STA 1013 : Statistics through Examples

Lecture 26: Interval Estimation 2

Hwiyoung Lee

November 13, 2019

Department of Statistics, Florida State University

Confidence Interval for μ (σ known)

Confidence Interval for population mean μ

Once a confidence level is specified, the interval estimate can be calculated using formulas;

- $(1-\alpha)\%$ **Confidence interval estimate of** μ

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

- which gives us the interval :

$$\left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\ ,\ \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

Confidence Intervals for population mean μ

- 90% Confidence Interval :

$$\left(\bar{X} \pm z_{0.1/2}\frac{\sigma}{\sqrt{n}}\right) = \left(\bar{X} - 1.645\frac{\sigma}{\sqrt{n}}\ ,\ \bar{X} + 1.645\frac{\sigma}{\sqrt{n}}\right)$$

- 95% Confidence Interval :

$$\left(\bar{X} \pm z_{0.05/2}\frac{\sigma}{\sqrt{n}}\right) = \left(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}}\ ,\ \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right)$$

- 99% Confidence Interval :

$$\left(\bar{X} \pm z_{0.01/2}\frac{\sigma}{\sqrt{n}}\right) = \left(\bar{X} - 2.576\frac{\sigma}{\sqrt{n}}\ ,\ \bar{X} + 2.576\frac{\sigma}{\sqrt{n}}\right)$$

Confidence Interval for population mean μ

Interpreting a 95% confidence interval

If we could repeat the sampling process many times, we would find that 95% of the confidence intervals would contain the population mean and 5% would not.

Interpreting a 95% confidence interval

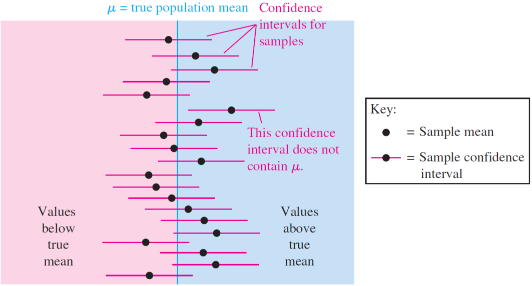


Figure 1: This figure illustrates the idea behind confidence intervals. The central vertical line represents the true population mean, μ. Each of the 20 horizontal lines represents the 95% confidence interval for a particular sample, with the sample mean marked by the dot in the center of the confidence interval. With a 95% confidence interval, we expect that 95% of all samples will give a confidence interval that contains the population mean, as is the case in this figure, for 19 of the 20 confidence intervals do indeed contain the population mean. We expect that the population mean will not be within the confidence interval in 5% of the cases; here, 1 of the 20 confidence intervals (the sixth from the top) does not contain the population mean.

Calculator - ZInterval

(ZInterval) - σ known, Statistics are given

Confidence Interval from statistics

1. Press the $\boxed{\text{stat}}$ and highlight **TESTS**
2. Scroll down to 7: **ZInterval**
3. Inpt : Highlight **Stats**
4. Enter values for
   - σ : population standard deviation
   - $\bar{X}$: Sample mean
   - n : Sample size
   - C-level : Confidence Level (Ex : 0.90. 0.95, 0.99)

Examples

A simple random sample of 50 items from a population with $\sigma = 6$ resulted in a sample mean of 32.

1. Provide a 90% confidence interval for the population mean.
2. Provide a 95% confidence interval for the population mean.
3. Provide a 99% confidence interval for the population mean.

Examples

The undergraduate grade point average (GPA) for students admitted to the top graduate business schools was 3.37. Assume this estimate was based on a sample of 120 students admitted to the top schools. Using past years' data, the population standard deviation can be assumed known with $\sigma = 0.28$.

1. What is the 90% confidence interval estimate of the mean undergraduate GPA for students admitted to the top graduate business schools?

2. What is the 95% confidence interval estimate of the mean undergraduate GPA for students admitted to the top graduate business schools?

(ZInterval) - σ known, Data is given

Confidence Interval from Data (If the problem only gives a list of data values with no statistics)

1. Press the $\boxed{\text{stat}}$ and highlight **TESTS**
2. Scroll down to 7: **ZInterval**
3. Inpt : Highlight **Data**
4. Enter values for σ, List, and C-level (Confidence Level)
    - σ : Population Standard deviation
    - List : You will need to enter the values in a list (preferably in L_1) beforehand
    - Freq : should always be set to 1
    - C-level : Confidence Level (Ex : 0.90. 0.95, 0.99)

Example

A doctor conducts a small survey with a random sample of his patients, measuring their cholesterol levels, and standard deviation of cholesterol levels can be assumed known with $\sigma = 1.75$. Here is his data (the measurements are in m.mol/L):

3.6	6.9	5.1	4.2	5.5	7.2	3.0	5.8	4.9	9.9
7.1	5.4	6.2	4.5	6.3	8.2	5.7	4.4	7.9	3.2

1. Find an 90% confidence interval for the mean cholesterol level of patients

2. Find an 95% confidence interval for the mean cholesterol level of patients

Summary (ZInterval)

1. Press the $\boxed{\text{stat}}$ and highlight **TESTS**
2. Scroll down to 7: **ZInterval**

- **Confidence Interval from statistics**
  3. Inpt : Highlight **Stats**
  4. Enter values for $\sigma, \bar{X}, n$, and C-level (Confidence Level)

- **Confidence Interval from Data**
  3. Inpt : Highlight **Data**
  4. Enter values for σ, List, and C-level (Confidence Level)
  5. Freq : should always be set to 1

Confidence Interval for μ (σ unknown)

Population standard deviation σ is unknown case

A doctor conducts a small survey with a random sample of his patients, measuring their cholesterol levels. Here is his data (the measurements are in m.mol/L):

3.6	6.9	5.1	4.2	5.5	7.2	3.0	5.8	4.9	9.9
7.1	5.4	6.2	4.5	6.3	8.2	5.7	4.4	7.9	3.2

- In the above example, the population standard deviation σ is not given
- How can we construct the Confidence Interval for μ ?

$$\bar{X} \pm z_{\alpha/2} \frac{\overbrace{\sigma}^{\textbf{unknown}}}{\sqrt{n}}$$

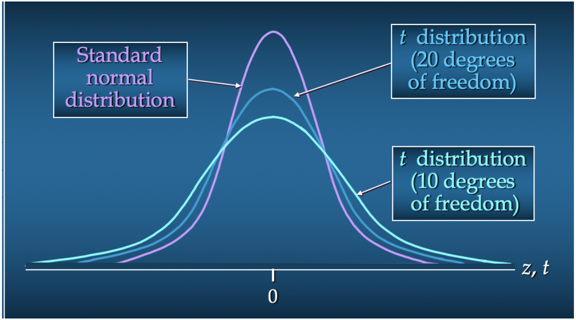
C.I for population mean (σ unknown case)

- In real world problems, both μ, and σ are **unknown**
- Thus, σ unknown case is more popular, and general
- If the population standard deviation σ is not given prior to sampling, **Use the sample standard deviation $\mathbf{s}$ to estimate σ**
- In this case, the interval estimate for μ is based on the t**-distribution**

t-distribution

- A t-distribution depends on a parameter known as the **degrees of freedom**
- Degrees of freedom refer to the sample size minus 1 (**d.f = n-1**)
- A t-distribution with more degrees of freedom has less dispersion
- As the number of degrees of freedom increases, the difference between the t distribution and the standard normal probability distribution becomes smaller and smaller

t-distribution



- t-distribution is similar to the normal distribution with its bell shape but has heavier tail

Confidence Interval (σ unknown)

- $(1-\alpha)\%$ **Confidence interval estimate of** μ

$$\bar{X} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}$$

- which gives us the interval :

$$\left(\bar{X} - t_{\alpha/2}\frac{s}{\sqrt{n}}\ ,\ \bar{X} + t_{\alpha/2}\frac{s}{\sqrt{n}}\right)$$

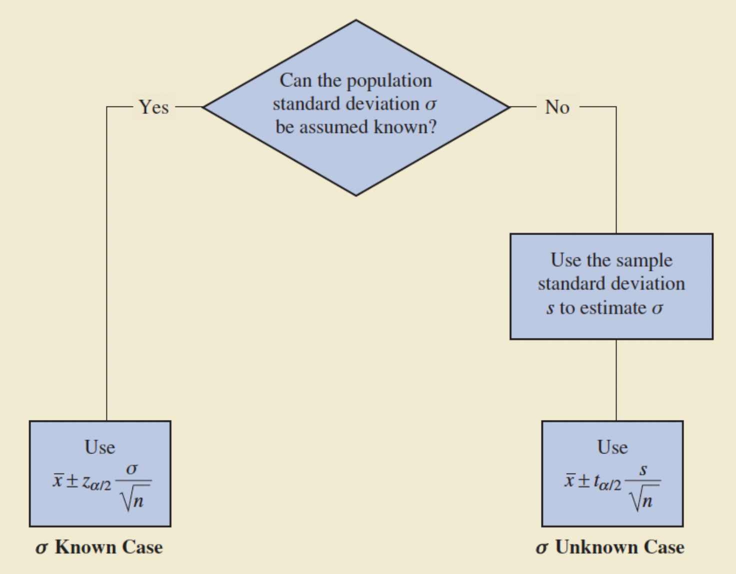
- $t_{\alpha/2}$: the t value providing an area of $\alpha/2$ in the upper tail of a t-distribution with $n-1$ degrees of freedom
- Note : $t_{\alpha/2}$ can be computed by **invT**$(1-\alpha/2, \textbf{df})$
- s : sample standard deviation

Critical values for t distribution

Critical Values (t*)			
	Confidence Level		
n − 1	0.900	0.950	0.990
10	1.812	2.228	3.169
20	1.725	2.086	2.845
30	1.697	2.042	2.750
40	1.684	2.021	2.704
50	1.676	2.009	2.678
60	1.671	2.000	2.660
70	1.667	1.994	2.648
80	1.664	1.990	2.639
90	1.662	1.987	2.632
100	1.660	1.984	2.626

- As sample size increases (df ↑), $t_{df,\alpha/2}$ converges to $z_{\alpha/2}$
- $z_{0.1/2} = 1.645, \quad z_{0.05/2} = 1.96, \quad z_{0.01/2} = 2.576$

Summary of Interval Estimation Procedures



Calculator - TInterval

(TInterval) - σ Unknown, Statistics are given

Confidence Interval from statistics

1. Press the $\boxed{\text{stat}}$ and highlight **TESTS**
2. Scroll down to 8: **TInterval**
3. Inpt : Highlight **Stats**
4. Enter values for
   - $\bar{X}$: Sample mean
   - S_X : Sample standard deviation
   - n : Sample size
   - C-level : Confidence Level (Ex : 0.90. 0.95, 0.99)

Example

A sample of 70 households provided the credit card balances shown in the table. The sample mean $\bar{X}$ is \$9,312 and the sample standard deviation (s) is \$4,007.

1. Estimate the mean credit card debt of all customers in the U.S. with a 90% confidence.

2. Estimate the mean credit card debt of all customers in the U.S. with a 95% confidence.

(TInterval) - σ Unknown, Data is given

Confidence Interval from Data (If the problem only gives a list of data values with no statistics)

1. Press the $\boxed{\text{stat}}$ and highlight **TESTS**
2. Scroll down to 8: **TInterval**
3. Inpt : Highlight **Data**
4. Enter values for List, and C-level (Confidence Level)
    - List : You will need to enter the values in a list (preferably in L_1) beforehand
    - Freq : should always be set to 1
    - C-level : Confidence Level (Ex : 0.90. 0.95, 0.99)

Example

cholesterol levels

A doctor conducts a small survey with a random sample of his patients, measuring their cholesterol levels. Here is his data (the measurements are in m.mol/L):

3.6	6.9	5.1	4.2	5.5	7.2	3.0	5.8	4.9	9.9
7.1	5.4	6.2	4.5	6.3	8.2	5.7	4.4	7.9	3.2

1. Find an 90% confidence interval for the mean cholesterol level of patients
2. Find an 95% confidence interval for the mean cholesterol level of patients

Example

Weight of Bears

The health of the bear population in Yellowstone National Park is monitored by periodic measurements taken from anesthetized bears. A sample of the weights (pounds) of such bears is given below.

80	344	416	348	166	220	262	360	204
144	332	34	140	180	105	166	204	26
120	436	125	132	90	40	220	46	154
116	182	150	65	356	316	94	86	150

Find a 95% confidence interval estimate of the mean of the population of all such bear weights.