Problem 1

iml02.sas

The variable group identies what treatment group each child is in, either the succimer (A) or the placebo group (P)

The variable lead0 is the child's baseline lead level in micrograms/dL. The variables lead1,lead4, and lead6 are the child's lead level after 1, 4, and 6 weeks in the study.

A variable y1 is created with all the lead readings for the succimer group.

A variable y0 is created with all the lead readings for the placebo group.

The sample sizes are 50 children for each group.

Two sample mean vectors are created, ybar0 and ybar1 for the placebo group and succimer group respectively. These vectors contain the treatment group mean lead level in micrograms/dL for each week. The sample mean for the succimer group drops by almost 50% from baseline to the reading after the first week. It then increases over the duration of the study. The lead levels in the placebo group drop slightly at each reading over the course of the study.

A sample covariance matrix is created for each treatment group s0 and s1. Two sample standard deviation vectors are created using the squareroot of the diagonal of the particular treatment groups sample covariance matrix. They are sd0 and sd1 for the placebo group and succimer group respectively. These vectors contain the treatment group standard deviation for each week. The standard deviation of the placebo group does not change dramatically over the readings while the standard deviations of the succimer group increase by over 50% from baseline to week one and continue to increase over the course of the study.

The group sample sizes, sample mean vectors, sample standard deviation vectors and sample covariance matrices are printed for each group.

iml03.sas

Variables for the sample size $n = 50$ and the standard deviation $\sigma = 2$ are created

A column vector β of 3 ones is created.
Three column vectors of length 50 are created: x_1, a the intercept which is all ones, x_2 a column of $\frac{1}{50}$ through $\frac{50}{50}$, and x_3, a column of $0, 1, 0, 1, \ldots$. An X matrix is created out of the three column vectors.
A 50 by 1 matrix y is created where $y = X\beta + \sigma * \text{random number vector}$ The random number vector is of length 50.

A matrix xy is which consists of X with the column y at the end. The first column is the intercept, the second column is the dose, the 3rd column is the group and the 4th column is y. A data set A is then created from the matrix xy

The parameter estimates $\hat{\beta}$ are calculated by taking the inverse of $X^T X$ and matrix multiplying it by $X^T y$

A linear regression is then run on the data, and an Anova table is printed in the output along with a table of Parameter estimates, with degrees of freedom, standard error, a t statistic, and a p-value. The parameter estimates in this table match up to those calculated using matrix operations which is to be expected. A table containing the square root of the mean square error, R^2 and Adjusted R^2, the dependent mean and the coefficient of variation is also printed.

Fit diagnostic graphs are plotted. They include: the residuals plotted against the fitted values, the studentized residuals against the fitted values, the studentized residuals against the leverage, the residuals against the quantiles, the y values against the fitted values, cooks distance is plotted for each observation, the residuals are fit with a normal distribution curve and a histogram. A side by side plot of the residuals by dose and by group is also printed.

Problem 11.2

1. The person time at risk is 2 weeks for everyone in our model

   Fitting a log linear regression model using the placebo group as the reference group and assuming a Poisson distribution for the counts of epileptic seizures.

   $\hat{\beta}_1 = 2.0794 \quad se(\hat{\beta}_1) = .0668$
   $\hat{\beta}_2 = -.1711 \quad se(\hat{\beta}_2) = .0962$

   $\ln(\hat{\mu}_i) = 2.0794 + -.1771 * Treatment_i$

2. β_1 is the log expected number of epileptic seizures when $Treatment_i = 0$ (Placebo group)
   β_2 is the change in the log expected number of epileptic seizures from the placebo group to the active treatment group (progabide group).

3. $exp(\beta_2)$ can be interpreted as the rate ratio even though we did not divide by the person time at risk, this would cancel out when taking the following ratio.

   $$\frac{(\mu_i|progabide)}{(\mu_i|placebo)}$$

   $exp(\beta_2) \approx .8377$
   From the model progabide appears to reduce the number of epileptic seizures by a factor of .8377

4. Rate ratio comparing progabide to placebo: .8377
   95% confidence interval for the rate ratio, comparing progabide to placebo:
   $(\exp(-.3599), \exp(.01733)) = (.6977, 1.018)$
   Since the confidence interval includes the null value 1, the results are not significant.

5. Fitting a log linear regression model adjusting for the effect of baseline age of the patient, using the placebo group as the reference group and assuming a Poisson distribution for the counts of epileptic seizures.

   $\hat{\beta}_1 = 2.5046 \quad se(\hat{\beta}_1) = .2105$
   $\hat{\beta}_2 = -.1797 \quad se(\hat{\beta}_2) = .0962$
   $\hat{\beta}_3 = -.0148 \quad se(\hat{\beta}_3) = .007$

   $\ln(\hat{\mu}_i) = 2.5046 + -.1797 * Treatment_i + -.0148 * age_i$

6. age-adjusted rate ratio comparing progabide to placebo=$\exp(\hat{\beta}_1) = .8355$

   95% CI for age-adjusted rate ratio comparing progabide to placebo
   $(\exp(-.3686), \exp(.0092)) = (.6917, 1.0092)$

7. The pearson chi-square statistic, Q_p, divided by its degrees of freedom is 12.8814. This is an indication of overdispersion since we expect the ratio of Q_p/df to be close to one.
   Q_p/df gives us the scaling factor for overdispersion ϕ

   $Q_p = 721.3566 \quad df = 56 \quad \hat{\phi} = 12.8814$ Since the estimated scale parameter for overdispersion $\hat{\phi}$ affects the standard errors only, the parameter estimates are the same as in 11.2.5. The standard errors are different.
   Multipling the standard errors from 11.2.5 by $\sqrt{p}$ gives us the new standard errors taking in account overdispersion.
   $\hat{\beta}_1 = 2.5046 \quad se(\hat{\beta}_1) = .7557$
   $\hat{\beta}_2 = -.1797 \quad se(\hat{\beta}_2) = .3459$
   $\hat{\beta}_3 = -.0148 \quad se(\hat{\beta}_3) = .0251$

   $\ln(\hat{\mu}_i) = 2.5046 + -.1797 * Treatment_i + -.0148 * age_i$

8. age-adjusted rate ratio comparing progabide to placebo, taking into account overdispersion= .8355 95% CI:
   $(\exp(-0.8576), \exp(0.4982)) = (.4242, 1.6458)$