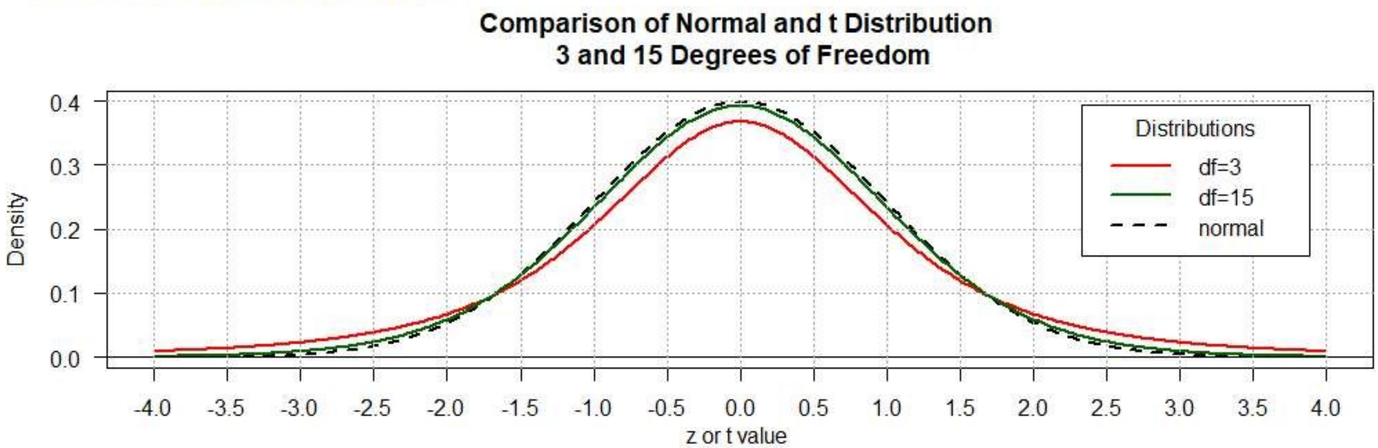
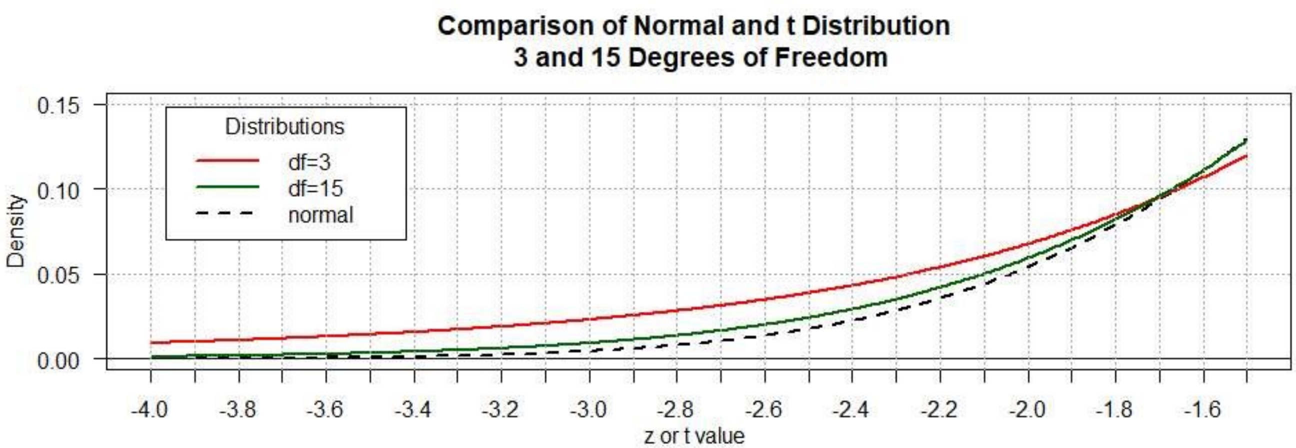
Topic 11 e: Student's - t Distribution

The Student's-t distribution is another continuous, symmetric, somewhat bell-shaped distribution. It is different from the standard normal distribution, N(0, 1), in that the standard Student's-t distribution has different forms each denoted by some "degrees of freedom" value.

The degrees of freedom is usually a positive integer. We can see the difference between two forms of the Student's-t, one with 3 degrees of freedom and one with 15 degrees of freedom, and the Normal distribution in the following graph.



Here are the key points that we can learn from that plot. The Student's-t is similar to the Normal. The higher the degrees of freedom the closer the Student's-t is to the Normal. The Student's-t has spread out the area under the curve so that more of the area is under the tails. We can see that final point by looking at the lower end of the plot in more detail.



Now, because the Student's-t distribution changes for every different number of degrees of of freedom, we would need a different table for each different number of degrees of freedom. (See web page for a link.)

Compare areas under the curve to the left of -2.8

For Student's-t df=3

-2.80	0.0339

For Student's-t df=15

-2.80	0.0067

For the Normal

-2.8	0.0026

Having different distributions for different numbers of degrees of freedom means that, if we follow the approach for N(0, 1), we need a different table for each different distribution. On the web it is easy to "publish" 100 or more different tables. A text book publisher would find it prohibitively expensive to do so in a textbook.

The solution was to publish a single table that just gave "**critical values**" for the Student's-t and for may different degrees of freedom. The choice of "critical values" was limited, generally to 0.40, 0.25, 0.20, 0.10, 0.05, 0.025, 0.01, 0.005, 0.0025, and 0.001. Note that because the **Student's-t** is symmetric, with a little computation, the table also gives values for 0.60, 0,75, 0.80, 0.90, 0.95, 0.975, 0.99, 0.995, 0.9975, and 0.999. Also, the table gave **positive t** values that had the specified area to the **right** of that value. Again, by symmetry, that meant that the **negative** of the table value had the specified area to the **left**. Here is a small fragment of such a Critival Values table.

Critical Values of Student's t

Values in the table represent the critical **t** value needed to have the desired area to the **right** of that critical value.

Degrees of Freedom	**0.4000**	**0.2500**	**0.2000**	**0.1000**	**0.0500**	**0.0250**	**0.0100**	**0.0050**	**0.0025**	**0.0010**
1	0.325	1.000	1.376	3.078	6.314	12.706	31.821	63.657	127.321	318.309
2	0.289	0.816	1.061	1.886	2.920	4.303	6.965	9.925	14.089	22.327
3	0.277	0.765	0.978	1.638	2.353	3.182	4.541	5.841	7.453	10.215
15	0.258	0.691	0.866	1.341	1.753	2.131	2.602	2.947	3.286	3.733
30	0.256	0.683	0.854	1.310	1.697	2.042	2.457	2.750	3.030	3.385

Being able to use these tables is an important outcome of this course. It is a skill worth acquiring! However, once we have R we no longer have a need for the tables.

In R if we want to find the area to the left of -1.43 for a Student's-t with 7 degrees of freedom we just use the command pt(-1.43, 7).

```
     # for Student's-t with 7 degrees of freedom
     # find P( X < -1.43 )
pt( -1.43, 7 )
```

```
>      # for Student's-t with 7 degrees of freedom
>      # find P( X < -1.43 )
> pt( -1.43, 7 )
[1] 0.09789789
```

For a Student's-t with 15 degrees of freedom,

1) find P(X < -1.3) =

2) find P(X > 2.6) =

3) find P(X < -1.83 or X > 1.54) =

4) find P(X < -2.45 or X > 2.45) =

5) find P(0.48 < X < 1.76) =

6) find P(-2.37 < X < 2.37) =

```
     # for Student's-t with 15 degrees of freedom
     #  1) find P( X < -1.3 )
pt( -1.3, 15 )
     #  2) find P( X > 2.6 )
1 - pt( 2.6 ,15 )    # old way
pt( 2.6, 15, lower.tail = FALSE)    # better way
     #  3) find P( X < -1.83 or X > 1.54 )
pt( -1.83, 15 ) + (1 - pt( 1.54, 15) )  #old way
pt( -1.83, 15 ) + pt( 1.54, 15,
                      lower.tail=FALSE ) #better way
     #  4) find P( X < -2.45 or X > 2.45 )
pt( -2.45, 15 ) + (1 - pt( 2.45, 15) )  #old way
pt( -2.45, 15 ) + pt( 2.45, 15,
                      lower.tail=FALSE ) #better way
             # or, using symmetry
2 * pt( -2.45, 15 )
```

```
>       # for Student's-t with 15 degrees of freedom
>       #  1) find P( X < -1.3 )
> pt( -1.3, 15 )
[1] 0.1066117
>       #  2) find P( X > 2.6 )
> 1 - pt( 2.6 ,15 )    # old way
[1] 0.0100495
> pt( 2.6, 15, lower.tail = FALSE)    # better way
[1] 0.0100495
>       #  3) find P( X < -1.83 or X > 1.54 )
> pt( -1.83, 15 ) + (1 - pt( 1.54, 15) )  #old way
[1] 0.1157904
> pt( -1.83, 15 ) + pt( 1.54, 15,
+                         lower.tail=FALSE ) #better way
[1] 0.1157904
>       #  4) find P( X < -2.45 or X > 2.45 )
> pt( -2.45, 15 ) + (1 - pt( 2.45, 15) )  #old way
[1] 0.02704137
> pt( -2.45, 15 ) + pt( 2.45, 15,
+                         lower.tail=FALSE ) #better way
[1] 0.02704137
>             # or, using symmetry
> 2 * pt( -2.45, 15 )
[1] 0.02704137
```

```
      #  5) find P( 0.48 < X < 1.76 )
pt( 1.76, 15) - pt( 0.48, 15 )
      #  6) find P( -2.37 < X < 2.37 )
pt( 2.37, 15) - pt( -2.37, 15 )
          #  or, using the complement
1 - 2*pt( -2.37, 15 )
```

```
>       #  5) find P( 0.48 < X < 1.76 )
> pt( 1.76, 15) - pt( 0.48, 15 )
[1] 0.2696865
>       #  6) find P( -2.37 < X < 2.37 )
> pt( 2.37, 15) - pt( -2.37, 15 )
[1] 0.9683769
>           #  or, using the complement
> 1 - 2*pt( -2.37, 15 )
[1] 0.9683769
```

For problems that used the "areas" given in the critical value table, solving problems such as "For a standard Student's-t distribution with 19 degrees of freedom, find the value for y such That $P(X > y) = 0.025$." However, for any "areas" not given in that table the task was much harder. For us, with R, the task is straight forward no matter what "area" is desired. We use the qt() function.

For a standard Student's-t distribution with 11 degrees of freedom find the value of y such that:

7) $P(X < y) = 0.33$
8) $P(X > y) = 0.075$
9) $P(X < -y \text{ or } X > y) = 0.08$
10) $P(-y < X < y) = 0.033$

```
      # For Student's-t with 11 degrees of freedom
      # find the value of  y  such that
      #  7) P( X < y ) = 0.33
qt( 0.33, 11 )
      #  8) P( X > y ) = 0.075
qt( 1 - 0.075, 11 )   # the old, ugly way
qt( 0.075, 11, lower.tail = FALSE )    # the better way
      #  9) P( X < -y  or X > y ) = 0.08
-qt( 0.08/2, 11 ) # note the leading -
qt( 0.08/2, 11, lower.tail=FALSE)  # more clear statement
      # 10) P( -y < X < y ) = 0.033
qt( (1 - 0.033)/2, 11, lower.tail=FALSE)
```

```
>       # For Student's-t with 11 degrees of freedom
>       # find the value of  y  such that
>       #  7) P( X < y ) = 0.33
> qt( 0.33, 11 )
[1] -0.4520743
>       #  8) P( X > y ) = 0.075
> qt( 1 - 0.075, 11 )   # the old, ugly way
[1] 1.54756
> qt( 0.075, 11, lower.tail = FALSE )    # the better way
[1] 1.54756
>       #  9) P( X < -y  or X > y ) = 0.08
> -qt( 0.08/2, 11 ) # note the leading -
[1] 1.928427
> qt( 0.08/2, 11, lower.tail=FALSE)  # more clear statement
[1] 1.928427
>       # 10) P( -y < X < y ) = 0.033
> qt( (1 - 0.033)/2, 11, lower.tail=FALSE)
[1] 0.04232258
```

This leaves us with the question of how do we solve a problem where the mean is not 0 and/or the standard deviation is not 1. Remember that for pnorm() and qnorm() we could specify parameters mean=46.3 and sd=7.2 to get those functions to behave as if the mean was 46.3 and the standard deviation was 7.2. There are no such parameters for pt() and qt().

For a Student's-t with mean=154.3 and standard deviation=13.2 and 24 degrees of freedom, find P(X < 140.0)? First, normalize the 140 value t = (140 - 154.3) / 13.2. Thus, t = -1.08333 so now use pt(t, 24).

```
      # when mean=154.3 and standard deviation = 13.2
      # and for 24 degrees of freedom, find P( X < 140.0 )
t <- (140 -154.3)/13.2
pt( t, 24 )
```

```
>       # when mean=154.3 and standard deviation = 13.2
>       # and for 24 degrees of freedom, find P( X < 140.0 )
> t <- (140 -154.3)/13.2
> pt( t, 24 )
[1] 0.1447123
```

Going in the other direction, for a Student's-t distribution with 27 degrees of freedom and mean=483.2 with standard deviation=26.4 find a value for y such that P(X < y) = 0.15? First, find the standard t value using qt(0.15, 27). Then use the inverse normalization to get the desired value, i.e., find t*26.4 + 483.2.

```
      # when mean=483.2 and standard deviation=26.4
      # with 27 degrees of freedom, find  y  such
      # that P( X < y ) = 0.15.
t <- qt( 0.15, 27 )
t
t*26.4 + 483.2
```

```
>       # when mean=483.2 and standard deviation=26.4
>       # with 27 degrees of freedom, find  y  such
>       # that P( X < y ) = 0.15.
> t <- qt( 0.15, 27 )
> t
[1] -1.056727
> t*26.4 + 483.2
[1] 455.3024
```