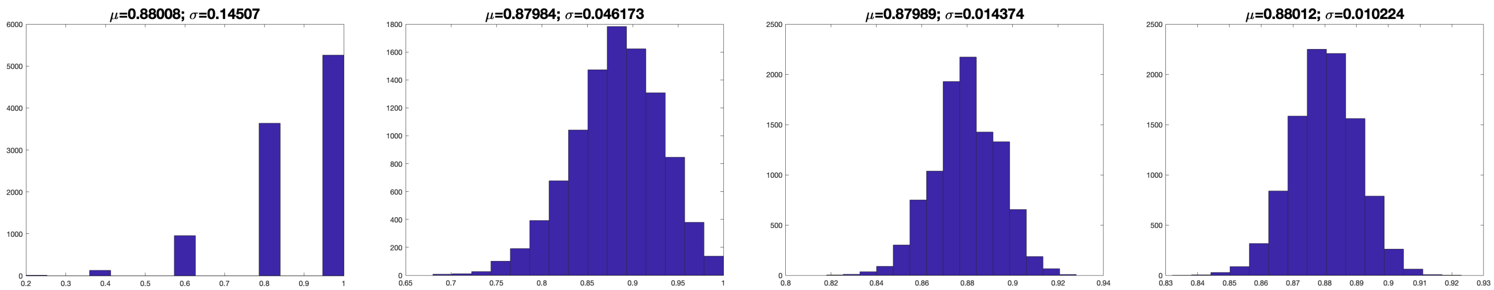
MAT1372, Classwork16, Fall2025

5.1 Point Estimates and Sampling Variability(Conti.)

8. What happens if the success-failure condition isn't satisfied?

If we do the simulations by executing the code when $p = 0.88$ with different sample size n:



$n = 5$	$n = 50$	$n = 500$	$n = 1000$
$np =$ 4.4 <10	$np =$ 44 ≥ 10	$np =$ 440 ≥ 10	$np =$ 880 ≥ 10
$n(1-p) =$ 0.6 <10	$n(1-p) =$ 6 <10	$n(1-p) =$ 60 ≥ 10	$n(1-p) =$ 120 ≥ 10

9. What do you observe the trend when n varies?

For the skewness and discreteness of the distributions, we have

(1) When either np or n(1 − p) is small, the distribution is more discrete (or not continuous)

(2) When np or n(1 − p) is smaller than 10, the skew in the distribution is more noteworthy

(3) The larger both np and n(1−p), the more normal the distribution. This may be a little harder to see for larger sample size

(4) When np and n(1 − p) are both very large, the distribution's discreteness is hardly evident, and it looks much more like a normal distribution

For the mean and standard error of the distributions, we have

(1) The centers of the distribution are always at the population proportion p. Because the sampling distribution of $\hat{p}$ is always centered at the population proportion p, it means $\hat{p}$ is unbias when the data are independent and drawn on a larger popultion

(2) For a particular population p, the variability in the sampling distribution decreases as the sample size becomes larger

5.2 Confidence Intervals for a Proportion

1. Confidence Interval and Confidence Level.

A confidence Interval (CI) is a plausible range of values used to estimate an unknown statistical parameter, such as population proportion. Rather than reporting a single estimate ("average screen time is 3hrs"), a confidence interval provides a range ("2 to 4 hrs"), along with a specified confidence level. If we report a point estimate $\hat{p}$, we probably will not hit the exact p. On the other hand, representing a CI, we have a good shot at capturing the p

2. Constructing a 95% Confidence Level.

The Center of the interval: sample proportion $\hat{p}$ is the most plausible value of p, so it makes sense to build a CI around $\hat{p}$

The size of the interval: The standard Error (SE) provides a guide for how large we should make the confidence Level

Build the interval: We can construct a CI that extends 1.96 standard error from $\hat{p}$ to be 95% confident that the interval captures p:

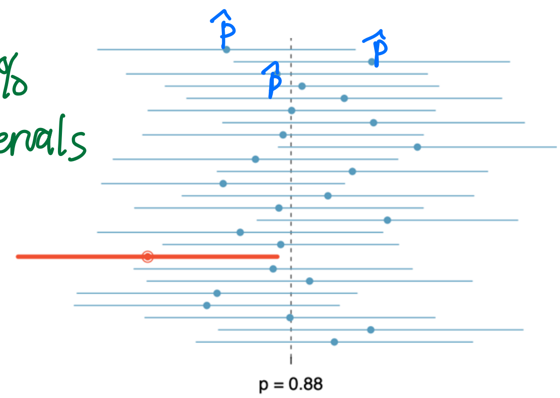
$$\text{point estimate} \pm 1.96\times SE = \hat{p} \pm 1.96\sqrt{\frac{p(1-p)}{n}}$$

Why 1.96×SE? The SE represents standard deviation for proportion $\hat{p}$, and when Central Limit Thm conditions are satisfied, then $\hat{p}$ closely follows normal distribution, 95% of the data is within 1.96 SE of $\hat{p}$

3. What does "95% confident" mean?

Suppose we took many samples and bulit a 95% CI from each. Then about 95% of those intervals would contain p.

Figure shows the process of creating 25 intervals from 25 samples from the simulation, only 24 of them contain p.



4. In Section 5.1 we learned that about 88.7% of a random sample of 1000 American adults supported solar power. Compute and interpret a 95% confidence interval for the population proportion.

$\hat{p} = 0.887$ $\quad SE_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.887\cdot 0.113}{1000}} = 0.01$

(since $\hat{p}$ follows normal distribution by Central limit Thm)

95% CI: $\hat{p} \pm 1.96\times SE_{\hat{p}}$ $\quad -1.96\times SE + 0.887 \quad \overset{\hat{p}}{0.887} \quad 1.96\times SE + 0.887$

(0.8674, 0.9066)