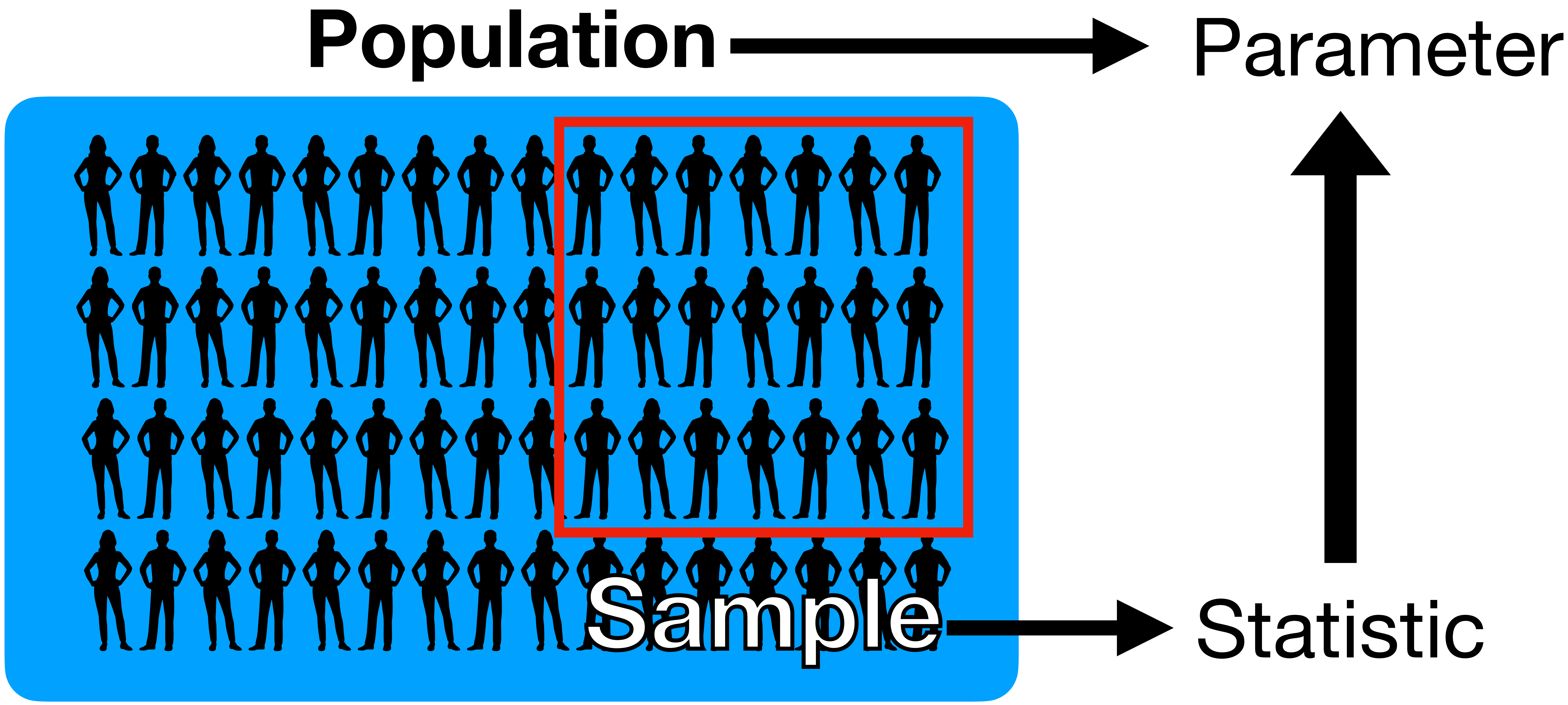
Lesson 025
Confidence Intervals

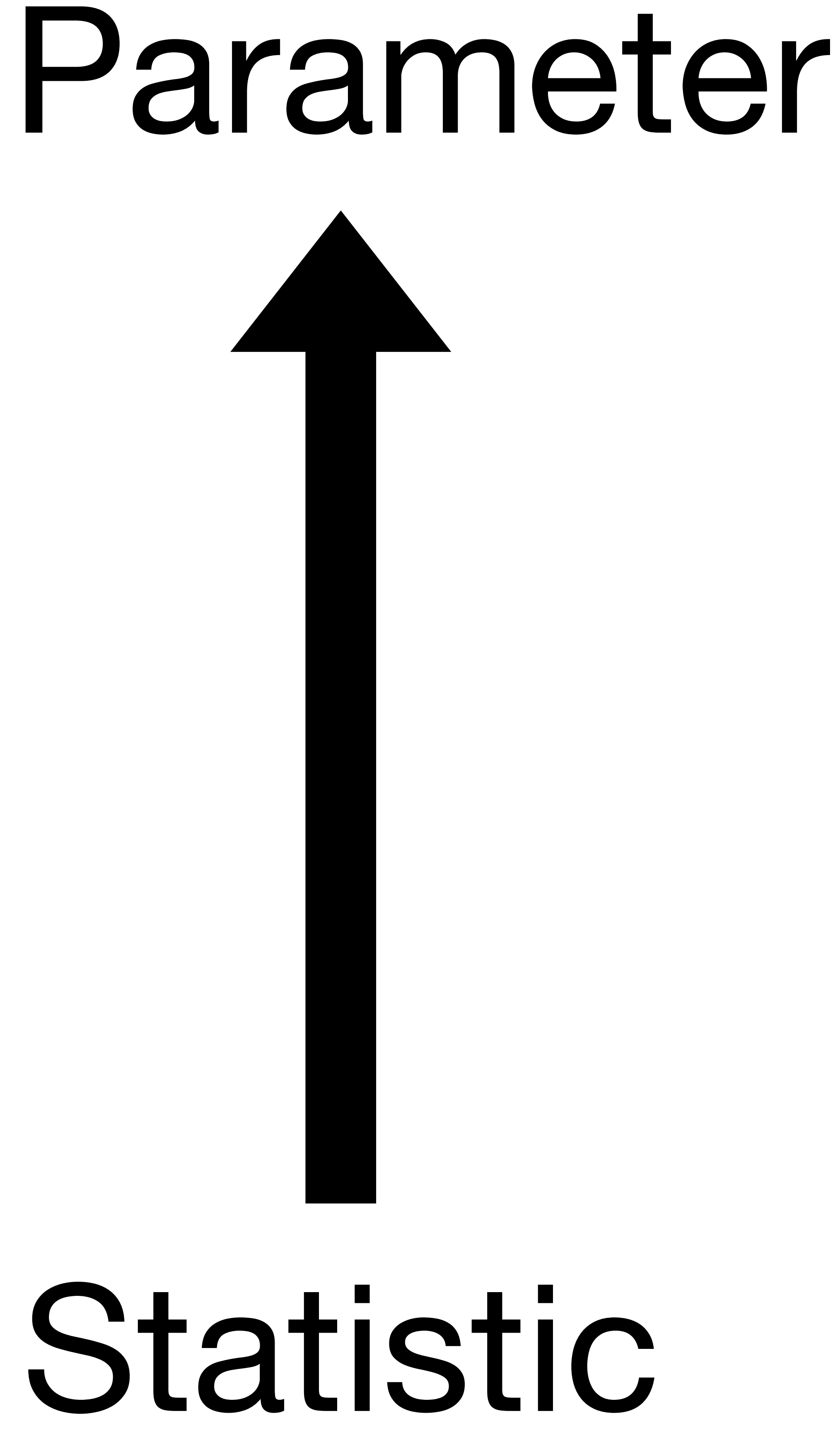
Monday, November 13


Population
Parameter
Sample
Statistic

We can use an estimate to get an single value that we think is reasonable.

The estimate will follow the estimator's sampling distribution.

How can we summarize this uncertainty?



Interval Estimation

- Point estimates fail to quantify uncertainty in our estimation procedure.
- We know that $\widehat{\theta} \neq \theta$, in general, we just hope that it's close.
- Often it is more useful to give a **range of plausible values**.
- These **interval estimates** are such that if they are more narrow, our estimate is more precise.

Confidence Intervals

- Confidence intervals (CIs) are one type of interval estimates.
- CIs allow you to indicate a threshold of "confidence" associated with a particular range of values.
- CIs are **random intervals,** with each endpoint being random.
- In general, a CI gives you a set of plausible values for a parameter, based on a sample.

Constructing Confidence Intervals

- Suppose that $\widehat{\theta}$ is an estimator for θ.
- Suppose that $f(x)$ and $F(X)$ are the pdf and cdf for the **sampling distribution.**
- We can use these quantities to find values a and b such that, for some $\alpha \in (0,1)$

$$P(a \leq \widehat{\theta} \leq b) = 1 - \alpha$$

Constructing Confidence Intervals (Cont.)

- If we take α to be small, say 0.05, then there is a high probability for $\widehat{\theta}$ to fall in $[a, b]$.
- These endpoints will typically depend on θ.
- If we then invert the inequality, $a(\theta) \leq \widehat{\theta} \leq b(\theta)$ we can derive an interval for θ.
- **This is a confidence interval for θ.**

Example: Normal Population, Known Variance

Interpretation for Confidence Intervals

- Before any data are collected, the process has a $1-\alpha$ chance of constructing an interval that contains the truth.

- **Important:** we **cannot** (ever!!) say "There is a $1-\alpha$ probability that the CI contains the truth."

- Once an interval is constructed it either does or it does not contain the truth.

"If we were to repeat this process many, many times, approximately $100 \times (1 - \alpha)\,\%$ of them will contain the truth."

If $[-3, 6]$ is a 95% confidence interval for θ, which of the following statements are true?

θ falls into the interval $[-3, 6]$.

0%

$P(-3 \leq \theta \leq 6) = 0.95$

0%

$P(-3 \leq \theta \leq 6) = 0.05$

0%

Approximately 95% of intervals constructed using this process will contain the truth.

0%

Suppose that a 99% confidence interval for μ is obtained as $[-4, 4]$. If $\mu = 5$, which statement is true?

$P(-4 \leq \mu \leq 4) = 0.99$

0%

$P(-4 \leq \mu \leq 4) = 0.01$

0%

$P(-4 \leq \mu \leq 4) = 0$

0%

$P(-4 \leq \mu \leq 4) = 1$

0%

Further Notes on Confidence Intervals

- We should select α based on the particular use case.
- Sometimes the interval will depend on unknown parameters: more on this later.
- Intervals can be symmetric or asymmetric.
- Intervals can be two-sided or one-sided.
  - Give an upper (lower) bound on the parameter of interest.

Margin of Error

- In the normal example we say that the interval was $\overline{X} \pm Z_{\alpha/2}\sigma \Big/ \sqrt{n}$.
- This has a width of $w = 2 \times Z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$.
- Half of this width is called the **margin of error**

$$\textbf{Margin of Error} = \frac{w}{2} = Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Margin of Error

- The margin of error is a measure of **precision of the estimate**.
- Smaller values are preferable, all else equal.
- The margin of error will shrink as the confidence decreases (α increases).
- Typically depends on the sample size, shrinking as n increases.

What is the margin of error for an interval $[-2, 8]$?

2

0%

8

0%

10

0%

5

0%

If the same process is used to construct two confidence intervals, which of the following statements is false?

If n increases, everything else equal, then the second interval will be shorter.

0%

If α increases, everything else equal, then the second interval will be shorter.

0%

If the confidence level increases, everything else equal, then the second interval will be longer.

0%

If the standard error decreases, everything else equal, then the second interval will be shorter.

0%

If the same process is used to construct two confidence intervals, which of the following statements is false?

In most settings, if n increases, everything else equal, then the second interval will be shorter.

0%

If α increases, everything else equal, then the second interval will be shorter.

0%

If the confidence level increases, everything else equal, then the second interval will be longer.

0%

If the standard error decreases, everything else equal, then the second interval will be shorter.

0%

All of the statements are true.

0%

Sample Sizing

- If you want a set level of precision (say ± 0.01) at a certain level of confidence (say $95\ \%$) then the margin of error can give you a particular sample size.

$$\text{ME} = |Z_{\alpha/2}| \frac{\sigma}{\sqrt{n}} \implies n = \left(Z_{\alpha/2} \frac{\sigma}{\text{ME}} \right)^2$$