12 Population-averaged models for nonnormal repeated measurements

12.1 Introduction

In the previous chapter, we discussed regression models for data that may not be normally distributed, such as count or binary data or data that take on positive values but that may have skewed distributions. These models, known as **generalized linear models**, have several features:

- A by-product of dealing with these types of variables is that the model for mean response may need to satisfy some restrictions. The most extreme case was that of models for binary data; here, the mean response is also the probability of seeing the event of interest, which must lie between 0 and 1. The main consequence is that models of interest are no longer necessarily **linear** in regression parameters $\boldsymbol{\beta}$ ($p \times 1$); instead, plausible models tend to be **nonlinear** functions f of $\boldsymbol{\beta}$ through a **linear predictor** $\boldsymbol{x}_j'\boldsymbol{\beta}$. Thus, the usual theory of linear models does not apply.
- The **variance** of the response is no longer legitimately viewed as being **constant** for all values of the mean response (that is, for all settings of the covariates). Rather, the distributional models that are sensible for these data impose a **relationship** between mean and variance; that is, the variance of a response taken at a particular value of the mean is some known function V of the mean.
- Because of the nonlinearity of mean response models and the fact that variance also is a function of the mean, it is no longer possible to derive an expression for the estimator of $\boldsymbol{\beta}$ in closed form. However, fortunately, it turns out that for all distributions in the class containing the relevant distributions, such as the Poisson, Bernoulli, and gamma, the (ML) estimator of $\boldsymbol{\beta}$ solves a set of p equations that is a sum of **weighted** deviations. Although these equations cannot be solved analytically, they may be solved via a general numerical algorithm (IRWLS). Furthermore, large sample approximations are available for the sampling distribution of the estimator $\widehat{\boldsymbol{\beta}}$, so that approximate inference may be carried out.

Generalized linear models may thus be viewed as an extension of ordinary linear regression models for normal data with constant variance. These models and methods are of course only applicable to the standard regression problem where independent scalar responses $Y_1, \ldots, Y_n$ have been observed at covariate settings $x_{j1}, \ldots, x_{jk}$ for the jth response, $j = 1, \ldots, n$.

In this chapter, we are concerned with how we might extend generalized linear models to the situation of longitudinal data, where now the responses are **vectors** $\boldsymbol{Y}_i$ of repeated count, binary, or other observations on each of m units.

- Recall in the the case of the linear model with the assumption of normality, the extension from ordinary regression problems to the longitudinal problem was facilitated by thinking about the **multivariate normal distribution**. That is, there is a natural generalization of the probability model we use for ordinary linear regression (the normal distribution) to that we use for longitudinal response vectors (multivariate normal).

- Specifically, if individual observations are assumed to be normally distributed, as they are in classical linear regression, then **vectors** of such observations have a multivariate normal distribution. Each component of the data vector is normally distributed individually, with mean determined by the regression model and variance that of the individual normal distribution. To fully characterize the multivariate normal distribution that is appropriate, the only **additional** piece of information we must specify is how the components of the vector are **correlated**. Put another way, as long as (i) we believe individual observations are normally distributed and (ii) are willing to specify the form of the **mean vector** through a regression model and the form of the **covariance matrix** of a data vector, either by outright assumption or using a mixed effects structure, we can **fully specify** the particular multivariate normal distribution that will be used as the basis for inference. Because of this, it was straightforward to contemplate models for longitudinal, normally distributed data. Moreover, because we thus had a full probability model, we could write down the joint probability distribution of the data and use the methods of maximum likelihood or restricted maximum likelihood to fit the model and make inference.

- By analogy, it is natural to hope that we could do something similar when the elements of a data vector $\boldsymbol{Y}_i$ are now counts, binary responses, or positive responses with constant CV. That is, it would be desirable if there were extensions of the Poisson, Bernoulli, and gamma distributions that could be **fully specified** by simply adding assumptions about **correlation** to the individual observation assumptions on mean and variance.

- Unfortunately, this is **not** the case. This same kind of generalization is not so easy for the other distributions in the scaled exponential family class, like the Poisson, Bernoulli, or gamma. In particular, multivariate extensions of these probability models are unwieldy or require **more** than just an assumption about the correlations among components of a data vector. Thus, sadly, trying to use multivariate extensions of the distributions used for ordinary regression (generalized linear models) to longitudinal data vectors is simply too complex to yield useful statistical models for real situations.

To make matters worse, still another problem complicates things **further**. We have noted two perspectives on modeling: **population-averaged** and **subject-specific**. For continuous, normally distributed data, it is often relevant, as we have seen, to specify models that are **linear**:

- With the **population-averaged** perspective, we modeled the **mean response** of the elements of a data vector by some function of **time** and possibly other covariates. This function was **linear** in parameters $\boldsymbol{\beta}$, e.g.

  $$E(Y_{ij}) = \beta_0 + \beta_1 t_{ij}.$$

  We then modeled the covariance matrix $\boldsymbol{\Sigma}_i$ of a data vector explicitly. This model would (hopefully) take into account variation from **all** sources, **among** and **within** individuals simultaneously.

- With the **subject-specific** perspective, we modeled the **individual trajectory** of the elements of a data vector by some function of **time**. This function was **linear** in individual-specific parameters; e.g. we wrote models like the straight-line random coefficient model

  $$Y_{ij} = \beta_{0i} + \beta_{1i} t_{ij} + e_{ij}.$$

  The individual-specific parameters β_{0i} and β_{1i} were in turn modeled as **linear** functions of a fixed parameter $\boldsymbol{\beta}$ and **random effects** $\boldsymbol{b}_i$, $\boldsymbol{\beta}_i = \boldsymbol{A}_i\boldsymbol{\beta} + \boldsymbol{b}_i$, that characterized respectively the "typical" values of the elements of $\boldsymbol{\beta}_i$ and how individual values deviated from these typical values. The result was **again** a model for mean response averaged across individuals that was a **linear** function of $\boldsymbol{\beta}$; e.g., with $\boldsymbol{A}_i = \boldsymbol{I}$,

  $$E(Y_{ij}) = \beta_0 + \beta_1 t_{ij}.$$

  The covariance model $\boldsymbol{\Sigma}_i$ arose from the combination of assumptions about $\boldsymbol{b}_i$ and $\boldsymbol{e}_i$, thus naturally taking into account variation from both sources separately.

Thus, in both cases, although the perspective starts out differently, we end up with a model for **mean response** $E(Y_{ij})$ that is a **linear** function of fixed parameters $\boldsymbol{\beta}$ of interest. We can end up with the **same** linear mean model from either perspective. So, even if two data analysts start out with these different perspectives, they are likely to arrive at the same mean model, and either of their interpretations of the model will be valid. The difference will be in what they end up assuming about **covariance**.

As we will discuss, when we consider models of the generalized linear model type that are **no longer linear**, it is **no longer the case** that the population-averaged and subject-specific perspectives necessarily can lead to the **same mean model**! Moreover, as a result, the **interpretations** of the different types of models are no longer both valid at the same time. This unfortunate problem is the result of the **nonlinearity** of the generalized linear models.

Historically, as a consequence of all of these issues, models and method for nonnormal responses that individually would follow generalized linear models were not widely available. The main impediments were that

- there are not easy multivariate generalizations of the necessary probability distributions, and
- population-averaged and subject-specific approaches do not necessarily lead to the same models for mean response.

Because there was no easy resolution to these problems, no one knew quite what to do. Then, in the mid-1980's, a paper appeared in the statistical literature that brought to the attention of statisticians an approach for modeling these data, along with an associated fitting method, that made good practical sense from a **population-averaged** perspective. The paper, Liang and Zeger (1986), generated a huge amount of interest in this approach.

In this chapter, we will introduce this approach and the associated fitting method known as **generalized estimating equations**, or **GEE**s. We will also show how to use `PROC GENMOD` in SAS to carry out such analyses. As we will detail in the next section, the modeling of data vectors follows from a **population-averaged** perspective, where the mean response of a data vector is modeled **explicitly** as a function of time, parameters $\boldsymbol{\beta}$, and possibly other covariates. No subject-specific random effects are involved. We will contrast this approach with one that does use subject-specific random effects in Section 12.5 and in the next chapter.

12.2 Population-averaged model

RECALL: The **population-averaged** approach is focused on modeling the **mean response** across the population of units at each time point as a function of time. Thus, the model describes how the averages across the population of responses at different time points are related over time. The model usually describes the mean response at any time t_{ij}, say, for unit i as a function of fixed parameters $\boldsymbol{\beta}$, time t_{ij}, and possibly additional covariates. The model is set up so that questions about how the mean response changes as a function of time and other covariates may be phrased in terms of questions about the value of **contrasts** of the elements of $\boldsymbol{\beta}$.

PROBLEM: In the case of **normally** distributed responses, if we specify such a mean response model **and** a model for the covariance matrix of a data vector, we have provided all the necessary ingredients to write down a **multivariate normal probability distribution** that we believe describes the population(s) of data vectors.

- Technically, if we can provide a mean vector and a covariance matrix, this is all we need to fully describe a corresponding multivariate normal distribution.
- This is a **desirable feature** of the multivariate normal distribution – it is **fully characterized** by a mean and covariance matrix.

In the case of **nonnormally** distributed response, if we specify such a mean response model and a model for the covariance matrix, we have **not necessarily** provided all the necessary ingredients to write down a corresponding **multivariate probability distribution** that we believe describes a population of data vectors. Here is a brief heuristic explanation:

- Technically, to develop **multivariate extensions** of probability distributions like the those underlying generalized linear models, it is **not enough** to provide just a mean vector and covariance matrix.
- Because in these probability distributions the **mean** and **variance** of an observation are **related** in a specific way, it turns out that it is much more difficult to fully describe a multivariate probability distribution for several such observations in a data vector. To do so requires not only **mean** and **covariance matrix** models, but **additional assumptions** about more complicated properties of observations taken three, four, ..., n at a time.
- With only the data at hand to guide the data analyst, it may be too **difficult** and **risky** to make

all of the assumptions required about these complicated properties. Furthermore, the resulting probability models can be so complex that fitting them to real data may be an insurmountable challenge.

APPROACH: The approach popularized by Liang and Zeger (1986) is to **forget** about trying to model the whole multivariate probability distribution of a data vector. Instead, the idea is just to model the **mean response** and the **covariance matrix** of a data vector as in the normal case, and leave it at that.

- The problem with this approach is that, consequently, there is no multivariate probability distribution upon which to base fitting methods and inference on parameters (like **maximum likelihood**).
- However, Liang and Zeger (1986) described an alternative approach to model fitting for such **mean-covariance** models for nonnormal longitudinal data that **does not require** specification of a full probability model but rather just requires the mean and covariance matrix. We discuss this method in the next section.

Here, we describe the modeling strategy.

MEAN–VARIANCE MODEL: The idea is to take **generalized linear models** for individual observations as the starting point.

- If we consider a **single component** of a data vector $\boldsymbol{Y}_i$ consisting of counts, binary responses, or continuous positive response with constant CV at different times, the distribution of possible values across the population of units might be well-represented by the Poisson, Bernoulli, and gamma probability models, respectively.
- Thus, the distribution of each observation in a data vector is taken to have ideally a **mean** and **variance** model of the type relevant to or imposed by these distributions.

EXAMPLE – EPILEPTIC SEIZURE DATA: Recall Example 4 from Chapter 1, given by Thall and Vail (1990). Here, 59 subjects suffering from epileptic seizures were assigned at random to receive either a placebo (subjects 1–28) or the anti-seizure drug progabide (subjects 29–59) in addition to a standard chemotherapy regimen all were taking. On each subject, the investigators recorded the subject's age, a_i, say for the ith subject, $i = 1, \ldots, 59$, a **baseline** number of seizures experienced by each subject over the 8-week period prior to the start of the study, and then the number of seizures over a 2 week period for four visits following initiation of assigned treatment. Let δ_i be the treatment indicator for the ith patient,

$$\begin{aligned} \delta_i &= 0 \quad \text{for placebo subjects} \\ &= 1 \quad \text{for progabide subjects} \end{aligned}$$

Before we consider a model for these data, we discuss an issue that has been of some debate among practitioners, that of "how to handle "baseline?"

In all of our examples up till now involving different groups, we have treated a baseline response, that is, a measure of the response taken at the start of a study (and prior to administration of treatment if there is one) as part of the overall response vector $\boldsymbol{Y}_i$. This takes automatic account of the information in the baseline response, its correlation with other responses, and the fact that different subjects have different baseline characteristics.

However, a common approach is to instead view the response vector as just the **post-baseline** responses and treat the baseline response as a **covariate** in a model for mean of this response vector. The idea is that this takes into account, or "adjusts for," the fact that different subjects have different baseline response characteristics.

Here, the baseline response and subsequent responses are not on the same scale; the baseline response is the number of seizures recorded over an **8-week** period prior to the start of the study (initiation of assigned treatment) while the post-baseline responses are the number recorded in the **2-week** period between the four visits. This discrepancy might especially motivate an analyst to treat baseline as a covariate, as it does not seem comparable with the rest of the response variables. In fact, the original analysis of these data by Thall and Vail (1980) did this.

However, this seems to be suboptimal, as it would seem to **ignore** the fact that baseline response would be expected to **vary** within subjects; that is, baseline response is a random variable. It is a simple matter to address the scaling issue; in the current study, one may divide the baseline responses by 4 to place them on a two-week basis.

The more fundamental issue is whether it is a good idea to treat a baseline response as a covariate in order to take into account the fact that units differ in their responses prior to treatment or whether it is preferable to treat the baseline value as part of the response vector for each unit. In the case of a **linear** mean response, it turns out that the two strategies can be **equivalent**, which is why we have not discussed this until now. However, when the model for mean response is **nonlinear**, this no longer holds.

Our position is that as a general strategy, it is preferable to treat a baseline response as part of the response vector rather than as a covariate. There are theoretical reasons, beyond our scope here, that support this position. We continue to follow this strategy for the rest of this course.

A very nice, detailed discussion of this issue is given by Fitzmaurice, Laird, and Ware (2004, Section 5.7).

Returning to the seizure data, adopting this view, we take the data vector corresponding to subject i to be $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{i5})'$, where Y_{i1} is the baseline response based on 8 weeks, and $Y_{i2}, \ldots, Y_{i5}$ are the responses at each of visits 1–4 based on 2 weeks (we discuss how to take into account the different time periods momentarily).

Before we specify the model, we consider some summary statistics. This was a **randomized** study, so we would expect subjects in the two groups to be similar in their characteristics prior to administration of the treatment. This seems plausible; the following table lists sample means (standard deviations) of age and baseline 8-week seizure counts (Y_{i1}) for each group.

	Age	Baseline
Placebo	29.6 (6.0)	30.8 (26.1)
Progabide	27.7 (6.6)	31.6 (27.9)

Notice that the subjects vary considerably in their baseline seizure counts.

Table 1 lists sample mean seizure counts at baseline and each visit time; those for baseline are divided by 4 to put them on the same 2-week scale as the others.

Table 1: *Sample mean seizure counts at baseline and each visit time for the 28 subjects assigned to placebo and 30 subjects assigned to progabide.*

Visit	*Placebo*	*Progabide*
0 (baseline)	7.70	7.90
1	9.35	8.58
2	8.29	8.42
3	8.79	8.13
4	7.96	6.71
average over visits 1–4	8.60	7.96

The raw sample means suggest a possible slight initial **increase** in 2-week seizure count followed by a "leveling-off," with a possible lowering by visit 4 in the progabide group.

Based on these observations, we might adopt a model for mean response that allows the possibility of a different mean at baseline and visits 1–4, where the mean at visits 1–4 is the same, and these might be different by group. Because the responses may be **small counts** for some subjects and are indeed counts for all, it is natural to consider a **loglinear** model.

Define $v_{ij} = 0$ if $j = 1$ (baseline) and $v_{ij} = 1$ otherwise (visits 1–4), and let $o_{ij} = 8$ if $j = 1$ and $o_{ij} = 2$ otherwise, so that o_{ij} records the observation period on which Y_{ij} is based (8 or 2 weeks). Then the following loglinear model incorporates these features:

$$E(Y_{ij}) = \exp(\log o_{ij} + \beta_0 + \beta_1 v_{ij} + \beta_2 \delta_i + \beta_3 \delta_i v_{ij}), \tag{12.1}$$

where thus $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_3)'$ is the vector of fixed regression parameters characterizing the mean response vector for any subject.

- The fixed quantity $\log o_{ij}$ cleverly takes account of the different observation periods for baseline and post-treatment visits. If we take the log of both sides of (12.1) and subtract $\log o_i$ from both sides, we get

  $$\log\{E(Y_{ij})\} - \log o_{ij} = \log\{E(Y_{ij}/o_{ij})\} = \beta_0 + \beta_1 v_{ij} + \beta_2 \delta_i + \beta_3 \delta_i v_{ij},$$

  so this is equivalent to modeling the means of $Y_{i1}/8$ and $Y_{ij}/2$ for $j = 2, \ldots, 5$.

- Model (12.1) says that, at baseline, the mean response is

$$\log\{E(Y_{i1}/8)\} = \beta_0 + \beta_2\delta_i$$

  while for visits 1–4 the mean is

$$\log\{E(Y_{ij}/2)\} = \beta_0 + \beta_1 + \beta_2\delta_i + \beta_3\delta_i,$$

  which is the same for all 4 post-baseline visits and may be viewed as reflecting the "overall" behavior averaged across them. Here, β_1 is the amount by which the logarithm of the mean "shifts" after the study begins. β_2 allows the baseline mean to be different by treatment, and β_3 reflects the additional amount by which the mean differs by treatment after treatment starts.

  As the study was randomized, we would not necessarily expect baseline mean responses to be different by treatment; certainly the sample means given above do not support this. We might thus eliminate this term from the model.

- A fancier model might allow the mean response to change smoothly with time (measured in weeks) following visit 1 somehow. One possibility would be to allow a straight-line relationship between baseline and visit 1, and then another straight-line relationship from visit 1 onward.

- Alternatively, the sample means seem to suggest that the effect of the progabide may not become apparent until the last visit. We consider such a model later in this chapter. We also consider taking into account age.

- On the original scale, note that as before that, for a loglinear model like (12.1), receiving treatment versus not has the effect of causing a **multiplicative** change in mean response. In particular, $\exp(\beta_3)$ is the multiplicative effect of progabide relative to placebo post-baseline. If β_3 is positive, then the multiplicative factor is **greater** than one, and the mean response increases; if β_2 is negative, then the multiplicative factor is **less** than one, and the mean response decreases.

EXAMPLE – WHEEZING DATA: Recall Example 5 from Chapter 1, given by Lipsitz, Laird, and Harrington (1992). These data are from a large public health study (the Six Cities study) and concerned the association between maternal smoking and respiratory health of children. In section 12.7, we will consider a subset of the full data set, data on 32 of these children. Each child was examined once a year at a clinic visit (visits at ages 9, 10, 11, and 12) for evidence of "wheezing" – the response was recorded as a binary variable (0=wheezing absent, 1=wheezing present).

In addition, the mother's current smoking status was recorded (0=none, 1=moderate, 2=heavy). For some children, visits were missed, so that both the response (wheezing indicator) and maternal smoking status were missing; for our purposes, we will assume that the reasons for this missingness are not related to the focus of study. (See Chapter 13 for more on missing data.)

Let Y_{ij} be the wheezing indicator (=0 or 1) on the ith child at the jth age t_{ij}, where t_{ij} ideally takes on all the values $9, 10, 11, 12$. Thus, $j = 1, \ldots, n_i$ for any child, with $n_i \leq 4$. As the response is binary, a **logistic** regression model would be appropriate for representing $E(Y_{ij})$. For child i, let

$$\begin{aligned}
\delta_{0ij} &= 1 \quad \text{if smoking=none at } t_{ij} \\
&= 0 \quad \text{otherwise} \\
\delta_{1ij} &= 1 \quad \text{if smoking=moderate at } t_{ij} \\
&= 0 \quad \text{otherwise} \\
c_i &= 0 \quad \text{if city=Portage} \\
&= 1 \quad \text{if city=Kingston}
\end{aligned}$$

Recall the discussion in Chapter 10 regarding **time-dependent covariates**. As maternal smoking is a time-dependent covariate, the considerations raised in that discussion are relevant. Here, we are interested in a model for mean response for the jth element of a data vector, $E(Y_{ij})$.

- As a mother's smoking behavior is something we only can **observe**, we should probably be more careful and acknowledge that it should be thought of as **random**; thus, we would think of the pair $\boldsymbol{\delta}_{ij} = (\delta_{0ij}, \delta_{1ij})'$ as a **random vector** characterizing the observed smoking behavior at age j. Thus, following the discussion in Chapter 10, we are really modeling the $E(Y_{ij}|\boldsymbol{\delta}_{i1}, \ldots, \boldsymbol{\delta}_{in_i})$.
- The model used by Lipsitz, Laird, and Harrington (1992) takes $E(Y_{ij})$ as depending on a mother's smoking status $(\delta_{0ij}, \delta_{1ij})$ at time j only; that is, they assume

  $$E(Y_{ij}|\boldsymbol{\delta}_{i1}, \ldots, \boldsymbol{\delta}_{in_i}) = E(Y_{ij}|\boldsymbol{\delta}_{ij}) = E(Y_{ij}|\delta_{0ij}, \delta_{1ij}).$$

  One possible rationale is that, because measurements are so far apart in time (one year), it might be believed that a mother's smoking behavior at one time is not associated with respiratory problems at another time. However, given the discussion in Chapter 10, this is something that must be considered critically.

In this example, an objective (see Chapter 1) is to understand whether maternal smoking behavior has an effect on wheezing.

A little thought suggests that this is indeed a complicated question; the children have not been subjected to a "one-time" treatment (smoking or not) that distinguishes them into groups, as in previous examples. Rather, the "treatment" changes with time and may be related to the response in a complicated way, as discussed in Chapter 10. It is not at all clear that a simple model like that above addresses this. Indeed, this question would seem to involve a **causal** interpretation! At best, all we can hope for is to understand **associations.**

Thus, writing down an appropriate model for $E(Y_{ij})$ requires considerable thought and a clear idea of how the model is to be used.

- It is sometimes argued that, if the goal is to use the model only to estimate a future child's risk of wheezing based on information at a particular time point only, then a model for $E(Y_{ij})$ as a function of $(\delta_{0ij}, \delta_{1ij})$ at j only may be of interest, even if it doesn't capture the true underlying mechanism leading to wheezing.
- However, this is almost always **not** the goal! Rather, the objective is as above: to assess and compare the effects of smoking patterns on wheezing patterns. Trying to do this based on the simple model we discuss next is likely to result in flawed and meaningless interpretations.

Further discussion is beyond the scope of this course; however, it is **critical** that the data analyst confronted with data such as these appreciate that there are profound issues involved in modeling them! Frankly, one should be **extremely careful** when dealing with **time dependent covariates** and **longitudinal data**.

- We again refer the reader to Fitzmaurice, Laird, and Ware (2004) for discussion. A very technical paper that also discusses this issue is from the literature on **causal inference** [Robins, Greenland, and Hu (1999)].

With the above **caveats** in mind, we show for illustration a model similar to that proposed by Lipsitz, Laird, and Harrington (1992). The model is

$$E(Y_{ij}) = \frac{\exp(\beta_0 + \beta_1 c_i + \beta_2 \delta_{0ij} + \beta_3 \delta_{1ij})}{1 + \exp(\beta_0 + \beta_1 c_i + \beta_2 \delta_{0ij} + \beta_3 \delta_{1ij})}, \tag{12.2}$$

where thus $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_3)'$ is the vector of fixed regression parameters characterizing the mean response vector for any subject. Of course, this implies (see the previous chapter) that the **log odds** is given by

$$\log\left(\frac{E(Y_{ij})}{1 - E(Y_{ij})}\right) = \beta_0 + \beta_1 c_i + \beta_2 \delta_{0ij} + \beta_3 \delta_{1ij}.$$

- Model (12.2) thus says that the **log odds** of having a wheezing response relative to not having it depends (linearly) on city and maternal smoking status. We could additionally add an "age" term to allow dependence on age (maybe as children grow older their tendency toward wheezing changes).

- Specifically, the model says that the log odds at age t_{ij} is equal to β_0 for a child from Portage whose mother is a heavy smoker at t_{ij}, since under these conditions $c_i = \delta_{0ij} = \delta_{1ij} = 0$. For a child from Kingston, the log odds would change by adding the amount β_1; for a child whose mother was a non (moderate) smoker, the log odds would change by adding the amount β_2 (β_3).

- With the model written as (12.2), we see that, because the logistic function increases (decreases) as the linear predictor increases (decreases), we see that the probability of wheezing at time t_{ij}, $E(Y_{ij})$, will, for example, increase if $\beta_1 > 0$ and a child is from Kingston ($c_i = 1$) rather than Portage ($c_i = 0$). If $\beta_1 < 0$, then the probability of wheezing is smaller for a child from Kingston than for one from Portage. Similarly, if $\beta_2 < 0$, this would say that the probability of wheezing is smaller for a child whose mother is a non- rather than heavy smoker (and similarly for $\beta_3 < 0$ and moderate smoking).

VARIANCE: The above examples illustrate how one might model the mean response as a function of time and other covariates using the types of models appropriate for nonnormal data. The next part of the modeling strategy is to model the **variance** of each element of the data vector.

- Recall that in the population-averaged approach, the covariance matrix of a data vector is modeled **directly**; i.e. the model selected incorporates the aggregate effects **both** of within- and among-unit variation. Thus, the diagonal elements of the covariance matrix represent the combined effects of variance from both sources.

- Thus, in the approach here, when we specify a model for variance of an element Y_{ij}, we are modeling the aggregate variance from both sources.

Thus, for the different types of data, the model for $\text{var}(Y_{ij})$ is meant to represent the overall variance of Y_{ij} from both sources. That is, the distribution of each observation in a data vector across the population of all units and including variability in taking measurements is assumed to have variance related to the assumed **mean** for Y_{ij} as in the models above. How variance is related to the mean depends on the type of data:

- For example, for **binary** responses Y_{ij} taken on unit i at times t_{ij}, variance would be taken to be that of a binary random variable as imposed by the Bernoulli distribution; i.e.

$$\text{var}(Y_{ij}) = E(Y_{ij})\{1 - E(Y_{ij})\}. \tag{12.3}$$

  Thus, for the wheezing data, variance would be modeled as in (12.3) with $E(Y_{ij})$ as in (12.1).

- For responses Y_{ij} in the form of **counts** taken at times t_{ij} on unit i, variance would be taken to be that of a Poisson random variable; i.e.

$$\text{var}(Y_{ij}) = E(Y_{ij}) \tag{12.4}$$

- For positive responses with constant coefficient of variation, variance would be modeled as $\text{var}(Y_{ij}) = \sigma^2\{E(Y_{ij})\}^2$, where $E(Y_{ij})$ is modeled by a suitable function like the loglinear or reciprocal model.

OVERDISPERSION: Sometimes, these models for variance turn out to be inadequate for representing all the variation in observations taken at a particular time across units. There are many reasons why this may be the case:

- The aggregate effects of (i) error introduced by taking measurements and (ii) variation because units differ add up to be more than would be expected if we only considered observations on a particular unit.

- There may be other factors involved in data collection that make things look more variable than the usual assumptions might indicate; e.g. the subjects in the seizure study may have not kept accurate records of the number of seizures that they experienced during a particular period, and perhaps recalled it as being greater or less than it actually was. This is usually not a problem for binary data, since it is generally easy to reliably record whether the event of interest occurred.

Theses issue could make the variance in the population of all possible observations across all units appear to be more variable than expected. Note that the second issue could arise even in the cases considered in Chapter 11. The extension we are about to discuss may be applied to ordinary generalized linear regression modeling as well in this case.

The phenomenon where variance may be greater than that dictated by a standard model based on one of these distributions is called **overdispersion**. To take this phenomenon into account, it is customary to be a little more flexible about modeling overall variance in some of these models.

- For example, for **count** data, it is standard to **modify** the variance model to allow for an additional **scale** or **overdispersion** parameter; i.e.

$$\text{var}(Y_{ij}) = \phi E(Y_{ij}). \tag{12.5}$$

- For **binary data**, this is not generally required; if we wrote a model

$$\text{var}(Y_{ij}) = \phi E(Y_{ij})\{1 - E(Y_{ij})\},$$

  we would expect ϕ to be estimated as equal to 1, as the variance of a binary response should be just $E(Y_{ij})\{1 - E(Y_{ij})\}$

Fancier ways to deal with "overdispersion" are described in, for example McCullagh and Nelder (1989).

"WORKING" CORRELATION MATRIX: The last requirement is to specify a model describing **correlation** among pairs of observations on the same data vector. Again, because the modeling is of the **population-averaged** type, the model for correlation is attempting to represent how **all** sources of variation that could lead to associations among observations "add up," the aggregate of

- Correlation due to the within-subject "fluctuations" on a particular unit (and possibly measurement error).
- Correlation due to the simple fact the observations on the same unit are "more alike" than those from different units.

The models that are chosen to represent the overall correlation are the same ones used in modeling normally distributed data that were discussed in Chapter 8. In the current context one thinks of associations exclusively in terms of correlations, as the variance is modeled by thinking about it **separately** from associations. Popular models are the ones in Chapter 8, which we write here in terms of the correlation matrices they dictate:

- **Unstructured correlation:** For observations taken at the **same** time points for different units, this assumption places **no restriction** on the nature of associations among elements of a data vector. If Y_{ij} and Y_{ik}, $j, k = 1, \ldots, n$, are two observations on the same unit where all units are observed at the same n times, and if ρ_{jk} represents the correlation between Y_{ij} and Y_{ik}, then $\rho_{jk} = 1$ if $j = k$ and $-1 \leq \rho_{jk} \leq 1$ if $j \neq k$. The implied correlation matrix for a data vector with all n observations is the $(n \times n)$ matrix

$$\begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{21} & 1 & \cdots & \rho_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ \rho_{n1} & \cdots & \rho_{n,n-1} & 1 \end{pmatrix},$$

  where of course $\rho_{jk} = \rho_{kj}$ for all j, k. Thus, the unstructured "working" correlation assumption depends on $n(n-1)/2$ **distinct** correlation parameters.

- **Compound symmetry (exchangeable) correlation:** This assumption says that the correlation between distinct observations on the same unit is **the same** regardless of when in time the observations were taken. In principle, this model could be used with balanced data, ideally balanced data with missing values, and unbalanced data where time points are different for different units. This structure may be written in terms of a single correlation parameter $0 < \rho < 1$; i.e.

$$\begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \vdots & \vdots \\ \rho & \cdots & \rho & 1 \end{pmatrix}.$$

- **One-dependent:** This assumption says that only observations adjacent in time are correlated by the same amount $-1 < \rho < 1$. In principle, this model could be used with any situation; however, for unbalanced data with different time points, it may not make sense, as we discussed in Chapter 8. The model may be written

$$\begin{pmatrix} 1 & \rho & 0 & \cdots & 0 \\ \rho & 1 & \rho & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & \rho & 1 \end{pmatrix}.$$

- **AR(1) correlation:** This assumption says that correlation among observations "tails off;" if $-1 < \rho < 1$, the model is

$$\begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho^{n-1} & \cdots & \rho^2 & \rho & 1 \end{pmatrix}.$$

  In principle, this model could be used with any situation; however, again, for unbalanced data with different time points, it may not make sense.

Note that in the case of ideally balanced data, if some data vectors are missing some observations, then the forms of these matrices must be constructed carefully to reflect this, as discussed in Chapter 8. E.g., for $n = 5$ and a vector missing the observations corresponding to $j = 2$ and 4, the unstructured matrix would be constructed as

$$\begin{pmatrix} 1 & \rho_{13} & \rho_{15} \\ \rho_{13} & 1 & \rho_{35} \\ \rho_{15} & \rho_{35} & 1 \end{pmatrix},$$

where we have used the fact that $\rho_{jk} = \rho_{kj}$.

For unbalanced data where the observations on each unit are taken at possibly **different** times, the models such as the **Markov** model discussed in Chapter 8 may be used in the obvious way; currently, this capability is not part of `PROC GENMOD` in SAS. The examples we consider in this chapter are from longitudinal studies designed (ideally) to be balanced.

The correlation model so specified is popularly referred to in the context of these models as the "**working** correlation matrix." This designation is given because it is well-recognized that such modeling carries with it much **uncertainty**; as we have discussed, we are attempting to capture variance and correlation from **all** sources with a **single model**. Thus, the model is considered to be only a "working" model rather than necessarily representing what is probably a very complex truth. "Working" correlation became popular in the context of modeling longitudinal data with generalized linear models; however, it is equally applicable when discussing the the modeling of Chapter 8 in the normal case. Thus, although this term gained popularity in nonnormal data situations, it has come to be used in the linear, normal case, too. As we have seen in the linear, normal case, introducing random effects is an **alternative** way to generate covariance models that may have an easier time at capturing both sources of variation.

ALL TOGETHER: Combining the models for variance and correlation gives a model for the **covariance** matrix for a data vector $\boldsymbol{Y}_i$. It is customary to represent this in the "alternative" form in Equation (3.7). Suppose that unit i has a vector of associated **covariates**, possibly including time t_{ij}, $\boldsymbol{x}_{ij}$.

- It may well be the case that $\boldsymbol{x}_{ij}$ does not vary with j, or varies with j only through t_{ij}. In this case, covariates are **time-independent**.
- Following our previous discussion, it may be that $\boldsymbol{x}_{ij}$ includes **time-dependent** covariates. It may even include values of such covariates or even responses at other j!

Thus, the notation $\boldsymbol{x}_{ij}$ is meant to include all components deemed relevant at j.

We write the mean response model as

$$\mu_{ij} = E(Y_{ij}) = f(\boldsymbol{x}_{ij}'\boldsymbol{\beta}),$$

where f is one of the functions such as the exponential (loglinear) or logistic regression models. Then the variance of Y_{ij} is modeled by some function of the mean response μ_{ij}; e.g.

$$\text{var}(Y_{ij}) = \phi V(\mu_{ij}),$$

where we include a dispersion parameter ϕ. The **standard deviation** of Y_{ij} is given by $\{\phi V(\mu_{ij})\}^{1/2}$.

Suppose that unit i has n_i observations, so that $j = 1, \ldots, n_i$. Define the **standard deviation** matrix for unit i as the $(n_i \times n_i)$ diagonal matrix whose diagonal elements are the standard deviations of the Y_{ij} under this model, except for the dispersion parameter; that is, let

$$\boldsymbol{T}_i^{1/2} = \begin{pmatrix} \{V(\mu_{i1})\}^{1/2} & 0 & \cdots & 0 \\ 0 & \{V(\mu_{i2})\}^{1/2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & \{V(\mu_{in_i})\}^{1/2} \end{pmatrix}. \tag{12.6}$$

Let $\boldsymbol{\Gamma}_i$ be the $(n_i \times n_i)$ **correlation** matrix under one of the assumptions above, properly constructed for this unit's time pattern. Then we may write the **covariance matrix** $\boldsymbol{\Sigma}_i$ for the data vector $\boldsymbol{Y}_i$ implied by the assumptions as (verify)

$$\boldsymbol{\Sigma}_i = \phi \boldsymbol{T}_i^{1/2} \boldsymbol{\Gamma}_i \boldsymbol{T}_i^{1/2};$$

note that we have multiplied by the overdispersion parameter $\phi = \phi^{1/2}\phi^{1/2}$ to complete the specification of the standard deviations in each matrix $\boldsymbol{T}_i^{1/2}$.

Note that the "i" subscript is needed on both $\boldsymbol{T}_i^{1/2}$ and $\boldsymbol{\Gamma}_i$ to remind us that the dimensions of these matrices and the diagonal elements of $\boldsymbol{T}_i^{1/2}$ depend on the particular unit i with its own mean response vector and number of observations n_i.

SUMMARY: We may now summarize the modeling strategy and resulting statistical model. To specify a population-averaged model for mean and covariance matrix of a data vector for nonnormal responses using this approach:

- The **mean response** of a data vector $\boldsymbol{Y}_i$ is modeled as a function of time, other covariates, and parameters $\boldsymbol{\beta}$ by using a **generalized linear model**-type mean structure to represent the mean response of each element of $\boldsymbol{Y}_i$.
- The **variance** of each element of $\boldsymbol{Y}_i$ is modeled by the function of the mean that is appropriate for the type of data; e.g. count data are taken to have the Poisson variance structure, which says that variance of any element of $\boldsymbol{Y}_i$ is equal to the corresponding model for the mean. These models are often modified to allow for the greater variation both within- and among-units by the addition of a **dispersion** parameter ϕ.
- **Correlation** among observations on the same unit (elements of $\boldsymbol{Y}_i$) is represented by choosing a model, such as the correlation structures corresponding to the AR(1), one-dependent, Markov, or other specifications. Because there is some uncertainty in doing this and (as we'll see) no formal way to check it, the chosen model is referred to as the "**working correlation matrix**" to emphasize this fact.

With these considerations, we have the following statistical model for the mean vector and covariance matrix of a data vector $\boldsymbol{Y}_i$ consisting of observations Y_{ij}, $j = 1, \ldots, n_i$ on unit i. If

- Mean response of Y_{ij} is modeled by a suitable function f of a **linear predictor** $\boldsymbol{x}_{ij}'\boldsymbol{\beta}$
- Variance is thus modeled as some function V of mean response times a dispersion parameter ϕ, which defines a standard deviation matrix $\boldsymbol{T}_i^{1/2}$ as in (12.6) above,
- Correlation is modeled by a "working" correlation assumption $\boldsymbol{\Gamma}_i$

$$E(\boldsymbol{Y}_i) = \begin{pmatrix} f(\boldsymbol{x}_{i1}'\boldsymbol{\beta}) \\ f(\boldsymbol{x}_{i2}'\boldsymbol{\beta}) \\ \vdots \\ f(\boldsymbol{x}_{in_i}'\boldsymbol{\beta}) \end{pmatrix} = \boldsymbol{f}_i(\boldsymbol{\beta}), \qquad \text{var}(\boldsymbol{Y}_i) = \phi \boldsymbol{T}_i^{1/2} \boldsymbol{\Gamma}_i \boldsymbol{T}_i^{1/2} = \boldsymbol{\Sigma}_i = \phi \boldsymbol{\Lambda}_i. \tag{12.7}$$

Let $\boldsymbol{\omega}$ refer to the distinct **unknown** parameters that fully describe the chosen "working" correlation matrix $\boldsymbol{\Gamma}_i$. For example, for the compound symmetry, AR(1), and one-dependent structure, $\boldsymbol{\omega} = \rho$; for the unstructured model, $\boldsymbol{\omega}$ consists of the **distinct** possible correlation parameters ρ_{jk} for the data vector of maximal size n.

As always, it is assumed that the individual data vectors $\boldsymbol{Y}_i$ are **independent** across individual units.

As noted above, however, we are not in a position to specify a full multivariate probability distribution corresponding to this mean and covariance model.

12.3 Generalized estimating equations

The considerations in the last section allow specification of a model for the mean and covariance of a data vector of the form (12.7). However, because this is not sufficient to specify an entire appropriate multivariate probability distribution, it is **not possible** to appeal immediately to the principle of **maximum likelihood** to develop a framework for estimation and testing.

IDEA: Although we do not have a basis for the maximum likelihood, why not try to emulate situations where there is such a basis? We have two situations to which we can appeal:

- The normal case with a **linear** mean model, discussed in Chapter 8. Here, the model was

$$E(\boldsymbol{Y}_i) = \boldsymbol{X}_i\boldsymbol{\beta}, \quad \text{var}(\boldsymbol{Y}_i) = \boldsymbol{\Sigma}_i$$

  for suitable choice of covariance matrix $\boldsymbol{\Sigma}_i$ depending on a vector of parameters $\boldsymbol{\omega}$, say. Assuming that the $\boldsymbol{Y}_i$ follow a multivariate normal, we were led to the estimator for $\boldsymbol{\beta}$

$$\widehat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{m} \boldsymbol{X}_i' \widehat{\boldsymbol{\Sigma}}_i^{-1} \boldsymbol{X}_i \right)^{-1} \sum_{i=1}^{m} \boldsymbol{X}_i' \widehat{\boldsymbol{\Sigma}}_i^{-1} \boldsymbol{Y}_i, \tag{12.8}$$

  where $\widehat{\boldsymbol{\Sigma}}_i$ is the covariance matrix with the estimator for $\boldsymbol{\omega}$ plugged in. It may be shown (try it!) that it is possible to **rewrite** (12.8) in the following form:

$$\sum_{i=1}^{m} \boldsymbol{X}_i' \widehat{\boldsymbol{\Sigma}}_i^{-1} (\boldsymbol{Y}_i - \boldsymbol{X}_i \widehat{\boldsymbol{\beta}}) = \boldsymbol{0}. \tag{12.9}$$

  That is, the estimator for $\boldsymbol{\beta}$ solves an a set of p **equations** for $\boldsymbol{\beta}$ ($p \times 1$) (with the estimator for $\boldsymbol{\omega}$ plugged in).

- In the case of ordinary generalized linear models, recall that considering maximum likelihood, which was possible in that case, led to solving a set of equations of the form (11.18); i.e.

$$\sum_{j=1}^{n} \frac{1}{V\{f(\boldsymbol{x}_j'\boldsymbol{\beta})\}}\{Y_j - f(\boldsymbol{x}_j'\boldsymbol{\beta})\}f'(\boldsymbol{x}_j'\boldsymbol{\beta})\boldsymbol{x}_j = \mathbf{0}, \tag{12.10}$$

  where $f'(u) = \dfrac{d}{du}f(u)$, the derivative of f with respect to its argument. The method of **iteratively reweighted least squares** was used to solve this equation. Note that if there is a scale parameter, it need not be taken into account in this calculation.

- Comparing (12.9) and (12.10), we see that there is a similar theme – the equations are **linear** functions of **deviations** of observations from their assumed mean are **weighted** in accordance with their covariance (for vectors) and variance (for individual observations). The variance or covariance matrix is not entirely known but is evaluated at estimates of the unknown quantities it contains ($\boldsymbol{\omega}$ in the first case and $\boldsymbol{\beta}$ in the second case).

GENERALIZED ESTIMATING EQUATION: From these observations, a natural approach for fitting model (12.7) is suggested: solve an **estimating equation** consisting of p equations for $\boldsymbol{\beta}$ ($p \times 1$) that (i) is a **linear** function of **deviations**

$$\boldsymbol{Y}_i - \boldsymbol{f}_i(\boldsymbol{\beta}),$$

and (ii) **weights** these deviations in the same way as in (12.9) and (12.10), using the inverse of the assumed covariance matrix $\boldsymbol{\Sigma}_i$ of a data vector with an estimator for the unknown parameters $\boldsymbol{\omega}$ in the "working" correlation matrix plugged in.

Note that even if there is a scale parameter, we really need only use the inverse of $\boldsymbol{\Lambda}_i$ in (12.7). As in (12.10), $\boldsymbol{\Sigma}_i$ and $\boldsymbol{\Lambda}_i$ will **also** depend on $\boldsymbol{\beta}$ through the variance functions $V\{f(\boldsymbol{x}_{ij}'\boldsymbol{\beta})\}$; more in a moment.

These results lead to consideration of the following equation to be solved for $\boldsymbol{\beta}$ (with a suitable estimator for $\boldsymbol{\omega}$ plugged in):

$$\sum_{i=1}^{m} \boldsymbol{\Delta}_i'\hat{\boldsymbol{\Lambda}}_i^{-1}\{\boldsymbol{Y}_i - \boldsymbol{f}_i(\widehat{\boldsymbol{\beta}})\} = \mathbf{0}, \tag{12.11}$$

where $\boldsymbol{\Delta}_i$ is the $(n_i \times p)$ matrix whose (j, s) element ($j = 1, \ldots, n_i$, $s = 1, \ldots, p$) is the derivative of $f(\boldsymbol{x}_{ij}'\boldsymbol{\beta})$ with respect to the sth element of $\boldsymbol{\beta}$, and $\hat{\boldsymbol{\Lambda}}_i$ is the matrix $\boldsymbol{\Lambda}_i$ in (12.7) with an estimator for $\boldsymbol{\omega}$ plugged in (see below). Note that ϕ can be disregarded here.

The matrix $\boldsymbol{\Delta}_i$ is a function of $\boldsymbol{\beta}$. It is also a function of $\boldsymbol{X}_i$, which here is defined as the $(n_i \times p)$ matrix whose rows are $\boldsymbol{x}_{ij}'$. It is possible to write out the form of $\boldsymbol{\Delta}_i$ precisely in terms of $\boldsymbol{X}_i$ and the elements $f'(\boldsymbol{x}_{ij}'\boldsymbol{\beta})$; this is peripheral to our discussion here; see Liang and Zeger (1986) for the gory details.

An equation of the form (12.11) to be solved to estimate a parameter $\boldsymbol{\beta}$ in a mean response model is referred to popularly as a **generalized estimating equation**, or GEE for short.

ESTIMATION OF $\boldsymbol{\omega}$: To use (12.11) to estimate $\boldsymbol{\beta}$, an estimator for $\boldsymbol{\omega}$ is required. There are a number of methods that have been proposed to obtain such estimators; the books by Diggle, Heagerty, Liang, and Zeger (2002) and Vonesh and Carter (1997) discuss this in detail. One intuitive way, and that used by `PROC GENMOD` in SAS and originally proposed by Liang and Zeger (1986), is to base the estimation on appropriate functions of deviations

$$\boldsymbol{Y}_i - \boldsymbol{f}_i(\widehat{\boldsymbol{\beta}}),$$

where $\widehat{\boldsymbol{\beta}}$ is some estimator for $\boldsymbol{\beta}$.

- For example, one could fit the mean model for all m individuals assuming **independence** among **all** observations using the techniques of Chapter 11 to obtain such an estimate. This estimate could be used to form deviations and thus to estimate $\boldsymbol{\omega}$.

To see how this might work, let

$$r_{ij} = \frac{Y_{ij} - f(\boldsymbol{x}_{ij}'\widehat{\boldsymbol{\beta}})}{[V\{f(\boldsymbol{x}_{ij}'\widehat{\boldsymbol{\beta}})\}^{1/2}}$$

be the deviation corresponding to the jth observation on unit i divided by an estimate of its standard deviation. Then the dispersion parameter ϕ is usually estimated by

$$\widehat{\phi} = (N-p)^{-1}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\frac{\{Y_{ij} - f(\boldsymbol{x}_{ij}'\widehat{\boldsymbol{\beta}})\}^2}{V\{f(\boldsymbol{x}_{ij}'\widehat{\boldsymbol{\beta}})\}} = (N-p)^{-1}\sum_{i=1}^{m}\sum_{j=1}^{n_i} r_{ij}^2. \tag{12.12}$$

Compare this to the **Pearson chi-square** in ordinary generalized linear models in Chapter 11; it is the same function but taken across **all** deviations for all units.

- If $\boldsymbol{\Gamma}_i$ corresponds to the **unstructured** correlation assumption, then estimate ρ_{jk} by

$$\widehat{\rho}_{jk} = m^{-1}\widehat{\phi}^{-1}\sum_{i=1}^{m} r_{ij}r_{ik}.$$

- If $\boldsymbol{\Gamma}_i$ corresponds to the **compound symmetry** structure, then the single parameter ρ may be estimated by

$$\widehat{\rho} = m^{-1}\widehat{\phi}^{-1}\sum_{i=1}^{m}(n_i-1)^{-1}\sum_{j=1}^{n_i-1} r_{ij}r_{i,j+1}.$$

  Note that the rationale here is to consider only **adjacent** pairs, as you might expect.

$\boldsymbol{\omega}$ for other covariance models may be estimated by a similar approach.

ALL TOGETHER: The above ideas may be combined to define an estimation scheme for $\boldsymbol{\beta}$, $\boldsymbol{\omega}$, and ϕ in the model (12.7). Heuristically, the scheme has the following form:

1. Obtain an initial estimator for $\boldsymbol{\beta}$ by assuming all observations across all individuals are **independent**. This may be carried out using the method of IRWLS for ordinary generalized linear models, as described in Chapter 11.

2. Using this estimator for $\boldsymbol{\beta}$, estimate ϕ and then $\boldsymbol{\omega}$ as appropriate for the assumed "working" correlation matrix.

3. Use these estimators for $\boldsymbol{\beta}$ and $\boldsymbol{\omega}$ to form an estimate of $\boldsymbol{\Lambda}_i$, $\hat{\boldsymbol{\Lambda}}_i$. Treat this as fixed in the generalized estimating equation (12.11). The resulting equation may then be solved by a numerical technique that is an **extended version** of the IRWLS method used in the ordinary case. Obtain a new estimator $\widehat{\boldsymbol{\beta}}$.

4. Return to step 2 if desired and repeat the process. Steps 2, 3, and 4 can be repeated until the results of two successive tries stay the same ("convergence").

The spirit of this scheme is implemented in the SAS procedure `PROC GENMOD`.

SAMPLING DISTRIBUTION: As before, it should not be surprising that we must appeal to **large sample theory** to obtain an approximation to the **sampling distribution** of the estimator $\widehat{\boldsymbol{\beta}}$ obtained by solving the GEE. Here, "large sample" refers to the number of units, m; this is sensible; each $\boldsymbol{Y}_i$ is from a different unit.

The results may be stated as follows: For m "large," the GEE estimator $\widehat{\boldsymbol{\beta}}$ for $\boldsymbol{\beta}$ satisfies

$$\widehat{\boldsymbol{\beta}} \dot{\sim} \mathcal{N}\left\{\boldsymbol{\beta}, \phi\left(\sum_{i=1}^{m} \boldsymbol{\Delta}_i' \boldsymbol{\Lambda}_i^{-1} \boldsymbol{\Delta}_i\right)^{-1}\right\}, \tag{12.13}$$

where $\boldsymbol{\Delta}_i$ is as defined previously. As in the ordinary generalized linear model case, $\boldsymbol{\Delta}_i$ and $\boldsymbol{\Lambda}_i$ depend on $\boldsymbol{\beta}$ and $\boldsymbol{\omega}$; moreover, ϕ is also unknown. Thus, for practical use, these quantities are replaced by estimates. Specifically, define

$$\widehat{\boldsymbol{V}}_\beta = \hat{\phi}\left(\sum_{i=1}^{m} \widehat{\boldsymbol{\Delta}}_i' \hat{\boldsymbol{\Lambda}}_i^{-1} \widehat{\boldsymbol{\Delta}}_i\right)^{-1},$$

where $\widehat{\boldsymbol{\Delta}}_i$ and $\hat{\boldsymbol{\Lambda}}_i$ are $\boldsymbol{\Delta}_i$ and $\boldsymbol{\Lambda}_i$ with the final estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\omega}$ plugged in and $\hat{\phi}$ is the estimate of ϕ. $\hat{\phi}$ would just be equal to 1 if no scale parameter is in the model. Again, we use the notation $\widehat{\boldsymbol{V}}_\beta$ to represent the estimated covariance matrix of $\widehat{\boldsymbol{\beta}}$.

As usual, **standard errors** for the elements of $\widehat{\boldsymbol{\beta}}$ may be obtained as the square roots of the diagonal elements of $\widehat{\boldsymbol{V}}_\beta$.

HYPOTHESIS TESTS: As in the ordinary generalized linear model case, **Wald** testing procedures are used to test null hypotheses of the form

$$H_0 : \boldsymbol{L}\boldsymbol{\beta} = \boldsymbol{h}.$$

As usual, we have the large sample approximation

$$\boldsymbol{L}\widehat{\boldsymbol{\beta}} \dot{\sim} \mathcal{N}(\boldsymbol{L}\boldsymbol{\beta}, \boldsymbol{L}\widehat{\boldsymbol{V}}_\beta\boldsymbol{L}'),$$

which may be used to construct test statistics and confidence intervals in a fashion identical to that discussed previously; for example, if $\boldsymbol{L}$ is a row vector, then the test may be based on comparing

$$z = \frac{\boldsymbol{L}\widehat{\boldsymbol{\beta}} - \boldsymbol{h}}{SE(\boldsymbol{L}\widehat{\boldsymbol{\beta}})}$$

to the critical values from the standard normal distribution. For more general $\boldsymbol{L}$, one may form the Wald χ^2 statistic More generally, the Wald χ^2 test statistic

$$(\boldsymbol{L}\widehat{\boldsymbol{\beta}} - \boldsymbol{h})'(\boldsymbol{L}\widehat{\boldsymbol{V}}_\beta\boldsymbol{L}')^{-1}(\boldsymbol{L}\widehat{\boldsymbol{\beta}} - \boldsymbol{h})$$

and compare to the appropriate χ^2 critical value with degrees of freedom equal to the number of rows of $\boldsymbol{L}$.

12.4 "Robust" estimator for sampling covariance

ISSUE: It is important to recognize that the GEE fitting method for estimating the parameters in model (12.7) is **not** a maximum likelihood method; rather, it was arrived at from an *ad hoc* perspective. As a result, it is not possible to derive quantities like AIC and BIC to compare different "working" correlation matrices to determine which assumption is most suitable. Consequently, it is sensible to be concerned that the validity of inferences on $\boldsymbol{\beta}$ such as the estimator itself, calculation of approximate confidence intervals, and tests may be compromised if the assumption on correlation is incorrect.

SOLUTION: One solution to this dilemma is to **modify** the estimated covariance matrix $\widehat{\boldsymbol{V}}_\beta$ to allow for the possibility that the choice of $\boldsymbol{\Gamma}_i$ used in the model is **incorrect**. The modified version of $\widehat{\boldsymbol{V}}_\beta$ is

$$\widehat{\boldsymbol{V}}_\beta^R = \left(\sum_{i=1}^m \widehat{\boldsymbol{\Delta}}_i'\hat{\boldsymbol{\Lambda}}_i^{-1}\widehat{\boldsymbol{\Delta}}_i\right)^{-1}\left(\sum_{i=1}^m \widehat{\boldsymbol{\Delta}}_i'\hat{\boldsymbol{\Lambda}}_i^{-1}\widehat{\boldsymbol{S}}_i\hat{\boldsymbol{\Lambda}}_i^{-1}\widehat{\boldsymbol{\Delta}}_i\right)\left(\sum_{i=1}^m \widehat{\boldsymbol{\Delta}}_i'\hat{\boldsymbol{\Lambda}}_i^{-1}\widehat{\boldsymbol{\Delta}}_i\right)^{-1}, \tag{12.14}$$

where

$$\widehat{\boldsymbol{S}}_i = \{\boldsymbol{Y}_i - \boldsymbol{f}_i(\widehat{\boldsymbol{\beta}})\}\{\boldsymbol{Y}_i - \boldsymbol{f}_i(\widehat{\boldsymbol{\beta}})\}'.$$

- Even if the model has a scale parameter. (12.14) does not require an estimate of it.
- Note that if $\widehat{\boldsymbol{S}}_i$ were equal to $\widehat{\boldsymbol{\Sigma}}_i = \widehat{\phi}\hat{\boldsymbol{\Lambda}}_i$, then (12.14) would be equivalent to $\widehat{\boldsymbol{V}}_\beta$ (verify).
- The rationale for the modification may be appreciated by considering the definition of the **true** covariance matrix for $\boldsymbol{Y}_i$; specifically,

  $$\text{var}(\boldsymbol{Y}_i) = E\{\boldsymbol{Y}_i - \boldsymbol{f}_i(\boldsymbol{\beta})\}\{\boldsymbol{Y}_i - \boldsymbol{f}_i(\boldsymbol{\beta})\}'.$$

  In the model, we have chosen $\boldsymbol{\Sigma}_i$ (through choosing $\boldsymbol{\Gamma}_i$ as our assumption about var($\boldsymbol{Y}_i$). By including the "middle" term in (12.14), we are thus hoping to "balance out" an alternative guess for var($\boldsymbol{Y}_i$) against the assumed model $\boldsymbol{\Sigma}_i$.
- It turns out that, for large m, $\widehat{\boldsymbol{V}}_\beta^R$ will provide a reliable estimate of the true sampling covariance matrix of $\widehat{\boldsymbol{\beta}}$ **even if** the chosen model $\boldsymbol{\Sigma}_i$ ($\boldsymbol{\Gamma}_i$) is incorrect. In contrast, if the model is incorrect, $\widehat{\boldsymbol{V}}_\beta$ will **not** provide a reliable estimate.

The alternative estimate of the sampling covariance matrix of $\widehat{\boldsymbol{\beta}}$ $\widehat{\boldsymbol{V}}_\beta^R$ is often referred to as the **robust** covariance matrix estimate. The term is derived from the fact that $\widehat{\boldsymbol{V}}_\beta^R$ is "robust" to the fact that we may be incorrect about $\boldsymbol{\Gamma}_i$. $\widehat{\boldsymbol{V}}_\beta$ is often referred to as the **model-based** covariance matrix estimate, because it uses the model assumption on $\boldsymbol{\Gamma}_i$ with no attempt to correct for the possibility it is wrong.

This "robust" modification may also be applied to the linear, normal models in Chapter 8. To get "robust" standard errors, use the `empirical` option in the `proc mixed` statement: `proc mixed empirical data=`;

The decision whether to use the **model-based** estimate $\widehat{\boldsymbol{V}}_\beta$ or the **robust** estimate $\widehat{\boldsymbol{V}}_\beta^R$ is an "art-form." No consensus exists on which one is to be preferred in **finite** samples in practical problems. If they are **very** different, some people take that as an indication that the original assumption is wrong. On the other hand, if one or more of the $\boldsymbol{Y}_i$ vectors contains "unusual" values that are very unlikely to be seen, this would be enough to "throw off" the estimate $\widehat{\boldsymbol{V}}_\beta^R$. Because there is no "iron-clad" rule, we offer no recommendation on which to use.

12.5 Contrasting population-averaged and subject-specific approaches

The model (12.7) is, as stated, a **population-averaged** model. The mean of a data vector and its covariance matrix are modeled **explicitly**. As a result, from our discussions in Chapter 9, we know that $\boldsymbol{\beta}$ has the interpretation as the parameters that describe the relationship of the **mean response** over time and other covariates.

An alternative perspective we discussed was that of the **subject-specific** approach. In this approach, one starts with thinking about **individual unit trajectories** rather than about the mean (average) across all units. In the linear model case, we did this by the introduction of **random effects**; e.g., the **random coefficient** model that says each unit has its own intercept and slope β_{0i} and β_{1i}, which in turn are represented as

$$\beta_{0i} = \beta_0 + b_{0i}, \quad \beta_{1i} = \beta_1 + b_{1i}, \qquad\qquad \boldsymbol{\beta} = (\beta_0, \beta_1)'.$$

In this model, the interpretation of $\boldsymbol{\beta}$ is as the "typical" value of intercept and slope in the population.

It just so happened that in the case of a **linear** model for either the mean response or individual trajectory, one arrives at the same mean response model. Thus, in this case, the distinction between these two interpretations was not important – either was valid.

SUBJECT-SPECIFIC GENERALIZED LINEAR MODEL: It is natural to consider the **subject-specific** approach in the case where the functions of generalized linear models are appropriate. For example, recall the seizure data, where the response is a **count**. By analogy to linear random coefficient and mixed effects models, suppose we decided to model the **individual trajectory** of counts for an individual subject as a **subject-specific** loglinear regression model. That is, suppose we wrote the "mean" for subject i as a function of subject-specific parameters β_{0i} and β_{3i} as

$$\exp(\beta_{0i} + \beta_{3i} t_{ij}) \tag{12.15}$$

In (12.15), β_{0i} and β_{3i} thus describe the subject's **own** (conditional) mean response as a function of time and **individual** "intercept" and "slope" on the log scale. Under this perspective, each subject has his/her own such parameters β_{0i} and β_{3i} that characterize his/her own mean response over time.

Now, just as we did earlier, suppose we thought of the β_{0i} and β_{4i} as arising from **populations** of such values. For example, suppose that

$$\beta_{3i} = \beta_3 + b_{3i},$$

where b_{3i} is a **random effect** for subject i with mean 0. b_{3i} describes how subject i deviates from the "typical" value β_3. Similarly, we might suppose that

$$\beta_{0i} = \beta_0 + b_{0i}$$

for another mean-zero random effect b_{0i}.

To incorporate the **covariate** information on treatment and age, we might assume that the "typical" **rate of change** of log mean with time does not depend on these covariates, but maybe the "typical" **intercept** does; e.g., we could write an alternative model depending on covariates a_i and δ_i, say, as

$$\beta_{0i} = \beta_0 + \beta_1 a_i + \beta_2 \delta_i + b_{0i}.$$

Putting all of this together, we arrive at a model for the "mean" for subject i, depending on the **random effect** vector $\boldsymbol{b}_i = (b_{0i}, b_{3i})'$:

$$E(Y_{ij} \,|\, \boldsymbol{b}_i) = \exp(\beta_0 + \beta_1 a_i + \beta_2 \delta_i + b_{0i} + \beta_3 t_{ij} + b_{3i} t_{ij}) \tag{12.16}$$

Following with the analogy, we could assume that the **random effects** $\boldsymbol{b}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{D})$ for some covariance matrix $\boldsymbol{D}$.

We could write this model another way. Let $\boldsymbol{\beta}_i = (\beta_{0i}, \beta_{3i})$. The we have a **first-stage** model that says the **conditional mean** for $\boldsymbol{Y}_i$, given $\boldsymbol{b}_i$ on which $\boldsymbol{\beta}_i$ depends is $\boldsymbol{f}_i(\boldsymbol{\beta}_i)$, where

$$\boldsymbol{f}_i(\boldsymbol{\beta}_i) = \begin{pmatrix} \exp(\beta_{0i} + \beta_{3i} t_{i1}) \\ \vdots \\ \exp(\beta_{0i} + \beta_{3i} t_{in_i}) \end{pmatrix}.$$

At the **second stage**, we could assume

$$\boldsymbol{\beta}_i = \boldsymbol{A}_i \boldsymbol{\beta} + \boldsymbol{b}_i;$$

for the model above, $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)'$ and, for subject i

$$\boldsymbol{A}_i = \begin{pmatrix} 1 & a_i & \delta_i & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

(Verify.)

ARE THE TWO MODELS THE SAME? All of this is very similar to what we did in the normal, linear case. In that case, both approaches led to the **same** representation of the ultimate mean response vector $E(\boldsymbol{Y}_i)$, but with different covariance matrices. The population-averaged model for mean response is $E(\boldsymbol{Y}_i) = \boldsymbol{X}_i \boldsymbol{\beta}$. In the subject-specific general linear mixed model, by contrast, the "individual mean" is

$$E(\boldsymbol{Y}_i \,|\, \boldsymbol{b}_i) = \boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{Z}_i \boldsymbol{b}_i. \tag{12.17}$$

But this "individual mean" has expectation

$$E\{\boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{Z}_i \boldsymbol{b}_i\} = \boldsymbol{X}_i \boldsymbol{\beta},$$

since $\boldsymbol{b}_i$ has mean zero, which is **identical** to the population-averaged model.

Here, our two competing models are the **population-averaged** model that says immediately that $E(\boldsymbol{Y}_i)$ has jth element

$$E(Y_{ij}) = \exp(\beta_0 + \beta_1 a_i + \beta_2 \delta_i + \beta_3 t_{ij}),$$

and, from (12.16), the **subject-specific** model that says $E(\boldsymbol{Y}_i \,|\, \boldsymbol{b}_i)$ has jth element

$$\exp(\beta_0 + \beta_1 a_i + \beta_2 \delta_i + b_{0i} + \beta_3 t_{ij} + b_{3i} t_{ij}).$$

If the models were **the same**, we would expect that the expectation of this would be **identical** to $E(Y_{ij})$ above. **However**, this is **not** the case. Note that we need to evaluate

$$E\left\{\exp(\beta_0 + \beta_1 b_i + \beta_2 a_i + \beta_3 \delta_i + b_{0i} + \beta_3 t_{ij} + b_{3i} t_{ij})\right\}.$$

Contrast this with the calculation in (12.17) above – because that function of $\boldsymbol{b}_i$ was **linear**, evaluating the expectation was straightforward. Here, however, evaluating the expectation is **not** straightforward, because it involves a complicated **nonlinear** function of $\boldsymbol{b}_i = (b_{0i}, b_{3i})'$. Even though $\boldsymbol{b}_i$ are normal, the expectation of this nonlinear function is not possible to evaluate by a simple rule as in the linear case. As a result, it is **not true** that the expectation is identical to $E(Y_{ij})$ above.

RESULT: This is a general phenomenon, although we showed it just for a specific model. In a **nonlinear** model, it is **no longer true** that the population-averaged and subject-specific perspectives lead to the **same** model for mean response $E(\boldsymbol{Y}_i)$. Thus, the two models are **different**. Furthermore, the parameter we called $\boldsymbol{\beta}$ in each model has a **different** interpretation; e.g. in the seizure example,

- $\boldsymbol{\beta}$ for the population-averaged model has the interpretation as the value that leads to the "typical" or mean response vector
- $\boldsymbol{\beta}$ for the subject-specific model has the interpretation as the value that is the "typical" value of "intercept" and "slope" of log mean.

This may seem like a subtle and difficult-to-understand difference, which it is. But the main point is that the two different modeling strategies lead to two different ways to describe the data with different interpretations. Obviously, in these more complex models, the distinction **matters**. See Chapter 13 for more.

12.6 Discussion

The presentation here just scratches the surface of the area of population-averaged modeling for longitudinal data that may not be normally distributed. In fact, this is still an area of active research, and papers on the subject may be found in current issues of *Journal of the American Statistical Association*, *Biometrics*, and others. See the books by Diggle, Liang, and Zeger (1995) and Vonesh and Carter (1997) for more extensive treatment.

12.7 Implementation with SAS

We illustrate how to carry out fitting of population-averaged generalized linear models for longitudinal data via the use of generalized estimating equations for the two examples discussed in this chapter:

1. The epileptic seizure data
2. Wheezing data from the Six Cities study

our main focus is on the use of `PROC GENMOD` to fit models like those in the examples. We show how to specify different "working" correlation models via the `repeated` statement in this procedure, both for balanced (the seizure data) and unbalanced (the wheezing data) cases and how to interpret the output.

ASIDE: It is possible to implement this fitting, and more variations on it, in SAS in other ways – one possibility is through use of the `GLIMMIX` SAS macro, developed at SAS, that is meant to be used for fitting **generalized linear mixed models**, which are **subject-specific** models for nonnormal longitudinal data incorporating random effects, as the name suggests (see Chapter 13). This is similar in spirit to using `PROC MIXED` to fit linear population-averaged regression models to normal data; these models contain no random effects, yet this procedure may be used to fit them, as we have seen. The details are beyond the scope of this course.

EXAMPLE 1 – EPILEPTIC SEIZURE DATA: We first consider the model (12.1),

$$E(Y_{ij}) = \exp(\log o_{ij} + \beta_0 + \beta_1 v_{ij} + \beta_2 \delta_i + \beta_3 v_{ij}\delta_i),$$

discussed earlier. We fit this model using several working correlation matrices. Here, the coefficient of greatest interest is β_3, which reflects whether post-baseline mean response is different in the two treatment groups.

There is one "unusual" subject (subject 207 in the progabide group) whose seizure counts are very high; this subject had a baseline count of 151 in the 8 week pre-treatment period. This subject's data are sufficiently unusual relative to those for the rest of the participants that it is natural to be concerned over whether the conclusions are sensitive to them. To investigate, we fit the model excluding the data for this subject.

Finally, we also allow for the possibility that the mean response changes at the 4th visit and include age as a covariate to take account of possible association of baseline seizure characteristics with age of the subject. For the first issue, we define an additional indicator variable $v4_{ij} = 0$ unless $j = 5$ corresponding to the visit 4. The model is modified to

$$E(Y_{ij}) = \exp(\log o_{ij} + \beta_0 + \beta_1 v_{ij} + \beta_2 \delta_i + \beta_3 v_{ij}\delta_i + \beta_4 v4_{ij} + \beta_5 v4_{ij}\delta_i).$$

The parameter β_5 reflects whether the difference in post-baseline mean response in fact changes at the fourth visit, while β_4 allows the possibility that the mean response "shifts" at the 4th visit relative to the earlier ones.

To incorporate o_{ij}, in the program we use the `offset` option in the `model` statement of `proc genmod`.

PROGRAM:

```
/*****************************************************************

  CHAPTER 12, EXAMPLE 1

  Fit a loglinear regression model to the epileptic seizure data.
  These are count data, thus we use the Poisson mean/variance
  assumptions.  This model is fitted with different working
  correlation matrics.

*****************************************************************/

options ls=80 ps=59 nodate; run;

/*****************************************************************

  The data look like (first 8 records on first 2 subjects)

       104 11 0  0 11 31
       104  5 1  0 11 31
       104  3 2  0 11 31
       104  3 3  0 11 31
       104  3 4  0 11 31
       106 11 0  0 11 30
       106  3 1  0 11 30
       106  5 2  0 11 30
       106  3 3  0 11 30
       106  3 4  0 11 30
          .
          .
          .

  column 1      subject
  column 2      number of seizures
  column 3      visit (baseine (0) and 1--4 biweekly visits)
  column 4      =0 if placebo, = 1 if progabide
  column 5      baseline number of seizures in 8 weeks prior to study
  column 6      age

*****************************************************************/

data seizure; infile 'seize.dat';
  input subject seize visit trt base age;
run;


/*****************************************************************

  Fit the loglinear regression model using PROC GENMOD and
  three different working correlation matrix assumptions:

  -  unstructured
  -  compound symmetry (exchangeable)
  -  AR(1)

  Subject 207 has what appear to be very unusual data -- for
  this subject, both baseline and study-period numbers of seizures
  are huge, much larger than any other subject.  In some published
  analyses, this subjectis deleted.  See Diggle, Heagerty, Liang,
  and Zeger (2002) and Thall and Vail (1990) for more on this subject.
  We carry out the analyses with and without this subject.

  We fit the mean model in equation (12.1) first. We then add age
  as a covariate to allow for systematic differences in baseline response
  due to age.  We use log(age) as has been the case in other analyses.

  The DIST=POISSON option in the model statement specifies
  that the Poisson requirement that mean = variance, be used.
  The LINK=LOG option asks for the loglinear model.  Other
  LINK= choices are available.

  The REPEATED statement specifies the "working" correlation
  structure to be assumed.    The CORRW option in the REPEATED
  statement prints out the estimated working correlation matrix
  under the assumption given in the TYPE= option.  The COVB
  option prints out the estimated covariance matrix of the estimate
  of beta -- both the usual estimate and the "robust" version
  are printed.  The MODELSE option specifies that the standard
  error estimates printed for the elements of betahat are based
  on the usual theory.  By default, the ones based on the "robust"
  version of the sampling covariance matrix are printed, too.

  The dispersion parameter phi is estimated rather then being held
  fixed at 1 -- this allows for the possibility of "overdispersion"

  The new version of SAS will not allow the response to be a noninteger
  when we declare dist = poisson.  Thus, analyzing seize/o is not
```

```
  possible.  Instead, one can use the OFFSET option in the MODEL
  statement.  This will fit the model exactly how it is written in
  model (12.1) -- the term log(o_ij) is the known "offset."  To get
  SAS to include this "offset," we form the variable logo in the
  data set and then declare logo to be an offset.

*****************************************************************/

data seizure; set seizure;
   logage=log(age);
   o=2; v=1;
   if visit=0 then o=8;
   if visit=0 then v=0;
   logo=log(o);
run;

title "UNSTRUCTURED CORRELATION";
proc genmod data=seizure;
  class subject;
  model seize = v trt trt*v /  dist = poisson link = log offset=logo;
  repeated subject=subject / type=un corrw covb modelse;
run;


title "EXCHANGEABLE (COMPOUND SYMMETRY) CORRELATION";
proc genmod data=seizure;
  class subject;
  model seize = v trt trt*v /  dist = poisson link = log offset=logo;
  repeated subject=subject / type=cs corrw covb modelse;
run;


title "AR(1) CORRELATION";
proc genmod data=seizure;
  class subject;
  model seize = v trt trt*v /  dist = poisson link = log offset=logo;
  repeated subject=subject / type=ar(1) corrw covb modelse;
run;

/*****************************************************************

  Delete the unusual subject and run again; we only use the
  compound symmetric covariance for the rest of the analyses.

*****************************************************************/


data weird; set seizure;
  if subject=207 then delete;
run;


title "SUBJECT 207 DELETED";
proc genmod data=weird;
  class subject;
  model seize = v trt trt*v /  dist = poisson link = log offset=logo;
  repeated subject=subject / type=cs corrw covb modelse;
run;

/*****************************************************************

  Now we fit two additional models on the full data (with 207).
  In the first, we add logage as a covariate.  In the second,
  we allow an additional shift at visit 4.  To do this,
  we define visit4 to be an indicator of the last visit.

*****************************************************************/

data seizure; set seizure;
  visit4=1;
  if visit<4 then visit4=0;
run;

title "AGE ADDED";
proc genmod data=seizure;
  class subject;
  model seize = logage v trt trt*v /  dist = poisson link = log offset=logo;
  repeated subject=subject / type=cs corrw covb modelse;
run;

title "MODIFIED MODEL";
proc genmod data=seizure;
  class subject;
  model seize = v visit4 trt trt*v trt*visit4 /
                 dist = poisson link = log offset=logo;
  repeated subject=subject / type=cs corrw covb modelse;
run;
```

OUTPUT: Following the output, we comment on a few aspects of the output.

```
                          UNSTRUCTURED CORRELATION                          1

                             The GENMOD Procedure

                              Model Information

                         Data Set              WORK.SEIZURE
                         Distribution               Poisson
                         Link Function                  Log
                         Dependent Variable           seize
                         Offset Variable               logo

                      Number of Observations Read         295
                      Number of Observations Used         295

                              Class Level Information

   Class        Levels    Values

   subject          59    101 102 103 104 106 107 108 110 111 112 113 114
                          116 117 118 121 122 123 124 126 128 129 130 135
                          137 139 141 143 145 147 201 202 203 204 205 206
                          207 208 209 210 211 213 214 215 217 218 219 220
                          221 222 225 226 227 228 230 232 234 236 238

                                Parameter Information

                               Parameter       Effect

                               Prm1            Intercept
                               Prm2            v
                               Prm3            trt
                               Prm4            v*trt

                       Criteria For Assessing Goodness Of Fit

          Criterion                 DF           Value        Value/DF

          Deviance                 291       3577.8316         12.2950
          Scaled Deviance          291       3577.8316         12.2950
          Pearson Chi-Square       291       5733.4815         19.7027
          Scaled Pearson X2        291       5733.4815         19.7027
          Log Likelihood                     6665.9803

  Algorithm converged.

                       Analysis Of Initial Parameter Estimates

                                 Standard       Wald 95%          Chi-
   Parameter  DF  Estimate        Error    Confidence Limits     Square  Pr > ChiSq

   Intercept   1    1.3476       0.0341     1.2809    1.4144  1565.44      <.0001
   v           1    0.1108       0.0469     0.0189    0.2027     5.58      0.0181

                          UNSTRUCTURED CORRELATION                          2
                            The GENMOD Procedure

                       Analysis Of Initial Parameter Estimates

                                 Standard       Wald 95%          Chi-
   Parameter  DF  Estimate        Error    Confidence Limits     Square  Pr > ChiSq

   trt         1    0.0265       0.0467    -0.0650    0.1180     0.32      0.5702
   v*trt       1   -0.1037       0.0651    -0.2312    0.0238     2.54      0.1110
   Scale       0    1.0000       0.0000     1.0000    1.0000

NOTE: The scale parameter was held fixed.

                               GEE Model Information

                    Correlation Structure                 Unstructured
                    Subject Effect                subject (59 levels)
                    Number of Clusters                             59
                    Correlation Matrix Dimension                    5
                    Maximum Cluster Size                            5
                    Minimum Cluster Size                            5

                             Covariance Matrix (Model-Based)

                          Prm1           Prm2           Prm3           Prm4

          Prm1         0.01205        0.01924       -0.01205       -0.01924
          Prm2         0.01924        0.03091       -0.01924       -0.03091
          Prm3        -0.01205       -0.01924        0.02220        0.03696
          Prm4        -0.01924       -0.03091        0.03696        0.06209

                             Covariance Matrix (Empirical)
```

	Prm1	Prm2	Prm3	Prm4
Prm1	0.23193	0.0007209	-0.23193	-0.000721
Prm2	0.0007209	0.01564	-0.000721	-0.01564
Prm3	-0.23193	-0.000721	0.32478	-0.03058
Prm4	-0.000721	-0.01564	-0.03058	0.06334

Algorithm converged.

Working Correlation Matrix

	Col1	Col2	Col3	Col4	Col5
Row1	1.0000	0.9435	0.7324	0.8213	0.6856
Row2	0.9435	1.0000	0.8187	0.9435	0.7819
Row3	0.7324	0.8187	1.0000	0.7146	0.5375
Row4	0.8213	0.9435	0.7146	1.0000	0.6841
Row5	0.6856	0.7819	0.5375	0.6841	1.0000

UNSTRUCTURED CORRELATION 3
The GENMOD Procedure

Analysis Of GEE Parameter Estimates
Empirical Standard Error Estimates

Parameter	Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept	1.1186	0.4816	0.1747	2.0625	2.32	0.0202
v	0.1233	0.1251	-0.1218	0.3684	0.99	0.3241
trt	0.0711	0.5699	-1.0459	1.1881	0.12	0.9007
v*trt	-0.1140	0.2517	-0.6072	0.3793	-0.45	0.6507

Analysis Of GEE Parameter Estimates
Model-Based Standard Error Estimates

Parameter	Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept	1.1186	0.1098	0.9034	1.3338	10.19	<.0001
v	0.1233	0.1758	-0.2213	0.4679	0.70	0.4831
trt	0.0711	0.1490	-0.2209	0.3631	0.48	0.6331
v*trt	-0.1140	0.2492	-0.6023	0.3744	-0.46	0.6474
Scale	4.9502	.	.	.	.	.

NOTE: The scale parameter for GEE estimation was computed as the square root of the normalized Pearson's chi-square.

EXCHANGEABLE (COMPOUND SYMMETRY) CORRELATION 4
The GENMOD Procedure

Model Information

Data Set	WORK.SEIZURE
Distribution	Poisson
Link Function	Log
Dependent Variable	seize
Offset Variable	logo

Number of Observations Read	295
Number of Observations Used	295

Class Level Information

Class	Levels	Values
subject	59	101 102 103 104 106 107 108 110 111 112 113 114 116 117 118 121 122 123 124 126 128 129 130 135 137 139 141 143 145 147 201 202 203 204 205 206 207 208 209 210 211 213 214 215 217 218 219 220 221 222 225 226 227 228 230 232 234 236 238

Parameter Information

Parameter	Effect
Prm1	Intercept
Prm2	v
Prm3	trt
Prm4	v*trt

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	291	3577.8316	12.2950
Scaled Deviance	291	3577.8316	12.2950
Pearson Chi-Square	291	5733.4815	19.7027
Scaled Pearson X2	291	5733.4815	19.7027
Log Likelihood		6665.9803	

```
  Algorithm converged.

                  Analysis Of Initial Parameter Estimates

                            Standard       Wald 95%          Chi-
   Parameter  DF  Estimate     Error   Confidence Limits   Square  Pr > ChiSq

   Intercept   1    1.3476    0.0341    1.2809    1.4144  1565.44      <.0001
   v           1    0.1108    0.0469    0.0189    0.2027     5.58      0.0181

               EXCHANGEABLE (COMPOUND SYMMETRY) CORRELATION                5
                           The GENMOD Procedure

                  Analysis Of Initial Parameter Estimates

                            Standard       Wald 95%          Chi-
   Parameter  DF  Estimate     Error   Confidence Limits   Square  Pr > ChiSq

   trt         1    0.0265    0.0467   -0.0650    0.1180     0.32      0.5702
   v*trt       1   -0.1037    0.0651   -0.2312    0.0238     2.54      0.1110
   Scale       0    1.0000    0.0000    1.0000    1.0000

NOTE: The scale parameter was held fixed.

                           GEE Model Information

            Correlation Structure                 Exchangeable
            Subject Effect                  subject (59 levels)
            Number of Clusters                              59
            Correlation Matrix Dimension                     5
            Maximum Cluster Size                             5
            Minimum Cluster Size                             5

                     Covariance Matrix (Model-Based)

                       Prm1           Prm2           Prm3           Prm4

      Prm1          0.02286        0.01051       -0.02286       -0.01051
      Prm2          0.01051        0.02393       -0.01051       -0.02393
      Prm3         -0.02286       -0.01051        0.04296        0.02132
      Prm4         -0.01051       -0.02393        0.02132        0.04838

                     Covariance Matrix (Empirical)

                       Prm1           Prm2           Prm3           Prm4

      Prm1          0.02476      -0.001152       -0.02476       0.001152
      Prm2        -0.001152        0.01348       0.001152       -0.01348
      Prm3         -0.02476       0.001152        0.04922        0.01525
      Prm4         0.001152       -0.01348        0.01525        0.04563

  Algorithm converged.

                         Working Correlation Matrix

                  Col1         Col2         Col3         Col4         Col5

     Row1       1.0000       0.7716       0.7716       0.7716       0.7716
     Row2       0.7716       1.0000       0.7716       0.7716       0.7716
     Row3       0.7716       0.7716       1.0000       0.7716       0.7716
     Row4       0.7716       0.7716       0.7716       1.0000       0.7716
     Row5       0.7716       0.7716       0.7716       0.7716       1.0000

               EXCHANGEABLE (COMPOUND SYMMETRY) CORRELATION                6
                           The GENMOD Procedure

                          Exchangeable Working
                               Correlation

                          Correlation    0.7715879669

                     Analysis Of GEE Parameter Estimates
                      Empirical Standard Error Estimates

                            Standard   95% Confidence
          Parameter Estimate    Error       Limits            Z Pr > |Z|

          Intercept   1.3476   0.1574   1.0392   1.6560    8.56   <.0001
          v           0.1108   0.1161  -0.1168   0.3383    0.95   0.3399
          trt         0.0265   0.2219  -0.4083   0.4613    0.12   0.9049
          v*trt      -0.1037   0.2136  -0.5223   0.3150   -0.49   0.6274

                     Analysis Of GEE Parameter Estimates
                     Model-Based Standard Error Estimates

                            Standard   95% Confidence
          Parameter Estimate    Error       Limits            Z Pr > |Z|

          Intercept   1.3476   0.1512   1.0513   1.6439    8.91   <.0001
```

```
v           0.1108   0.1547  -0.1924   0.4140    0.72   0.4739
trt         0.0265   0.2073  -0.3797   0.4328    0.13   0.8982
v*trt      -0.1037   0.2199  -0.5348   0.3274   -0.47   0.6374
Scale       4.4388    .        .         .        .      .

NOTE: The scale parameter for GEE estimation was computed as the square root
      of the normalized Pearson's chi-square.

                              AR(1) CORRELATION                              7
                             The GENMOD Procedure

                              Model Information

                        Data Set              WORK.SEIZURE
                        Distribution              Poisson
                        Link Function                 Log
                        Dependent Variable          seize
                        Offset Variable              logo

                      Number of Observations Read         295
                      Number of Observations Used         295

                             Class Level Information

   Class        Levels    Values

   subject          59    101 102 103 104 106 107 108 110 111 112 113 114
                          116 117 118 121 122 123 124 126 128 129 130 135
                          137 139 141 143 145 147 201 202 203 204 205 206
                          207 208 209 210 211 213 214 215 217 218 219 220
                          221 222 225 226 227 228 230 232 234 236 238

                             Parameter Information

                           Parameter       Effect

                           Prm1            Intercept
                           Prm2            v
                           Prm3            trt
                           Prm4            v*trt

                      Criteria For Assessing Goodness Of Fit

          Criterion                 DF           Value        Value/DF

          Deviance                 291       3577.8316         12.2950
          Scaled Deviance          291       3577.8316         12.2950
          Pearson Chi-Square       291       5733.4815         19.7027
          Scaled Pearson X2        291       5733.4815         19.7027
          Log Likelihood                     6665.9803

  Algorithm converged.

                    Analysis Of Initial Parameter Estimates

                              Standard       Wald 95%          Chi-
   Parameter  DF  Estimate     Error   Confidence Limits   Square  Pr > ChiSq

   Intercept   1    1.3476    0.0341    1.2809    1.4144  1565.44      <.0001
   v           1    0.1108    0.0469    0.0189    0.2027     5.58      0.0181

                              AR(1) CORRELATION                              8
                             The GENMOD Procedure

                    Analysis Of Initial Parameter Estimates

                              Standard       Wald 95%          Chi-
   Parameter  DF  Estimate     Error   Confidence Limits   Square  Pr > ChiSq

   trt         1    0.0265    0.0467   -0.0650    0.1180     0.32      0.5702
   v*trt       1   -0.1037    0.0651   -0.2312    0.0238     2.54      0.1110
   Scale       0    1.0000    0.0000    1.0000    1.0000

NOTE: The scale parameter was held fixed.

                             GEE Model Information

                Correlation Structure                          AR(1)
                Subject Effect                 subject (59 levels)
                Number of Clusters                              59
                Correlation Matrix Dimension                     5
                Maximum Cluster Size                             5
                Minimum Cluster Size                             5

                          Covariance Matrix (Model-Based)

                            Prm1           Prm2           Prm3           Prm4

          Prm1           0.02046       0.007458       -0.02046      -0.007458
          Prm2          0.007458        0.02829      -0.007458       -0.02829
```

Prm3	-0.02046	-0.007458	0.03859	0.01571
Prm4	-0.007458	-0.02829	0.01571	0.05781

Covariance Matrix (Empirical)

	Prm1	Prm2	Prm3	Prm4
Prm1	0.02620	-0.003809	-0.02620	0.003809
Prm2	-0.003809	0.01248	0.003809	-0.01248
Prm3	-0.02620	0.003809	0.04494	0.01198
Prm4	0.003809	-0.01248	0.01198	0.06782

Algorithm converged.

Working Correlation Matrix

	Col1	Col2	Col3	Col4	Col5
Row1	1.0000	0.8131	0.6611	0.5375	0.4371
Row2	0.8131	1.0000	0.8131	0.6611	0.5375
Row3	0.6611	0.8131	1.0000	0.8131	0.6611
Row4	0.5375	0.6611	0.8131	1.0000	0.8131
Row5	0.4371	0.5375	0.6611	0.8131	1.0000

AR(1) CORRELATION 9
The GENMOD Procedure

Analysis Of GEE Parameter Estimates
Empirical Standard Error Estimates

Parameter	Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept	1.3119	0.1619	0.9947	1.6292	8.10	<.0001
v	0.1515	0.1117	-0.0675	0.3704	1.36	0.1751
trt	0.0188	0.2120	-0.3968	0.4343	0.09	0.9295
v*trt	-0.1283	0.2604	-0.6388	0.3821	-0.49	0.6222

Analysis Of GEE Parameter Estimates
Model-Based Standard Error Estimates

Parameter	Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept	1.3119	0.1430	1.0316	1.5923	9.17	<.0001
v	0.1515	0.1682	-0.1782	0.4811	0.90	0.3678
trt	0.0188	0.1965	-0.3663	0.4038	0.10	0.9240
v*trt	-0.1283	0.2404	-0.5996	0.3429	-0.53	0.5935
Scale	4.4907	.	.	.	.	.

NOTE: The scale parameter for GEE estimation was computed as the square root of the normalized Pearson's chi-square.

SUBJECT 207 DELETED 10
The GENMOD Procedure

Model Information

Data Set	WORK.WEIRD
Distribution	Poisson
Link Function	Log
Dependent Variable	seize
Offset Variable	logo

Number of Observations Read	290
Number of Observations Used	290

Class Level Information

Class	Levels	Values
subject	58	101 102 103 104 106 107 108 110 111 112 113 114 116 117 118 121 122 123 124 126 128 129 130 135 137 139 141 143 145 147 201 202 203 204 205 206 208 209 210 211 213 214 215 217 218 219 220 221 222 225 226 227 228 230 232 234 236 238

Parameter Information

Parameter	Effect
Prm1	Intercept
Prm2	v
Prm3	trt
Prm4	v*trt

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF

Deviance	286	2413.0245	8.4371
Scaled Deviance	286	2413.0245	8.4371
Pearson Chi-Square	286	3015.1555	10.5425
Scaled Pearson X2	286	3015.1555	10.5425
Log Likelihood		5631.7547	

Algorithm converged.

Analysis Of Initial Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	1.3476	0.0341	1.2809	1.4144	1565.44	<.0001
v	1	0.1108	0.0469	0.0189	0.2027	5.58	0.0181

SUBJECT 207 DELETED 11
The GENMOD Procedure

Analysis Of Initial Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
trt	1	-0.1080	0.0486	-0.2034	-0.0127	4.93	0.0264
v*trt	1	-0.3016	0.0697	-0.4383	-0.1649	18.70	<.0001
Scale	0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

GEE Model Information

Correlation Structure	Exchangeable
Subject Effect	subject (58 levels)
Number of Clusters	58
Correlation Matrix Dimension	5
Maximum Cluster Size	5
Minimum Cluster Size	5

Covariance Matrix (Model-Based)

	Prm1	Prm2	Prm3	Prm4
Prm1	0.01223	0.001520	-0.01223	-0.001520
Prm2	0.001520	0.01519	-0.001520	-0.01519
Prm3	-0.01223	-0.001520	0.02495	0.005427
Prm4	-0.001520	-0.01519	0.005427	0.03748

Covariance Matrix (Empirical)

	Prm1	Prm2	Prm3	Prm4
Prm1	0.02476	-0.001152	-0.02476	0.001152
Prm2	-0.001152	0.01348	0.001152	-0.01348
Prm3	-0.02476	0.001152	0.03751	-0.002999
Prm4	0.001152	-0.01348	-0.002999	0.02931

Algorithm converged.

Working Correlation Matrix

	Col1	Col2	Col3	Col4	Col5
Row1	1.0000	0.5941	0.5941	0.5941	0.5941
Row2	0.5941	1.0000	0.5941	0.5941	0.5941
Row3	0.5941	0.5941	1.0000	0.5941	0.5941
Row4	0.5941	0.5941	0.5941	1.0000	0.5941
Row5	0.5941	0.5941	0.5941	0.5941	1.0000

SUBJECT 207 DELETED 12

The GENMOD Procedure

Exchangeable Working Correlation

Correlation 0.5941485833

Analysis Of GEE Parameter Estimates
Empirical Standard Error Estimates

Parameter	Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept	1.3476	0.1574	1.0392	1.6560	8.56	<.0001
v	0.1108	0.1161	-0.1168	0.3383	0.95	0.3399
trt	-0.1080	0.1937	-0.4876	0.2716	-0.56	0.5770
v*trt	-0.3016	0.1712	-0.6371	0.0339	-1.76	0.0781

Analysis Of GEE Parameter Estimates
Model-Based Standard Error Estimates

Parameter	Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept	1.3476	0.1106	1.1309	1.5644	12.19	<.0001
v	0.1108	0.1233	-0.1308	0.3524	0.90	0.3687
trt	-0.1080	0.1579	-0.4176	0.2015	-0.68	0.4940
v*trt	-0.3016	0.1936	-0.6811	0.0779	-1.56	0.1193
Scale	3.2469	.	.	.	.	.

NOTE: The scale parameter for GEE estimation was computed as the square root of the normalized Pearson's chi-square.

AGE ADDED 13
The GENMOD Procedure

Model Information

Data Set	WORK.SEIZURE
Distribution	Poisson
Link Function	Log
Dependent Variable	seize
Offset Variable	logo

Number of Observations Read	295
Number of Observations Used	295

Class Level Information

Class	Levels	Values
subject	59	101 102 103 104 106 107 108 110 111 112 113 114 116 117 118 121 122 123 124 126 128 129 130 135 137 139 141 143 145 147 201 202 203 204 205 206 207 208 209 210 211 213 214 215 217 218 219 220 221 222 225 226 227 228 230 232 234 236 238

Parameter Information

Parameter	Effect
Prm1	Intercept
Prm2	logage
Prm3	v
Prm4	trt
Prm5	v*trt

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	290	3520.0007	12.1379
Scaled Deviance	290	3520.0007	12.1379
Pearson Chi-Square	290	5476.2836	18.8837
Scaled Pearson X2	290	5476.2836	18.8837
Log Likelihood		6694.8957	

Algorithm converged.

AGE ADDED 14
The GENMOD Procedure

Analysis Of Initial Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	3.2206	0.2482	2.7340	3.7071	168.30	<.0001
logage	1	-0.5616	0.0740	-0.7066	-0.4166	57.61	<.0001
v	1	0.1108	0.0469	0.0189	0.2027	5.58	0.0181
trt	1	-0.0043	0.0469	-0.0962	0.0876	0.01	0.9271
v*trt	1	-0.1037	0.0651	-0.2312	0.0238	2.54	0.1110
Scale	0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

GEE Model Information

Correlation Structure	Exchangeable
Subject Effect	subject (59 levels)
Number of Clusters	59
Correlation Matrix Dimension	5
Maximum Cluster Size	5
Minimum Cluster Size	5

Covariance Matrix (Model-Based)

Prm1	Prm2	Prm3	Prm4	Prm5

Prm1	1.88238	-0.56242	0.009622	-0.05729	-0.009622
Prm2	-0.56242	0.17001	-4.92E-18	0.01073	-4.7E-17
Prm3	0.009622	-4.92E-18	0.02306	-0.009622	-0.02306
Prm4	-0.05729	0.01073	-0.009622	0.04165	0.01956
Prm5	-0.009622	-4.7E-17	-0.02306	0.01956	0.04657

Covariance Matrix (Empirical)

	Prm1	Prm2	Prm3	Prm4	Prm5
Prm1	1.88843	-0.56699	-0.02199	0.01540	0.03990
Prm2	-0.56699	0.17266	0.006605	-0.01262	-0.01202
Prm3	-0.02199	0.006605	0.01348	0.0005524	-0.01348
Prm4	0.01540	-0.01262	0.0005524	0.04566	0.01574
Prm5	0.03990	-0.01202	-0.01348	0.01574	0.04563

Algorithm converged.

AGE ADDED 15

The GENMOD Procedure

Working Correlation Matrix

	Col1	Col2	Col3	Col4	Col5
Row1	1.0000	0.7617	0.7617	0.7617	0.7617
Row2	0.7617	1.0000	0.7617	0.7617	0.7617
Row3	0.7617	0.7617	1.0000	0.7617	0.7617
Row4	0.7617	0.7617	0.7617	1.0000	0.7617
Row5	0.7617	0.7617	0.7617	0.7617	1.0000

Exchangeable Working Correlation

Correlation 0.7617417343

Analysis Of GEE Parameter Estimates
Empirical Standard Error Estimates

Parameter	Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept	4.4338	1.3742	1.7404	7.1272	3.23	0.0013
logage	-0.9275	0.4155	-1.7419	-0.1131	-2.23	0.0256
v	0.1108	0.1161	-0.1168	0.3383	0.95	0.3399
trt	-0.0266	0.2137	-0.4454	0.3923	-0.12	0.9011
v*trt	-0.1037	0.2136	-0.5223	0.3150	-0.49	0.6274

Analysis Of GEE Parameter Estimates
Model-Based Standard Error Estimates

Parameter	Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept	4.4338	1.3720	1.7447	7.1228	3.23	0.0012
logage	-0.9275	0.4123	-1.7356	-0.1194	-2.25	0.0245
v	0.1108	0.1519	-0.1869	0.4084	0.73	0.4656
trt	-0.0266	0.2041	-0.4266	0.3735	-0.13	0.8965
v*trt	-0.1037	0.2158	-0.5266	0.3193	-0.48	0.6309
Scale	4.3350	.	.	.	.	

NOTE: The scale parameter for GEE estimation was computed as the square root of the normalized Pearson's chi-square.

MODIFIED MODEL 16
The GENMOD Procedure

Model Information

Data Set	WORK.SEIZURE
Distribution	Poisson
Link Function	Log
Dependent Variable	seize
Offset Variable	logo

Number of Observations Read	295
Number of Observations Used	295

Class Level Information

Class	Levels	Values
subject	59	101 102 103 104 106 107 108 110 111 112 113 114 116 117 118 121 122 123 124 126 128 129 130 135 137 139 141 143 145 147 201 202 203 204 205 206 207 208 209 210 211 213 214 215 217 218 219 220 221 222 225 226 227 228 230 232 234 236 238

Parameter Information

Parameter	Effect
Prm1	Intercept
Prm2	v
Prm3	visit4
Prm4	trt
Prm5	v*trt
Prm6	visit4*trt

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	289	3567.6314	12.3447
Scaled Deviance	289	3567.6314	12.3447
Pearson Chi-Square	289	5673.2719	19.6307
Scaled Pearson X2	289	5673.2719	19.6307
Log Likelihood		6671.0804	

Algorithm converged.

MODIFIED MODEL 17

The GENMOD Procedure

Analysis Of Initial Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	1.3476	0.0341	1.2809	1.4144	1565.44	<.0001
v	1	0.1351	0.0501	0.0369	0.2333	7.27	0.0070
visit4	1	-0.1009	0.0764	-0.2506	0.0489	1.74	0.1867
trt	1	0.0265	0.0467	-0.0650	0.1180	0.32	0.5702
v*trt	1	-0.0769	0.0694	-0.2129	0.0591	1.23	0.2676
visit4*trt	1	-0.1210	0.1092	-0.3350	0.0931	1.23	0.2679
Scale	0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

GEE Model Information

Correlation Structure	Exchangeable
Subject Effect	subject (59 levels)
Number of Clusters	59
Correlation Matrix Dimension	5
Maximum Cluster Size	5
Minimum Cluster Size	5

Covariance Matrix (Model-Based)

	Prm1	Prm2	Prm3	Prm4	Prm5	Prm6
Prm1	0.02277	0.01031	0.001711	-0.02277	-0.01031	-0.001711
Prm2	0.01031	0.02436	-0.004423	-0.01031	-0.02436	0.004423
Prm3	0.001711	-0.004423	0.02569	-0.001711	0.004423	-0.02569
Prm4	-0.02277	-0.01031	-0.001711	0.04280	0.02052	0.005259
Prm5	-0.01031	-0.02436	0.004423	0.02052	0.04828	-0.006694
Prm6	-0.001711	0.004423	-0.02569	0.005259	-0.006694	0.05315

Covariance Matrix (Empirical)

	Prm1	Prm2	Prm3	Prm4	Prm5	Prm6
Prm1	0.02476	-0.000931	-0.000952	-0.02476	0.0009314	0.0009516
Prm2	-0.000931	0.01770	-0.01079	0.0009314	-0.01770	0.01079
Prm3	-0.000952	-0.01079	0.01447	0.0009516	0.01079	-0.01447
Prm4	-0.02476	0.0009314	0.0009516	0.04922	0.01554	-0.001292
Prm5	0.0009314	-0.01770	0.01079	0.01554	0.05058	-0.01277
Prm6	0.0009516	0.01079	-0.01447	-0.001292	-0.01277	0.01681

Algorithm converged.

MODIFIED MODEL 18

The GENMOD Procedure

Working Correlation Matrix

	Col1	Col2	Col3	Col4	Col5
Row1	1.0000	0.7772	0.7772	0.7772	0.7772
Row2	0.7772	1.0000	0.7772	0.7772	0.7772
Row3	0.7772	0.7772	1.0000	0.7772	0.7772
Row4	0.7772	0.7772	0.7772	1.0000	0.7772
Row5	0.7772	0.7772	0.7772	0.7772	1.0000

Exchangeable Working

```
                          Correlation

                    Correlation    0.7771671618

                Analysis Of GEE Parameter Estimates
                Empirical Standard Error Estimates

                       Standard   95% Confidence
Parameter  Estimate     Error       Limits            Z Pr > |Z|

Intercept    1.3476   0.1574   1.0392   1.6560    8.56   <.0001
v            0.1351   0.1330  -0.1257   0.3958    1.02   0.3099
visit4      -0.1009   0.1203  -0.3366   0.1349   -0.84   0.4017
trt          0.0265   0.2219  -0.4083   0.4613    0.12   0.9049
v*trt       -0.0769   0.2249  -0.5177   0.3639   -0.34   0.7323
visit4*trt  -0.1210   0.1297  -0.3751   0.1331   -0.93   0.3507

              Analysis Of GEE Parameter Estimates
              Model-Based Standard Error Estimates

                       Standard   95% Confidence
Parameter  Estimate     Error       Limits            Z Pr > |Z|

Intercept    1.3476   0.1509   1.0518   1.6434    8.93   <.0001
v            0.1351   0.1561  -0.1708   0.4410    0.87   0.3868
visit4      -0.1009   0.1603  -0.4150   0.2133   -0.63   0.5292
trt          0.0265   0.2069  -0.3790   0.4320    0.13   0.8980
v*trt       -0.0769   0.2197  -0.5076   0.3537   -0.35   0.7262
visit4*trt  -0.1210   0.2305  -0.5728   0.3308   -0.52   0.5997
Scale        4.4307    .        .        .         .

NOTE: The scale parameter for GEE estimation was computed as the square root
      of  the normalized Pearson's chi-square.
```

INTERPRETATION:

- Pages 1–3 report the fit of the first model assuming the unstructured "working" correlation structure; pages 4–6 show the results for the compound symmetry assumption, and pages 7–9 show the results for the AR(1) assumption.

- On pages 1, 4, and 7, the table `Analysis of Initial Parameter Estimates` gives the estimates of $\boldsymbol{\beta}$ under the **independence** assumption (thus, these tables are the same for each fit).

- The results of solving the GEE begin on pages 2, 5, and 8 with the `Model Information` heading. The `Covariance Matrix (Model Based)` is the estimate $\widehat{\boldsymbol{V}}_\beta$; the `Covariance Matrix (Empirical)` is the "robust" estimate $\widehat{\boldsymbol{V}}_\beta^R$. They are somewhat similar for each fit, but different enough. How different can be seen in the tables `Analysis of GEE Parameter Estimates` that follow; that labeled `Empirical Standard Error Estimates` uses $\widehat{\boldsymbol{V}}_\beta^R$ to compute standard errors; that labeled `Model-Based Standard Error Estimates` uses $\widehat{\boldsymbol{V}}_\beta$.

- The fits are qualitatively very similar. In all cases, there does not seem to be any evidence that β_3 is different from zero.

- We have no formal method of choosing among the various "working" correlation assumptions. A practical approach is to inspect the results as above for each one – if they are in qualitative agreement, then we feel reasonably confident that results are not too dependent on the correlation assumption.

- Pages 10–12 show the results of the fit with the compound symmetric assumption and "unusual"subject 207 deleted. Note that now the results are suggestive of an effect of progabide; $\widehat{\beta}_3 = -0.30$ with a (robust) standard error of 0.17, yielding a p-value for a test of $\beta_3 = 0$ of 0.08.

- Adding age to the model [as log(age)] does not alter the results. Taking special account of the 4th visit does not yield any additional insight. It seems that, perhaps due to the magnitude of variation in the data and probable lack of a strong treatment effect, there is little evidence favoring progabide over placebo.

EXAMPLE 2 – WHEEZING DATA FROM THE SIX CITIES STUDY: Here, we consider fitting the model (12.2) similar to that fitted in Lipsitz, Laird, and Harrington (1992),

$$E(Y_{ij}) = \frac{\exp(\beta_0 + \beta_1 c_i + \beta_2 \delta_{0ij} + \beta_3 \delta_{1ij})}{1 + \exp(\beta_0 + \beta_1 c_i + \beta_2 \delta_{0ij} + \beta_3 \delta_{1ij})}.$$

We consider as in the seizure example several different "working" correlation assumptions. The output is in the same form as for the seizure example.

Recall, of course, our previous discussion about time-dependent covariates. The model for $E(Y_{ij})$ may well suffer the flaws we mentioned earlier; this fitting is mainly for illustration.

A difference between this fit and that in the seizure example is that there are **missing** values for some subjects. To make sure that SAS uses the correct convention to construct the covariance matrix for each individual (and hence the estimate of $\boldsymbol{\omega}$), the `within=` option of the `repeated` statement is used with the `class` variable `time`, which is identically equal to the numerical variable `age`. This has the effect of telling the program that it should consult the variable `time` to make sure each observation is classified correctly at its appropriate level of `age`.

Because these are binary data, we do not consider an overdispersion scale parameter. This is held fixed at 1.0 in the analyses by default for binary data.

PROGRAM:

```
/*****************************************************************

  CHAPTER 12, EXAMPLE 2

  Fit a logistic regression model to the "wheezing" data.
  These are binary data, thus, we use the Bernoulli (bin)
  mean/variance assumptions.  The model is fitted with different
  working correlation matrices.

*****************************************************************/

options ls=80 ps=59 nodate; run;

/*****************************************************************

  The data look like (first 4 records):

    1 portage   9 0 1  10 0 1  11 0 1  12 0 0
    2 kingston  9 1 1  10 2 1  11 2 0  12 2 0
    3 kingston  9 0 1  10 0 0  11 1 0  12 1 0
    4 portage   9 0 0  10 0 1  11 0 1  12 1 0

          .
          .
          .

  column 1      child
  column 2      city
  columns 3-5   age=9, smoking indiciator, wheezing response
  columns 6-8   age=10, smoking indiciator, wheezing response
  columns 9-11  age=11, smoking indiciator, wheezing response
  columns 12-14 age=12, smoking indiciator, wheezing response

  Some of the children have missing values for smoking and wheezing,
  as shown in Chapter 1.  There are 32 children all together.  See the
  output for the full data printed out one observation per line.

  We read in the data using the "@@" symbol so that SAS will continue
  to read for data on the same line and the OUTPUT statement to
  write each block of three observations for each age in as a separate
  data record.  The resulting data set is one with a separate line for
  each observation.  City is a character variable, so the dollar
  sign is used to read it in as such.

*****************************************************************/

data wheeze; infile 'wheeze.dat';
  input child city $ @@;
  do i=1 to 4;
    input age smoke wheeze @@;
    output;
  end;
run;

proc print data=wheeze; run;

/*****************************************************************

  Fit the logistic regression model using PROC GENMOD and
  three different working correlation matrix assumptions:

  -  unstructured
  -  compound symmetry (exchangeable)
  -  AR(1)

  We fit a model with linear predictor allowing effects of
  city and maternal smoking status but no "interaction"
  terms among these.

  The DIST=BIN option in the MODEL statement specifies that the
  Bernoulli mean-variance relationship be assumed.  The LINK=LOGIT
  option asks for the logistic mean model.

  The REPEATED statement specifies the "working" correlation
  structure to be assumed.    The CORRW option in the REPEATED
  statement prints out the estimated working correlation matrix
  under the assumption given in the TYPE= option.  The COVB
  option prints out the estimated covariance matrix of the estimate
  of beta -- both the usual estimate and the "robust" version
  are printed.  The MODELSE option specifies that the standard
  error estimates printed for the elements of betahat are based
  on the usual theory.  By default, the ones based on the "robust"
  version of the sampling covariance matrix are printed, too.

  The dispersion parameter phi is held fixed at 1 by default.

  The missing values are coded in the usual SAS way by periods (.).
```

```
  We delete these from the full data set, so that the data set input
  to PROC GENMOD contains only the observed data.  We assume that the
  fact that these observations are missing has nothing to do with the
  thing under study (which may or may not be true).  Thus,
  because these data are not balanced, we use the WITHIN option of
  the REPEATED statement to give SAS the time variable AGE as a
  classification variable so that it can figure out where the missing
  values are and use this information in estimating the correlation
  matrix.

  In versions 7 and higher of SAS, PROC GENMOD will model by
  default the probability that the response y=0 rather than
  the conventional y=1!  To make PROC GENMOD model probability
  y=1, as is standard, one must include the DESCENDING option in
  the PROC GENMOD statement.  In earlier versions of SAS, the
  probability y=1 is modeled by default, as would be expected.

  If the user is unsure which probability is being modeled, one
  can check the .log file.  In later versions of SAS, an explicit
  statement about what is being modeled will appear. PROC GENMOD
  output should also contain a statement about what is being
  modeled.

*****************************************************************/

data wheeze; set wheeze;
  if wheeze=. then delete;
  time=age;
run;

title "UNSTRUCTURED CORRELATION";
proc genmod data=wheeze descending;
  class child city smoke time;
  model wheeze = city smoke / dist=bin link=logit;
  repeated subject=child / type=un corrw covb modelse within=time;
run;

title "COMPOUND SYMMETRY (EXCHANGEABLE) CORRELATION";
proc genmod data=wheeze descending;
  class child city smoke time;
  model wheeze = city smoke / dist=bin link=logit;
  repeated subject=child / type=cs corrw covb modelse within=time;
run;

title "AR(1) CORRELATION";
proc genmod data=wheeze descending;
  class child city smoke time;
  model wheeze = city smoke / dist=bin link=logit;
  repeated subject=child / type=ar(1) corrw covb modelse within=time;
run;
```

OUTPUT: Following the output, we comment on a few aspects of the output.

The SAS System 1

Obs	child	city	i	age	smoke	wheeze
1	1	portage	1	9	0	1
2	1	portage	2	10	0	1
3	1	portage	3	11	0	1
4	1	portage	4	12	0	0
5	2	kingston	1	9	1	1
6	2	kingston	2	10	2	1
7	2	kingston	3	11	2	0
8	2	kingston	4	12	2	0
9	3	kingston	1	9	0	1
10	3	kingston	2	10	0	0
11	3	kingston	3	11	1	0
12	3	kingston	4	12	1	0
13	4	portage	1	9	0	0
14	4	portage	2	10	0	1
15	4	portage	3	11	0	1
16	4	portage	4	12	1	0
17	5	kingston	1	9	0	0
18	5	kingston	2	10	1	0
19	5	kingston	3	11	1	0
20	5	kingston	4	12	1	0
21	6	portage	1	9	0	0
22	6	portage	2	10	1	0
23	6	portage	3	11	1	0
24	6	portage	4	12	1	0
25	7	kingston	1	9	1	0
26	7	kingston	2	10	1	0
27	7	kingston	3	11	0	0
28	7	kingston	4	12	0	0
29	8	portage	1	9	1	0
30	8	portage	2	10	1	0
31	8	portage	3	11	1	0
32	8	portage	4	12	2	0
33	9	portage	1	9	2	1
34	9	portage	2	10	2	0
35	9	portage	3	11	1	0
36	9	portage	4	12	1	0
37	10	kingston	1	9	0	0
38	10	kingston	2	10	0	0
39	10	kingston	3	11	0	0
40	10	kingston	4	12	1	0
41	11	kingston	1	9	1	1
42	11	kingston	2	10	0	0
43	11	kingston	3	11	0	1
44	11	kingston	4	12	0	1
45	12	portage	1	9	1	0
46	12	portage	2	10	0	0
47	12	portage	3	11	0	0
48	12	portage	4	12	0	0
49	13	kingston	1	9	1	0
50	13	kingston	2	10	0	1
51	13	kingston	3	11	1	1
52	13	kingston	4	12	1	1
53	14	portage	1	9	1	0
54	14	portage	2	10	2	0
55	14	portage	3	11	1	0

The SAS System 2

Obs	child	city	i	age	smoke	wheeze
56	14	portage	4	12	2	1
57	15	kingston	1	9	1	0
58	15	kingston	2	10	1	0
59	15	kingston	3	11	1	0
60	15	kingston	4	12	2	1
61	16	portage	1	9	1	1
62	16	portage	2	10	1	1
63	16	portage	3	11	2	0
64	16	portage	4	12	1	0
65	17	portage	1	9	2	1
66	17	portage	2	10	2	0
67	17	portage	3	11	1	0
68	17	portage	4	12	1	0
69	18	kingston	1	9	0	0
70	18	kingston	2	10	0	0
71	18	kingston	3	11	0	0
72	18	kingston	4	12	0	0
73	19	portage	1	9	0	0
74	19	portage	2	10	.	.
75	19	portage	3	11	.	.
76	19	portage	4	12	.	.
77	20	kingston	1	9	.	.
78	20	kingston	2	10	0	1

79	20	kingston	3	11	.	.
80	20	kingston	4	12	.	.
81	21	portage	1	9	.	.
82	21	portage	2	10	.	.
83	21	portage	3	11	2	1
84	21	portage	4	12	.	.
85	22	kingston	1	9	.	.
86	22	kingston	2	10	.	.
87	22	kingston	3	11	.	.
88	22	kingston	4	12	1	0
89	23	portage	1	9	2	0
90	23	portage	2	10	1	1
91	23	portage	3	11	.	.
92	23	portage	4	12	.	.
93	24	kingston	1	9	2	0
94	24	kingston	2	10	.	.
95	24	kingston	3	11	0	0
96	24	kingston	4	12	.	.
97	25	portage	1	9	0	1
98	25	portage	2	10	.	.
99	25	portage	3	11	.	.
100	25	portage	4	12	0	0
101	26	portage	1	9	.	.
102	26	portage	2	10	0	0
103	26	portage	3	11	1	0
104	26	portage	4	12	.	.
105	27	portage	1	9	.	.
106	27	portage	2	10	1	0
107	27	portage	3	11	.	.
108	27	portage	4	12	1	0
109	28	kingston	1	9	.	.
110	28	kingston	2	10	.	.

Obs	child	city	i	age	smoke	wheeze
111	28	kingston	3	11	2	0
112	28	kingston	4	12	1	1
113	29	portage	1	9	1	0
114	29	portage	2	10	0	0
115	29	portage	3	11	0	0
116	29	portage	4	12	.	.
117	30	kingston	1	9	1	1
118	30	kingston	2	10	1	0
119	30	kingston	3	11	.	.
120	30	kingston	4	12	1	1
121	31	kingston	1	9	1	0
122	31	kingston	2	10	.	.
123	31	kingston	3	11	1	0
124	31	kingston	4	12	2	1
125	32	portage	1	9	.	.
126	32	portage	2	10	1	1
127	32	portage	3	11	1	0
128	32	portage	4	12	1	0

The GENMOD Procedure

Model Information

Data Set	WORK.WHEEZE
Distribution	Binomial
Link Function	Logit
Dependent Variable	wheeze

Number of Observations Read	100
Number of Observations Used	100
Number of Events	29
Number of Trials	100

Class Level Information

Class	Levels	Values
child	32	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32
city	2	kingston portage
smoke	3	0 1 2
time	4	9 10 11 12

Response Profile

Ordered Value	wheeze	Total Frequency
1	1	29
2	0	71

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

Parameter Information

Parameter	Effect	city	smoke
Prm1	Intercept		
Prm2	city	kingston	
Prm3	city	portage	
Prm4	smoke		0
Prm5	smoke		1
Prm6	smoke		2

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	96	117.9994	1.2292
Scaled Deviance	96	117.9994	1.2292
Pearson Chi-Square	96	99.6902	1.0384

UNSTRUCTURED CORRELATION 5
The GENMOD Procedure

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Scaled Pearson X2	96	99.6902	1.0384
Log Likelihood		-58.9997	

Algorithm converged.

Analysis Of Initial Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept		1	-0.4559	0.5285	-1.4917	0.5799	0.74
city	kingston	1	0.2382	0.4479	-0.6398	1.1161	0.28
city	portage	0	0.0000	0.0000	0.0000	0.0000	.
smoke	0	1	-0.4494	0.6159	-1.6565	0.7577	0.53
smoke	1	1	-0.8751	0.6029	-2.0568	0.3067	2.11
smoke	2	0	0.0000	0.0000	0.0000	0.0000	.
Scale		0	1.0000	0.0000	1.0000	1.0000	

Analysis Of Initial Parameter Estimates

Parameter		Pr > ChiSq
Intercept		0.3883
city	kingston	0.5950
city	portage	.
smoke	0	0.4656
smoke	1	0.1467
smoke	2	.
Scale		

NOTE: The scale parameter was held fixed.

GEE Model Information

Correlation Structure	Unstructured
Within-Subject Effect	time (4 levels)
Subject Effect	child (32 levels)
Number of Clusters	32
Correlation Matrix Dimension	4
Maximum Cluster Size	4
Minimum Cluster Size	1

UNSTRUCTURED CORRELATION 6

The GENMOD Procedure

Covariance Matrix (Model-Based)

	Prm1	Prm2	Prm4	Prm5
Prm1	0.25733	-0.09887	-0.19993	-0.18313
Prm2	-0.09887	0.22799	-0.02525	-0.02022
Prm4	-0.19993	-0.02525	0.36412	0.20072
Prm5	-0.18313	-0.02022	0.20072	0.27654

Covariance Matrix (Empirical)

	Prm1	Prm2	Prm4	Prm5
Prm1	0.19295	-0.05378	-0.16907	-0.23162
Prm2	-0.05378	0.21935	-0.03901	-0.06092
Prm4	-0.16907	-0.03901	0.32007	0.30071
Prm5	-0.23162	-0.06092	0.30071	0.46706

```
  Algorithm converged.

                          Working Correlation Matrix

                            Col1         Col2         Col3         Col4

            Row1          1.0000       0.1967       0.1807      -0.1604
            Row2          0.1967       1.0000       0.5531      -0.1131
            Row3          0.1807       0.5531       1.0000       0.2524
            Row4         -0.1604      -0.1131       0.2524       1.0000

                      Analysis Of GEE Parameter Estimates
                      Empirical Standard Error Estimates

                                Standard   95% Confidence
    Parameter          Estimate    Error       Limits            Z Pr > |Z|

    Intercept           -0.6197   0.4393  -1.4806   0.2413   -1.41   0.1583
    city      kingston   0.3126   0.4683  -0.6053   1.2306    0.67   0.5044
    city      portage    0.0000   0.0000   0.0000   0.0000       .       .
    smoke     0         -0.3851   0.5657  -1.4940   0.7237   -0.68   0.4960
    smoke     1         -0.4098   0.6834  -1.7493   0.9296   -0.60   0.5487
    smoke     2          0.0000   0.0000   0.0000   0.0000       .       .

                     Analysis Of GEE Parameter Estimates
                     Model-Based Standard Error Estimates

                                Standard   95% Confidence
    Parameter          Estimate    Error       Limits            Z Pr > |Z|

    Intercept           -0.6197   0.5073  -1.6139   0.3745   -1.22   0.2219
    city      kingston   0.3126   0.4775  -0.6232   1.2485    0.65   0.5126

                           UNSTRUCTURED CORRELATION                          7
                             The GENMOD Procedure

                     Analysis Of GEE Parameter Estimates
                     Model-Based Standard Error Estimates

                                Standard   95% Confidence
    Parameter          Estimate    Error       Limits            Z Pr > |Z|

    city      portage    0.0000   0.0000   0.0000   0.0000       .       .
    smoke     0         -0.3851   0.6034  -1.5678   0.7976   -0.64   0.5233
    smoke     1         -0.4098   0.5259  -1.4405   0.6209   -0.78   0.4358
    smoke     2          0.0000   0.0000   0.0000   0.0000       .       .
    Scale                1.0000        .        .        .       .       .
NOTE: The scale parameter was held fixed.

                  COMPOUND SYMMETRY (EXCHANGEABLE) CORRELATION                 8
                             The GENMOD Procedure

                               Model Information

                            Data Set              WORK.WHEEZE
                            Distribution             Binomial
                            Link Function               Logit
                            Dependent Variable         wheeze

                      Number of Observations Read         100
                      Number of Observations Used         100
                      Number of Events                     29
                      Number of Trials                    100

                              Class Level Information

    Class      Levels    Values

    child          32    1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
                         21 22 23 24 25 26 27 28 29 30 31 32
    city            2    kingston portage
    smoke           3    0 1 2
    time            4    9 10 11 12

                                Response Profile

                            Ordered                 Total
                              Value    wheeze    Frequency

                                  1    1              29
                                  2    0              71

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

                             Parameter Information

                  Parameter       Effect       city        smoke

                  Prm1            Intercept
```

Prm2	city	kingston	
Prm3	city	portage	
Prm4	smoke		0
Prm5	smoke		1
Prm6	smoke		2

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	96	117.9994	1.2292
Scaled Deviance	96	117.9994	1.2292
Pearson Chi-Square	96	99.6902	1.0384

COMPOUND SYMMETRY (EXCHANGEABLE) CORRELATION 9

The GENMOD Procedure

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Scaled Pearson X2	96	99.6902	1.0384
Log Likelihood		-58.9997	

Algorithm converged.

Analysis Of Initial Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept		1	-0.4559	0.5285	-1.4917	0.5799	0.74
city	kingston	1	0.2382	0.4479	-0.6398	1.1161	0.28
city	portage	0	0.0000	0.0000	0.0000	0.0000	.
smoke	0	1	-0.4494	0.6159	-1.6565	0.7577	0.53
smoke	1	1	-0.8751	0.6029	-2.0568	0.3067	2.11
smoke	2	0	0.0000	0.0000	0.0000	0.0000	.
Scale		0	1.0000	0.0000	1.0000	1.0000	

Analysis Of Initial Parameter Estimates

Parameter		Pr > ChiSq
Intercept		0.3883
city	kingston	0.5950
city	portage	.
smoke	0	0.4656
smoke	1	0.1467
smoke	2	.
Scale		

NOTE: The scale parameter was held fixed.

GEE Model Information

Correlation Structure	Exchangeable
Within-Subject Effect	time (4 levels)
Subject Effect	child (32 levels)
Number of Clusters	32
Correlation Matrix Dimension	4
Maximum Cluster Size	4
Minimum Cluster Size	1

COMPOUND SYMMETRY (EXCHANGEABLE) CORRELATION 10

The GENMOD Procedure

Covariance Matrix (Model-Based)

	Prm1	Prm2	Prm4	Prm5
Prm1	0.30777	-0.11319	-0.24502	-0.22930
Prm2	-0.11319	0.25956	-0.02313	-0.01878
Prm4	-0.24502	-0.02313	0.40717	0.24963
Prm5	-0.22930	-0.01878	0.24963	0.35226

Covariance Matrix (Empirical)

	Prm1	Prm2	Prm4	Prm5
Prm1	0.20021	-0.08869	-0.15237	-0.23871
Prm2	-0.08869	0.24782	-0.03222	-0.005869
Prm4	-0.15237	-0.03222	0.33433	0.28719
Prm5	-0.23871	-0.005869	0.28719	0.45634

Algorithm converged.

Working Correlation Matrix

	Col1	Col2	Col3	Col4
Row1	1.0000	0.1251	0.1251	0.1251
Row2	0.1251	1.0000	0.1251	0.1251
Row3	0.1251	0.1251	1.0000	0.1251
Row4	0.1251	0.1251	0.1251	1.0000

Exchangeable Working Correlation

Correlation 0.1251298267

Analysis Of GEE Parameter Estimates
Empirical Standard Error Estimates

Parameter		Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept		-0.4771	0.4475	-1.3541	0.3999	-1.07	0.2863
city	kingston	0.2456	0.4978	-0.7301	1.2213	0.49	0.6217
city	portage	0.0000	0.0000	0.0000	0.0000	.	.
smoke	0	-0.4006	0.5782	-1.5338	0.7327	-0.69	0.4885
smoke	1	-0.8492	0.6755	-2.1732	0.4748	-1.26	0.2087
smoke	2	0.0000	0.0000	0.0000	0.0000	.	.

Analysis Of GEE Parameter Estimates
Model-Based Standard Error Estimates

Parameter		Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept		-0.4771	0.5548	-1.5644	0.6102	-0.86	0.3898
city	kingston	0.2456	0.5095	-0.7529	1.2442	0.48	0.6297
city	portage	0.0000	0.0000	0.0000	0.0000	.	.
smoke	0	-0.4006	0.6381	-1.6512	0.8501	-0.63	0.5302
smoke	1	-0.8492	0.5935	-2.0125	0.3141	-1.43	0.1525
smoke	2	0.0000	0.0000	0.0000	0.0000	.	.
Scale		1.0000	.	.	.	.	.

NOTE: The scale parameter was held fixed.

Model Information

Data Set	WORK.WHEEZE
Distribution	Binomial
Link Function	Logit
Dependent Variable	wheeze

Number of Observations Read	100
Number of Observations Used	100
Number of Events	29
Number of Trials	100

Class Level Information

Class	Levels	Values
child	32	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32
city	2	kingston portage
smoke	3	0 1 2
time	4	9 10 11 12

Response Profile

Ordered Value	wheeze	Total Frequency
1	1	29
2	0	71

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

Parameter Information

Parameter	Effect	city	smoke
Prm1	Intercept		
Prm2	city	kingston	
Prm3	city	portage	
Prm4	smoke		0
Prm5	smoke		1
Prm6	smoke		2

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	96	117.9994	1.2292
Scaled Deviance	96	117.9994	1.2292
Pearson Chi-Square	96	99.6902	1.0384

AR(1) CORRELATION 13
The GENMOD Procedure

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Scaled Pearson X2	96	99.6902	1.0384
Log Likelihood		-58.9997	

Algorithm converged.

Analysis Of Initial Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept		1	-0.4559	0.5285	-1.4917	0.5799	0.74
city	kingston	1	0.2382	0.4479	-0.6398	1.1161	0.28
city	portage	0	0.0000	0.0000	0.0000	0.0000	.
smoke	0	1	-0.4494	0.6159	-1.6565	0.7577	0.53
smoke	1	1	-0.8751	0.6029	-2.0568	0.3067	2.11
smoke	2	0	0.0000	0.0000	0.0000	0.0000	.
Scale		0	1.0000	0.0000	1.0000	1.0000	

Analysis Of Initial Parameter Estimates

Parameter		Pr > ChiSq
Intercept		0.3883
city	kingston	0.5950
city	portage	.
smoke	0	0.4656
smoke	1	0.1467
smoke	2	.
Scale		

NOTE: The scale parameter was held fixed.

GEE Model Information

Correlation Structure	AR(1)
Within-Subject Effect	time (4 levels)
Subject Effect	child (32 levels)
Number of Clusters	32
Correlation Matrix Dimension	4
Maximum Cluster Size	4
Minimum Cluster Size	1

AR(1) CORRELATION 14

The GENMOD Procedure

Covariance Matrix (Model-Based)

	Prm1	Prm2	Prm4	Prm5
Prm1	0.31680	-0.12039	-0.24953	-0.22783
Prm2	-0.12039	0.27022	-0.02180	-0.01881
Prm4	-0.24953	-0.02180	0.42144	0.24916
Prm5	-0.22783	-0.01881	0.24916	0.34094

Covariance Matrix (Empirical)

	Prm1	Prm2	Prm4	Prm5
Prm1	0.22402	-0.08293	-0.18320	-0.26011
Prm2	-0.08293	0.23368	-0.02015	-0.007078
Prm4	-0.18320	-0.02015	0.34711	0.30564
Prm5	-0.26011	-0.007078	0.30564	0.45771

Algorithm converged.

Working Correlation Matrix

	Col1	Col2	Col3	Col4
Row1	1.0000	0.2740	0.0751	0.0206
Row2	0.2740	1.0000	0.2740	0.0751
Row3	0.0751	0.2740	1.0000	0.2740

```
          Row4       0.0206       0.0751       0.2740       1.0000

                    Analysis Of GEE Parameter Estimates
                    Empirical Standard Error Estimates

                                 Standard   95% Confidence
Parameter          Estimate       Error       Limits            Z Pr > |Z|

Intercept           -0.5442   0.4733  -1.4719   0.3835   -1.15   0.2502
city      kingston   0.2755   0.4834  -0.6720   1.2230    0.57   0.5687
city      portage    0.0000   0.0000   0.0000   0.0000       .      .
smoke     0         -0.3776   0.5892  -1.5323   0.7771   -0.64   0.5216
smoke     1         -0.6861   0.6765  -2.0121   0.6399   -1.01   0.3105
smoke     2          0.0000   0.0000   0.0000   0.0000       .      .

                    Analysis Of GEE Parameter Estimates
                    Model-Based Standard Error Estimates

                                 Standard   95% Confidence
Parameter          Estimate       Error       Limits            Z Pr > |Z|

Intercept           -0.5442   0.5629  -1.6474   0.5590   -0.97   0.3336
city      kingston   0.2755   0.5198  -0.7433   1.2943    0.53   0.5961

                          AR(1) CORRELATION                                15
                         The GENMOD Procedure

                    Analysis Of GEE Parameter Estimates
                    Model-Based Standard Error Estimates

                                 Standard   95% Confidence
Parameter          Estimate       Error       Limits            Z Pr > |Z|

city      portage    0.0000   0.0000   0.0000   0.0000       .      .
smoke     0         -0.3776   0.6492  -1.6500   0.8948   -0.58   0.5608
smoke     1         -0.6861   0.5839  -1.8305   0.4583   -1.18   0.2400
smoke     2          0.0000   0.0000   0.0000   0.0000       .      .
Scale                1.0000        .        .        .       .      .

NOTE: The scale parameter was held fixed.
```

INTERPRETATION:

- In this example, the analyses in each "working" case appear to be far less sensitive to whether $\widehat{\boldsymbol{V}}_\beta$ or $\widehat{\boldsymbol{V}}_\beta^R$ is used to construct standard errors; comparison of these matrices in each case shows that they are fairly similar.

- It is perhaps because it does not appear that there is any effect of any of the covariates on probability of wheezing that the analyses all seem to agree. Note from `Analysis of GEE Parameter Estimates` in each case that the signs (positive or negative) appear to be intuitively in the right direction; e.g., the coefficients for the "smoking" indicators are negative, suggesting that probability of wheezing is lower for children whose mothers do not smoke or only moderately smoke versus those who have heavy-smokers for mothers. However, in no case is there evidence to suggest these are different than zero. As there are only 32 children on which this analysis is based, perhaps the sample size is too small to detect departures from the various null hypotheses being tested.

- Keep in mind that this interpretation only makes sense under the assumption that the model for $E(Y_{ij})$ is correct!