

# STA258H5

Al Nosedal  
and Alison Weir

Winter 2017

CONFIDENCE INTERVAL  
(ONE MEAN  
NORMAL POPULATION  
LARGE SAMPLE  
VARIANCE KNOWN).

# Point and Interval Estimates

The sample mean,  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ , is a number we use to estimate the population mean,  $\mu$ . This is called a **point estimate**.

But, we know it's not equal to  $\mu \rightarrow$  we'd rather estimate the population mean using an **interval estimate** that gives a range of real numbers that we hope contains the population mean,  $\mu$ .

# Confidence Interval for $\mu$

CONFIDENCE INTERVAL FOR THE MEAN (NORMAL SAMPLE, VARIANCE KNOWN).

Let  $X_1, X_2, \dots, X_n$  be iid  $N(\mu, \sigma^2)$ , where  $\mu$  is unknown and  $\sigma$  is known.

We know that  $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$ .

We also know that  $P(-1.96 < Z < 1.96) = 0.95$ .

$\Rightarrow P(-1.96 < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < 1.96) = 0.95$ .

$\Rightarrow P\left(\bar{X} - 1.96 \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1.96 \frac{\sigma}{\sqrt{n}}\right) = 0.95$

- This is a random interval  $\bar{X} \pm 1.96 \frac{\sigma}{\sqrt{n}}$
- The interval is random since  $\bar{X}$  is random due to sampling.
- The population mean  $\mu$  is a fixed, but unknown, number.
- The probability  $\mu$  is inside the random interval is 0.95. You can think about it as the success rate for the method.
- 95% of all samples give an interval that captures  $\mu$ , and 5% of all samples give an interval that does not capture  $\mu$ .

- This is NOT a random interval  $\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}$
- The probability  $\mu$  is inside this interval is either 1 or 0.

# Confidence interval isn't always right

The fact that not all confidence intervals contain the true value of the parameter is often illustrated by plotting a number of random confidence intervals at once and observing. Let's do it!

```
## Step 1. Generate random samples;
```

```
set.seed(2017)
```

```
m=50;
```

```
# m= number of samples;
```

```
n=25;
```

```
# n= number of obs in sample;
```

```
mu.i=0;
```

```
# mu.i = mean of obs;
```

```
sigma.i=5;
```

```
# sigma.i =std. dev. of obs;
```

```
mu.total=n*mu.i;
```

```
# mean of Total;
```

```
sigma.total=sqrt(n)*sigma.i;
```

```
# std. dev. of Total;
```

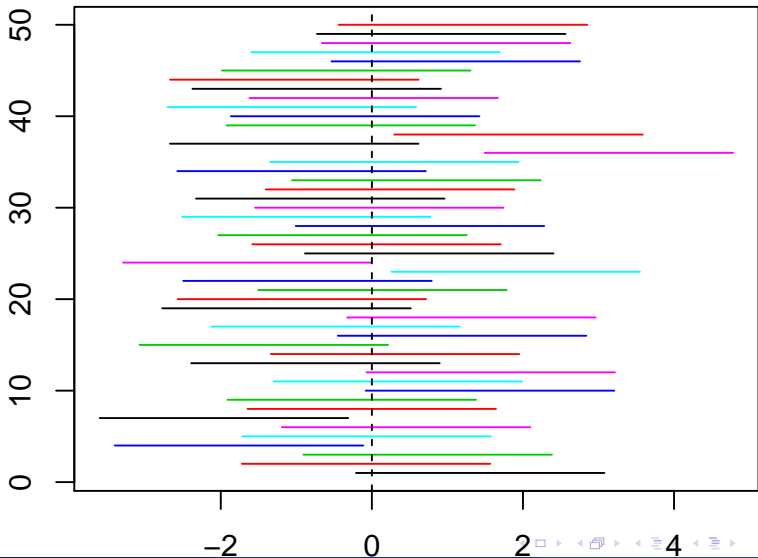


```
## Step 2. Construct CIs;  
  
xbar=rnorm(m,mu.total,sigma.total)/n;  
  
SE=sigma.i/sqrt(n);  
  
alpha=0.10;  
  
z.star=qnorm(1-alpha/2);
```

```
## Step 3. Graph CIs;
```

```
matplot(rbind(xbar-z.star*SE,xbar+z.star*SE),rbind(1:m,1:m),  
type="l",lty=1, xlab=" ",ylab=" ");
```

```
abline(v=0,lty=2);
```

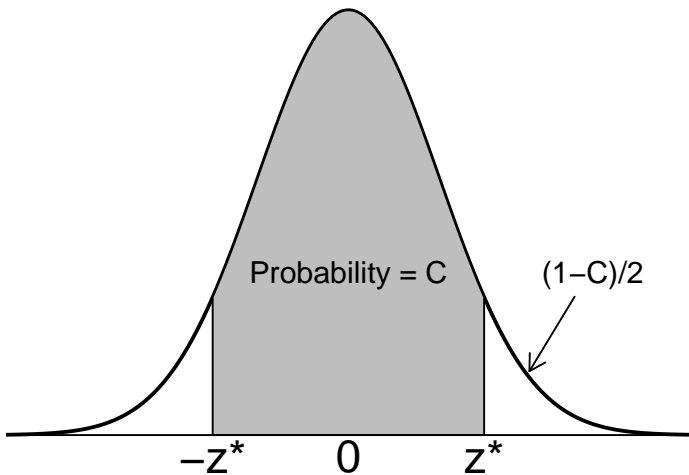


# Confidence Interval for the mean of a Normal population

Draw an SRS (Simple Random Sample) of size  $n$  from a Normal population having unknown mean  $\mu$  and **known** standard deviation  $\sigma$ . A level  $C$  confidence interval for  $\mu$  is

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

The critical value  $z_*$  is illustrated in a Figure below and depends on  $C$ .



# Large Sample CI for $\mu$ (Normal data)

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

Valid if

- n large
- random sample from a Normal distribution
- independent observations

Some definitions

- $1 - \alpha$  is the confidence coefficient
- $100(1 - \alpha)\%$  is the confidence level

Confidence coefficient	Confidence level	$z$
0.90	90%	1.645
0.95	95%	1.96
0.99	99%	2.576

# Example

Playbill magazine reported that the mean annual household income of its readers is \$ 119,155. Assume this estimate of the mean annual household income is based on a sample of 80 households, and based on past studies, the population standard deviation is known to be  $\sigma = \$ 30,000$ .

- Develop a 90 % confidence interval estimate of the population mean.
- Develop a 95 % confidence interval estimate of the population mean.
- Develop a 99 % confidence interval estimate of the population mean.



In this case,  $\mu$  = mean annual household income of **all** its readers.

$\sigma = 30,000$ ,  $n = 80$ ,  $1 - \alpha = 0.90$ ,  $\alpha = 0.10$ .

a) margin of error =  $Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = (1.64) \left( \frac{30000}{\sqrt{80}} \right) = 5500.727 \approx 5500.73$

Confidence Interval is given by: estimate  $\pm$  margin of error. That is:

$$\bar{x} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}.$$

$$119,155 \pm 5500.73$$

$$(113,654.27, 124,655.73)$$

In this case,  $\mu$  = mean annual household income of **all** its readers.  
 $\sigma = 30,000$ ,  $n = 80$ ,  $1 - \alpha = 0.95$ ,  $\alpha = 0.05$ .

b) margin of error =  $Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = (1.96) \left( \frac{30000}{\sqrt{80}} \right) = 6574.04$

Confidence Interval is given by: estimate  $\pm$  margin of error. That is:

$$\bar{x} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$119,155 \pm 6574.04$$

$$(112,580.96, 125,729.04)$$

In this case,  $\mu$  = mean annual household income of **all** its readers.

$\sigma = 30,000$ ,  $n = 80$ ,  $1 - \alpha = 0.99$ ,  $\alpha = 0.01$ .

c) margin of error =  $Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = (2.57) \left( \frac{30000}{\sqrt{80}} \right) = 8620.042 \approx 8620.04$

Confidence Interval is given by: estimate  $\pm$  margin of error. That is:

$$\bar{x} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$119,155 \pm 8620.04$$

$$(110,534.96, 127,775.04)$$

# Example

The number of cars sold annually by used car salespeople is Normally distributed with a standard deviation of 15. A random sample of 15 salespeople was taken, and the number of cars each sold is listed here. Find the 95% confidence interval estimate of the population mean. Interpret the interval estimate.

# Example

79	43	58	66	101
63	79	33	58	71
60	101	74	55	88

*# R Code;*

```
simple.z.test = function(x,sigma,conf.level=0.95) {  
n = length(x);  
xbar=mean(x);  
alpha = 1 - conf.level;  
zstar = qnorm(1-alpha/2);  
SE = sigma/sqrt(n);  
xbar + c(-zstar*SE,zstar*SE);  
}
```

```
# Step 1. Entering data;

cars=c(79, 43, 58, 66, 101, 63, 79,
33, 58, 71, 60, 101, 74, 55, 88);

# Step 2. Finding CI;

simple.z.test(cars,15);

## [1] 61.00909 76.19091
```

We estimate that the mean number of cars sold annually by all used car salespersons lies between 61 and 76, approximately. This type of estimate is correct 95% of the time.



## Example

Suppose a student measuring the boiling temperature of a certain liquid observes the readings (in degrees Celsius) 102.5, 101.7, 103.1, 100.9, 100.5, and 102.2 on 6 different samples of the liquid. He calculates the sample mean to be 101.82. If he knows that the distribution of boiling points is Normal, with standard deviation 1.2 degrees, what is the confidence interval for the population mean at a 95% confidence level?

A **confidence interval** uses sample data to estimate an unknown population parameter with an indication of how accurate the estimate is and of how confident we are that the result is correct.

The **interval** often has the form  
estimate  $\pm$  margin of error

The **confidence level** is the success rate of the method that produces the interval.

A level  $C$  **confidence interval for the mean**  $\mu$  of a Normal population with **known** standard deviation  $\sigma$ , based on an SRS of size  $n$ , is given by

$$\bar{x} \pm z^* \frac{\sigma}{n}$$

The **critical value**  $z^*$  is chosen so that the standard Normal curve has area  $C$  between  $-z^*$  and  $z^*$ .

Other things being equal, the **margin of error** of a confidence interval gets smaller as

- the confidence level  $C$  decreases,
- the population standard deviation  $\sigma$  decreases, and
- the sample size  $n$  increases.