Homework 3 - Logistic Regression

Quantile Analysis
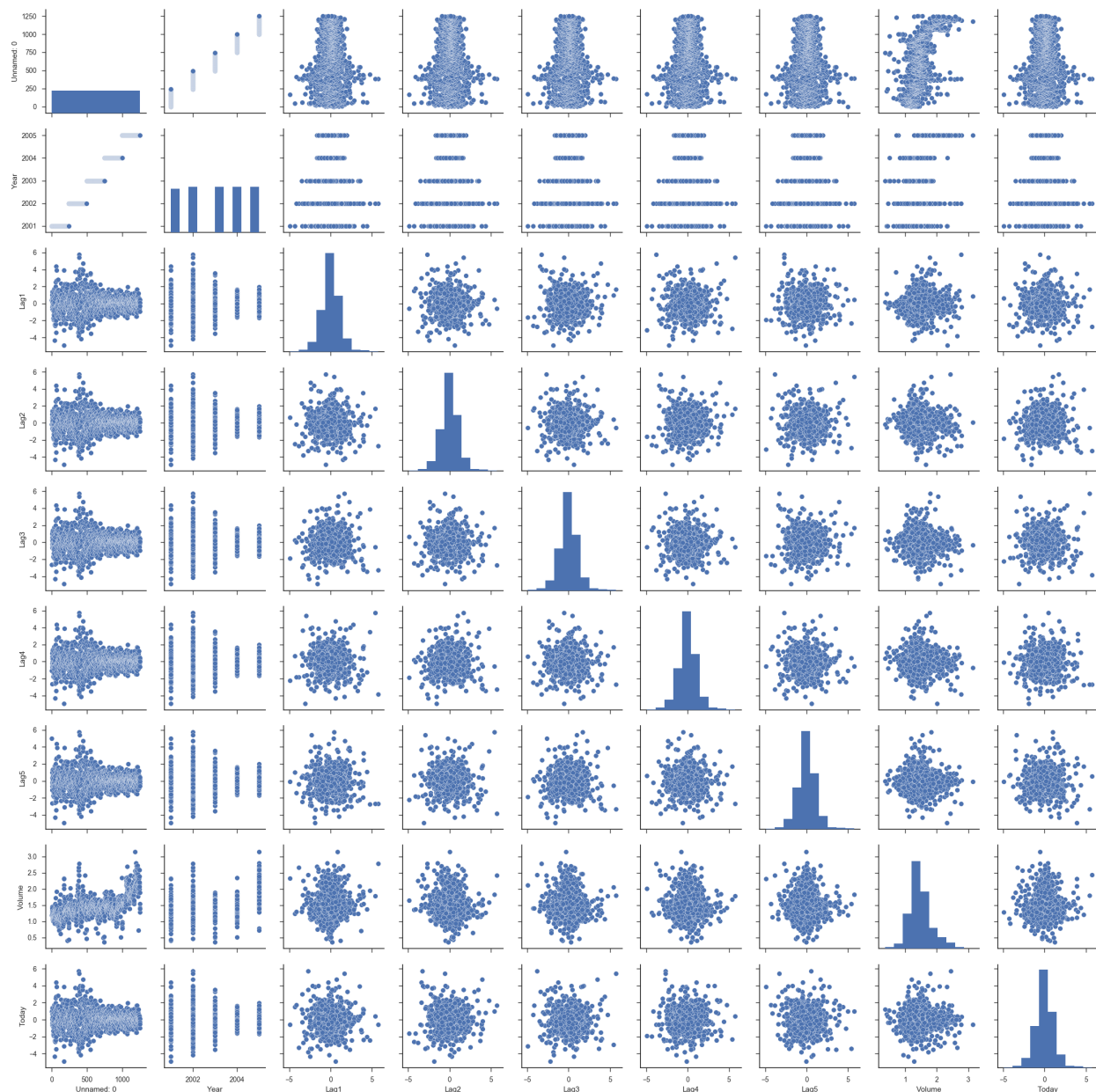
	Year	Lag1	Lag2	Lag3	Lag4	Lag5	Volume	Today	Direction
min	2001	-4.922	-4.922	-4.922	-4.922	-4.922	0.35607	-4.922	Up: 642
1st Qu.	2002	-0.6395	-0.6395	-0.64	-0.64	-0.64	1.2574	-0.6395	Down:608
Median	2003	0.039	0.039	0.0385	0.0385	0.0385	1.42295	0.0385	
Mean	2003	0.003834	0.003919	0.001716	0.001636	0.00561	1.478305	0.003138	
3rd Qu.	2004	0.59675	0.59675	0.59675	0.59675	0.59675	1.641675	0.59675	
Max	2005	5.733	5.733	5.733	5.733	5.733	3.152470	5.733	

1. Quantify the Spread

	Evenly Spread	Skewed	Quantile Useful?
Year	yes	no	Not really
Lag1	yes	df['Lag5'], to upper end	A little
Lag2	yes	Slightly to upper end	A little
Lag3	yes	Slightly to upper	A little

		end	
Lag4	yes	Slightly to upper end	A little
Lag5	yes	Slightly to upper end	A little
Volume	no	Significantly to lower end	Not much
Today	yes	Slightly to upper end	A little
Direction	no	Slightly to upper end	some

Pairwise Correlation

	Year	Lag1	Lag2	Lag3	Lag4	Lag5	Volume	Today
Year		Fairly flat	Fairly flat	Fairly flat	Fairly flat	Fairly flat	Fairly flat	Fairly flat
Lag1	Fairly		Fairly	Fairly	Fairly	Fairly	Fairly	Fairly

	flat		flat, slightly up	flat	flat	flat	flat	flat
Lag2	Fairly flat	Fairly flat		Fairly flat	Fairly flat	Fairly flat	Fairly flat	Fairly flat
Lag3	Fairly flat	Fairly flat	Fairly flat		Fairly flat	Fairly flat	Fairly flat	Fairly flat
Lag4	Fairly flat	Fairly flat	Fairly flat	Fairly flat		Fairly flat	Fairly flat	Fairly flat
Lag5	Fairly flat	Fairly flat	Fairly flat	Fairly flat	Fairly flat		Fairly flat	Fairly flat
Volume	Fairly flat	Fairly flat	Fairly flat	Fairly flat	Fairly flat	Fairly flat		Fairly flat
Today	Fairly flat	Fairly flat	Fairly flat	Fairly flat	Fairly flat	Fairly flat	Fairly flat	

Pairwise Correlation Coefficients

	Year	Lag1	Lag2	Lag3	Lag4	Lag5	Volume	Today
Year	1	0.029700	0.030596	0.033195	0.035689	0.029788	0.539006	0.030095
Lag1	0.029700	1	-0.026294	-0.010803	-0.002986	-0.005675	0.040910	-0.026155
Lag2	0.030596	-0.026294	1	-0.025897	-0.010854	-0.003558	-0.043383	-0.010250
Lag3	0.033195	-0.010803	-0.025897	1	-0.024051	-0.018808	-0.041824	-0.002448
Lag4	0.035689	-0.002986	-0.010854	-0.024051	1	-0.027084	-0.048414	-0.006900
Lag5	0.029788	-0.005675	-0.003558	-0.018808	-0.027084	1	-0.022002	-0.034860
Volume	0.539006	0.040910	-0.043383	-0.041824	-0.048414	-0.022002	1	0.014592
Today	0.030095	-0.026155	-0.010250	-0.002448	-0.006900	-0.034860	0.014592	1

	Year	Lag1	Lag2	Lag3	Lag4	Lag5	Volume	Today
Year		Slightly up	Slightly up	Slightly up	Slightly up	Slightly up	Slightly up	Slightly up
Lag1	Slightly up		Slightly down	Slightly down	Slightly down	Slightly down	Slightly up	Slightly down
Lag2	Slightly up	Slightly down		Slightly down	Slightly down	Slightly down	Slightly down	Slightly down
Lag3	Slightly up	Slightly down	Slightly down		Slightly down	Slightly down	Slightly down	Slightly down
Lag4	Slightly up	Slightly down	Slightly down	Slightly down		Slightly down	Slightly down	Slightly down
Lag5	Slightly up	Slightly down	Slightly down	Slightly down	Slightly down		Slightly down	Slightly down
Volume	Slightly up	Slightly up	Slightly down	Slightly down	Slightly down	Slightly down		Slightly up
Today	Slightly up	Slightly down	Slightly down	Slightly down	Slightly down	Slightly down	Slightly up	

Interesting correlations between Volume and Today and Lags

The only positive correlation between volume and any lags is with Lag1.

There is very little correlation between today and previous days, the less so the more days you go back into the past.

Prediction

It seems likely that lags could reasonably predict volume and today

Compare with outcomes in the sample data set

1. Count when Up was predicted correctly and incorrectly
   a. Correct: 550

   b. Incorrect: 98
2. Count when Down was predicted correctly and incorrectly
   a. Correct: 116
   b. Incorrect: 486
3. Determine what percent of outcomes were predicted correctly
   a. Correctly: 53.4%
   b. Incorrectly: 46.7%
4. Which is more correct Up or Down? Is there a difference?
   a. Up seems to be correct more often. This would lead me to trust predictions of up more often than predictions of down. Down tends to be correct 20% of the time while Up is correct 84% of the time.

Compare with outcomes in the sample data set - 2005 Data

1. Count when Up was predicted correctly and incorrectly
   a. Correct: 111
   b. Incorrect: 30
2. Count when Down was predicted correctly and incorrectly
   a. Correct: 37
   b. Incorrect: 74
3. Determine what percent of outcomes were predicted correctly
   a. Correctly: 69.8%
   b. Incorrectly: 49.1%
4. Which is more correct Up or Down? Is there a difference?
   a. Up seems to be correct more often. This would lead me to trust predictions of up more often than predictions of down. Down tends to be correct 21% of the time while Up is correct 79% of the time.