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"!*

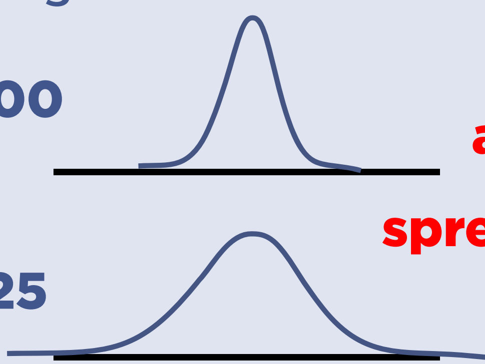
SAMPLE SIZE vs SPREAD

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?

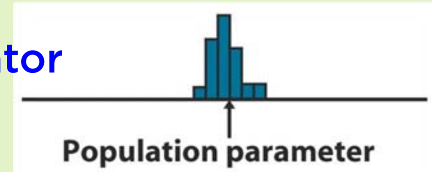
46% of respondents approved

*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

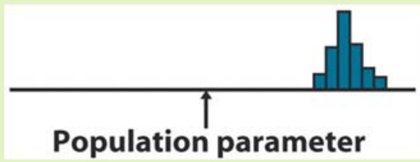
If we sample properly…

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.

10. Contacts. Assume that 30% of students at a university wear contact lenses.

a) We randomly pick 100 students. Let $\hat{p}$ represent the proportion of students in this sample who wear contacts. What's the appropriate model for the distribution of $\hat{p}$? Specify the name of the distribution, the mean, and the standard deviation. Be sure to verify that the conditions are met.

b) What's the approximate probability that more than one third of this sample wear contacts?

p = true (population) proportion of students at a university who wear contact lenses (assumed to be 0.30) $p = 0.3$ $q = 0.7$ $n = 100$

a) Conditions to use the Normal model:

- Randomization: It is given that we have a random sample of 100 students.
- 10% condition: We assume that 100 students represents less than 10% of the total student population at the university (in other words, the university has at least 1000 students - this is realistic)
- success/failure condition:

$np = 100(0.3) = 30 \geq 10$ ✓

$nq = 100(0.7) = 70 \geq 10$ ✓

The conditions are met, so we may use the Normal model to approximate the sampling distribution of $\hat{p}$.

Mean: $\mu_{\hat{p}} = p = 0.30$

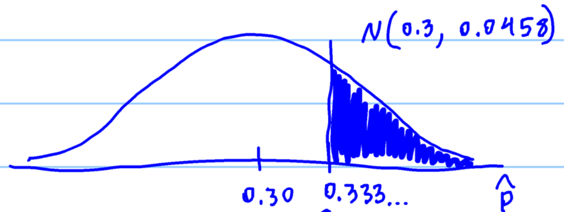
Standard Deviation:

$$\sigma_{\hat{p}} = \sqrt{\frac{p \cdot q}{n}} = \sqrt{\frac{(0.3)(0.7)}{100}}$$

$$SD(\hat{p}) = \sigma_{\hat{p}} = 0.0458$$

$$N(0.3, 0.0458)$$

$\hookrightarrow P\left(\hat{p} > \frac{1}{3}\right)$



b) $$z = \frac{\text{obs} - \text{exp}}{SD} = \frac{\hat{p} - p}{\sigma_{\hat{p}}}$$

$$= \frac{\frac{1}{3} - 0.3}{0.0458}$$

$$z = 0.7278$$

$$P(\hat{p} > 0.3333) = P(z > 0.73)$$

$$= 1 - 0.7673$$

$$= \boxed{0.2327}$$

calculator: 0.2333

$\hookrightarrow$ Normalcdf(0.7278, 99)

☆ There is about a 23.3% probability that more than a third of this sample of university students wears contact lenses.

15. Polling. 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? Be sure to verify that the assumptions and conditions necessary for your analysis are met.

p = population proportion of voters that support the referendum on the school budget.

(assumed to be 0.52)

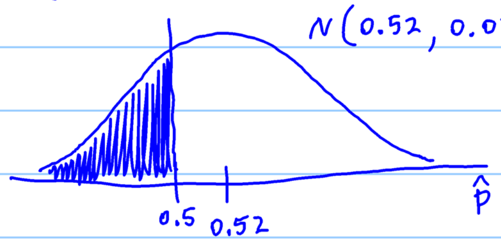
$p = 0.52$ $\quad$ $n = 400$

$q = 0.48$

↳ this means that $\hat{p} < 0.50$!

$P(\hat{p} < 0.5)$?

$N(0.52, 0.025)$



$$z = \frac{\text{obs} - \text{mean}}{SD}$$

$$= \frac{\hat{p} - p}{\sigma_{\hat{p}}} = \frac{0.5 - 0.52}{0.025}$$

$$z = -0.8$$

$$P(\hat{p} < 0.5) = P(z < -0.8)$$

☆ Don't do "1 - minus"!

$$= \boxed{0.2119}$$

There is roughly a 21.2% probability that less than half of the newspaper's sample will approve of the referendum - thus leading the newspaper to predict defeat.

Conditions / Assumptions:

- Randomization: We must assume that the 400 polled voters were selected randomly. This should be reasonable.
- 10% condition: We assume that there are at least 10×400 (or 4000) eligible voters. This is reasonable.
- Success/failure:

$np = 400(0.52) = 208 \geq 10$ ✓

$nq = 400(0.48) = 192 \geq 10$ ✓

The conditions are satisfied, so we may use the Normal model to approximate the sampling distribution for $\hat{p}$.

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

Standard Deviation: $\sigma_{\hat{p}} = \sqrt{\frac{p \cdot q}{n}}$

$$= \sqrt{\frac{(0.52)(0.48)}{400}}$$

$SD(\hat{p}) = \sigma_{\hat{p}} = 0.02498$

≈ 0.025

$$N(0.52, 0.025)$$