

## Answers

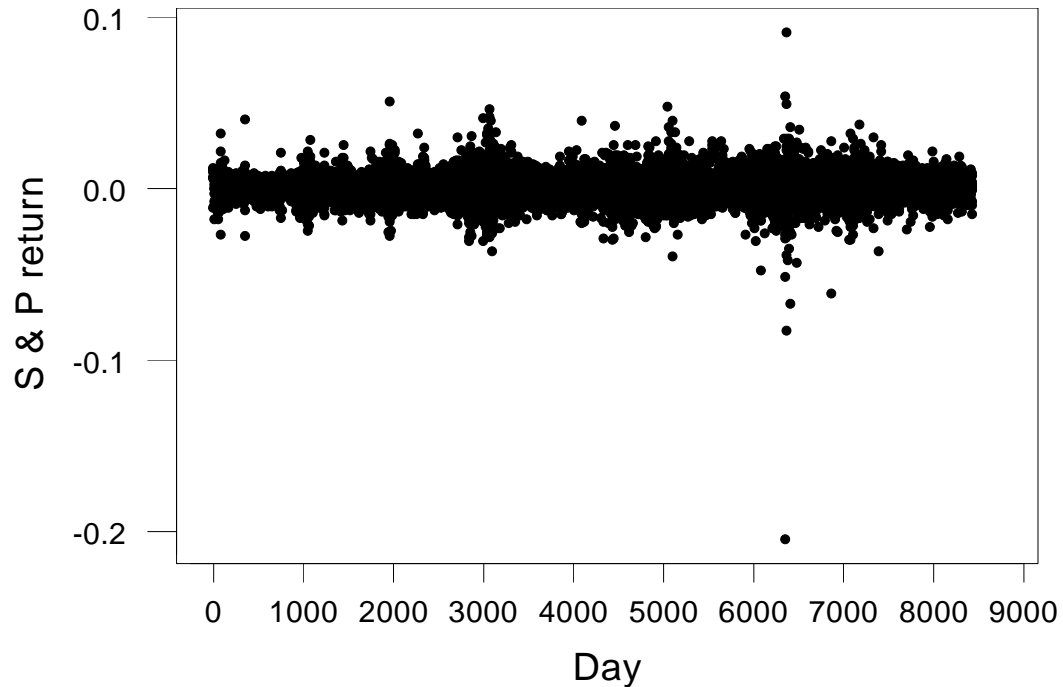
1. The mean daily S&P return is .00032:

### Descriptive Statistics

Variable	N	N*	Mean	Median	Tr Mean	StDev	SE Mean
S & P re	8431	1	0.00032	0.00036	0.00031	0.00859	0.00009

Variable	Min	Max	Q1	Q3
S & P re	-0.20467	0.09099	-0.00394	0.00457

With roughly 250 trading days per year, this works out to an annualized return of  $(1.00032)^{250} = 1.083$ , or an 8.3% annualized return. Use of the arithmetic mean requires that the data be a random sample from a population that is at least reasonably Gaussian. Here is a time series plot of the values. There doesn't appear to be autocorrelation, but there are unusual values in the period around the stock market crash of 1987:



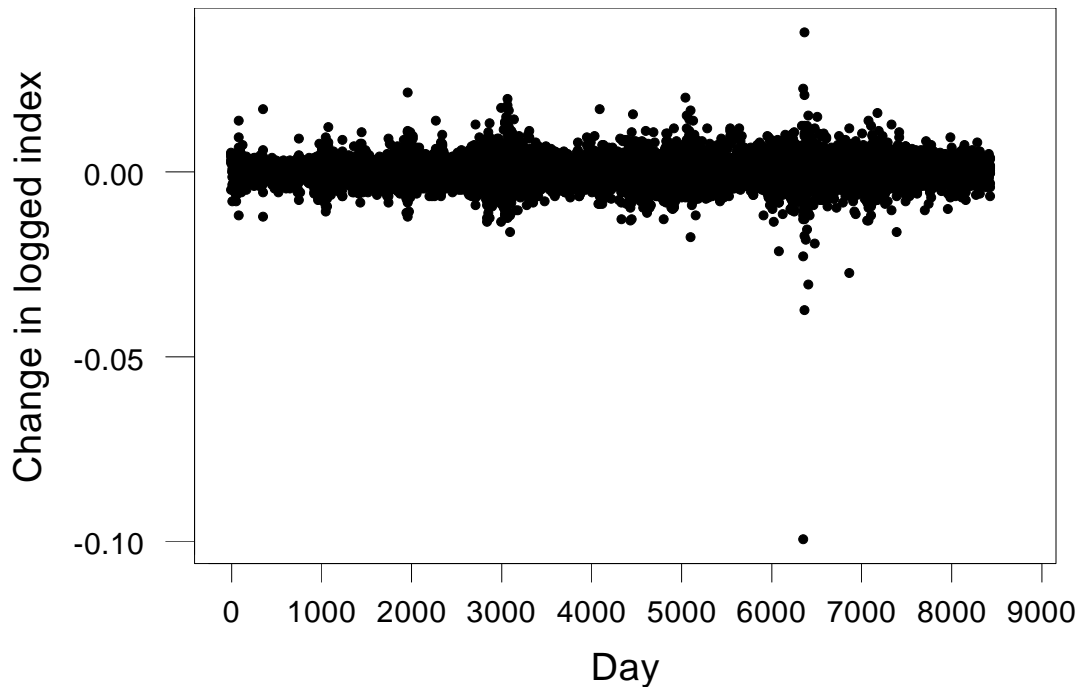
Even more importantly, averaging returns this way doesn't make much sense, since a 10% increase followed by a 10% decrease does not result in zero change in principal (that is,  $(1.1)(.9) \neq 1$ ). To get the geometric mean of the returns, we calculate the change in the logged S&P index values, determine the mean of these values, and then antilog. Here is the required information:

### Descriptive Statistics

Variable	N	N*	Mean	Median	Tr Mean	StDev	SE Mean
Change i	8431	1	0.00012	0.00016	0.00013	0.00376	0.00004

Variable	Min	Max	Q1	Q3
Change i	-0.09945	0.03782	-0.00172	0.00198

The geometric mean of the returns is  $10^{-0.0012} - 1 = .00028$ , or smaller than what would be predicted by the arithmetic mean. It results in an annualized return of  $(1.00028)^{250} = 1.072$ , or a 7.2% annualized return. The geometric mean requires that the changes in logged index be a random sample from a population that is at least reasonably Gaussian. Here is a time series plot of the values, which is similar to that of the returns:



The semilog estimate of the return is obtained by regressing the log of the S&P index on time:

## Regression Analysis

The regression equation is

Logged S & P index = 1.74 + 0.000103 Day

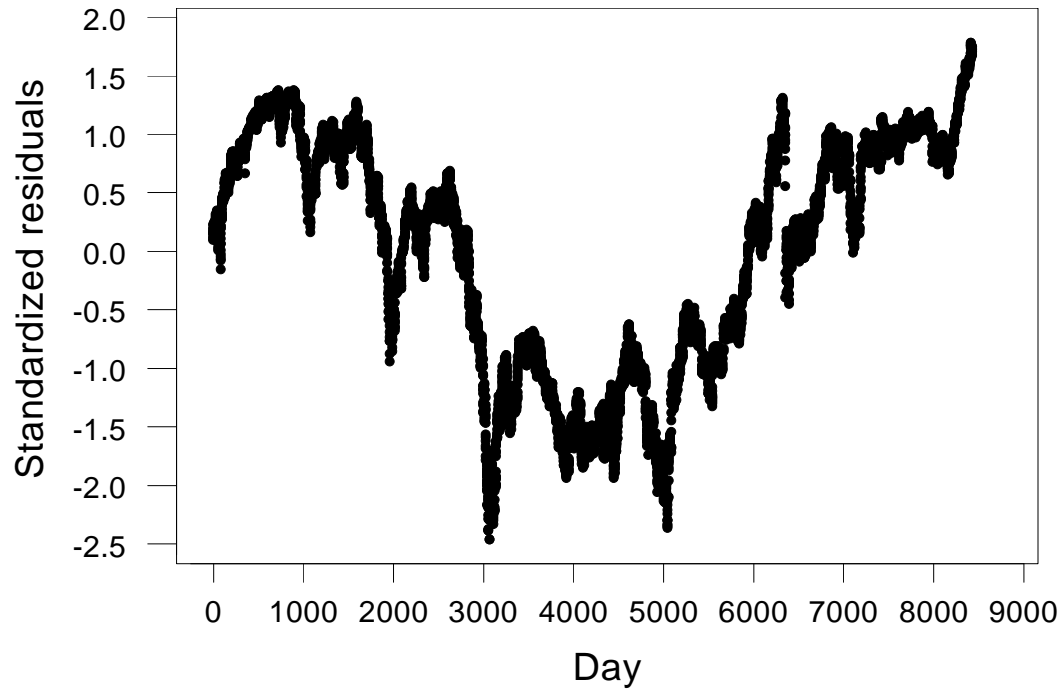
Predictor	Coef	StDev	T	P
Constant	1.73659	0.00229	759.30	0.000
Day	0.00010333	0.00000047	219.97	0.000

S = 0.1050      R-Sq = 85.2%      R-Sq(adj) = 85.2%

## Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	533.45	533.45	48387.29	0.000
Error	8430	92.94	0.01		
Total	8431	626.38			

The semilog estimate is  $10^{0.000103} - 1 = .00024$ , or smaller than what would be predicted by the geometric mean. It results in an annualized return of  $(1.00024)^{250} = 1.061$ , or a 6.1% annualized return. The semilog estimate requires that the errors from the regression be a random sample from a population that is at least reasonably Gaussian. Here is a time series plot of the standardized residuals, which clearly shows autocorrelation. The semilog estimate is **not** appropriate here, and we prefer the geometric mean of the returns as the best estimate.



It is very apparent that the period around the stock market crash of October 1987 is unusually volatile. This can be quantified by looking at the changes in logged index for all of the days except October 13, 1987 — February 10, 1988, and the values for those 84 days:

Descriptive Statistics

Variable	N	N*	Mean	Median	Tr Mean	StDev	SE Mean
Change i	8347	1	0.00013	0.00015	0.00012	0.00345	0.00004

Variable	Min	Max	Q1	Q3
Change i	-0.02745	0.02128	-0.00170	0.00197

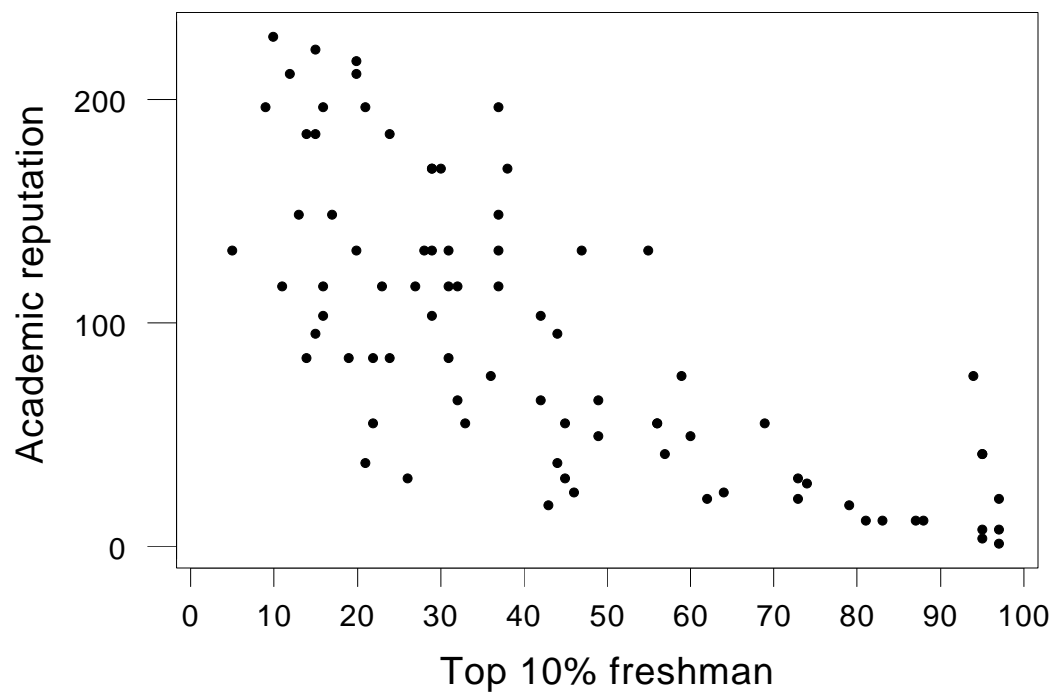
-----

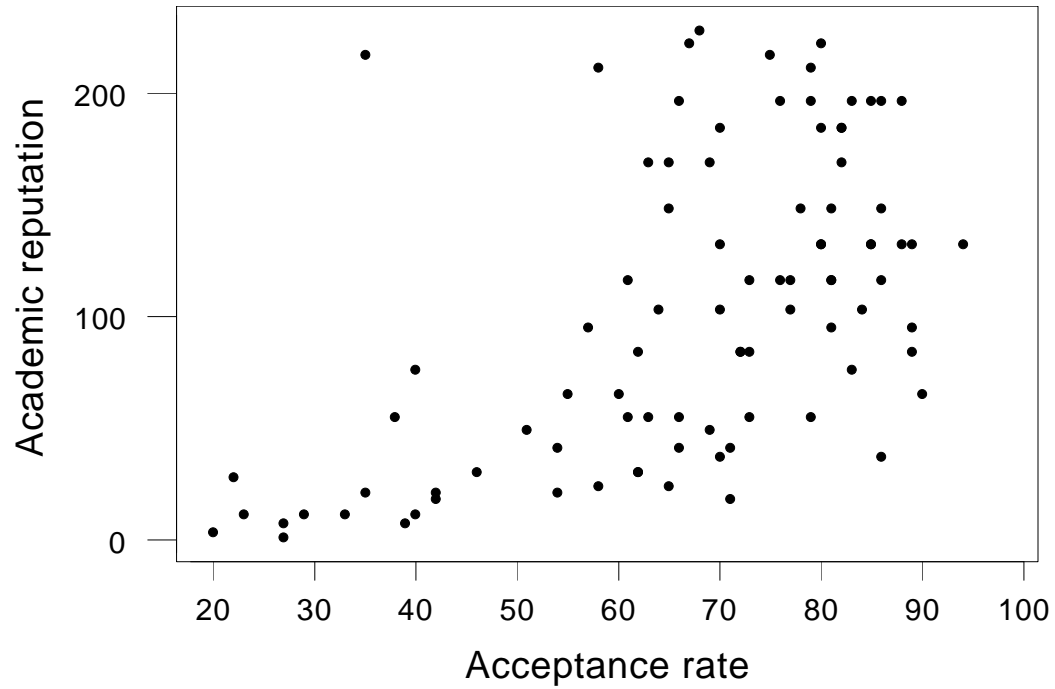
Variable	N	Mean	Median	Tr Mean	StDev	SE Mean
Change i	84	-0.00097	0.00068	0.00017	0.01534	0.00167

Variable	Min	Max	Q1	Q3
Change i	-0.09945	0.03782	-0.00558	0.00668

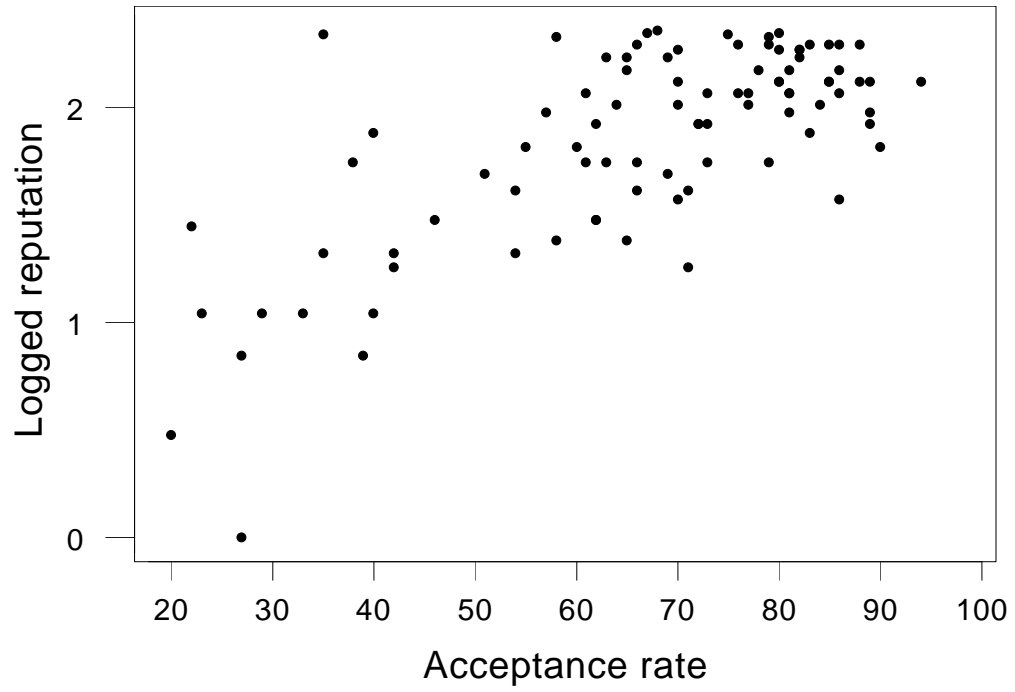
The estimated return without those days was  $10^{.00013} - 1 = .0003$ , for an annualized return of 7.77%. More importantly, the standard deviation of the change in logged index has dropped from .00376 to .00345. In contrast, the 84 days around the crash were money losing (geometric mean  $10^{-.00097} - 1 = -.00223$ , or an annualized return of -42.8%), and tremendously volatile (standard deviation of change in logged index of .0153, or more than 44 times larger than the standard deviation of the surrounding days).

2. First, let's look at a few scatter plots. Remember that a lower value for academic reputation is a good thing.

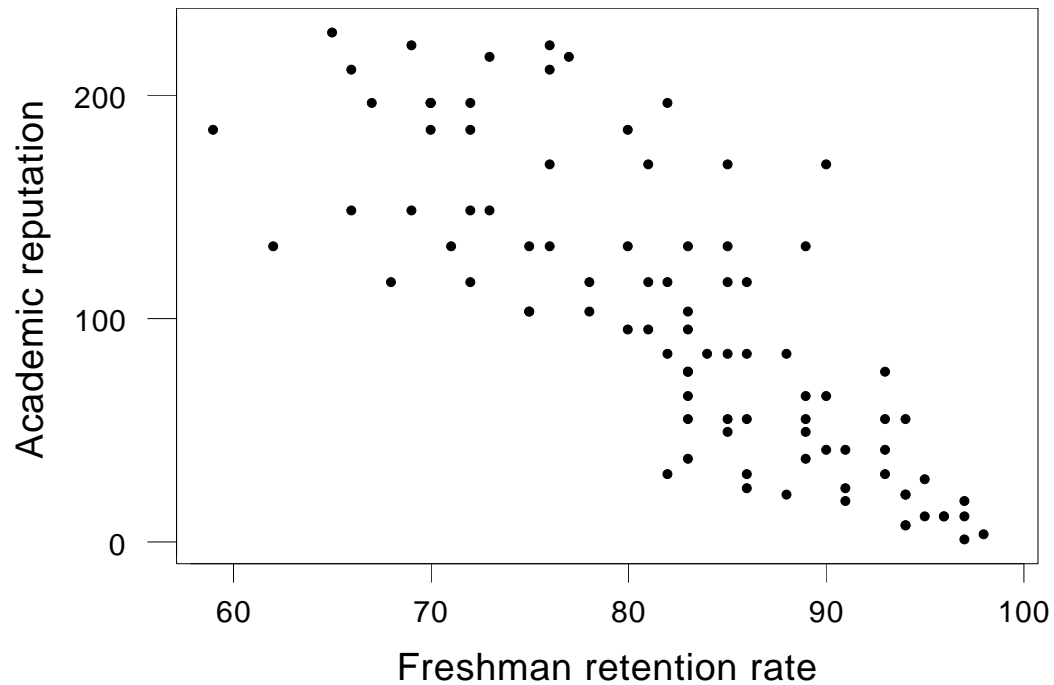




We see the expected relationship of a higher percentage of students from the top 10% of their high school classes being associated with better reputation, and higher acceptance rate being associated with worse reputation. Each of the plots hints at possible nonlinearity, though — might logged reputation be a better choice here?



Logs doesn't really seem to have helped much. A plot of reputation versus freshman retention rate looks particularly promising:



Let's first try a regression on all of the predictors (after logging the long right-tailed expenditure variable):

#### Regression Analysis

The regression equation is

$$\begin{aligned} \text{Academic reputation} = & 716 - 0.173 \text{ Top 10\% freshman} - 0.204 \text{ Acceptance rate} \\ & - 57.0 \text{ Logged expenditure} - 4.76 \text{ Freshman retention rate} \\ & + 0.601 \text{ Graduation rate} - 26.5 \text{ Top 50 school?} \end{aligned}$$

81 cases used 10 cases contain missing values

Predictor	Coef	StDev	T	P	VIF
Constant	716.0	133.2	5.37	0.000	
Top 10%	-0.1729	0.2904	-0.60	0.553	4.0
Acceptan	-0.2041	0.3389	-0.60	0.549	2.6
Logged e	-56.96	27.97	-2.04	0.045	2.4
Freshman	-4.760	1.078	-4.41	0.000	7.0
Graduati	0.6006	0.5260	1.14	0.257	6.2
Top 50 s	-26.51	13.53	-1.96	0.054	2.7

S = 34.71      R-Sq = 73.4%      R-Sq(adj) = 71.2%

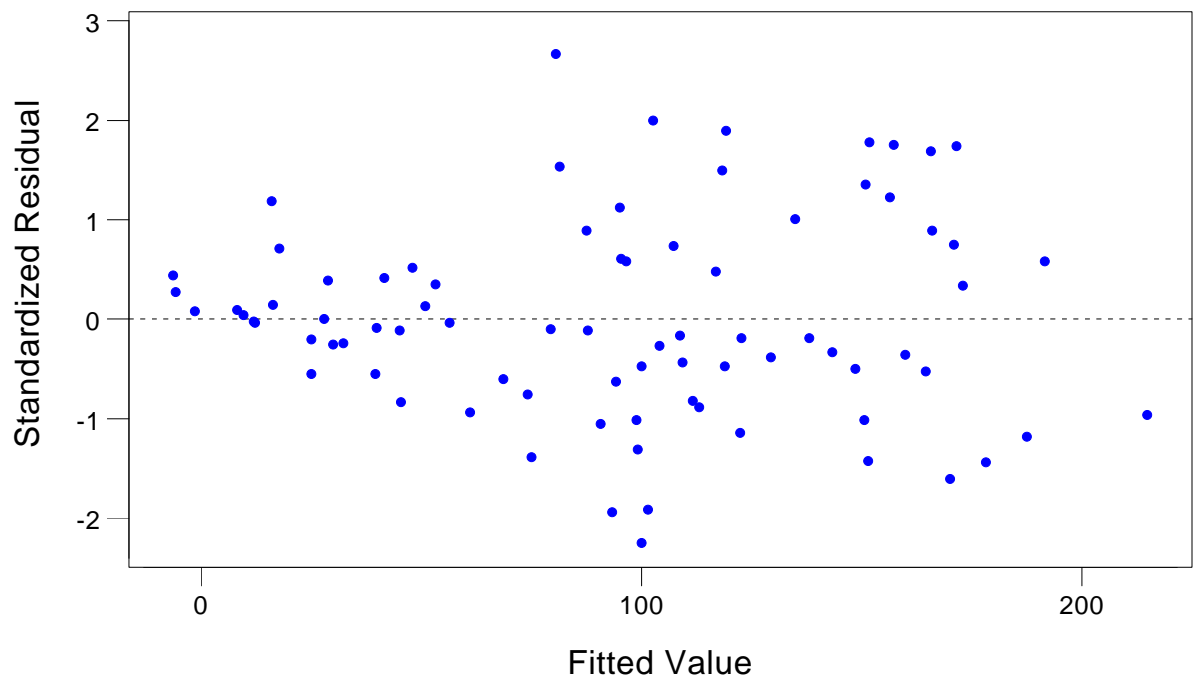
### Analysis of Variance

Source	DF	SS	MS	F	P
Regression	6	245426	40904	33.95	0.000
Error	74	89146	1205		
Total	80	334571			

Collinearity isn't a problem, but several variables are apparently not needed. There aren't any obvious problems with assumptions, although the residuals versus fitted values plot shows a bit of structure (more on that later).

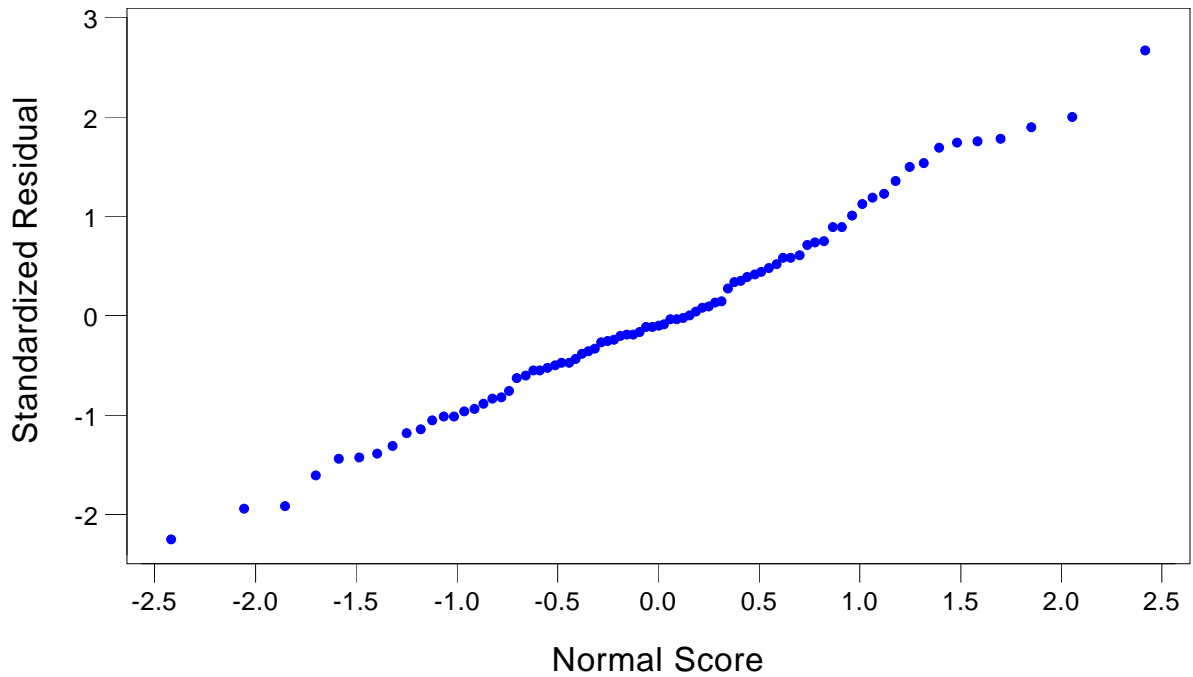
### Residuals Versus the Fitted Values

(response is Academic)



## Normal Probability Plot of the Residuals

(response is Academic)



Here is the regression fit on the variables that seem to help:

### Regression Analysis

The regression equation is

Academic reputation = 659 - 55.4 Logged expenditure  
- 3.91 Freshman retention rate - 28.5 Top 50 school?

90 cases used 1 cases contain missing values

Predictor	Coef	StDev	T	P	VIF
Constant	659.43	99.99	6.60	0.000	
Logged e	-55.45	26.60	-2.08	0.040	2.2
Freshman	-3.9131	0.6172	-6.34	0.000	2.2
Top 50 s	-28.49	12.85	-2.22	0.029	2.3

S = 36.66      R-Sq = 70.7%      R-Sq(adj) = 69.7%

## Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	278610	92870	69.09	0.000
Error	86	115601	1344		
Total	89	394211			

The regression is highly significant ( $F = 69.09$ ,  $p < .001$ ), with the three predictors accounting for roughly 70% of the variability in academic reputation. Given freshman retention rate and top 50 status are held fixed, multiplying expenditure per student by 10 is associated with reputation status that is 55 places higher; holding expenditure and top 50 status fixed, a 1 point increase in freshman retention rate is associated with 4 places higher in reputation; and holding expenditure and freshman retention rate fixed, a top 50 school is rated almost 30 places higher than a non-top 50 school. Regression diagnostics look okay here (the California Institute of Technology is a leverage point, due to its high expenditure per student, but removing it changes little):

Row	University	SRES2	HI2	COOK2
1	Andrews University	1.68025	0.050869	0.0378279
2	Auburn University	-1.07186	0.031728	0.0094117
3	Baylor University	-1.13422	0.030319	0.0100560
4	Biola University	2.22055	0.017085	0.0214267
5	Boston College	0.52842	0.071760	0.0053967
6	Boston University	-0.99318	0.046509	0.0120287
7	Brandeis University	-0.06231	0.040400	0.0000409
8	California Institute of Technology	0.41833	0.182775	0.0097847
9	Carnegie Mellon University	-0.58690	0.048899	0.0044274
10	Catholic University of America	0.67056	0.035166	0.0040973
11	Clark University	-0.09418	0.023746	0.0000539
12	Clarkson University	0.94266	0.029187	0.0066791
13	Dartmouth College	0.25317	0.051202	0.0008648
14	DePaul University	0.48469	0.025332	0.0015264
15	Duke University	0.14699	0.050638	0.0002881
16	Duquesne University	2.29324	0.062634	0.0878498
17	Florida Institute of Technology	1.20399	0.030313	0.0113288
18	George Mason University	-1.30861	0.023468	0.0102884
19	George Washington University	0.46512	0.042687	0.0024116
20	Georgetown University	0.23295	0.039371	0.0005560
21	Georgia Institute of Technology	-0.99929	0.058475	0.0155045
22	Idaho State University	-0.90558	0.105862	0.0242733
23	Illinois Institute of Technology	-0.57812	0.026424	0.0022678
24	Indiana Univ. of Pennsylvania	1.82196	0.020234	0.0171387
25	Indiana Univ.--Purdue Univ.--Indianapolis	-1.54729	0.128758	0.0884544
26	Kansas State University	-1.33226	0.025120	0.0114339
27	Lehigh University	0.51895	0.045910	0.0032396
28	Louisiana State Univ.-- Baton Rouge	-0.59861	0.021923	0.0020079
29	Massachusetts Inst. of Technology	0.08930	0.064201	0.0001368
30	Montana State University	0.28565	0.040339	0.0008575
31	North Carolina State Univ.--Raleigh	-0.21818	0.042412	0.0005271
32	Northeastern University	-1.04444	0.028271	0.0079340
33	Northwestern University	0.10525	0.047087	0.0001368
34	Nova Southeastern University	1.50975	0.089581	0.0560691
35	Ohio University	-0.67269	0.025761	0.0029913

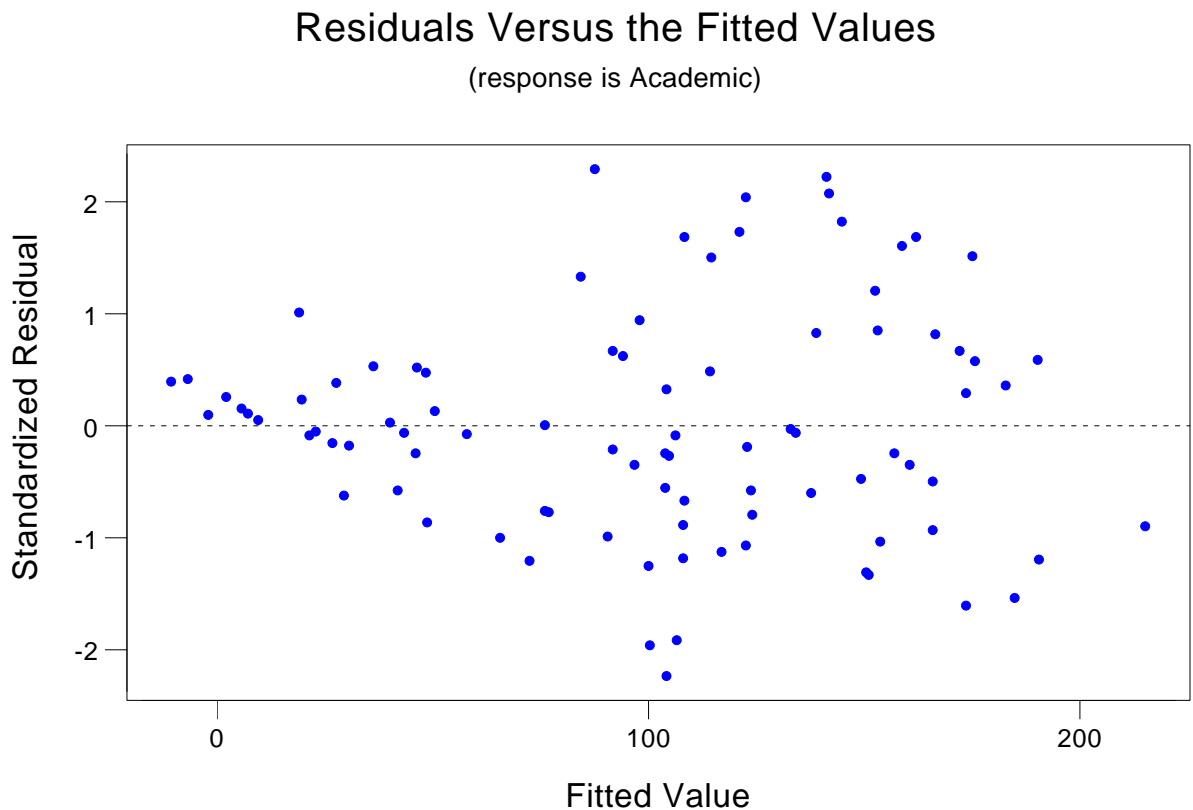
36	Oklahoma State University	-0.47903	0.030158	0.0017839
37	Oregon State University	-0.80213	0.016965	0.0027760
38	Pace University	1.73194	0.017383	0.0132664
39	Pennsylvania State Univ.	-1.20681	0.077390	0.0305410
40	San Diego State University	-0.02765	0.033841	0.0000067
41	St. John's University	-0.19426	0.021256	0.0002049
42	St. Louis University	0.62431	0.088660	0.0094794
43	Stevens Institute of Technology	-0.06412	0.037510	0.0000401
44	SUNY--Albany	-0.55543	0.028999	0.0023033
45	SUNY--Binghamton	-0.00132	0.087377	0.0000000
46	Syracuse University	0.12678	0.046540	0.0001962
47	Tennessee State University	2.06822	0.022086	0.0241520
48	Texas A&M Univ.--College Station	-0.76588	0.084636	0.0135587
49	Texas Woman's University	0.36178	0.047970	0.0016487
50	Tulane University	-0.07682	0.046822	0.0000725
51	U. of North Carolina--Chapel Hill	-0.15710	0.041066	0.0002642
52	Union Institute	*	*	*
53	Univ. of Arkansas-- Fayetteville	-0.35374	0.030967	0.0009997
54	Univ. of California--Los Angeles	-0.05251	0.037823	0.0000271
55	Univ. of California--San Diego	0.37495	0.037295	0.0013616
56	Univ. of California--Santa Cruz	-0.88745	0.022198	0.0044699
57	Univ. of Maryland-- Baltimore County	1.68578	0.036449	0.0268749
58	Univ. of Massachusetts--Lowell	0.81561	0.032123	0.0055195
59	Univ. of Minnesota--Twin Cities	-1.97042	0.052780	0.0540853
60	Univ. of Missouri--Columbia	-1.18602	0.022306	0.0080233
61	Univ. of Southern California	-0.24971	0.040279	0.0006543
62	Univ. of Southern Mississippi	1.60447	0.027169	0.0179736
63	Univ. of Wisconsin--Madison	-0.86643	0.067687	0.0136255
64	Univ. of Wisconsin--Milwaukee	-1.60707	0.041705	0.0280996
65	University of California--Berkeley	-0.62704	0.046643	0.0048091
66	University of California--Davis	0.02285	0.041390	0.0000056
67	University of Delaware	-0.35407	0.031116	0.0010065
68	University of Denver	1.50304	0.020850	0.0120264
69	University of Florida	-0.78026	0.052699	0.0084671
70	University of Houston	-0.94164	0.030537	0.0069824
71	University of Iowa	-1.92135	0.023519	0.0222287
72	University of Louisville	-0.50049	0.041024	0.0026790
73	University of Memphis	2.03459	0.031499	0.0336583
74	University of Miami	-0.25031	0.054889	0.0009097
75	University of Montana	-1.19567	0.058618	0.0222549
76	University of Nevada--Reno	0.84790	0.028859	0.0053410
77	University of North Texas	0.56801	0.045508	0.0038456
78	University of Northern Colorado	0.58112	0.058180	0.0052153
79	University of Pennsylvania	0.04530	0.051137	0.0000276
80	University of Pittsburgh	-1.25636	0.037733	0.0154737
81	University of Rochester	1.00844	0.047331	0.0126312
82	University of San Francisco	0.32700	0.029124	0.0008019
83	University of South Dakota	0.66489	0.036864	0.0042301
84	University of South Florida	0.82758	0.017424	0.0030363
85	University of Texas-- Dallas	-0.25195	0.025252	0.0004111
86	University of Texas--Austin	-2.23808	0.041097	0.0536692
87	University of Vermont	-0.26870	0.026230	0.0004862
88	University of Virginia	-0.09142	0.067705	0.0001518

89	Vanderbilt University	-0.18188	0.041484	0.0003579
90	Worcester Polytechnic Inst.	1.33110	0.045037	0.0208903
91	Yale University	0.39436	0.086363	0.0036752

Here is output that gives the desired prediction interval for NYU. The estimated academic reputation is 51, which is too high — NYU was actually rated at #41. Note that the prediction interval includes impossible negative values, which reflects the high variability and right tail in the reputation variable (building the model in the log scale would have avoided that problem).

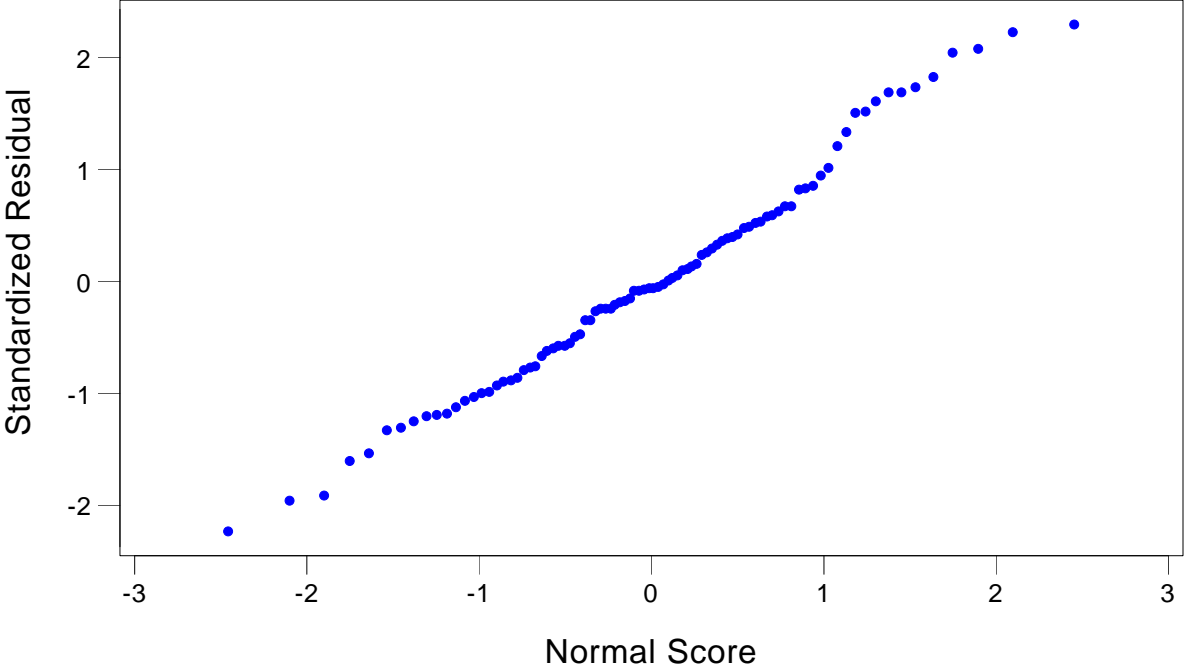
Fit	StDev Fit	95.0% CI	95.0% PI
51.41	8.42	( 34.68, 68.15)	( -23.38, 126.21)

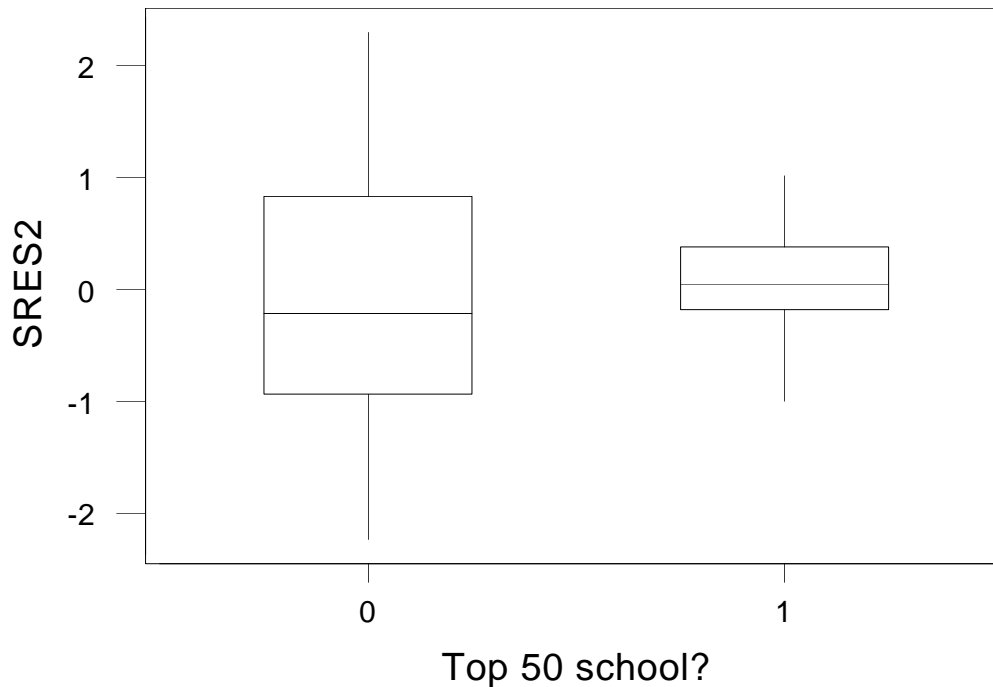
Residual plots indicate that this model still has problems. In particular, the residuals versus fitted values plot shows some structure, and less variability, for lower fitted reputation values. This corresponds to the top 50 schools, as side-by-side boxplots clearly show:



# Normal Probability Plot of the Residuals

(response is Academic)





To correct these problems would require methods beyond the scope of this course, including weighted least squares. To get a sense of what might be happening here, consider the following output, which summarizes the best two-predictor model for the top 50 schools. Note that it includes two different variables than the model on all schools does, and (despite a lower  $R^2$ ) has smaller standard error of the estimate, since there is less inherent variability of reputation in the top 50 schools than there is in the other schools (not surprisingly).

#### Regression Analysis

The regression equation is

Academic reputation = 39.1 - 0.422 Top 10% freshman + 0.430 Acceptance rate

Predictor	Coef	StDev	T	P	VIF
Constant	39.11	21.76	1.80	0.085	
Top 10%	-0.4220	0.1843	-2.29	0.031	1.9
Acceptan	0.4298	0.2132	2.02	0.055	1.9

S = 13.95      R-Sq = 54.9%      R-Sq(adj) = 51.2%

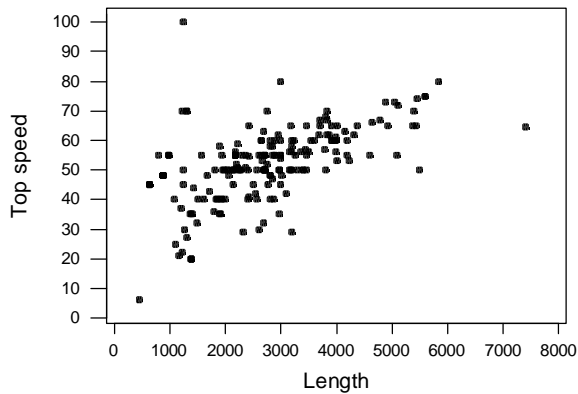
### Analysis of Variance

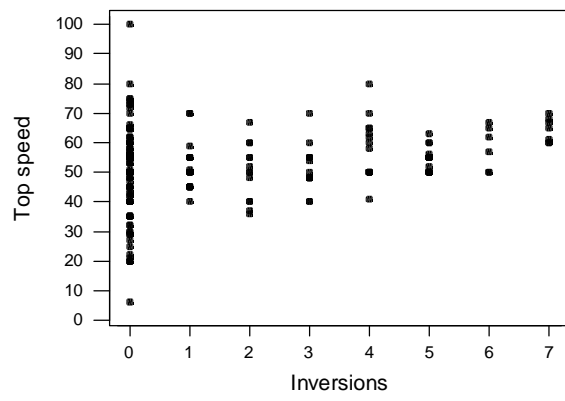
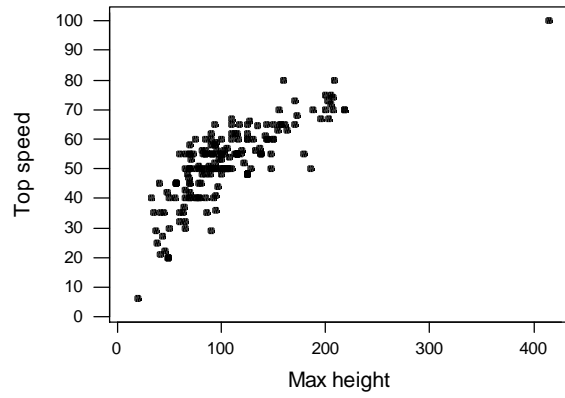
Source	DF	SS	MS	F	P
Regression	2	5692.7	2846.4	14.63	0.000
Error	24	4670.9	194.6		
Total	26	10363.6			

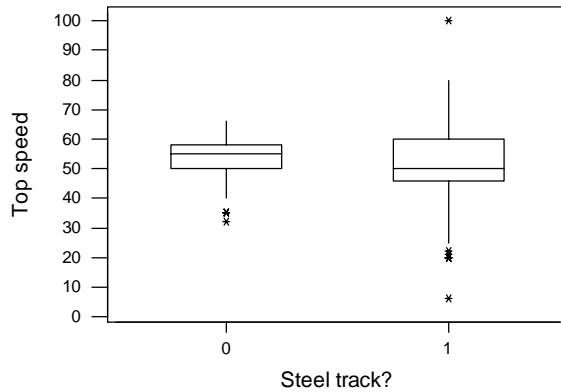
Fit	StDev	Fit	95.0% CI	95.0% PI
33.99	3.57	(	26.62, 41.36)	( 4.26, 63.72)

The prediction for NYU is now better than the truth (although very close to the *U.S. News and World Report* rating of #35), and the prediction interval does not include negative ratings.

- The first step is to look at the data using histograms, scatter plots, and side-by-side boxplots. Several show apparent relationships involving top speed; longer and higher coasters have higher top speeds, which makes sense. The number of inversions is weakly related to top speed, if at all. Note, however, the big difference between coasters with zero inversions and nonzero inversions; the reason for this is that all wooden track roller coasters have zero inversions (the cars can't go upside down on wooden tracks!). There doesn't seem to be much difference in top speed between wooden track and steel track rides, although there is more variability in the steel track coasters. Two rides also show up as potentially unusual: Beast, which is unusually long (7400 feet), and especially Superman The Escape, which is short (1235 feet), very high (415 feet), and very fast (100 miles per hour).







In the following regression all four predictors provide significant predictive power, although most of the fit actually comes from knowing height and length.

#### Regression Analysis

The regression equation is

$$\text{Top speed} = 29.8 - 5.75 \text{ Steel track?} + 0.00203 \text{ Length} + 0.184 \text{ Max height} + 0.734 \text{ Inversions}$$

210 cases used 1 cases contain missing values

Predictor	Coef	StDev	T	P	VIF
Constant	29.758	1.552	19.18	0.000	
Steel tr	-5.752	1.304	-4.41	0.000	1.7
Length	0.0020270	0.0004445	4.56	0.000	1.5
Max heig	0.18387	0.01207	15.24	0.000	1.5
Inversio	0.7340	0.2623	2.80	0.006	1.5

S = 6.632      R-Sq = 70.5%      R-Sq(adj) = 69.9%

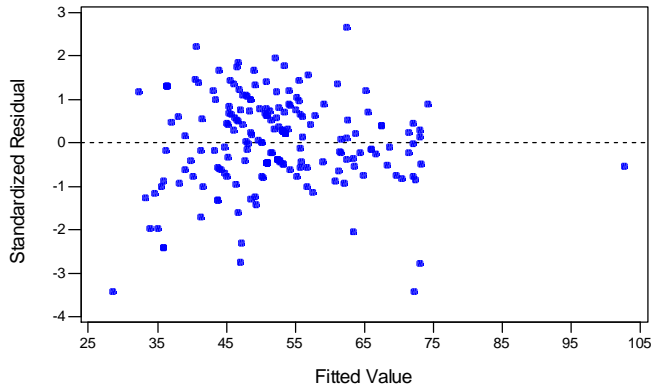
#### Analysis of Variance

Source	DF	SS	MS	F	P
Regression	4	21547.1	5386.8	122.46	0.000
Residual Error	205	9017.7	44.0		
Total	209	30564.8			

All of the coefficients are statistically significant, but Superman The Escape shows up as a very large leverage point:

### Residuals Versus the Fitted Values

(response is Top speed)



Row	Name	SRES1	HI1
1	Adventure Express	-1.00819	0.027069
2	Afterburner	1.32152	0.020102
3	Alpengeist	-0.77080	0.034240
4	American Eagle	0.52859	0.023166
5	Anaconda	-1.13245	0.024508
6	Apollo's Chariot	1.20895	0.043846
7	Bat	-0.45156	0.019742
8	Batman & Robin; The Chiller	0.69987	0.048858
9	Batman The Escape	1.37213	0.011118
10	Batman The Ride (Great America)	-0.23375	0.018578
11	Batman The Ride (Magic Mountain)	-0.37350	0.017688
12	Batman The Ride (Great Adventure)	-0.37134	0.017664
13	Batman The Ride (St. Louis)	-0.37134	0.017664
14	Batman The Ride (Georgia)	-0.37350	0.017688
15	Batman The Ride (Texas)	-0.37350	0.017688
16	Beast	-0.75563	0.078758
17	Big Bad Wolf	-0.01041	0.018243
18	Big Dipper	-2.30355	0.017751
19	Black Widow	1.32152	0.020102
20	Blue Streak (Cedar Point)	-1.41224	0.017367
21	Boomerang (Knott's Berry Farm)	-0.45156	0.019742
22	Boomerang (Marine World)	-0.45156	0.019742
23	Boomerang (Great Escape)	-0.45156	0.019742
24	Boomerang (Wild Adventures)	-0.45156	0.019742
25	Boomerang (Fiesta Texas)	-0.45156	0.019742
26	Boomerang (Elitch Gardens)	-0.45156	0.019742
27	Boomerang Coast to Coaster	-0.45156	0.019742
28	Cannonball	0.42103	0.018982
29	Cannonball Run	-0.32891	0.023741
30	Canyon Blaster	-1.23444	0.012220
31	Cedar Creek Mine Ride	0.61472	0.027874
32	Chang	-0.21580	0.023834
33	Coaster	-0.15480	0.017240

34	Colossus (Magic Mountain)	0.07489	0.020665
35	Colossus (Lagoon)	1.23183	0.012144
36	Comet (Great Escape)	-0.11162	0.019764
37	Comet (Lincoln Park)	1.09674	0.017664
38	Corkscrew (Cedar Point)	0.30418	0.009504
39	Corkscrew (Valleyfair!)	0.52770	0.013345
40	Corkscrew (Michigan's Adventure)	*	*
41	Corkscrew (Playland)	-0.62619	0.011637
42	Crazy Mouse (Steel Pier)	-2.41038	0.022397
43	Crazy Mouse (Myrtle Beach)	-2.41038	0.022397
44	Cyclone (Astroland)	1.40853	0.017079
45	Cyclops	1.75618	0.021826
46	Dahlonega Mine Train	-0.99793	0.030275
47	Demon (Great America)	-0.00215	0.010972
48	Demon (Paramount's Great America)	-0.00215	0.010972
49	Desperado	0.89442	0.069726
50	Diamond Back	1.32152	0.020102
51	Double Loop	-1.60335	0.008254
52	Drachen Fire	-0.37327	0.018792
53	Dragon Fyre	0.65900	0.014979
54	Dragon Mountain	-3.43127	0.040649
55	Dueling Dragons	0.43411	0.017478
56	Excalibur (Valleyfair!)	1.09657	0.016666
57	Excalibur (Funtown/Splashtown USA)	0.21029	0.017896
58	Exterminator	-2.41038	0.022397
59	Face/Off	0.21669	0.021732
60	Flashback (Texas)	-0.45156	0.019742
61	Flashback (Magic Mountain)	-1.31827	0.016721
62	Gemini	0.76973	0.027618
63	Georgia Cyclone	-0.49342	0.016317
64	Georgia Scorcher	0.30860	0.009522
65	Ghost Rider	-0.88122	0.022374
66	Ghoster Coaster	-0.77180	0.028584
67	Giant Dipper (Belmont Park)	0.99615	0.017338
68	Giant Dipper (Santa Cruz)	1.06775	0.017391
69	Great American Scream Machine (Great Adventure)	-0.10098	0.040218
70	Great American Scream Machine (Georgia)	0.14329	0.016363
71	Great Bear	1.34190	0.014740
72	Great Nor'Easter	0.27068	0.016425
73	Great White (Sea World)	-0.41469	0.016880
74	Great White (Morey's Piers)	-1.01469	0.016896
75	Grizzly (Great America)	0.29214	0.015878
76	Gwazi	-0.48610	0.016049
77	Hangman	0.27068	0.016425
78	Hercules	-0.26495	0.026793
79	High Roller (Valleyfair!)	0.20175	0.017004
80	Hurler	-0.21560	0.016049
81	Incredible Hulk	1.56361	0.045777
82	Invertigo	0.21669	0.021732
83	Iron Dragon	-0.55726	0.021772
84	Iron Wolf	0.79711	0.009771
85	Jack Rabbit (Kennywood)	0.54395	0.023171
86	Jaguar	-1.71313	0.022855

87	Joker's Revenge	-0.70601	0.010513
88	Jr. Gemini	-3.43542	0.033749
89	Judge Roy Scream	0.72627	0.017245
90	King Cobra	1.86360	0.010578
91	Kong	0.27068	0.016425
92	Kumba	-0.53924	0.041828
93	La Vibora	-0.92272	0.020032
94	Laser	0.75189	0.008005
95	Le Boomerang	-0.45156	0.019742
96	Le Monstre	-0.93172	0.020286
97	Loch Ness Monster	0.61461	0.010170
98	Mad Mouse (Valleyfair!)	-0.87638	0.022345
99	Magnum XL-200	-0.00760	0.055905
100	Mamba	0.30382	0.063891
101	Manhattan Express	-0.84174	0.035621
102	Mantis	-0.22932	0.016493
103	Mean Streak	-0.82309	0.035411
104	Medusa	-0.35911	0.042022
105	Mighty Canadian Minebuster	1.20699	0.017625
106	Mind Eraser (Six Flags America)	0.27068	0.016425
107	Mind Eraser (Geauga Lake)	-0.45156	0.019742
108	Mind Eraser (Elitch Gardens)	0.27068	0.016425
109	Mind Eraser (Riverside Park)	0.27068	0.016425
110	Mind Eraser (Darien Lake)	0.27068	0.016425
111	Mindbender	-0.20972	0.017960
112	Montezooma's Revenge	0.21866	0.033341
113	Montu	-0.73881	0.041205
114	Mr. Freeze (Texas)	0.39731	0.069811
115	Mr. Freeze (St. Louis)	0.39731	0.069811
116	Ninja (Georgia)	-0.55723	0.015971
117	Ninja (St. Louis)	1.95612	0.010301
118	Ninja (Magic Mountain)	2.21268	0.025225
119	Orient Express	1.44475	0.014672
120	Outlaw	0.03759	0.017378
121	Phoenix	-0.78759	0.016316
122	Predator	-0.62582	0.015985
123	Python	1.38398	0.012518
124	Ragin' Cajun	-0.45156	0.019742
125	Raging Bull	0.24835	0.054192
126	Rampage	-0.44380	0.018010
127	Raptor	-0.65536	0.029285
128	Rattler	-2.77842	0.037569
129	Raven	0.66071	0.018788
130	Rebel Yell	0.57539	0.016208
131	Red Devil	0.69475	0.010804
132	Revolution (Magic Mountain)	0.37629	0.016103
133	Revolution (Libertyland)	0.67819	0.010276
134	Riddler's Revenge	-0.14600	0.034030
135	Roar	-0.42444	0.015876
136	Roller Coaster	-0.41022	0.017853
137	Rolling Thunder	0.31981	0.015964
138	Runaway Mountain	0.15257	0.019222
139	Scooby Doo's Ghoster Coaster	-0.61218	0.029066

140	Scorpion	1.46538	0.014166
141	Screaming Eagle	0.63397	0.017486
142	Sea Serpent	-0.45156	0.019742
143	Serial Thriller	0.63377	0.017395
144	Shivering Timbers	0.20609	0.031922
145	Shockwave (Great America)	-0.50906	0.040520
146	Shockwave (Texas)	0.89578	0.013801
147	Shockwave (King Dominion)	0.49571	0.010592
148	Sidewinder (Hersheypark)	-0.45156	0.019742
149	Sidewinder (Elitch Gardens)	1.32152	0.020102
150	Silver Bullet	1.67756	0.011415
151	Sky Princess	-0.59825	0.021589
152	SkyRider	0.84907	0.011135
153	Space Mountain	-2.75028	0.022738
154	Steamin' Demon	-0.76989	0.010226
155	Steel Eel	0.90376	0.026620
156	Steel Force	0.44676	0.062517
157	Steel Phantom	2.66546	0.013590
158	Superman - Ride of Steel	-0.49747	0.061370
159	Superman The Escape	-0.54182	0.386365
160	Swamp Fox	0.24148	0.018165
161	T2	0.27068	0.016425
162	Taz's Texas Tornado	0.98302	0.017098
163	Tennessee Tornado	0.09887	0.014837
164	Texas Cyclone	1.77769	0.015900
165	Texas Giant	-0.15657	0.027182
166	Thunder Road	0.71100	0.015895
167	Thunder Run	1.18178	0.027065
168	Thunderation	0.45376	0.022744
169	Thunderbolt (Kennywood)	0.99140	0.017006
170	Thunderbolt Express	1.32152	0.020102
171	Thunderhawk	-0.77180	0.016599
172	Tidal Wave	-0.45156	0.019742
173	Timber Terror	0.62956	0.016827
174	Timber Wolf	-0.56635	0.019647
175	Top Gun (Canada's Wonderland)	0.78649	0.017396
176	Top Gun (Kings Island)	1.19995	0.018441
177	Top Gun (Great America)	0.06907	0.007593
178	Top Gun - The Jet Coaster	1.04244	0.026723
179	Tornado	0.81882	0.016615
180	Tree Topper	-1.31891	0.027156
181	Tremors	0.87789	0.016581
182	Two Face: The Flip Side	0.21669	0.021732
183	Ultra Twister	-0.10421	0.018472
184	Vampire	-0.45156	0.019742
185	Vapor Trail	-1.15183	0.023958
186	Viper (Great America)	-0.78349	0.016134
187	Viper (Magic Mountain)	-0.22728	0.042577
188	Viper (Great Adventure)	0.42053	0.008946
189	Viper (Astroland)	0.99487	0.011773
190	Volcano, The Blast Coaster	1.36121	0.012870
191	Vortex (Canada's Wonderland)	1.44105	0.016954
192	Vortex (Kings Island)	-2.03846	0.028576

193	Vortex (Great America)	-0.95101	0.008156
194	Vortex (Carowinds)	0.58179	0.008231
195	Whirlwind	-0.40943	0.013647
196	Whizzer	-0.17733	0.026483
197	Wild Chipmunk	-1.96572	0.024688
198	Wild Maus	-1.97276	0.023321
199	Wild One	-0.43902	0.018124
200	Wild Thing	0.13477	0.062035
201	Wildcat (Cedar Point)	0.46930	0.022675
202	Wildcat (Hersheypark)	-0.41921	0.015878
203	Wildcat (Frontier City)	1.68415	0.017080
204	Wilde Beaste	0.72652	0.016096
205	Windjammer	-0.17894	0.013418
206	Wolverine Wildcat	0.53688	0.016053
207	Woodstock's Express	-1.25593	0.025587
208	Zeus	1.18809	0.016292
209	Zippin Pippin	-1.28293	0.017019
210	Zoomerang	-0.45156	0.019742
211	Zyklon	-0.17857	0.024512

---

Row	Name	COOK1
1	Adventure Express	0.0056559
2	Afterburner	0.0071653
3	Alpengeist	0.0042128
4	American Eagle	0.0013253
5	Anaconda	0.0064440
6	Apollo's Chariot	0.0134043
7	Bat	0.0008213
8	Batman & Robin; The Chiller	0.0050322
9	Batman The Escape	0.0042336
10	Batman The Ride (Great America)	0.0002069
11	Batman The Ride (Magic Mountain)	0.0005024
12	Batman The Ride (Great Adventure)	0.0004959
13	Batman The Ride (St. Louis)	0.0004959
14	Batman The Ride (Georgia)	0.0005024
15	Batman The Ride (Texas)	0.0005024
16	Beast	0.0097627
17	Big Bad Wolf	0.0000004
18	Big Dipper	0.0191791
19	Black Widow	0.0071653
20	Blue Streak (Cedar Point)	0.0070500
21	Boomerang (Knott's Berry Farm)	0.0008213
22	Boomerang (Marine World)	0.0008213
23	Boomerang (Great Escape)	0.0008213
24	Boomerang (Wild Adventures)	0.0008213
25	Boomerang (Fiesta Texas)	0.0008213
26	Boomerang (Elitch Gardens)	0.0008213
27	Boomerang Coast to Coaster	0.0008213
28	Cannonball	0.0006860
29	Cannonball Run	0.0005261

30	Canyon Blaster	0.0037705
31	Cedar Creek Mine Ride	0.0021670
32	Chang	0.0002274
33	Coaster	0.0000841
34	Colossus (Magic Mountain)	0.0000237
35	Colossus (Lagoon)	0.0037308
36	Comet (Great Escape)	0.0000502
37	Comet (Lincoln Park)	0.0043259
38	Corkscrew (Cedar Point)	0.0001776
39	Corkscrew (Valleyfair!)	0.0007533
40	Corkscrew (Michigan's Adventure)	*
41	Corkscrew (Playland)	0.0009233
42	Crazy Mouse (Steel Pier)	0.0266209
43	Crazy Mouse (Myrtle Beach)	0.0266209
44	Cyclone (Astroland)	0.0068945
45	Cyclops	0.0137635
46	Dahlongega Mine Train	0.0062182
47	Demon (Great America)	0.0000000
48	Demon (Paramount's Great America)	0.0000000
49	Desperado	0.0119922
50	Diamond Back	0.0071653
51	Double Loop	0.0042790
52	Drachen Fire	0.0005337
53	Dragon Fyre	0.0013208
54	Dragon Mountain	0.0997736
55	Dueling Dragons	0.0006704
56	Excalibur (Valleyfair!)	0.0040759
57	Excalibur (Funtown/Splashtown USA)	0.0001612
58	Exterminator	0.0266209
59	Face/Off	0.0002086
60	Flashback (Texas)	0.0008213
61	Flashback (Magic Mountain)	0.0059106
62	Gemini	0.0033656
63	Georgia Cyclone	0.0008077
64	Georgia Scorcher	0.0001831
65	Ghost Rider	0.0035543
66	Ghoster Coaster	0.0035055
67	Giant Dipper (Belmont Park)	0.0035017
68	Giant Dipper (Santa Cruz)	0.0040357
69	Great American Scream Machine (Great Adventure)	0.0000855
70	Great American Scream Machine (Georgia)	0.0000683
71	Great Bear	0.0053877
72	Great Nor'Easter	0.0002447
73	Great White (Sea World)	0.0005905
74	Great White (Morey's Piers)	0.0035389
75	Grizzly (Great America)	0.0002754
76	Gwazi	0.0007708
77	Hangman	0.0002447
78	Hercules	0.0003865
79	High Roller (Valleyfair!)	0.0001408
80	Hurler	0.0001516
81	Incredible Hulk	0.0234580
82	Invertigo	0.0002086

83	Iron Dragon	0.0013823
84	Iron Wolf	0.0012539
85	Jack Rabbit (Kennywood)	0.0014037
86	Jaguar	0.0137287
87	Joker's Revenge	0.0010592
88	Jr. Gemini	0.0824440
89	Judge Roy Scream	0.0018511
90	King Cobra	0.0074264
91	Kong	0.0002447
92	Kumba	0.0025387
93	La Vibora	0.0034807
94	Laser	0.0009125
95	Le Boomerang	0.0008213
96	Le Monstre	0.0035949
97	Loch Ness Monster	0.0007762
98	Mad Mouse (Valleyfair!)	0.0035109
99	Magnum XL-200	0.0000007
100	Mamba	0.0012601
101	Manhattan Express	0.0052342
102	Mantis	0.0001764
103	Mean Streak	0.0049742
104	Medusa	0.0011313
105	Mighty Canadian Minebuster	0.0052275
106	Mind Eraser (Six Flags America)	0.0002447
107	Mind Eraser (Geauga Lake)	0.0008213
108	Mind Eraser (Elitch Gardens)	0.0002447
109	Mind Eraser (Riverside Park)	0.0002447
110	Mind Eraser (Darien Lake)	0.0002447
111	Mindbender	0.0001609
112	Montezooma's Revenge	0.0003298
113	Montu	0.0046915
114	Mr. Freeze (Texas)	0.0023695
115	Mr. Freeze (St. Louis)	0.0023695
116	Ninja (Georgia)	0.0010079
117	Ninja (St. Louis)	0.0079648
118	Ninja (Magic Mountain)	0.0253395
119	Orient Express	0.0062160
120	Outlaw	0.0000050
121	Phoenix	0.0020577
122	Predator	0.0012725
123	Python	0.0048564
124	Ragin' Cajun	0.0008213
125	Raging Bull	0.0007068
126	Rampage	0.0007224
127	Raptor	0.0025915
128	Rattler	0.0602677
129	Raven	0.0016717
130	Rebel Yell	0.0010909
131	Red Devil	0.0010543
132	Revolution (Magic Mountain)	0.0004635
133	Revolution (Libertyland)	0.0009550
134	Riddler's Revenge	0.0001502
135	Roar	0.0005813

136	Roller Coaster	0.0006118
137	Rolling Thunder	0.0003319
138	Runaway Mountain	0.0000912
139	Scooby Doo's Ghoster Coaster	0.0022438
140	Scorpion	0.0061710
141	Screaming Eagle	0.0014306
142	Sea Serpent	0.0008213
143	Serial Thriller	0.0014221
144	Shivering Timbers	0.0002801
145	Shockwave (Great America)	0.0021888
146	Shockwave (Texas)	0.0022458
147	Shockwave (King Dominion)	0.0005261
148	Sidewinder (Hersheypark)	0.0008213
149	Sidewinder (Elitch Gardens)	0.0071653
150	Silver Bullet	0.0064993
151	Sky Princess	0.0015795
152	SkyRider	0.0016236
153	Space Mountain	0.0351980
154	Steamin' Demon	0.0012248
155	Steel Eel	0.0044675
156	Steel Force	0.0026621
157	Steel Phantom	0.0195767
158	Superman - Ride of Steel	0.0032361
159	Superman The Escape	0.0369683
160	Swamp Fox	0.0002158
161	T2	0.0002447
162	Taz's Texas Tornado	0.0033618
163	Tennessee Tornado	0.0000294
164	Texas Cyclone	0.0102118
165	Texas Giant	0.0001370
166	Thunder Road	0.0016331
167	Thunder Run	0.0077703
168	Thunderation	0.0009584
169	Thunderbolt (Kennywood)	0.0034009
170	Thunderbolt Express	0.0071653
171	Thunderhawk	0.0020109
172	Tidal Wave	0.0008213
173	Timber Terror	0.0013567
174	Timber Wolf	0.0012856
175	Top Gun (Canada's Wonderland)	0.0021902
176	Top Gun (Kings Island)	0.0054104
177	Top Gun (Great America)	0.0000073
178	Top Gun - The Jet Coaster	0.0059673
179	Tornado	0.0022657
180	Tree Topper	0.0097114
181	Tremors	0.0025989
182	Two Face: The Flip Side	0.0002086
183	Ultra Twister	0.0000409
184	Vampire	0.0008213
185	Vapor Trail	0.0065131
186	Viper (Great America)	0.0020133
187	Viper (Magic Mountain)	0.0004594
188	Viper (Great Adventure)	0.0003193

189	Viper (Astroland)	0.0023583
190	Volcano, The Blast Coaster	0.0048315
191	Vortex (Canada's Wonderland)	0.0071627
192	Vortex (Kings Island)	0.0244473
193	Vortex (Great America)	0.0014874
194	Vortex (Carowinds)	0.0005619
195	Whirlwind	0.0004639
196	Whizzer	0.0001711
197	Wild Chipmunk	0.0195621
198	Wild Maus	0.0185858
199	Wild One	0.0007115
200	Wild Thing	0.0002403
201	Wildcat (Cedar Point)	0.0010220
202	Wildcat (Hersheypark)	0.0005671
203	Wildcat (Frontier City)	0.0098575
204	Wilde Beaste	0.0017270
205	Windjammer	0.0000871
206	Wolverine Wildcat	0.0009405
207	Woodstock's Express	0.0082840
208	Zeus	0.0046755
209	Zippin Pippin	0.0056992
210	Zoomerang	0.0008213
211	Zyklon	0.0001603

Does removing this one very unusual ride (it's at Six Flags Magic Mountain in Valencia, California, for you crazy people) change things? Let's see. Remember that anything we now see doesn't apply to unusual rides of this type.

#### Regression Analysis

The regression equation is

$$\text{Top speed} = 29.6 - 5.83 \text{ Steel track?} + 0.00192 \text{ Length} + 0.189 \text{ Max height} + 0.697 \text{ Inversions}$$

209 cases used 1 cases contain missing values

Predictor	Coef	StDev	T	P	VIF
Constant	29.637	1.570	18.87	0.000	
Steel tr	-5.827	1.314	-4.44	0.000	1.7
Length	0.0019214	0.0004862	3.95	0.000	1.8
Max heig	0.18893	0.01528	12.36	0.000	1.8
Inversio	0.6969	0.2716	2.57	0.011	1.6

S = 6.644      R-Sq = 68.1%      R-Sq(adj) = 67.5%

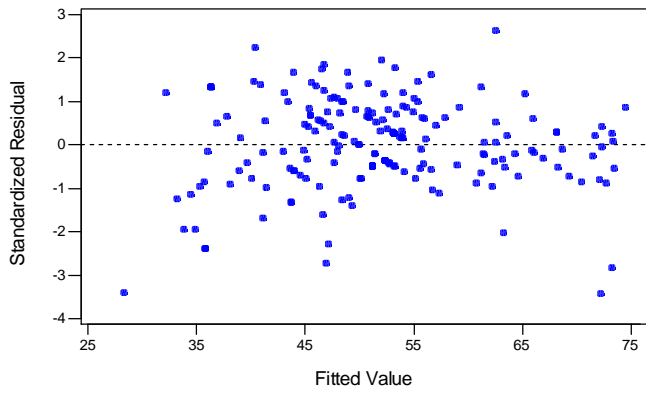
#### Analysis of Variance

Source	DF	SS	MS	F	P
Regression	4	19240.0	4810.0	108.97	0.000
Residual Error	204	9004.8	44.1		
Total	208	28244.8			

Not much has changed. Once again, by the way, just looking at a ride will tell you a lot about how fast it is — almost two-thirds of the variability in speed is accounted for by height and length. The regression coefficients reflect that additional length and additional height (given the other variables) are associated with higher speed. In addition, given the type of track, length, and height, each inversion is associated with about .7 MPH in increased speed. An interesting result is that, given everything else, steel track rides are almost 6 miles per hour slower on average than wood track rides. Residual plots look a bit better, although there is evidence of nonconstant variance related to type of track and the number of inversions. Given the nature of the residuals versus fitted values plot, trying a weighted least squares analysis would probably not make that much of a difference. There are four rides that are noticeably slower than we would have predicted (Dragon Mountain, Junior Gemini, Rattler, and Space Mountain), but removing these points changes little. The two Mr. Freeze coasters are marginally leverage points (they are the highest rides left in the sample, and are among the shortest), but similarly omitting them changes little.

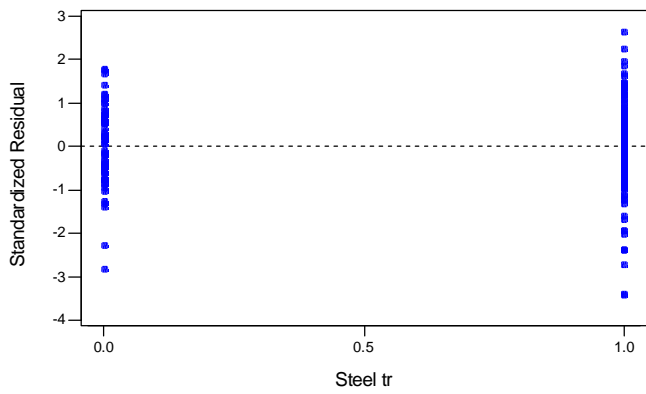
### Residuals Versus the Fitted Values

(response is Top spee)



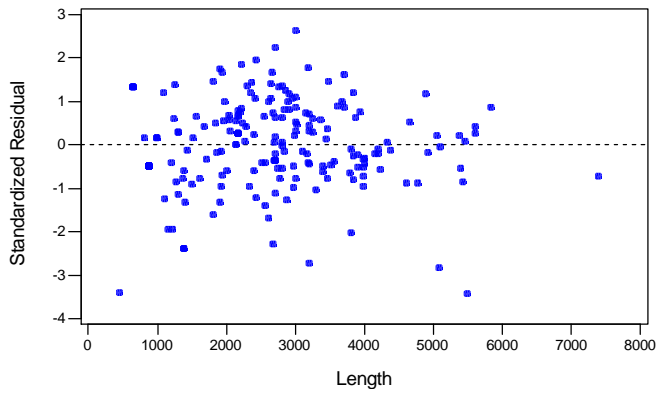
### Residuals Versus Steel tr

(response is Top spee)



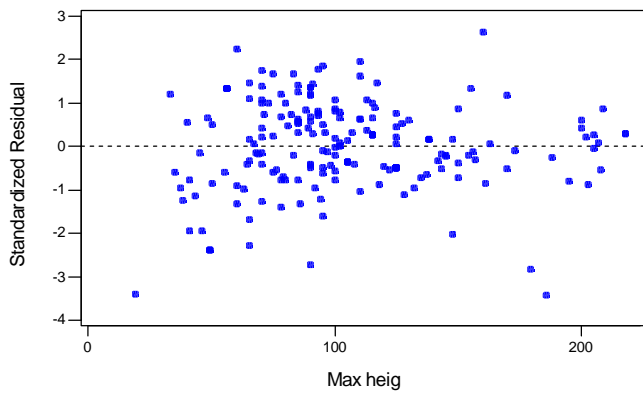
Residuals Versus Length

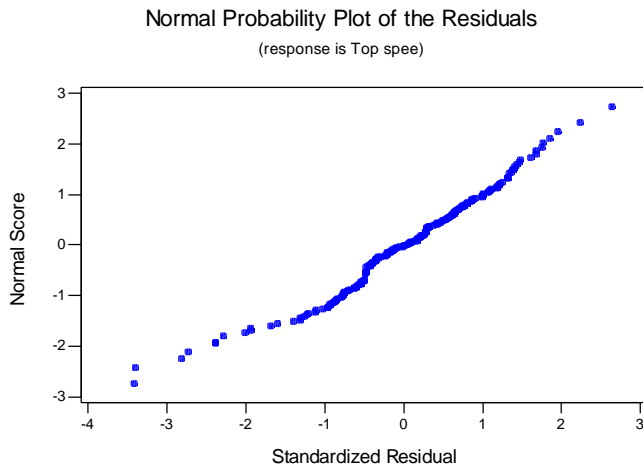
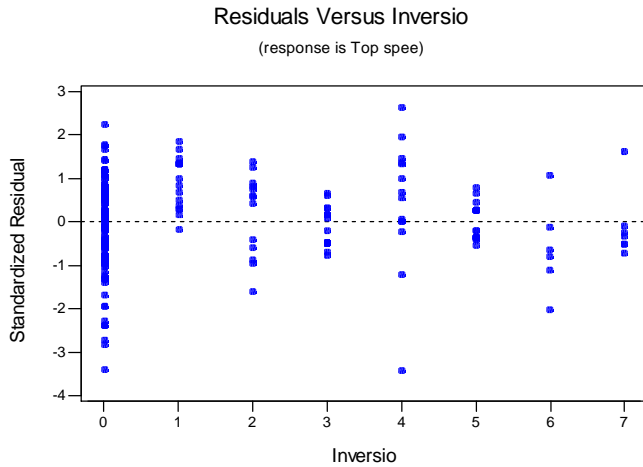
(response is Top spee)



Residuals Versus Max heig

(response is Top spee)





What about the 27 other coasters? Here are predicted top speeds for them, along with prediction interval limits:

Data Display

Row		C11	PFIT1	PLIM1	PLIM2
1	Alpine Bobsled	38.7638	25.5376	51.9901	
2	Arkansas Twister	54.0025	40.7991	67.2059	
3	Blue Streak (Conneaut Lake)	49.9453	36.7383	63.1523	
4	Cyclone (Lakeside)	52.0203	38.8117	65.2289	
5	Disaster Transport	39.4242	26.1927	52.6557	
6	Doo Wopper	32.1225	18.8410	45.4040	
7	Dragon	46.5301	33.3743	59.6860	
8	Dragon Coaster	51.2839	38.0728	64.4950	
9	Gold Rusher	42.0302	28.7835	55.2769	
10	Grizzly (Kings Dominion)	52.1260	38.9226	65.3294	
11	Jack Rabbit (Clementon)	41.7348	28.4557	55.0139	

12	Jack Rabbit (Seabreeze)	47.8990	34.6689	61.1291
13	Leap-the-Dips	42.5098	29.2598	55.7598
14	Mad Mouse (Michigan's Adventure)	39.0930	25.8661	52.3199
15	Mini Mine Train	29.3366	16.0127	42.6605
16	Racer	47.6762	34.4523	60.9000
17	Rollo Coaster	36.4671	23.1320	49.8022
18	Run-A-Way Mine Train	35.0335	21.6897	48.3772
19	Santa Monica West Coaster	36.6984	23.4589	49.9379
20	Sea Dragon	38.7856	25.4894	52.0817
21	Skyliner	45.5839	32.3578	58.8100
22	Texas Tornado	44.2566	31.0913	57.4219
23	Thunderbolt (Riverside)	48.3666	35.1537	61.5795
24	Tree Top Racers	32.1513	18.8700	45.4326
25	Twisted Sisters	50.5153	37.3096	63.7211
26	Twister	56.3254	43.1119	69.5390
27	XLR-8	46.4141	33.0806	59.7476

The daredevils among you will be heading to Knoebels Park in Elysburg Pennsylvania to ride Twister (it was scheduled to open sometime in 1999) or to the Magic Springs Family Theme Park to ride the Arkansas Twister (although it was not operating as of mid-1999). Me, I'll be gritting my teeth on the Mini Mine Train at Six Flags Over Texas in Arlington, Texas!

4. The process is the same as in the previous questions, but I'll only outline the results here. First of all, I would recommend using logged median household income rather than median household income, since the relationships are a bit more linear when using logs. I'll proceed using the unlogged target variable here, however. The expected marginal relationships are evident: a direct relationship between income and percent white, percent with college degrees and percent with a graduate degree, and an inverse relationship between income and percent black and percent with only a high school degree. There is also a direct relationship between income and median age, with older areas having higher income. There is apparently little relationship between income and percent of the residents that are women. None of the ZIP areas show up as being particularly unusual. Here is a regression on all predictors:

#### Regression Analysis

The regression equation is

$$\text{Median Household Income} = -53598 + 763 \% \text{ Female} + 100 \% \text{ White} - 36.1 \% \text{ Black} \\ + 763 \text{ Median Age} + 5 \% \text{ HS Max} + 3 \% \text{ College Max} + 660 \% \text{ Grad}$$

Predictor	Coef	StDev	T	P	VIF
Constant	-53598	12526	-4.28	0.000	
% Female	762.9	272.4	2.80	0.009	2.5
% White	100.4	101.2	0.99	0.330	22.6
% Black	-36.10	73.39	-0.49	0.627	10.7
Median A	763.1	260.4	2.93	0.007	2.5
% HS Max	4.7	540.1	0.01	0.993	25.5
% Colleg	3.3	271.1	0.01	0.990	21.7
% Grad	660.2	260.4	2.54	0.018	19.0

S = 3774      R-Sq = 93.6%      R-Sq(adj) = 91.8%

### Analysis of Variance

Source	DF	SS	MS	F	P
Regression	7	5375320996	767902999	53.92	0.000
Residual Error	26	370267662	14241064		
Total	33	5745588658			

Here are results from a regression on percent female, median age, and percent with a graduate degree:

### Regression Analysis

The regression equation is

Median Household Income = - 42588 + 445 % Female + 930 Median Age + 927 % Grad

Predictor	Coef	StDev	T	P	VIF
Constant	-42588	12323	-3.46	0.002	
% Female	444.6	182.4	2.44	0.021	1.0
Median A	930.3	212.6	4.38	0.000	1.5
% Grad	926.51	76.74	12.07	0.000	1.5

S = 3977      R-Sq = 91.7%      R-Sq(adj) = 90.9%

### Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	5271188230	1757062743	111.11	0.000
Residual Error	30	474400428	15813348		
Total	33	5745588658			

With an  $R^2$  of roughly 91%, these three variables are very good at modeling household income. While the marginal relationship between income and percent female wasn't very strong, given median age and percent with a graduate degree, a higher percent of female residents is associated with higher per capita income. The only problem in residual plots or diagnostics is that observation 34 (ZIP area 10044) is a bit of an outlier, apparently because of a high income (\$44518) with moderate percentage of people with graduate degrees (17.8%). Omitting it doesn't affect things very much.