Forest Fires

June 5, 2016

```
Regression Modeling in Practice Course
by Wesleyan University

Linear Regression Model
Mario Colosso V.
```

```
In [1]: %matplotlib inline

        import pandas
        import matplotlib.pyplot as plt
        import statsmodels.formula.api as smf

        pandas.set_option('display.mpl_style', 'default') # Make the graphs a bit prettier
        pandas.set_option('display.float_format', lambda x:'%.3f'%x)
        plt.rcParams['figure.figsize'] = (15, 5)
```

```
C:\Anaconda3\lib\site-packages\IPython\core\interactiveshell.py:2885: FutureWarning:
mpl_style had been deprecated and will be removed in a future version.
Use `matplotlib.pyplot.style.use` instead.

  exec(code_obj, self.user_global_ns, self.user_ns)
```

```
In [2]: #plt.style.use('ggplot')
        #plt.rcParams['figure.figsize'] = (15, 5)

        #print(plt.style.available)
```

0.0.1 Load Forest Fires .csv file

```
In [3]: fires = pandas.read_csv('forestfires.csv')
```

0.1 1. Lets have a brief look of Fires DataFrame

```
In [4]: fires.head()    #Show first rows
```

Out[4]:

	X	Y	month	day	FFMC	DMC	DC	ISI	temp	RH	wind	rain	area
0	7	5	mar	fri	86.200	26.200	94.300	5.100	8.200	51	6.700	0.000	0.000
1	7	4	oct	tue	90.600	35.400	669.100	6.700	18.000	33	0.900	0.000	0.000
2	7	4	oct	sat	90.600	43.700	686.900	6.700	14.600	33	1.300	0.000	0.000
3	8	6	mar	fri	91.700	33.300	77.500	9.000	8.300	97	4.000	0.200	0.000
4	8	6	mar	sun	89.300	51.300	102.200	9.600	11.400	99	1.800	0.000	0.000

0.1.1 Get some descriptive statistic of the data

In [5]:
```
fires_attributes = fires.columns.values.tolist()
number_of_columns = len(fires_attributes)
```

In [6]:
```
statistics = pandas.DataFrame(index=range(0, number_of_columns - 4),
                              columns=('name', 'min', 'max', 'mean'))
```

In [7]:
```
for attr in range(4, number_of_columns):
    idx = attr - 4
    statistics.loc[idx] = {'name': fires_attributes[attr],
                           'min':  min(fires[fires_attributes[attr]]),
                           'max':  max(fires[fires_attributes[attr]]),
                           'mean': fires[fires_attributes[attr]].mean()}
```

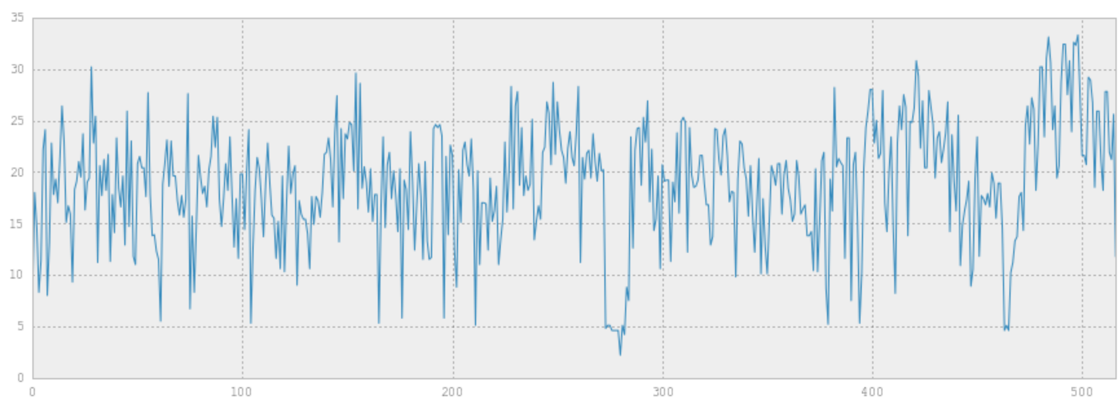
In [8]:
```
statistics   #Show min, max and mean
```

Out[8]:

	name	min	max	mean
0	FFMC	18.700	96.200	90.645
1	DMC	1.100	291.300	110.872
2	DC	7.900	860.600	547.940
3	ISI	0.000	56.100	9.022
4	temp	2.200	33.300	18.889
5	RH	15	100	44.288
6	wind	0.400	9.400	4.018
7	rain	0.000	6.400	0.022
8	area	0.000	1090.840	12.847

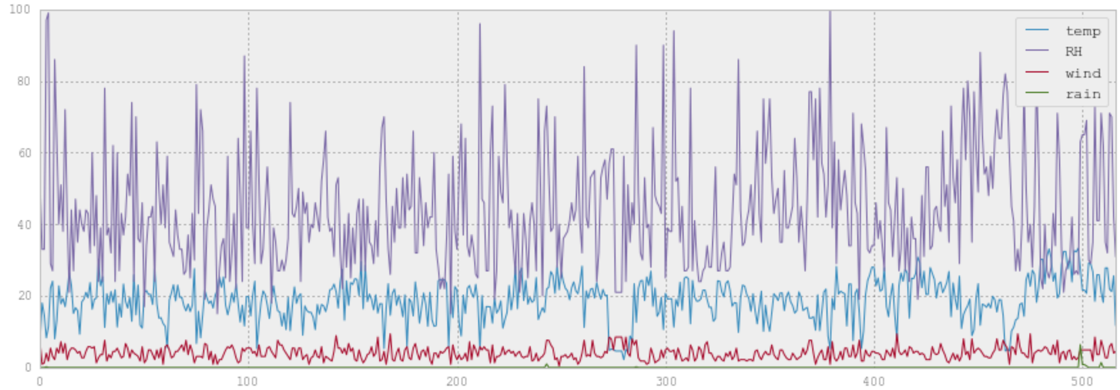
In [9]:
```
fires['temp'].plot()   #Plot temperature graph
```

Out[9]: <matplotlib.axes._subplots.AxesSubplot at 0x1f797faaef0>



In [10]:
```
fires[['temp', 'RH', 'wind', 'rain']].plot()   #Plot temperature, relative humidity, wind
                                               #and rain graphs
```

Out[10]: <matplotlib.axes._subplots.AxesSubplot at 0x1f79806aef0>




```
In [11]: print(fires.corr())    #Show correlation between variables
```

```
X       Y   FFMC    DMC     DC    ISI   temp     RH   wind   rain  \
X      1.000  0.540 -0.021 -0.048 -0.086  0.006 -0.051  0.085  0.019  0.065
Y      0.540  1.000 -0.046  0.008 -0.101 -0.024 -0.024  0.062 -0.020  0.033
FFMC -0.021 -0.046  1.000  0.383  0.331  0.532  0.432 -0.301 -0.028  0.057
DMC  -0.048  0.008  0.383  1.000  0.682  0.305  0.470  0.074 -0.105  0.075
DC   -0.086 -0.101  0.331  0.682  1.000  0.229  0.496 -0.039 -0.203  0.036
ISI   0.006 -0.024  0.532  0.305  0.229  1.000  0.394 -0.133  0.107  0.068
temp -0.051 -0.024  0.432  0.470  0.496  0.394  1.000 -0.527 -0.227  0.069
RH    0.085  0.062 -0.301  0.074 -0.039 -0.133 -0.527  1.000  0.069  0.100
wind  0.019 -0.020 -0.028 -0.105 -0.203  0.107 -0.227  0.069  1.000  0.061
rain  0.065  0.033  0.057  0.075  0.036  0.068  0.069  0.100  0.061  1.000
area  0.063  0.045  0.040  0.073  0.049  0.008  0.098 -0.076  0.012 -0.007

       area
X      0.063
Y      0.045
FFMC   0.040
DMC    0.073
DC     0.049
ISI    0.008
temp   0.098
RH    -0.076
wind   0.012
rain  -0.007
area   1.000
```

0.2 2. Linear regression

0.2.1 Convert categorical variables (months and days) into numerical values

```
In [12]: months_table = ['jan', 'feb', 'mar', 'apr', 'may', 'jun',
                          'jul', 'aug', 'sep', 'oct', 'nov', 'dec']
         days_table =   ['sun', 'mon', 'tue', 'wed', 'thu', 'fri', 'sat']

         fires['month'] = [months_table.index(month) for month in fires['month'] ]
         fires['day'] =   [days_table.index(day)     for day   in fires['day']   ]
```

```
fires.head()
```

Out[12]:

	X	Y	month	day	FFMC	DMC	DC	ISI	temp	RH	wind	rain	area
0	7	5	2	5	86.200	26.200	94.300	5.100	8.200	51	6.700	0.000	0.000
1	7	4	9	2	90.600	35.400	669.100	6.700	18.000	33	0.900	0.000	0.000
2	7	4	9	6	90.600	43.700	686.900	6.700	14.600	33	1.300	0.000	0.000
3	8	6	2	5	91.700	33.300	77.500	9.000	8.300	97	4.000	0.200	0.000
4	8	6	2	0	89.300	51.300	102.200	9.600	11.400	99	1.800	0.000	0.000

0.2.2 Center each explanatory variable

```
In [13]: for idx in range(0, number_of_columns - 1):
             fires[fires_attributes[idx]] = fires[fires_attributes[idx]] -
                                             fires[fires_attributes[idx]].mean()
```

```
In [14]: for idx in range(0, number_of_columns):
             statistics.loc[idx] = {'name': fires_attributes[idx],
                                    'min':  min(fires[fires_attributes[idx]]),
                                    'max':  max(fires[fires_attributes[idx]]),
                                    'mean': fires[fires_attributes[idx]].mean()}
```

```
In [15]: statistics   #Only explanatory variables were centered
```

Out[15]:

	name	min	max	mean
0	X	-3.669	4.331	0.000
1	Y	-2.300	4.700	0.000
2	month	-6.476	4.524	0.000
3	day	-2.973	3.027	-0.000
4	FFMC	-71.945	5.555	0.000
5	DMC	-109.772	180.428	-0.000
6	DC	-540.040	312.660	0.000
7	ISI	-9.022	47.078	-0.000
8	temp	-16.689	14.411	0.000
9	RH	-29.288	55.712	0.000
10	wind	-3.618	5.382	-0.000
11	rain	-0.022	6.378	0.000
12	area	0.000	1090.840	12.847

0.2.3 Generate models to test each variable

```
In [16]: for idx in range(4, number_of_columns - 1):
             model = smf.ols(formula = "area ~ " +
                             fires_attributes[idx], data = fires).fit()
             print(model.summary())
             print()
```

```
OLS Regression Results
==============================================================================
Dep. Variable:                   area   R-squared:                       0.002
Model:                            OLS   Adj. R-squared:                 -0.000
Method:                 Least Squares   F-statistic:                    0.8304
Date:                Fri, 03 Jun 2016   Prob (F-statistic):              0.363
Time:                        21:45:10   Log-Likelihood:                -2880.0
No. Observations:                 517   AIC:                             5764.
Df Residuals:                     515   BIC:                             5773.
```

```
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [95.0% Conf. Int.]
------------------------------------------------------------------------------
Intercept     12.8473      2.800      4.588      0.000         7.346    18.348
FFMC           0.4627      0.508      0.911      0.363        -0.535     1.460
==============================================================================
Omnibus:                      983.137   Durbin-Watson:                   1.649
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           808340.065
Skew:                          12.793   Prob(JB):                         0.00
Kurtosis:                     195.015   Cond. No.                         5.51
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

                            OLS Regression Results
==============================================================================
Dep. Variable:                   area   R-squared:                       0.005
Model:                            OLS   Adj. R-squared:                  0.003
Method:                 Least Squares   F-statistic:                     2.759
Date:                Fri, 03 Jun 2016   Prob (F-statistic):             0.0973
Time:                        21:45:10   Log-Likelihood:                -2879.1
No. Observations:                 517   AIC:                             5762.
Df Residuals:                     515   BIC:                             5771.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [95.0% Conf. Int.]
------------------------------------------------------------------------------
Intercept     12.8473      2.795      4.597      0.000         7.357    18.338
DMC            0.0725      0.044      1.661      0.097        -0.013     0.158
==============================================================================
Omnibus:                      982.803   Durbin-Watson:                   1.649
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           811231.935
Skew:                          12.780   Prob(JB):                         0.00
Kurtosis:                     195.368   Cond. No.                         64.0
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

                            OLS Regression Results
==============================================================================
Dep. Variable:                   area   R-squared:                       0.002
Model:                            OLS   Adj. R-squared:                  0.001
Method:                 Least Squares   F-statistic:                     1.259
Date:                Fri, 03 Jun 2016   Prob (F-statistic):              0.262
Time:                        21:45:10   Log-Likelihood:                -2879.8
No. Observations:                 517   AIC:                             5764.
Df Residuals:                     515   BIC:                             5772.
Df Model:                           1
Covariance Type:            nonrobust
```

```
==============================================================================
                 coef    std err          t      P>|t|      [95.0% Conf. Int.]
------------------------------------------------------------------------------
Intercept     12.8473      2.799      4.590      0.000         7.349    18.346
DC             0.0127      0.011      1.122      0.262        -0.010     0.035
==============================================================================
Omnibus:                      982.892   Durbin-Watson:                   1.645
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           807312.305
Skew:                          12.786   Prob(JB):                         0.00
Kurtosis:                     194.893   Cond. No.                         248.
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

                            OLS Regression Results
==============================================================================
Dep. Variable:                   area   R-squared:                       0.000
Model:                            OLS   Adj. R-squared:                 -0.002
Method:                 Least Squares   F-statistic:                   0.03512
Date:                Fri, 03 Jun 2016   Prob (F-statistic):              0.851
Time:                        21:45:11   Log-Likelihood:                -2880.4
No. Observations:                 517   AIC:                             5765.
Df Residuals:                     515   BIC:                             5773.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [95.0% Conf. Int.]
------------------------------------------------------------------------------
Intercept     12.8473      2.802      4.585      0.000         7.342    18.352
ISI            0.1153      0.615      0.187      0.851        -1.093     1.324
==============================================================================
Omnibus:                      983.625   Durbin-Watson:                   1.649
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           809992.277
Skew:                          12.806   Prob(JB):                         0.00
Kurtosis:                     195.211   Cond. No.                         4.56
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

                            OLS Regression Results
==============================================================================
Dep. Variable:                   area   R-squared:                       0.010
Model:                            OLS   Adj. R-squared:                  0.008
Method:                 Least Squares   F-statistic:                     4.978
Date:                Fri, 03 Jun 2016   Prob (F-statistic):             0.0261
Time:                        21:45:11   Log-Likelihood:                -2878.0
No. Observations:                 517   AIC:                             5760.
Df Residuals:                     515   BIC:                             5768.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [95.0% Conf. Int.]
```

```
------------------------------------------------------------------------------
Intercept     12.8473      2.789      4.607      0.000         7.368    18.326
temp           1.0726      0.481      2.231      0.026         0.128     2.017
==============================================================================
Omnibus:                      979.270   Durbin-Watson:                   1.650
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           793772.021
Skew:                          12.687   Prob(JB):                         0.00
Kurtosis:                     193.275   Cond. No.                         5.80
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

                            OLS Regression Results
==============================================================================
Dep. Variable:                   area   R-squared:                       0.006
Model:                            OLS   Adj. R-squared:                  0.004
Method:                 Least Squares   F-statistic:                     2.954
Date:                Fri, 03 Jun 2016   Prob (F-statistic):             0.0863
Time:                        21:45:11   Log-Likelihood:                -2879.0
No. Observations:                 517   AIC:                             5762.
Df Residuals:                     515   BIC:                             5770.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [95.0% Conf. Int.]
------------------------------------------------------------------------------
Intercept     12.8473      2.794      4.598      0.000         7.358    18.337
RH            -0.2946      0.171     -1.719      0.086        -0.631     0.042
==============================================================================
Omnibus:                      980.422   Durbin-Watson:                   1.642
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           795947.965
Skew:                          12.720   Prob(JB):                         0.00
Kurtosis:                     193.531   Cond. No.                         16.3
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

                            OLS Regression Results
==============================================================================
Dep. Variable:                   area   R-squared:                       0.000
Model:                            OLS   Adj. R-squared:                 -0.002
Method:                 Least Squares   F-statistic:                   0.07815
Date:                Fri, 03 Jun 2016   Prob (F-statistic):              0.780
Time:                        21:45:11   Log-Likelihood:                -2880.4
No. Observations:                 517   AIC:                             5765.
Df Residuals:                     515   BIC:                             5773.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [95.0% Conf. Int.]
------------------------------------------------------------------------------
Intercept     12.8473      2.802      4.585      0.000         7.342    18.352
```

```
wind           0.4376      1.565      0.280      0.780      -2.638      3.513
==============================================================================
Omnibus:                      983.721   Durbin-Watson:                   1.647
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           810324.708
Skew:                          12.809   Prob(JB):                         0.00
Kurtosis:                     195.251   Cond. No.                         1.79
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

                            OLS Regression Results
==============================================================================
Dep. Variable:                   area   R-squared:                       0.000
Model:                            OLS   Adj. R-squared:                 -0.002
Method:                 Least Squares   F-statistic:                   0.02794
Date:                Fri, 03 Jun 2016   Prob (F-statistic):              0.867
Time:                        21:45:11   Log-Likelihood:                -2880.4
No. Observations:                 517   AIC:                             5765.
Df Residuals:                     515   BIC:                             5773.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [95.0% Conf. Int.]
------------------------------------------------------------------------------
Intercept     12.8473      2.802      4.585      0.000        7.342    18.352
rain          -1.5842      9.477     -0.167      0.867      -20.203    17.035
==============================================================================
Omnibus:                      983.726   Durbin-Watson:                   1.649
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           810320.385
Skew:                          12.809   Prob(JB):                         0.00
Kurtosis:                     195.250   Cond. No.                         3.38
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

```
In [17]: print((smf.ols(formula = "area ~ temp", data = fires).fit()).summary())
```

```
OLS Regression Results
==============================================================================
Dep. Variable:                   area   R-squared:                       0.010
Model:                            OLS   Adj. R-squared:                  0.008
Method:                 Least Squares   F-statistic:                     4.978
Date:                Fri, 03 Jun 2016   Prob (F-statistic):             0.0261
Time:                        21:45:11   Log-Likelihood:                -2878.0
No. Observations:                 517   AIC:                             5760.
Df Residuals:                     515   BIC:                             5768.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [95.0% Conf. Int.]
------------------------------------------------------------------------------
Intercept     12.8473      2.789      4.607      0.000        7.368    18.326
temp           1.0726      0.481      2.231      0.026        0.128     2.017
```

```
==============================================================================
Omnibus:                      979.270   Durbin-Watson:                   1.650
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           793772.021
Skew:                          12.687   Prob(JB):                         0.00
Kurtosis:                     193.275   Cond. No.                         5.80
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

The results of the linear regression models indicated than only temperature (Beta = 1.0726, p = 0.026) was significantly and positively associated with the total burned area due to forest fires. 'p-value' of other models are greater than treshold value of 0.05 so results are not statistically significant to reject null hypothesis.

0.2.4 Create a Linear Regression Model for a combination of variables

In [18]:
```
explanatory_variables = "FFMC + DMC + DC + ISI + temp + RH + wind + rain"
response_variable =     "area"

model = smf.ols(formula = response_variable + " ~ " + explanatory_variables,
                data = fires).fit()
```

In [19]:
```
print(model.summary())
```

```
OLS Regression Results
==============================================================================
Dep. Variable:                   area   R-squared:                       0.016
Model:                            OLS   Adj. R-squared:                  0.001
Method:                 Least Squares   F-statistic:                     1.033
Date:                Fri, 03 Jun 2016   Prob (F-statistic):              0.410
Time:                        21:45:11   Log-Likelihood:                -2876.3
No. Observations:                 517   AIC:                             5771.
Df Residuals:                     508   BIC:                             5809.
Df Model:                           8
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [95.0% Conf. Int.]
------------------------------------------------------------------------------
Intercept     12.8473      2.799      4.590      0.000         7.349    18.346
FFMC          -0.0233      0.661     -0.035      0.972        -1.322     1.275
DMC            0.0765      0.067      1.145      0.253        -0.055     0.208
DC            -0.0057      0.016     -0.349      0.727        -0.038     0.026
ISI           -0.6984      0.772     -0.905      0.366        -2.215     0.818
temp           0.8480      0.787      1.077      0.282        -0.699     2.394
RH            -0.1963      0.237     -0.829      0.407        -0.661     0.269
wind           1.5271      1.670      0.914      0.361        -1.754     4.808
rain          -2.5400      9.676     -0.263      0.793       -21.549    16.469
==============================================================================
Omnibus:                      978.059   Durbin-Watson:                   1.645
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           792201.920
Skew:                          12.652   Prob(JB):                         0.00
Kurtosis:                     193.092   Cond. No.                         871.
==============================================================================
```

```
Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

p-value of combination model (p = 0.410) is bigger than treshold value, so the combination of the Canadian Forest Fire Weather Index (FWI) system plus temperature, humidity, wind and rain are not significantly associated with the total burned area due to forest fires. p-value of temperature in combination model (p = 0.282) is not longer statistically significant, a confounder variable?

```
In [ ]:
```