Name Spring, 2012

Applied Statistics Comprehensive Examination
Regression & Linear Models

1. (40 Points) Consider the following main effects model without replicates:

$$y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$$

where Factor A has three levels ($i = 1, 2, 3$) and Factor B has two levels ($j = 1, 2$).

   (a) Construct contrasts for testing the differences in the levels of Factor A such that the sums of squares for the contrasts would sum to the total sum of squares for Factor A.

   (b) Write the design matrix using sum to zero restrictions.

   (c) Write out a basis set of estimable functions.

   (d) Is $\alpha_1 - 2\alpha_2 + \alpha_3$ estimable? Explain.

2. (10 Points) When would Type II and Type III Sums of Squares be the same, but differ from the Type I Sum of Squares?

3. (10 points) Derive the least squares estimators for β_0 and β_1 for simple linear regression, $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ (do not use matrix form).

4. (40 points) Various races are run on the Scottish hills. For each race, the record time (in seconds), the distance (in miles), and the climb (in feet) were recorded. The following output data, regression equation, and plots were produced by predicting time from distance and climb.

   a. (10 points) State the assumptions that are necessary for this multiple regression. For each assumption, state whether it appears to be true from the following output or whether it's not possible to assess it from the output.
   b. (5 points) The value of one of the parameter estimates is 0.66 (0.66289 in the output). Interpret this value.
   c. (4 points) State and interpret the value of R-squared.
   d. (8 points) Identify high leverage points. State evidence to show that they are high leverage.
   e. (8 points) Identify high influence points. State evidence to show that they are high influence.
   f. (5 points) Based on the information you have, do you think that this regression equation correctly fits the model? If yes, state evidence. If no, suggest how the model might be modified.

Obs	Hill	Time	Distance	Climb
1	Greenman	965	2.5	650
2	Carnethy	2901	6.0	2500
3	CraigDun	2019	6.0	900
4	BenRha	2736	7.5	800
5	BenLomon	3736	8.0	3070
6	Goatfell	4393	8.0	2866
7	BensofJu	12277	16.0	7500
8	Cairnpap	2182	6.0	800
9	Scolty	1785	5.0	800
10	Traprain	2385	6.0	650
11	LairigGh	11560	28.0	2100
12	Dollar	2583	5.0	2000
13	Lomondso	3900	9.5	2200
14	CairnTab	2648	6.0	500
15	EildonTw	1616	4.5	1500
16	Cairngor	4335	10.0	3000
17	SevenHil	5905	14.0	2200
18	KnockHil	4719	3.0	350
19	BlackHil	1045	4.5	1000
20	CreagBea	1954	5.5	600
21	Kildoon	957	3.0	300
22	MeallAnt	1674	3.5	1500
23	HalfBenN	2859	6.0	2200
24	CowHill	1076	2.0	900
25	NorthBer	1121	3.0	600
26	CreagDub	1573	4.0	2000
27	Burnswar	2066	6.0	800
28	Largo	1714	5.0	950
29	Criffel	3030	6.5	1750
30	Achmony	1257	5.0	500
31	BenNevis	5135	10.0	4400
32	Knockfar	1943	6.0	600
33	TwoBrewe	10215	18.0	5200
34	Cocklero	1686	4.5	850
35	MoffatCh	9590	20.0	5000

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: Time

Number of Observations Read	35
Number of Observations Used	35

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	281686567	140843283	181.66	<.0001
Error	32	24810082	775315		
Corrected Total	34	306496649			

Root MSE	880.51977	R-Square	0.9191
Dependent Mean	3472.57143	Adj R-Sq	0.9140
Coeff Var	25.35642		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-539.48291	258.16073	-2.09	0.0447
Distance	1	373.07268	36.06841	10.34	<.0001
Climb	1	0.66289	0.12305	5.39	<.0001

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: Time

Output Statistics

Obs	Dependent Variable	Predicted Value	Std Error Mean Predict	Residual	Std Error Residual	Student Residual
1	965.0000	824.0760	204.1510	140.9240	856.5	0.165
2	2901	3356	195.8323	-455.1731	858.5	-0.530
3	2019	2296	172.5559	-276.5523	863.4	-0.320
4	2736	2789	193.8918	-52.8726	858.9	-0.0616
5	3736	4480	207.0087	-744.1646	855.8	-0.870
6	4393	4345	190.4951	48.0645	859.7	0.0559
7	12277	10401	570.9372	1876	670.3	2.798
8	2182	2229	178.3640	-47.2635	862.3	-0.0548
9	1785	1856	176.6706	-71.1909	862.6	-0.0825
10	2385	2130	188.2520	255.1697	860.2	0.297
11	11560	11299	731.3172	261.3832	490.4	0.533
12	2583	2652	183.5489	-68.6565	861.2	-0.0797
13	3900	4463	158.2947	-563.0611	866.2	-0.650
14	2648	2030	199.3622	617.6029	857.7	0.720
15	1616	2134	173.3782	-517.6761	863.3	-0.600
16	4335	5180	185.4588	-844.9078	860.8	-0.982
17	5905	6142	253.8882	-236.8881	843.1	-0.281
18	4719	811.7459	207.1660	3907	855.8	4.566
19	1045	1802	172.7751	-757.2321	863.4	-0.877
20	1954	1910	188.6629	43.8504	860.1	0.0510
21	957.0000	778.6015	209.4354	178.3985	855.2	0.209
22	1674	1761	193.4296	-86.6034	859.0	-0.101
23	2859	3157	175.6053	-298.3067	862.8	-0.346
24	1076	803.2616	212.8334	272.7384	854.4	0.319
25	1121	977.4679	198.3082	143.5321	857.9	0.167
26	1573	2279	206.4935	-705.5838	856.0	-0.824
27	2066	2229	178.3640	-163.2635	862.3	-0.189
28	1714	1956	170.6968	-241.6241	863.8	-0.280
29	3030	3046	152.3278	-15.5435	867.2	-0.0179
30	1257	1657	193.4086	-400.3245	859.0	-0.466
31	5135	6108	307.0252	-972.9510	825.3	-1.179
32	1943	2097	191.8296	-153.6859	859.4	-0.179
33	10215	9623	364.7359	592.1572	801.4	0.739
34	1686	1703	176.8190	-16.7989	862.6	-0.0195
35	9590	10236	384.8074	-646.4106	792.0	-0.816

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: Time

Output Statistics

Obs	-2-1 0 1 2	Cook's D	RStudent	Hat Diag H	Cov Ratio	DFFITS
1	| | |	0.001	0.1620	0.0538	1.1595	0.0386
2	*| |	0.005	-0.5242	0.0495	1.1269	-0.1196
3	| |	0.001	-0.3158	0.0384	1.1329	-0.0631
4	| |	0.000	-0.0606	0.0485	1.1556	-0.0137
5	*| |	0.015	-0.8661	0.0553	1.0837	-0.2095
6	| |	0.000	0.0550	0.0468	1.1536	0.0122
7	| |***** |	1.893	3.1688	0.4204	0.8179	2.6990
8	| |	0.000	-0.0540	0.0410	1.1467	-0.0112
9	| |	0.000	-0.0812	0.0403	1.1453	-0.0166
10	| |	0.001	0.2924	0.0457	1.1431	0.0640
11	| * |	0.211	0.5270	0.6898	3.4524	0.7858
12	| |	0.000	-0.0785	0.0435	1.1492	-0.0167
13	*| |	0.005	-0.6441	0.0323	1.0922	-0.1177
14	| * |	0.009	0.7146	0.0513	1.1039	0.1661
15	*| |	0.005	-0.5936	0.0388	1.1062	-0.1192
16	*| |	0.015	-0.9810	0.0444	1.0501	-0.2114
17	| |	0.002	-0.2769	0.0831	1.1908	-0.0834
18	| |******|	0.407	7.6110	0.0554	0.0493	1.8424
19	*| |	0.010	-0.8738	0.0385	1.0634	-0.1749
20	| |	0.000	0.0502	0.0459	1.1526	0.0110
21	| |	0.001	0.2054	0.0566	1.1611	0.0503
22	| |	0.000	-0.0992	0.0483	1.1546	-0.0223
23	| |	0.002	-0.3409	0.0398	1.1327	-0.0694
24	| |	0.002	0.3147	0.0584	1.1571	0.0784
25	| |	0.000	0.1647	0.0507	1.1557	0.0381
26	*| |	0.013	-0.8201	0.0550	1.0914	-0.1978
27	| |	0.001	-0.1865	0.0410	1.1431	-0.0386
28	| |	0.001	-0.2756	0.0376	1.1345	-0.0545
29	| |	0.000	-0.0176	0.0299	1.1338	-0.0031
30	| |	0.004	-0.4603	0.0482	1.1323	-0.1036
31	**| |	0.064	-1.1865	0.1216	1.0960	-0.4414
32	| |	0.001	-0.1761	0.0475	1.1513	-0.0393
33	| * |	0.038	0.7335	0.1716	1.2609	0.3338
34	| |	0.000	-0.0192	0.0403	1.1461	-0.0039
35	*| |	0.052	-0.8118	0.1910	1.2764	-0.3945

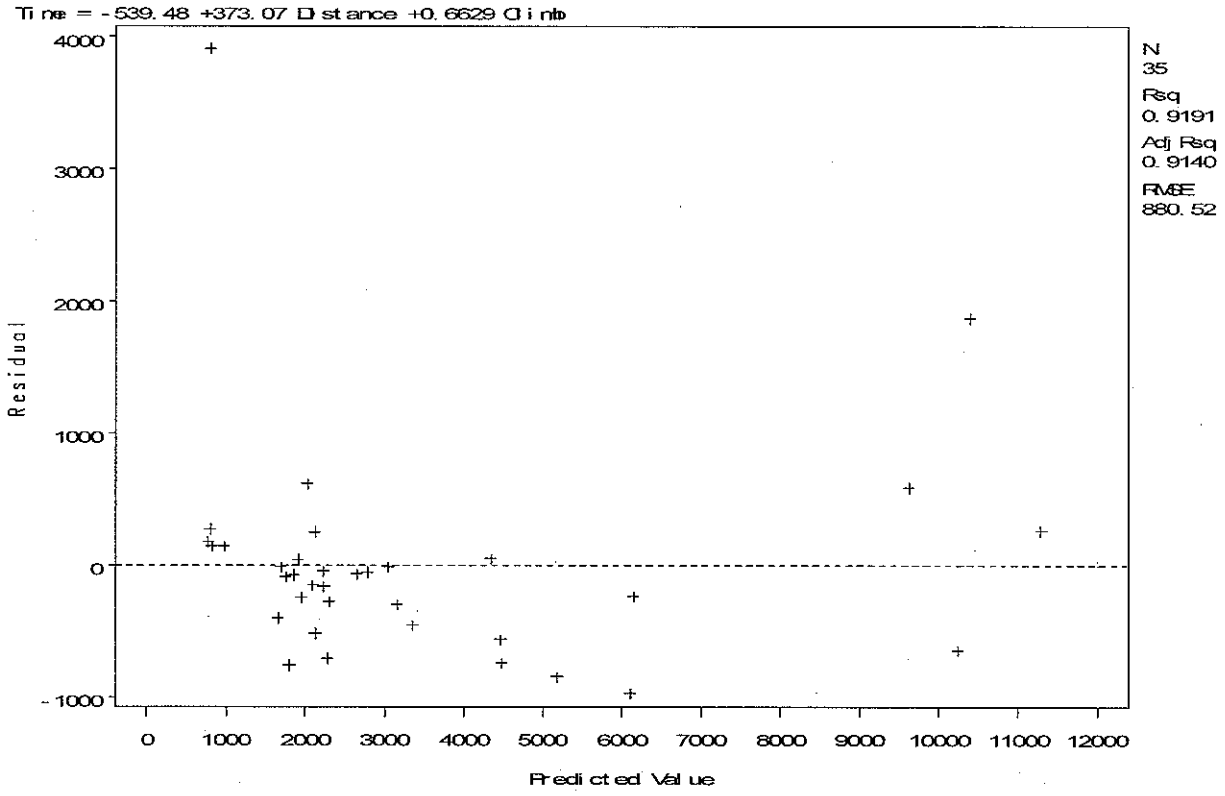
The SAS System

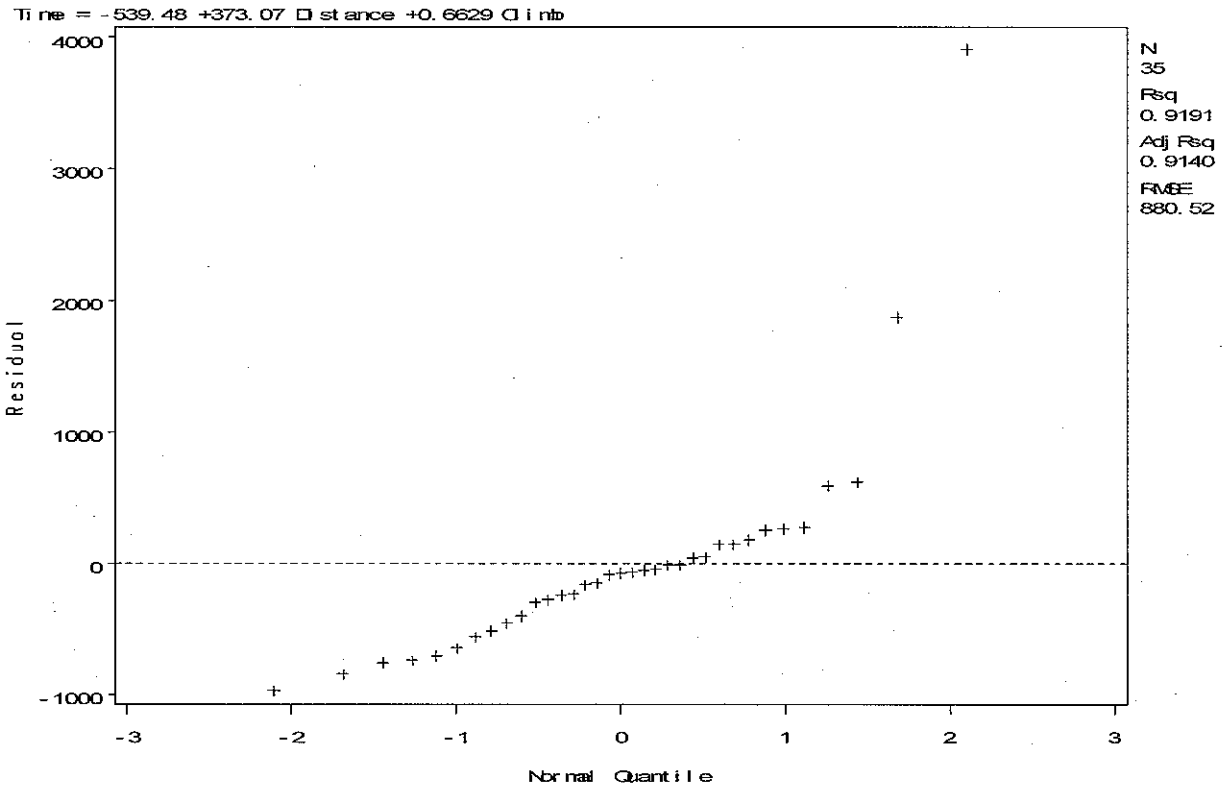
The REG Procedure
Model: MODEL1
Dependent Variable: Time

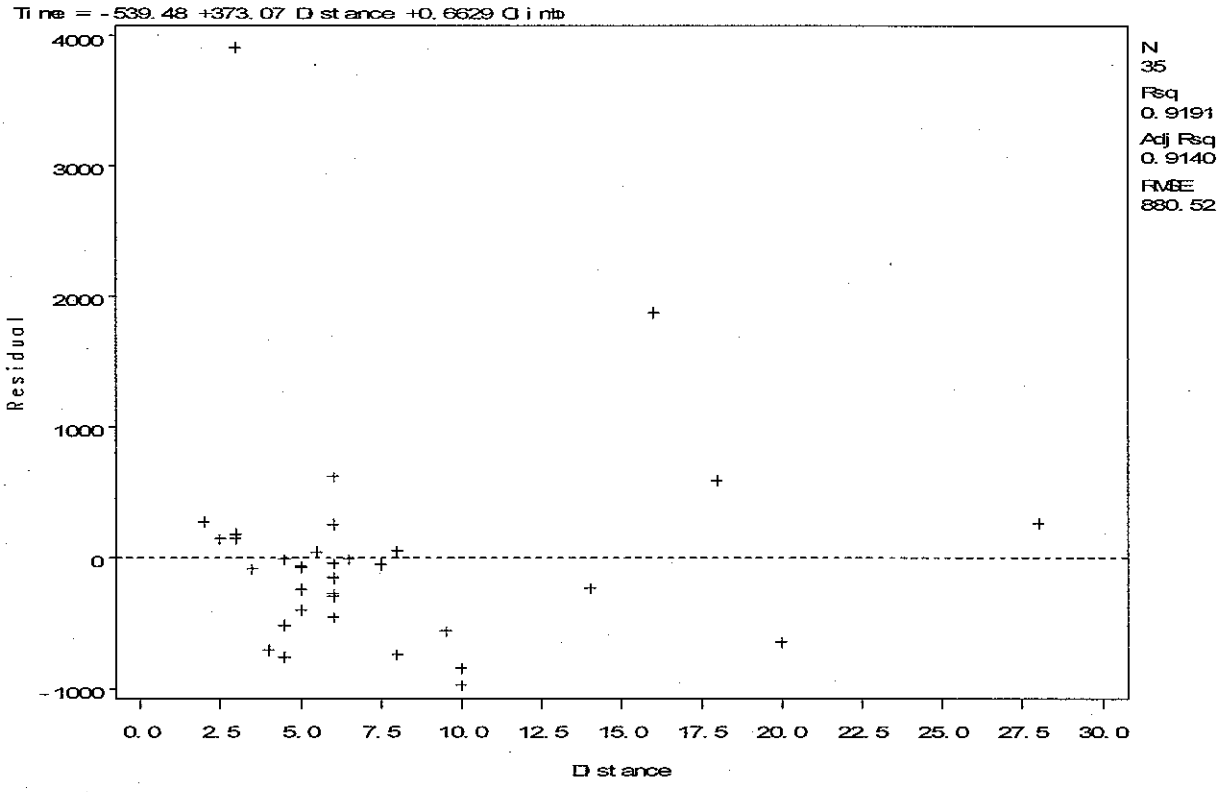
Output Statistics

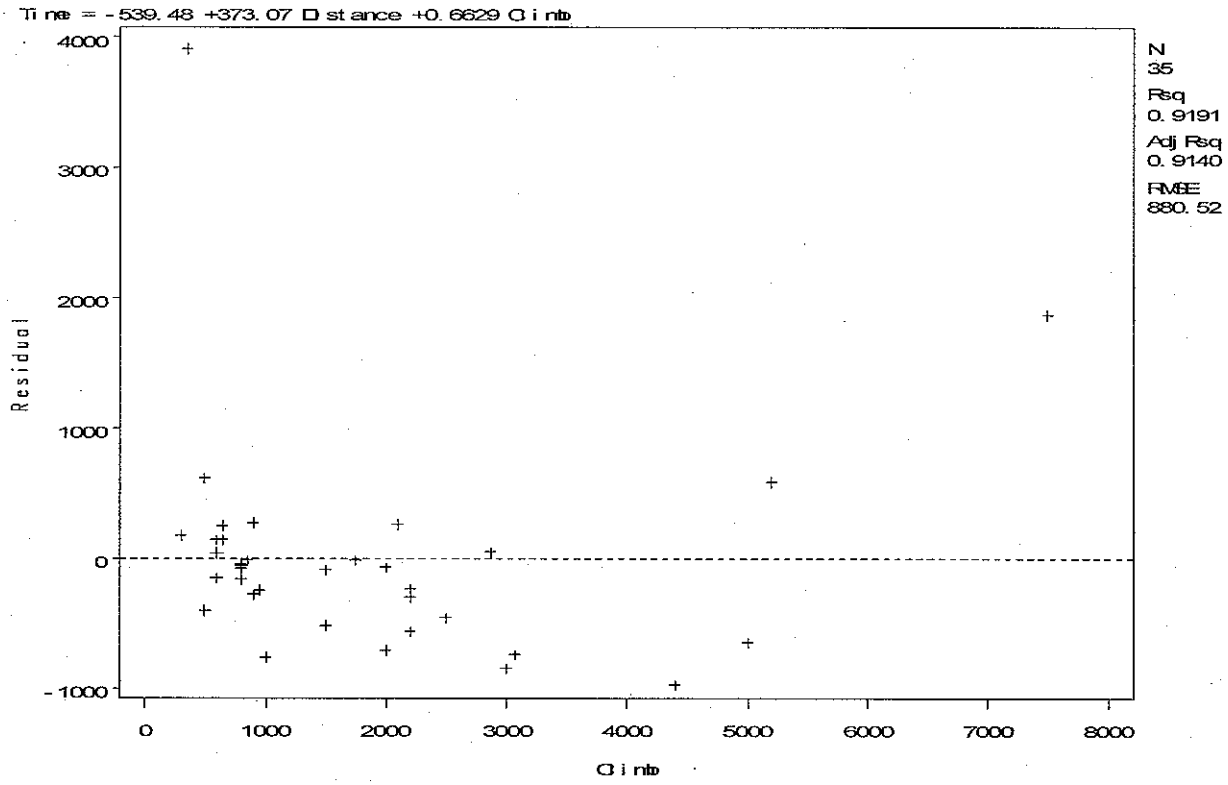
Obs	DFBETAS Intercept	Distance	Climb
1	0.0378	-0.0166	-0.0047
2	-0.0596	0.0672	-0.0734
3	-0.0486	-0.0067	0.0280
4	-0.0077	-0.0057	0.0088
5	-0.0505	0.0847	-0.1450
6	0.0035	-0.0043	0.0076
7	-0.8906	-0.7127	2.3645
8	-0.0085	-0.0016	0.0056
9	-0.0144	0.0009	0.0062
10	0.0470	0.0131	-0.0365
11	-0.3012	0.7689	-0.4799
12	-0.0115	0.0097	-0.0075
13	-0.0317	-0.0299	-0.0007
14	0.1180	0.0420	-0.1049
15	-0.1004	0.0577	-0.0223
16	-0.0185	0.0068	-0.0999
17	0.0120	-0.0665	0.0345
18	1.7583	-0.4066	-0.6559
19	-0.1589	0.0443	0.0294
20	0.0087	0.0014	-0.0059
21	0.0478	-0.0100	-0.0192
22	-0.0189	0.0139	-0.0065
23	-0.0412	0.0340	-0.0329
24	0.0748	-0.0464	0.0064
25	0.0369	-0.0126	-0.0083
26	-0.1377	0.1361	-0.1013
27	-0.0292	-0.0057	0.0192
28	-0.0477	0.0069	0.0150
29	-0.0021	0.0006	-0.0003
30	-0.0853	-0.0077	0.0548
31	0.0210	0.1701	-0.3737
32	-0.0286	-0.0087	0.0233
33	-0.1582	0.0970	0.1557
34	-0.0036	0.0007	0.0011
35	0.2087	-0.1990	-0.1009

Sum of Residuals	0
Sum of Squared Residuals	24810082
Predicted Residual SS (PRESS)	35336725


Time = -539.48 +373.07 Distance +0.6629 Climb
N
35
Rsq
0.9191
Adj Rsq
0.9140
RMSE
880.52
Residual
4000
3000
2000
1000
0
-1000
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
11000
12000
Predicted Value


Time = -539.48 +373.07 Distance +0.6629 Climb
Residual
4000
3000
2000
1000
0
-1000
-3
-2
-1
0
1
2
3
Normal Quantile
N
35
Rsq
0.9191
Adj Rsq
0.9140
RMSE
880.52


Time = -539.48 +373.07 Distance +0.6629 Climb
Residual
4000
3000
2000
1000
0
-1000
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
22.5
25.0
27.5
30.0
Distance
N
35
Rsq
0.9191
Adj Rsq
0.9140
RMSE
880.52


Time = -539.48 +373.07 Distance +0.6629 Climb
N
35
Rsq
0.9191
Adj Rsq
0.9140
RMSE
880.52
Residual
4000
3000
2000
1000
0
-1000
0
1000
2000
3000
4000
5000
6000
7000
8000
Climb