MULTIPLE REGRESSION

1. **Add categorical variables:**

Example 1: Use the same data as the previous handout (tumor data). Fit the model using Days and Sex. Then look at the output and interpret the model.

*summary(name of your model)*

*Anova(name of your model*

*contrasts(name of your categorical varable)* # this is to see how R codes the categorical variable

*model.matrix(name of your model)* # this is to see the design matrix in R

Example 2: Plot the result of the model above for each sex. First, you need to plot your data with ratio in the y axis and Days in the x axis. You can add this in the plot command *pch=as.numeric(Name of your Categorical variable)* to have different marker type for each sex.

Then, add the line for female using *abline (the value of the intercept, the value of the slope)*

Then, add the line for male using *abline(the value of the intercept, the value of the slope).* Make sure to have different color for the male line.

To access the intercept, use Name of your *model$coefficients[1]*

To access the other slope values, use Name of your *model$coefficients[i]* where i is the index of the predictor.

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 3: From the model above, 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*

*contrasts(name of your categorical varable)* # this is to see how R codes the categorical variable

*model.matrix(name of your model)* # this is to see the design matrix in R

1. **Add categorical predictor with more than 2 levels**

Example 4: Fit the data using Days and Time. Note that the Time variable in R currently is a numerical variable. First, we want to treat this as factor using *factor(name of your variable)*. Then, fit your model, interpret the result, and plot the regression line for each Time level.