In [1]:
```
import numpy as np
import numpy.random as npr
import matplotlib.pyplot as plt
%matplotlib inline
import pandas as pd
import scipy.stats as stats
```

In [2]:
```
df=pd.read_csv("firearms-urban.csv")
```

In [3]:
```
urban=df[df['Percent Urban']>80]['RATE-2014']
rural=df[df['Percent Urban']<=80]['RATE-2014']
pooled=df['RATE-2014']
```

In [4]:
```
diff=rural.mean()-urban.mean() # the difference in the means of rural an
d urban
```

In [5]:
```
pooled_var=np.var(pooled,ddof=1) # the variance of the pooled data
pooled_var
```

Out[5]:
```
17.220408163265308
```

In [6]:
```
sm_var=pooled_var*(1/len(urban)+1/len(rural))
sm_var
```

Out[6]:
```
1.494827097505669
```

In [7]:
```
dof= (len(urban)-1) + (len(rural)-1) # degrees of freedom
dof
```

Out[7]:
```
48
```

In [8]:
```
myt=stats.t(48,scale=np.sqrt(sm_var)) # create a gaussian RV with the de
grees of freedom and sigma ?
```

In [9]:
```
t = np.linspace(-4,4,100)
# plt.plot(t,myt.pdf(t))
```

Lecture 29 Assignment

Use the Student's T random variable to determine a 95% confidence interval for the mean difference under the null hypothesis. Is the resulting confidence interval compatible with the observed difference of means?

Hint: The inverse CDF function in `scipy.stats` is called the Percent point function (PPF) and is given by the `ppf` method of random variable objects.

In [10]:
```
lower = myt.ppf(.025)
upper = myt.ppf(.975)
lower,upper
```

Out[10]:
```
(-2.458264820960342, 2.4582648209603413)
```

The observed diff is 4.31, which clearly falls outside of the range of this interval by a significant portion. This leads me to believe that the observed diff is indeed statistically significant.