

The analysis in Fig. 2 shows that there is a significantly increasing effect of dose (on a log scale): the Wald statistic for testing $H_0:\beta_1 = 0$ vs. $H_a:\beta_1 > 0$ is $W = 0.3705/0.0253 = 14.6446$; one-sided $P < 0.0001$. Notice, however, that the model fit may be in question: the ratio of D to df is $D/df = 10.77$, which is very large. Some further study of this effect would be necessary before using this model for predictive purposes.

The GENMOD Procedure					
Model Information					
Distribution	BINOMIAL				
Link Function	USER				
Dependent Variable	Y				
Dependent Variable	N				
Observations Used	6				
Number Of Events	330				
Number Of Trials	576				
Criteria For Assessing Goodness Of Fit					
Criterion	DF	Value	Value/DF		
Deviance	4	43.0610	10.7653		
Scaled Deviance	4	43.0610	10.7653		
Pearson Chi-Square	4	33.8032	8.4508		
Scaled Pearson X2	4	33.8032	8.4508		
Log Likelihood	.	-273.0489	.		
Analysis Of Parameter Estimates					
Parameter	DF	Estimate	Std Err	ChiSquare	Pr>Chi
INTERCEPT	1	0.5712	0.0355	258.6139	0.0001
X	1	0.3705	0.0253	214.4642	0.0001
SCALE	0	1.0000	0.0000	.	.
NOTE: The scale parameter was held fixed.					

Fig. 7.7 SAS PROC GENMOD output (edited) for complementary-log analysis of 9AA data.

12.1 Extending the exponential class

The exponential class specification given by Dobson (2002, §3.2) as

$$f(y) = \exp\{a(y)b(\theta) + c(\theta) + d(y)\} \quad (12.1)$$

may be extended to include a possibly-unknown dispersion parameter, $\psi > 0$. The extension involves the following specifications:

- $a(y)$ may be extended to $a(y,\psi)$, but this is then typically restricted to the simple canonical form

$$a(y,\psi) = y;$$

- $b(\theta)$ becomes $b(\theta,\psi)$ and is typically restricted to a scaled natural form

$$b(\theta,\psi) = \frac{\theta}{\psi},$$

where θ is the natural parameter of the p.d.f. We continue to say that a p.d.f satisfying $b(\theta,\psi) \propto \theta$ is in its *natural parameterization*;

- $c(\theta)$ becomes $c(\theta,\psi)$ and is typically restricted to a scaled form

$$c(\theta, \psi) = \frac{-\gamma(\theta)}{\psi}$$

(notice the change to a negative sign);

- $d(y)$ becomes $d(y, \psi)$.

Thus the class becomes

$$f(y) = \exp \left\{ \frac{y\theta - \gamma(\theta)}{\psi} + d(y, \psi) \right\}. \quad (12.2)$$

(In some cases, a set of prior, known weights, w , may be included in the analysis. To do so, simply change all instances where ψ appears to the inverse-weighted dispersion ψ/w . SAS PROC GENMOD handles such a setting via the SCWGT statement. If no weights are to be included, let $w = 1$ for all y .) An additional, important restriction continues to be that the support space for Y cannot depend upon any unknown parameters.

Recall that for any $f(y)$ satisfying (12.1), $E[Y] = -c(\theta)/b(\theta)$. It is common to denote this simply as $\mu = E[Y]$. Under (12.2), this becomes

$$\mu = E[Y] = \gamma(\theta). \quad (12.3)$$

Similarly, under (12.1) $\text{var}[Y] = [b(\theta)c(\theta) - b(\theta)c(\theta)]/[b(\theta)]^3$. Under (12.2), this simplifies to

$$\text{var}[Y] = \psi \frac{2\gamma(\theta)}{\theta^2}, \quad (12.4)$$

i.e., $\text{var}[Y] = \psi\gamma(\theta)$. We often find that the variance simplifies to $\text{var}[y] = \psi V(\mu)$, for some function $V(\mu)$ depending only on $\mu = \gamma(\theta)$. In such instances, we refer to $V(\mu)$ as the *variance function*.

12.2 Dispersion parameters

Incorporating an unknown dispersion parameter into GLiMs

Four important examples of exponential class p.d.f.s are: the normal with known σ^2 , the binomial, the Poisson, and the exponential. Notice that all four of these are in canonical form; i.e., $a(y) = y$ in (12.1). Under the dispersion extension in (12.2) this feature is essentially unaffected. In particular, the latter three p.d.f.s all satisfy (12.2) with $\psi = 1$, so we continue to observe $a(y, \psi) = y$.

All of the constructions necessary for building generalized linear regression models (GLiMs) from the exponential class in (12.1) can be extended to the class in (12.2). Indeed, in PROC GENMOD (12.2) is the default probability model. [SAS defines a separate *scale parameter* as a function of the dispersion parameter, ψ , for each possible p.d.f. in (12.2). For instance, when $Y \sim N(\mu, \sigma^2)$ SAS defines the scale as $\sigma^2 = \sqrt{\psi}$ but when $Y \sim \text{Gamma}(\alpha, \beta)$ it defines the scale as $\alpha = 1/\psi$. See SAS (1998, Ch. 10).]

Scaled deviance functions for measuring model adequacy

Historically, GLiMs were defined in terms of (12.2) before they were applied to p.d.f.s in (12.1); see McCullagh and Nelder (1989, §2.2). Treatments such as that given by Dobson (2002) fit well into this scheme, but also require slight alterations to address the full capabilities of GLiMs under (12.2). One important modification regards the nature of the deviance function to measure model adequacy and its relation to likelihood ratio (LR) statistics.

In its full and proper form, the deviance function is again defined as the discrepancy in log-likelihoods between the fitted model and the most-fully parameterized (*maximal*) model. When the dispersion parameter ψ in (12.2) is not equal to 1.0, however, its value may enter into the calculation. That is, the difference in log-likelihoods is now

$$\begin{aligned} \ell(\hat{\boldsymbol{\theta}}_{\max}; \mathbf{y}) - \ell(\hat{\boldsymbol{\theta}}; \mathbf{y}) &= \sum_i \left[\frac{y_i \hat{\theta}_{i, \max} - \gamma(\hat{\theta}_{i, \max})}{\psi} + d(y_i, \psi) - \frac{y_i \hat{\theta}_i - \gamma(\hat{\theta}_i)}{\psi} - d(y_i, \psi) \right] \\ &= \frac{\left\{ \sum_i [y_i(\hat{\theta}_{i, \max} - \hat{\theta}_i) - \gamma(\hat{\theta}_{i, \max}) + \gamma(\hat{\theta}_i)] \right\}}{\psi}, \end{aligned} \quad (12.5)$$

where here and in the sequel the sum is taken over the index $i = 1, \dots, N$, $\hat{\boldsymbol{\theta}}_{\max}$ denotes the maximum likelihood estimate (MLE) of the unknown natural parameters $\boldsymbol{\theta} = [\theta_1 \dots \theta_N]^T$, and $\hat{\boldsymbol{\theta}}$ is the corresponding MLE under the current model. Notice that ψ is assumed constant; i.e., it does not change over values of y_i . Versions of (12.5) incorporating prior weights are also possible; see McCullagh and Nelder (1989, §2.3).

Technically, the deviance function, $D(\mathbf{y}, \hat{\boldsymbol{\theta}})$, for the current model is twice the *numerator* in (12.5):

$$D(\mathbf{y}, \hat{\boldsymbol{\theta}}) = 2 \left\{ y_i(\hat{\theta}_{i,\max} - \hat{\theta}_i) - \gamma(\hat{\theta}_{i,\max}) + \gamma(\hat{\theta}_i) \right\}. \quad (12.6)$$

If $\psi = 1$, then (12.5) and (12.6) coincide (to a factor of 2) and constructions using either form will be equivalent. If the dispersion parameter is not equal to one, however, then its value must be incorporated into the log-likelihood ratio. Traditional GLiMs do so by defining the *scaled deviance* $D^*(\mathbf{y}, \hat{\boldsymbol{\theta}})$:

$$D^*(\mathbf{y}, \hat{\boldsymbol{\theta}}) = \frac{D(\mathbf{y}, \hat{\boldsymbol{\theta}})}{\psi} \quad (12.7)$$

which is simply (12.5) multiplied by two. [A better phrase for (12.7) would be to call it the dispersion-adjusted deviance, since as noted above the term “scale” is used in many different ways for models satisfying (12.2).] Thus $D^*(\mathbf{y}, \hat{\boldsymbol{\theta}}) = 2\{\ell(\hat{\boldsymbol{\theta}}_{\max}; \mathbf{y}) - \ell(\hat{\boldsymbol{\theta}}; \mathbf{y})\}$ is the traditional log-likelihood ratio. This is known to converge in distribution to $\chi^2(\nu)$, where ν is the difference in numbers of parameters fit in $\hat{\boldsymbol{\theta}}_{\max}$ and $\hat{\boldsymbol{\theta}}$. Note, however, that as was the case with no dispersion parameter, this asymptotic approximation can be poor, and it is not recommended for more than informal use.

For testing two nested models, say

$$H_0: \boldsymbol{\theta} = \boldsymbol{\theta}_0 = [\theta_1 \dots \theta_Q \ 0 \dots 0]^T \quad \text{vs.} \quad H_a: \boldsymbol{\theta} = [\theta_1 \dots \theta_Q \ \theta_{Q+1} \dots \theta_P]^T,$$

a LR statistic may be constructed from the corresponding pair of scaled deviances: $G^2 = D^*(\mathbf{y}, \hat{\boldsymbol{\theta}}_0) - D^*(\mathbf{y}, \hat{\boldsymbol{\theta}})$. This statistic converges in distribution to $\chi^2(P-Q)$, and this asymptotic approximation is typically more stable than that for $D^*(\mathbf{y}, \hat{\boldsymbol{\theta}})$.

The forms of the deviance function for different exponential class p.d.f.s are given in numerous sources, including McCullagh and Nelder (1989, §2.3) and SAS (1998, Ch. 10). One special case is notable: for $Y \sim N(\mu, \sigma^2)$, the deviance is simply the residual sum of squares: $D(\mathbf{y}, \hat{\boldsymbol{\theta}}) = \sum [y_i - \mu_i(\hat{\boldsymbol{\theta}})]^2$.

Recall also that another important measure of model adequacy is Pearson’s X^2 statistic, which in its general form is based on the generalized Pearson residuals

$$\frac{y_i - \mu_i(\hat{\boldsymbol{\theta}})}{\sqrt{V[\mu_i(\hat{\boldsymbol{\theta}})]}},$$

where $V(\mu)$ is the variance function of the p.d.f. in (12.2). Then, summing the squared residuals results in

$$X^2 = \sum \frac{[y_i - \mu_i(\hat{\boldsymbol{\theta}})]^2}{V[\mu_i(\hat{\boldsymbol{\theta}})]}.$$

This is the *generalized Pearson X^2 statistic* and it serves as a competitor to the deviance and scaled deviance for assessing model adequacy. In PROC GENMOD, SAS defines a *scale parameter* to account for cases where $\psi \neq 1$. It also reports a Scaled deviance and a Scaled Pearson X^2 under Criteria For Assessing Goodness Of Fit. When $\psi \neq 1$, the scaled deviance and/or scaled X^2 should be used instead of their corresponding unscaled values for assessing model adequacy, constructing LR statistics and Analysis of Deviance tables, etc.

If the value of $\psi \neq 1$ is known, it is easily incorporated into the PROC GENMOD analysis: simply call the NOSCALE option to prevent the scale parameter from being estimated from the data *and* call the SCALE= option (both in the MODEL statement) to identify the fixed value of ψ . (Be sure to specify ψ carefully: for most models PROC GENMOD defines its scale parameter as $\sqrt{\psi}$, but for the Gamma p.d.f. it sets its scale parameter to $1/\psi$.)

If the scale parameter is *unknown*, PROC GENMOD can estimate it. One can use the ML approach, however, it is often recommended to estimate the scale parameter using information in the scaled deviance or in the scaled X^2 statistic. Recall that as $N \rightarrow \infty$, $D^*(\mathbf{y}, \hat{\boldsymbol{\theta}})$ is approximately distributed as $\chi^2(\nu)$, where $\nu = N-P$ are the *df* associated with the current model. As such, the first moment is $E[D^*(\mathbf{y}, \hat{\boldsymbol{\theta}})] = \nu$. But since $D^*(\mathbf{y}, \hat{\boldsymbol{\theta}}) = D(\mathbf{y}, \hat{\boldsymbol{\theta}})/\psi$, this suggests $E[D(\mathbf{y}, \hat{\boldsymbol{\theta}})/\psi] = \nu$. From this, a method-of-moments estimator for ψ is $\hat{\psi}_D = \nu^{-1}D(\mathbf{y}, \hat{\boldsymbol{\theta}})$. This is achieved in PROC GENMOD via the DSCALE option in the MODEL statement. A similar operation with the scaled X^2 statistic — which is also distributed approximately as $\chi^2(\nu)$ — gives another method-of-moments estimator for ψ as $\hat{\psi}_P = X^2/\nu$. This is achieved in PROC GENMOD via the PSCALE option in the MODEL statement. The latter is preferred when the variance function is not constant with respect to μ , since the deviance-based estimate can exhibit instabilities in these instances.

Example: $Y \sim N(\mu, \sigma^2)$

For the normal case, the addition of an explicit dispersion parameter now allows consideration of cases where σ^2 is unknown. We write the p.d.f. as

$$\begin{aligned} f(y) &= \exp \left\{ -\frac{y^2 - 2y\mu + \mu^2}{2\sigma^2} - \frac{1}{2} \log(2\sigma^2) \right\} \\ &= \exp \left\{ \frac{y\mu - \frac{1}{2}\mu^2}{\sigma^2} - \frac{y^2}{2\sigma^2} - \frac{1}{2} \log(2\sigma^2) \right\}. \end{aligned} \quad (12.8)$$

In this form, we take the natural parameter to be $\theta = \mu$ and the dispersion parameter to be $\psi = \sigma^2$. Thus we see $\gamma(\theta) = \frac{1}{2}\theta^2$ and $d(y, \psi) = -\frac{1}{2}[y^2 + \log(2\sigma^2)]$. Since the support space for Y is the real line, and since this is independent of the unknown parameters, we can say that (12.8) is a member of the exponential class as given by (12.2). Notice that from (12.3), $E[Y] = \gamma'(\theta) = \theta = \mu$, and from (12.4), $\text{var}[Y] = \gamma''(\theta)\psi = (1/\sigma^2)\sigma^2 = 1$, as expected.

For a normal-based GLiM, the deviance was noted above to be the usual residual sum of squares, thus if there are P regression parameters being estimated in the current model, $D(\mathbf{y}, \hat{\boldsymbol{\beta}})/(N - P)$ is the usual mean square for error and acts as the desired, unbiased estimator of $\psi = \sigma^2$.

Example: $Y \sim \text{Gamma}(\alpha, \beta)$

A common model for positive-valued continuous data where the variance is not constant involves a gamma parent distribution. The standard formulation for the gamma is $Y \sim \text{Gamma}(\alpha, \beta)$, where α and β are positive parameters. The associated p.d.f. is

$$f(y) = \frac{y^{\alpha-1} e^{-y/\beta}}{\Gamma(\alpha) \beta^\alpha}.$$

The mean parameter here is $\mu = E[Y] = \alpha\beta$, and the dispersion parameter is $\psi = 1/\alpha$. The variance is $\text{var}[Y] = \alpha\beta^2 = (\alpha\beta)^2/\alpha = \mu^2\psi$. Written in terms of μ and ψ , the p.d.f. is

$$f(y) = \frac{y^{1/\psi} \exp\{-(\mu\psi)^{-1}y\}}{y^{1/\psi} (\mu\psi)^{1/\psi}}, \quad (12.9)$$

which is equivalent to

$$f(y) = \exp \left\{ \frac{-(y/\mu) - \log(\mu)}{\psi} + \frac{\log(y/\psi)}{\psi} - \log[y^{1/\psi} (\mu\psi)^{-1/\psi}] \right\}.$$

Writing the natural parameter as $\theta = -1/\mu$ shows that this density is in the form (12.2), with $\gamma(\theta) = \log\{-1/\theta\}$, and $d(y, \psi) = \psi^{-1} \log(y/\psi) - \log[y^{1/\psi} (\mu\psi)^{-1/\psi}]$. Since the support space for Y — the positive real line — is independent of the unknown parameters, we see that this is a member of the exponential class as given by (12.2). As expected, $\gamma'(\theta) = -1/\theta = \mu$, and $\text{var}[Y] = \gamma''(\theta)\psi = \theta^{-2}\psi = \mu^2\psi$. The variance function is $V(\mu) = \mu^2$. This is clearly quadratic in μ and as a result, the ratio of the population standard deviation to the mean is constant in μ , $(\psi\mu^2)^{1/2}/\mu = \psi^{1/2}$. This ratio is referred to as the population coefficient of variation (CV). For those situations where continuous, positive-valued data are thought to have a CV constant with respect to μ , the gamma model often is considered (McCullagh and Nelder, 1989, Ch. 8).

A special case of the Gamma parent model is the Exponential p.d.f.: $f(y) = \mu^{-1} \exp\{-y/\mu\}$ (Piegorisch and Bailer, 1997, §1.3). This is (12.9) with $\psi = 1$ (as noted above). We know that this is a (canonical) member of the exponential class in (12.1). But, through its connection to the Gamma p.d.f. in (12.9), it must also be a member of the exponential class (12.2).

12.3 Gamma regression models

Since GLiMs may be fit to any data arising from an exponential class p.d.f. from (12.2), it is possible to fit a GLiM to continuous, positive-valued data that exhibit a constant CV using (12.9) and PROC GENMOD. Assuming a vector of explanatory variables, \mathbf{x}_i , is associated with each observed response y_i , form the linear predictor, $\eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$. This can represent a simple or multiple linear regression model, or some form of ANOVA- or ANCOVA-type model, etc. Given an appropriate link function, one can compute parameter estimates, standard errors, LR or Wald tests, and Analyses of Deviance, using the entire arsenal of methods described by Dobson (2002).

Common choices for the link function include the log link

$$\eta_i = \log\{\mu_i\} \quad (12.10)$$

the power link, $\eta_i = \mu_i$, or the reciprocal link (a special case of the power link), $\eta_i = 1/\mu_i$. The latter case is the canonical link function for gamma GLiMs (since it mimics the relationship between the mean μ and the natural parameter θ), and it is the default in PROC GENMOD. However, under the reciprocal link it is necessary to truncate η_i if it reaches or drops below zero, in order to satisfy the model constraint that $\mu > 0$. An alternative link function that does not suffer this problem is the natural logarithm in (12.10), with inverse link $\mu = e^\eta$. In the latter case, we could refer to this as a log-linear model for μ , even though the data are continuous.

Example 12.A: Gamma GLiM for mutant frequencies

To illustrate these methods with gamma-distributed data, consider again the data on mutant frequencies (MF) given by Finette *et al.* (1994) and presented in Example 4.X. Recall that these data were frequencies of *hprt* mutations in human females. The data were

$x_i = \text{Age}$	$y_i = \text{MF}$	$x_i = \text{Age}$	$y_i = \text{MF}$	$x_i = \text{Age}$	$y_i = \text{MF}$
0.75	0.3	3.20	1.0	9.00	9.07
0.83	0.4	3.20	1.8	10.9	0.8
1.00	0.4	4.00	0.4	11.0	1.9
1.20	0.7	5.00	0.4	13.2	4.0
2.00	2.0	6.50	8.0	13.9	1.2
2.00	2.9	7.50	1.4	13.9	1.2
3.10	0.7	7.70	8.7	15.2	4.35

We assumed in Example 4.X that these data were distributed as Gamma, with unknown means and known scale set equal to 1.5. (That is, we set the PROC GENMOD scale parameter to 1.5, which corresponds to $\psi = 1/1.5 = 0.6667$.) That specification was made for convenience, however, and in fact the true value of ψ is unknown. A re-analysis of these data that allows ψ to vary freely can be accomplished via the SAS code in Fig. 12.1. Notice that the code assumes a logarithmic link as per (12.10) and a simple linear predictor $\eta_i = \beta_0 + \beta_1 x_i$, $i=1, \dots, 21$.

```
* SAS code to fit MF data with varying scale;
options linesize=69 nodate nonumber;

data MutFrg;
input age mf @@; lines;
0.75 0.3 0.83 0.4 1.00 0.4 1.20 0.77
2.00 2.0 2.00 2.9 3.10 0.7 3.20 1.0
3.20 1.8 4.00 0.45 5.00 0.4 6.50 8.0
7.50 1.4 7.70 8.7 9.00 9.0 10.9 0.8
11.0 1.9 13.2 4.0 13.9 1.2 13.9 1.2
15.2 4.3
proc genmod;

model mf = age / dist=gamma link=log expected;

run;
```

Fig. 12.1 SAS PROC GENMOD gamma-GLiM code for MF data.

The resulting output (edited) appears in Fig. 12.2. There, we see that the predictor variable (Age) is associated with significant increases in the mean MF (at the 5% significance level): the Wald statistic is $W = b_1/se[b_1] = 0.1089/0.0519 = 2.099$. The approximate one-sided P -value is $1 - \Phi(2.099) = 0.0179$, where $\Phi(\cdot)$ is the standard normal c.d.f. In addition, the MLE for the scale parameter is $1/\hat{\psi} = 1.2302$, with $se[1/\hat{\psi}] = 0.3403$. Roughly speaking, therefore, our earlier specification of $1/\psi = 1.5$ does not appear to be grossly invalid.

The GENMOD Procedure					
Model Information					
Description	Value				
Distribution	GAMMA				
Link Function	LOG				
Dependent Variable	MF				
Observations Used	21				
Criteria For Assessing Goodness Of Fit					
Criterion	DF	Value	Value/DF		
Deviance	19	19.2617	1.0138		
Scaled Deviance	19	23.6952	1.2471		
Pearson Chi-Square	19	23.0964	1.2156		
Scaled Pearson X2	19	28.4126	1.4954		
Log Likelihood	.	-37.7466	.		
Analysis Of Parameter Estimates					
Parameter	DF	Estimate	Std Err	ChiSquare	Pr>Chi
INTERCEPT	1	0.1098	0.3289	0.1114	0.7386
AGE	1	0.1089	0.0410	7.0604	0.0079
SCALE	1	1.2302	0.3403	.	.
NOTE: The scale parameter was estimated by maximum likelihood.					

Fig. 12.2 SAS PROC GENMOD output (edited) for gamma-GLiM analysis of MF data.

Example 12.B: Gamma ANCOVA for mutant frequencies

The data in Example 12.A were actually part of a larger experiment covering both male and female subjects (Finette *et al.*, 1994). The entire data set of *hprt* mutant frequencies (MF) is given in Table 12.1.

Table 12.1. Mutant *hprt* frequencies ($\times 10^6$) for human infants, children, and teenagers.

Gender	Age	MF	Gender	Age	MF	Gender	Age	MF
F	0.75	0.3	F	13.2	4.0	M	4.2	2.5
F	0.83	0.4	F	13.9	1.2	M	4.4	2.6
F	1.0	0.4	F	13.9	1.2	M	4.7	0.7
F	1.2	0.77	F	15.2	4.3	M	5.16	0.7
F	2.0	2.0				M	5.2	2.2
F	2.0	2.9	M	0.08	1.3	M	6.4	5.4
F	3.1	0.7	M	0.2	0.3	M	8.1	4.1
F	3.2	1.0	M	0.75	0.5	M	8.2	0.81
F	3.2	1.8	M	0.75	1.9	M	8.7	0.5
F	4.0	0.45	M	0.8	1.1	M	8.9	2.2
F	5.0	0.4	M	1.0	0.32	M	9.7	2.1
F	6.5	8.0	M	1.0	1.3	M	10.0	2.3
F	7.5	1.4	M	1.0	4.4	M	10.7	2.0
F	7.7	8.7	M	1.3	3.1	M	12.0	5.5
F	9.0	9.0	M	1.4	2.2	M	12.5	4.0
F	10.9	0.8	M	1.5	0.85	M	13.1	0.2
F	11.0	1.9	M	1.7	1.7	M	13.2	7.0

It is of interest with these data to determine if mutant frequencies differ between Males and Females, adjusting for the Age covariate. (As seen in Ex. 12.A and as in fact is known in mutagenesis research, mutant frequencies tend to increase as humans age.) This is a standard form of (equal-slopes) ANCOVA, for which the SAS code in Fig. 12.3 is useful. Notice the centering of the Age covariate about its grand mean $\bar{x}_{++} = 5.8344$, as is appropriate for the ANCOVA. Once again, a log link is chosen, as per (12.10). The code also includes a CONTRAST statement, used to test specifically for a Gender difference.

```

* SAS code to fit ANCOVA to MF data;
options linesize=69 nodate nonumber;

data MutFrqX;
  input gender $ age y @@;
      agebar = 5.8344; x = age - agebar;
  lines;
f 0.75 0.3 f 0.83 0.4 f 1.00 0.4
  ... other data omitted for space considerations ...
m 13.1 0.2 m 13.2 7.0

proc genmod;
  class gender;
  model y = x gender / dist=gamma link=log type1 expected;
  contrast 'f vs. m' gender 1 -1;

run;

```

Fig. 12.3 SAS PROC GENMOD gamma-GLiM code for ANCOVA of MF data.

The corresponding output (edited) appears in Fig. 12.4. The analysis therein indicates that gender does not appear to exert a significant influence on mean MF rates, after adjusting for Age. The Analysis of Deviance from the TYPE1 output shows a difference in scaled deviances of $G^2 = 0.0028$. On 1 *df* this is insignificant ($P = 0.96$). The CONTRAST analysis exhibits a similar result. In fact, it exhibits an *identical* result. This is normally not guaranteed, but since the gender classification factor had only two levels, there is only 1 *df* of information available in the analysis. Thus the TYPE1 fit of Gender summarizes completely the comparison between the two levels. Another consequence of this effect is seen under Analysis Of Parameter Estimates, where the parameter estimate and standard error for GENDER = M are given as identically zero. SAS has allocated the single *df* of information between genders to the F level (it comes before M alphabetically) and has set the M-level parameter to zero. (This is known as a *corner-point constraint*.)

12.4 Quasi-likelihood

An interesting realization was made by Wedderburn (1974) regarding estimation of the unknown regression parameters, $\boldsymbol{\beta}$, in a GLiM: the ML estimating equation for each β_j has the same form. From Dobson (2002, Eqn. 4.18) we have that the log-likelihood derivative (“score”) is

$$\frac{\ell}{\beta_j} = \frac{[y_i - \mu_i(\boldsymbol{\beta})] x_{ij}}{\text{var}[Y_i]} \frac{\mu_i}{\eta_i}, \quad (12.11)$$

where $\mu_i(\boldsymbol{\beta})$ is written to represent the mean as a function of $\boldsymbol{\beta}$. Recall that we denote the score as $U_j = \ell / \beta_j$. Extended to the larger exponential class in (12.2), U_j simplifies to

$$U_j = \frac{\ell}{\beta_j} \propto \frac{[y_i - \mu_i(\boldsymbol{\beta})] x_{ij}}{V[\mu_i]} \frac{\mu_i}{\eta_i}$$

(since ψ is treated as a constant). $V(\mu_i)$ is the variance function from $\text{var}[Y_i] = \psi V(\mu_i)$. Now, recognize that when $\eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$, η_i / β_j is simply x_{ij} , so from the chain rule

$$\frac{\mu_i}{\beta_j} = \frac{\mu_i}{\eta_i} \frac{\eta_i}{\beta_j} = \frac{\mu_i}{\eta_i} x_{ij}. \quad (12.12)$$

From this, one can write U_j as

$$U_j = \frac{\ell}{\beta_j} = \frac{[y_i - \mu_i(\boldsymbol{\beta})]}{V[\mu_i]} \frac{\mu_i}{\beta_j} \quad (12.13)$$

(Zeger and Liang, 1992). To find the MLE of β_j , $j=1, \dots, P$, set $U_j = 0$ and solve the equation numerically.

```

The GENMOD Procedure

      Model Information
Distribution          GAMMA
Link Function        LOG
Dependent Variable   Y
Observations Used    50

      Class Level Information
Class      Levels  Values
GENDER      2    f m

      Criteria For Assessing Goodness Of Fit
Criterion      DF      Value      Value/DF
Deviance       47      37.2654    0.7929
Scaled Deviance 47      55.3918    1.1785
Pearson Chi-Square 47      37.7115    0.8024
Scaled Pearson X2 47      56.0549    1.1927
Log Likelihood  .      -86.4991    .

      Analysis Of Parameter Estimates
Parameter      DF      Estimate      Std Err      ChiSquare      Pr>Chi
INTERCEPT    1      0.7780      0.1527      25.9567    0.0001
X              1      0.0830      0.0253      10.7430    0.0010
GENDER      f    1      -0.0127     0.2365      0.0029    0.9571
GENDER      m    0      0.0000      0.0000      .          .
SCALE        1      1.4864      0.2705      .          .
NOTE:  The scale parameter was estimated by maximum likelihood.

      LR Statistics For Type 1 Analysis
Source          Deviance      DF      ChiSquare      Pr>Chi
INTERCEPT    43.6376      0      .              .
X              37.2673      1      8.8004      0.0030
GENDER        37.2654      1      0.0028      0.9577

      CONTRAST Statement Results
Contrast      DF      ChiSquare      Pr>Chi      Type
f vs. m      1      0.0028      0.9577      LR

```

Fig. 12.4 SAS PROC GENMOD output (edited) for gamma-GLiM ANCOVA of MF data.

Wedderburn recognized that even if the underlying distribution of Y_i is *not* of the exponential-class form in (12.2), the estimating equations in (12.13) can still yield consistent estimators of β . Use of (12.13) still requires the user to specify the mean as a function of β , via $\mu_i(\beta)$, and the variance as a function of μ_i , via $\text{var}[Y_i] = \psi V(\mu_i)$. Notice, however, that the model framework can extend beyond a form of GLiM, since (12.13) is written to allow $\mu_i(\beta)$ to be any differentiable function of β .

Wedderburn coined the term *quasi-likelihood* to describe a situation where (12.13) is used to estimate regression coefficients with a specific form for the mean and variance (or equivalent quantities), but where the underlying probability distribution is left unspecified. The quantity U_j in (12.13) is the *quasi-score*, the integral of which is called the *quasi-likelihood function* (McCullagh and Nelder, 1989, §9.2). Estimators achieved by setting U_j to zero are called maximum quasi-likelihood estimators (MQLEs). They possess asymptotic features similar to those of MLEs (McCullagh, 1983), allowing for a broad range of statistical inferences to be performed on β . Heyde (1997) gives a modern overview.

If the dispersion parameter ψ in a quasi-likelihood model is unknown, it must be estimated. It is typical to appeal to moment estimators similar to those described in §12.2; i.e., since the scaled Pearson X^2 statistic remains distributed as approximately χ^2 , a moment estimator is available by equating X^2/ψ to its *df*, say ν . This produces the Pearson-based scale estimator $\hat{\psi}_p = X^2/\nu$.