A confidence interval describes how we can use a single random sample to make inference about an unknown population parameter.

1 Large sample confidence intervals

A basic usage is to find an interval for the value of μ using $\bar{x}$. For large samples we have the probability that this random interval

$$(\bar{x} - 1.96\mathsf{SE}(\bar{x}), \bar{x} + 1.96\mathsf{SE}(\bar{x}))$$

contains μ is 0.95.

Here $\mathsf{SE}(\bar{x}) = S/\sqrt{n}$. The value of 1.96 comes from the value for the probability 0.95.

Let's see how we might use simulation to find out the probability for different values of the factor 1.96.

For instance, how often is μ in the interval $\bar{x} \pm 1\mathsf{SE}(\bar{x})$? A simulation can tell.

To simulate, we need to specify a value of μ. Let's take $\mu = 0$, and use the standard normal for our population. We take $n = 50$.

We define a function that we can easily edit:

```
> sim = function() {
+     res = c()
+     n = 50
+     k = 1
+     for (i in 1:500) {
+         x = rnorm(n)
+         xbar = mean(x)
+         SE = sd(x)/sqrt(n)
+         res[i] = ((xbar - k * SE) < 0) & ((xbar + k * SE) > 0)
+     }
+     sum(res)/500
+ }
```

To call the function, we just use its name:

```
> sim()
```

```
[1] 0.718
```

To make changes to the function use the command

```
> fix(sim)
```

and use the editor to effect the changes.

The key to this function is the logical expression that returns TRUE if μ is in the random interval $\bar{x} \pm k\mathsf{SE}$ and FALSE otherwise. Then `sum()` turns TRUE into 1, and FALSE into 0.

1. Run the above simulation with $k = 1$. Is your answer the same? Similar? (What does similar mean?)

2. Run the simulation with $k = 2$. What proportion of the time is the mean within 2 standard errors?

3. There is a relationship between k and the probability that a random interval chosen this way contains 0:

$$P(-k \leq Z \leq k) = 1 - \alpha.$$

   When $\alpha = 0.05$, $k = 1.96$ (about 2). When $k = 1$, $\alpha = 1 - 0.68$. To exactly solve for k from a given α is done with `qnorm()`:

```
> alpha = 0.05
> qnorm(1 - alpha/2)

[1] 1.96
```

   To solve for α from k is done with `pnorm()`

```
> k = 2
> 2 * (1 - pnorm(k))

[1] 0.0455
```

   For $k = 1.96$ compare your proportion from a simulation to the proper value of α. Repeat for $k = 1.5$.

2 small sample CIs

For small n, the same thing can be done, but in this case the normal distribution isn't used for the sampling distribution. Let's see if our simulations show that the relationship between k and α isn't quite right anymore.

1. Let $n = 5$, $k = 1.96$. Do a simulation and see if the value simulated for α is close to 0.05.

2. Repeat with $n = 3$, 10 and 20.

The correct relationship between k and α involves the t-distribution with $n-1$ *degrees of freedom*. The `qt()` and `pt()` function deliver the differences.

For instance, for a given k, the correct value of α *depends* on n, and is found with

```
> k = 2
> n = 5
> 2 * (1 - pt(k, df = n - 1))

[1] 0.1161
```

For comparison, the reverse is

```
> alpha = 0.05
> n = 5
> qt(1 - alpha/2, df = n - 1)

[1] 2.776
```

1. Do the simulations for $\alpha = 0.05$ and $n = 3, 5, 10, 20$. Find the appropriate k. Compare the sample proportion with α.

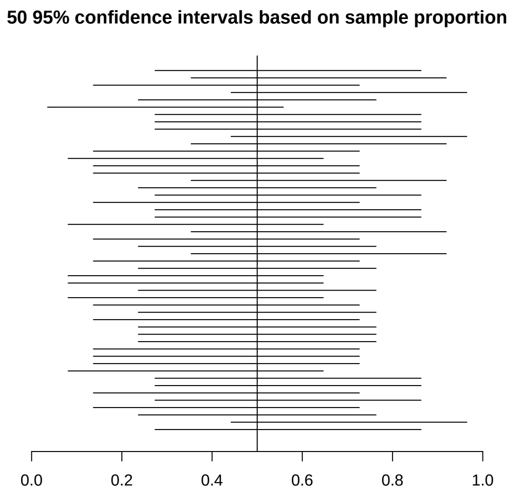
3 Viewing CIs

Download the function `plot.CI()` with this command:

```
> source("http://wiener.math.csi.cuny.edu/st/R/plotCI.R")
```

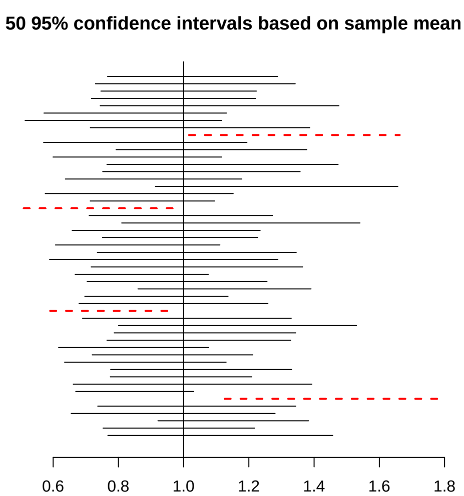
This function will produce a graphic showing simulated confidence intervals. The default is to simulate CIs based on proportions:

```
> plotCI()
```



By changing the default arguments, you can plot others. For instance the value `type="mean"` we get CIs for $\bar{x}$, and if we make family = ”exp” we will have an exponential population (with rate 1.) Does this produce the expected number of CIs containing 1 (the mean)?

```
> plotCI(n = 50, type = "mean", family = "exp", mu = 1)
```



For the defaults of $n = 10$, look at the following populations (families) to see if for $\alpha = 0.05$, the correct proportion of CIs contain the mean.

1. Exponential (`"exp"`,$\mu = 1$)
2. t with `"t"` and `df=5`
3. Uniform with `"unif"` and $\mu = 1/2$.