Analysis of Olby's Chiari data, 3rd document

Jason A. Osborne, Feb 6, 2006

To investigate the dependence of severity on study factors, regression models were fit, with NEUROgrade and SHMgrade taken as continuous responses, even though they are really ordinal. This analysis is meaningful only if the GRADE measurement of severity is on a ratio scale, so that grade 2 is twice as severe as grade 1, and the difference between grades 1 and 2 is the same as the difference between 2 and 3. Otherwise, perhaps these models, while not conforming strictly to the usual assumptions, can at least provide some explanation of the variance in severity. Regression models were fit only to those dogs with NEURO grade exceeding 0, then again only for dogs with SHM grade exceeding 0. The study factors appear to explain a substantial amount of variability for the former. In the output below, models with low C(p) and high R-Square are preferred. More information about the estimation of models with an (*) follows the model selection output. The first model listed is well-determined (all coefficients are significant at level .05), and $r^2 = 57\%$ of the variation in NEUROgrade is explained by the model. The fitted model is

$$\text{NEURO grade} = 5.3 + 1.1 * (\text{Compression}) - 1.2 * (\text{Kinking} = 1) - 0.5 * \text{cdratio}$$

There is still a substantial amount of variability in NEUROgrade not explained by the model. It is interesting to note that marginally, only one of these three factors (CDRATIO) exhibits significant correlation (negative) with NEUROgrade given NEUROgrade > 0, but that other associations are revealed in the multiple linear regression.

For the SHMgrade, there are more observations, but less evidence of any associations. There is a borderline significant difference in SHMgrade between the sexes ($p = .0537$, mean grade is 2.25 for sex=1, 2.8 for sex=2). (I don't know which sex is M and which is F.)

Possible interactions weren't investigated. For example, if the probability of NEURO is increasing in CDFOSSA for low CDRATIO but decreasing in CDFOSSA for high CDRATIO, then the current models won't be able to describe such an interaction. However, there are too many ways to propose models with interactions, so I won't proceed without the suggestion of some candidate models to consider.

The REG Procedure
Model: MODEL1
Dependent Variable: neurograde

C(p) Selection Method

	Number in Model	C(p)	R-Square	Variables in Model
*	3	3.2919	0.5733	Compression kink1 cdratio
	4	3.3355	0.6740	Compression kink1 hydrocephalus cdratio
	4	3.3747	0.6719	sex1 Compression kink1 cdratio
	3	4.1442	0.5294	sex1 kink1 cdratio
*	5	4.2761	0.7285	sex1 Compression kink1 hydrocephalus cdratio
	5	4.6221	0.7107	Compression kink1 hydrocephalus cdfossa cdratio
	4	4.7797	0.5996	Compression kink1 Herniation cdratio
	5	5.0975	0.6862	Compression kink1 hydrocephalus Herniation cdratio
	4	5.2181	0.5771	Compression kink1 cdfossa cdratio
	5	5.2273	0.6795	sex1 Compression kink1 cdfossa cdratio
	5	5.3309	0.6742	sex1 Compression kink1 Herniation cdratio
			...	
	6	6.0228	0.7415	sex1 Compression kink1 hydrocephalus cdfossa cdratio

Best 5-predictors model for NEUROGRADE 5
The REG Procedure
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	9.97460	1.99492	3.76	0.0568
Error	7	3.71770	0.53110		
Corrected Total	12	13.69231			

Root MSE	0.72877	R-Square	0.7285
Dependent Mean	1.84615	Adj R-Sq	0.5345
Coeff Var	39.47485		

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Type II SS
Intercept	Intercept	1	8.55878	2.29810	3.72	0.0074	7.36652
sex1		1	0.56248	0.47444	1.19	0.2745	0.74647
Compression	K_Compression	1	0.90129	0.44736	2.01	0.0838	2.15568
kink1		1	-1.33160	0.50122	-2.66	0.0326	3.74853
hydrocephalus	S_Hdrocephalus	1	-0.31754	0.26302	-1.21	0.2665	0.77409
cdratio	W_Ratio_caud_total	1	-0.52105	0.17379	-3.00	0.0200	4.77391

The REG Procedure
Model: MODEL2
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	7.84957	2.61652	4.03	0.0451
Error	9	5.84274	0.64919		
Corrected Total	12	13.69231			

Root MSE	0.80573	R-Square	0.5733
Dependent Mean	1.84615	Adj R-Sq	0.4310
Coeff Var	43.64346		

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Type II SS
Intercept	Intercept	1	7.67505	2.38499	3.22	0.0105	6.72301
Compression	K_Compression	1	1.11536	0.45829	2.43	0.0377	3.84530
kink1		1	-1.19573	0.51686	-2.31	0.0460	3.47455
cdratio	W_Ratio_caud_total	1	-0.50183	0.18883	-2.66	0.0261	4.58507

The CORR Procedure

1 With Variables: neurograde
5 Variables: sex1 Compression kink1 hydrocephalus cdratio

Pearson Correlation Coefficients, N = 59
Prob > |r| under H0: Rho=0

	sex1	Compression	kink1	hydrocephalus	cdratio
neurograde	0.05951	0.09109	-0.13537	0.03138	-0.29600
G_Neuro_Grade	0.6543	0.4926	0.3067	0.8135	0.0228