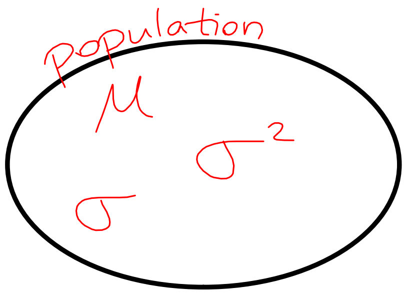
Sections 5.1 and 5.2

In Section 5.1, read pages 172 to 174.
On page 175 read the paragraph labeled "Limitations of Point Estimates."
In Section 5.2 read pages 176 to 184.
C A R E F U L L Y

Why do we take a sample?

It's difficult to sample the entire population due to time and money...



Which STATISTICS do a good job estimating PARAMETERS?

$$Bias = \mu_{\hat{\theta}} - \theta \quad = 0$$

where θ is a parameter and $\hat{\theta}$ is an estimator of θ.

If the bias is equal to zero, the statistic is doing it's job.

Is $\bar{x}$ a good estimator of μ?

$$\text{Bias} = \mu_{\bar{x}} - \mu = \mu - \mu = 0$$

Is s^2 a good estimator of σ^2?

Yes, $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

Is $\hat{p}$ a good estimator of p?

$$\hat{p} = \frac{X}{n} \qquad X \sim \text{Binomial}(n,p)$$

$$\text{Bias} = \mu_{\hat{p}} - p = \mu_{X/n} - p$$

$$= \frac{1}{n}\mu_X - p$$

$$= \frac{1}{\cancel{n}} \cdot \cancel{n} \cdot p - p$$

$$= p - p = 0$$

Which STATISTICS are BIASED?

median

range

standard deviation $\sqrt{\ }$ ing creates bias :-(
* Jensen's Inequality

Bias is small when the sample size is large, so we use s **to estimate** σ**.**

A statistic is a point estimate.

It will almost NEVER equal the parameter it is estimating.

There is uncertainty.

We will create an interval we HOPE will capture the parameter with a specified level of CONFIDENCE.

Central Limit Theorem: $\bar{X} \sim N\left(\mu_x, \frac{\sigma_x^2}{n}\right)$

if n is large enough!

Point Estimate: Use $\bar{X}$ to estimate μ_x.

Goal: Get an interval that we HOPE captures μ_x.

Confidence? Why not probability?

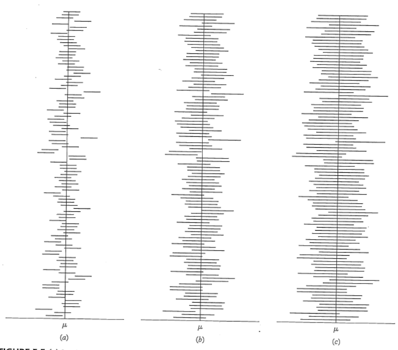
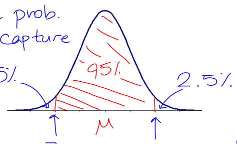


FIGURE 5.5 *(a)* One hundred 68% confidence intervals for a population mean, each computed from a different sample. Although precise, they fail to cover the population mean 32% of the time. This high failure rate makes the 68% confidence interval unacceptable for practical purposes. *(b)* One hundred 95% confidence intervals computed from these samples. This represents a good compromise between reliability and precision for many purposes. *(c)* One hundred 99.7% confidence intervals computed from these samples. These intervals fail to cover the population mean only three times in 1000. They are extremely reliable, but imprecise.

Goal: Estimate the unknown population mean value μ.

How: Create an interval after selecting how "confident" we want to be.

α is the prob. we do not capture μ.



$-1.96 = -z_{\alpha/2}$ $\quad$ $z_{\alpha/2} = 1.96$

$\bar{X} \sim N\left(\mu_x, \frac{\sigma_x^2}{n}\right)$

$P(-1.96 < Z < 1.96) = 0.95$

$P\left(-1.96 < \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} < 1.96\right) = 0.95$

$P\left(-1.96 \cdot \frac{\sigma}{\sqrt{n}} < \bar{x} - \mu < 1.96 \cdot \frac{\sigma}{\sqrt{n}}\right) = 0.95$

$P\left(-\bar{x} - 1.96 \cdot \frac{\sigma}{\sqrt{n}} < -\mu < -\bar{x} + 1.96 \frac{\sigma}{\sqrt{n}}\right) = 0.95$

$P\left(\bar{x} - 1.96 \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}\right) = 0.95$

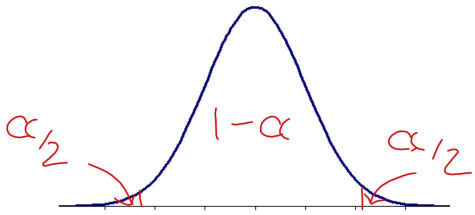
95% Confidence Interval

$$\underbrace{\bar{x} \pm \underbrace{1.96}_{\text{Critical value}} \cdot \underbrace{\frac{\sigma}{\sqrt{n}}}_{\text{Standard Error}}}_{\text{Margin of Error}}$$

General Confidence Intervals

How often are you willing to fail? α

So...Confidence Level $= 1-\alpha$



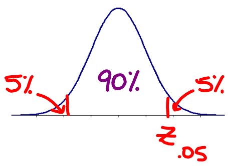
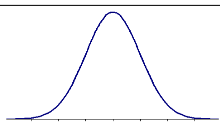
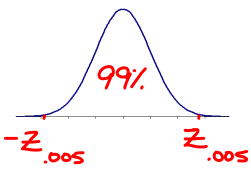
$(1-\alpha)\%$ CI formula

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Confidence Level: **1 - α**

Success rate **of the confidence interval procedure.**

We pick how many times out of 100 we can tolerate ***failing*** **to trap the true parameter.**

Confidence Level, 1 - α	Critical Value: *Z*-value Cutoff	$Z_{\alpha/2}$
90% $= 1-\alpha$ $\alpha = 10\%$ $\alpha/2 = 5\%$		$z_{.05} = 1.645$
95%		$z_{.025} = 1.96$
99% $\alpha = 1\%$ $\alpha/2 = 0.5\%$		$z_{.005} = 2.576$

The 95% Confidence Interval for the mean lifetime of light bulbs was found to be (334 hours, 366 hours).

Point Estimate: $\bar{x} = 350$

Margin of Error: 16 (Half of 32 ... distance from $\bar{x}$ to endpt)

Standard Error:

$16 \rightarrow ME = \underset{\text{value}}{\overset{1.96}{\text{Critical}}} \cdot SE \qquad SE = 8.163$

Interpretation: We can be 95% confident that the true mean lifetime of lightbulbs is between 334 and 366 hours.

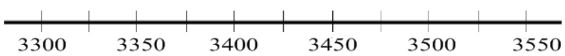
Example: A simple random sample of baby birth weights in the U.S. has a mean of 3433 g and a standard deviation of 495 g.

If the sample size had been 75, construct a 95% CI estimate of the mean birth weight in the U.S.

$$3433 \pm 1.96 \cdot \frac{495}{\sqrt{75}}$$

If the sample size had been 7500, construct a 95% CI estimate of the mean birth weight in the U.S.

$$3433 \pm 1.96 \cdot \frac{495}{\sqrt{7500}}$$



5.1 and 5.2 Due Tuesday,
Project 3 due Thursday (2/16)