A Plot Method for "htest" Objects

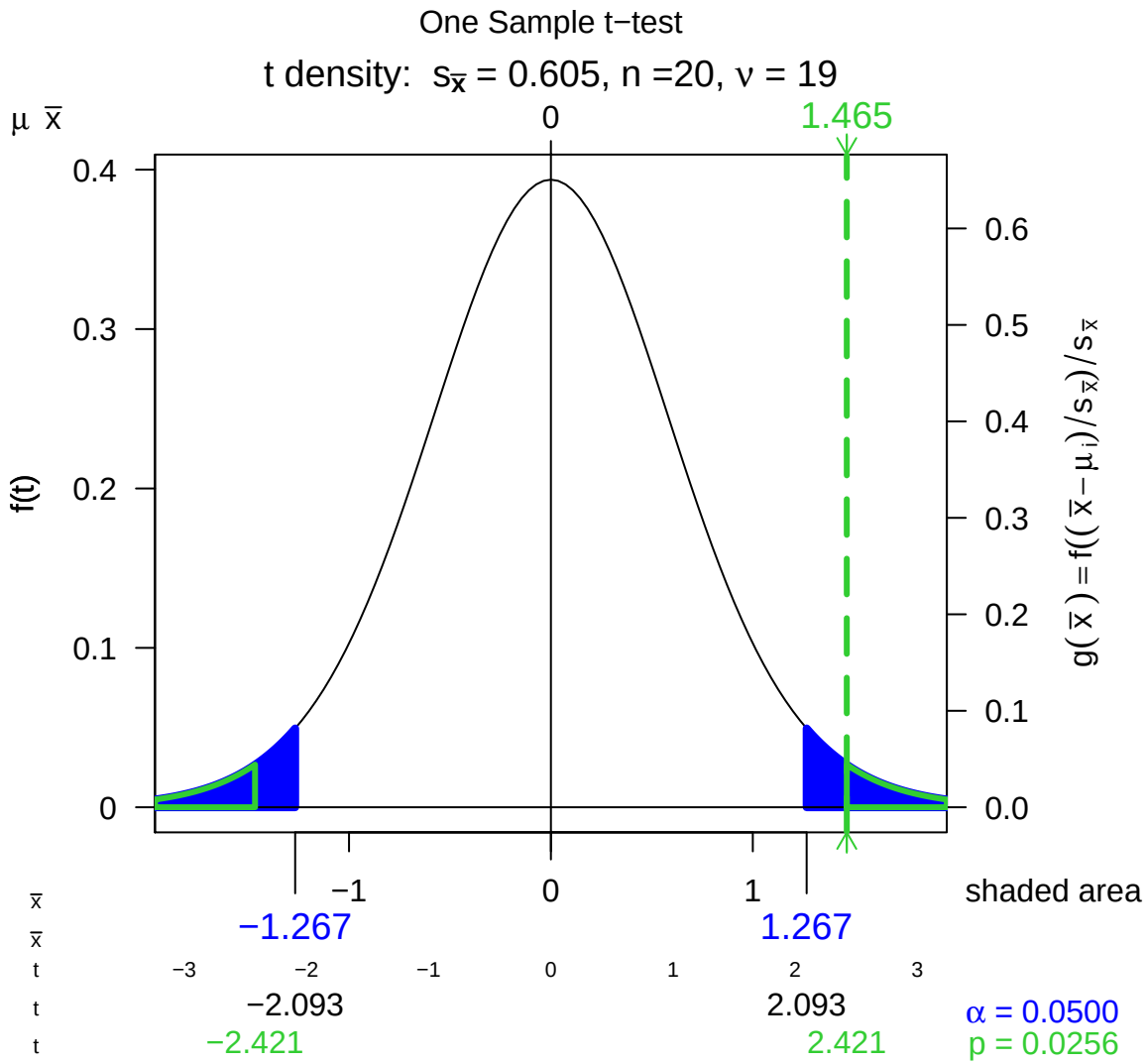
Richard M. Heiberger G. Jay Kerns

Temple University Youngstown State University

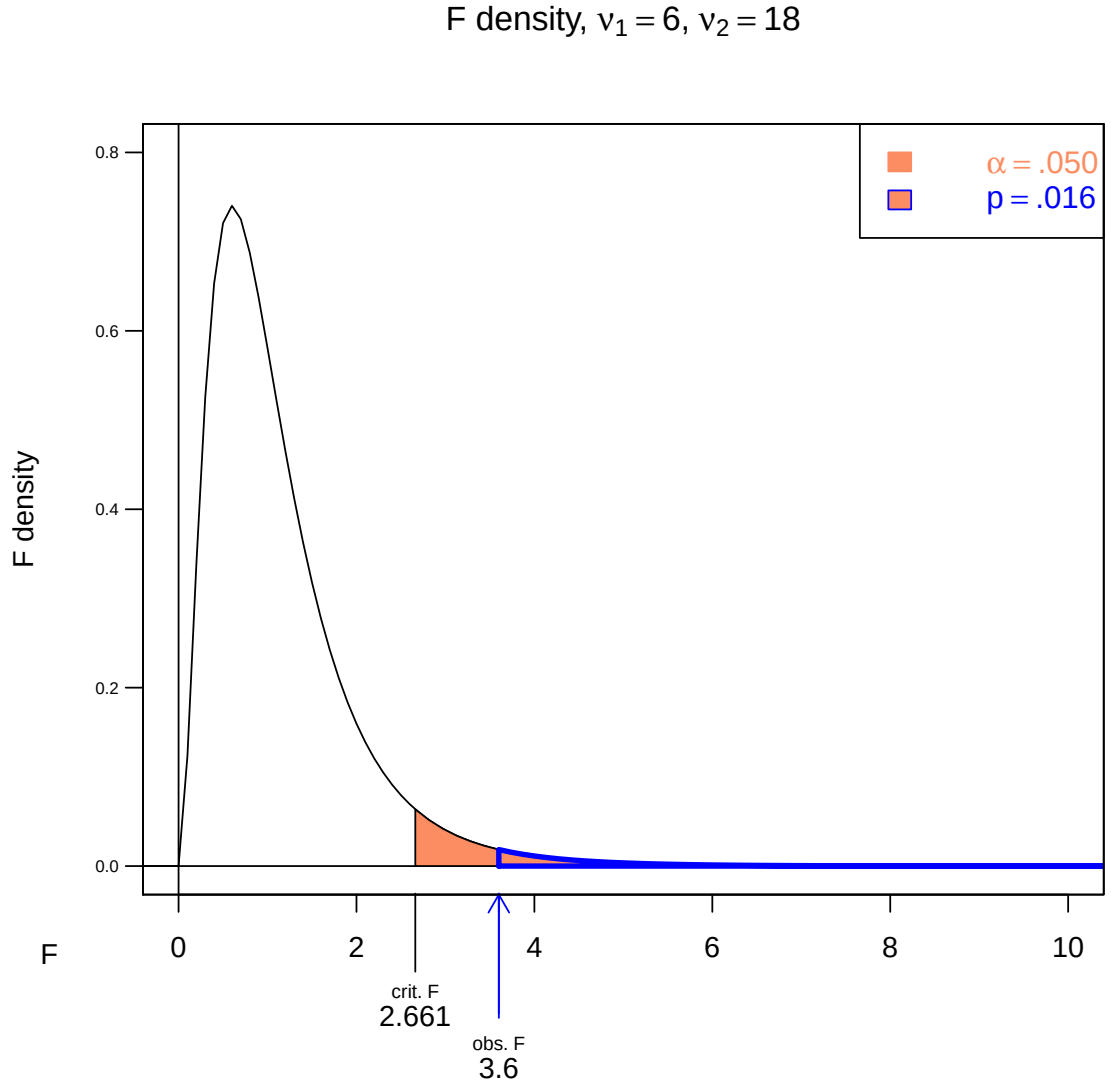
The numerical results of many statistical tests in R are stored in an "htest" object. The print method for the class displays a table. We have written a generic plot.htest function for the class that calls the graphing functions in the HH package.

```
> z <- rnorm(20, 2, 3)
> t.test(z)

        One Sample t-test

data:  z
t = 2.4213, df = 19, p-value = 0.02564
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 0.1986526 2.7321318
sample estimates:
mean of x
 1.465392
```

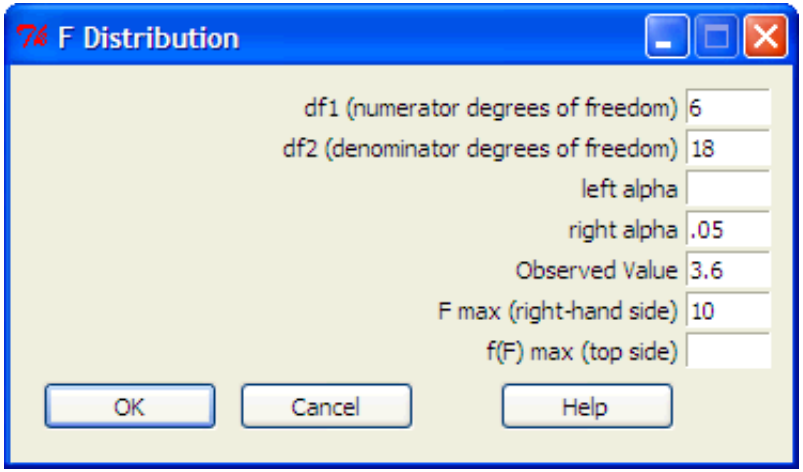
```
> z <- rnorm(20, 2, 3)
> t.test(z)

        One Sample t-test

data:  z
t = 2.4213, df = 19, p-value = 0.02564
alternative hypothesis:
     true mean is not equal to 0
95 percent confidence interval:
 0.1986526 2.7321318
sample estimates:
mean of x
 1.465392

> plot(t.test(z))
>
```

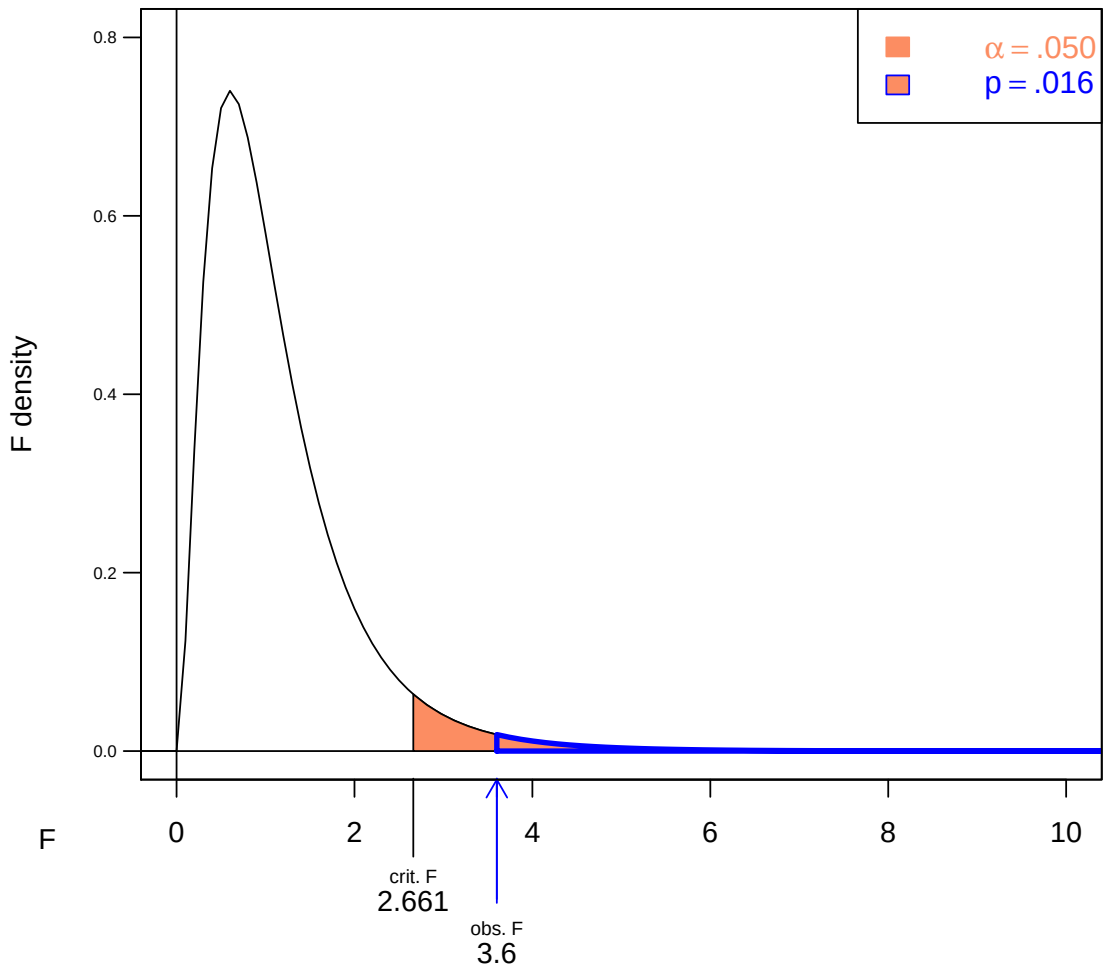


```
> vt <- var.test(x, y, alt="greater")
> vt

        F test to compare two variances

data:  x and y
F = 3.6, num df = 6, denom df = 18,
     p-value = 0.01598
alternative hypothesis:
    true ratio of variances
     is greater than 1
95 percent confidence interval:
 1.35272     Inf
sample estimates:
ratio of variances
                3.6

> plot(vt)
```



We can specify the graph from the Rcmdr menu.





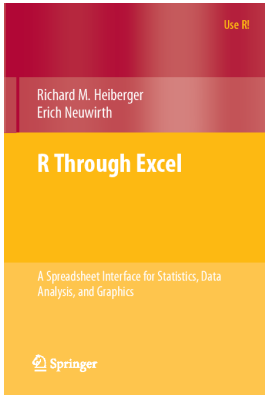
RExcel

Microsoft Excel is the most widely used spreadsheet program. Many of our clients and students use it as their data management system and as their working environment.

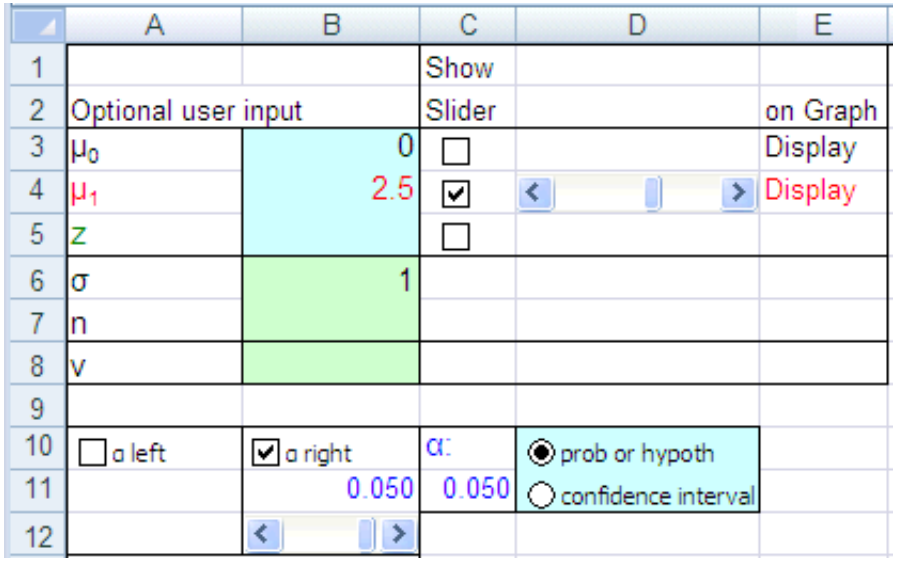
On Windows `RExcel` and `statconnDCOM` (`http://rcom.univie.ac.at`) access COM—the Microsoft interprocess communications system, and seamlessly integrates the entire set of R's statistical and graphical methods into Excel.

Therefore these `"htest"` graphing methods are available through RExcel.

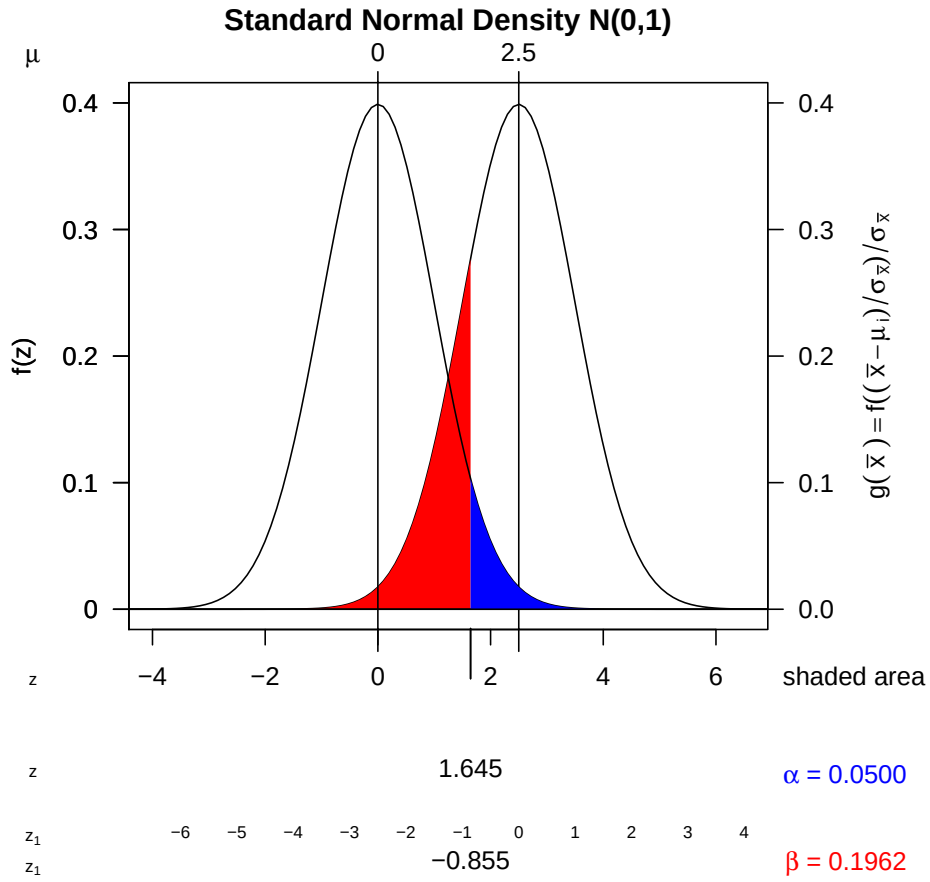
The normal example is from the recent book *R through Excel* by Richard Heiberger and Erich Neuwirth. It is possible to build half the introductory course on this one graph.



There are many more examples included with the software. There are several other talks at this UserR! 2010 conference illustrating applications of RExcel.

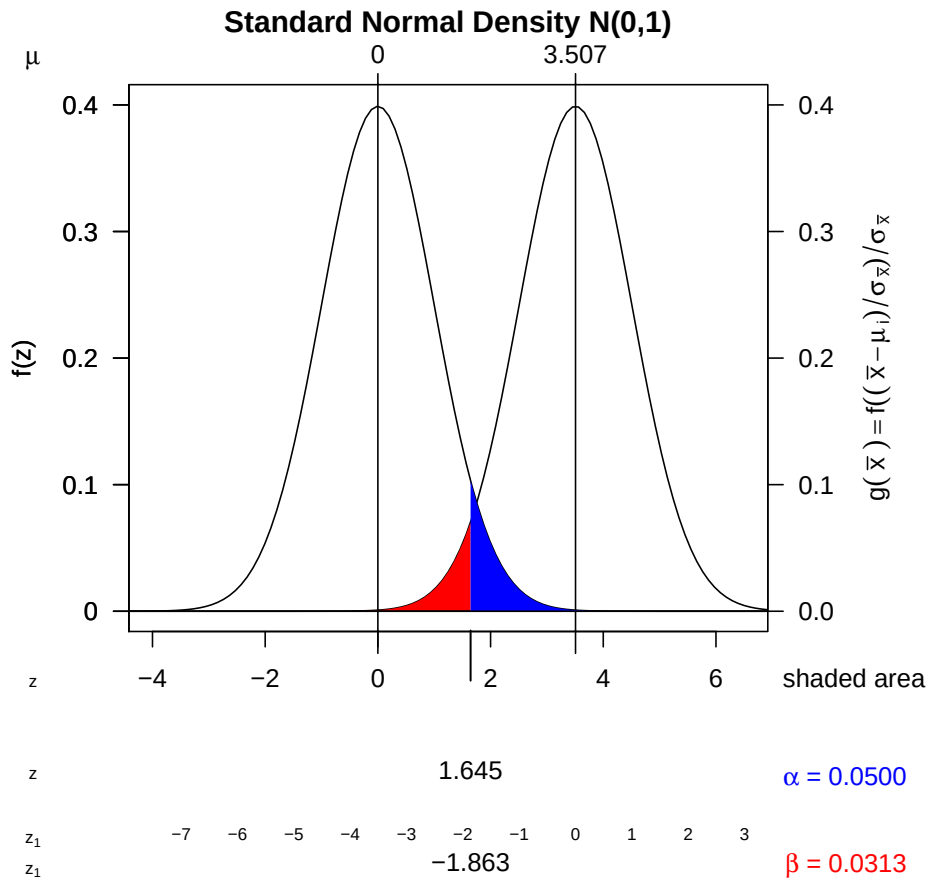


	A	B	C	D	E
1			Show		
2	Optional user input		Slider		on Graph
3	μ_0	0	☐		Display
4	μ_1	2.5	☑	◀ ▶	Display
5	z		☐		
6	σ	1			
7	n				
8	v				
9					
10	☐ a left	☑ a right	α:	◉ prob or hypoth	
11		0.050	0.050	○ confidence interval	
12		◀ ▶			

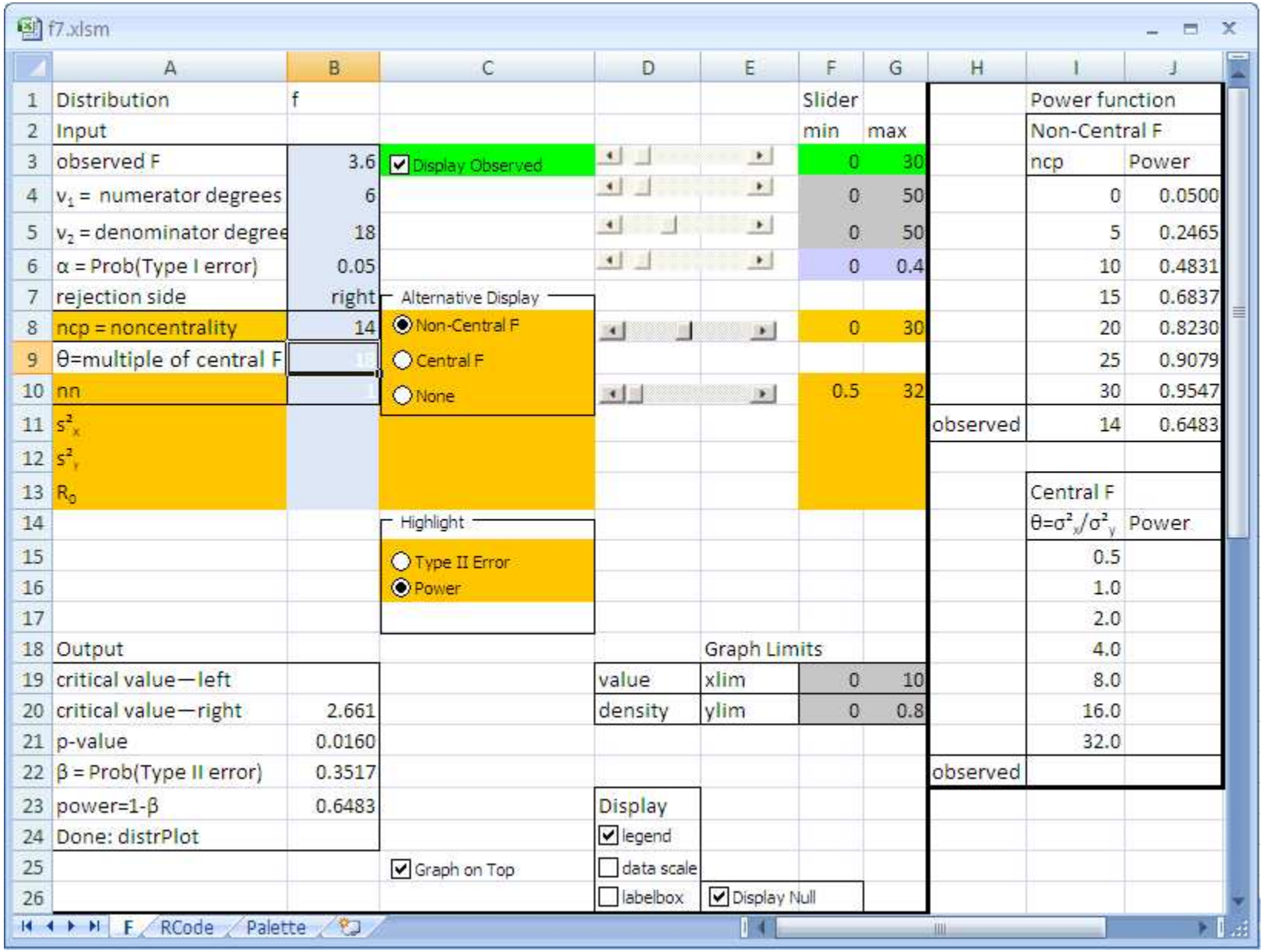


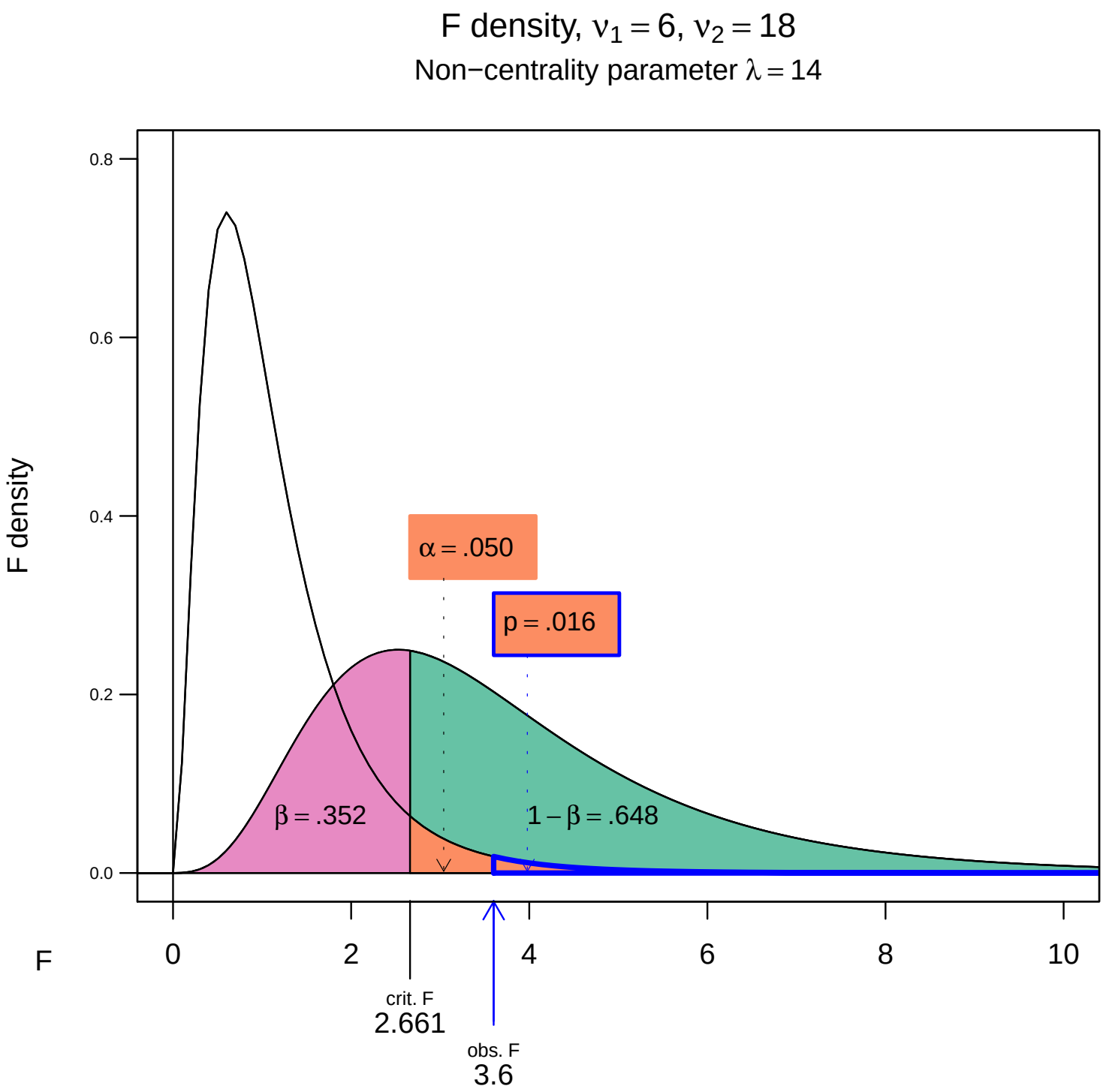
Placing values in the cells in Excel provides live control of the curve displayed in the R graph. The slider on μ_1 in Excel, smoothly moves the normal curve centered at μ_1 and adjusts the corresponding area illustrating β, the probability of the Type II Error.

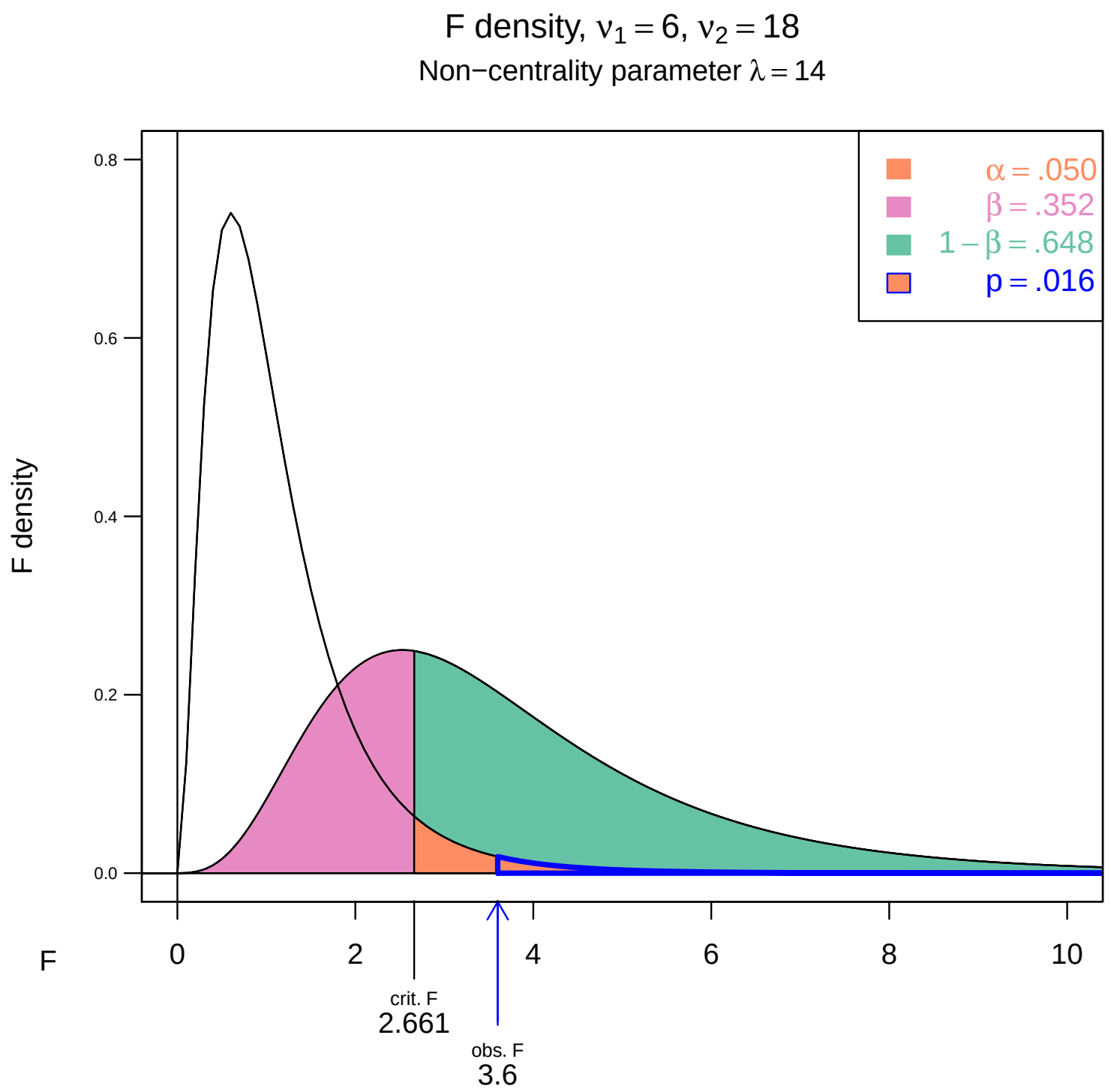
	A	B	C	D	E
1			Show		
2	Optional user input		Slider		on Graph
3	μ_0	0	☐		Display
4	μ_1	3.5075	☑	◀ ▶	Display
5	z		☐		
6	σ	1			
7	n				
8	v				
9					
10	☐ a left	☑ a right	α:	◉ prob or hypoth	
11		0.050	0.050	○ confidence interval	
12		◀ ▶			

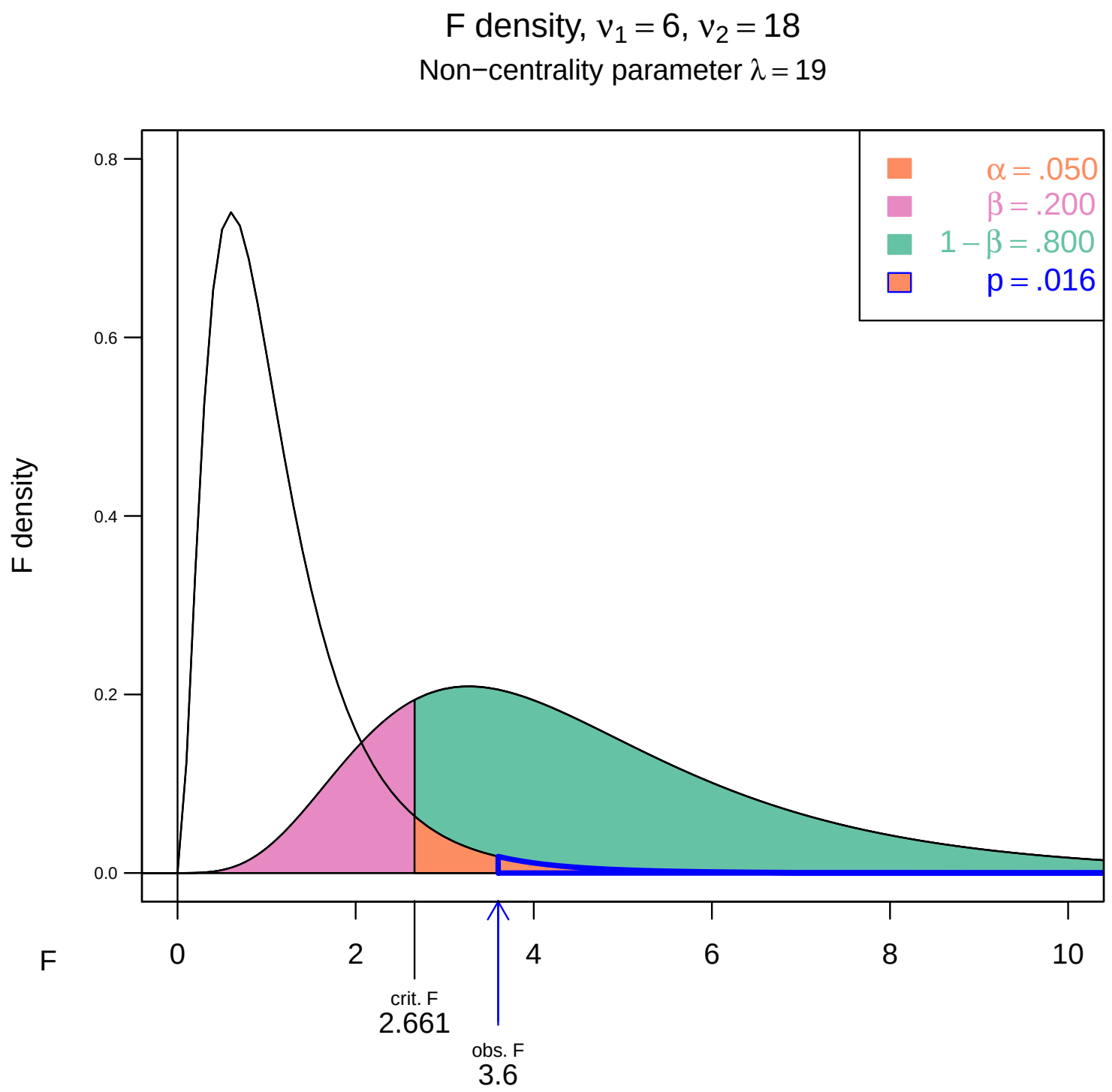


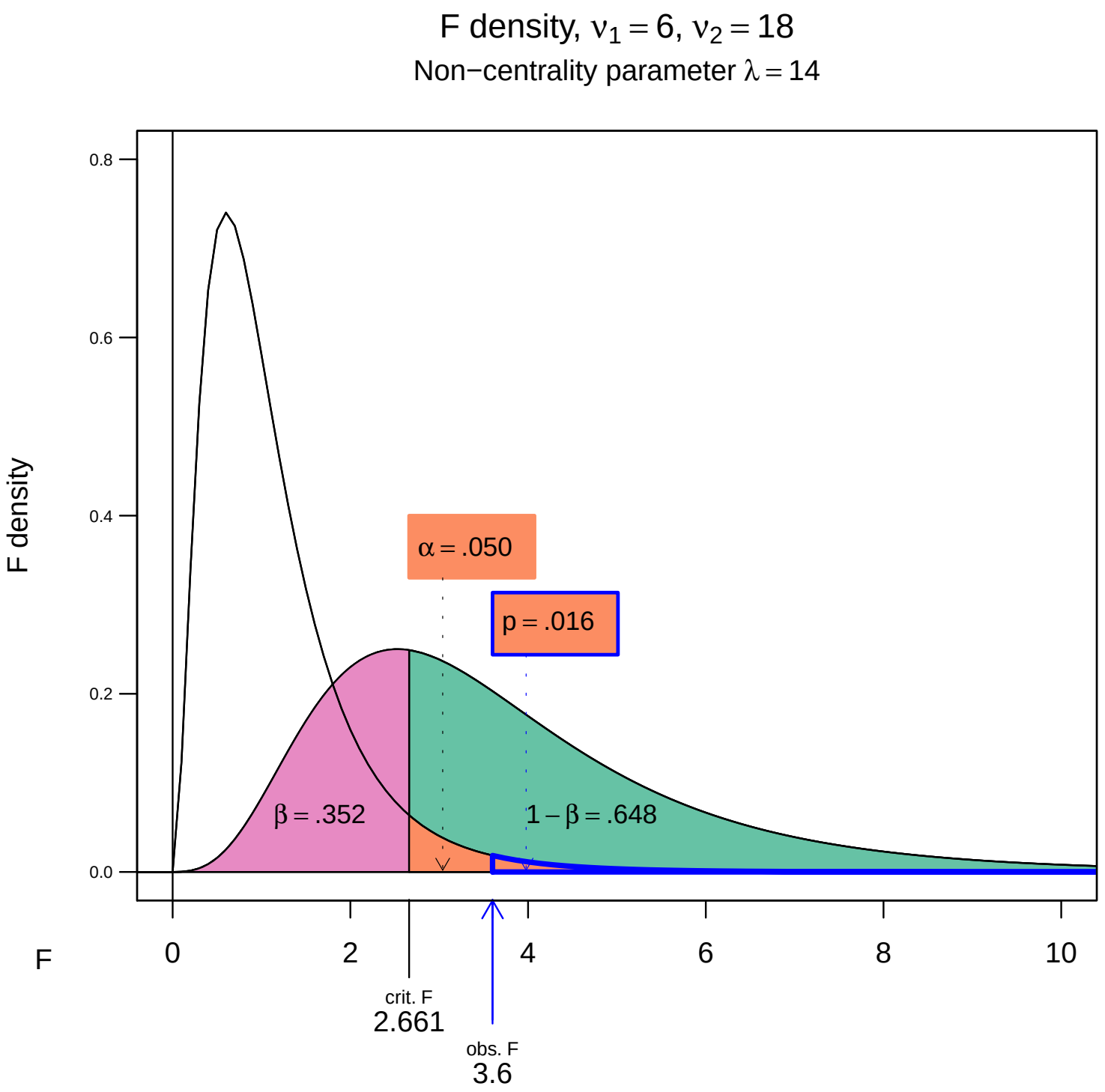
Placing values in the cells in Excel provides live control of the curve displayed in the R graph. The slider on μ_1 in Excel, smoothly moves the normal curve centered at μ_1 and adjusts the corresponding area illustrating β, the probability of the Type II Error.

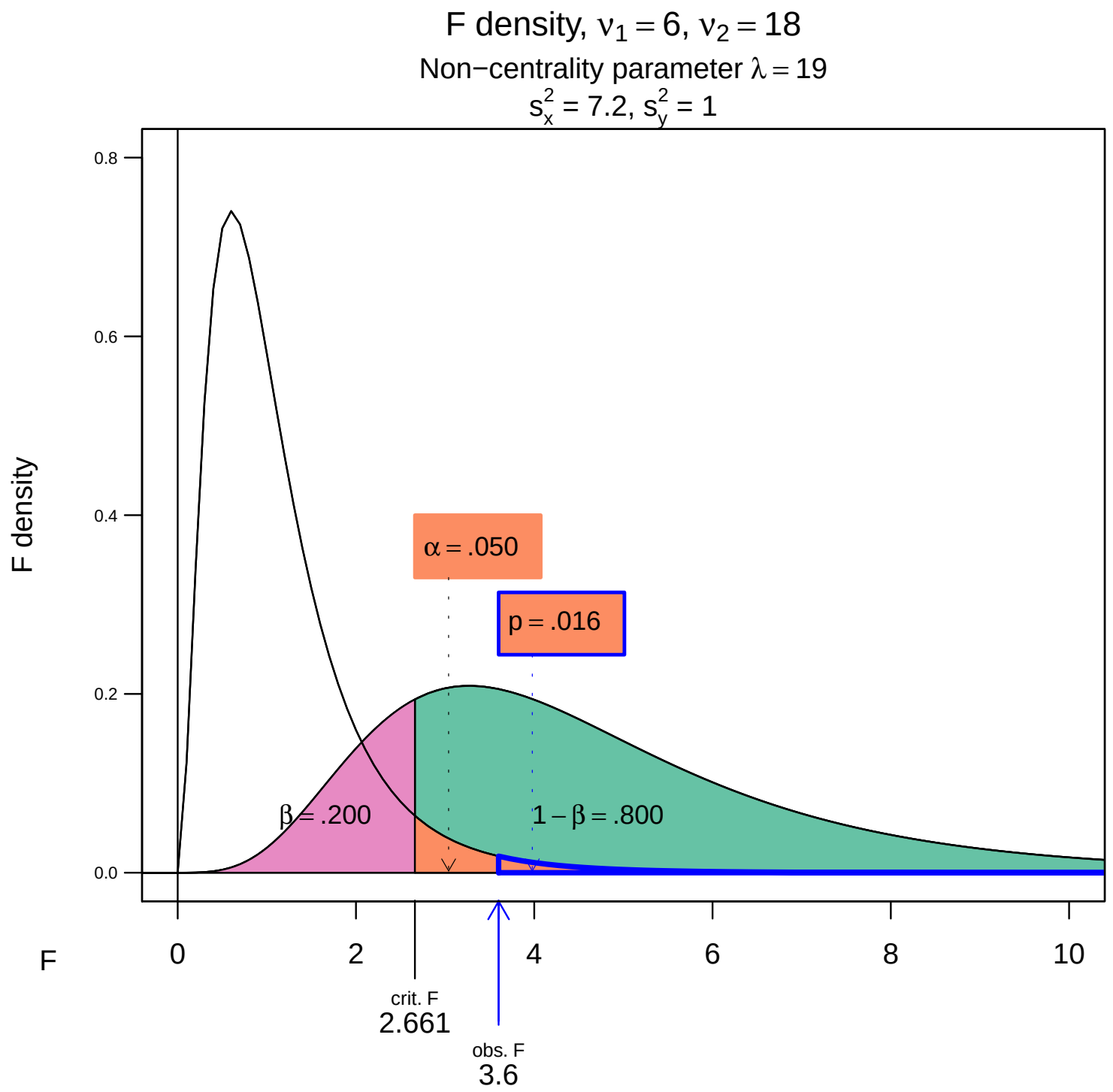

f7.xlsm
Distribution
f
Input
observed F
3.6
v_1 = numerator degrees
6
v_2 = denominator degree
18
α = Prob(Type I error)
0.05
rejection side
right
ncp = noncentrality
14
θ=multiple of central F
nn
s^2_x
s^2_y
R_0
Display Observed
Alternative Display
Non-Central F
Central F
None
Highlight
Type II Error
Power
Graph on Top
Output
critical value—left
critical value—right
2.661
p-value
0.0160
β = Prob(Type II error)
0.3517
power=1-β
0.6483
Done: distrPlot
Slider
min
max
0
30
0
50
0
50
0
0.4
0
30
0.5
32
Graph Limits
value
xlim
0
10
density
ylim
0
0.8
Display
legend
data scale
labelbox
Display Null
Power function
Non-Central F
ncp
Power
0
0.0500
5
0.2465
10
0.4831
15
0.6837
20
0.8230
25
0.9079
30
0.9547
observed
14
0.6483
Central F
θ=σ²x/σ²y
Power
0.5
1.0
2.0
4.0
8.0
16.0
32.0
observed
F
RCode
Palette

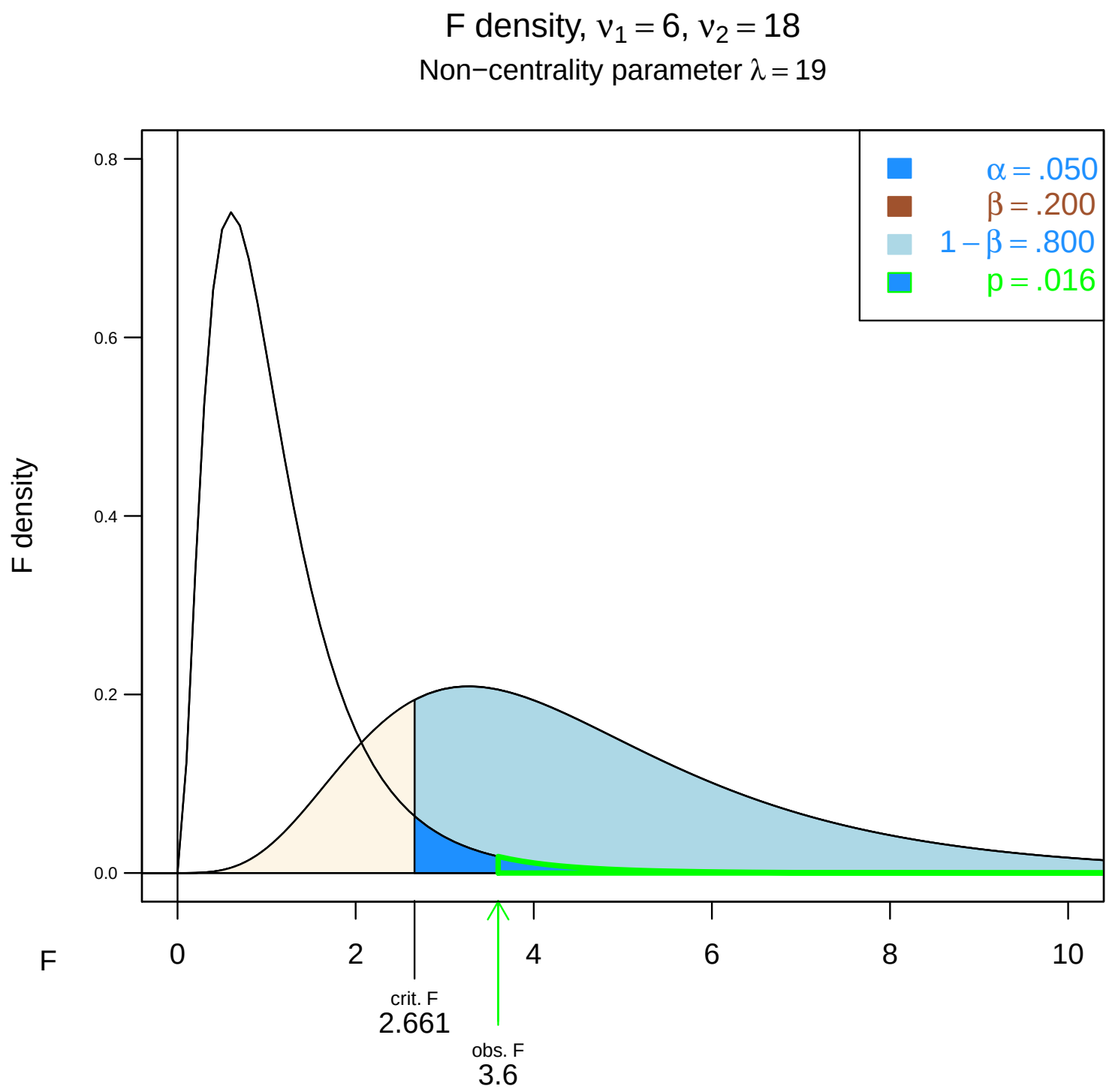

F density, $\nu_1 = 6$, $\nu_2 = 18$
Non−centrality parameter $\lambda = 14$
F density
0.0
0.2
0.4
0.6
0.8
$\alpha = .050$
$p = .016$
$\beta = .352$
$1 - \beta = .648$
F
0
2
4
6
8
10
crit. F
2.661
obs. F
3.6


F density, $\nu_1 = 6$, $\nu_2 = 18$
Non−centrality parameter $\lambda = 14$
$\alpha = .050$
$\beta = .352$
$1 - \beta = .648$
p = .016
F density
0.0
0.2
0.4
0.6
0.8
F
0
2
4
6
8
10
crit. F
2.661
obs. F
3.6


F density, $\nu_1 = 6$, $\nu_2 = 18$
Non−centrality parameter $\lambda = 19$
$\alpha = .050$
$\beta = .200$
$1 - \beta = .800$
p = .016
F density
0.0
0.2
0.4
0.6
0.8
F
0
2
4
6
8
10
crit. F
2.661
obs. F
3.6


F density, $\nu_1 = 6$, $\nu_2 = 18$
Non−centrality parameter $\lambda = 14$
F density
0.0
0.2
0.4
0.6
0.8
$\alpha = .050$
$p = .016$
$\beta = .352$
$1 - \beta = .648$
F
0
2
4
6
8
10
crit. F
2.661
obs. F
3.6


F density, $\nu_1 = 6$, $\nu_2 = 18$
Non−centrality parameter $\lambda = 19$
$s_x^2 = 7.2$, $s_y^2 = 1$
F density
0.0
0.2
0.4
0.6
0.8
$\alpha = .050$
$p = .016$
$\beta = .200$
$1 - \beta = .800$
F
0
2
4
6
8
10
crit. F
2.661
obs. F
3.6


F density, $\nu_1 = 6$, $\nu_2 = 18$
Non−centrality parameter $\lambda = 19$
$\alpha = .050$
$\beta = .200$
$1 - \beta = .800$
$p = .016$
F density
0.0
0.2
0.4
0.6
0.8
F
0
2
4
6
8
10
crit. F
2.661
obs. F
3.6

Graphical Design

1. Color choice
2. Outline of p-value area
3. Alternate axes

   In the normal and t plot, we show the $\bar{x}$-scale, the z-scale under the null, and the z_1-scale under the alternative. We can show the data scale for the F and χ^2.

Design Questions

I invite discussion afterwards on these topics.

1. Color scheme
2. Legend vs Labels
3. Static vs Dynamic
4. Paper vs Screen

Conclusions

Dynamic graphs of hypothesis tests are an excellent way to understand the material and to teach the material.

They can be used in production as part of experimental design. Inspecting these graphs can help in the determination of sample size.

References

1. Baier, T. and Neuwirth, E. (2007) "Excel :: Com :: R". *Computational Statistics*, 22 (1): 91–108. `http://www.springerlink.com/content/uv6667814108258m/fulltext.pdf` You can download this paper for no charge if your library subscribes.

2. Heiberger, Richard M., and Erich Neuwirth (2009). *R through Excel: A Spreadsheet Interface for Statistics, Data Analysis, and Graphics*, Springer–Verlag, New York. Series: Use R!
   `http://www.springer.com/978-1-4419-0051-7`

3. R Development Core Team (2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, `http://www.R-project.org`