Diet

December 10, 2017

```
In [1]: import pandas as pd
        import numpy as np
        import matplotlib.pyplot as plt
        import seaborn as sns
        %matplotlib inline
```

```
/usr/local/lib/python2.7/dist-packages/pandas/core/computation/__init__.py:18: UserWarning: The
The minimum supported version is 2.4.6

  ver=ver, min_ver=_MIN_NUMEXPR_VERSION), UserWarning)
```

```
In [23]: diet = pd.read_csv('diet.csv')[1:]
```

```
In [24]: diet.head()
```

Out[24]:

	Sales	Advertising	Lagged Sales	delta A(t)
1	20.5	16	12.0	0.066667
2	21.0	18	20.5	0.125000
3	15.5	27	21.0	0.500000
4	15.3	21	15.5	-0.222222
5	23.5	49	15.3	1.333333

```
In [25]: diet.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 35 entries, 1 to 35
Data columns (total 4 columns):
Sales           35 non-null float64
Advertising     35 non-null int64
Lagged Sales    35 non-null float64
delta A(t)      35 non-null float64
dtypes: float64(3), int64(1)
memory usage: 1.2 KB
```

```
In [26]: diet.describe()
```

Out[26]:

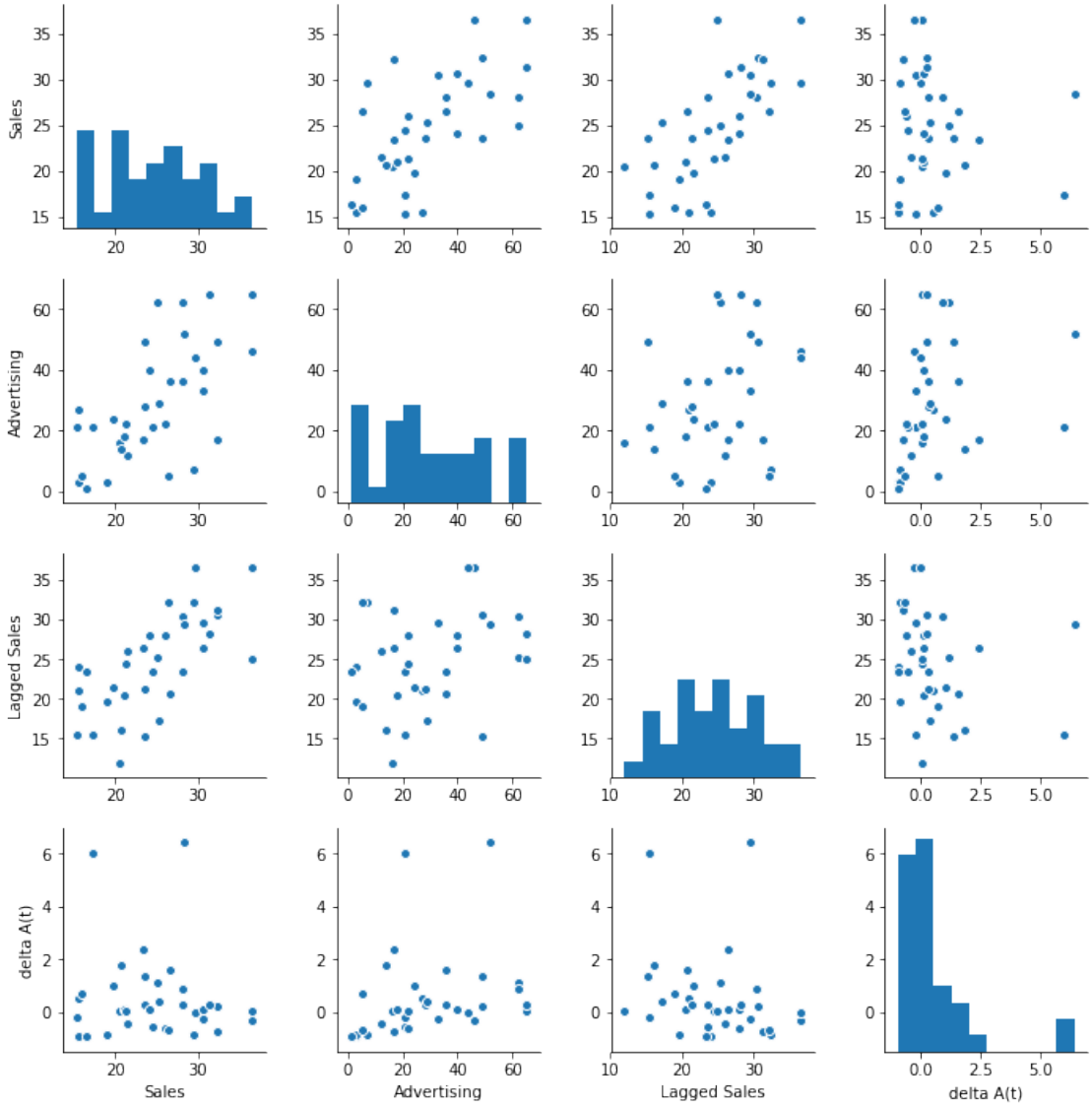
	Sales	Advertising	Lagged Sales	delta A(t)
count	35.000000	35.000000	35.000000	35.000000
mean	24.602857	28.914286	24.477143	0.517691
std	5.898479	18.905915	6.121134	1.634668
min	15.300000	1.000000	12.000000	-0.941176
25%	20.600000	16.500000	20.600000	-0.373427
50%	24.500000	24.000000	24.500000	0.111111
75%	28.900000	42.000000	28.900000	0.772727
max	36.500000	65.000000	36.500000	6.428571

In [27]:
```
diet.columns
```

Out[27]:
```
Index([u'Sales', u'Advertising', u'Lagged Sales', u'delta A(t)'], dtype='object')
```

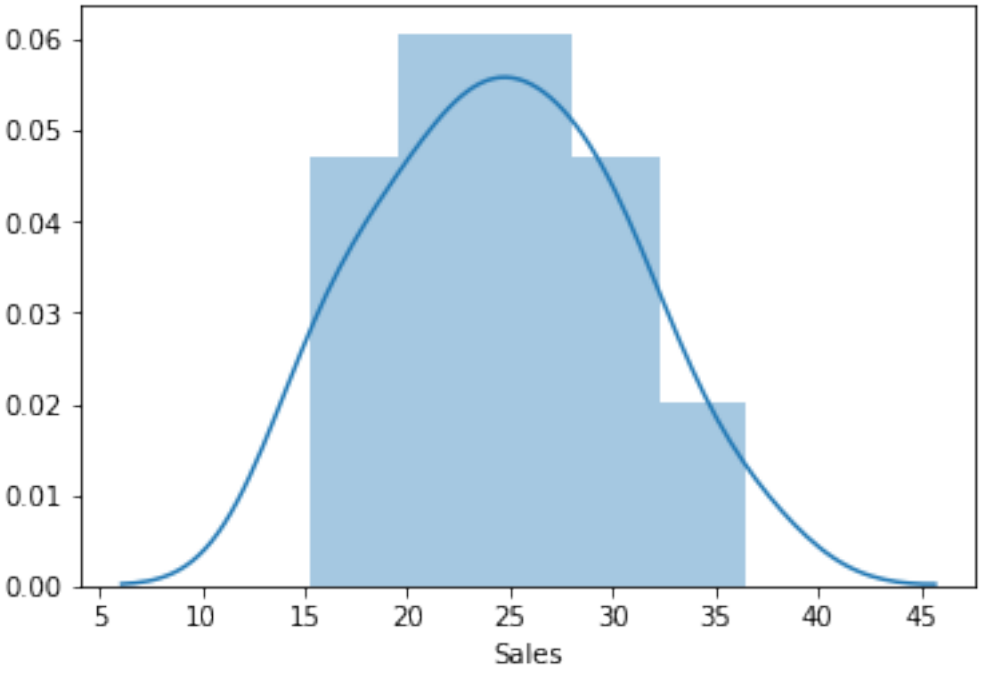
In [28]:
```
sns.pairplot(diet.dropna())
```

Out[28]:
```
<seaborn.axisgrid.PairGrid at 0x7fbd1bdb6e50>
```



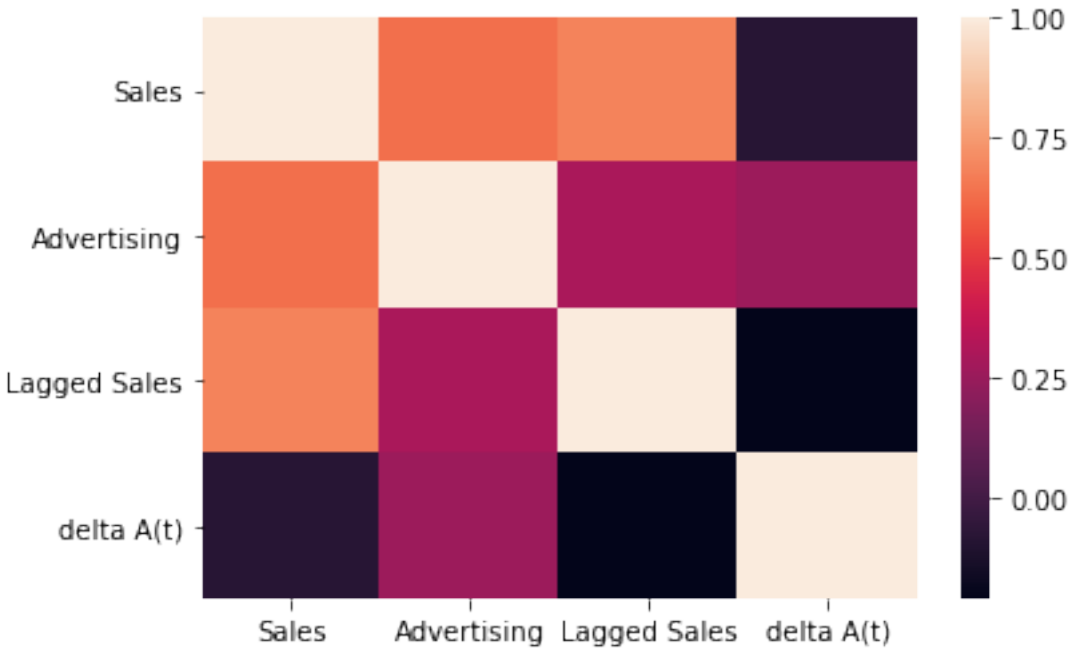
In [29]:
```
sns.distplot(diet['Sales'])
```

Out[29]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x7fbd1ca952d0>
```



In [30]:
```
sns.heatmap(diet.corr())
```

Out[30]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x7fbd1c071050>
```

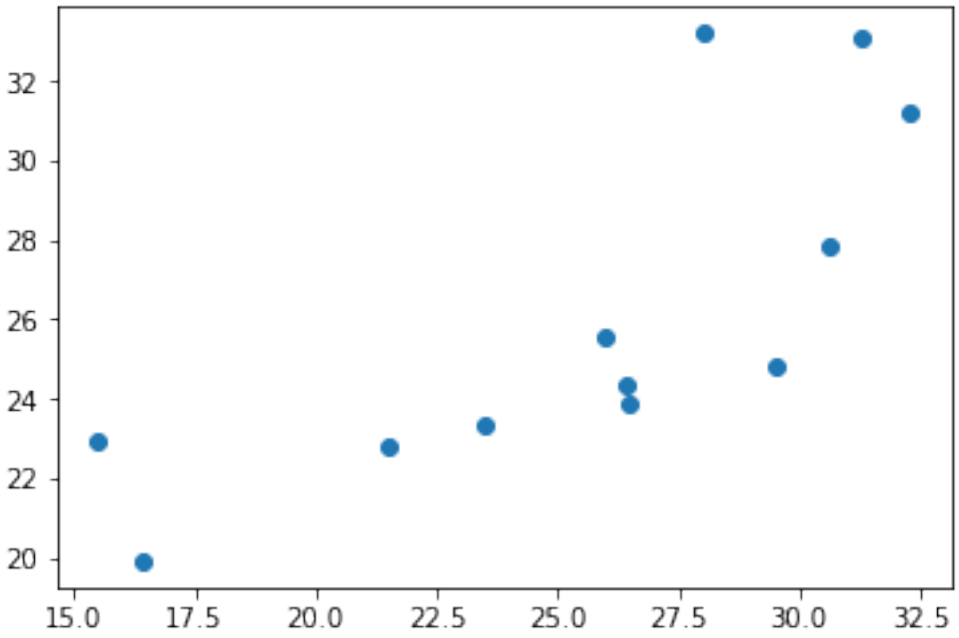



```
In [31]: X = diet.drop('Sales', axis=1)
         y = diet['Sales']

In [32]: from sklearn.model_selection import train_test_split

In [40]: X_train, X_test, y_train, y_test = train_test_split(X, y, test_size= 0.33, random_state

In [41]: from sklearn.linear_model import LinearRegression

In [42]: lm = LinearRegression()

In [43]: lm.fit(X_train,y_train)

Out[43]: LinearRegression(copy_X=True, fit_intercept=True, n_jobs=1, normalize=False)

In [44]: print(lm.intercept_)

9.16670760627


In [45]: coeff_df = pd.DataFrame(lm.coef_,X.columns,columns=['Coefficient'])
         coeff_df

Out[45]:              Coefficient
         Advertising      0.181874
         Lagged Sales     0.432434
         delta A(t)      -0.482796
```

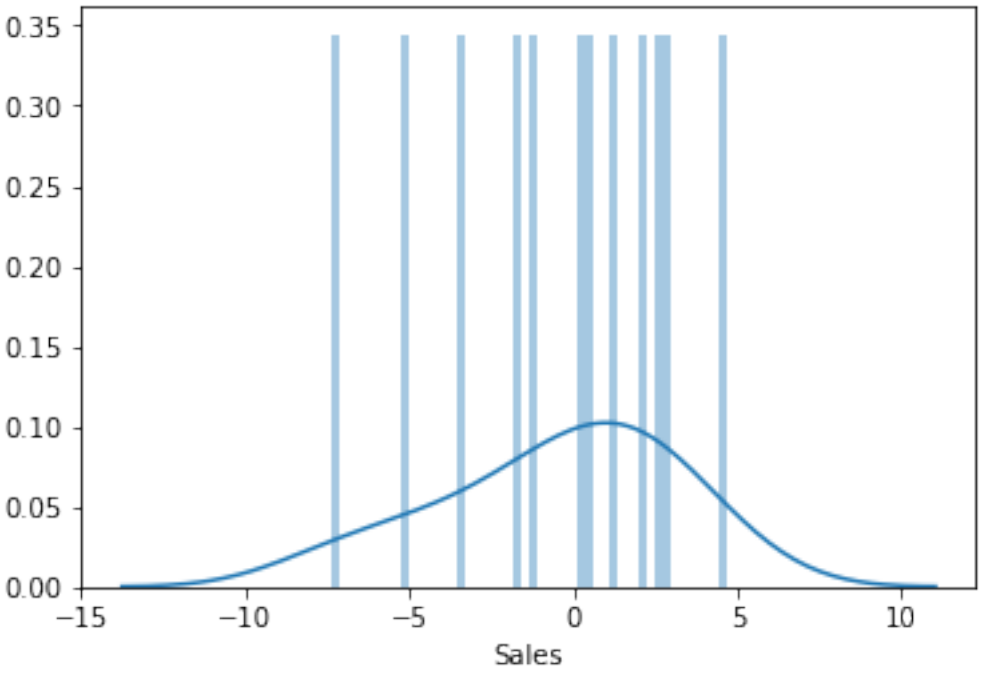
In [46]:
```
predictions = lm.predict(X_test)
```

In [47]:
```
plt.scatter(y_test,predictions)
```

Out[47]:
```
<matplotlib.collections.PathCollection at 0x7fbd1bf32810>
```



In [48]:
```
sns.distplot((y_test-predictions),bins=50);
```



```
In [49]: from sklearn import metrics

In [50]: print('MAE:', metrics.mean_absolute_error(y_test, predictions))
         print('MSE:', metrics.mean_squared_error(y_test, predictions))
         print('RMSE:', np.sqrt(metrics.mean_squared_error(y_test, predictions)))

('MAE:', 2.7516006884560427)
('MSE:', 11.779176453467235)
('RMSE:', 3.43208048470126)
```