



## Topic 5: Confidence Intervals

Ethan P. Marzban    University of California, Santa Barbara    PSTAT 120B



# Outline

1. Confidence Intervals

2. Normal Confidence Intervals



## Leadup

- Let's go on a fishing trip!
- Say we saw a fish in a specific spot in a lake.
- To catch a fish, we could simply throw a spear right where we saw our last fish.
- But... is that really the most efficient way to catch a fish?
- Wouldn't it be better to cast a *net*, somewhere around where we saw our last fish, to try and increase our odds of catching at least one fish?



## Leadup

- In many ways, estimation is like fishing.
- The true value of the population parameter is analogous to our fish.
- Constructing a single estimate from a single estimator (constructed from a single sample) is like throwing a spear - our estimate is a single value, which we hope is pretty close to the true value of the population parameter.
- So, what's the analog of casting a net in estimation?

# Confidence Intervals

---



## Interval Estimators

- To make things more mathematical, suppose we have an i.i.d. sample  $\vec{Y} := \{Y_i\}_{i=1}^n$  from a population with parameter  $\theta$ .
- Previously, we constructed point estimators  $\hat{\theta}_n$  which, in the language of our textbook, are rules we can use to generate numerical estimates of  $\theta$ .
- We'll now turn our attention to interval estimators (often referred to as confidence intervals), which we hope will cover the true value of  $\theta$ .



## Interval Estimators

- As the name suggests, an interval estimator is a random interval. That is, it is an interval of the form

$$\left[ \hat{\theta}_L, \hat{\theta}_U \right]$$

where  $\hat{\theta}_L$  and  $\hat{\theta}_U$  are random variables.

- Note that the *endpoints* of a confidence interval are random.
  - The lower endpoint,  $\hat{\theta}_L$ , is often referred to as the **lower confidence limit** and the upper endpoint,  $\hat{\theta}_U$ , is often referred to as the **upper confidence limit**.



## Interval Estimators

- As I mentioned at the start of this discussion, we want to have some degree of certainty that the interval  $[\hat{\theta}_L, \hat{\theta}_U]$  covers the true value of  $\theta$ .
- Since our interval is random, it makes sense to talk about the **coverage probability** (aka **confidence coefficient**), defined to be

$$\mathbb{P}(\hat{\theta}_L \leq \theta \leq \hat{\theta}_U)$$

- So, for example, a **95% confidence interval** (i.e. a confidence interval with 95% coverage probability) is one such that

$$\mathbb{P}(\hat{\theta}_L \leq \theta \leq \hat{\theta}_U) = 0.95$$

i.e. an interval  $[\hat{\theta}_L, \hat{\theta}_U]$  that we are 95% certain covers the true value of  $\theta$ .





## Example: Population Mean

- As a somewhat more concrete example, let's return to our problem of trying to estimate the true average weight of a randomly-selected DSH cat.
- Assuming  $Y_1, \dots, Y_n$  denotes an i.i.d. sample of cat weights following some distribution with unknown parameter  $\mu$ , a  $(1 - \alpha) \times 100\%$  confidence interval for  $\mu$  is an interval  $[\hat{\mu}_L, \hat{\mu}_U]$  such that

$$\mathbb{P}(\hat{\mu}_L \leq \mu \leq \hat{\mu}_U) = 1 - \alpha$$

where  $1 - \alpha$  denotes the coverage probability. (The reason why we use  $(1 - \alpha)$  will become clear next week, after we discuss Hypothesis Testing.)



## Example: Population Mean

- Can anyone give me a 100% confidence interval for  $\mu$ ?
- Sure:  $(-\infty, \infty)$ . I am 100% sure that the true average weight of a randomly-selected DSH cat lies somewhere between  $-\infty$  and  $\infty$ .
  - Indeed, even  $[0, \infty)$  would be a 100% confidence interval for  $\mu$ , based on the physical constraints of our problem but we can ignore that for now.
- Alright, but this is an (effectively) useless interval! But, this highlights something important: there is a tradeoff between coverage probability and the *width* of our confidence interval. Higher coverage probabilities necessitate larger and larger confidence intervals - this is why it's not really practical to construct a 100% confidence interval.



# Confidence Intervals

- Alright, so let's start constructing confidence intervals!
- I'm going to break our considerations into two: first we'll talk about constructing confidence intervals assuming a normally-distributed population, and then we'll relax the normality assumption and discuss ways to construct confidence intervals for more general distributions.

# Normal Confidence Intervals

---



# First Goal

## Goal

Given  $Y_1, \dots, Y_n \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$  for an unknown  $\mu \in \mathbb{R}$  but a known  $\sigma^2 > 0$ , we want to construct a  $(1 - \alpha) \times 100\%$  confidence interval for  $\mu$ .



## Sample Mean

- Now, again, we know that the sample mean  $\bar{Y}_n$  is a very good point estimator for  $\mu$ .
  - Again, it's an unbiased and consistent estimator for  $\mu$ , as well as a sufficient statistic for  $\mu$ .
- Of course, there's no guarantee that for any *particular* sample,  $\bar{Y}_n$  will be exactly equal to  $\mu$  (hence why we're trying to construct *intervals* now!)
- But, consistency more or less tells us that  $\bar{Y}_n$  will probably be quite *close* to the true value of  $\mu$ .



## Sample Mean

- So, it makes sense to construct our interval by taking  $\bar{Y}_n$  (which, again, will likely be very close to the true value of  $\mu$ ), and adding and subtracting some **margin of error** (think of it like padding).
- In other words, we'll take our interval to be

$$\bar{Y}_n \pm \text{m.e.} = [\bar{Y}_n - \text{m.e.}, \bar{Y}_n + \text{m.e.}]$$

where “m.e.” stands for margin of error (i.e. the half-width of our confidence interval).

- Since we're constructing a  $(1 - \alpha) \times 100\%$  confidence interval, we want

$$\mathbb{P}(\bar{Y}_n - \text{m.e.} \leq \mu \leq \bar{Y}_n + \text{m.e.}) = 1 - \alpha$$



## Sample Mean

- So, our problem essentially boils down to finding the appropriate value of m.e. such that the above equation holds.
- Let's try and simplify our probability on the LHS a bit. I find it useful to consider each inequality separately.
- $\mathbb{P}(\bar{Y}_n - \text{m.e.} \leq \mu) = \mathbb{P}(\bar{Y}_n \leq \mu + \text{m.e.})$
- $\mathbb{P}(\mu \leq \bar{Y}_n + \text{m.e.}) = \mathbb{P}(\bar{Y}_n \geq \mu - \text{m.e.})$
- So, what we have is

$$\mathbb{P}(\bar{Y}_n - \text{m.e.} \leq \mu \leq \bar{Y}_n + \text{m.e.}) = \mathbb{P}(\mu - \text{m.e.} \leq \bar{Y}_n \leq \mu + \text{m.e.})$$





## Sample Mean

- Again, we are trying to select m.e. such that this whole probability equals  $1 - \alpha$ :

$$\mathbb{P}(\mu - \text{m.e.} \leq \bar{Y}_n \leq \mu + \text{m.e.}) = 1 - \alpha$$

- Now, we know that  $\bar{Y}_n \sim \mathcal{N}(\mu, \sigma^2/n)$ . So, it seems tempting to standardize the RHS!



## Sample Mean

- That is:

$$\begin{aligned}\mathbb{P}(\mu - \text{m.e.} \leq \bar{Y}_n \leq \mu + \text{m.e.}) &= \mathbb{P}\left(-\frac{\text{m.e.}}{\sigma/\sqrt{n}} \leq \frac{\bar{Y}_n - \mu}{\sigma/\sqrt{n}} \leq \frac{\text{m.e.}}{\sigma/\sqrt{n}}\right) \\ &= \Phi\left(\frac{\text{m.e.}}{\sigma/\sqrt{n}}\right) - \Phi\left(-\frac{\text{m.e.}}{\sigma/\sqrt{n}}\right) \\ &= 2\Phi\left(\frac{\text{m.e.}}{\sigma/\sqrt{n}}\right) - 1\end{aligned}$$

- So, our margin of error must satisfy

$$2\Phi\left(\frac{\text{m.e.}}{\sigma/\sqrt{n}}\right) - 1 = 1 - \alpha \implies \text{m.e.} = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \cdot \frac{\sigma}{\sqrt{n}}$$

# Confidence Interval for the Mean; Known Variance



## Theorem (CI for $\mu$ ; Known Variance)

Given  $Y_1, \dots, Y_n \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$  where  $\mu \in \mathbb{R}$  is unknown but  $\sigma^2 > 0$  is known, a  $(1 - \alpha) \times 100\%$  confidence interval for  $\mu$  is given by

$$\begin{aligned} \bar{Y}_n \pm \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \cdot \frac{\sigma}{\sqrt{n}} \\ = \left[ \bar{Y}_n - \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \cdot \frac{\sigma}{\sqrt{n}}, \bar{Y}_n + \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \cdot \frac{\sigma}{\sqrt{n}} \right] \end{aligned}$$



## Example

The weight of a croissant from *Le Gauchon* (in grams) is normally distributed about some unknown mean  $\mu$  and known standard deviation 2 grams. An i.i.d. sample of 8 croissants from *Le Gauchon* is taken, and their weights (in grams) are as follows:

63.5, 64.5, 65.1, 68.9, 69.9, 70.1, 72.3, 72.4

- Construct a 90% confidence interval for  $\mu$ , based on the data that was collected. You may leave your answer in terms of  $\Phi^{-1}(\cdot)$ , the inverse of the standard normal CDF.
- Would a 80% confidence interval for  $\mu$  be wider or narrower than the interval you constructed in part (a)?



## Solutions

- We only need to plug into our formula from above!
- Firstly, the sample mean is easily computed to be  $\bar{y}_8 = 68.3375$  g.
- Now, a 90% confidence interval is equivalent to a  $(1 - 0.1) \times 100\%$  confidence interval, meaning we plug  $\alpha = 0.1$  into our CI formula from above:

$$\bar{y}_n \pm \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \cdot \frac{\sigma}{\sqrt{n}} = 68.3375 \pm \Phi^{-1}(0.95) \cdot \frac{2}{\sqrt{8}}$$

- With a computer software, we can compute this to be  $[67.17441, 69.50059]$  - that is, we are 95% confident that the true average weight of a *Le Gauch* croissant is between 67.17441 grams and 69.50059 grams (**notice the wording of our conclusion!**)



## Solutions

- For part (b), we need only to remember our discussion from earlier, about the relationship between the width of a CI and our coverage probability.
- Higher coverage probabilities necessitate wider intervals.
- Since an 80% coverage probability is *less* than a 95% coverage probability, we expect an 80% confidence interval to be **narrower** than a 95% confidence interval.
- If you're curious, you can construct an 80% confidence interval which you should find to be around  $[67.43131, 69.24369]$ , which is indeed narrower than our interval from part (a).



## Second Goal

### Goal

Given  $Y_1, \dots, Y_n \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$  for an unknown  $\mu \in \mathbb{R}$  and an unknown  $\sigma^2 > 0$ , we want to construct a  $(1 - \alpha) \times 100\%$  confidence interval for  $\mu$ .



## Second Goal

- Let's still consider an interval for the form

$$\bar{Y}_n \pm \text{m.e.}$$

- Re-using some of the work we did in the previous case (where  $\sigma^2$  was known), we have

$$\mathbb{P}(\mu - \text{m.e.} \leq \bar{Y}_n \leq \mu + \text{m.e.}) = 1 - \alpha$$

- Before, we standardized because we knew that  $\bar{Y}_n \sim \mathcal{N}(\mu, \sigma^2/n)$ .
- But, even though this is true, this fact doesn't really help us in practice since the value of  $\sigma$  is unknown!
- So, here's our clever idea - let's replace  $\sigma$  with a "good" estimator for it - namely,  $S_n$  (the sample standard deviation).





## Second Goal

- This works out well, because

$$\left( \frac{\bar{Y}_n - \mu}{S_n/\sqrt{n}} \right) \sim t_{n-1}$$

by our “Modified Standardization Result” from our lecture on multivariate transformations involving the normal distribution.



## Second Goal

- So:

$$\begin{aligned}\mathbb{P}(\mu - \text{m.e.} \leq \bar{Y}_n \leq \mu + \text{m.e.}) &= \mathbb{P}\left(-\frac{\text{m.e.}}{S_n/\sqrt{n}} \leq \frac{\bar{Y}_n - \mu}{S_n/\sqrt{n}} \leq \frac{\text{m.e.}}{S_n/\sqrt{n}}\right) \\ &= F_{t_{n-1}}^{-1}\left(\frac{\text{m.e.}}{S_n/\sqrt{n}}\right) - F_{t_{n-1}}^{-1}\left(-\frac{\text{m.e.}}{S_n/\sqrt{n}}\right) \\ &= 2F_{t_{n-1}}^{-1}\left(\frac{\text{m.e.}}{S_n/\sqrt{n}}\right) - 1\end{aligned}$$

- Thus, our margin of error must satisfy

$$2F_{t_{n-1}}^{-1}\left(\frac{\text{m.e.}}{\sigma/\sqrt{n}}\right) - 1 = 1 - \alpha \implies \text{m.e.} = F_{t_{n-1}}^{-1}\left(1 - \frac{\alpha}{2}\right) \cdot \frac{\sigma}{\sqrt{n}}$$

# Confidence Interval for the Mean; Unknown Variance



## Theorem (CI for $\mu$ ; Unknown Variance)

Given  $Y_1, \dots, Y_n \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$  where both  $\mu \in \mathbb{R}$  and  $\sigma^2 > 0$  are unknown, a  $(1 - \alpha) \times 100\%$  confidence interval for  $\mu$  is given by

$$\begin{aligned} \bar{Y}_n \pm F_{t_{n-1}}^{-1} \left( 1 - \frac{\alpha}{2} \right) \cdot \frac{S_n}{\sqrt{n}} \\ = \left[ \bar{Y}_n - F_{t_{n-1}}^{-1} \left( 1 - \frac{\alpha}{2} \right) \cdot \frac{S_n}{\sqrt{n}}, \bar{Y}_n + F_{t_{n-1}}^{-1} \left( 1 - \frac{\alpha}{2} \right) \cdot \frac{S_n}{\sqrt{n}} \right] \end{aligned}$$



## Example

Assume the same setup as the previous croissant example, except now assume that  $\sigma^2$  is unknown. Construct a 95% CI for  $\mu$ , the true average weight of a *Le Gauchon* croissant.

- We still have  $\bar{y}_n = 68.3375$  g. We also have  $s_8 = 3.518903$ . Therefore, plugging into our formula from the previous slide, our CI is

$$68.3375 \pm F_{t_7}^{-1}(0.975) \cdot \frac{3.518903}{\sqrt{8}}$$

which, using a computer software, amounts to around [65.39562 , 71.27938].

# Asymptotic Confidence Intervals for the Mean



- Note that the CLT enables us to relatively easily construct *large-sample* (i.e. *asymptotic*) confidence intervals for the mean.
- That is, we know that regardless of our population distribution (assuming finite mean and variance),

$$\frac{\sqrt{n}(\bar{Y}_n - \mu)}{\sigma} \rightsquigarrow \mathcal{N}(0, 1)$$

- You'll work through some problems relating to this on the next (and final!) homework.



## Interpreting Confidence Intervals

- One interpretation of an  $(1 - \alpha) \times 100\%$  confidence interval  $[a, b]$  is: “we are  $(1 - \alpha) \times 100\%$  certain that the interval  $[a, b]$  contains the true value of  $\theta$ .”
  - So, for example, a 95% CI for a population mean  $\mu$  can be interpreted as an interval that we are 95% certain covers the true value of  $\mu$ .
- There is another interesting way to interpret CIs: If the same procedure was used many times, each individual interval would either contain or fail to contain the true value of  $\theta$ , but the percentage of all intervals that capture  $\theta$  would be very close to  $(1 - \alpha) \times 100\%$ . (This is the wording taken from the textbook.) Let's see this in action by way of a live demo.