PubH 7407, Spring 2009: Exam II

There are three data analysis problems, each with multiple parts; there's 100 points total. Please start a new page for each problem. Put your name at the top of each page.

1. Over the course of one school year, third graders from three different schools were exposed to three different styles of mathematics instruction: a **self**-paced computer-learning style, a **team** approach, and a traditional **class** approach. The students were asked which style they prefer and their responses, classified by the type of program they are in (a **regular** school day versus a regular day supplemented with an **afternoon** school program); the data are below (Stokes, Davis, and Koch, 2000).

School	Program	Learning style preference		
		Self=1	Team=2	Class=3
1	Regular	10	17	26
1	Afternoon	5	12	50
2	Regular	21	17	26
2	Afternoon	16	12	36
3	Regular	15	15	16
3	Afternoon	12	12	20

A generalized baseline category logit model was fit to these data with preference as the outcome variable and school and program as predictors. Let (π_s, π_t, π_c) – which must sum to one – denote the probabilities of self, team, and traditional class approach being preferred (for fixed levels of school and program).

```
 data school;
 input school program$ preference$ count @@;
* preference: self=1, team=2, class=3;
 datalines;
   1 reg 1 10  1 reg 2 17  1 reg 3 26
   1 aft 1  5  1 aft 2 12  1 aft 3 50
   2 reg 1 21  2 reg 2 17  2 reg 3 26
   2 aft 1 16  2 aft 2 12  2 aft 3 36
   3 reg 1 15  3 reg 2 15  3 reg 3 16
   3 aft 1 12  3 aft 2 12  3 aft 3 20
;

proc logistic data=school; freq count;
 class school program / param=ref;
 model preference=school program / link=glogit scale=none aggregate;
```

```
                                        The LOGISTIC Procedure

                                           Model Information
                                Response Variable             preference
                                Number of Response Levels     3
                                Frequency Variable            count
                                Model                         generalized logit

                                            Response Profile
                                   Ordered                           Total
                                     Value     preference        Frequency
                                         1     1                        79
                                         2     2                        85
                                         3     3                       174

                         Logits modeled use preference='3' as the reference category.

                                        Class Level Information
                                                                 Design
                                     Class        Value     Variables
                                     school       1           1      0
                                                  2           0      1
                                                  3           0      0
                                     program      aft         1
                                                  reg         0

                                Deviance and Pearson Goodness-of-Fit Statistics

                         Criterion          Value        DF      Value/DF      Pr > ChiSq
                         Deviance          1.7776         4        0.4444          0.7766
                         Pearson           1.7589         4        0.4397          0.7800

                                       Number of unique profiles: 6
```

```
                    Type 3 Analysis of Effects
                                  Wald
           Effect      DF    Chi-Square    Pr > ChiSq
           school       4       14.8424        0.0050
           program      2       10.9160        0.0043

              Analysis of Maximum Likelihood Estimates
                                             Standard          Wald
Parameter        preference     DF     Estimate        Error    Chi-Square     Pr > ChiSq
Intercept        1               1       0.0914       0.2925        0.0977         0.7546
Intercept        2               1       0.0894       0.2901        0.0949         0.7580
school     1     1               1      -1.3147       0.3839       11.7262         0.0006
school     1     2               1      -0.6556       0.3395        3.7296         0.0535
school     2     1               1      -0.2319       0.3327        0.4859         0.4858
school     2     2               1      -0.4755       0.3436        1.9145         0.1665
program    aft   1               1      -0.7474       0.2820        7.0272         0.0080
program    aft   2               1      -0.7426       0.2706        7.5332         0.0061
```

(a) (10) Write down the fitted model.

$$\log \frac{\hat{\pi}_s}{\hat{\pi}_c} = 0.0914 - 1.3147\, I\{s=1\} - 0.2319\, I\{s=2\} - 0.7474\, I\{p=\text{aft}\}$$
$$\log \frac{\hat{\pi}_t}{\hat{\pi}_c} = 0.0894 - 0.6556\, I\{s=1\} - 0.4755\, I\{s=2\} - 0.7426\, I\{p=\text{aft}\}$$

Although not asked for (and nontrivial to compute in a short time), here are the fitted probabilities under the model – just for your information – note that each row sums to one:

School	Program	Learning style preference Self=1	Team=2	Class=3
1	Regular	0.16	0.30	0.54
1	Afternoon	0.10	0.19	0.71
2	Regular	0.34	0.27	0.39
2	Afternoon	0.24	0.18	0.58
3	Regular	0.35	0.34	0.31
3	Afternoon	0.25	0.26	0.49

(b) (10) Using only the fitted model in (a), for what covariate combination(s) are $\hat{\pi}_s > \hat{\pi}_c$ *and* $\hat{\pi}_t > \hat{\pi}_c$ simultaneously? Hint: What does $\hat{\pi}_t > \hat{\pi}_c$ imply for $\log\left(\frac{\hat{\pi}_t}{\hat{\pi}_c}\right)$ and $\hat{\pi}_s > \hat{\pi}_c$ imply for $\log\left(\frac{\hat{\pi}_s}{\hat{\pi}_c}\right)$? For most covariate combinations, which style is preferred?

$$\hat{\pi}_s > \hat{\pi}_c \text{ implies } \log\left(\frac{\hat{\pi}_s}{\hat{\pi}_c}\right) > 0 \text{ and } \hat{\pi}_t > \hat{\pi}_c \text{ implies } \log\left(\frac{\hat{\pi}_t}{\hat{\pi}_c}\right) > 0.$$

This happens only for school 3 ($s = 3$) and regular program ($p = reg$); i.e. when the right side of the fitted model above is positive. For the other 5 covariate combinations $\hat{\pi}_c > \hat{\pi}_s$ *and* $\hat{\pi}_c > \hat{\pi}_t$; i.e. the traditional class approach is generally preferred.

(c) (10) Is there a significant program effect? How does the program affect the odds of choosing self versus class and team versus class, adjusting for the school? (The baseline category is the traditional class approach.)

Yes, the Type 3 test for dropping program yields a p-value of 0.0043, significant at the 1% level. π_s/π_c for the afternoon program is (estimated to be) approximately half ($e^{-0.7474} \approx 0.47$) that of the regular class day, and the same holds for π_t/π_c ($e^{-0.7426} \approx 0.48$). These are both significantly different from 1 ($p = 0.008$ and $p = 0.006$).

That is, the odds of choosing traditional versus either self or team increase by about 2 for for regular versus afternoon program.

2. Bell et al. (1989) considered data on lumbar laminectomy, a spinal surgery performed on children for tumor and congenital or developmental abnormalities such as syrinx, diastematomyelia, and tethered cord. Of interest is to determine the probability of spinal deformity (i.e. incidence) following surgery as a function of the child's age and the starting range of vertebrae levels involved in the operation. The data are from 83 patients. The outcome is presence or absence of kyphosis, defined as forward flexion of the spine of at least 40 degrees from vertical. We saw in class that both effects were nonlinear in the predictor. Consider the model

$$\text{logit } P(y_i = 1) = \beta_0 + \beta_1 a_i + \beta_2 a_i^2 + \beta_3 s_i + \beta_4 s_i^2,$$

where (a_i, s_i, y_i) are the data for the i^{th} subject, $y_i = 1$ indicating kyphosis is present, a_i the age in months, and s_i the starting range of vertebrae. Below is output from the model fit. Starting range ranges from 1 to 18 vertebrae, and age ranges from 1 to 206 months.

The LOGISTIC Procedure

Response Profile

Ordered Value	kyphosis	Total Frequency
1	absent	64
2	present	17

Probability modeled is kyphosis='present'.

Deviance and Pearson Goodness-of-Fit Statistics

Criterion	Value	DF	Value/DF	Pr > ChiSq
Deviance	51.2975	73	0.7027	0.9748
Pearson	53.0661	73	0.7269	0.9619

Number of unique profiles: 78

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-4.1854	1.7820	5.5165	0.0188
age	1	0.0816	0.0344	5.6264	0.0177
age*age	1	-0.00041	0.000196	4.3804	0.0364
start	1	0.5619	0.3216	3.0525	0.0806
start*start	1	-0.0482	0.0201	5.7417	0.0166

Hosmer and Lemeshow Goodness-of-Fit Test

Chi-Square	DF	Pr > ChiSq
3.2098	8	0.9205

(a) (10) Holding age constant, which starting range maximizes the probability of kyphosis?

$$\frac{d}{ds}[-4.1854 + 0.0816a - 0.00041a^2 + 0.5619s - 0.0483s^2] = 0.5619 - 0.0966s,$$

$$0.5619 - 0.0966s \stackrel{set}{=} 0 \text{ yields } s_{max} = 5.82.$$

(b) (5) Holding starting range constant, which age *in years* maximizes the probability of kyphosis?

$$\frac{d}{da}[-4.1854 + 0.0816a - 0.00041a^2 + 0.5619s - 0.0483s^2] = 0.0816 - 0.00082a,$$

$$0.0816 - 0.00082a \stackrel{set}{=} 0 \text{ yields } a_{max} = 99.5 \text{ months},$$

or $99.5/12 = 8.3$ years.

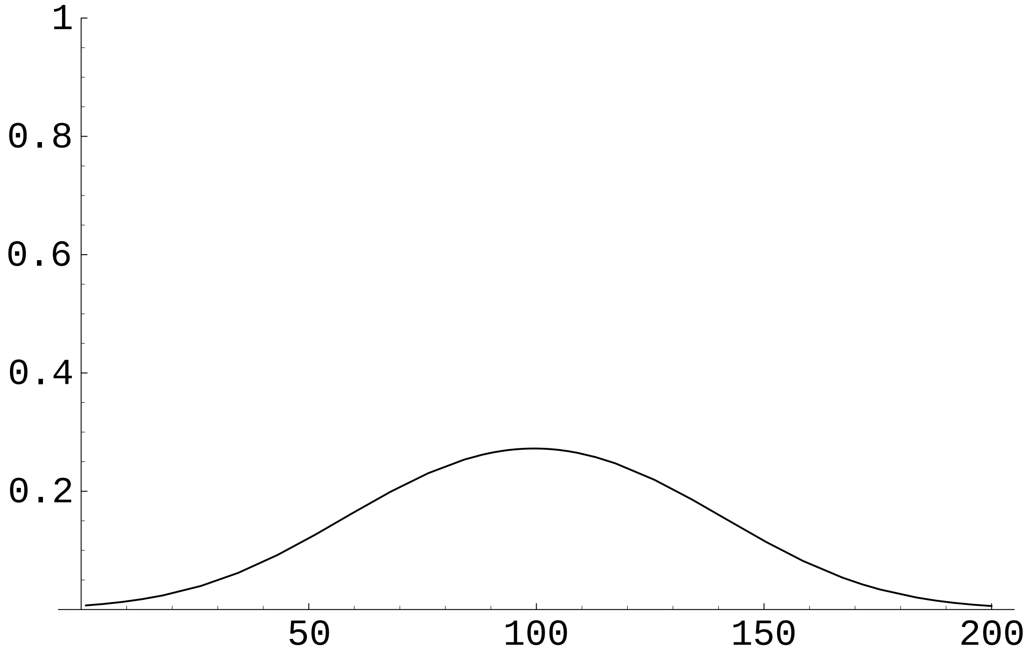

Figure 1: Probability of kyphosis for starting range equal to 13, across the age range seen in the data in months.

(c) (10) What is the probability of kyphosis $p(a)$ for a given age a (in months), fixing starting range at 13 (the median in the data set)? Roughly sketch this function across the observed range of age values in the data. Hint: only explicitly compute this function at a few ages.

Let

$$
\begin{aligned}
p(a) &= \frac{\exp(-4.1854 + 0.0816a - 0.00041a^2 + 0.5619(13) - 0.0483(13)^2)}{1 + \exp(-4.1854 + 0.0816a - 0.00041a^2 + 0.5619(13) - 0.0483(13)^2)} \\
&= \frac{\exp(-5.0434 + 0.0816 - 0.00041a^2)}{1 + \exp(-5.0434 + 0.0816 - 0.00041a^2)}.
\end{aligned}
$$

Then $p(1) = 0.0069$, $p(99.5) = 0.2722$, and $p(206) = 0.0035$. The endpoint values are essentially zero, so also take a few more $p(60) = 0.1647$ and $p(140) = 0.1604$. You can sketch it from this; see Figure 1.

(d) (10) Comment on overall lack-of-fit.

The deviance and Pearson GOF test statistics are less than their degrees of freedom (and so much less than twice the *df*), indicating (very informally) no gross LOF. The p-values for these tests cannot be trusted because the predictors are continuous and there is essentially no replication (78 unique profiles out of 81). The Hosmer and Lemeshow test gives a p-value of 0.92 indicating no gross LOF.

---

3. Consider the clinical trial data from eight centers, where two creams were compared to cure infection. $y_{ij} = 1$ indicates success (infection cure) for subject j at clinic i with covariate $x_{ij} = 0$ for drug and $x_{ij} = 1$ for control. A very tiny table is reproduced below to remind you of the data:

Center	Treatment X	Response Y Success	Failure
a	Drug	11	25
	Control	10	27
b	Drug	16	4
	Control	22	10
c	Drug	14	5
	Control	7	12
d	Drug	2	14
	Control	1	16
e	Drug	6	11
	Control	0	12
f	Drug	1	10
	Control	0	10
g	Drug	1	4
	Control	1	8
h	Drug	4	2
	Control	6	1

Several models were fit in SAS PROC LOGISTIC, GENMOD, and NLMIXED, including fixed and random effects models, and a marginal model. The fixed effects model fit in PROC LOGISTIC is

$$\text{logit } P(y_{ij} = 1) = \beta_0 + \beta_1 x_{ij} + u_i,$$

where $u_1, \ldots, u_8$ are fixed clinic effects constrained so that $\sum_{i=1}^{8} u_i = 0$. The random effects model instead assumes $u_1, \ldots, u_8 \overset{iid}{\sim} N(0, \sigma^2)$.

```
data ctr1; input center$ treat s n @@; f=n-s; treat=treat-1;
 datalines;
 a 1 11 36 a 2 10 37  b 1 16 20 b 2 22 32
 c 1 14 19 c 2 7  19  d 1 2  16 d 2 1  17
 e 1 6  17 e 2 0  12  f 1 1  11 f 2 0  10
 g 1 1   5 g 2 1   9  h 1 4   6 h 2 6   7
;
data ctr2; set ctr1; do i=1 to n; if i<=s then y=1; else y=0; output; end;
proc logistic data=ctr2; class center;
  model y(event='1') = treat center;
proc genmod descending data=ctr2; class center;
  model y=treat / dist=bin link=logit;
  repeated subject=center / type=exch corrw;
proc nlmixed data=ctr2 qpoints=100;
  eta=alpha+beta*treat+u; p=exp(eta)/(1+exp(eta));
  model y ~ binary(p);
  random u ~ normal(0,sig*sig) subject=center out=re;
proc print data=re;
```

```
                              The LOGISTIC Procedure

                            Probability modeled is y=1.

                            Type 3 Analysis of Effects
                                             Wald
                      Effect      DF    Chi-Square    Pr > ChiSq
                      treat        1        6.4174        0.0113
                      center       7       58.4897        <.0001

                       Analysis of Maximum Likelihood Estimates
                                           Standard          Wald
   Parameter       DF    Estimate       Error    Chi-Square    Pr > ChiSq
   Intercept        1     -0.4784      0.2592        3.4059        0.0650
   treat            1     -0.7769      0.3067        6.4174        0.0113
   center    a      1     -0.0667      0.3133        0.0453        0.8315
   center    b      1      1.9888      0.3556       31.2789        <.0001
   center    c      1      1.0862      0.3596        9.1236        0.0025
   center    d      1     -1.4851      0.5707        6.7711        0.0093
   center    e      1     -0.5866      0.4582        1.6390        0.2005
   center    f      1     -2.2136      0.9171        5.8260        0.0158
   center    g      1     -0.8644      0.7016        1.5178        0.2180
```

```
                               The GENMOD Procedure

PROC GENMOD is modeling the probability that y='1'.

                               GEE Model Information

                 Correlation Structure               Exchangeable
                 Subject Effect                   center (8 levels)
                 Number of Clusters                               8
                 Correlation Matrix Dimension                    73
                 Maximum Cluster Size                            73
                 Minimum Cluster Size                            13

                               Exchangeable Working
                                    Correlation
                            Correlation    0.2168926903
```

Analysis Of GEE Parameter Estimates
Empirical Standard Error Estimates

Parameter	Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept	-0.3201	0.4111	-1.1259	0.4858	-0.78	0.4363
treat	-0.5540	0.2330	-1.0106	-0.0974	-2.38	0.0174

The NLMIXED Procedure

Fit Statistics

-2 Log Likelihood	302.9
AIC (smaller is better)	308.9

Parameter Estimates

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper	Gradient
alpha	-0.4591	0.5508	7	-0.83	0.4320	0.05	-1.7616	0.8433	0.000013
beta	-0.7385	0.3004	7	-2.46	0.0436	0.05	-1.4489	-0.02808	2.115E-6
sig	1.4008	0.4261	7	3.29	0.0133	0.05	0.3934	2.4083	0.000033

Obs	center	Effect	Estimate	StdErr Pred	DF	tValue	Probt	Alpha	Lower	Upper
1	a	u	-0.09886	0.57554	7	-0.17177	0.86848	0.05	-1.45980	1.26208
2	b	u	1.85011	0.60147	7	3.07598	0.01792	0.05	0.42786	3.27235
3	c	u	0.99147	0.60198	7	1.64702	0.14355	0.05	-0.43199	2.41493
4	d	u	-1.29471	0.69606	7	-1.86006	0.10520	0.05	-2.94062	0.35121
5	e	u	-0.55775	0.64815	7	-0.86052	0.41800	0.05	-2.09038	0.97488
6	f	u	-1.60169	0.81836	7	-1.95719	0.09120	0.05	-3.53681	0.33343
7	g	u	-0.70444	0.76815	7	-0.91706	0.38961	0.05	-2.52081	1.11194
8	h	u	1.73721	0.74864	7	2.32047	0.05336	0.05	-0.03306	3.50747

(a) (10) Focus on the fixed and random clinic effects models. How many parameters (and thus degrees of freedom) does the fixed effects model add to the model including an intercept and treatment effect? How many parameters does the random effects model add? Is there a significant treatment effect in either model? Provide estimates and 95% CIs of the odds ratio for drug versus control and interpret.

The fixed effects model adds 7 parameters: offsets to the intercept for 7 of the 8 clinics; the random effects model adds one parameter σ. The treatment effect is significant in both models. In the random effects model the odds of success change by a factor of $e^{0.7385} = 2.1$ for treatment versus control within a clinic; a 95% CI is $(1.2, 3.8)$. In the fixed effects model, the odds of success change by a factor of $e^{0.7769} = 2.2$ for treatment versus control within a clinic; a 95% CI is $(1.2, 4.0)$.

(b) (5) Now focus on the marginal model. What does exchangeable correlation within a clinic imply? Provide an estimate of this correlation. Provide an estimate and 95% CI of the odds ratio for drug versus control and interpret.

Exchangeable correlation implies that responses within a clinic are correlated, but that this correlation is the same for all pairs of observations (from different subjects). The estimated correlation is $\hat{\alpha} = 0.22$.

The odds of success change by a factor of $e^{0.554} = 1.7$ for treatment versus control; a 95% CI is $(1.1, 2.7)$. This interpretation is for the population as a whole, averaged across clinics.

(c) (5) Is success independent of treatment, conditional on the clinic? Which analyses address this?

Success significantly depends on treatment within clinic ($p = 0.01$ for fixed effects, $p = 0.04$ for random effects); the conditional analyses address this.

(d) (5) Is success *marginally* independent of treatment? Which analyses address this?

Success significantly depends on treatment marginally ($p = 0.02$); the marginal analysis address this.

(e) (5) Random effects models are said to "shrink" population estimates. Compare the intercept and treatment effects from the two conditional models and comment on this statement. How about the estimated clinic effects?

Both estimated regression effects are smaller under the random effects model; clinic effects are similar.

(f) (5) Below is a fit of a simple logistic regression model without random effects. Formally test $H_0 : \sigma = 0$ in the conditional model. Does this agree with the AIC?

The difference in $-2 \log L$ is $358.2 - 302.9 = 55.3$. The p-value is $0.5P(\chi^2_1 > 55.3) < 0.5P(\chi^2_1 > 3.84) = 0.025$. So we reject at the 5% level. The AIC confirms this; AIC=362.2 for the model without random effects and AIC=308.9 for the model with random effects.

```
proc nlmixed data=ctr2;
eta=alpha+beta*treat; p=exp(eta)/(1+exp(eta));
model y ~ binary(p);
                                  The NLMIXED Procedure

                                     Fit Statistics

                          -2 Log Likelihood              358.2
                          AIC (smaller is better)        362.2
```