AP Statistics

Section 9.2: Sample Proportions

The objective of some statistical applications is to reach a conclusion about a population proportion, p. For example, we may try to estimate an approval rating through a survey, or test a claim about the proportion of defective light bulbs in a shipment based on a random sample. Since p is unknown to us we must base our conclusion on a sample proportion, $\hat{p}$. We have seen that the value of $\hat{p}$ will vary from sample to sample. The amount of variability will depend on the size of our sample.

If we take a simple random sample (SRS) from the population of interest, our estimator is the sample proportion of successes defined as

$$\hat{p} = \frac{\text{count of "successes" in sample}}{\text{size of sample}} = \frac{X}{n}.$$

The count X will follow a binomial distribution; the proportion $\hat{p}$ does *not* have a binomial distribution. Using properties of linear combinations of random variables, however, we can demonstrate the following relationships for the proportion $\hat{p}$.

The Sampling Distribution for Proportions

If we take repeated random samples of size n from a population, the sample proportion $\hat{p}$ will have the following distribution and properties:

population proportion

- The ***mean*** of the sampling distribution is exactly p.
- The ***standard deviation*** of the sampling distribution is $\sqrt{\frac{p(1-p)}{n}}$.
- Rule of Thumb 1: We will use this formula for the standard deviation of $\hat{p}$ only when the population is at least 10 times as large as the sample.
- Rule of Thumb 2: We will use the ***normal*** approximation to the sampling distribution of $\hat{p}$ for values of n and p that satisfy $np \geq 10$ and $n(1-p) \geq 10$.

Because the mean of the sampling distribution of $\hat{p}$ is always equal to the parameter p, the sample proportion $\hat{p}$ is an unbiased estimator of p.

* The standard deviation of $\hat{p}$ gets smaller as the sample size n increases. Thus, $\hat{p}$ is less variable in larger samples.

population proportion

Example:

Based on census data, we know that 11% of US adults are black. Therefore, $p = 0.11$. Further, we would expect a sample to contain roughly 11% black representation. Suppose a sample of 1500 adults contains 138 black individuals. Should we suspect "undercoverage" in the sampling method?

n X

Here $\hat{p} = 138/1500 = 0.092$. Is this lower than what would be expected by chance? That is, we know it is possible that a sample could contain 9.2% black representation, but is it likely that would happen due to natural variation in a random sampling method?

- Check assumptions

~ Population of US adults $\geq 10(1500)$ ✓
$\geq 15{,}000$

~ $np = 1500(0.11) = 165 \geq 10$ ✓
$n(1-p) = 1500(0.89) = 1335 \geq 10$ ✓

normal approx. is okay

- Describe the sampling distribution

$N(\mu, \sigma)$ $N(0.11, 0.008)$

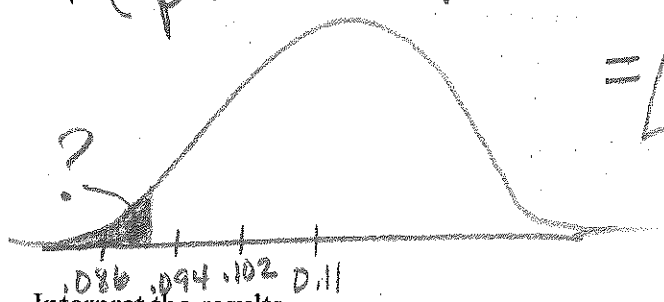
$\mu_{\hat{p}} = p = 0.11$

$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.11(0.89)}{1500}} \approx 0.008$

- Calculate the probability

$P(\hat{p} \leq 0.092) = P(z \leq -2.25)$ $\quad z = \frac{0.092 - 0.11}{0.008} = -2.25$

$= \boxed{0.0122}$



- Interpret the results

Only 1.22% of samples of size 1500 would have 9.2% or less black representation. Since this is so unlikely, we have reason to suspect possible undercoverage in the sample.

Example:

The development of viral hepatitis subsequent to a blood transfusion can cause serious complications for a patient. The article "Hepatitis in Patients with Acute Non-lymphatic Leukemia" (*American Journal of Medicine*, 1983: 413-421) reported that in spite of careful screening for those having a hepatitis antigen, viral hepatitis occurs in 7% of blood recipients. Suppose a new treatment is believed to reduce the incidence of viral hepatitis. The treatment is given to 200 blood recipients and only 6 of them contract hepatitis. Does it appear that the treatment is effective? That is, is it very likely that we would observe only 6/200 recipients contract hepatitis when 7% of the population is known to do so?

$p = 0.07$

n

$X = 6$

$X/n = \hat{p} = 0.03$

- Check assumptions

~ population of blood recipients $\geq 10(200)$
$\geq 2{,}000$ ✓

~ $np = 200(0.07) = 14 \geq 10$ ✓
$n(1-p) = 200(0.93) = 186 \geq 10$ ✓

- Describe the sampling distribution

$N(0.07, 0.018)$

$\mu_{\hat{p}} = p = 0.07$

$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{.07(0.93)}{200}} \approx 0.018$

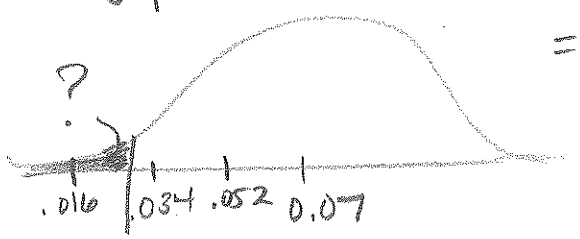
- Calculate the probability

$P(\hat{p} \leq 0.03) = P(z \leq -2.22)$

$= \boxed{0.0132}$

$z = \frac{0.03 - 0.07}{0.018}$

$= -2.22$



- Interpret the results

Only 1.32% of samples of size 200 would have 6 or less recipients contract hepatitis.

Assignment: Read p. 504-511
p. 511 **#9.19, 9.20, 9.21**