*You are currently looking at **version 1.0** of this notebook. To download notebooks and datafiles, as well as get help on Jupyter notebooks in the Coursera platform, visit the Jupyter Notebook FAQ (https://www.coursera.org/learn/python-data-analysis/resources/0dhYG) course resource.*

Distributions in Pandas

In [ ]:
```
import pandas as pd
import numpy as np
```

In [ ]:
```
np.random.binomial(1, 0.5)
```

In [ ]:
```
np.random.binomial(1000, 0.5)/1000
```

In [ ]:
```
chance_of_tornado = 0.01/100
np.random.binomial(100000, chance_of_tornado)
```

In [ ]:
```
chance_of_tornado = 0.01

tornado_events = np.random.binomial(1, chance_of_tornado, 1000000)

two_days_in_a_row = 0
for j in range(1,len(tornado_events)-1):
    if tornado_events[j]==1 and tornado_events[j-1]==1:
        two_days_in_a_row+=1

print('{} tornadoes back to back in {} years'.format(two_days_in_a_row, 1000000/3
```

In [ ]:
```
np.random.uniform(0, 1)
```

In [ ]:
```
np.random.normal(0.75)
```

Formula for standard deviation

$$\sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

In [ ]:
```
distribution = np.random.normal(0.75,size=1000)

np.sqrt(np.sum((np.mean(distribution)-distribution)**2)/len(distribution))
```

In [ ]:
```
np.std(distribution)
```

In [ ]:
```
import scipy.stats as stats
stats.kurtosis(distribution)
```

In [ ]:
```
stats.skew(distribution)
```

In [ ]:
```
chi_squared_df2 = np.random.chisquare(2, size=10000)
stats.skew(chi_squared_df2)
```

In [ ]:
```
chi_squared_df5 = np.random.chisquare(5, size=10000)
stats.skew(chi_squared_df5)
```

In [ ]:
```
%matplotlib inline
import matplotlib
import matplotlib.pyplot as plt

output = plt.hist([chi_squared_df2,chi_squared_df5], bins=50, histtype='step',
                  label=['2 degrees of freedom','5 degrees of freedom'])
plt.legend(loc='upper right')
```

Hypothesis Testing

In [ ]:
```
df = pd.read_csv('grades.csv')
```

In [ ]:
```
df.head()
```

In [ ]:
```
len(df)
```

In [ ]:
```
early = df[df['assignment1_submission'] <= '2015-12-31']
late = df[df['assignment1_submission'] > '2015-12-31']
```

In [ ]:
```
early.mean()
```

In [ ]:
```
late.mean()
```

In [ ]:
```
from scipy import stats
stats.ttest_ind?
```

In [ ]:
```
stats.ttest_ind(early['assignment1_grade'], late['assignment1_grade'])
```

In [ ]:
```
stats.ttest_ind(early['assignment2_grade'], late['assignment2_grade'])
```

In [ ]:
```
stats.ttest_ind(early['assignment3_grade'], late['assignment3_grade'])
```