Recall :-

Confidence Interval

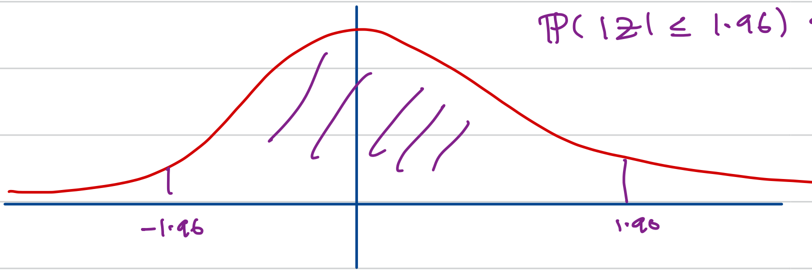
$X_1, X_2, \ldots X_n$ — μ – mean (unknown)

(known) σ^2 – variance

$$\frac{\sqrt{n}\,(\bar{X}-\mu)}{\sigma} \;\underset{\text{Theorem}}{\overset{\text{Central limit}}{\equiv}}\; \underset{(0,1)}{\text{Normal}} - Z$$

$$\mathbb{P}\left(\frac{\sqrt{n}\,(\bar{X}-\mu)}{\sigma} \le x\right) \text{ "} = \text{" } \mathbb{P}(Z \le x)$$

$\mathbb{P}(|Z| \le 1.96) \simeq 0.95$



$$\bar{X} = \frac{X_1 + \ldots + X_n}{n}$$

$$\left|\frac{\sqrt{n}\,(\bar{X}-\mu)}{\sigma}\right| \;\underset{\text{Theorem}}{\overset{\text{Central limit}}{\equiv}}\; \underset{(0,1)}{\text{Normal}} - |Z|$$

$$\left|\frac{\sqrt{n}\,(\bar{X}-\mu)}{\sigma}\right| \le 1.96 \iff -1.96 \le \frac{\sqrt{n}\,(\bar{X}-\mu)}{\sigma} \le 1.96$$

$$\iff (-1.96)\sigma \le \sqrt{n}\,(\bar{X}-\mu) \le (1.96)\,\sigma$$

$$\iff \bar{X} - \frac{1.96\,\sigma}{\sqrt{n}} \le \mu \le \bar{X} + \frac{1.96\,\sigma}{\sqrt{n}}$$

Confidence Intervals

Sample $x_1, x_2, \ldots x_n$ from population

• Assume μ & x

Unknown $E[X] = \mu$

$Var[X] = \sigma^2$ - Known

Compute : $\bar{X} = \sum_{i=1}^{n} \frac{x_i}{n}$

Using the Central Limit Theorem for large n we have

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

which is the same as saying

$$P\left(\mu \in \left(-\frac{1.96\sigma}{\sqrt{n}} + \bar{X}, \frac{1.96\sigma}{\sqrt{n}} + \bar{X}\right)\right) \approx 0.95$$

The interval $\left(-\frac{1.96\sigma}{\sqrt{n}} + \bar{X}, \frac{1.96\sigma}{\sqrt{n}} + \bar{X}\right)$ is called the 95% confidence interval for μ.

↳ dependent on sample.
and is valid if σ is known.

Confidence Intervals - Important - Precise meaning.

$$\mathbb{P}\left(\sqrt{n}\frac{|\bar{X}-\mu|}{\sigma} \le 1.96\right) \simeq 0.95$$

↕ interpretation

95% confidence interval for μ is $\left(-\frac{1.96\sigma}{\sqrt{n}} + \bar{X}, \frac{1.96\sigma}{\sqrt{n}} + \bar{X}\right)$

Meaning: for n large if we did m (large) repeated trials and computed the above interval for each trial then true mean would belong to approximately 95% of m intervals calculated.

Confidence Intervals ; $\sigma^2 = 1$.

The below is code for finding the confidence interval for a data x.

```
> cifn = function(x, alpha=0.95){
+ z = qnorm( (1-alpha)/2, lower.tail=FALSE)
+ sdx = sqrt(1/length(x))
+ c(mean(x) - z*sdx, mean(x) + z*sdx)
+ }
```

$\alpha = 0.95$

$$\left(\bar{X} - \frac{1.96}{\sqrt{n}}, \bar{X} + \frac{1.96}{\sqrt{n}}\right)$$

Three Confidence Intervals for Normal(0,1)

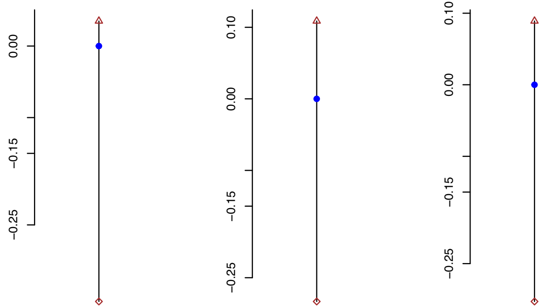
```
> x1 = rnorm(100,0,1);y = cifn(x1)
> y
[1] -0.35705304  0.03493976
> x2 = rnorm(100,0,1);z = cifn(x2)
> z
[1] -0.2832489  0.1087439
> x3 = rnorm(100,0,1);w = cifn(x3)
> w
[1] -0.30294682  0.08904598
```

Does 0 belong to all the three confidence intervals ?

Confidence Intervals Plots

rnorm(100, 0, 1) ≡ data.

The below is a plot of the three confidence intervals computed in the previous slide.



True mean $\mu = 0$

Contained in all 3 c. intervals

Confidence Intervals : 10 Trials

We generate 10 trials of 100 samples from Normal(0,1) and compute the confidence intervals using the function defined earlier.

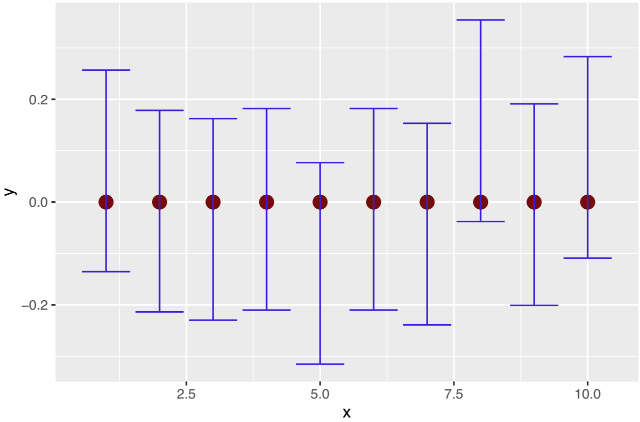
```
> normaldata = replicate(10, rnorm(100,0,1),
+ simplify=FALSE)
> cidata = sapply(normaldata, cifn)
```

It is easy to check how many of them contain 0.

```
> TRUEIN = cidata[1,]*cidata[2,]<0
> table(TRUEIN)
TRUEIN
TRUE
  10
```

$(a,b) \ni 0 \Leftrightarrow a \cdot b < 0$

Confidence Intervals : 10 Trials



Confidence Intervals: 40 Trials

We generate 10 trials of 100 samples from Normal(0,1) and compute the confidence intervals using the function defined earlier.

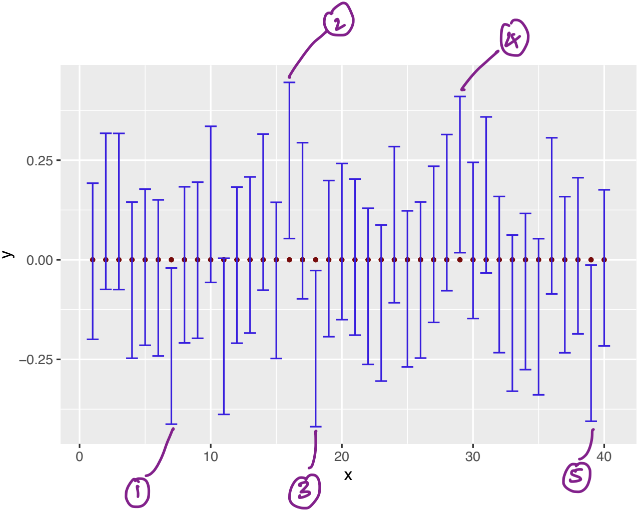
```
> normaldata = replicate(40, rnorm(100,0,1),
+ simplify=FALSE)
> cidata = sapply(normaldata, cifn)
```

It is easy to check how many of them contain 0.

```
> TRUEIN = cidata[1,]*cidata[2,]<0
> table(TRUEIN)
```

```
TRUEIN
FALSE  TRUE
    5    35
```

Confidence Intervals: 40 trials Plot



Confidence Intervals : 100 Trials

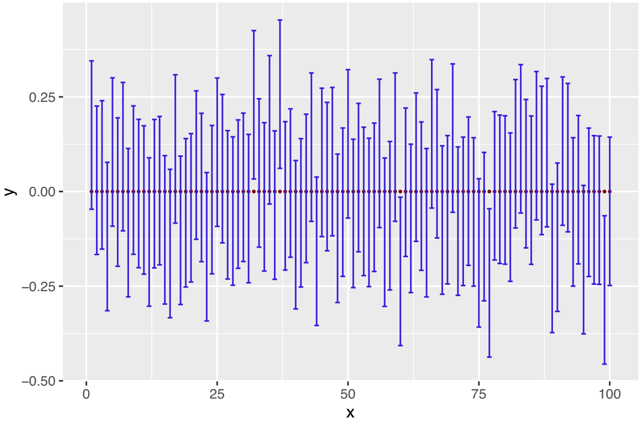
We generate 100 trials of 100 samples from Normal(0,1) and compute the confidence intervals using the function defined earlier.

```
> normaldata = replicate(100, rnorm(100,0,1),
+ simplify=FALSE)
> cidata = sapply(normaldata, cifn)
```

It is easy to check how many of them contain 0.

```
> TRUEIN = cidata[1,]*cidata[2,]<0
> table(TRUEIN)
TRUEIN
FALSE  TRUE
    5    95
```

Confidence Intervals : 100 Trials



Confidence Intervals : 1000 Trials

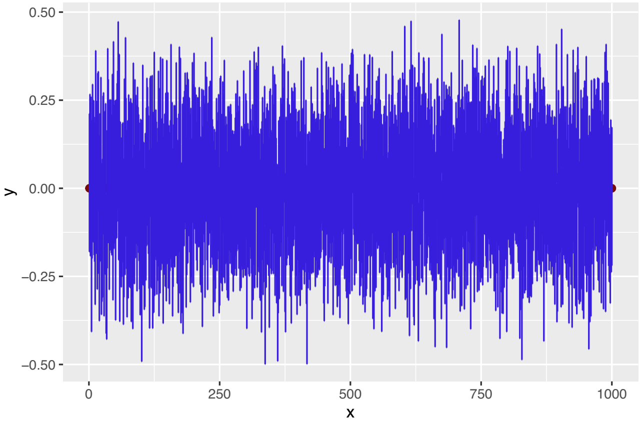
We generate 1000 trials of 100 samples from Normal(0,1) and compute the confidence intervals using the function defined earlier.

```
> normaldata = replicate(1000, rnorm(100,0,1),
+ simplify=FALSE)
> cidata = sapply(normaldata, cifn)
```

It is easy to check how many of them contain 0.

```
> TRUEIN = cidata[1,]*cidata[2,]<0
> table(TRUEIN)
TRUEIN
FALSE  TRUE
   54   946
```

Confidence Intervals : 1000 Trials



Confidence Intervals

95% confidence interval for μ is $\left(-\frac{1.96\sigma}{\sqrt{n}} + \bar{X}, \frac{1.96\sigma}{\sqrt{n}} + \bar{X}\right)$

Meaning: for n large if we did m (large) repeated trials and computed the above interval for each trial then true mean would belong to approximately 95% of m intervals calculated.

Thus numerically the above meaning seems to hold for a Normal population.