MAT1372, Classwork15, Fall2025

5.1 Point Estimates and Sampling Variability

1. Point estimates. Point estimate involves the use of sample data to calculate a single value which is to serve as a "best guess" or "best estimate" of an unknown population parameter

2. Error: sampling error and bias.

   Sampling error: It's also called sampling uncertainty and describes how much an estimate will tend to vary from one sample to the next.

   Bias: It describes a systematic tendcy to over- or under-estimate the true population value (ch 1)

3. Example of the variability of a point estimate.

   Suppose the proportion of American adults who support the expansion of solar energy is $p = 0.88$, which is our parameter of interest. How does the sample proportion $\hat{p}$ behave when the true population proportion is 0.88 (which we are **Not** supposed to know)?

   Here's how we might go about constructing such a *simulation*:

   (1) There were about 250 million American adults in 2018. On 250 million cards, write "support" on 88% of them and "not" on 12% of them.

   (2) Mix up the card and pull out 1000 cards to represent our sample of 1000 adults

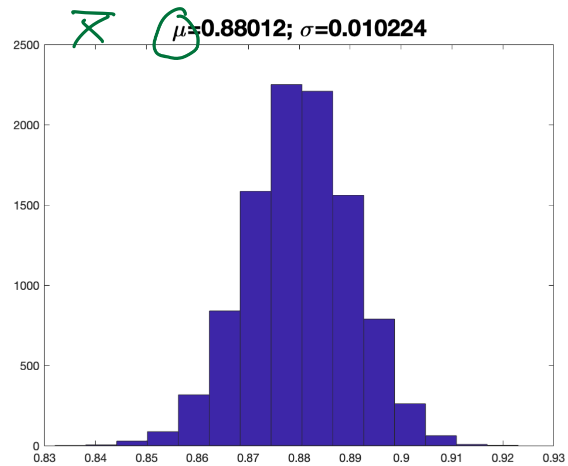
   (3) Compute the fraction of the sample that say "support".

   (4) Repeat (2) and (3) many, many times.

250×10^6

```
population = 250e6; n = 1e3; % sample size
num_simulation = 10000;%number of simulation
random_array = randperm(population);
mean_simulation =[];
for i=1: num_simulation
        x1=random_array(randi([1, population], n, 1));
        sample=[x1<=0.88*population*ones(size(x1))];
        mean = sum(sample)/n;
        mean_simulation =[mean_simulation mean];
end
hist(mean_simulation,30);
```

$\bar{X}$



This code gives us a distribution of sample proportion which is called a sampling distribution:

Center. The center of this distribution is $\bar{x}_{\hat{p}} = 0.88012$, which is the same as the parameter of interest. (This simulation is a simple random sample)

Spread. The standard deviation (SD) is $S_{\hat{p}} = 0.010224$. For a sample distribution, we typically use the term standard error, denoted by $SE_{\hat{p}}$

Shape. It's symmetric and bell-shape, and it resembles a normal distribution.

4. **Central Limit Theorem and the Success-Failure Condition**

When observations are independent and the sample size is sufficiently large, the sample proportion $\hat{p}$ will tend to follow a normal distribution with the following:

mean $\mu_{\hat{p}} = p$, and Standard Error $SE_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$

In order for the Central Limit Theorem to hold, the sample size (n) is typically considered sufficiently large when $np \geq 10$ and $n(1-p) \geq 10$, which is called the success-failure condition.

5. In 3., we estimated the mean and standard error of $\hat{p}$ using simulated data when $p = 0.88$ and $n = 1000$.

Confirm that the Central Limit Theorem applies and the sampling distribution is approximately normal.

① **Independence**. The poll is a simple random sample of American adults, which means that the observations are independent.

② **Success-failure condition**. We can confirm the size is sufficiently large by checking the success-failure condition and comfirming the two calculated value:

$np = 1000 \cdot 0.88 = 880 \geq 10$. $n(1-p) = 1000 \cdot 0.12 = 120 \geq 10$

Based on ① ②, the Central Limit Thm applies, it's reasonable to model this using a normal distribution

6. Applying the Central Limit Theorem to a real-world setting.

In the real setting, we could NOT know what the population proportion p is for supporting solar energy. The thing we can do is a poll of 1000 people which gives us the sample proportion $\hat{p}$. Assume $\hat{p} = 0.887$.

Does the sample proportion from the poll approximately follow a normal distribution?

We can check the conditions from the Central Limit Theorem.

Independence. Yes, since it is a simple random sample

Success-failure condition. To check this condition, we need p (population proportion) to check $np \geq 10$, $n(1-p) \geq 10$. However, we do not know p. Thus, we use the $\hat{p}$ (sample proportion) to check success-failure condition: $n\hat{p} = 1000 \cdot 0.887 = 887 \geq 10$, $n \cdot (1-\hat{p}) = 113 \geq 10$. $\hat{p}$ acts as a reasonable substitue for p during this check.

7. Substitution Approximation of using $\hat{p}$. The substitution approximation of using $\hat{p}$ in place of p is also useful when computing the standard error of the sample distribution

$$SE_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} \approx \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.887 \cdot 0.113}{1000}} = 0.010\cdots$$

"plug-in principle". In this case, $SE_{\hat{p}}$ didn't change enough to be detected using only 3 decimal places versus the $SE_{\hat{p}} = 0.01~~0224~~$