ECO220Y
Estimation:
Confidence Interval Estimator for Population Mean

Readings: Chapter 13 (section 13.2) and re-read Chapter 11(skip 11.5)

Fall 2011

Lecture 11

Classroom Rules

⋆ No Side Conversations! ⋆

Why?

Too costly for yourselves and others:

1. Distracts other students in class
2. Leads to using time unproductively
3. Develops a feeling of disrespect
4. Reduces participation in class

⋆ Check our course web-site regularly ⋆

1. not to miss important announcements
2. keep track of where we are in the course

Recap: CI for Population Proportion

- Standard Error - $SE_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
- Significance level - α
- Confidence level - $1 - \alpha$
- Critical value - $z_{\alpha/2}$ (re-read section 11.3 in your textbook)
- Confidence interval - $\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

Aside: How to Derive CI for Population Proportion

$$1-\alpha = P\left(p - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} < \hat{p} < p + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right)$$

Aside: How to Derive CI for Population Proportion

$$1 - \alpha = P\left(p - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} < \hat{p} < p + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right)$$

$$1 - \alpha = P\left(p - \hat{p} - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} < \hat{p} - \hat{p} < p - \hat{p} + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right)$$

Aside: How to Derive CI for Population Proportion

$$1 - \alpha = P\left(p - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} < \hat{p} < p + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right)$$

$$1 - \alpha = P\left(p - \hat{p} - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} < \hat{p} - \hat{p} < p - \hat{p} + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right)$$

$$1 - \alpha = P\left(-\hat{p} - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} < 0 - p < -\hat{p} + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right)$$

Aside: How to Derive CI for Population Proportion

$$1-\alpha = P\left(p - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} < \hat{p} < p + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right)$$

$$1-\alpha = P\left(p - \hat{p} - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} < \hat{p} - \hat{p} < p - \hat{p} + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right)$$

$$1-\alpha = P\left(-\hat{p} - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} < 0 - p < -\hat{p} + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right)$$

$$1-\alpha = P\left(\hat{p} + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} > p > \hat{p} - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right)$$

Aside: How to Derive CI for Population Proportion

$$1-\alpha = P\left(p - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} < \hat{p} < p + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right)$$

$$1-\alpha = P\left(p - \hat{p} - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} < \hat{p} - \hat{p} < p - \hat{p} + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right)$$

$$1-\alpha = P\left(-\hat{p} - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} < 0 - p < -\hat{p} + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right)$$

$$1-\alpha = P\left(\hat{p} + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} > p > \hat{p} - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right)$$

$$1-\alpha = P\left(\hat{p} - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right)$$

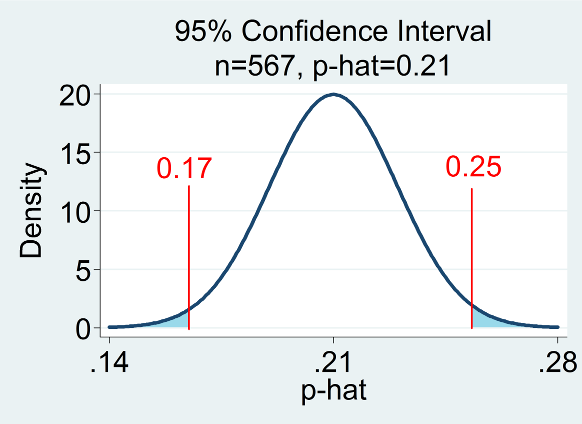
Example

To determine how many Americans smoke, annual surveys are conducted. In 2000, the survey asked a random sample of 567 Americans 18 years of age and over whether they smoked on some day. 119 individuals said "yes". Estimate with 95% confidence the proportion of Americans who smoke.

- $n = 567$ and $X = 119$
- $\hat{p} = \frac{X}{n} = \frac{119}{567} = 0.2116 \approx 0.21$
- Can use normal approximation?
  - Check rule of thumb:
  - $n\hat{p} = 567 * 0.21 = 119 > 10$
  - and $n(1 - \hat{p}) = 567 * 0.79 = 448 > 10$
- $\alpha = 1 - 95 = 5$, $z_{\alpha/2} = 1.96$
- $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{0.21 * 0.79/567} = 0.02$

Example Cont'd

$$\text{CI Estimator} = \hat{p} \pm 1.96 * 0.02 = 0.21 \pm 0.039$$



Interpretation? Margin of error?

Selecting Sample Size

For any desired level of *ME* we can choose sample size:

$$z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = ME$$

$$ME\sqrt{n} = z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})}$$

$$\boxed{n = \left[\frac{z_{\alpha/2}}{ME}\right]^2 \hat{p}(1-\hat{p})}$$

Note that in your aid sheet for the term test and final exam *ME* will be denoted as τ

Example Cont'd

To estimate proportion of Americans who smoke within ± 0.02 and with 95% confidence, sample how many?

$$n = \left[\frac{z_{\alpha/2}}{ME}\right]^2 \hat{p}(1-\hat{p})$$

$$n = \left[\frac{1.96}{0.02}\right]^2 0.5(1-0.5) = 2401$$

Why use 0.5?

$$n = \left[\frac{1.96}{0.02}\right]^2 0.21(1-0.79) \approx 1594$$

Always round up!

Preview: Inference about μ

- When population variance (σ^2) is known
- Unrealistic assumption

Learn it first

because it is simpler

- Use Standard Normal table

- When population variance (σ^2) is unknown
- Realistic assumption

Learn it second

because it is harder

- Use Student t table (Lecture 12)

Example: Computing CI for μ

A company wants to estimate the average Friday lunch break taken by its salaried executives. One Friday, 25 executives are monitored and it is found that their average lunch break is 95 minutes. If the standard deviation is assumed to be 25 minutes, what is a 95 confidence interval for the average Friday lunch break for all the company's executives?

- Let the average lunch break be X; then $X \sim (\mu_x, 25^2)$
- and $\bar{X} \sim N(\mu_x, \frac{25^2}{25})$
  - What implicit assumption we are making here?
- For this specific sample, $\bar{X} = 95$
- There is a 95% chance that $\bar{X}$ is within 1.96 st. deviations from μ_x
- $P(\mu_x - 1.96 * 5 < \bar{X} < \mu_x + 1.96 * 5) = 0.95$
- In standard notation, $P(-1.96 < Z < 1.96) = 0.95$

$\alpha = 5\%$ - chance that interval excludes parameter.
$1 - \alpha = 95\%$ - chance that interval includes parameter.

$$1 - \alpha = P[-z_{\alpha/2} < Z < z_{\alpha/2}]$$

Because of symmetry $\underbrace{-z_{\alpha/2}}_{\text{bottom percentile}} = \underbrace{z_{\alpha/2}}_{\text{top percentile}}$

$$Z = \frac{\bar{X} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

Un-standardize:

$$1 - \alpha = P[-z_{\alpha/2} < \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} < z_{\alpha/2}]$$

$$1 - \alpha = P[\mu - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \bar{X} < z_{\alpha/2}\frac{\sigma}{\sqrt{n}} + \mu]$$

Interpretation?

Derive Confidence Interval

$$1 - \alpha = P[\mu - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \bar{X} < z_{\alpha/2}\frac{\sigma}{\sqrt{n}} + \mu]$$

Derive Confidence Interval

$$1 - \alpha = P[\mu - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \bar{X} < z_{\alpha/2}\frac{\sigma}{\sqrt{n}} + \mu]$$

$$1 - \alpha = P[\mu - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} - \bar{X} < 0 < z_{\alpha/2}\frac{\sigma}{\sqrt{n}} + \mu - \bar{X}]$$

Derive Confidence Interval

$$1-\alpha = P[\mu - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \bar{X} < z_{\alpha/2}\frac{\sigma}{\sqrt{n}} + \mu]$$

$$1-\alpha = P[\mu - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} - \bar{X} < 0 < z_{\alpha/2}\frac{\sigma}{\sqrt{n}} + \mu - \bar{X}]$$

$$1-\alpha = P[-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} - \bar{X} < -\mu < z_{\alpha/2}\frac{\sigma}{\sqrt{n}} - \bar{X}]$$

Derive Confidence Interval

$$1-\alpha = P[\mu - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \bar{X} < z_{\alpha/2}\frac{\sigma}{\sqrt{n}} + \mu]$$

$$1-\alpha = P[\mu - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} - \bar{X} < 0 < z_{\alpha/2}\frac{\sigma}{\sqrt{n}} + \mu - \bar{X}]$$

$$1-\alpha = P[-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} - \bar{X} < -\mu < z_{\alpha/2}\frac{\sigma}{\sqrt{n}} - \bar{X}]$$

$$1-\alpha = P[z_{\alpha/2}\frac{\sigma}{\sqrt{n}} + \bar{X} > \mu > -z_{\alpha/2}\frac{\sigma}{\sqrt{n}} + \bar{X}]$$

Derive Confidence Interval

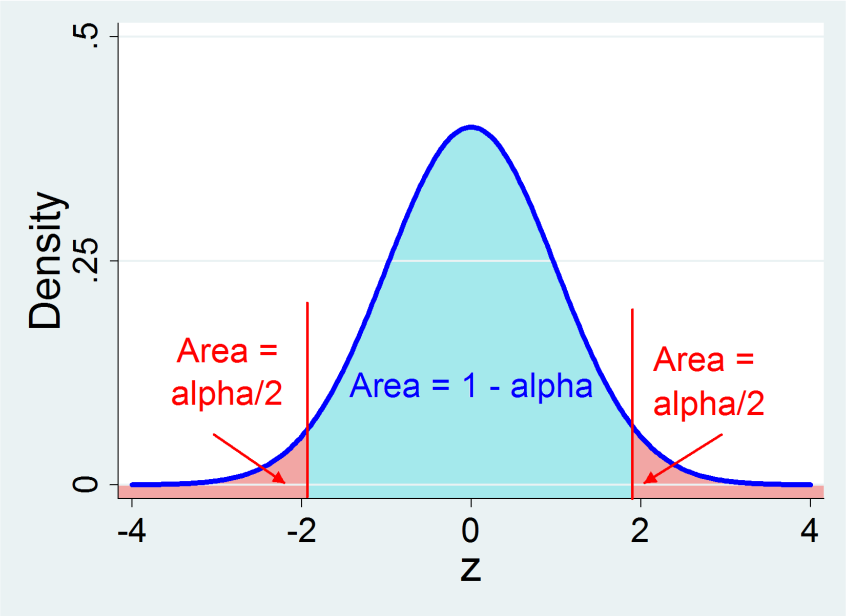
$$1 - \alpha = P[\mu - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \bar{X} < z_{\alpha/2}\frac{\sigma}{\sqrt{n}} + \mu]$$

$$1 - \alpha = P[\mu - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} - \bar{X} < 0 < z_{\alpha/2}\frac{\sigma}{\sqrt{n}} + \mu - \bar{X}]$$

$$1 - \alpha = P[-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} - \bar{X} < -\mu < z_{\alpha/2}\frac{\sigma}{\sqrt{n}} - \bar{X}]$$

$$1 - \alpha = P[z_{\alpha/2}\frac{\sigma}{\sqrt{n}} + \bar{X} > \mu > - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} + \bar{X}]$$

$$1 - \alpha = P[\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}]$$


.5
.25
0
Density
Area =
alpha/2
Area = 1 - alpha
Area =
alpha/2
-4
-2
0
2
4
z

Confidence Interval (CI) and its Interpretation

$$1 - \alpha = P[\mu - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \bar{X} < z_{\alpha/2}\frac{\sigma}{\sqrt{n}} + \mu]$$

$$\downarrow \qquad \qquad \downarrow$$

$$1 - \alpha = P[\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}]$$

Interpretation: For a random sample n from a population with mean μ and standard deviation σ there is 1-α chance that this interval contains μ.

CI Estimator of μ when σ^2 is Known

- Confidence interval estimator of μ is:

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

- Confidence level: 1-α
- Lower confidence limit (LCL):$\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$
- Upper confidence limit (UCL):$\bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$

Critical Values

Z	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
1.6	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.9	0.4738	0.47448	0.4750	0.4756	0.9761	0.4767
2.5	0.4945	0.9946	0.4948	0.4949	0.4951	0.4952

$z_{0.05} = 1.645$ $z_{0.025} = 1.96$ $z_{0.005} = 2.575$

Alternative Critical Values

1. $\alpha = 5\%$, confidence level=1-α=95%
   - $P(-1.96 < Z < 1.96) = 0.95$
   - $P(\bar{X} - 1.96 * \frac{\sigma_x}{\sqrt{n}} < \mu < \bar{X} + 1.96 * \frac{\sigma_x}{\sqrt{n}}) = 0.95$
2. $\alpha = 10\%$, confidence level=1-α=90%
   - $P(-1.645 < Z < 1.645) = 0.90$
   - $P(\bar{X} - 1.645 * \frac{\sigma_x}{\sqrt{n}} < \mu < \bar{X} + 1.645 * \frac{\sigma_x}{\sqrt{n}}) = 0.90$
3. $\alpha = 1\%$, confidence level=1-α=99%
   - $P(-2.575 < Z < 2.575) = 0.99$
   - $P(\bar{X} - 2.575 * \frac{\sigma_x}{\sqrt{n}} < \mu < \bar{X} + 2.575 * \frac{\sigma_x}{\sqrt{n}}) = 0.99$

Interval Determinants

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

- What happens to CI if n increases?
- If population variance increases?
- If we change the desired level of confidence?

Margin of Error

$$\bar{X} \pm \underbrace{z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}_{\text{margin of error}}$$

- What is the reason for our uncertainty about the estimate of μ?
  - Sampling error!
- The spread in a confidence interval is a <u>margin of sampling error</u>.
- The margin of sampling error is $z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$

Example: Computing CI

A company wants to estimate the average Friday lunch break taken by its salaried executives. One Friday, 25 executives are monitored and it is found that their average lunch break is 95 minutes. If the standard deviation is assumed to be 25 minutes, what is a 95 confidence interval for the average Friday lunch break for all the company's executives?

Back to Example: Average Lunch Break

- $\bar{X} = 95$ mins and $n = 25$
- s.e. $= \frac{\sigma}{\sqrt{n}} = \frac{25}{5} = 5$ mins
- $1 - \alpha$=0.95, $z_{\alpha/2}$=1.96
- LCL=$\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 95 - 1.96 * 5 = 85.2$ mins
- UCL=$\bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 95 + 1.96 * 5 = 104.8$ mins
- Interpretation?

Comparing Means of Two Populations

- Suppose we want to compare average salary of male (denoted as population 1) and female (denoted as population 2) employees in the same job in the same industry.
- Population one has mean salary μ_1 and standard deviation σ_1.
- Population two has mean salary μ_2 and standard deviation σ_2.
- A random sample of size n_1 from population one and n_2 from population two gives sample means of $\bar{X}_1$ and $\bar{X}_2$ and standard errors $\sigma_{\bar{X}_1}$ and $\sigma_{\bar{X}_2}$.
- We are interested to know whether $\mu_1 = \mu_2$?

Comparing Means of Two Populations

- We need to know the sampling distribution of difference between two sample means.
- Sample means are random variables, difference between two sample means is a linear combination of two random variables.
- A linear combination of independent normal random variables yields a normal variable.
- We can standardize the difference!

Mean and Variance of Sampling Distribution of Difference

$$\bar{X}_1 \sim N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right) \qquad \bar{X}_2 \sim N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right)$$

$$E[\bar{X}_1 - \bar{X}_2] = E[\bar{X}_1] - E[\bar{X}_2] = \mu_1 - \mu_2$$

$$V[\bar{X}_1 - \bar{X}_2] = 1^2 V[\bar{X}_1] + (-1)^2 V[\bar{X}_2] = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

$$\boxed{\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)}$$

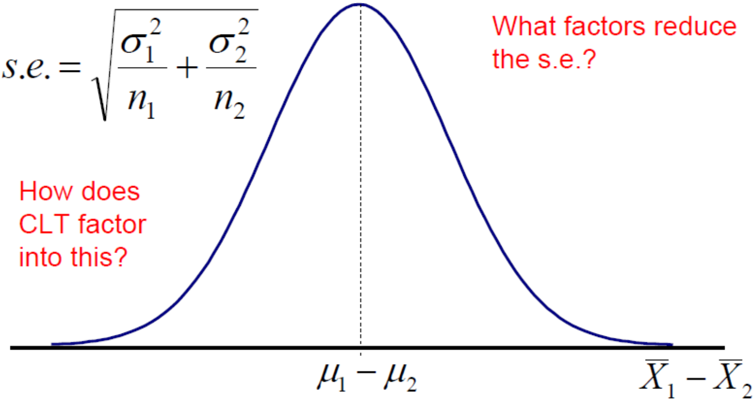
What is assumed about the relationship between two sample means?

Sampling Distribution of $\bar{X}_1 - \bar{X}_2$

$$s.e. = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

What factors reduce the s.e.?

How does CLT factor into this?



Wendy's vs. McDonald's

Between Wendy's and McDonald's, which fast-food drive-through window is faster? To answer the question, a random sample of service times for each restaurant was measured. Can we infer from these data that there are differences in service time between the two chains?

The data:

	Wendy's	McDonalds
n	213	202
$\bar{X}$	150	154
σ_x	21	24

Wendy's vs. McDonald's

- Find the sampling distribution of the difference between two sample means
  - $\bar{X}_W - \bar{X}_M \sim N(\mu_W - \mu_M, \frac{\sigma_W^2}{n_W} + \frac{\sigma_M^2}{n_M})$
  - $\bar{X}_W - \bar{X}_M \sim N(\mu_W - \mu_M, \frac{21^2}{213} + \frac{24^2}{202})$
  - $\bar{X}_W - \bar{X}_M \sim N(\mu_W - \mu_M, 4.92)$

- Construct a 95% confidence interval estimator and see whether 0 belongs to that interval
  - LCL:$(\bar{X}_W - \bar{X}_M) - 1.96 * \sigma_{\bar{X}_W - \bar{X}_M} = (150 - 154) - 1.96 * \sqrt{4.92} = -8.35$
  - UCL:$(\bar{X}_W - \bar{X}_M) + 1.96 * \sigma_{\bar{X}_W - \bar{X}_M} = (150 - 154) + 1.96 * \sqrt{4.92} = 0.35$
  - CI: (-8.35, 0.35)
- 95% chance that the interval above will include population mean (difference between population means)