BLOOD-BRAIN BARRIER EXPERIMENT

7. Selecting a Regression Model

In this section we will use some variable selection techniques to obtain a new regression model for predicting antibody concentration ratio. These techniques include backward elimination procedure, forward regression, and stepwise regression. The best subsets method is not supported by SPSS. The techniques are based on adding independent variables (one at a time) to a regression model or removing independent variables (one at a time) from the model.

Forward variable selection enters the variables into the model one at a time based on entry criteria. At each step, the hypothesis that the coefficient of the entered variable is 0 is tested using its t statistic (actually an F statistic that is the square of the t). *Backward* variable elimination begins with all independent variables in the model, and at each step, removes the least useful predictor. Variables are removed until an established criterion holds. *Stepwise* selection begins like forward method, but at each step, tests variables already in the model for removal.

SPSS provides two criteria for moving variables. They are based on an F statistic that is the square of the t statistic. The first criterion for removing variables is the minimum F value that a variable must have to remain in the model. Variables with F statistics less than the value specified for removal are eligible for removal. Some texts and software packages call this statistic *F-to-remove*. The second criterion is the *maximum probability of F-to-remove*. The default F-to-remove is 2.71, and the default probability is 0.10.

We start with the stepwise selection method applied to our data. The stepwise regression applied to our data can be summarized in the following table:

STEPWISE SELECTION METHOD					
MODEL	**VARIABLES ENTERED**	**VARIABLES REMOVED**	**R^2**	**ADJ. R^2**	**ST. ERROR**
1	LNTIME		.890	.886	.7584
2	TREAT		.926	.921	.6307

The criterion *Probability-of-F-to-enter <= .050* is used to enter a variable into the model, the condition *Probability-of-F-to-remove>= .100* is used to remove a variable from the model. SPSS also provides the ANOVA table for each model.

ANOVA[c]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	148.400	1	148.400	258.000	.000[a]
	Residual	18.406	32	.575		
	Total	166.806	33			
2	Regression	154.475	2	77.238	194.187	.000[b]
	Residual	12.330	31	.398		
	Total	166.806	33			

a. Predictors: (Constant), LNTIME

b. Predictors: (Constant), LNTIME, TREAT

c. Dependent Variable: LNRATIO

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-3.421	.181		-18.853	.000
	LNTIME	1.092	.068	.943	16.062	.000
2	(Constant)	-3.855	.187		-20.574	.000
	LNTIME	1.098	.057	.948	19.416	.000
	TREAT	.846	.216	.191	3.908	.000

a. Dependent Variable: LNRATIO

You can check that forward selection method applied to our data is here equivalent to the stepwise selection method. For the forward selection method, the criterion *Probability-of-F-to-enter <= .050* is used to enter a variable into the model.

The backward elimination method applied to our set of independent variables can be summarized in the following table:

BACKWARD ELIMINATION METHOD					
MODEL	**VARIABLES REMOVED**	**VARIABLES IN THE MODEL**	**R^2**	**ADJ. R^2**	**ST. ERROR**
1		LNTIME, TREAT, DAYS, SEX, WEIGHT, LOSS, TUMOR	.943	.928	.6049
2	WEIGHT	LNTIME, TREAT, DAYS, SEX, LOSS, TUMOR	.943	.930	.5942

BACKWARD ELIMINATION METHOD					
MODEL	**VARIABLES REMOVED**	**VARIABLES IN THE MODEL**	**R^2**	**ADJ. R^2**	**ST. ERROR**
3	SEX	LNTIME, TREAT, DAYS, LOSS, TUMOR	.940	.929	.5996
4	DAYS	LNTIME, TREAT, LOSS, TUMOR	.935	.926	.6104
5	LOSS	LNTIME, TREAT, TUMOR	.929	.922	.6263
6	TUMOR	LNTIME, TREAT	.926	.921	.6307

As the criterion to remove a variable, we used the probability of F to be at least 0.100. SPSS also provides the analysis of the coefficients for each of the six regression models:

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	-7.972	3.142		-2.538	.017
	LNTIME	1.009	.074	.872	13.691	.000
	TREAT	.875	.218	.198	4.021	.000
	DAYS	.427	.267	.087	1.600	.122
	SEX	.419	.352	.080	1.190	.245
	WEIGHT	-1.15E-03	.005	-.014	-.227	.822
	LOSS	-3.40E-02	.029	-.064	-1.178	.249
	TUMOR	1.596E-03	.001	.073	1.254	.221
2	(Constant)	-8.395	2.483		-3.381	.002
	LNTIME	1.005	.070	.868	14.298	.000
	TREAT	.872	.213	.197	4.088	.000
	DAYS	.443	.253	.091	1.753	.091
	SEX	.385	.313	.074	1.231	.229
	LOSS	-3.52E-02	.028	-.066	-1.261	.218
	TUMOR	1.620E-03	.001	.075	1.300	.205
3	(Constant)	-7.529	2.403		-3.133	.004
	LNTIME	1.054	.058	.911	18.079	.000
	TREAT	.903	.214	.204	4.229	.000
	DAYS	.348	.243	.071	1.434	.163
	LOSS	-4.10E-02	.028	-.077	-1.477	.151
	TUMOR	2.097E-03	.001	.096	1.754	.090
4	(Constant)	-4.104	.272		-15.079	.000
	LNTIME	1.081	.056	.934	19.252	.000
	TREAT	.929	.217	.210	4.290	.000
	LOSS	-4.52E-02	.028	-.085	-1.607	.119
	TUMOR	2.200E-03	.001	.101	1.812	.080
5	(Constant)	-4.104	.279		-14.697	.000
	LNTIME	1.082	.058	.935	18.780	.000
	TREAT	.913	.222	.206	4.110	.000
	TUMOR	1.338E-03	.001	.062	1.197	.241
6	(Constant)	-3.855	.187		-20.574	.000
	LNTIME	1.098	.057	.948	19.416	.000
	TREAT	.846	.216	.191	3.908	.000

a. Dependent Variable: LNRATIO

The backward elimination method has produced the same regression model with the two independent variables LNTIME, and TREAT. The two predictors in the model are significant with the p-value reported as zero.

The regression model obtained via the three variable selection techniques has the form:

$LNRATIO = \beta_0 + \beta_1 * LNTIME + \beta_2 * TREAT + ERROR.$

The SPSS output for the model is displayed below:

MULTIPLE LINEAR REGRESSION

Multiple R	.962
R Square	.926
Adjusted R Square	.921
Standard Error	.6307

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	2	154.475	77.238
Residual	31	12.330	.3980

F = 194.187 Signif F = .0000

According to the above output, the estimated regression line of log concentration ratio on the two predictors is

$\mu\{LNRATIO\} = 1.098 * LNTIME + .846 * TREAT - 3.855.$

From the above equation, we obtain that

$\mu\{LNRATIO \mid BD\} - \mu\{LNRATIO \mid NS\} = .846.$

Under the assumption, that LNRATIO follows approximately a symmetric distribution, μ{ln(RATIO)}=Median{ln(RATIO)}=ln{Median(RATIO)}.

Using the above equalities, we have

ln{Median(RATIO|BD)} - ln{Median(RATIO|NS)} = 0.846. Therefore,

$$\ln\{\frac{Median(RATIO \mid BD)}{Median(RATIO \mid NS)}\} = 0.846.$$

Thus, the median ratio of antibody concentration in the brain tumor to antibody concentration in the liver is estimated to be exp(0.846) = 2.33 times greater for the blood-brain barrier diffusion treatment than for the saline control.