Week 4 Quiz

Perrin Anto - paj2117

Due Saturday Oct 5 11:59pm

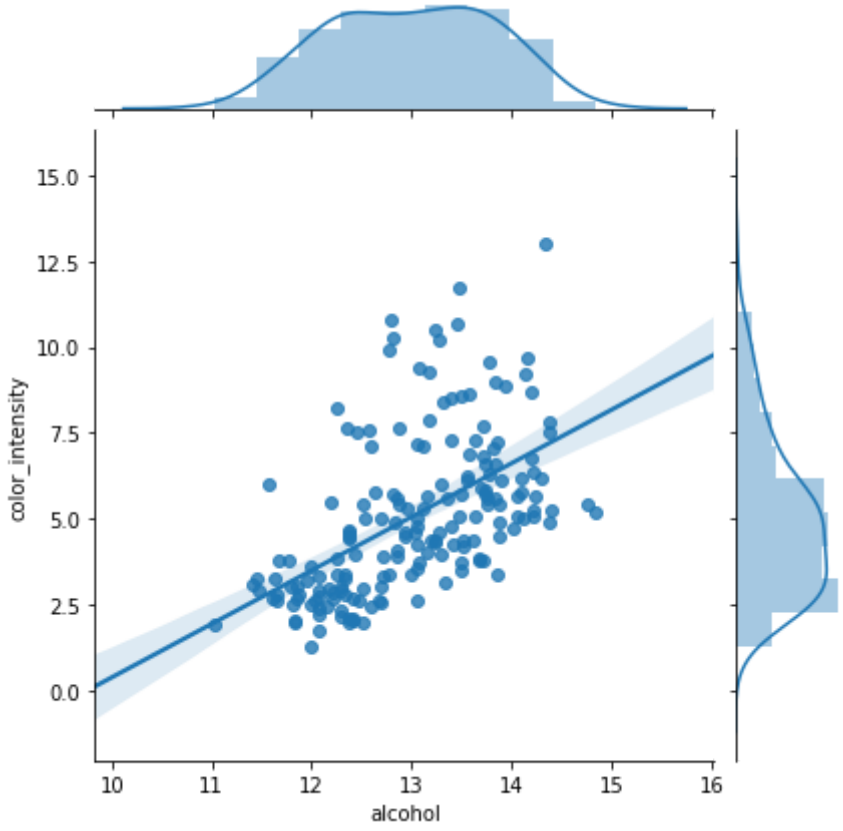
In [1]:
```
import pandas as pd
import seaborn as sns
import statsmodels.api as sm

import warnings
warnings.simplefilter(action='ignore', category=FutureWarning)

%matplotlib inline
```

We're going to see if there is a linear relationship between alcohol level and color intensity in wine using a simple linear regression model.

In [2]:
```
# read in ../data/wine_dataset.csv as df
df = pd.read_csv('../data/wine_dataset.csv')
```

In [3]:
```
# using jointplot with kind='reg', plot alcohol (x) against color_intens
ity (y) including a regression line
sns.jointplot(x="alcohol",y='color_intensity',data=df,kind='reg')
```

Out[3]: <seaborn.axisgrid.JointGrid at 0x1c225f4f10>



In [4]:
```
# Create a new variable X that is just the alcohol column from df
X = df['alcohol']
```

In [5]:
```
# Add a constant of 1 to X using sm.add_constant (Recall this is our int
ercept or bias term)
X = sm.add_constant(X)
```

In [6]:
```
# Print out the first five rows of X. There should be two columns: cons
t, alcohol
X.head(5)
```

Out[6]:

	const	alcohol
0	1.0	14.23
1	1.0	13.20
2	1.0	13.16
3	1.0	14.37
4	1.0	13.24

In [7]:
```
# Create a new variable y that is just the color_intensity column from d
f
y = df['color_intensity']
```

In [8]:
```
# Using sm.OLS, instantiate and fit the simple linear model (note: the o
rder of paramters is y,X)
model_slr = sm.OLS(y,X).fit()
```

In [9]:
```
# Print out the learned paramters the model
model_slr.params
```

Out[9]:
```
const     -15.225741
alcohol     1.560220
dtype: float64
```

In [10]:
```
# according to the model, what will color_intensity be when alcohol = 0?
#-15.225741
#

# according to the model, does color_intensity go up or down as alcohol
 increases?
#Up
#
```

In [11]:
```
# Print out the rsquared value (Recall that this is amount of variance i
n y explained by the model)
model_slr.rsquared
```

Out[11]:
```
0.2985138336694634
```