7.2 Lecture Notes & Examples

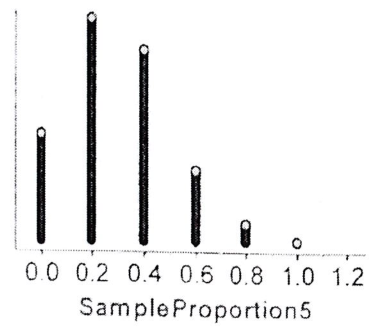
Section 7.2 - Sample Proportions (pp. 432-439)

1. The Sampling Distribution of $\hat{p}$

Let's turn once again to the hyena experiment on the first day of the course. Suppose a team took one sample and found the proportion of males to be $\hat{p}$ =0.20. Since another random sample would likely result in a different estimate, we can only say that "about" 20% of the population of hyenas in the Croatan NF are males. In this section, we are going to use sampling distributions to clarify what "about" means.

Activity

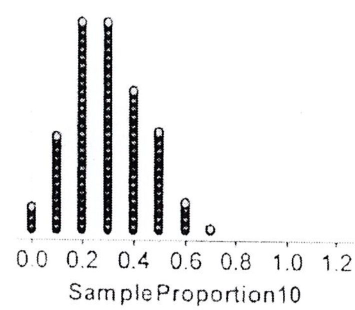
Suppose a team performed the hyena experiment again. First they chose repeated samples of size 5. The distribution of sample proportions is shown at the right.



Describe the distribution:

skewed right
center ≈ 0.3
Range from 0 to 1
No Outliers

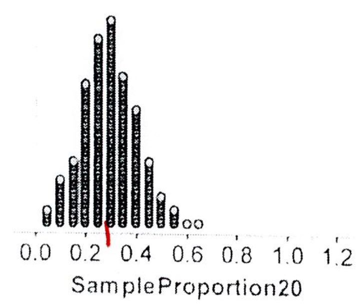
The team then took repeated samples of size 10. The distribution of sample proportions is shown at the right.



Describe the distribution:

Roughly symmetric
Center ≈ 0.25
Range: 0 to 0.7
No Outliers

Finally, the team took repeated samples of size 20. The distribution of sample proportions is shown at the right.



Describe the distribution:

Roughly Symmetric
Center ≈ 0.25
Range from 0 to 0.65
No Outliers

Summarize what happened to the center, shape and spread as the sample size was increased from 5 to 20.

As sample size increased from 5 to 20, centered stayed about the same, the shape became more symmetric, the spread became less (less variability)

Binomial Distribution - Is the hyena experiment binomial? Let X = the number of males obtained in each sample. Is X a binomial random variable?

B: Yes s = male f = female

I: No, not replacing.

N: Fixed n? Yes, sample size was constant

S: prob success fixed? no since not replacing

* However, If population size is large we still can be close to a binomial distribution (10% condition)

This means that $\hat{p} = \dfrac{\text{count of \# of successes}}{\text{sample size}} = \dfrac{X}{n}$

From Chapter 6, we know that the mean and standard deviation of a binomial random variable X are:

$\mu_X = np$ and $\sigma_X = \sqrt{np(1-p)}$ or $\sqrt{npq}$ where $q = (1-p)$ prob of failure

Since $\hat{p} = \frac{X}{n} = \frac{1}{n}(X)$, we are just multiplying a random variable by a constant $(\frac{1}{n})$: $\mu_{\hat{p}} = \frac{1}{n}\mu_X = \frac{1}{n}(np) = p \Rightarrow \boxed{\mu_{\hat{p}} = p}$

sample proportion

(This means $\hat{p}$ is an unbiased estimator of p)

Spread: $\sigma_{\hat{p}} = \frac{1}{n}\sqrt{np(1-p)} = \sqrt{\frac{np(1-p)}{n^2}} = \sqrt{\frac{p(1-p)}{n}}$

* This means as $n \uparrow$, $\sigma_{\hat{p}} \downarrow$

sample size increases, spread decreases

Shape: Multiplying by a constant does not change shape

Sampling Distribution of a Sample Proportion

Choose an SRS of size n from a population of size N with proportion p of successes. Let $\hat{p}$ be the sample proportion of successes. Then:

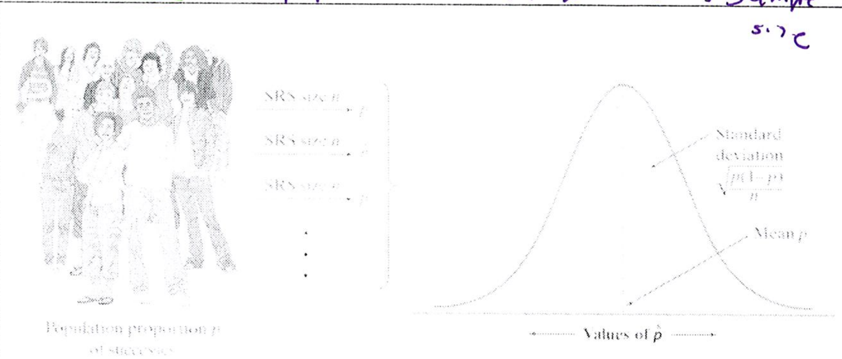
AP Formula Sheet

- The **mean** of the sampling distribution of $\hat{p}$ is p. $\mu_{\hat{p}} = p$
- The **standard deviation** of the sampling distribution of $\hat{p}$ is $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$

$n \leq \frac{1}{10}(N)$ or $10(n) \leq N$

** as long as the *10% condition* is satisfied: $n \leq (1/10)N$. ** population is at least 10 times sample size

- As n increases, the sampling distribution of $\hat{p}$ becomes **approximately Normal**.
  - Before you perform Normal calculations, check that the *Normal condition* is satisfied: * $np \geq 10$ and $n(1-p) \geq 10$. *



Check Your Understanding - About 75% of young adult internet users (ages 18-29) watch online video. Suppose that a sample survey contacts an SRS of 1000 young adult internet users and calculates the proportion $\hat{p}$ in this sample who watch online video. $n = 1000$

(a) What is the mean of the sampling distribution of $\hat{p}$? Explain.

Since $p = .75$, then $\mu_{\hat{p}} = 0.75$

(b) Find the standard deviation of the sampling distribution of $\hat{p}$. Check that the 10% condition is met. $n \leq \frac{1}{10}N$ or $10n \leq N$

✓ We can assume more than $10(1{,}000) = 10{,}000$ internet users.

$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{.75(.25)}{1000}} \approx \boxed{.0137}$

(c) Is the sampling distribution of $\hat{p}$ approximately Normal? Check that the Normal condition is met.

✓ $np \geq 10$ $1000(.75) \geq 10$ $750 \geq 10$ ✓

✓ $n(1-p) \geq 10$ $1000(.25) \geq 10$ $250 \geq 10$ ✓

Yes, since both $np \geq 10$ and $n(1-p) \geq 10$.

(d) If the sample size were 9000 instead of 1000, how would this change the sampling distribution of $\hat{p}$?

$n = 9000$ It would still be approximately normal with $\mu_{\hat{p}} = 0.75$ but $\sigma_{\hat{p}} = \sqrt{\frac{(.75)(.25)}{9000}} \approx .0046 \Rightarrow$ less variability

2. Using the Normal Approximation of $\hat{p}$

Example - The superintendent of a large school district wants to know what proportion of middle school students in her district are planning on attending a four-year college or university. Suppose that 80% of all middle school students in her district are planning to attend a four-year college or university. What is the probability that an SRS of size 125 will give a result within 7 percentage points of the true value?

① State: We want to find the prob. that the % of middle school students who plan to attend a 4 year college or university falls between 73% and 87%.
$P(.73 < \hat{p} < .87)$

② Plan: Find $\mu_{\hat{p}}$ and $\sigma_{\hat{p}}$. For $\sigma_{\hat{p}}$ must ✓ 10% condition
$\mu_{\hat{p}} = .80$

$10(n) \le$ Population

It is reasonable to assume $10(125) \le$ Population ✓

$1{,}250 \le$ population

so $\sigma_{\hat{p}} = \sqrt{\dfrac{.80(.20)}{125}} = .036$

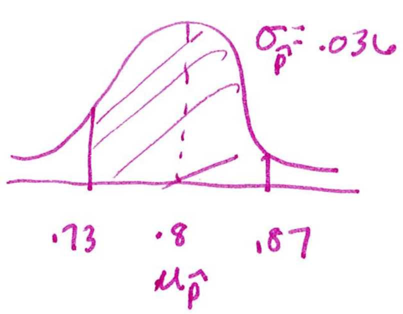
Now ✓ if normal conditions are met.
$np \ge 10$ and $n(1-p) \ge 10$

$n = 125$
$p = .8$

∴ $125(.8) \ge 10$ and $125(.20) \ge 10$
$100 \ge 10$ ✓ $25 \ge 10$ ✓

We can consider the distribution of $\hat{p}$ to be ≈ Normal
$N(.8, .036)$

③ DO: $P(0.73 \le \hat{p} \le .87)$



normalcdf(.73, .87, .8, .036) ≈ .948 ≈ 95%

④ Conclude: About 95% of all SRS's of size 125 will give a sample proportion within 7 points of true proportion.