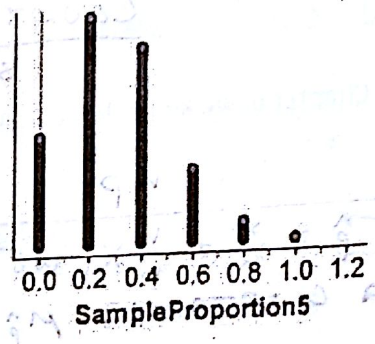
Section 7.2 - Sample Proportions (pp. 440-449)

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.

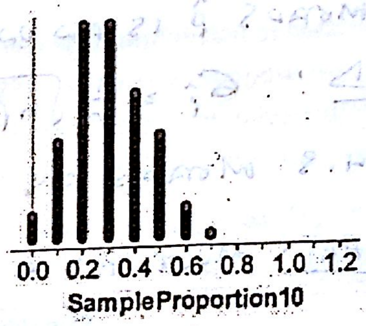
Activity

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.



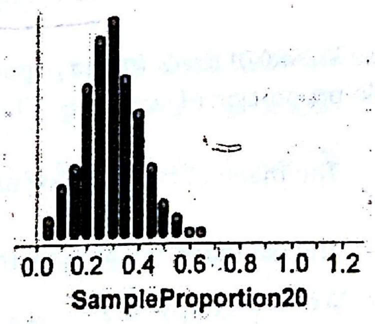
Describe the distribution: SKEWED RIGHT
CENTER ≈ 0.3
RG FROM 0 TO 1
NO OUTLIERS

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



Describe the distribution: R. SYMMETRIC
CENTER ≈ 0.25
RG FROM 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: R. SYMMETRIC
CENTER ≈ 0.25
RG 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.

CENTER STAYED ABOUT THE SAME.
SHAPE BECAME MORE SYMMETRIC.
LESS SPREAD (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: BINARY? YES, MALE + FEMALE

I: INDEPENDENT? YES, RANDOM SAMPLING

N: FIXED n? YES, SAMPLE SIZE WAS CONSTANT

S: SUCCESS PROB FIXED? NO, SINCE W/O REPLACEMENT,

* HOWEVER, IF POP. SIZE IS LARGE WE WILL BE CLOSE TO BINOMIAL (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 $\sigma_X = \sqrt{np(1-p)}$

SINCE $\hat{p} = \frac{X}{n} = \frac{1}{n}X$, WE ARE JUST MULTIPLYING A RANDOM VARIABLE BY A CONSTANT: $\mu_{\hat{p}} = \frac{1}{n}\mu_X = \frac{1}{n}(np) = p \Rightarrow \therefore \boxed{\mu_{\hat{p}} = p}$

(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 $n \uparrow$, $\sigma_{\hat{p}} \downarrow$.

SHAPE: MULT. BY A CONSTANT DOES NOT Δ SHAPE.

Sampling Distribution of a Sample Proportion

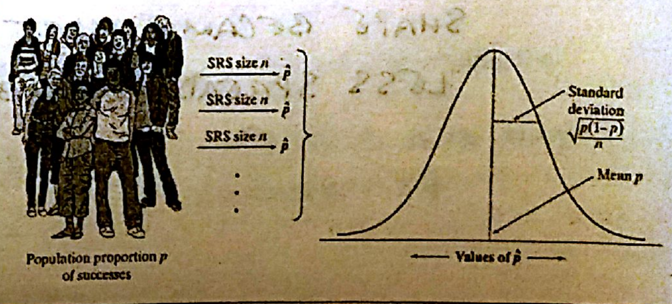
AP FORMULA SHEET

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:

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

as long as the *10% condition* is satisfied: $n \leq (1/10)N$.

- 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.

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

SINCE $p = 0.75$, THEN $\mu_{\hat{p}} = p = \boxed{0.75}$

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

WE CAN ASSUME MORE THAN 10(1000) = 10,000 INTERNET USERS.

$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.75(0.25)}{1000}} \approx \boxed{0.0137}$

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

YES, SINCE $np = (0.75)(1000) = 750$ AND $n(1-p) = (0.25)(1000) = 250$ BOTH > 10.

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

IT WOULD STILL BE APPROX. NORMAL WITH $\mu_{\hat{p}} = 0.75$ BUT

$\sigma_{\hat{p}} = \sqrt{\frac{(0.75)(0.25)}{9000}} = \boxed{0.0046} \Rightarrow$ 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?

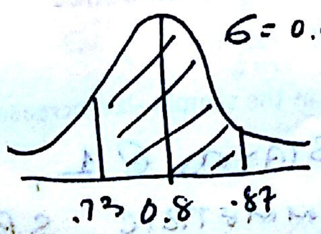
① STATE: WE WANT TO FIND PROB. THAT THE % OF M/S STUDENTS WHO PLAN TO ATTEND A 4-YR COLLEGE/UNIV. FALLS BETWEEN 73% AND 87%. $P(0.73 < \hat{p} < 0.87)$

② PLAN: $\mu_{\hat{p}} = 0.8$ SINCE $p = 0.8$. SINCE DISTRICT IS LARGE WE CAN ASSUME THAT THERE ARE MORE THAN 10(125) = 1250 M/S STUDENTS. (INDEPENDENCE). $\Rightarrow$

$\sigma_{\hat{p}} = \sqrt{\frac{(0.8)(0.2)}{125}} \approx 0.036$

SINCE $np = 125(0.8) = 100 > 10$ AND $125(0.2) = 25 > 10$, WE CONSIDER $\hat{p}$ TO BE APPROX. NORMAL $(0.8, 0.036)$.

③ DO:



$\sigma = 0.036$

$P(0.73 \le \hat{p} \le 0.87) = 0.948$

④ CONCLUDE: ABOUT 95% OF ALL SRS'S OF SIZE 125 WILL

HW: read pp. 440-447; do problems: p. 439 - 21-24; pp. 447 - 27, 29, 33, 35, 37, 41, 47*, 48*.

GIVE A SAMPLE PROPORTION W/IN 7 PTS OF TRUE PROPORTION.

Scanned by CamScanner