# T-Distribution

## Introduction

This week, we are introduced to the t-distribution. The t-distribution describes the distribution of sample means around their mean value. Contrast this with the Z-distribution, which describes the distribution of individuals around the population mean.

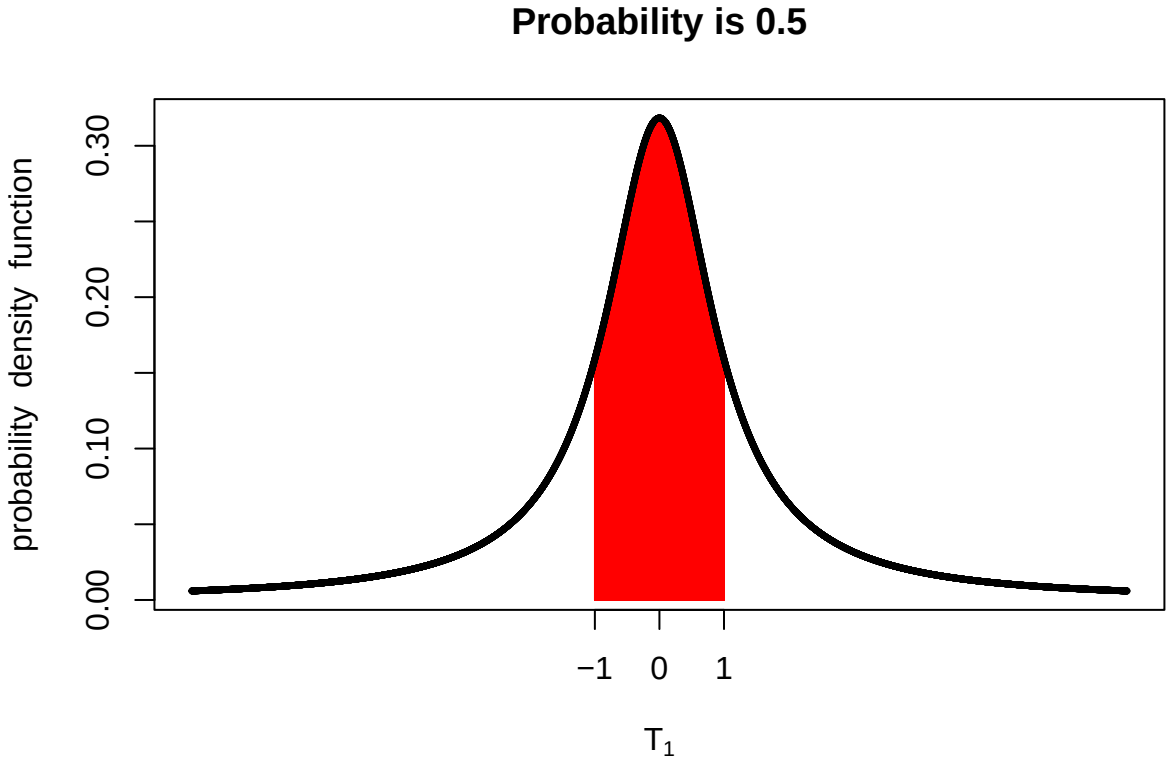
The most critical way in which the t-distribution is different from the Z-distribution is the shape of the t-distribution changes shape with the number of samples used to calculate the sample mean. As the number of samples increases, the t-distribution narrows in width, and expands in height.

## Plotting the Distribution

In the lecture, we once again used the "shadeDist" function from the fastGraph package. Use install.packages("fastGraph") to install fastGraph in your account if you did not do so in the last unit. Then use library(fastgraph) to tell R to run that package in the current section.

We can use the shadeDist function to observe how the difference of the t-Distribution changes with the number of samples taken. Say we want to see the t-distribution for a sample size of 2. We will shade the area from t=-1 to t=1 for reference.

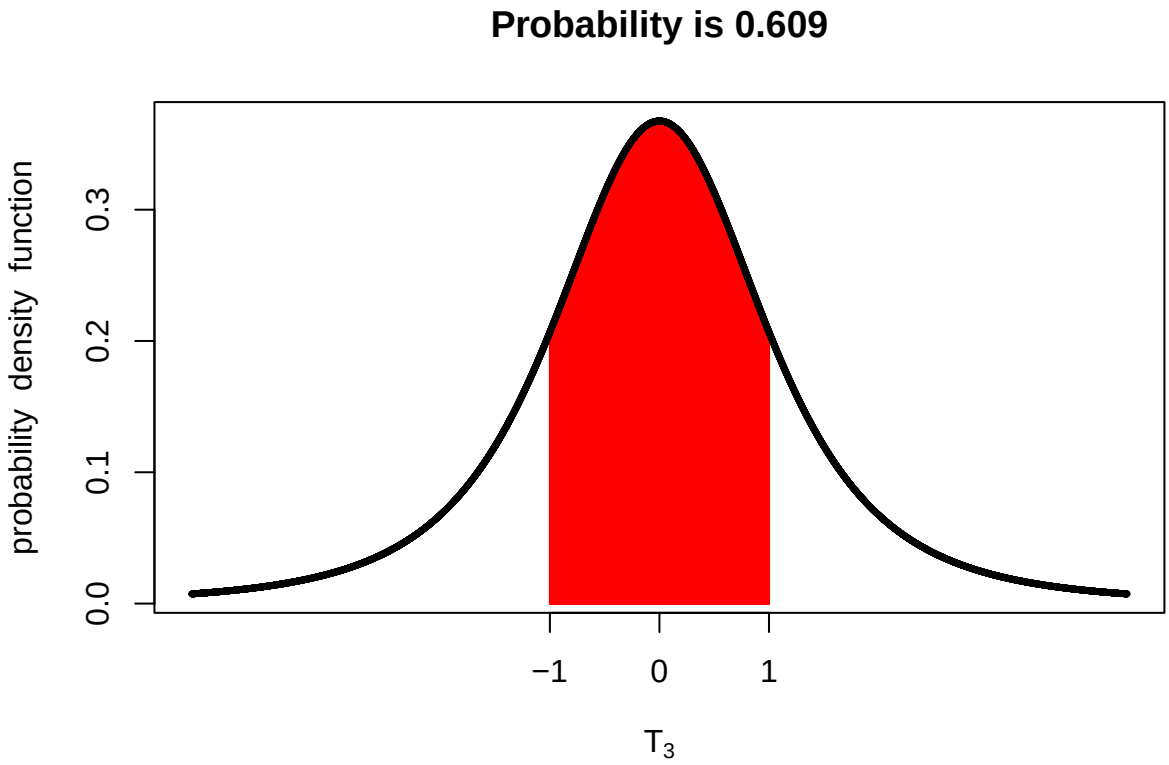
```
library(fastGraph)
shadeDist(c(-1,1), "dt", 1, lower.tail=FALSE )
```



Remember, we must provide four arguments to the shadeDist function. The first, "c(-1,1)", is the range of t-values. The second, "dt", tells R we are modeling the t-distribution. The third, 3 is the degrees of freedom associated with t, which is always one less than the number of samples. the final argument, "lower.tail=FALSE", tells R to shade the middle of the plot.

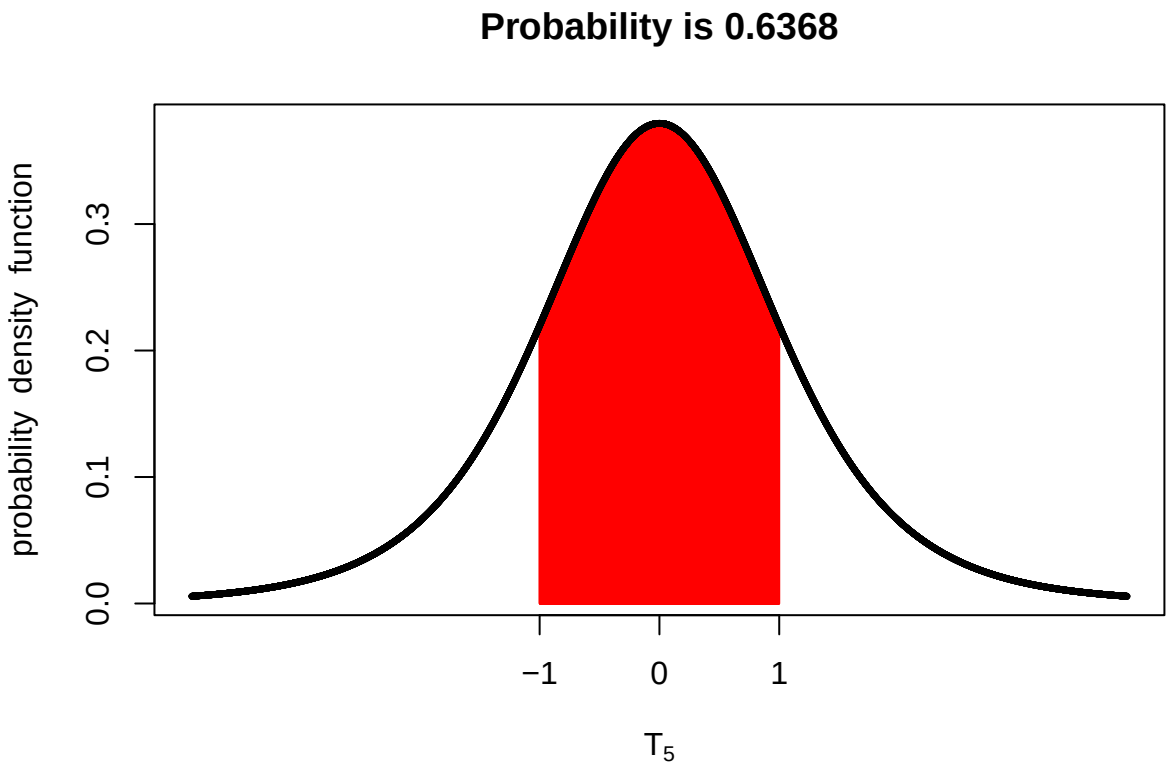
Now lets change the number of samples to 4.

```
shadeDist(c(-1,1), "dt", 3, lower.tail=FALSE )
```



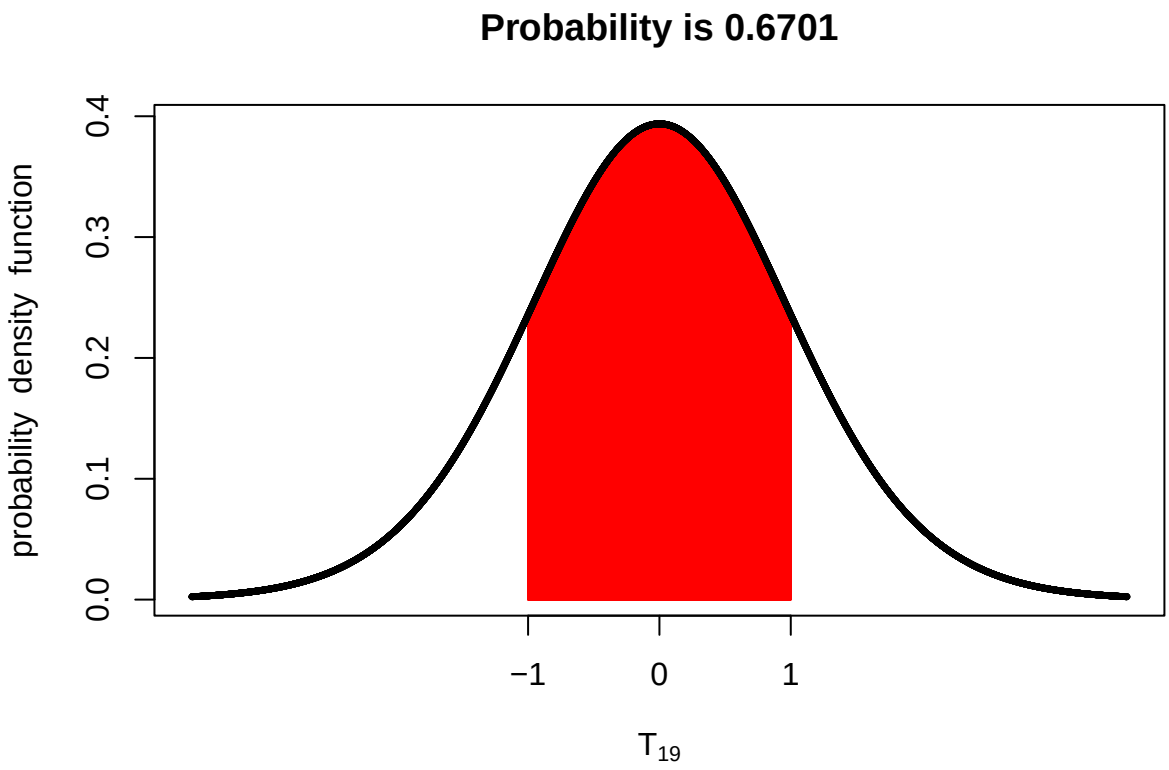

We can see the shaded area gets wider, meaning the distribution is getting narrower. In we increase to 6, the shape doesn't seem to change much, yet the proportion of the curve that is between t=-1 and t=1 increases.

```
shadeDist(c(-1,1), "dt", 5, lower.tail=FALSE )
```




If we increase to n=20, the curve continues to narrow.

```
shadeDist(c(-1,1), "dt", 19, lower.tail=FALSE )
```



## Calculating T

To create the plot above, we entered t-values, and the degree of freedom. R not only drew the plot, but calculated the probability of observing sample means within the given range.

To calculate t, we need to supply the degrees of freedom and the target probability to which t should respond. We will use the "qt" function in R to do this.

Let's say we want to know the t-values that, given four samples per mean, would be expected to include 50% of potential sample means? In that case, we are interested in the middle 50% of the distribution. There will be 25% of the distribution below this range, and 25% above. So, for the lower t-value, we will tell R to calculate the t-value below which 25% of sample means are expected to occur.

```
qt(0.25,3)
```

```
## [1] -0.7648923
```

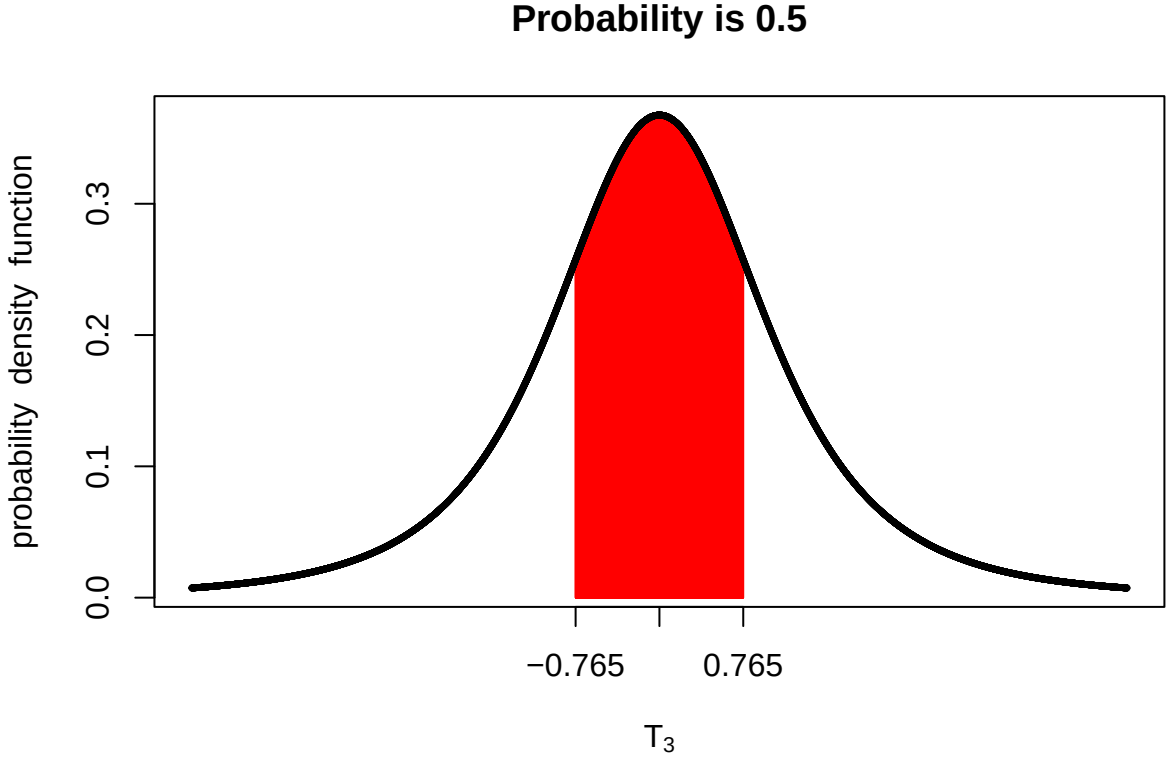
In the argument above, we told R to calculate the t-value associated with the lower 0.25 of the population, and 3 degrees of freedom (for 4 samples). For the upper t-value, we want the value where 75% of the distribution is lower.

```
qt(0.75,3)
```

```
## [1] 0.7648923
```

As you can see, the upper and lower limit are symmetrical – they only differ in sign. We can use the shadeDist from above to check our work, by plugging in the t-values we just calculated.

```
shadeDist(c(-0.7649,0.7649), "dt", 3, lower.tail=FALSE ) # t with 15 d.f. and non-centrality parameter=
```




As we can see, the defined proportion equals our target.

What if we want the t-values that bind the middle 68% of potential sample means? Then we would have 32% of the distribution outside this range. 16% of sample means would be expected to be below the range, and 16% percent above.

The t-value associated with the lower range, again assuming 3 df, would be:

```
qt(0.16,3)
```

```
## [1] -1.188929
```

And the t-value associated with the upper range would be:

```
qt(0.84,3)
```

```
## [1] 1.188929
```

How about the t-values that would be expected to define the range where 90% of sample means would be expected?

```
qt(0.05,3)
```

```
## [1] -2.353363
```

```
qt(0.95,3)
```

```
## [1] 2.353363
```

And, finally, 95%?

```
qt(0.025,3)
```

```
## [1] -3.182446
```

```
qt(0.975,3)
```

```
## [1] 3.182446
```

## Practice

Calculate the t-values that would define the middle 60% of potential sample means if there were 10 samples.

The lower limit is:

```
qt(0.2,9)
```

```
## [1] -0.8834039
```

What is the upper limit?

Q: What is the range of t-values associated with 70% of potential sample means, if the number of samples is 12? A: (-1.09, 1.09)

Q: What is the range of t-values associated with 80% of potential sample means, if the number of samples is 14? A: (-1.35, 1.35)

Q: What is the range of t-values associated with 90% of potential sample means, if the number of samples is 17? A: (-1.75, 1.75)

Q: What is the range of t-values associated with 95% of potential sample means, if the number of samples is 20? A: (-2.09, 2.09)

Q: What is the range of t-values associated with 99% of potential sample means, if the number of samples is 20? A: (-2.86, 2.86)