Univariate Statistical Analysis

Lecture 6

Confidence Interval for Population Mean (Chapter 9, Section 9.3)

Today

- Define Confidence Interval
- Confidence Interval for a Population Mean

Revisit Lecture 5 Example 1 - Finding Area

The life of AA brand batteries is normally distributed with an average of 25 hours and a standard deviation of 4 hours. What is the probability that a randomly selected battery can be last for:

a. between 23 and 28 hours
b. between 13 and 37 hours

Example 1 - Finding Area Cont'

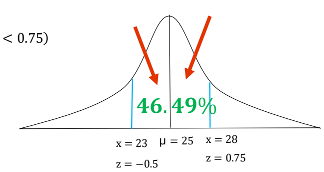
Given $\mu = 25, \sigma = 4$ and $z = \frac{x - \mu}{\sigma}$

$a.\ \boldsymbol{P(23 \le x < 28)}$

$$z = \frac{x - \mu}{\sigma} = \frac{23 - 25}{4} = \frac{-2}{4} = -0.5000$$

$$z = \frac{x - \mu}{\sigma} = \frac{28 - 25}{4} = \frac{3}{4} = 0.7500$$

So, $P(23 \le x < 28) = P(-0.5 \le z < 0.75)$
$= P(z < 0.75) - P(z \le -0.5)$
$= 0.7734 - 0.3085$
$= 0.4649$



Example 1 - Finding Area Cont'

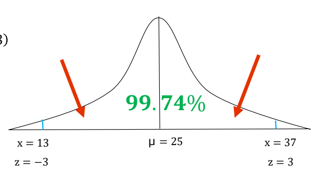
Given $\mu = 25, \sigma = 4$ and $z = \frac{x - \mu}{\sigma}$

$b.\ \boldsymbol{P(13 \le x < 37)}$

$$z = \frac{x - \mu}{\sigma} = \frac{13 - 25}{4} = \frac{-12}{4} = -3$$

$$z = \frac{x - \mu}{\sigma} = \frac{37 - 25}{4} = \frac{12}{4} = 3$$

So, $P(13 \le x < 37) = P(-3 \le z < 3)$
$= P(z < 3) - P(z \le -3)$
$= 0.9987 - 0.0013$
$= 0.9974$



Example 1 - Finding Area Cont'

a. between 23 and 28 hours
$\boldsymbol{P(23 \le x < 28)} = 0.4649$
There is a 46.49% chance of getting a battery, which can be last for between 23 and 28 hours.

b. between 13 and 37 hours
$\boldsymbol{P(13 \le x < 37)} = 0.9974$
There is a 99.74% chance of getting a battery, which can be last for between 13 and 37 hours.

That is an example of showing how we use the sample data to describe the population scenario.

Given the **x-values** → z-scores→ **area (probability)**

The reverse way is sorted of the process of constructing the Confidence Interval. Given the **area (probability)** → z-scores → **the interval (lower and upper bound values)**

Confidence Interval
Population Mean μ
z score

Sample data likely include the population parameter. Hence, based on some confidence levels, we can predict or construct the (mean) interval for the population.

The One-Sample z Confidence Interval for μ

The general formula for a confidence interval for a population mean μ when

1. $\bar{x}$ is the sample mean from a **simple random sample**,
2. the **sample size n is large** (generally $n \geq 30$), and
3. **σ, the population standard deviation, is known**

is

$$\bar{x} \pm (z \text{ critical value})\left(\frac{\sigma}{\sqrt{n}}\right)$$

Confidence Interval, μ
z score

$$\mu = \bar{x} \pm z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)$$

μ = Population Mean

$\bar{x}$ = Sample Mean

σ = Population Standard Deviation

n = Number of Samples

α = Significant Level = 1 − Confidence Level = 1 - CL

$z_{\frac{\alpha}{2}}$ = Critical Value

Example 2 - Scores

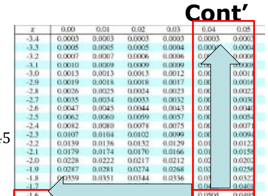
Scores on a test are normally distributed with a population standard deviation of 15. A random sample of 30 scores on the test has a mean of 50. Estimate the population mean with

a. 90% confidence interval
b. 95% confidence interval
c. 99% confidence interval

Example 2 – Scores Cont'

Estimate the population mean with
a. 90% confidence interval
Given $\sigma = 15$, $n = 30$, $\bar{x} = 50$,
CI = 90% = 0.9 → $\alpha = 1 - 0.9 = 0.1$

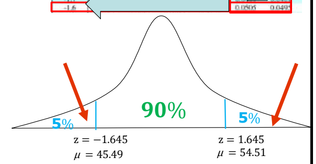
Critical value = $z_{\frac{\alpha}{2}} = z_{\frac{0.1}{2}} = z_{0.05} = -1.645$



$\mu = \bar{x} \pm z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)$

$= 50 \pm (1.645)(\frac{15}{\sqrt{30}})$
$= 50 \pm (1.645)(2.74)$
$= 50 \pm 4.51$

$45.49 \leq \mu \leq 54.51$

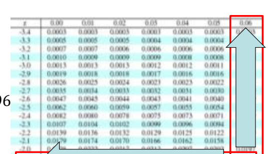


We are 90% confident that the population mean score is between 45.49 and 54.51.

Example 2 – Scores Cont'

Estimate the population mean with
b. 95% confidence interval
Given $\sigma = 15$, $n = 30$, $\bar{x} = 50$,
CI = 95% = 0.95 → $\alpha = 1 - 0.95 = 0.05$

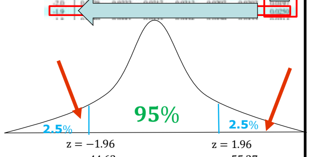
Critical value = $z_{\frac{\alpha}{2}} = z_{\frac{0.05}{2}} = z_{0.025} = -1.96$



$\mu = \bar{x} \pm z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)$

$= 50 \pm (1.96)(\frac{15}{\sqrt{30}})$
$= 50 \pm (1.96)(2.74)$
$= 50 \pm 5.37$

$44.63 \leq \mu \leq 55.37$

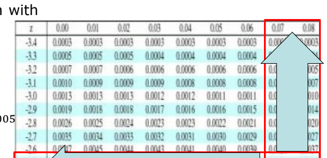


We are 95% confident that the population mean score is between 44.63 and 55.37.

Example 2 – Scores Cont'

Estimate the population mean with
c. 99% confidence interval
Given $\sigma = 15$, $n = 30$, $\bar{x} = 50$,
CI = 99% = 0.99 → $\alpha = 1 - 0.99 = 0.01$

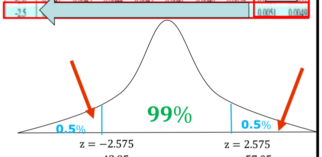
Critical value = $z_{\frac{\alpha}{2}} = z_{\frac{0.01}{2}} = z_{0.005}$ = −2.575



$\mu = \bar{x} \pm z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)$

$= 50 \pm (2.575)(\frac{15}{\sqrt{30}})$
$= 50 \pm (2.575)(2.74)$
$= 50 \pm 7.05$

$42.95 \leq \mu \leq 57.05$



We are 99% confident that the population mean score is between 42.95 and 57.05.

Example 2 – Scores Cont'

	90% C.I.	95% C.I.	99% C.I.
Z-value	1.645	1.96	2.575
Population Mean	$45.49 \le \mu \le 54.51$	$44.63 \le \mu \le 55.37$	$42.95 \le \mu \le 57.05$

Stop and Think:

Assume other factors are remained the same, what' s happen when C.I. increases?

Can we have 100% C.I. possible in real life?

Practical Question 1 – *z score*

Suppose that a random sample of 30 observations from a population whose standard deviation is 50 with a sample mean of 1,000. Estimate the population mean with

a. 90% confidence interval
b. 95% confidence interval
c. 99% confidence interval

Answer:

a. $984.98 \le \mu \le 1015.02$
b. $982.11 \le \mu \le 1017.89$
c. $976.49 \le \mu \le 1023.51$

Confidence Interval Population Mean μ *t Distribution*

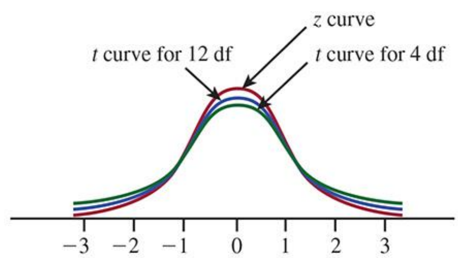
To be realistic, it is hard to know the population standard deviation, σ. Therefore, we use sample standard deviation, s, to estimate the population standard deviation, σ.

Hence, we need another distribution when the σ is UNKNOWN
→ t distributions.

Important Properties of *t* Distributions

1. The t distribution corresponding to any particular number of degrees of freedom is bell shaped and centered at zero (just like the standard normal (z) distribution).
2. Each t distribution is more spread out than the standard normal (z) distribution.
3. As the number of degrees of freedom increases, the variability of the corresponding t distribution decreases.
4. As the number of degrees of freedom increases, the corresponding sequence of t distributions approaches the standard normal (z) distribution.

Confidence Interval Population Mean μ *t Distribution*



Confidence Interval Population Mean μ *t Distribution*

The One-Sample *t* Confidence Interval for μ

The general formula for a confidence interval for a population mean μ based on a sample of size n when

1. $\bar{x}$ is the sample mean from a **simple random sample**,
2. the **population distribution is normal**, *or* **the sample size n is large** (generally $n \ge 30$), and
3. **σ, the population standard deviation, is unknown**

is

$$\bar{x} \pm (t \text{ critical value})\left(\frac{s}{\sqrt{n}}\right)$$

where the t critical value is based on df $= n - 1$. Appendix Table 3 gives critical values appropriate for each of the confidence levels 90%, 95%, and 99%, as well as several other less frequently used confidence levels.

Confidence Interval, μ *t Distribution*

$$\mu = \bar{x} \pm t_{df}\left(\frac{s}{\sqrt{n}}\right)$$

μ = Population Mean

$\bar{x}$ = Sample Mean

s = Sample Standard Deviation

n = Number of Samples

df = Degree of Freedom= $n-1$ (The pieces of information required to work out n, e.g. traffic light)

t = Critical Value

Example 3 – Drive-Through Medicine (p.477)

Given $n = 38$, $\bar{x} = 26$ and $s = 1.57$, normally distributed. Estimate the population mean with a 95% confidence interval. $\frac{\alpha}{2} = \frac{0.05}{2} = 0.025$

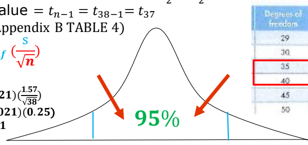
Critical value $= t_{n-1} = t_{38-1} = t_{37}$
≈ 2.021 (Appendix B TABLE 4)

Degrees of freedom	$t_{0.100}$	$t_{0.050}$	$t_{0.025}$	$t_{0.010}$
29	1.311	1.699	[illegible]	2.462
30	1.310	1.697	[illegible]	2.457
35	[illegible]	[illegible]	2.030	2.438
40	1.303	[illegible]	2.021	2.423
45	1.301	1.679	2.014	2.412
50	1.299	1.676	2.009	2.403

$\mu = \bar{x} \pm t_{df}\left(\frac{s}{\sqrt{n}}\right)$

$= 26 \pm (2.021)\left(\frac{1.57}{\sqrt{38}}\right)$
$= 26 \pm (2.021)(0.25)$
$= 26 \pm 0.51$

$25.49 \leq \mu \leq 26.51$



$t = -2.021$
$\mu = 25.49$

$t = 2.021$
$\mu = 26.51$

We are 95% confident that the population mean is between 25.49 and 26.51, $(25.49, 26.51)$.

Example 4

Given $n = 5$, $\bar{x} = 30$ and $s = 1.45$, normally distributed. Estimate the population mean with a 90% confidence interval. $\frac{\alpha}{2} = \frac{0.1}{2} = 0.05$

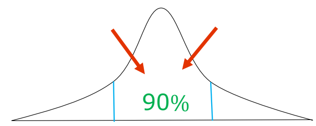
Degrees of freedom	$t_{0.10}$	$t_{0.05}$	$t_{0.025}$	$t_{0.01}$	$t_{0.005}$
1	3.078	[illegible]	12.706	31.821	63.657
2	1.886	[illegible]	4.303	6.965	9.925
3	1.638	[illegible]	3.182	4.541	5.841
4	[illegible]	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032

Critical value $= t_{n-1} = t_{5-1} = t_4$
$= 2.132$

$\mu = \bar{x} \pm t_{df}\left(\frac{s}{\sqrt{n}}\right)$

$= 30 \pm (2.132)\left(\frac{1.45}{\sqrt{5}}\right)$
$= 30 \pm (2.132)(0.648)$
$= 30 \pm 1.38$

$28.62 \leq \mu \leq 31.38$



$t = -2.132$
$\mu = 28.62$

$t = 2.132$
$\mu = 31.38$

We are 90% confident that the population mean is between 28.62 and 31.38, $(28.62, 31.38)$.

Practical Question - *t Distributions*

Exercise 9.37 What is the appropriate *t critical value* for part a, c and e (page. 481)

a. $CI = 95\%$, $n = 17$
t critical value $= \mathbf{2.12}$

c. $CI = 99\%$, $n = 24$
t critical value $= \mathbf{2.807}$

e. $CI = 90\%$, $n = 13$
t critical value $= \mathbf{1.782}$

Conclusion

- Define Confidence Interval
- Confidence Interval for a Population Mean, **μ**
  - Population Standard deviation, σ is known
    - z score
    - $\mu = \bar{x} \pm z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)$
  - Population Standard deviation, σ is NOT known
    - t distribution
    - $\mu = \bar{x} \pm t_{df}\left(\frac{s}{\sqrt{n}}\right)$