Sampling Distributions

part i: proportions

AP Statistics
Chapter 18

SAMPLING* DISTRIBUTION

Collection of a LARGE NUMBER of sample statistics (means or proportions) of a given sample size.

**not "sample distribution"!*

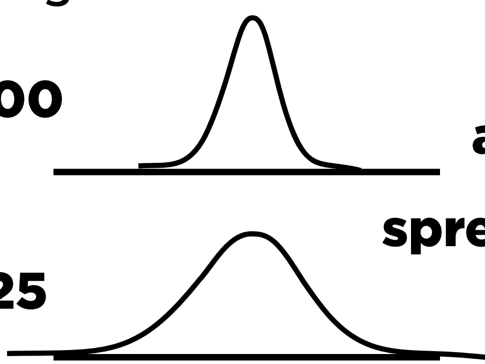
SAMPLE SIZE vs SPREAD

What do we notice about the sampling models for…

n = 100

vs

n = 25



as n ↑, spread ↓.

REMEMBER BIAS? (THIS WAS FROM THE FIRST WEEK OF SCHOOL…)

a) After 9/11, President Bush authorized government wiretaps on some phone calls in the U.S. without getting court warrants, saying this was necessary to reduce the threat of terrorism. Do you approve or disapprove of this?

53% of respondents approved

b) After 9/11, George W. Bush authorized government wiretaps on some phone calls in the U.S. without getting court warrants. Do you approve or disapprove of this?

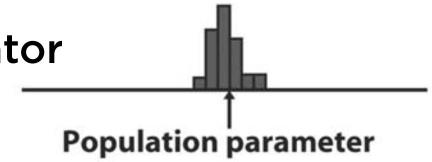
46% of respondents approved

*THIS WAS JUST ONE FORM OF BIAS (RESPONSE BIAS)… THERE WERE OTHER FORMS OF BIAS AS WELL (VOLUNTARY, NON-RESPONSE, ETC)

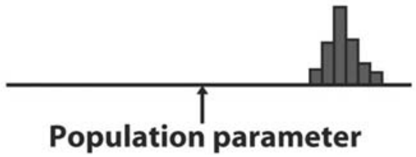
UNBIASED VS BIASED ESTIMATORS

If we sample properly…

UNBIASED estimator



If there is BIAS in sampling method…



BIASED estimator

(sampling model does not properly represent the "truth")

SAMPLING DISTRIBUTION FOR PROPORTIONS:

These are found on the formula chart!

$$\mu_{\hat{p}} = p$$

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

If certain conditions are met, we can use the Normal model to approximate the sampling model…

Assumptions (you **DON'T** need to write these down…)

- Sampled values are independent of one another
- Sample size is large enough.

Assumptions are hard — often impossible — to check. That's why we *assume* them.

CONDITIONS (TO SEE IF THE ASSUMPTIONS ARE REASONABLE)

1. **Random Sample:**
   The sample should be a SRS of the population
   *(or at least **representative** of the population… hopefully we have no reason to suspect bias???)*

2. **10% condition:**
   If sampling has been made **WITHOUT REPLACEMENT**, then the sample size, *n*, should not exceed 10% of the population size.
   *(this is so that sampling without replacement is **SIMILAR ENOUGH** to sampling WITH replacement)*

3. **Success/failure condition (NORMALITY):**
   np and *nq* should be at least 10.

Basic steps for probability problems with sample proportions

1. Define the parameter (using "p", NOT "p-hat"!)
2. Check the necessary conditions
3. Find the mean and standard deviation of "p-hat"
4. Use the normal model to find the probability
5. Interpret the probability in context *(write out a sentence explaining what the probability that you found means)*

I. The polling problem

Just before a referendum on a school budget, a local newspaper polls 400 voters in an attempt to predict whether the budget will pass. Suppose that the budget actually has the support of 52% of the voters. What's the probability the newspaper's sample will lead them to predict defeat?
(in other words, what is the probability that less than half of the sample approves of the budget?)

sample size "n"

true proportion p

p = the true proportion of ALL voters who favor the budget

Mean & St. Dev of $\hat{p}$:

$$\mu_{\hat{p}} = p = 0.52$$

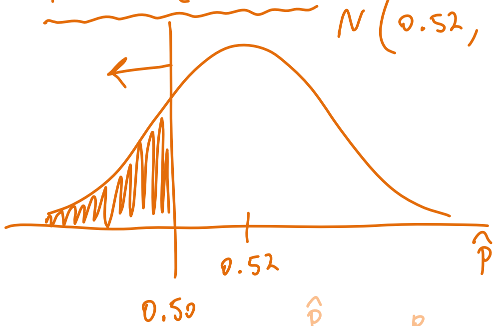
$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.52(1-0.52)}{400}} = 0.02498$$

Conditions:

- Random sample?
  Hopefully the newspaper took a random sample of 400 voters. (We don't know this, but no obvious reason to suspect bias)
- 10% condition
  400 voters is reasonably < 10% of all voters in this area
- Normal approximation
  $np = 400(0.52) = 208 \geq 10$ ✓
  $nq = 400(1-0.52) = 192 \geq 10$ ✓

Find $P(\hat{p} < 0.5)$:

$N(0.52, 0.02498)$



$$z = \frac{\overset{\hat{p}}{\text{obs}} - \overset{p}{\text{mean}}}{\underset{\sqrt{\frac{pq}{n}}}{\text{st. dev}}} = \frac{0.50 - 0.52}{0.02498}$$

$$= -0.8006$$

$$P(z < -0.80) = \boxed{0.2119}$$

★ The probability that fewer than half of the 400 sampled voters favor the budget is about 0.2119.

II. The lefty-desk problem

extraneous info !!

Suppose that about 13% of the students at a large college are left-handed. A 200-seat auditorium has been built with 15 "lefty seats." In a class of 90 students, what's the probability that there will **not** be enough seats for the left-handed students? → this means $> \frac{15}{90}$ are lefty... or $\hat{p} > \frac{15}{90}$

p = the true proportion of all students at this college who are lefty.

(We need to find probability that $\hat{p} > \frac{15}{90}$)

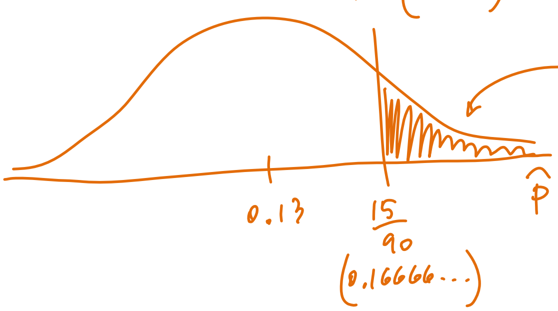
Mean & St. Dev of $\hat{p}$:

$$\mu_{\hat{p}} = p = 0.13$$

$$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.13(0.87)}{90}} = 0.03545$$

Find $P\left(\hat{p} > \frac{15}{90}\right)$ ← 0.166667

$N(0.13,\ 0.03545)$



$$z = \frac{\text{obs} - \text{mean}}{\text{st. dev}}$$

$$= \frac{\frac{15}{90} - 0.13}{0.03545} = 1.0343\ldots$$

$$p\left(\hat{p} > \frac{15}{90}\right)$$

$$= p(z > 1.0343)$$

$$= 1 - 0.8485 \text{ (ish)}$$

$$\approx \boxed{0.1515 \text{ (ish)}}$$

☆ The probability that more than 15 of the 90 students are left-handed is about 0.1515.

Conditions:

- The 90 students in this class may not be a proper random sample, but (hopefully) are representative of all students (in terms of them being lefty or not)
- 90 students is surely < 10% of all students at this "large college" (therefore even though we are sampling withOUT replacement, it is mathematically similar enough to sampling WITH replacement)
- Normality:

  $np = 90(0.13) = 11.7 \geq 10$ ✓

  $nq = 90(1-0.13) = 78.3 \geq 10$ ✓