Logistic Regression

Model and Description

Next, we will perform logistic regression with the remission data set. We are using logistic regression because our dependent variable is binary. With logistic regression, you are usually using your depend variables to predict odds. In using the remission data set, it appears we are using the data to predict odds of going back into remission, given some health factors. Below is the code showing that we made "remiss"(our dependent variables) a factor, and our model:

Model 1

```
`remission.(1)`$remiss<-factor(`remission.(1)`$remiss)
```

```
Deviance Residuals:
     Min        1Q    Median        3Q       Max
-1.95165  -0.66491  -0.04372   0.74304   1.67069

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  58.0385    71.2364   0.815   0.4152
cell         24.6615    47.8377   0.516   0.6062
smear        19.2936    57.9500   0.333   0.7392
infil       -19.6013    61.6815  -0.318   0.7507
li            3.8960     2.3371   1.667   0.0955 .
blast         0.1511     2.2786   0.066   0.9471
temp        -87.4339    67.5735  -1.294   0.1957
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 34.372  on 26  degrees of freedom
Residual deviance: 21.751  on 20  degrees of freedom
AIC: 35.751

Number of Fisher Scoring iterations: 8
```

As shown above, all of our independent variables do not pass the t test, given an alpha of .05. It can be interpreted that given the log odds of "li", our results change by the estimate of 3.8960. The same aforementioned logic with "li" can be used with the other independent variables as well, although we probably wouldn't use their estimation because they do not pass the t-test. The model will be looked further into, to fix the muddy t-tests.

glm Function vs. lm function

The difference between glm and lm, is that glm is used to form generalized linear models that have symbolic (binary) descriptions in their y axis, while lm is used to form linear models that can be used to for multiple forms of regression, excluding logistic. In other words, glm is used to show the binary response variables, and lm is used for standard regression models.

Model 11

Final Logistic Model

```
Call:
glm(formula = remiss ~ li, family = "binomial", data = `remission

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.9448  -0.6465  -0.4947   0.6571   1.6971

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   -3.777      1.379  -2.740  0.00615 **
li             2.897      1.187   2.441  0.01464 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 34.372  on 26  degrees of freedom
Residual deviance: 26.073  on 25  degrees of freedom
AIC: 30.073

Number of Fisher Scoring iterations: 4
```

After removing the independent variables from highest p-value to lowest, we ended up with one independent variable that passed the t-test. Our Final model contains “li” with a p-value of .01464. It can be interpreted that given the log odds of “li”, our results change by the estimate of 2.897. We can assume we would use this estimation because now it passes the t-tests.