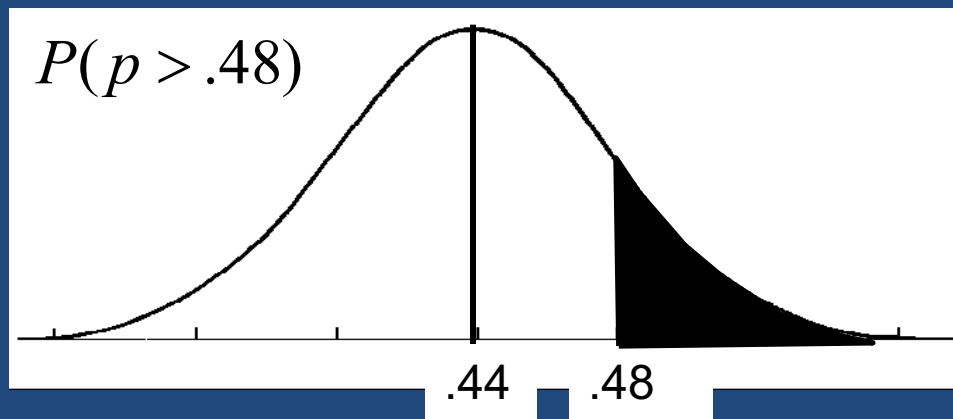
One more question...

- What is the probability that the professor's survey would have more than 48% admitting they have engaged in binge drinking?

$p = .44$

$n = 244$

$p = .48$

$P(p > .48)$

.44   .48

$z = \dfrac{-}{\rule{2cm}{0.4pt}}$

$P(p > .48) =$

Homework 18-1

- p. 432

  #12, 14, 16, 21, 24