STA 103 Lecture 9: Estimation

Instructor: Wookyeong Song

Department of Statistics, University of California, Davis

Aug 25th, 2025

Announcement

WS **Wookyeong Song** AUTHOR | TEACHER

Created Aug 24 11:46pm | Posted Aug 24 11:46pm

Midterm 2 Format Announcement

Hi all,

The midterm 2 will be held next Tue (Aug 26th) from 12:10 pm - 1:00 pm (50 minutes). Below is the general information about the Midterm 2 format.

There will be **four Big problems.** We have a total of 10 subproblems, i.e., 1(a), 1(b), 1(c), 2(a), 2(b), 3(a), 3(b), 3(c), 4(a), 4(b). The subproblems within the same main problem **may or may not** build on each other (e.g., (a) or (b) might or might not be used to solve (c)), but each main problem (1, 2, 3, 4) is **independent** from the others.

Details:

- Midterm 2 will cover **Lecture 5 - 8**.
- Each subproblems will be similar to one of (lecture note examples, assignment 2, and problem set 2).
- You will not encounter completely unfamiliar topics on the exam.
- Each question will have a clear, well-defined answer. Some questions may ask you to justify your answer based on your results.
- **Showing your work is essential**: grader will check your derivations and may award partial credit even if your final answer is not correct.
- Likewise, you may not receive full credit if you just drop the answer without any derivation process.
- A printed **formula sheet** and **Z-table** (with everything you need) will be provided at the start of the exam—**no need to print** your own.
- Please bring your **own calculator**.
- Please bring your **student ID**.

Tips for preparation:

I recommend mastering first both the material and the worked examples on the **lecture notes**. This will give you the solid foundation you need for the exams. But, to complete **10** subproblems in 50 minutes, you'll need not only a clear understanding of each concept but also enough practice to solve each problem in **about 5 minutes**. To build that speed, working through Assignment 2 and Problem Set 2 will be especially helpful.

This week: Office hour is today (8/25) 2:00 – 3:00 pm, and there will be no office hour on Thursday (8/28).

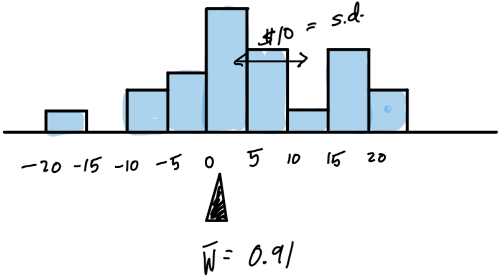
Types of Statistical Analysis (Lecture 1)

- Descriptive Statistics: Summarize data using graphs, charts, averages, and percentages. (Midterm 1, Midterm 2)
  - Example: A marketing team reports the average spending per customer during a promotional event.
- Inferential Statistics: Make generalizations about a population based on sample data. Estimation, confidence interval, hypothesis testing. (After Midterm 2)
  - A retailer surveys 200 shoppers to estimate the satisfaction level of all store visitors. Find 95% confidence interval of the satisfaction level given 200 samples.
  - For A/B test, is new product B statistically better than old product A? (Youtube 2 minutes 1 ad vs 1 minute 2 ads).

Motivation

Start with a motivating example.

- **Set up**: You are watching someone repeatedly play a slot machine in Las Vegas.
- They played n times and left. Let RV W be the earnings from each play. Let $W_1, \ldots, W_n$ be the earnings of ith play.
- You recorded the winnings of each play. Suppose the histogram (PDF) of these W_i's looks like this:



Motivation

- **Question**: Can you use this data to determine if this machine has a positive expected payout in the next $n+1$th trial, i.e., if

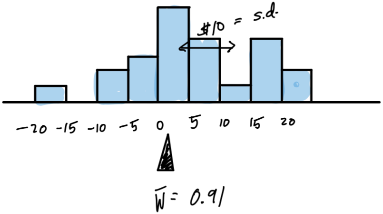
$$\mu = E(W_{n+1}) > 0.$$

- If so, you should sit down and play as much as possible.

- **However**, μ (population parameter) is unknown and we want to use the previous play to conjecture if $\mu > 0$ or not.

- Is it reasonable to think μ is near $\$0.91$? If so, how wrong could this be
  - If it can be wrong by $\pm\$10$, then we can't be sure if true $\mu > 0$ or not.
  - If it can be wrong by only $\pm\$0.1$, then we can be confident that $\mu > 0$.

Motivation



- Suppose we know $E(W_i) = \mu$, $\text{sd}(W_i) = \$10$, $i = 1, \ldots, n$.

- **Case 1**: Let $n = 100$. Then,

$$E(\bar{W}) = E(W_1) = \mu$$

$$\text{sd}(\bar{W}) = \frac{\text{sd}(W_1)}{\sqrt{n}} = \frac{10}{\sqrt{100}} = 1,$$

and applying the CLT,

$$\bar{W} \approx N(E(\bar{W}), \text{sd}(\bar{W})^2) = N(\mu, 1^2)$$

Motivation

- **Case 1**: Let $n = 100$. Then,

$$E(\bar{W}) = E(W_1) = \mu$$
$$\mathrm{sd}(\bar{W}) = \frac{\mathrm{sd}(W_1)}{\sqrt{n}} = \frac{10}{\sqrt{100}} = 1,$$

  and applying the CLT,

$$\bar{W} \approx N(E(\bar{W}), \mathrm{sd}(\bar{W})^2) = N(\mu, 1^2).$$

- So $\bar{W}$ behaves like a random draw from $N(\mu, 1^2)$ so we would expect $\bar{W}$ to be fallen within $E(\bar{W}) \pm \mathrm{sd}(\bar{W}) = \mu \pm 1$ approximately 68%. Similarly, $\bar{W}$ to be within $E(\bar{W}) \pm \mathrm{sd}(\bar{W}) = \mu \pm 2$ approximately 95%.

- Moving things around, we "expect" μ to be $\bar{W} \pm 2$ approximately 95%.

- **Takeaway**: It is possible that $\mu \leq 0$.

Motivation

- **Case 2**: Let $n = 10000$. Then,

$$E(\bar{W}) = E(W_1) = \mu$$
$$\mathrm{sd}(\bar{W}) = \frac{\mathrm{sd}(W_1)}{\sqrt{n}} = \frac{10}{\sqrt{10000}} = \frac{10}{100} = 0.1,$$

  and applying the CLT,

$$\bar{W} \approx N(E(\bar{W}), \mathrm{sd}(\bar{W})^2) = N(\mu, 0.1^2).$$

- So we would expect $\bar{W}$ to be $\mu \pm 2 \times 0.1$ approximately 95%.

- Moving things around, we "expect" μ to be $\bar{W} \pm 0.2$ approximately 95%.

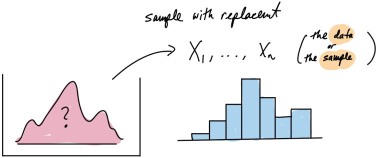
- **Takeaway**: It is likely that $\mu > 0$.

- How likely? Well, if $\mu \leq 0$, we just observed a $\bar{W}$ that is more than $\frac{0.91}{0.1} = 9.1$ standard deviations away from what we expect (Z-score is 9.1!!). This happens with extremely small probability $P(|Z| \geq 9.1) < 10^{-5}$.

Basics of Estimation

- The previous motivation example demonstrated the basics of **estimation**, **confidence intervals**, **p-values**, and **hypothesis testing**, without all the technical fuss.

- The general setup is as follows:



The left panel is some population of numbers. You want to investigate some parameter θ (for example, population mean μ or population variance σ^2), describing the shape of the PMF or PDF.

The right panel shows histogram obtained by samples (data) with replacement $X_1, X_2, \ldots, X_n$. Use this "shape" of this distribution of numbers to estimate θ, call it $\hat{\theta}$ (for example, sample mean $\bar{X} = \frac{X_1+X_2+\cdots+X_n}{n}$, or sample variance $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$). $\hat{\theta}$ is a function of the data $X_1, X_2, \ldots, X_n$, i.e., $\hat{\theta}(X_1, X_2, \ldots, X_n)$.

Basics of Estimation

- The main problem is figuring out how wrong $\hat{\theta}$ could be.
- The basic tools used for this are the **CLT**.
- **Example**: Estimating a population proportion. Interview random 1000 UCD students. 34% said they like stats. What does this tell us about all UCD students?
- Could 34% misrepresent the whole campus by as much as $\pm 20\%$?

Basics of Estimation

- **Statistical Thinking**: Let X_i be whether ith student like stats or not, (i.e., $X_i = 1$ if ith student likes stats, and $X_i = 0$ if does not). Then, we observe $X_1, X_2, \ldots, X_{1000}$. One may use the box model.

- Here, $\theta =$ proportion of students who like stats over all students. Our estimator $\hat{\theta} =$ proportion of students who like stats in the 1000 samples (data).

$$\hat{\theta} = \frac{X_1 + X_2 + \cdots X_{1000}}{1000} = \bar{X}.$$

- We have $\hat{\theta} = 0.34$ from the data (34% students like stats over 1000 students). How close do we expect this to be to θ?

Evaluating a Good Point Estimator

- We have the following measure to :

$$
\begin{aligned}
\hat{\theta} - \theta &= \text{Error}, \\
(\hat{\theta} - \theta)^2 &= \text{Squared Error}, \\
E\left[(\hat{\theta} - \theta)^2\right] &= \text{Mean Squared Error, called } \mathrm{MSE}(\hat{\theta}), \\
\sqrt{E\left[(\hat{\theta} - \theta)^2\right]} &= \text{Root Mean Squared Error, called } \mathrm{RMSE}(\hat{\theta}).
\end{aligned}
$$

- RMSE($\hat{\theta}$) tells you the typical error when using $\hat{\theta}$ to estimate θ.

Evaluating a Good Point Estimator

- Recall our example, X_i be whether ith student like stats or not, (i.e., $X_i = 1$ if ith student likes stats, and $X_i = 0$ if does not). The X_i has only 0 and 1 with probability of students who like stats over all students θ. Then, $X_i \sim \text{Bernoulli}(\theta)$.

- Then, we have

$$E(\hat{\theta}) = E\left(\frac{X_1 + X_2 + \cdots + X_{1000}}{1000}\right) = E(X_1) = \theta.$$

- So $E(\hat{\theta}) = \theta$ which means if a "million" other people interviewed 1000 random UCD student and got their own values for $\hat{\theta}$, the average value of these $\hat{\theta}$ would be the true θ.

- Another way to say this is that the estimate $\hat{\theta}$ has no systematic bias. For estimates that do have bias we can quantify it by

$$\text{Bias}(\hat{\theta}) = E(\hat{\theta}) - \theta.$$

Evaluating a Good Point Estimator

- For our UCD student example Bias$(\hat{\theta}) = 0$, i.e., $\theta = E(\hat{\theta})$, so

$$\text{RMSE}(\hat{\theta}) = \sqrt{E\left[(\hat{\theta} - \theta)^2\right]} = \sqrt{E\left[(\hat{\theta} - E(\hat{\theta}))^2\right]} = \mathrm{sd}(\hat{\theta}) = \mathrm{SE}(\hat{\theta}).$$

- **Remark**: One may use "Standard Error (SE)" for the standard deviation of the sample statistic, $\mathrm{SE}(\hat{\theta}) = \mathrm{sd}(\hat{\theta})$. The standard deviation sd usually measures variability of individual observations $\mathrm{sd}(X_1)$.

- Therefore, since $X_1 \sim$ Bernoulli(θ), $\mathrm{sd}(X_1) = \theta(1-\theta)$, we have

$$\begin{aligned}\text{RMSE}(\hat{\theta}) = \mathrm{sd}(\hat{\theta}) = \mathrm{sd}(\bar{X}) = \frac{\mathrm{sd}(X_1)}{\sqrt{n}} = \frac{\mathrm{sd}(X_1)}{\sqrt{1000}} \\ = \frac{\sqrt{\theta(1-\theta)}}{\sqrt{1000}}.\end{aligned}$$

- Here, the typical error $\mathrm{SE}(\hat{\theta}) = \text{RMSE}(\hat{\theta})$ when using $\hat{\theta}$ to estimate θ is $\frac{\sqrt{\theta(1-\theta)}}{\sqrt{1000}}$, but this depends on θ which we don't know.

Evaluating a Good Point Estimator

There are two sensible ways to proceed:

- **Method 1 (plug-in estimator)**: Typical error using $\hat{\theta} = 0.34$ to estimate θ is

$$\mathrm{SE}(\hat{\theta}) = \frac{\sqrt{\theta(1-\theta)}}{\sqrt{1000}} \approx \frac{\sqrt{\hat{\theta}(1-\hat{\theta})}}{\sqrt{1000}} = \frac{\sqrt{0.34 \times 0.66}}{\sqrt{1000}} = 0.015.$$

- Since we have "best prediction $\pm$ typical error" = "$E(\hat{\theta}) \pm \mathrm{sd}(\hat{\theta})$" formula, given by

$$\hat{\theta} \approx \theta \pm \frac{\sqrt{0.34 \times 0.66}}{\sqrt{1000}} \approx \theta \pm 0.015.$$

  Moving things around,

$$\theta \approx \hat{\theta} \pm \frac{\sqrt{0.34 \times 0.66}}{\sqrt{1000}} \approx \hat{\theta} \pm 0.015 = 0.34 \pm 0.015.$$

- **Conclusion**: There is almost no way the true value of $\theta > 0.4$. If $\theta > 0.4$, then $\hat{\theta}$ was observed to be more than 4 standard errors below what we expect, $0.4 = 0.34 + 4 \times 0.015$, $P(Z > 4) < 10^{-5}$.

Evaluating a Good Point Estimator

There are two sensible ways to proceed:

- **Method 2 (conservative approach)**: Function $\mathrm{SE}(\hat{\theta}) = \frac{\sqrt{\theta(1-\theta)}}{\sqrt{1000}}$ is maximized when $\theta = \frac{1}{2}$. (**Why?** draw a plot)

- Then, typical error using $\hat{\theta}$ to estimate θ is

$$\mathrm{SE}(\hat{\theta}) = \frac{\sqrt{\theta(1-\theta)}}{\sqrt{1000}} < \frac{\sqrt{0.5 \times 0.5}}{\sqrt{1000}} = 0.0158.$$

- Then, conservative error estimate:

$$\theta \approx \hat{\theta} \pm 0.0158 = 0.34 \pm 0.0158.$$

- **Same Conclusion**: the true value of θ is probably **not less than** $0.34 - 3 \times 0.0158 = 0.2926$ and **not greater than** $0.34 + 3 \times 0.0158 = 0.3874$. The interval $(0.2926, 0.3874)$ is a 99.7% confidence interval for θ.

Example of Good Estimation

- **Example**: A business wants to estimate the average time customers spend on its website. A sample of 50 users has a mean of 8.2 minutes, i.e., i.i.d. $X_1, \dots, X_{50}$, with $\bar{X} = 8.2$. Assume that standard deviation of X is 2.1 minutes, $\mathrm{sd}(X_1) = \mathrm{sd}(X_2) = \cdots \mathrm{sd}(X_{50}) = 2.1$.

- We use the sample mean $\bar{X} = 8.2$ minutes as our best estimate of the true average time customers spend on the site.

- The standard error is $\mathrm{SE}(\bar{X}) = \frac{\mathrm{sd}(X_1)}{\sqrt{n}} = \frac{2.1}{\sqrt{50}} = 0.297$ minutes quantifies the variability in our estimate due to sampling.

- If we took many batches of such 50 users samples, most sample means $\bar{X}$ would fall within about $\pm 2\mathrm{SE}(\bar{X}) = \pm 0.6$ minutes of 8.2 (roughly a 95% range).