BLOOD-BRAIN BARRIER EXPERIMENT

6. Tentative Multiple Linear Regression Model

The graphical tools (see Section 4) showed that the logarithmic transformation applied to both the response (concentration ratio) and the explanatory variable sacrifice time makes the relationship between the two variables approximately linear. In this section we will examine the relationship between the logarithm of concentration ratio (*LNRATIO*) and the eight explanatory variables with the following multiple regression model:

$$LNRATIO = \beta_0 + \beta_1 * LNTIME + \beta_2 * TREAT + \beta_3 * LNTIMTRE + \beta_4 * DAYS + \beta_5 * SEX + \beta_6 * WEIGHT + \beta_7 * LOSS + \beta_8 * TUMOR + ERROR.$$

We have included the log sacrifice time by treatment interaction term because the scatterplot in Section 4 suggested that the difference between the two treatments may be greater for the shorter sacrifice times than for the longer ones.

The random variable *ERROR* is assumed to follow a normal distribution with the mean of zero and an unknown standard deviation σ. The standard deviation is constant at all levels of the response variable *LNRATIO* under a range of settings of the eight independent variables *LNTIME*, *TREAT*, *LNTIMTRE*, *DAYS*, *SEX*, *WEIGHT*, *LOSS*, and *TUMOR*. Here, *LNRATIO* is the natural logarithm of the antibody concentration ratio, *LNTIME* is the natural logarithm of the sacrifice time, *LNTIMTRE* is the interaction term of *LNTIME* and *TREAT*.

The multiple linear regression model can be stated equivalently as follows:

$$\mu\{LNRATIO\} = \beta_0 + \beta_1 * LNTIME + \beta_2 * TREAT + \beta_3 * LNTIMTRE + \beta_4 * DAYS + \beta_5 * SEX + \beta_6 * WEIGHT + \beta_7 * LOSS + \beta_8 * TUMOR + ERROR.$$

The above model with the eight predictors is useful only if at least one slope β_i is different from zero. The hypothesis that the model is useful can be tested using F test.

The regression of the log antibody concentration ratio (*LNRATIO*) can now be done using the eight predictor variables. If the model explains a large portion of the variation in infant birth weight, it would be expected that at least some regression coefficients would not be significantly different from zero.

The following table displays the initial regression results for this data set.

MODEL SUMMARY

R	.971
R Square	.943
Adjusted R Square	.925
Standard Error	.6168

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	8	157.295	19.662
Residual	25	9.511	.380
Total	33	166.806	

F = 51.681 Signif F = .0000

The squared correlation coefficient R^2 (0.943) says that a significant portion (over 94 %) of the variation in antibody concentration ratio is explained by these eight predictors. The adjusted squared multiple correlation coefficient is 0.925

We analyze the ANOVA table associated with the multiple linear regression. The sum of squares due to the regression model is reported as 157.295, and the sum of squares due to error (residual sum of squares) is 9.511. The residual mean square is an estimate of the variance σ^2 and is equal to 0.380.

The value of the F statistic is equal to 51.681 with the corresponding p-value of 0 provides very strong evidence of the utility of the model.

Now we analyze the part of the output providing the estimates of the regression parameters.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-7.985	3.210		-2.488	.020
	LNTIME	1.005	.098	.868	10.244	.000
	TREAT	.862	.301	.195	2.868	.008
	LNTIMTRE	7.24E-03	.114	.005	.064	.950
	DAYS	.428	.272	.087	1.570	.129
	SEX	.417	.360	.080	1.158	.258
	WEIGHT	-1.1E-03	.005	-.014	-.214	.833
	LOSS	-3.4E-02	.030	-.065	-1.156	.259
	TUMOR	1.61E-03	.001	.074	1.218	.235

a. Dependent Variable: LNRATIO

According to the output, the estimated regression line of birth weights on the nine predictors is

$$\mu\{BWT\} = 1.005 * LNTIME - .862 * TREAT + .00724 * LNTIMTRE + .428 * DAYS + .417 * SEX$$
$$- .0011 * WEIGHT - .034 * LOSS + .00161 * TUMOR - 7.985.$$

The comparison of the eight independent variables by means of individual t-statistics indicates the relative magnitude of the unique contribution of each variable to the overall variability in the response. According to the above SPSS output, log sacrifice time is the largest contributor to the explained variation in log concentration ratio. The regression coefficient associated with log sacrifice time is 1.005 with a corresponding t ratio of 10.244, indicating a very strong effect of sacrifice time on the concentration ratio after accounting for the effect of the type of treatment and the covariates.

The treatment type is the next most important contributor. The remaining variables have significance probabilities greater than .12, and therefore can be considered nonsignificant contributors. Thus we can conclude that the design variables are significant when the covariates are also included in the model, and the covariates are not significant when the design variables are also included in the model.

The high p-value for the log sacrifice time by treatment interaction term indicates lack of evidence of significant interaction between the two variables. The final model used to estimate the treatment effect should therefore include only LNTIME and TREAT.