<u>8.2: Distribution of the Sample Proportion</u>

<u>Recall</u>: A *parameter* is a numerical summary of a population; a *statistic* is a numerical summary of a sample. (For example, the population mean and population standard deviation are parameters; the sample mean and sample standard deviation are statistics.)

The *sampling distribution of a statistic* is the probability distribution of all possible values for that statistic computed from all possible samples of fixed size n.

The sampling distribution of the sample proportion:

In this section, the parameter we are interested in is the *population proportion*:

The proportion is the percentage p (in decimal form) of the population that possesses some characteristic of interest.

For example, we may be interested in the proportion of children who have a certain medical condition, the proportion of U.S. citizens who received a tax refund, the proportion of students at a certain high school that decide to go to college, or the proportion of nurse candidates who pass the nursing licensure exam.

The *sampling distribution of the sample proportion* is the probability distribution of all possible values for the sample proportion, denoted $\hat{p}$, computed from all possible samples of fixed size n.

If x is the number of data points in a sample of size n that have the characteristic of interest, then the sample proportion is

$$\hat{p} = \frac{x}{n}.$$

In the same manner as for the sample mean, we use the sample proportion $\hat{p}$ to make inferences about the population proportion p.

Shape, mean and standard deviation of the sampling distribution of the sample proportion:

<u>Sampling distribution of the sample proportion</u>:

Suppose random samples of size *n* are taken from a population with population proportion *p*.

Also suppose that the sample size is small compared to the size of the population.
(Rule of thumb: The sample must be less than 5% of the population size; otherwise we must use a finite population correction factor, which is beyond the scope of this class.)

Then:

The shape of the sampling distribution of $\hat{p}$ is approximately normal, provided that $n\cancel{\hat{p}}(1-p) \geq 10$. $np(1-p) \geqslant 10$

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

Calculating the probability of a sample's results: $\sigma_{\hat{p}} = \sqrt{\dfrac{p(1-p)}{n}}$

Recall: $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$

<u>Example 1:</u> In the U.S., approximately 12.3% of adults have diabetes.
http://www.cdc.gov/diabetes/pubs/statsreport14/national-diabetes-report-web.pdf

In a random sample of 100 Americans, find the probability that:

a) At least 10% of those in the sample have diabetes.
b) At least 20% of those in the sample have diabetes.

(a) population proportion: $p = 0.123$

sample proportion: $\hat{p} = 0.10$ ← we're really interested in $P(\hat{p} \geqslant 0.10)$

$n = 100$

calculate the mean and standard deviation (standard error) of the sampling distribution of the sample proportions

$\mu_{\hat{p}} = p = 0.123$

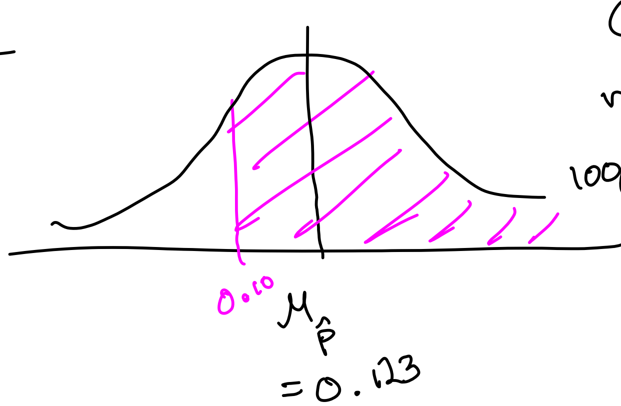
$\sigma_{\hat{p}} = \sqrt{\dfrac{p(1-p)}{n}} = \sqrt{\dfrac{0.123(1-0.123)}{100}} = \sqrt{\dfrac{0.123(0.877)}{100}}$

(standard error)

≈ 0.03284

Ex 1 cont'd

Calculate z-score for 0.10.



Check that
$np(1-p) \geq 10$
$100(0.123)(0.877)$
$= 10.7871$
≥ 10

(so the sample distribution is normal)

we can use z-table

$$Z = \frac{\hat{p} - \mu_{\hat{p}}}{\sigma_{\hat{p}}}$$

$$= \frac{0.10 - 0.123}{0.0328437}$$

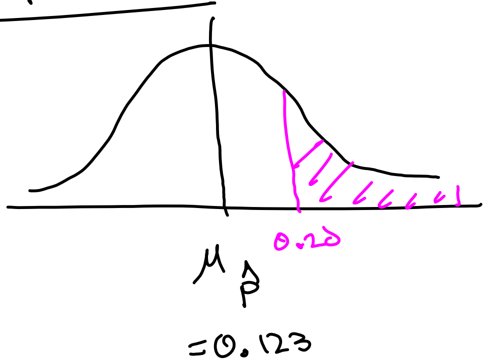
$$\approx -0.700286 \approx -0.70$$

$$P(\hat{p} \geq 0.10) = 0.5 + 0.2580$$

$$= \boxed{0.7580}$$

← from z-table, $P(0 < Z < 0.70) = 0.2580$

ⓑ Find $P(\hat{p} \geq 0.20)$.



Calculate the z-score for $\hat{p} = 0.20$

$$Z = \frac{\hat{p} - \mu_{\hat{p}}}{\sigma_{\hat{p}}}$$

$$= \frac{0.20 - 0.123}{0.032844}$$

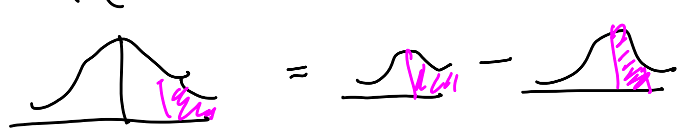
$$\approx 2.34444 \approx 2.34$$

Look this up in z-table:

$$P(0 < Z < 2.34) = 0.4904$$

$$P(\hat{p} > 0.20) = 0.5 - 0.4904$$

$$= \boxed{0.0096}$$



Example 2: Women comprise about 61% of the LSC-North Harris student body. In a random sample of 55 LSC-North Harris students, compute the probability that:

$n = 55$

population proportion: $p = 0.61$ (proportion of women)

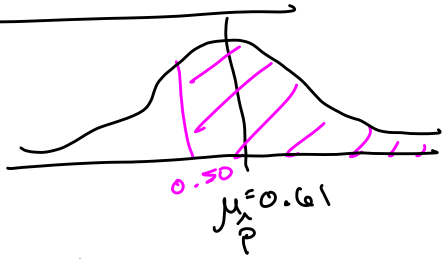
a) At least 50% of the students in the sample are women.
b) At least 60% of the students in the sample are men.
c) At least 40 of the students in the sample are women.
d) At least 45 of the students in the sample are women.
e) Would any of these results be considered unusual?

Check for normality using rule of thumb

$np(1-p) = 55(0.61)(0.39) = 13.08 \geq 10$

OK to use normal table

a) Find $P(\hat{p} \geq 0.50)$



Find standard error:

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.61(0.39)}{55}} \approx 0.065768$$

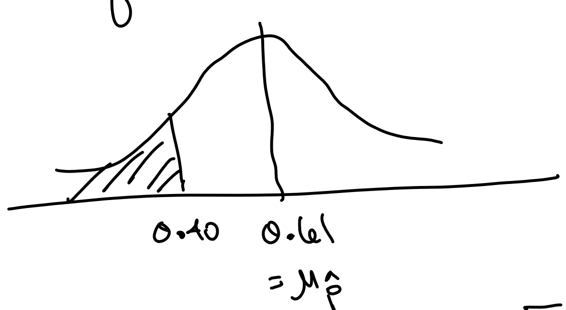
Find z-score for $\hat{p} = 0.50$:

$$z = \frac{\hat{p} - \mu_{\hat{p}}}{\sigma_{\hat{p}}} = \frac{0.50 - 0.61}{0.065768} \approx -1.67$$

From z-table, $P(0 < z < 1.67) = 0.4525$

$$P(\hat{p} \geq 0.50) = 0.5 + 0.4525 = \boxed{0.9525}$$

b) Find prob. that at least 60% are men.
This is equivalent to: 40% or less are women



Calculate z-score for $\hat{p} = 0.40$:

$$z = \frac{0.40 - 0.61}{0.065768} = -3.19$$

From normal table, $P(0 < z < 3.19) = 0.4993$

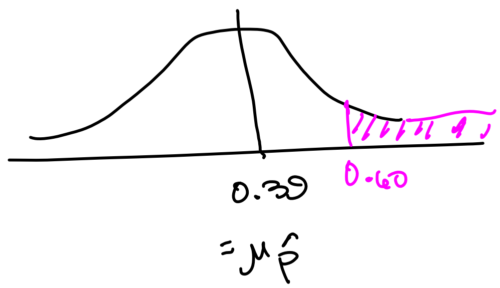
$$P(\hat{p} \leq 0.40) = 0.5 - 0.4993 = \boxed{0.0007}$$

Alternatively: revise our population proportion to represent the population of men:

now $p = 0.39$

standard error: $\sigma_{\hat{p}} = \sqrt{\dfrac{p(1-p)}{n}} = \sqrt{\dfrac{0.39(0.61)}{55}}$

≈ 0.065768



Find z-score for $\hat{p} = 0.60$

$$z = \frac{\hat{p} - \mu_{\hat{p}}}{\sigma_{\hat{p}}} = \frac{0.60 - 0.39}{0.065768}$$

≈ 3.19

From normal table,

$P(0 < z < 3.19) = 0.4993$

$P(\hat{p} \geq 0.60)$

$= 0.5 - 0.4993$

$= \boxed{0.0007}$

(c) Find prob. that sample contains at least 40 women.

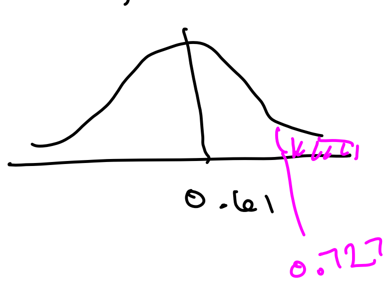
For 40 women,

$\hat{p} = \dfrac{40}{55} = 0.72727 \Rightarrow 72.727\%$ are women (of sample)

Find $P(\hat{p} \geq 0.72727)$

$$z = \frac{0.72727 - 0.61}{0.065768} \approx 1.78$$

From normal table, $P(0 < z < 1.78) = 0.4625$



$P(\hat{p} \geq 0.727) = 0.5 - 0.4625$

$= \boxed{0.0375}$

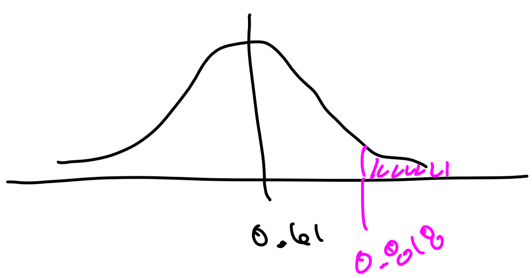
(d) at least 45 of those in the sample are women.

$$\hat{p} = \frac{45}{55} = 0.81818 \Rightarrow 81.818\% \text{ are women (in sample)}$$

Calculate z-score for $\hat{p} = 0.81818$

$$z = \frac{\hat{p} - \mu_{\hat{p}}}{\sigma_{\hat{p}}} = \frac{0.81818 - 0.61}{0.065768} \approx 3.165 \approx 3.17$$

From normal table, $P(0 < z < 3.17) = 0.4992$



$$P(\hat{p} \geq 0.81818) = 0.5 - 0.4992 = 8 \times 10^{-4} = \boxed{0.0008}$$

(e) The situations of part (b) and part (d) (at least 60% men and at least 45 women) are considered unusual because they are more than 2 standard deviations away from $p = \mu_{\hat{p}}$.

(Their z-scores are more extreme than $z = 2$ or $z = -2$).