

Unit 3: Foundations for inference

2. Confidence Intervals

Sta 101 – Spring 2019

Duke University, Department of Statistical Science



Dr. Ellison

Slides posted at
<https://www2.stat.duke.edu/courses/Spring19/sta101.001/>

Outline

1. Housekeeping

2. Main ideas

1. Making a Confidence Interval: Use confidence intervals to estimate population parameters
2. Conditions for using the CLT to make a Conf. Interval:
 Statistical inference methods based on the CLT depend on the same conditions as the CLT
3. What sample size do we need? Calculate the sample size a priori to achieve desired margin of error

3. Summary

Coming up...

- ▶ Lab Assignment 4 is due **Thursday just before your lab section time.**
- ▶ Sample Midterm 1+Key Posted on Sakai Resources!
- ▶ Midterm 1 Review **Wednesday 2/13**
- ▶ Midterm 1 (Unit 1-Unit 3.2) **Monday 2/18**

1

Outline

1. Housekeeping

2. Main ideas

1. Making a Confidence Interval: Use confidence intervals to estimate population parameters
2. Conditions for using the CLT to make a Conf. Interval:
 Statistical inference methods based on the CLT depend on the same conditions as the CLT
3. Common misconceptions about confidence intervals
4. What sample size do we need? Calculate the sample size a priori to achieve desired margin of error

3. Summary

Outline

What is the general structure of a confidence interval (for any population parameter)?

2. Use confidence intervals to estimate population parameters

General Confidence Interval Set-Up

point estimate ± *margin of error*

3

2. Use confidence intervals to estimate population parameters

General Confidence Interval Set-Up

point estimate ± *margin of error*

point estimate ± (critical value)(standard error)

3

2. Use confidence intervals to estimate population parameters

General Confidence Interval Set-Up

point estimate ± *margin of error*

point estimate ± (critical value)(standard error)

The sample statistic

- (see upcoming slides for info)

Standard deviation of:

- the distribution sample statistics
- (aka: the sampling distribution)

3

Outline

What is the confidence interval for a population mean?

2. Use confidence intervals to estimate population parameters

(1 - α) Confidence Interval for μ

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

*right conditions must be met to use this equation

3

2. Use confidence intervals to estimate population parameters

(1 - α) Confidence Interval for μ

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

*right conditions must be met to use this equation

How to find the critical value $z_{\alpha/2}^*$:

- Look up the positive Z-score in Standard Normal Distribution (z-tables) that creates a right tail with area/probability $\alpha/2$.

• Z-Distribution aka: Standard Normal Distribution
 • Aka: N(0,1)

Normal probability table

Normal distribution		Normal deviate (area of Z)									
Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5200	0.5240	0.5280	0.5320	0.5360	0.5400
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753	0.5793
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141	0.6179
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517	0.6554
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879	0.6915
0.5	0.6915	0.6950	0.6985	0.7020	0.7054	0.7089	0.7124	0.7159	0.7193	0.7227	0.7261
0.6	0.7261	0.7294	0.7327	0.7359	0.7391	0.7423	0.7454	0.7485	0.7516	0.7547	0.7578
0.7	0.7578	0.7608	0.7638	0.7667	0.7696	0.7725	0.7754	0.7782	0.7811	0.7839	0.7867
0.8	0.7867	0.7895	0.7923	0.7950	0.7977	0.8004	0.8031	0.8058	0.8084	0.8111	0.8137
0.9	0.8137	0.8163	0.8189	0.8214	0.8239	0.8264	0.8289	0.8314	0.8338	0.8362	0.8386
1.0	0.8386	0.8410	0.8434	0.8458	0.8481	0.8504	0.8527	0.8549	0.8571	0.8593	0.8615
1.1	0.8615	0.8636	0.8657	0.8677	0.8697	0.8717	0.8736	0.8755	0.8773	0.8791	0.8810
1.2	0.8810	0.8828	0.8846	0.8863	0.8881	0.8898	0.8915	0.8932	0.8948	0.8964	0.8979
1.3	0.8979	0.8994	0.9009	0.9023	0.9038	0.9052	0.9066	0.9079	0.9092	0.9105	0.9117
1.4	0.9117	0.9129	0.9141	0.9152	0.9163	0.9174	0.9184	0.9194	0.9204	0.9213	0.9222
1.5	0.9222	0.9231	0.9240	0.9249	0.9257	0.9265	0.9273	0.9281	0.9289	0.9296	0.9304
1.6	0.9304	0.9311	0.9318	0.9325	0.9332	0.9339	0.9345	0.9351	0.9357	0.9363	0.9368
1.7	0.9368	0.9374	0.9379	0.9384	0.9389	0.9394	0.9398	0.9403	0.9407	0.9411	0.9415
1.8	0.9415	0.9419	0.9423	0.9427	0.9431	0.9435	0.9438	0.9441	0.9444	0.9447	0.9450
1.9	0.9450	0.9453	0.9456	0.9459	0.9462	0.9464	0.9467	0.9469	0.9471	0.9473	0.9475
2.0	0.9475	0.9477	0.9479	0.9480	0.9481	0.9482	0.9483	0.9484	0.9485	0.9485	0.9486
2.1	0.9486	0.9486	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487
2.2	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487
2.3	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487
2.4	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487
2.5	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487
2.6	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487
2.7	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487
2.8	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487
2.9	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487
3.0	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487

3

2. Use confidence intervals to estimate population parameters

92% Confidence Interval for μ

$$\bar{x} \pm z_{0.04}^* \frac{\sigma}{\sqrt{n}}$$

*right conditions must be met to use this equation

How to find the critical value $z_{0.04}^*$:

- Look up the positive Z-score in Standard Normal Distribution (z-tables) that creates a right tail with area/probability 0.04 (aka: left area of 1 - 0.04 = 0.96).

• Z-Distribution aka: Standard Normal Distribution
 • Aka: N(0,1)

Normal probability table

Normal distribution		Normal deviate (area of Z)									
Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5200	0.5240	0.5280	0.5320	0.5360	0.5400
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753	0.5793
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141	0.6179
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517	0.6554
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879	0.6915
0.5	0.6915	0.6950	0.6985	0.7020	0.7054	0.7089	0.7124	0.7159	0.7193	0.7227	0.7261
0.6	0.7261	0.7294	0.7327	0.7359	0.7391	0.7423	0.7454	0.7485	0.7516	0.7547	0.7578
0.7	0.7578	0.7608	0.7638	0.7667	0.7696	0.7725	0.7754	0.7782	0.7811	0.7839	0.7867
0.8	0.7867	0.7895	0.7923	0.7950	0.7977	0.8004	0.8031	0.8058	0.8084	0.8111	0.8137
0.9	0.8137	0.8163	0.8189	0.8214	0.8239	0.8264	0.8289	0.8314	0.8338	0.8362	0.8386
1.0	0.8386	0.8410	0.8434	0.8458	0.8481	0.8504	0.8527	0.8549	0.8571	0.8593	0.8615
1.1	0.8615	0.8636	0.8657	0.8677	0.8697	0.8717	0.8736	0.8755	0.8773	0.8791	0.8810
1.2	0.8810	0.8828	0.8846	0.8863	0.8881	0.8898	0.8915	0.8932	0.8948	0.8964	0.8979
1.3	0.8979	0.8994	0.9009	0.9023	0.9038	0.9052	0.9066	0.9079	0.9092	0.9105	0.9117
1.4	0.9117	0.9129	0.9141	0.9152	0.9163	0.9174	0.9184	0.9194	0.9204	0.9213	0.9222
1.5	0.9222	0.9231	0.9240	0.9249	0.9257	0.9265	0.9273	0.9281	0.9289	0.9296	0.9304
1.6	0.9304	0.9311	0.9318	0.9325	0.9332	0.9339	0.9345	0.9351	0.9357	0.9363	0.9368
1.7	0.9368	0.9374	0.9379	0.9384	0.9389	0.9394	0.9398	0.9403	0.9407	0.9411	0.9415
1.8	0.9415	0.9419	0.9423	0.9427	0.9431	0.9435	0.9438	0.9441	0.9444	0.9447	0.9450
1.9	0.9450	0.9453	0.9456	0.9459	0.9462	0.9464	0.9467	0.9469	0.9471	0.9473	0.9475
2.0	0.9475	0.9477	0.9479	0.9480	0.9481	0.9482	0.9483	0.9484	0.9485	0.9485	0.9486
2.1	0.9486	0.9486	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487
2.2	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487
2.3	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487
2.4	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487
2.5	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487
2.6	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487
2.7	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487
2.8	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487
2.9	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487
3.0	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487	0.9487

3

2. Use confidence intervals to estimate population parameters

92% Confidence Interval for μ

$$\bar{x} \pm z_{0.04}^* \frac{\sigma}{\sqrt{n}}$$

*right conditions must be met to use this equation

How to find the critical value $z_{0.04}^*$:

- Look up the positive Z-score in Standard Normal Distribution (z-tables) that creates a right tail with area/probability **0.04 (aka: left area of $1 - 0.04 = 0.96$)**.
- Z-Distribution aka: Standard Normal Distribution
- Aka: $N(0,1)$

Normal probability table only gives areas to left of z-scores

3

2. Use confidence intervals to estimate population parameters

92% Confidence Interval for μ

$$\bar{x} \pm z_{0.04}^* \frac{\sigma}{\sqrt{n}}$$

*right conditions must be met to use this equation

What does the critical value represent in a 92% confidence interval for μ :

- Number of standard deviations ($\frac{\sigma}{\sqrt{n}}$) \bar{x} needs to be away from μ (on both sides) of the **sampling distribution** to capture the **middle 92%** of all possible randomly sampled \bar{x} 's.

3

Outline

What is the margin of error for a population mean?

Outline

What is the margin of error for a population mean?

*right conditions must be met to use this equation

Margin of Error for a $(1 - \alpha)$ Confidence Interval for μ

$$\text{Margin of Error} = z_{\alpha/2}^* \frac{\sigma}{\sqrt{n}}$$

3. Critical value depends on the confidence level

Clicker question

What is the critical value (Z^*) for a confidence interval at the 91% confidence level?

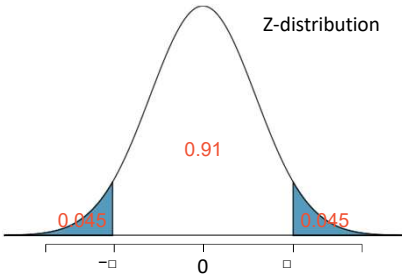
- (a) $Z^* = 1.34$
- (b) $Z^* = 1.65$
- (c) $Z^* = 1.70$
- (d) $Z^* = 1.96$
- (e) $Z^* = 2.33$

4

3. Critical value depends on the confidence level

Clicker question

What is the critical value (Z^*) for a confidence interval at the 91% confidence level?



Z-distribution

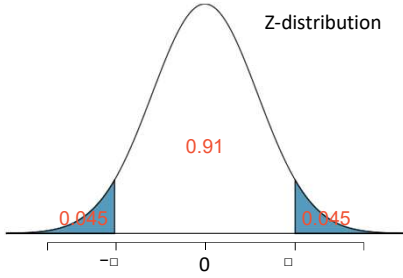
- (a) $Z^* = 1.34$
- (b) $Z^* = 1.65$
- (c) $Z^* = 1.70$
- (d) $Z^* = 1.96$
- (e) $Z^* = 2.33$

4

3. Critical value depends on the confidence level

Clicker question

What is the critical value (Z^*) for a confidence interval at the 91% confidence level?



Z-distribution

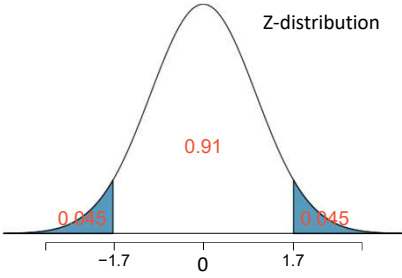
- (a) $Z^* = 1.34$
- (b) $Z^* = 1.65$
- (c) $Z^* = 1.70$
- (d) $Z^* = 1.96$
- (e) $Z^* = 2.33$

4

3. Critical value depends on the confidence level

Clicker question

What is the critical value (Z^*) for a confidence interval at the 91% confidence level?



Z-distribution

- (a) $Z^* = 1.34$
- (b) $Z^* = 1.65$
- (c) $Z^* = 1.70$
- (d) $Z^* = 1.96$
- (e) $Z^* = 2.33$

4

(1 - α) Confidence Interval
 = (point estimate) ± z*_{α/2} (standard error)

*if using the z-distribution

Standard Normal Distribution
 Z ~ N(0,1)

The diagram shows a bell-shaped curve centered at 0. The area under the curve between -z*_{α/2} and z*_{α/2} is labeled 1-α. The two tail areas outside these points are shaded blue and each labeled α/2.

Outline

1. Housekeeping
2. Main ideas
 1. [Making a Confidence Interval](#): Use confidence intervals to estimate population parameters
 2. [Conditions for using the CLT to make a Conf. Interval](#): Statistical inference methods based on the CLT depend on the same conditions as the CLT
 3. [Common misconceptions about confidence intervals](#)
 4. [What sample size do we need?](#) Calculate the sample size a priori to achieve desired margin of error
3. Summary

Outline

Case 1: If we know σ, when can we create a confidence interval (or margin of error) for μ (using CLT methods)?

Outline

Case 1: If we know σ, when can we create a confidence interval (or margin of error) for μ (using CLT methods)?

▲ Aka: use these equations

(1 - α) Confidence Interval for μ

$$\bar{x} \pm z^*_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Margin of Error in a (1 - α) Confidence Interval for μ

$$ME = z^*_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

NEW 2. Use confidence intervals to estimate population parameters

Case 1: We know population standard deviation σ .

Central Limit Theorem Conditions

- Independence:** Sampled observations must be independent.
 - * This is difficult to verify, but is more likely if
 - random sampling/assignment is used, and,
 - if sampling without replacement, $n < 10\%$ of the population.
- Sample size/skew:**
 - Either the population distribution is normal OR
 - $n > 30$ and the population distribution is not extremely skewed (the more skewed the distribution, the higher n necessary for the CLT to apply).

Outline

Case 2: If we don't know σ , when can we create a confidence interval (or margin of error) for μ (using CLT methods)?

Outline

Case 2: If we don't know σ , when can we create a confidence interval (or margin of error) for μ (using CLT methods)?

Aka: use these equations

$(1 - \alpha)$ Confidence Interval for μ

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

Margin of Error in a $(1 - \alpha)$ Confidence Interval for μ

$$ME = z_{\alpha/2}^* \frac{\sigma}{\sqrt{n}}$$

NEW 2. Use confidence intervals to estimate population parameters

Case 2: We don't know population standard deviation σ .

Central Limit Theorem Conditions AND **Condition: n needs to be a good enough approximation for σ**

- Independence:** Sampled observations must be independent.
 - * This is difficult to verify, but is more likely if
 - random sampling/assignment is used, and,
 - if sampling without replacement, $n < 10\%$ of the population.
- Sample size/skew:**
 - Either the population distribution is normal OR
 - $n > 30$ and the population distribution is not extremely skewed (the more skewed the distribution, the higher n necessary for the CLT to apply).

$n > 30$

NEW 2. Use confidence intervals to estimate population parameters

Case 2: We don't know population standard deviation σ .

Central Limit Theorem Conditions AND **Condition: s needs to be a good enough approximation for σ**

1. **Independence:** Sampled observations must be independent.

* This is difficult to verify, but is more likely if

- random sampling/assignment is used, and,
- if sampling without replacement, $n < 10\%$ of the population.

2. **Sample size/skew:**

1. Either the population distribution is normal OR
2. $n > 30$ and the population distribution is not extremely skewed (the more skewed the distribution, the higher n necessary for the CLT to apply).

$n > 30$

Combining both sets of conditions is equivalent to saying...

NEW 2. Use confidence intervals to estimate population parameters

Case 2: You don't know population standard deviation σ .

Central Limit Theorem Conditions AND **Condition: s needs to be a good enough approximation for σ**

1. **Independence:** Sampled observations must be independent.

* This is difficult to verify, but is more likely if

- random sampling/assignment is used, and,
- if sampling without replacement, $n < 10\%$ of the population.

2. **Sample size/skew:**

2. $n > 30$ and the population distribution is not extremely skewed (the more skewed the distribution, the higher n necessary for the CLT to apply).

Outline

1. Housekeeping

2. Main ideas

1. Making a Confidence Interval: NEW Use confidence intervals to estimate population parameters
2. Conditions for using the CLT to make a Conf. Interval: Statistical inference methods based on the CLT depend on the same conditions as the CLT
3. Common misconceptions about confidence intervals ☹️
4. What sample size do we need? NEW Calculate the sample size a priori to achieve desired margin of error

3. Summary

Outline

XX% Confidence Interval Interpretation:

“We are XX% confident that the true *insert population parameter* is between and *insert confidence interval upper bound*.”

Outline

XX% Confidence Interval *Interpretation:*

“We are XX% confident that the true *insert population parameter* is between *insert confidence interval lower bound* and *insert confidence interval upper bound*.

What does this mean?

Outline

XX% Confidence Interval *Interpretation:*

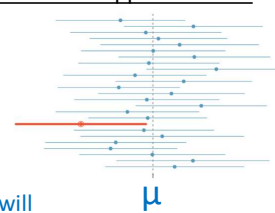
“We are XX% confident that the true *insert population parameter* is between *insert confidence interval lower bound* and *insert confidence interval upper bound*.

What does this mean?

If we do the following many times

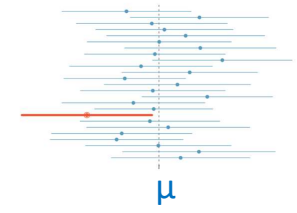
1. Randomly sample n observations.
2. Calculate the sample statistic (ex \bar{x}).
3. Calculate a XX% confidence interval using this sample statistic (ex \bar{x}).

Then XX% of these confidence intervals will contain the true population parameter (ex μ).



Common misconceptions about confidence intervals

1. The confidence level of a confidence interval is the probability that the true population parameter is in the confidence interval you construct for a single sample. *The confidence level is equal to the proportion of random samples that result in confidence intervals that contain the true pop. parameter.*



5

Common misconceptions about confidence intervals

1. The confidence level of a confidence interval is the probability that the true population parameter is in the confidence interval you construct for a single sample. *The confidence level is equal to the proportion of random samples that result in confidence intervals that contain the true pop. parameter.*
2. A narrower confidence interval is always better. *This is incorrect since the width is a function of both the confidence level and the standard error.*

$$\bar{x} \pm \underbrace{z_{\alpha/2}}_{\text{small} \rightarrow \text{lower confidence level}} \underbrace{\frac{\sigma}{\sqrt{n}}}_{\text{small} \rightarrow \text{lower variability in sample means}}$$

5

Common misconceptions about confidence intervals

- The confidence level of a confidence interval is the probability that the true population parameter is in the confidence interval you construct for a single sample.
The confidence level is equal to the proportion of random samples that result in confidence intervals that contain the true pop. parameter.
- A narrower confidence interval is always better.
This is incorrect since the width is a function of both the confidence level and the standard error.
- A wider interval means less confidence.
This is incorrect since it is possible to make very precise statements with very little confidence.

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

Annotations:
 - $z_{\alpha/2}^*$: small \rightarrow lower confidence level
 - $\frac{\sigma}{\sqrt{n}}$: small \rightarrow lower variability in sample means

5

Outline

- Housekeeping
- Main ideas
 - Making a Confidence Interval: Use confidence intervals to estimate population parameters
 - Conditions for using the CLT to make a Conf. Interval: Statistical inference methods based on the CLT depend on the same conditions as the CLT
 - Common misconceptions about confidence intervals
 - What sample size do we need? Calculate the sample size a priori to achieve desired margin of error
- Summary

4. Calculate the sample size a priori to achieve desired margin of error

Know

$ME \leq \#$ $ME = z^* \frac{s}{\sqrt{n}}$

Want

So if we know the desired ME , and confidence level (and hence z^*), and the sample standard deviation, we can solve for n .

6

Application exercise: 3.1 Confidence interval for a single mean

See course website for details.

7

Summary of main ideas

1. Statistical inference methods based on the CLT depend on the same conditions as the CLT
2. Use confidence intervals to estimate population parameters
3. Critical value depends on the confidence level
4. Calculate the sample size a priori to achieve desired margin of error

8