ECO 391 Economics and Business Statistics

Lecture 4: Interval Estimation

Xiaozhou Ding

February 6, 2019

Overview

1. Confidence Interval for the Population Mean When σ Is Known
    - Constructing a Confidence Interval for μ When σ is Known

2. Confidence Interval for the Population Mean When σ is Unknown
    - Constructing a Confidence Interval for μ When σ is Unknown

Confidence Interval for the Population Mean When σ Is Known

Several Concepts

Definition

Point estimator is a function of the random sample used to make inferences about the value of an unknown population parameter.

For example, $\overline{X}$ is a point estimator for μ.

Definition

Point estimate is the value of the point estimator derived from a given sample.

For example, $\bar{x} = 60,000$ is a point estimate of the mean starting salary of business graduates.

Several Concepts

Definition

Confidence interval provides a range of values that, with a certain level of confidence, contains the population parameter of interest. It is also referred to as an interval estimate. Construct a confidence interval as:

$$\textbf{Point estimate} \pm \textbf{Margin of error}$$

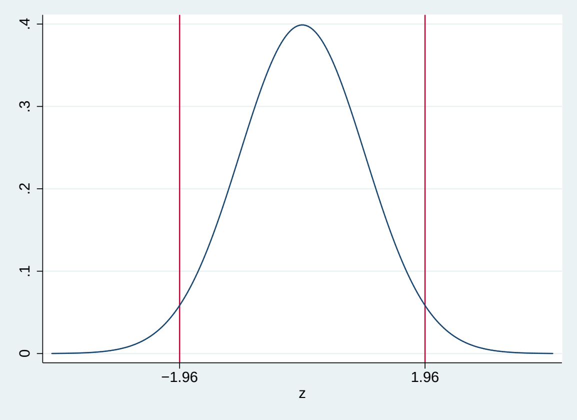
Margin of error accounts for the variability of the estimator and the desired confidence level of the interval.

Constructing a Confidence Interval for μ When σ is Known

Consider a standard normal random variable Z,

$$P(-1.96 \leq Z \leq 1.96) = 0.95$$

as illustrated here.



Constructing a Confidence Interval for μ When σ is Known

Since

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}},$$

we get

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

which, after algebraically manipulating, is equal to

$$P\left(\mu - 1.96 \times \frac{\sigma}{\sqrt{n}} \leq \overline{X} \leq \mu + 1.96 \times \frac{\sigma}{\sqrt{n}}\right) = 0.95.$$

Constructing a Confidence Interval for μ When σ is Known

- Note that
$$P\left(\mu - 1.96 \times \frac{\sigma}{\sqrt{n}} \leq \overline{X} \leq \mu + 1.96 \times \frac{\sigma}{\sqrt{n}}\right) = 0.95$$
implies there is a 95% probability that the sample mean $\overline{X}$ will fall within the interval
$$\mu \pm 1.96\sigma/\sqrt{n}.$$
- Thus, if samples of size n are drawn repeatedly from a given population, 95% of the computed sample means, $\bar{x}$'s, will fall within the interval and the remaining 5% will fall outside the interval.

Constructing a Confidence Interval for μ When σ is Known

- Since we do not know μ, we cannot determine if a particular $\bar{x}$ falls within the interval or not.
- However, we do know that $\bar{x}$ will fall within the interval $\mu \pm 1.96\sigma/\sqrt{n}$ if and only if μ falls within the interval $\bar{x} \pm 1.96\sigma/\sqrt{n}$.
- This will happen 95% of the time given the interval construction. Thus, this is a 95% confidence interval for the population mean.

Constructing a Confidence Interval for μ When σ is Known

- Level of significance (i.e., probability of error = α).
- Confidence coefficient = $(1 - \alpha)$
  $\alpha = 1 -$ confidence coefficient
- A $100(1 - \alpha)\%$ confidence interval of the population mean μ when the standard deviation σ is known is computed as $\bar{x} \pm z_{\alpha/2}\sigma/\sqrt{n}$, or equivalently,

$$[\bar{x} - z_{\alpha/2}\sigma/\sqrt{n}, \bar{x} + z_{\alpha/2}\sigma/\sqrt{n}].$$

Constructing a Confidence Interval for μ When σ is Known

- $z_{\alpha/2}$ is the z value associated with the probability of $\alpha/2$ in the upper-tail.
- Confidence intervals:
  - 90%, $\alpha = 0.1$, $\alpha/2 = 0.05$, $z_{\alpha/2} = z_{.05} = 1.645$.
  - 95%, $\alpha = 0.05$, $\alpha/2 = 0.025$, $z_{\alpha/2} = z_{.025} = 1.96$.
  - 99%, $\alpha = 0.01$, $\alpha/2 = 0.005$, $z_{\alpha/2} = z_{.005} = 2.575$.

Interpreting a Confidence Interval

- Interpreting a confidence interval requires care.
- Incorrect: The probability that μ falls in the interval is 0.95.
- Correct: If numerous samples of size n are drawn from a given population, then 95% of the intervals formed by the formula $\bar{x} \pm z_{\alpha/2}\sigma/\sqrt{n}$ will contain μ.
  - Since there are many possible samples, we will be right 95% of the time, thus giving us 95% confidence.

Confidence Interval (CI) Calculation

Example

To estimate the mean age of subscribers to Sports Illustrated magazine, a random sample of 100 subscribers is taken.

- The sample mean=31
- The population variance=144
- Calculate a 95% confidence interval for μ

1. The sample mean $\overline{X} = 31$
2. $n = 100$, $(1 - \alpha) = 0.95$, so $\alpha = 0.05$.
   $\sigma^2 = 144 \Rightarrow \sigma = 12$.
   Therefore,
   $$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{12}{10} = 1.2$$
3. Find the values $+Z_{\alpha/2}$ and $-Z_{\alpha/2}$
   At 95% confidence, $\alpha = 0.05$. $Z_{\alpha/2} = Z_{0.05/2} = Z_{0.025}$. So 1.96.
4. The two-sided confidence interval is:
   $$\left[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$
   $$\left[31 - 1.96\frac{12}{\sqrt{100}}, \bar{3}1 + 1.96\frac{12}{\sqrt{100}}\right] = [28.65, 33.35]$$
5. Provide a full interpretation of your result.
   Based on this sample, I can say with 95% confidence that the true population mean age of subscribers to Sports Illustrated is between 28.65 and 33.35 years.

Factors that Influence the Width of a Confidence Interval

- **Sample size** n.
  For a given confidence level $100(1-\alpha)\%$ and population standard deviation σ, the smaller the sample size n, the wider the width of the interval.
- **Standard deviation** σ.
  For a given confidence level $100(1-\alpha)\%$ and sample size n, the greater the population standard deviation σ, the wider the confidence interval.
- **Confidence level** $100(1-\alpha)\%$.
  For a given sample size n and population standard deviation σ, the greater the confidence level $100(1-\alpha)\%$, the wider the width of the interval.

Caution: increasing the CI increases the likelihood of capturing μ, but decreases precision.

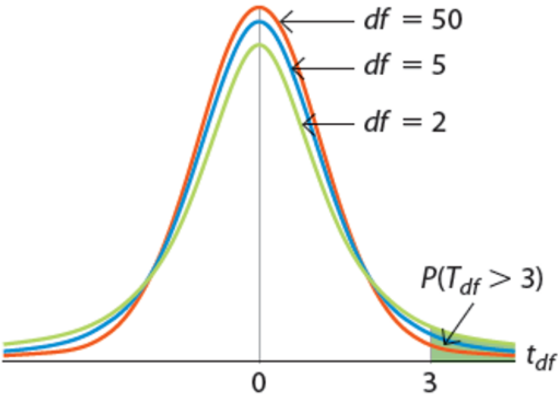
Confidence Interval for the Population Mean When σ is Unknown

The t-Distribution

- If repeated samples of size n are taken from a normal population with a finite variance, then the statistic T follows the t-distribution with $(n-1)$ degrees of freedom, df.
- Degrees of freedom=# observations (n)-1
- *df* determines the extent of the broadness of the tails of the distribution; the fewer the degrees of freedom, the broader the tails.

Summary of the t_{df}-Distribution

- Bell-shaped and symmetric around 0 with asymptotic tails (the tails get closer and closer to the horizontal axis, but never touch it).
- Has slightly broader tails than the z distribution.
- Consists of a family of distributions where the actual shape of each one depends on the df. As df increases, the t_{df} distribution becomes similar to the z-distribution; it is identical to the z-distribution when df approaches infinity. `http://demonstrations.wolfram.com/ComparingNormalAndStudentsTDistributions/`


df = 50
df = 5
df = 2
P(T_df > 3)
t_df
0
3

Degrees of freedom (ν)	Amount of area in one tail (α)							
	0.0005	**0.001**	**0.005**	**0.010**	**0.025**	**0.050**	**0.100**	**0.200**
1	636.6192	318.3088	63.65674	31.82052	12.70620	6.313752	3.077684	1.376382
2	31.59905	22.32712	9.924843	6.964557	4.302653	2.919986	1.885618	1.060660
3	12.92398	10.21453	5.840909	4.540703	3.182446	2.353363	1.637744	0.978472
4	8.610302	7.173182	4.604095	3.746947	2.776445	2.131847	1.533206	0.940965
5	6.868827	5.893430	4.032143	3.364930	2.570582	2.015048	1.475884	0.919544
6	5.958816	5.207626	3.707428	3.142668	2.446912	1.943180	1.439756	0.905703
7	5.407883	4.785290	3.499483	2.997952	2.364624	1.894579	1.414924	0.896030
8	5.041305	4.500791	3.355387	2.896459	2.306004	1.859548	1.396815	0.888890
9	4.780913	4.296806	3.249836	2.821438	2.262157	1.833113	1.383029	0.883404
10	4.586894	4.143700	3.169273	2.763769	2.228139	1.812461	1.372184	0.879058
11	4.436979	4.024701	3.105807	2.718079	2.200985	1.795885	1.363430	0.875530
12	4.317791	3.929633	3.054540	2.680998	2.178813	1.782288	1.356217	0.872609
13	4.220832	3.851982	3.012276	2.650309	2.160369	1.770933	1.350171	0.870152
14	4.140454	3.787390	2.976843	2.624494	2.144787	1.761310	1.345030	0.868055
15	4.072765	3.732834	2.946713	2.602480	2.131450	1.753050	1.340606	0.866245
16	4.014996	3.686155	2.920782	2.583487	2.119905	1.745884	1.336757	0.864667
17	3.965126	3.645767	2.898231	2.566934	2.109816	1.739607	1.333379	0.863279

Example

Compute $t_{\alpha/2,df}$ for $\alpha/2 = 0.025$ using 2, 5, and 50 degrees of freedom.

Turning to the Student's t Distribution table in Appendix A, or using the T.INV() function:

- For $df = 2$, $t_{0.025,2} = 4.303$
- For $df = 5$, $t_{0.025,5} = 2.571$
- For $df = 50$, $t_{0.025,50} = 2.009$

Note that the t_{df} values change with the degrees of freedom. Further, as df increases, the t_{df} distribution begins to resemble the z distribution.

Constructing a Confidence Interval for μ When σ is Unknown

A $100(1-\alpha)\%$ confidence interval of the population mean μ when the population standard deviation σ is not known is computed as

$$\bar{x} \pm t_{\alpha/2,df} \frac{s}{\sqrt{n}}$$

or equivalently,

$$\left[\bar{x} - t_{\alpha/2,df} \frac{s}{\sqrt{n}}, \bar{x} + t_{\alpha/2,df} \frac{s}{\sqrt{n}}\right],$$

where s is the sample standard deviation.

Constructing a CI if Population Variance is Unknown

1. Determine the sample mean
2. Record the sample size n, the level of confidence $(1-\alpha)$, and the sample standard deviation, s.
3. Find the values $+t_{\alpha/2,v}$, and $-t_{\alpha/2,v}$ with degrees of freedom $v = n-1$ from t-table.
4. The two-sided confidence interval is

$$\left[\bar{x} - t_{\alpha/2,df}\frac{s}{\sqrt{n}}, \bar{x} + t_{\alpha/2,df}\frac{s}{\sqrt{n}}\right]$$

5. Interpret your result. We are $100(1-\alpha)\%$ confident that the true population mean falls in this interval.

Example

The Summerhill Trucking Company owns a large fleet of rental trucks. Many of the trucks need substantial repairs from time to time. The company president takes a random sample of 64 trucks and finds the sample mean annual repair bill is $1,245 and the sample standard deviation, S=$288. However, the president does not have information regarding the population parameters.

- Calculate a 99% confidence interval for μ.

1. The sample mean is \$1,245
2. $n = 64$, $(1 - \alpha) = 0.99 \Rightarrow \alpha = 0.01$. $S = 288$, $se = \frac{S}{\sqrt{n}} = \frac{288}{\sqrt{64}} = 36$.
3. t values: $\alpha/2 = 0.005$, $v = 64 - 1 = 63$, $t_{\alpha/2,63} = t_{0.005,63}$. You can find it in the t-table at the intersection of column 60 and row 0.005, so $t = 2.660$.
4. The two-sided confidence interval is:

$$\left[1245 - 2.660\frac{288}{\sqrt{64}}, 1245 + 2.660\frac{288}{\sqrt{64}}\right] = [1149.24, 1340.76]$$

5. Interpret your result: We can say with 99% confidence, that the true population mean annual repair bill for Summerhill's trucks is between \$1,149.24 and \$1,340.76.

Using Excel to construct confidence intervals. The easiest way to estimate the mean when the population standard deviation is unknown is as follows: