MULTIPLE REGRESSION

The human brain is protected from bacteria and toxins, which course through the bloodstream, by a single layer of cells called the *blood–brain barrier*. This barrier normally allows only a few substances, including some medications, to reach the brain. Because chemicals used to treat brain cancer have such large molecular size, they cannot pass through the barrier to attack tumor cells. At the Oregon Health Sciences University, Dr. E. A. Neuwelt developed a method of disrupting the barrier by infusing a solution of concentrated sugars. As a test of the disruption mechanism, researchers conducted a study on rats, which possess a similar barrier. (Data from P. Barnett et al., “Differential Permeability and Quantitative MR Imaging of a Human Lung Carcinoma Brain Xenograft in the Nude Rat,” *American Journal of Pathology* 146(2) (1995): 436–49.) The rats were inoculated with human lung cancer cells to induce brain tumors. After 9 to 11 days, they were infused with either the barrier disruption (BD) solution or, as a control, a normal saline (NS) solution. Fifteen minutes later, the rats received a standard dose of the therapeutic antibody L6-F(ab0)2. After a set time they were sacrificed, and the amounts of antibody in the brain tumor and in normal tissue were measured. The timeline for the experiment is shown in Figure 1. Measurements for the 4 rats (out of 34 rats) are listed in Figure 2.

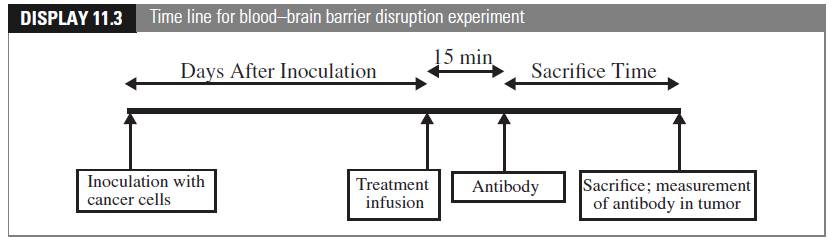


Figure 1: The timeline for the experiment

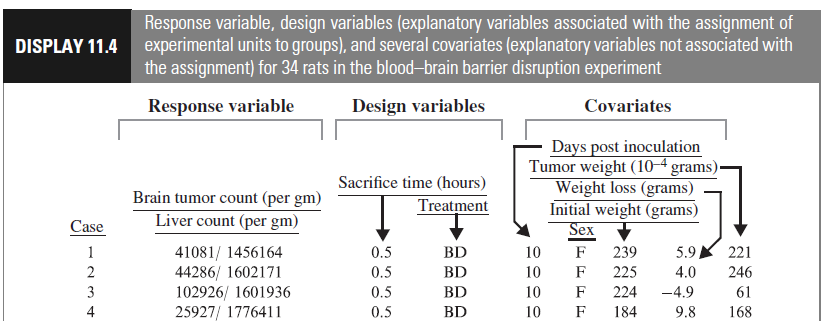


Figure 2: Measures for 4 rats

Since the amount of the antibody in normal tissue indicates how much of it the rat actually received, a key measure of the effectiveness of transmission across the blood–brain barrier is the ratio of the antibody concentration in the brain tumor to the antibody concentration in normal tissue outside of the brain. The brain tumor concentration divided by the liver concentration is a measure of the amount of the antibody that reached the brain relative to the amount of it that reached other parts of the body. This is the response variable: both the numerator and denominator of this ratio are listed in Figure 2. The explanatory variables in the table comprise two categories: *design variables* are those that describe manipulation by the researcher; *covariates* are those measuring characteristics of the subjects that were not controllable by the researcher.

1. **Important libraries:**

*library(Sleuth2)*

*library(corrplot)*

*library(car) # to do some important stat calculation*

*library(olsrr) # to do variable selections*

1. **Correlation plot:**

To do correlation plot, use

*A=cor (YourDataName)*

*corrplot(A)*

To do pairwise correlation plot, use:

*pairs(YourDataName)*

Example 1: Load the data from the library Sleuth2. The name of the data is case1102. Compute the log of the ratio of brain tumor antibody count to liver antibody count. Then, compute the correlation matrix for the log ratio, days after inoculation, tumor weight, weight loss, and initial weight. These are the numerical variables of your data. Then plot the correlation using two methods above. What do you notice from the plots?

Hint: You may need to put these variables together in a new data frame and find its correlation.

1. **Multiple regression:**

To compute a multiple regression model in R, use:

*model= lm(yData ~ xData\_1+xData\_2+….. , the name of you data)*

or if you plan to include all variables in your design matrix (full model)

*model= lm(yData ~. , the name of you data)*

To compute variance inflation factor, use:

*vif(model)*

*mean(vif(model))*

Example 2: Construct a multiple regression model that use days after inoculation, tumor weight, weight loss, and initial weight to predict the log ratio. Output and interpret the model: write down its formula, interpret each parameter (what it means and if it is significant or not). State the hypothesis and the result of the F-statistics in the output.

Example 2 (Cont’d): Compute the variance inflation factor for each variable and the model. Is this concerning?

1. **Variable selection:**

To do variable selection, you can use:

*ols\_step\_forward\_p(model)*

*ols\_step\_backward\_p(model)*

*ols\_step\_both\_p(model)*

Here is a typical output from forward selection:

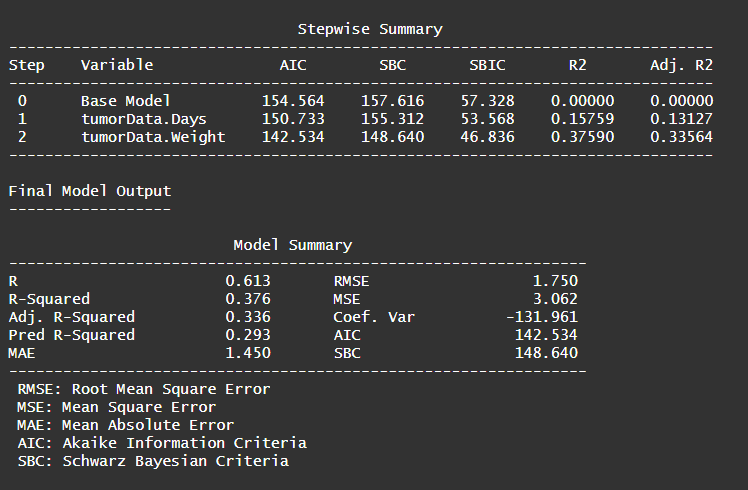


Figure 3: First half of the output of forward selection. It also provides some of the criteria-based values such as AIC, adjusted , and soon.

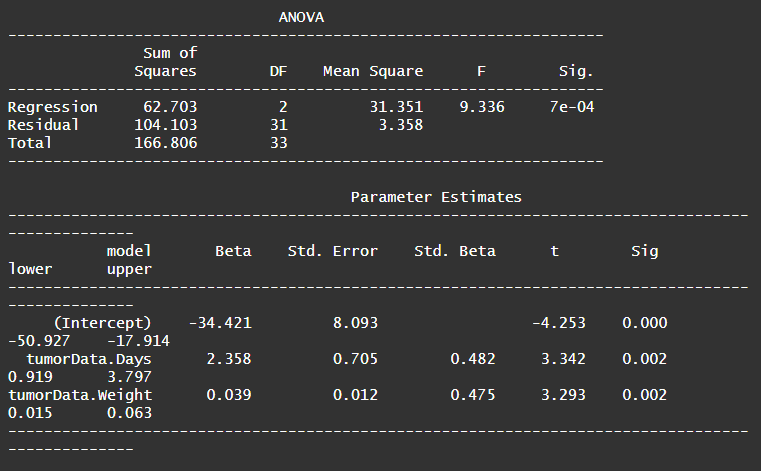


Figure 4: This is the second half of the output from forward selection with the selected model, the inference steps (t test for one predictor and its confidence interval; look at slides 18-20 of the lecture note 08\_Inference with Linear Regression\_01), and the ANOVA table. This ANOVA table is the F test of all predictors (look at slides 13-16 of the lecture note 08\_Inference with Linear Regression\_01)

Example 3: Conduction variable selection using forward, backward, and stepwise methods. What is the output from each method? Arrive at your choice for your model.

1. **Add categorical variables:**

From the model you obtain in example 3, we will now add some categorical variable. In R, you can just add them into the model.

Example 4: Add the variable sex into the model that you obtained in example 3. Interpret the result. Use both commands below to get the output as Figure 3

*summary(name of your model)*

*Anova(name of your model*

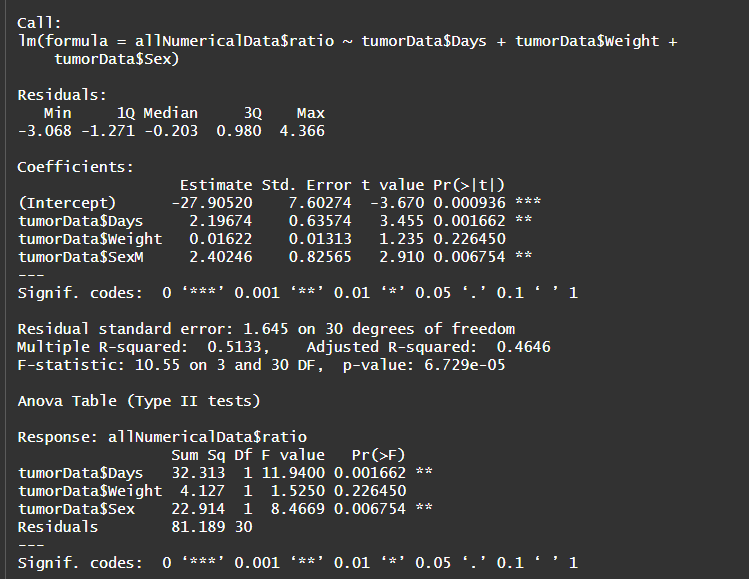


Figure 5: Output for example 4. Here you see that one of the coefficients is called tumorData$SexM. This is to interpret the contrast between F and M. This means being a M is (significantly) associated with an average increase of 2.40246 in the log ratio as compared to being a F. The Anova table tells you if each parameter (Day, Weight, and Sex) is significant or not. Why do you think the Weight parameter is not significant here anymore even though it was significant in Example 3?

Example 4: There are two treatments: either the barrier disruption (BD) solution or, as a control, a normal saline (NS) solution. From the model in example 3, include the Treatment into your model as well. Interpret the result.

Example 5: From the model in example 3, include both sex and treatment into the model. Interpret your result. Check the VIF of the model.

Example 6: Do a variable selection (3 methods) for the model in Example 5. What is the output from each method? Arrive at your choice for your model.

1. **Add interaction term**

Sometimes, there are interaction between the variables, and we want to take that into account for the model as well. To add interaction term between categorical values , the interaction term is coded as

XData\_1: XData2

Example 7: From the model that you choose in example 6, add the interaction between sex and treatment. Interpret your result. Use both commands below to see the output.

*summary(name of your model)*

*Anova(name of your model*

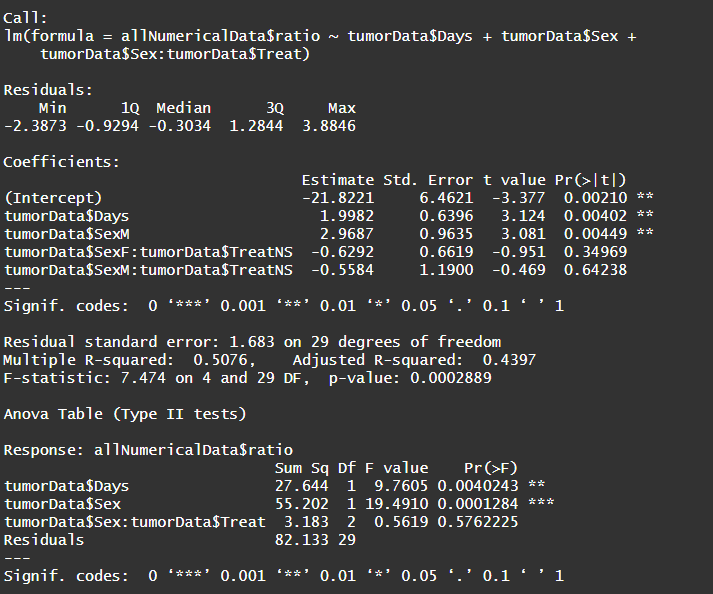


Figure 6: A typical output when you have interaction term between categorical variables.