STAT 8630, Spring 2013 Hwk #4 Solutions

1. See the R script hwk4-1.R and its output, hwk4-1.pdf.

10.1.1 These residuals are called m1.res0 in the R script.

10.1.2 The histogram appears on p. 2 of the output. It looks symmetric & approximately normal.

10.1.3 The Q-Q plot appears on p. 3 of the output. This plot also looks good and does not suggest any serious departures from normality in the raw population-level residuals.

10.1.4 - 10.1.6 & 10.1.7 These plots appear on pp. 4, 5, 6, respectively. In each plot there is a wave pattern that suggests a systematic misspecification of the mean in model m1. The mean response does not seem to be linearly related to age or log height. Recall that in FEVExample1.sas I tried models with higher order terms in age and had to include a quartic term in age to achieve a good fit. Alternatively, we can try spline models as in FEVExample2.sas

10.1.8 The cumulative residual plots from model m1 (without random effects) appear on pp. 7-11 of the output. These plots do not look good. The wave in the

resids vs fitteds, resids vs age, and resids vs light plots show up as an unusual path in the cumulative resids plots.

10.1.9 I fit model 3 as m3 and plotted the residuals from m3 vs fitteds on p.13, resids vs age from m3 on p.14, and the cumulative resids vs each predictor in model m3 on pp. 15-20. Unfortunately, these plots don't look much better. There is still a wave in the residuals vs fitteds & vs age that gives unusual paths in the cumulative residual plots.

To try to improve on the model, I changed from a two-piece linear spline to a two-piece quadratic spline. Specifically, I fit a model of the form

(m3a)
$$y_{ij} = (\beta_1 + b_{1i}) + (\beta_2 + b_{2i}) X_{1ij} + \beta_3 X_{1ij}^2$$
$$+ (\beta_4 + b_{3i})(X_{1ij} - K)_+ + \beta_5 X_{1i1} + \beta_6 X_{2ij} + \beta_7 X_{2i1}$$
$$+ \varepsilon_{ij}, \quad \begin{pmatrix} b_{1i} \\ b_{2i} \\ b_{3i} \end{pmatrix} \overset{iid}{\sim} N(0, D)$$ ← unstructured

$$\varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$$

Plots from model m3a appear on pp. 22-30. These look substantially better, & indicate a reasonably well specified mean. Of course, this model is not uniquely correct. Other specifications of the mean are possible which may fit equally well or better.

10.1.10 A plot of the semivariogram for model m3a appears on p. 31. It has a distinct wave pattern to it, which suggests that the error ~~residual~~ correlation structure (which is independent here) may be misspecified. Therefore, I fit one final model in which a Gaussian spatial correlation structure with a nugget was used for $var(\varepsilon_i) = R_i$. To fit this model, which I called m4, I had to drop the b_{2i} and b_{3i} random effects from m3a. The resulting fitted model had a much smaller AIC than m3a (AIC(m4) = -4705.1 vs AIC(m3a) = -4659.3) and a better looking semivariogram (p. 32 of the output).

STAT 8630 – Homework # 4 – Solution

Spring, 2013

1.) See hwk4-13.1.sas and hwk4-13.1.lst

13.1.1

a) This model is fit in the 1st call to PROC GENMOD in hwk4-13.1.sas Parameter estimates and hypothesis tests appear on p.3 of the output. A score test of the requested hypothesis has test statistic 10.99 on 4 d.f. (p=.0267), so we reject H_0, and conclude that the treatment does have an effect on changes in the log odds of good respiratory status (changes over time from baseline).

13.1.2

b) I've used effect coding in this call to PROC GENMOD, so the changes from baseline in the active treatment group are 1.00, 1.09, 1.18, and 0.675 in y1, y2, y3, y4, respectively. In the placebo group, these changes are $1.00 - 0.86 = 0.14$, $1.09 - 1.38 = -.29$, $1.18 - 1.18 = 0.00$, and $0.675 - .746 = -.071$. So much larger changes from baseline in the active treatment group, and these changes are all increases in the log odds of "good" respiratory status.

13.1.3

a) See 2nd call to PROC GENMOD (model 2). There is insufficient evidence to suggest that treatment effects differ across centers (trt*ctr effect non-significant w/ score test stat = 0.16, p=.6887) nor that the

trt * visit effect depends on center (trt*ctr*visit effect is non-significant w/ score test stat = 3.09, p=.5433)

b.) Reducing this model via backward elimination leads to model M5, which is the same as model 1 except that an additive center effect has been added

Referring to ~~the parameter~~ Model 5, the model is of the form

$$\text{logit}(\pi_{hijk}) = \mu + \alpha_h + \beta_i + \gamma_k + (\beta\gamma)_{ik} \qquad (*)$$

where π_{hijk} = Probability of good resp. health for the j^{th} subject in the i^{th} treatment, at the h^{th} center, at the k^{th} measurment occasion.

$h = 1,2$ centers; $i = 1,2$ corresponding to active, placebo;

~~k~~ $k = 0,1,2,3,4$ corresponding to baseline, years 1-4.

In this model we use the side conditions

$$\alpha_1 = \beta_1 = \gamma_1 = (\beta\gamma)_{10} = (\beta\gamma)_{11} = (\beta\gamma)_{12} = (\beta\gamma)_{13} = (\beta\gamma)_{14}$$

$$= (\beta\gamma)_{20} = 0$$

So μ = log odds of good respir. health for subjects in center 1 in the active trt at baseline ($\hat{\mu} = -.7352$)

α_2 = difference in log odds between center 2 and center 1 for active trt at baseline ($\hat{\alpha}_2 = .9696$)

β_2 = difference in log odds between placebo and active treatments at center 1 at baseline ($\hat{\beta}_2 = .0779$)

γ_k = difference in log odds between year k and year 0 for active treatment (either center), $k = 1, 2, 3, 4$.

$(\beta\gamma)_{2k}$ = difference in the change in log odds from baseline value between placebo and active treatments (either center), $k = 1, 2, 3, 4$

13.1.4

In the final call to PROC GENMOD I left model (*) using the following reparameterization

$$\text{logit}(\pi_{hijk}) = \mu_{ik} + \beta I(h=1)$$

So the $\hat{\mu}_{ik}$ are the estimated log odds of good respir health in the i^{th} trt at k^{th} time point for center 2, and $\hat{\beta}$ + the $\hat{\mu}_{ik}$'s are the corresponding values for center 1.

Applying the inverse logit to these values gives the following table of probabilities of good health:

		Visit				
Center	Trt	0	1	2	3	4
1	A	.324	.582	.609	.631	.498
	P	.341	.374	.276	.339	.325
2	A	.558	.786	.804	.819	.723
	P	.577	.612	.501	.575	.559

In both centers, the probability stays close to the baseline value for years 1-4 under the placebo treatment. However, for the active treatment the probabilities increase from baseline by an ~~uncertain~~ amount that increases over years 1-3.
The overall probability of good health is substantially higher in center 2 for all trt x year combinations

13.3 See hwk4-13.3.sas and its output, hwk4-13.3.lst.

13.3.1 This model is fit in the parameterization of the problem in the 1st and second calls to PROC GENMOD. These models are identical, but in the second call I asked for Wald tests instead of generalized score tests for the Type III effects in the model.

13.3.2 β_2 is the log odds ratio comparing the odds of a lower response (quicker to sleep) between subjects at week 2 vs week 0 in the placebo group.

$$\beta_2 = \frac{\text{log odds of lower response at wk 2, placebo}}{\text{log odds of lower response at wk 0, placebo}}$$

13.3.3 β_3 is the difference between two log odds ratios

β_3 = log OR comparing wk 2 vs wk 0 for treated

− log OR comparing wk 2 vs wk 0 for placebo

$= (\alpha_K + \beta_1 + \beta_2 + \beta_3) - (\alpha_K + \beta_1)$ ← log odds of lower response, wk 2, treated / log odds of lower response, wk 0, treated

$- [(\alpha_K + \beta_2) - \alpha_K]$ ← log odds of lower response, wk 2, placebo / log odds of lower response, wk 0, placebo

β_3 quantifies the treatment × occasion interaction on the log odds of a lower response.

13.3.4. This is a test of no interaction, or $H_0: \beta_3 = 0$. In the 1st call to PROC GENMOD, a score test of this hypothesis gives (generalized) test statistic 8.33 on 1 df, $p = .0039$. The 2nd call to PROC GENMOD gives Wald test statistic = 8.45 on 1 df, $p = .0037$. The score test is recommended. Either way, we conclude that the change in the log odds of a lower response depends on treatment. In particular, the log odds of a lower response (quicker to sleep) are estimated to be $\hat{\beta}_3 = .7078$ higher in the treated group. This is evidence that the treatment works (produces a greater increase in the odds of going to sleep quickly than does the placebo).

13.3.5 I did this based on a reparameterization of model M1 fit in the 3rd call to PROC GENMOD, but it could be done based on any equivalent parameterization.

$\hat{OR} = .7385$ w/ 95% Wald interval $(.6703, .7968)$

13.3.6 $\hat{OR} = .8514$ " " " " $(.7956, .8940)$

13.3.7 $\exp.t(-.4876) = .3805$

13.3.8 $\exp.t(2.1312) - \exp.t(.8280) = .8939 - .6959 = .1980$