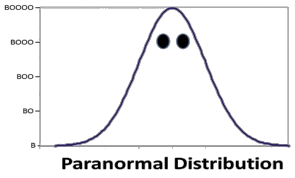
Lesson 31: What is a sampling distribution model?



Do you believe in ghosts?

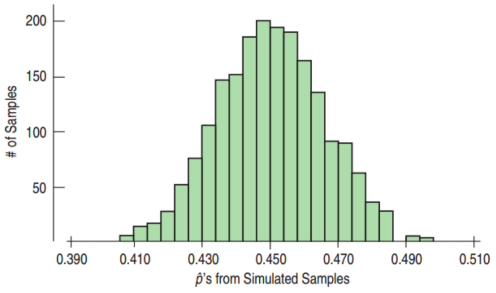
Two polls at the same time

Do you believe in ghosts?

1. 2005 Harris Poll
   - 889 adults
   - 40% said they believed in ghosts

2. CBS News Poll
   - 808 adults
   - 48% said they believed in ghosts

Q	Why do sample proportions vary?
A	samples are composed of diff. people
Q	What is the true population proportion???
Q	How much variability can I expect from sample to sample?
A	Simulation - pretend samples to get an understanding. Imagine taking thousands of samples and plotting the sample proportions.



A histogram of sample proportions for 2000 simulated samples of 808 adults drawn from a population with p=.45

The sample proportions vary, but their distribution is centered at the true proportion, *p*

True population proportion is our goal

Sampling distribution

shows what we could expect if we could see all the proportions from all the possible samples.

(Note: Sample proportion is a random variable.)

Expect the histogram to be unimodal, symmetric, and centered at *p*

↓

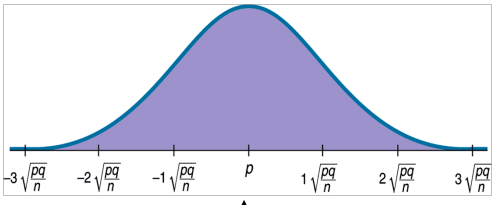
shape reminds of normal model

↓

call it a sampling distribution model : allows us to think about how likely it is for us to observe a sample proportion in any particular interval.

Recall, normal modal requires two parameters $N(\mu,\sigma)$

Sampling distribution model requires $N\left(p,\sqrt{\frac{pq}{n}}\right)$



But we run simulations supposing we know the population proportion, so....

We don't know the true population proportion.

$$\hat{p} = \frac{\#\,successes}{\#\,trials}$$

proportion of successes

observed proportion from our data

our estimate for actual *p*

$$\mu(\hat{p}) = p$$
$$\sigma(\hat{p}) = \sqrt{\frac{pq}{n}}$$

Recall Binomial Probability

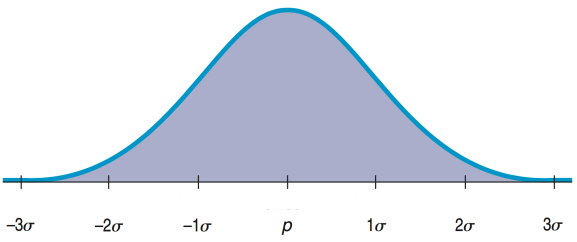
$$\sigma(\hat{p}) = \frac{\sqrt{npq}}{n} = \sqrt{\frac{pq}{n}}$$

Back to ghosts

suppose p = 45%

$$\mu(\hat{p}) = p = .45$$
$$\sigma(\hat{p}) = \sqrt{\frac{pq}{n}} =$$

$\}$ $N(.45,.0175)$



CBS poll is within two standard deviations of *p* guess.
Our model says 95% of all sample proportions will fall between 41% and 48.5%.
What we observed in ghosts poll was just sampling variablility.

** Remember: a model is a model

This model is better with larger sample size.

Ghosts -> 2 or 3 individuals are just not a good size for sampling proportions.

What's a good sample size then?????

Assumptions and conditions to use the normal model for sampling proportions.

1. Independence Assumption - sampled values are independent of each other
2. Sample size n is large enough

These may be hard to check...so instead check the following conditions to use the normal model.

1. Randomization Condition: Randomization was used in the experiment, surveys to make representative samples

2. 10% condition: n must be no larger than 10% of the entire population (just ask if the entire population is 10 times the sample size or larger)

3. Success/failure condition: sample is big enough for us to expect at least 10 successes and 10 failures

CBS survey np = (808)*(.45) = 364 successes

nq=(808)*(.55)=444 failures

Example from page 418

Suppose that about 13% of the population is left-handed.8 A 200-seat school auditorium has been built with 15 "lefty seats," seats that have the built-in desk on the left rather than the right arm of the chair. (For the right-handed readers among you, have you ever tried to take notes in a chair with the desk on the left side?)

Question: In a class of 90 students, what's the probability that there will not be enough seats for the left-handed students?