Exercise 1

The LOGISTIC Procedure

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	297.9576	28	<.0001
Score	250.7647	28	<.0001
Wald	176.4223	28	<.0001

Type 3 Analysis of Effects			
Effect	**DF**	**Wald Chi-Square**	**Pr > ChiSq**
AGE	1	0.2355	0.6275
SEX	1	18.9480	<.0001
RACE	3	2.3001	0.5125
RTRTN	3	33.3488	<.0001
SITE	19	38.3971	0.0053
BASE	1	154.6231	<.0001

Analysis of Maximum Likelihood Estimates						
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	25.0208	2.3390	114.4328	<.0001
AGE		1	0.00271	0.00559	0.2355	0.6275
SEX	**2**	1	-0.8352	0.1919	18.9480	<.0001
SEX	**1**	0	0	.	.	.
RACE	**White**	1	0.3972	0.6688	0.3527	0.5526
RACE	**Native A**	1	0.3113	0.8698	0.1281	0.7204
RACE	**Black**	1	0.6819	0.6863	0.9872	0.3204
RACE	**Asian**	0	0	.	.	.
RTRTN	**4**	1	1.6760	0.2945	32.3803	<.0001
RTRTN	**3**	1	1.1655	0.2840	16.8473	<.0001
RTRTN	**2**	1	0.9029	0.2842	10.0960	0.0015
RTRTN	**1**	0	0	.	.	.
SITE	**20**	1	2.0470	0.6852	8.9256	0.0028
SITE	**19**	1	2.1204	0.6782	9.7761	0.0018
SITE	**18**	1	0.7455	0.6790	1.2056	0.2722
SITE	**17**	1	2.3301	0.6722	12.0156	0.0005
SITE	**16**	1	1.9646	0.6637	8.7626	0.0031

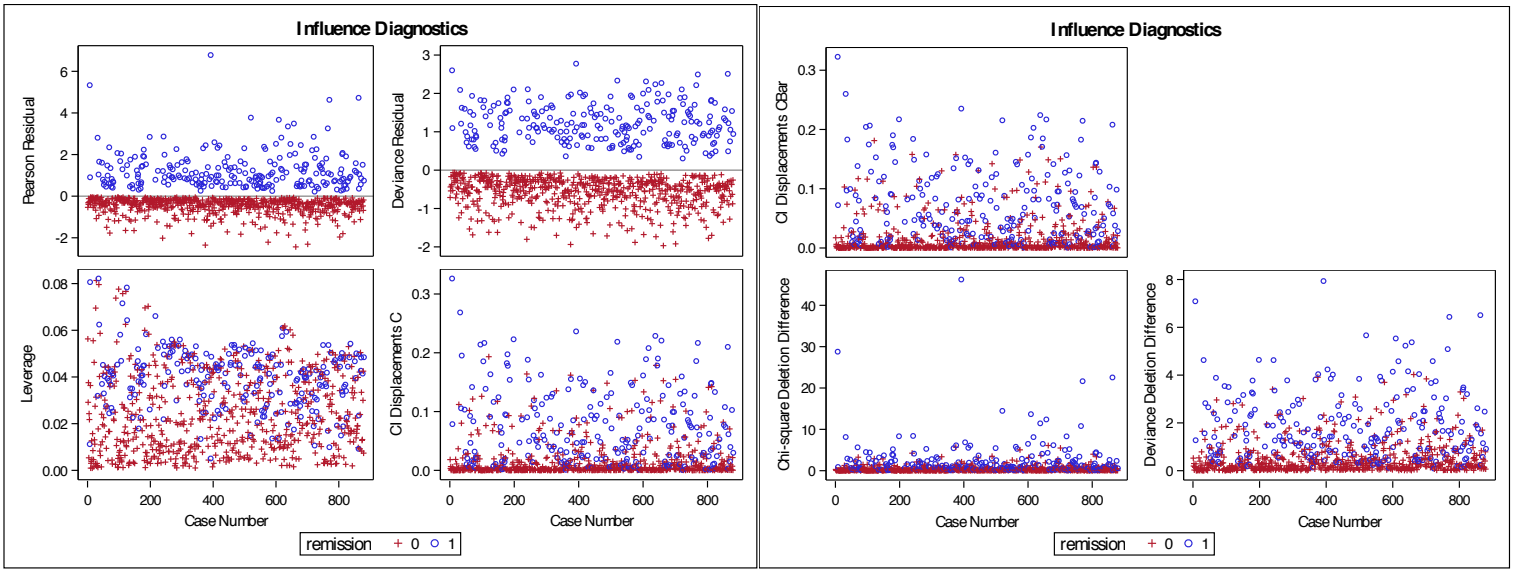
Analysis of Maximum Likelihood Estimates						
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
SITE	**15**	1	0.7474	0.7093	1.1105	0.2920
SITE	**14**	1	2.1244	0.6784	9.8077	0.0017
SITE	**13**	1	1.8108	0.6673	7.3634	0.0067
SITE	**12**	1	1.7822	0.6930	6.6139	0.0101
SITE	**11**	1	1.0318	0.6871	2.2550	0.1332
SITE	**10**	1	1.7760	0.6698	7.0312	0.0080
SITE	**9**	1	2.4802	0.6712	13.6528	0.0002
SITE	**8**	1	0.9804	0.6828	2.0620	0.1510
SITE	**7**	1	1.0368	0.6908	2.2528	0.1334
SITE	**6**	1	1.3508	0.6970	3.7561	0.0526
SITE	**5**	1	0.9784	0.7261	1.8157	0.1778
SITE	**4**	1	1.8307	0.6715	7.4334	0.0064
SITE	**3**	1	0.7855	0.7503	1.0961	0.2951
SITE	**2**	1	1.9156	0.6649	8.2996	0.0040
SITE	**1**	0	0	.	.	.
BASE		1	-0.1770	0.0142	154.6231	<.0001

a. The global test shows p-values less than 0.05 for Likelihood Ratio, Score and Wald test, so we conclude that there is at least one predictor whose coefficient is significantly different from zero. From the table with parameter estimates, we can see that predictors sex, rtrtn, site, and base are significant in the model with p-values less than 0.05. Even though indicator variables for some sites have p-value greater than 0.05, predictor site should be kept in the model because indicator variables for other sites are significant. Based on Type 3 analysis, we may want to retain sex, rtrtn, site and base in the model and remove age and race from the model.

The LOGISTIC Procedure

Summary of Stepwise Selection							
Step	Effect Entered	Effect Removed	DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq
1	BASE		1	1	185.5663		<.0001
2	RTRTN		3	2	28.3783		<.0001
3	SEX		1	3	20.4470		<.0001
4	SITE		19	4	41.4935		0.0021

b. Based on the result of stepwise selection, we choose predictors sex, rtrtn, site and base.



c. In diagnostic plots, there are no influential points. All absolute values of deviance residuals are around or less than 2 and all Cbar measures are less than 1. Thus we will keep all data. To test goodness of fit for a model we refer to the result from the Hosmer and Lemeshow test. The p- value is 0.52 which is greater than 0.05, thus we conclude that there is no lack of fit issue and our fitted model is adequate.

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
6.3552	8	0.6075

The p- value is 0.6075 which is greater than 0.05, thus we conclude that there is no lack of fit issue and our fitted model is adequate.

Analysis of Maximum Likelihood Estimates						
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	25.3938	2.2507	127.2934	<.0001
SEX	**2**	1	-0.8103	0.1906	18.0667	<.0001
SEX	**1**	0	0	.	.	.
RTRTN	**4**	1	1.6327	0.2910	31.4893	<.0001
RTRTN	**3**	1	1.1482	0.2834	16.4124	<.0001
RTRTN	**2**	1	0.8759	0.2832	9.5696	0.0020
RTRTN	**1**	0	0	.	.	.
SITE	**20**	1	2.0777	0.6813	9.2990	0.0023
SITE	**19**	1	2.1306	0.6774	9.8935	0.0017
SITE	**18**	1	0.7565	0.6755	1.2543	0.2627
SITE	**17**	1	2.3704	0.6690	12.5544	0.0004
SITE	**16**	1	1.9822	0.6620	8.9645	0.0028
SITE	**15**	1	0.7522	0.7080	1.1289	0.2880
SITE	**14**	1	2.0978	0.6764	9.6204	0.0019
SITE	**13**	1	1.8600	0.6654	7.8142	0.0052
SITE	**12**	1	1.7593	0.6925	6.4544	0.0111
SITE	**11**	1	1.1063	0.6837	2.6179	0.1057
SITE	**10**	1	1.7423	0.6674	6.8161	0.0090
SITE	**9**	1	2.5037	0.6700	13.9644	0.0002
SITE	**8**	1	0.9828	0.6805	2.0857	0.1487
SITE	**7**	1	1.0814	0.6898	2.4578	0.1169
SITE	**6**	1	1.3885	0.6931	4.0137	0.0451
SITE	**5**	1	0.9826	0.7269	1.8273	0.1764
SITE	**4**	1	1.8408	0.6697	7.5545	0.0060
SITE	**3**	1	0.8063	0.7483	1.1608	0.2813
SITE	**2**	1	1.9384	0.6621	8.5711	0.0034
SITE	**1**	0	0	.	.	.
BASE		1	-0.1757	0.0141	155.0772	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
SEX 2 vs 1	0.445	0.306	0.646
RTRTN 4 vs 1	5.118	2.893	9.052
RTRTN 3 vs 1	3.152	1.809	5.494
RTRTN 2 vs 1	2.401	1.378	4.182
SITE 20 vs 1	7.986	2.101	30.360
SITE 19 vs 1	8.420	2.232	31.763
SITE 18 vs 1	2.131	0.567	8.008
SITE 17 vs 1	10.701	2.884	39.708
SITE 16 vs 1	7.259	1.983	26.570
SITE 15 vs 1	2.122	0.530	8.498
SITE 14 vs 1	8.148	2.165	30.675
SITE 13 vs 1	6.424	1.743	23.669
SITE 12 vs 1	5.809	1.495	22.570
SITE 11 vs 1	3.023	0.792	11.547
SITE 10 vs 1	5.711	1.544	21.122
SITE 9 vs 1	12.227	3.289	45.459
SITE 8 vs 1	2.672	0.704	10.141
SITE 7 vs 1	2.949	0.763	11.396
SITE 6 vs 1	4.009	1.031	15.594
SITE 5 vs 1	2.671	0.643	11.103
SITE 4 vs 1	6.302	1.696	23.418
SITE 3 vs 1	2.240	0.517	9.709
SITE 2 vs 1	6.947	1.898	25.433
BASE	0.839	0.816	0.862

d. For sex variable, the expected decrease in log (remission) by changing sex=1 to sex=2 is -0.8103. Consequently, the expected count would decrease by a multiplicative factor of e^-0. 8103, which is the odds ratio 0.445. The decrease is significant. For rtrtn variable, the expected increase in log (remission) by changing rtrtn=1 to rtrtn=2 is 0.8759, to rtrtn=3 is 1.1482, and to rtrtn=4 is 1.6327. Consequently, the expected count would increase by a multiplicative factor of e^-0.8759(=odds ratio 2.401), e^1.1482(=odds ratio 3.152), and e^1.6327(=odds ratio 5.118). The increases are all significant. For variable base, the expected decrease in log(remission) for a one-unit increase in baseline is -0.1757, consequently the expected remission would decrease by a multiplicative factor of e^-0.1757, which is the odds ratio 0.839. The decrease is significant. The MLE for treatment 1 is equal to zero because treatment 1 is the reference. One unit increase in baseline will make odds of having value less than 120 decrease by a factor of 0.839. The odds ratio comparison between treatment 4 and treatment 1 is 5.118, which means that the odds of having value equal or less than 120

is 5.118 for treatment 4 compared to treatment 1. The 95% CI of sex 2 vs 1 is (0.306, 0.646), of treatment 4 vs 1 is (2.893, 9.052), of treatment 3 vs 1 is (1.809, 5.494), of treatment 2 vs 1 is (1.378, 4.182), and of base is (0.816, 0.862). These 95% CIs do not contain 1, which means the differences of remission between sex, between treatment group, and between baseline are significant. Treatment 4 is most effective, treatment 3 is second most effective, treatment 1 is the least effective.

The FREQ Procedure

Table of remission by _INTO_			
remission	**_INTO_(Formatted Value of the Predicted Response)**		
Frequency	**0**	**1**	**Total**
0	583	63	646
1	113	121	234
Total	696	184	880

e. The percent of the total number of observations that are misclassified: (63+113)/880=20%

Exercise 2
The LOGISTIC Procedure

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	122.4327	28	<.0001
Score	114.4530	28	<.0001
Wald	101.5228	28	<.0001

Type 3 Analysis of Effects			
Effect	**DF**	**Wald Chi-Square**	**Pr > ChiSq**
AGE	1	0.7340	0.3916
SEX	1	11.6230	0.0007
RACE	3	1.3847	0.7091
RTRTN	3	40.1682	<.0001
SITE	19	41.2339	0.0022
BASE	1	17.2208	<.0001

Analysis of Maximum Likelihood Estimates						
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	-7.6843	1.6406	21.9395	<.0001
AGE		1	-0.00385	0.00449	0.7340	0.3916
SEX	**2**	1	-0.5154	0.1512	11.6230	0.0007
SEX	**1**	0	0	.	.	.
RACE	**White**	1	-0.3803	0.5071	0.5625	0.4532
RACE	**Native A**	1	-0.3781	0.6608	0.3273	0.5672
RACE	**Black**	1	-0.2144	0.5206	0.1695	0.6805
RACE	**Asian**	0	0	.	.	.
RTRTN	**4**	1	1.3612	0.2188	38.7084	<.0001
RTRTN	**3**	1	0.8890	0.2176	16.6947	<.0001
RTRTN	**2**	1	0.6432	0.2203	8.5268	0.0035
RTRTN	**1**	0	0	.	.	.
SITE	**20**	1	0.9288	0.5217	3.1695	0.0750
SITE	**19**	1	1.4668	0.5159	8.0846	0.0045
SITE	**18**	1	0.1174	0.5718	0.0421	0.8373
SITE	**17**	1	1.5918	0.5129	9.6322	0.0019

Exercise 2
The LOGISTIC Procedure

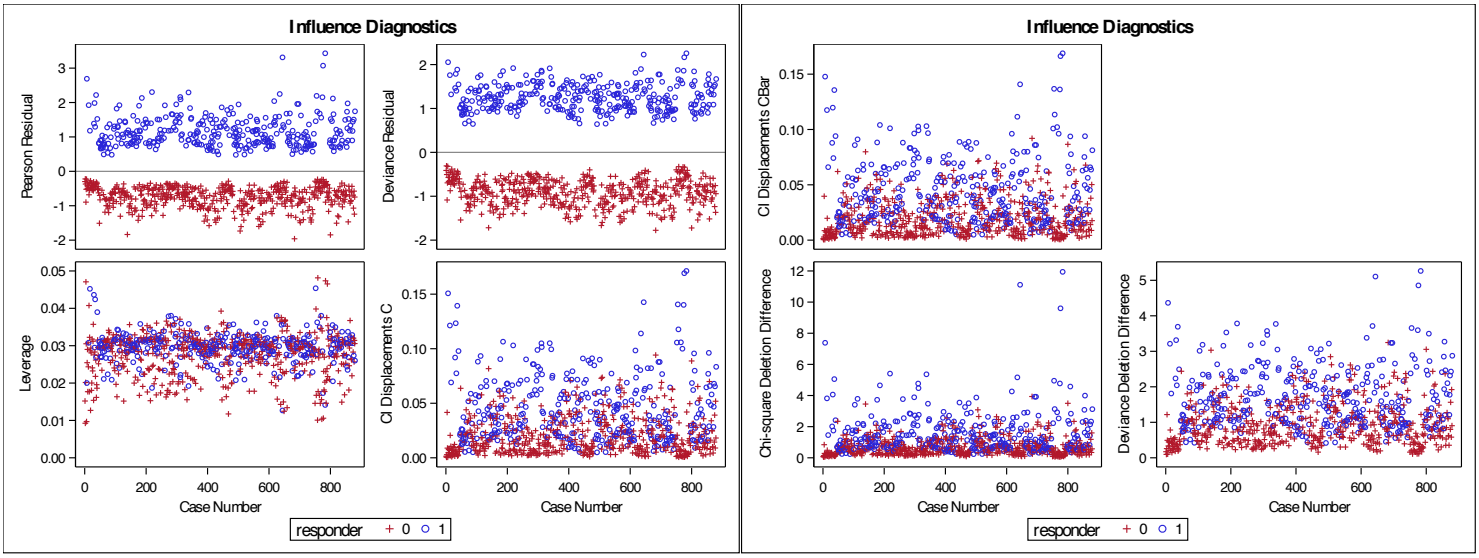
Analysis of Maximum Likelihood Estimates						
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
SITE	**16**	1	1.5714	0.5167	9.2481	0.0024
SITE	**15**	1	0.6604	0.5375	1.5100	0.2191
SITE	**14**	1	1.4217	0.5116	7.7229	0.0055
SITE	**13**	1	1.3207	0.5159	6.5547	0.0105
SITE	**12**	1	1.5934	0.5125	9.6664	0.0019
SITE	**11**	1	0.6382	0.5330	1.4334	0.2312
SITE	**10**	1	1.6329	0.5143	10.0797	0.0015
SITE	**9**	1	1.2136	0.5143	5.5681	0.0183
SITE	**8**	1	0.8350	0.5223	2.5556	0.1099
SITE	**7**	1	0.3722	0.5398	0.4752	0.4906
SITE	**6**	1	1.0088	0.5196	3.7687	0.0522
SITE	**5**	1	0.9086	0.5225	3.0234	0.0821
SITE	**4**	1	1.4957	0.5121	8.5292	0.0035
SITE	**3**	1	0.7840	0.5221	2.2545	0.1332
SITE	**2**	1	1.7822	0.5139	12.0241	0.0005
SITE	**1**	0	0	.	.	.
BASE		1	0.0372	0.00896	17.2208	<.0001

a. The global test shows p-values less than 0.05 for Likelihood Ratio, Score and Wald test, so we conclude that there is at least one predictor whose coefficient is significantly different from zero. From the table with parameter estimates, we can see that predictors sex, rtrtn, site, and base are significant in the model with p-values less than 0.05. Even though indicator variables for some sites have p-value greater than 0.05, predictor site should be kept in the model because indicator variables for other sites are significant. Based on Type 3 analysis, we may want to retain sex, rtrtn, site and base in the model and remove age and race from the model.

The LOGISTIC Procedure

Summary of Stepwise Selection							
Step	Effect Entered	Effect Removed	DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq
1	RTRTN		3	1	40.7963		<.0001
2	BASE		1	2	19.3036		<.0001
3	SEX		1	3	13.8281		0.0002
4	SITE		19	4	43.6922		0.0010

b. Based on the result of stepwise selection, we choose predictors sex, rtrtn, site and base.



c. In diagnostic plots, there are no influential points. All absolute values of deviance residuals are around or less than 2 and all Cbar measures are less than 1. Thus we will keep all data. To test goodness of fit for a model we refer to the result from the Hosmer and Lemeshow test. The p- value is 0.52 which is greater than 0.05, thus we conclude that there is no lack of fit issue and our fitted model is adequate.

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
8.7844	8	0.3608

The p- value is 0.3608 which is greater than 0.05, thus we conclude that there is no lack of fit issue and our fitted model is adequate.

Analysis of Maximum Likelihood Estimates						
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	-8.1463	1.5516	27.5653	<.0001
SEX	**2**	1	-0.5234	0.1505	12.0906	0.0005
SEX	**1**	0	0	.	.	.
RTRTN	**4**	1	1.3551	0.2177	38.7391	<.0001
RTRTN	**3**	1	0.8824	0.2173	16.4886	<.0001
RTRTN	**2**	1	0.6364	0.2198	8.3848	0.0038
RTRTN	**1**	0	0	.	.	.
SITE	**20**	1	0.9530	0.5195	3.3651	0.0666
SITE	**19**	1	1.4478	0.5153	7.8936	0.0050
SITE	**18**	1	0.1129	0.5704	0.0392	0.8430
SITE	**17**	1	1.5951	0.5121	9.7013	0.0018
SITE	**16**	1	1.5669	0.5148	9.2644	0.0023
SITE	**15**	1	0.6634	0.5347	1.5394	0.2147
SITE	**14**	1	1.4270	0.5108	7.8055	0.0052
SITE	**13**	1	1.3118	0.5148	6.4923	0.0108
SITE	**12**	1	1.6177	0.5113	10.0121	0.0016
SITE	**11**	1	0.6400	0.5324	1.4451	0.2293
SITE	**10**	1	1.6425	0.5126	10.2677	0.0014
SITE	**9**	1	1.2568	0.5125	6.0131	0.0142
SITE	**8**	1	0.8449	0.5216	2.6245	0.1052
SITE	**7**	1	0.4009	0.5381	0.5550	0.4563
SITE	**6**	1	1.0476	0.5171	4.1041	0.0428
SITE	**5**	1	0.9099	0.5213	3.0467	0.0809
SITE	**4**	1	1.5098	0.5109	8.7334	0.0031
SITE	**3**	1	0.7916	0.5211	2.3074	0.1288
SITE	**2**	1	1.7895	0.5129	12.1719	0.0005
SITE	**1**	0	0	.	.	.
BASE		1	0.0369	0.00894	17.0461	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
SEX 2 vs 1	0.593	0.441	0.796
RTRTN 4 vs 1	3.877	2.530	5.941
RTRTN 3 vs 1	2.417	1.579	3.700
RTRTN 2 vs 1	1.890	1.228	2.907
SITE 20 vs 1	2.593	0.937	7.179
SITE 19 vs 1	4.254	1.549	11.678
SITE 18 vs 1	1.120	0.366	3.424
SITE 17 vs 1	4.929	1.806	13.448
SITE 16 vs 1	4.792	1.747	13.142
SITE 15 vs 1	1.941	0.681	5.537
SITE 14 vs 1	4.166	1.531	11.337
SITE 13 vs 1	3.713	1.354	10.184
SITE 12 vs 1	5.041	1.851	13.732
SITE 11 vs 1	1.896	0.668	5.384
SITE 10 vs 1	5.168	1.892	14.115
SITE 9 vs 1	3.514	1.287	9.596
SITE 8 vs 1	2.328	0.838	6.470
SITE 7 vs 1	1.493	0.520	4.287
SITE 6 vs 1	2.851	1.035	7.854
SITE 5 vs 1	2.484	0.894	6.901
SITE 4 vs 1	4.526	1.663	12.319
SITE 3 vs 1	2.207	0.795	6.129
SITE 2 vs 1	5.987	2.191	16.360
BASE	1.038	1.020	1.056

d. For sex variable, the expected decrease in log (responder) by changing sex=1 to sex=2 is -0.5234. Consequently, the expected count would decrease by a multiplicative factor of e^-0. 5234, which is the odds ratio 0.593. The decrease is significant. For rtrtn variable, the expected increase in log (responder) by changing rtrtn=1 to rtrtn=2 is 0.6364, to rtrtn=3 is 0.8824, and to rtrtn=4 is 1.3551. Consequently, the expected count would increase by a multiplicative factor of e^0.6364 (=odds ratio 1.890), e^0.8824(=odds ratio 2.417), and e^1.3551(=odds ratio 3.877). The increases are all significant. For variable base, the expected increase in log(responder) for a one-unit increase in baseline is 0.0369, consequently the expected remission would increase by a multiplicative factor of e^0.0369, which is the odds ratio 1.038. The increase is significant. The MLE for treatment 1 is equal to zero because treatment 1 is the reference. One unit increase in baseline will make odds of having

chg less than -40 increase by a factor of 1.038. The odds ratio comparison between treatment 4 and treatment 1 is 3.877, which means that the odds of having chg equal or less than -40 is 3.877 for treatment 4 compared to treatment 1. The 95% CI of sex 2 vs 1 is (0.441, 0.796), of treatment 4 vs 1 is (2.530, 5.941), of treatment 3 vs 1 is (1.579, 3.700), of treatment 2 vs 1 is (1.228, 2.907), and of base is (1.020, 1.056). These 95% CIs do not contain 1, which means the differences of responder between sex, between treatment group, and between baseline are significant. Treatment 4 is most effective, treatment 3 is second most effective, treatment 1 is the least effective.

The FREQ Procedure

Table of responder by _INTO_			
responder	_INTO_(Formatted Value of the Predicted Response)		
Frequency	**0**	**1**	**Total**
0	436	101	537
1	190	153	343
Total	626	254	880

e. The percent of the total number of observations that are misclassified: (101+190)/880=33.07%.