

#### Question 4:

First we set up the data as in example 11 but kept the age variable and defined  $\text{logage} = \log(\text{age})$ . Now we consider whether the relationships between  $\text{logcnt}$  and  $\text{logbase}$  and between  $\text{logcnt}$  and  $\text{logage}$  are linear.  $\text{logage}$  clearly has a linear relationship but  $\text{logbase}$  could also be used with a high order term but we will stick with it being a linear relationship since it is still reasonable.

We then estimate the dispersion parameter  $\phi$  using the model  $\text{cnt} = \text{trt}|\text{visit}$   $\text{logbase}|\text{logage}|\text{trt}|\text{visit}$  and store it to select an error structure in the next step. The SAS output tells us the estimate is 2.1243.

We will now choose a covariance structure. Using the unstructured covariance we the error: "ERROR: Error in computing the variance function," so we will not use the unstructured covariance as it is not a good fit for our model. We get three very similar values but the AR(1) gave us a slightly smaller value than the CS so we will continue with the AR(1) variance structure.

```
Indep. : -1302.3390
AR(1)  : -1302.4522
CS      : -1302.4424
```

Now we will create models and choose the one with the lowest QICu as our model. We reduce the mean model with backward selection and by eliminating higher-order interactions. We also estimate  $\phi$  using the working models to have a better estimate of the standard error. Of the 14 models we tested, the model with the lowest QICu is number 7 which has the following score statistics and QICu:

Score Statistics For Type 3 GEE Analysis			
Source	DF	Chi-Square	Pr > ChiSq
trt	1	6.35	0.0117
visit	3	7.08	0.0695
trt*visit	3	6.68	0.0828
logbase	0	.	.
logbase*trt	1	4.08	0.0435
logbase*visit	3	7.94	0.0473
logbase*trt*visit	3	6.10	0.1067
logage	1	7.39	0.0065
logage*visit	3	6.14	0.1048

```

mean1: -1278.0590
mean2: -1293.4990
mean3: -1307.0317
mean4: -1266.2944
mean5: -1179.3371
mean6: -1307.0731
mean7: -1331.1432
mean8: -1297.7000
mean9: -1308.9752
mean10: -1201.4321
mean11: -1308.5561
mean12: -1193.1612
mean13: -1221.3559
mean14: -1229.1075

```

In our model, logage does not have an interaction term with trt so it is unnecessary and not possible to make a contrast statement including logage\*trt. The results show the point estimate of the mean ratio at the mean of logbase is 1.4201 although this would fail to be statistically significant from 0 at the 5% significance level.

#### The GENMOD Procedure

Contrast Estimate Results										
Label	Mean Estimate	Mean		L'Beta Estimate	Standard Error	Alpha	L'Beta		Chi-Square	Pr > ChiSq
		Confidence Limits					Confidence Limits			
trt-wald	3.8500	1.7109	8.6636	1.3481	0.4138	0.05	0.5370	2.1591	10.61	0.0011
trt-wald at mean	1.4201	0.9986	2.0196	0.3507	0.1797	0.05	-0.0014	0.7029	3.81	0.0509

Contrast Results for GEE Analysis				
Contrast	DF	Chi-Square	Pr > ChiSq	Type
trt-score	1	6.35	0.0117	Score
trt-wald	1	10.61	0.0011	Wald
trt-score at mean	1	3.84	0.0502	Score
trt-wald at mean	1	3.81	0.0509	Wald