Constructing Confidence Intervals for Population Means

with Example of Paired Data
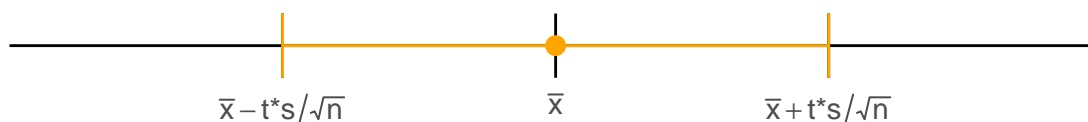
Goal: A $(1 - \alpha) \times 100$ % confidence interval for μ. Examples:

$\alpha = 0.05 \Leftrightarrow 95\%$ confidence interval.

$\alpha = 0.01 \Leftrightarrow 99\%$ confidence interval.

$\alpha = 0.1 \Leftrightarrow 90\%$ confidence interval.

Interval: $\bar{x} \pm t^* \frac{s}{\sqrt{n}}$

$$\bar{x} - t^*s/\sqrt{n} \qquad \bar{x} \qquad \bar{x} + t^*s/\sqrt{n}$$

- The **margin of error** is $t^* s/\sqrt{n}$: the amount we add and subtract from $\bar{x}$.
- $\bar{x}$ is the sample mean
- s is the sample standard deviation
- n is the sample size
- t^* is the $(1 - \frac{\alpha}{2}) * 100$ percentile of the t_{n-1} distribution. Examples:
    - $\alpha = 0.05 \Leftrightarrow 95\%$ CI $\Leftrightarrow t^* = $ 97.5th Percentile of t_{n-1} distribution.
    - $\alpha = 0.01 \Leftrightarrow 99\%$ CI $\Leftrightarrow t^* = $ 99.5th Percentile of t_{n-1} distribution.
    - $\alpha = 0.1 \Leftrightarrow 90\%$ CI $\Leftrightarrow t^* = $ 95th Percentile of t_{n-1} distribution.

Example with α = 0.05 (95% CI)

Total area to left of t* is 0.975
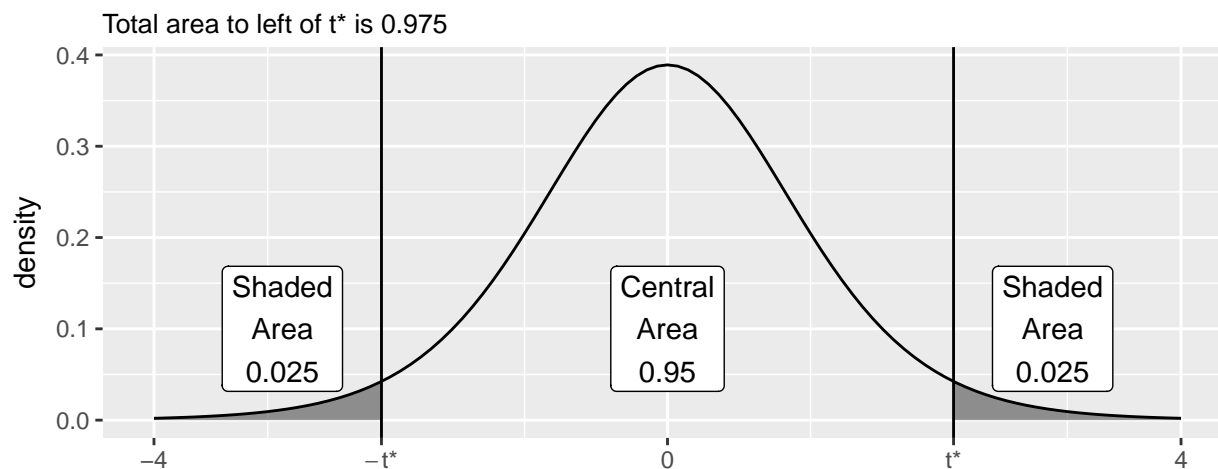
In R, to look up t^*:

```r
qt(0.975, df = 10) # For a 95% CI, sample size is n = 11
```

```
## [1] 2.228139
```

Example 1

This example was presented in Rice (2007):

> To study the effect of cigarette smoking on platelet aggregation, Levine (1973) drew blood samples from 11 individuals before and after they smoked a cigarette and measured the extent to which the blood platelets aggregated. Platelets are involved in the formation of blood clots, and it is known that smokers suffer more often from disorders involving blood clots than do nonsmokers. The data are shown in the following table, which gives the maximum percentage of all the platelets that aggregated afer being exposed to a stimulus.

```
platelets <- read_csv("http://www.evanlray.com/stat343_s2018/data/rice/Chapter%2011/platelet.txt")
```

```
## Parsed with column specification:
## cols(
##   `'before'` = col_integer(),
##   `'after'` = col_integer()
## )
```

```
colnames(platelets) <- c("before", "after")
head(platelets)
```

```
## # A tibble: 6 x 2
##   before after
##    <int> <int>
## 1     25    27
## 2     25    29
## 3     27    37
## 4     44    56
## 5     30    46
## 6     67    82
```

```
nrow(platelets)
```

```
## [1] 11
```

This is an example of **paired data**:

- We have two measurements on each person (these are **not independent**!)
- We are interested in the **difference** between these measurements on the same person
- These **differences are independent** across different people (or at least, this seems plausible)

```
platelets <- platelets %>%
  mutate(difference = after - before)
head(platelets)
```
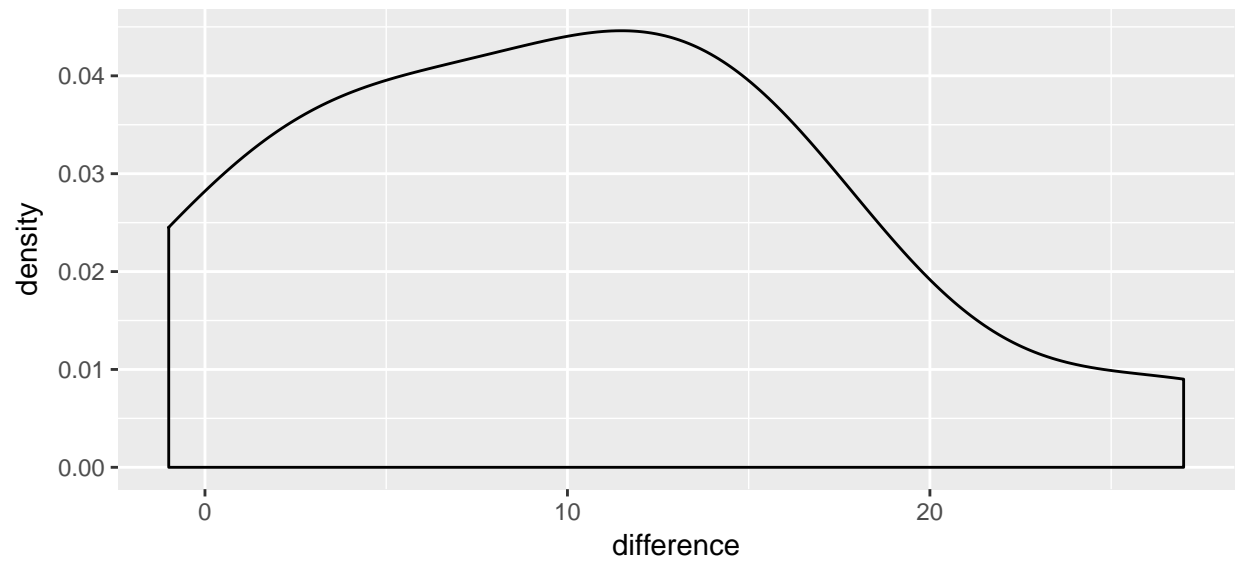
```
## # A tibble: 6 x 3
##   before after difference
##    <int> <int>      <int>
## 1     25    27          2
## 2     25    29          4
## 3     27    37         10
## 4     44    56         12
## 5     30    46         16
## 6     67    82         15
```

(a) What is the population parameter of interest?

(b) Check the conditions for inference with these data

```
ggplot(data = platelets, mapping = aes(x = difference)) +
  geom_density()
```

(c) Find a 95% confidence interval for the population parameter, and verify that it agrees with the output from `t.test`.

```
platelets %>%
  summarize(
    mean_difference = mean(difference),
    sd_difference = sd(difference))
```

```
## # A tibble: 1 x 2
##   mean_difference sd_difference
##             <dbl>         <dbl>
## 1        10.27273      7.976101
```

```
qt(0.975, df = 11 - 1)
```

```
## [1] 2.228139
```

```
10.27273 -  2.228139 * 7.976101 / sqrt(11)
```

```
## [1] 4.914312
```

```
10.27273 +  2.228139 * 7.976101 / sqrt(11)
```

```
## [1] 15.63115
```

```
t.test(~ difference, data = platelets)
```

```
## ~difference
##
##   One Sample t-test
##
## data:  difference
## t = 4.2716, df = 10, p-value = 0.001633
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##    4.91431 15.63114
## sample estimates:
## mean of x
##  10.27273
```