

# Chapter 8 Confidence Intervals

## Ch 8.1 Confidence Interval for mean with one sample when $\sigma$ is known.

Terms: Population mean:  $\mu$ , sample mean:  $\bar{x}$   
 sample standard deviation:  $s$ , sample size:  $n$   
 Population standard deviation:  $\sigma$   
 Confidence Level: Clevel,  
 Significant Level  $\alpha$ :  $1 - \text{Clevel}$  (unlikely)

### A) Estimate $\mu$ (when $\sigma$ is given)

Use one sample with size  $n$ ,  $\bar{x}$ ,  $s$  or raw data.

- 1) point estimate of  $\mu$ :  $\bar{x}$
- 2) Interval estimate of  $\mu$ :  $\bar{x} - E < \mu < \bar{x} + E$   
 When  $E$  (EBM) =  $z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$  when  $\sigma$  is given.

Use Online calculator statdisk to find confidence interval of mean when  $\sigma$  is given.

use <https://www.statdisk.com/#>

Analysis/Confidence Intervals/Mean one sample/

If summary statistics ( $\bar{x}$ ,  $s$ ,  $n$ ) are given.

Select "use Summary Statistics" tab otherwise use "use data" tab.

Enter C\_level,  $n$ ,  $\bar{x}$ ,  $s$  or select with data

Must enter  $\sigma$  that is given.

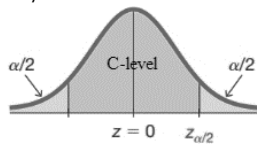
Output:  $E$  is the margin of error or Error bound for

Mean. Confidence interval estimate is

*lower limit*  $< \mu <$  *upper limit*

Note: the requirement for the confidence interval is CLT applies. ( $n > 30$  or  $X$  is Normal). Sample is SRS.

$z_{\alpha/2}$  is the critical value with C-level in the middle



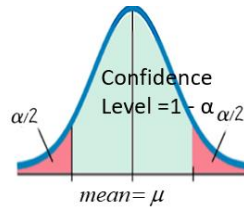
To find  $z_{\alpha/2}$   
 Inverse Normal Calculator  
 with right tail =  $\alpha/2$

Explanation:

According to Central Limit Theorem, if  $n > 30$  or  $X$  is normal,  $\bar{x}$  has a sampling distribution that is Normal with mean =  $\mu$  and SD =  $\frac{\sigma}{\sqrt{n}}$ .

Given a C-level, the maximum sampling error of  $\bar{x}$  from  $\mu$  is  $E$  with a right tail of  $\alpha/2$ .

Since standard error is  $\frac{\sigma}{\sqrt{n}}$ ,  $E = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$



So the range of values from  $\bar{x} - E$  to  $\bar{x} + E$  should include the real  $\mu$  at a confidence level of C-level.

Ex1. Suppose scores on a standardized exam are normally distributed with standard deviation of 35. A sample of 14 scores are collected with sample mean 137.3 and sample standard deviation 37.2.  
 a) Use the sample to estimate the population mean score of the standardized test at a 98% confidence level.

Identify: C-level = 0.98,  $n = 14$ ,  $\bar{x} = 137.3$ ,  $s = 37.2$ ,  $\sigma = 35$

Statdisk/Analysis/Confidence intervals/Mean one sample/

```
Margin of Error, E = 21.76103
98% Confident the population mean is within the range:
115.53897 < mean < 159.06103
```

b) At a confidence level of 98%, what is the significant level and the critical value.

$\alpha = 0.02$ , use Inverse-Normal calculator with right tail = 0.01, critical value =  $z_{\alpha/2} = 2.326$

c) A previous study shows that the mean scores was 144.2. Is the population mean score below 144.2?

d) Explain what will happen to margin of error  $E$  if sample size increase.

e) Explain what will happen to margin of error  $E$  if confidence level increase to 99%.

Ex2. The specific Absorption Rate (SAR) for a cell phone measures the amount of radio frequency energy absorbed by the user's body when using the handset. The legal SAR level is no more than 1.6 watts per kilogram. The following SAR data are collected from 30 phones.

- 1.11, 1.48, 1.43, 1.3, 1.09, 0.455, 1.41, 0.82, 0.78, 1.25, 1.36, 1.34, 1.18, 1.3, 1.26, 1.29, 0.36, 0.52, 1.6, 1.39, 0.74, 0.5, 0.4, 0.867, 0.68, 0.51, 1.13, 0.3, 1.48, 1.38.

a) Use the sample to find the 95% confidence interval of the mean SAR level for all phones given that population standard deviation is  $\sigma = 0.337$ .

Ans: Use Statdisk/Analysis/Confidence Interval/  
One sample mean/ with data tab

output:

Margin of Error, E = 0.12059

95% Confident the population mean is within the range:  
0.90314 < mean < 1.14432

b) Write the confidence interval in the form of  
interval notation and +/- notation.

c) Can we conclude that the mean SAR level is below  
the legal allowable value of 1.6?

d) Find the critical value of  $z_{\alpha/2}$ .

e) If we want to decrease the margin of error, what  
can be done?

### B) Find sample size n for a given Margin of error (EBM)

From central limit Theorem, we have the equation  
between E and n.  $E = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ . Algebra can be used  
to show that for a given C-level, E and  $\sigma$ ,

$$n = \left( \frac{z_{\alpha/2} \cdot \sigma}{E} \right)^2$$

Find sample size by online calculator:

Statdisk.com / Analysis/sample size Determination/  
Estimate mean/

Enter Clevel, E, population standard deviation.

Note: when sample size increase, E decrease.  
when E increase, sample size decrease.

Ex3.

The population standard deviation of student's age in  
a community college is 15 years. If we want to be 95%  
confidence that the error in confidence interval  
estimate is within 2 years from the real mean age of  
all students, how many students must be surveyed?

Use Statdisk/ Analysis / Sample size Determination/  
Estimate Mean Clevel = 0.95, E = 2, pop. sd = 15  
Required sample size = 217.

### Ch 8.2 Confidence Interval for mean with one sample when $\sigma$ is not known.

Use one sample with size n,  $\bar{x}$ , s

- 1) point estimate of  $\mu$ :  $\bar{x}$
- 2) Interval estimate of  $\mu$ :  $\bar{x} - E < \mu < \bar{x} + E$   
When  $E$  (EBM) =  $t_{\alpha/2} \frac{s}{\sqrt{n}}$  when  $\sigma$  is not given.

$E = t_{\alpha/2} \frac{s}{\sqrt{n}}$  when  $\sigma$  is not known or given.

Use Online calculator to find confidence interval of  
mean when  $\sigma$  is not given.

- use online statdisk: <https://www.statdisk.com/#>  
Analysis/Confidence Intervals/Mean one sample/  
If summary statistics ( $\bar{x}$ , s, n) are given, select "use  
Summary Statistics" tab otherwise use "use data" tab.  
Enter Clevel, n,  $\bar{x}$ , s, or select with data tab.

Do not enter population standard deviation.

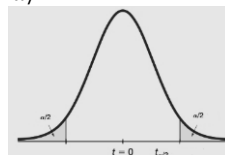
Output: E is the margin of error or Error bound for  
Mean.

Confidence interval estimate is  
 $lower\ limit < \mu < upper\ limit$

Note: the requirement for this confidence interval is  
CTL applies ( $n > 30$  or X is normal). Sample is SRS.

### B) T-distribution

$t_{\alpha/2}$  is the critical value with C-level in the middle.



To find  $t_{\alpha/2}$

Use Online Stat book Inverse-t  
Calculator:

[http://onlinestatbook.com/2/calculators/inverse\\_t\\_dist.html](http://onlinestatbook.com/2/calculators/inverse_t_dist.html)

the degree of freedom =  $n - 1$ , Select confidence  
interval from 90%, 95% or 99%. "Calculate"  
The "t for confidence interval" is the critical value  $t_{\alpha/2}$ .

Properties of t-distributions:

- 1)  $t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$  = no. of times of  $s/\sqrt{n} \bar{x}$  is from mean  $\mu$   
when  $\bar{x}$  has a normal distribution.
- 2) Each t-distribution is different for different sample  
size according to the degree of freedom  $df = n - 1$ .
- 3) Each t-distribution has the same general symmetric  
bell shape as standard normal, but with more  
variability.
- 4) As sample size increase, the student t- distribution  
approaches the standard normal distribution.

Ex1: Patients with insomnia are treated with zopiclone. After the treatment, the 16 subjects had a mean wake time of 98.9 min. and a sample standard deviation of 42.3 min. Assume that wake times are normally distributed.

a) Construct a 98% confidence interval estimate of mean wake time for all patients treated by zopiclone.  
- since X is normal,  $\bar{x}$  is normal,  $\sigma$  is not known, so use t-distribution.

- Identify C-level = 0.98,  $n = 16$ ,  $\bar{x} = 98.9$ ,  $s = 42.3$ ,

- Use statdisk/ Analysis/Confidence interval/ Mean one sample/

- output: E = 27.521,  $71.379 < \mu < 126.421$   
interval format (71.379, 126.421),  
+/- format:  $98.9 \pm 27.521$

b) Can we conclude the mean wake time is less than 100 min.?

No, the mean wake time can have values more than 100min.

c) If confidence level is decreased to 95%, what will happen to the margin of error?

### C) Interpret the Confidence interval of mean:

- We estimate with \_\_\_\_\_ confidence that the true population mean for \_\_\_\_\_ is between \_\_\_\_ and \_\_\_\_ .
- 90% of confidence interval constructed in this way contains the true \_\_\_\_\_ population mean.

#### Make conclusion from a confidence Interval:

- Any value in the confidence Interval can be  $\mu$ .
- If the whole interval  $> a$ , we can conclude  $\mu > a$
- If the whole interval  $< a$ , we can conclude  $\mu < a$ .
- When two confidence interval overlap, we can conclude that the two  $\mu$  may be the same. We cannot conclude one of the  $\mu$  is higher.
- Never make conclusion about population mean based solely on value of  $\bar{x}$  .

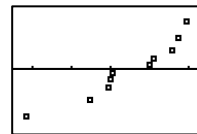
Ex2. Listed below are amounts of arsenic (in ug, per serving) in samples of brown rice from California.

5.4, 5.6, 7.3, 4.5, 7.5, 1.5, 5.5, 9.1, 8.7, 8.4

a) Can we assume that amount of arsenic in brown rice is from a normal distribution?

Use Normal Quantile plot to assess Normality.

Copy data to statdisk. Data/Normal Assessment



The points are reasonably close to a straight line. So we can assume population X is normal.

b) Construct 90% confidence interval of mean arsenic in all brown rice from California. What distribution is used?

-Since X is normal, so  $\bar{x}$  is normal,  $\sigma$  is not known, so use t distribution.

-identify Clevel = 0.9, column in statdisk.

- Statdisk Analysis/Confidence intervals/Mean One sample /use data tab.

output: E = 1.348 ,  $5.002 < \text{mean} < 7.698$

-Interval estimate:  $5.0 \text{ ug} < \mu < 7.7 \text{ ug}$

c) Find the critical value.

Since t-distribution is used when  $\sigma$  is not known, use online calculator inverse t-calculator.

degree of freedom = 9, C-level = 90%

critical value  $t_{0.05} = 1.833$

d) Interpret the meaning of the confidence Interval:

"We estimate with 90% confidence that the mean amount of arsenic is between 5.0 ug and 7.7 ug."

"90% of all confidence interval collected from sample of size 10 will contain the true mean arsenic."

d) Can we conclude the mean amount is less than 8 ug?

Yes, because all values in the interval are less than 8 ug. We can conclude that the mean amount of arsenic in all brown rice are less than 9 ug. at 90% confidence.

### Ch 8.3 Confidence Interval for population proportion:

Terms: Population proportion:  $p$

Sample proportion:  $\hat{p} = (x/n)$

number of success:  $x$ ; sample size:  $n$

Confidence Level: C-level

significant Level:  $\alpha = 1 - C\text{-level}$   
(probability of unlikely)

EBP: Error Bound for proportion, margin of error

#### A) To Estimate $p$ :

1) point estimate:  $\hat{p} = x/n$

2) Interval estimate:  $\hat{p} - E < p < \hat{p} + E$

$$E (EBP) = z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

Note:  $E$  is the approximated value of sampling error of sampling distribution of  $\hat{p}$ , which has a normal distribution when  $x$  and  $n-x \geq 5$ .

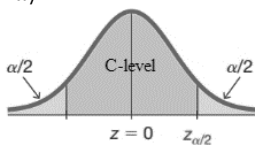
Use Statdisk online calculator to find the confidence interval.

- identify C-level, sample size  $n$  and success =  $x$ .

- <https://www.statdisk.com/#>, Analysis/ Confidence Intervals/Proportion one sample/

output:  $E$  (EBP) and lower  $< p <$  upper

$z_{\alpha/2}$  is the critical value with C-level in the middle.



To find  $z_{\alpha/2}$

Use online Inverse Normal calculator with right tail =  $\alpha/2$

Explanation:

$X$  = number of success is a binomial distribution with

mean =  $np$  and  $sd = \sqrt{npq}$ , when  $np$  and  $nq \geq 5$ ,

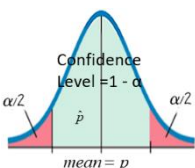
distribution of  $X$  is normal with mean =  $np$  and

$sd = \sqrt{npq}$ . So distribution of  $\hat{p} = \frac{x}{n}$  is normal with

mean =  $p$  and  $SD = \sqrt{\frac{pq}{n}}$

At a given C-level, the maximum error of  $\hat{p}$  and  $p$  is  $E$

where  $E(EBP) \approx z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}}$



Note: the requirement for this confidence interval is not  $n > 30$  but  $np$  and  $nq \geq 5$ .

#### Interpret a 95% Confidence Interval:

We are 95% confidence that the interval from \_\_\_\_ to \_\_\_\_ actual contain the true value of the population proportion of the category of interest.

The Confidence interval also shows that 95% of all confidence Intervals contain the true value of  $p$ .

Make conclusion from a confidence Interval:

- 1) Any value in the confidence Interval can be  $p$ .
- 2) If the whole interval  $> a$ , we can conclude  $p > a$
- 3) If the whole interval  $< a$ , we can conclude  $p < a$ .
- 4) When two confidence interval overlap, we can conclude that the two  $p$  may be the same. We cannot conclude one of the  $p$  is higher.

Ex1. A research is conducted to determine how many household use Netflix to stream videos. A random sample of 500 households show that 442 households use Netflix.

a) Use a 90% confidence level to compute a confidence interval estimate of true proportion of households using Netflix.

use Statdisk Analysis/Confidence Intervals/Proportion one sample : Clevel = 0.9,  $n = 500$ ,  $x = 442$ ,

Margin of Error,  $E = 0.02356$

90% Confidence Interval (using normal approx):  
 $0.86044 < p < 0.90756$

b) Find the critical value.

Use Inverse Normal calculator:

$z_{0.05} =$

c) Interpret the confidence interval in non-technical term.

We estimate with 90% confidence that the true proportion of all households that use Netflix is between 86.0% to 90.8%

d) Can we conclude with 90% confidence that more than 80% of households use Netflix?

Ex2: A Gallup poll of 1487 adults showed that 43% of the respondents have Facebook pages.

a) Find the number in the sample who have Facebook pages.  $x = n\hat{p} = 1487 (0.43) = 639$

c) Find the interval estimate of  $p$  at 95% confidence level and margin of error  $E$ .

Use statdisk with Clevel = 0.95,  $n = 1487$ ,  $x = 639$   
Interval estimate is  $0.405 < p < 0.455$ ,  $E = 0.025$

d) Write a non-technical interpretation of the above  
We are 95% confidence that the true percent of adults who have Facebook pages are between 40.5% to 45.5%.

e) Can we claim that less than 60% of all adults have Facebook pages?

Since the whole interval is less than 0.6, yes, we can conclude less than 60% of all adults have Facebook pages.

### B) Determine sample size for a desired E

Since  $E(EBP) \approx z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}}$

$n = \frac{(z_{\alpha/2})^2 \hat{p}\hat{q}}{E^2}$  when an estimate  $\hat{p}$  is known.

$n = \frac{(z_{\alpha/2})^2 0.25}{E^2}$  when no estimate  $\hat{p}$  is known.

Use statdisk / Analysis/Sample size Determination/  
Estimate proportion to find sample size for a given error E.

Input C-Level, Desired E, Estimate of  $p = \hat{p} = x/n$   
"Evaluate"

Ex3. What sample size should be used if we want the keep the margin of error within 2.1% when estimating a proportion at a 90% confidence interval. use  $\hat{p} = 0.43$  as an old estimation of p.

C-level = 0.9, E = 0.021, estimate of p = 0.43,  
Evaluate  
sample size p = 1504.

b) If no previous study has been done , what sample size will be needed. (do not use p-hat = 0.43)

Use statdisk / Analysis/Sample size Determination/  
Estimate proportion

C-level = 0.9, E = 0.021, estimate of p = blank  
Evaluate  
sample size p = 1534.