

Resources: Plus Google Provides Many More ...

http://support.sas.com/documentation/cdl/en/statug/67523/HTML/default/viewer.htm#statug_genmod_overview.htm
<http://www.math.wpi.edu/saspdf/stat/chap29.pdf>

Examples of Generalized Linear Models

You construct a generalized linear model by deciding on response and explanatory variables for your data and choosing an appropriate link function and response probability distribution. Some examples of generalized linear models follow. Explanatory variables can be any combination of continuous variables, classification variables, and interactions.

Traditional Linear Model

- response variable: a continuous variable
- distribution: normal
- link function: identity

$$g(\mu) = \mu$$

Logistic Regression

- response variable: a proportion
- distribution: binomial
- link function: logit

$$g(\mu) = \log\left(\frac{\mu}{1-\mu}\right)$$

Poisson Regression in Log Linear Model

- response variable: a count
- distribution: Poisson
- link function: log

$$g(\mu) = \log(\mu)$$

Gamma Model with Log Link

- response variable: a positive, continuous variable
- distribution: gamma
- link function: log

$$g(\mu) = \log(\mu)$$

Note:

The response variables y_i are independent for $i = 1, 2, \dots$ and have a probability distribution from an exponential family. This implies that the variance of the response depends on the mean μ through a *variance function* V :

$$\text{var}(y_i) = \frac{\phi V(\mu_i)}{w_i}$$

where ϕ is a constant and w_i is a known weight for each observation. The *dispersion parameter* ϕ is either known (for example, for the binomial or Poisson distribution, $\phi = 1$) or it must be estimated.

The GENMOD Procedure

The GENMOD procedure fits a generalized linear model to the data by maximum likelihood estimation of the parameter vector β . There is, in general, no closed form solution for the maximum likelihood estimates of the parameters. The GENMOD procedure estimates the parameters of the model numerically through an iterative fitting process. The dispersion parameter ϕ is also estimated by maximum likelihood or, optionally, by the residual deviance or by Pearson's chi-square divided by the degrees of freedom. Covariances, standard errors, and are computed for the estimated parameters based on the asymptotic normality of maximum likelihood estimators.

A number of popular link functions and probability distributions are available in the GENMOD procedure. The built-in link functions are

- identity: $g(\mu) = \mu$
- logit: $g(\mu) = \log(\mu/(1 - \mu))$
- probit: $g(\mu) = \Phi^{-1}(\mu)$, where Φ is the standard normal cumulative distribution function
- power: $g(\mu) = \begin{cases} \mu^\lambda & \text{if } \lambda \neq 0 \\ \log(\mu) & \text{if } \lambda = 0 \end{cases}$
- log: $g(\mu) = \log(\mu)$
- complementary log-log: $g(\mu) = \log(-\log(1 - \mu))$

The available distributions and associated variance functions are

- normal: $V(\mu) = 1$
- binomial (proportion): $V(\mu) = \mu(1 - \mu)$
- Poisson: $V(\mu) = \mu$
- gamma: $V(\mu) = \mu^2$
- inverse Gaussian: $V(\mu) = \mu^3$
- negative binomial: $V(\mu) = \mu + k\mu^2$
- multinomial

```
proc genmod options;  
class categorical predictors;  
model response = predictors / dist = distribution link = link function ...;  
...
```

The PROC GENMOD statement invokes the procedure.

All statements other than the MODEL statement are optional.

The CLASS statement, if present, must precede the MODEL statement, and the CONTRAST statement must come after the MODEL statement.

OBSTATS

specifies that an additional table of statistics be displayed. For each observation, the following items are displayed:

- the value of the response variable (variables if the data are binomial), frequency, and weight variables
- the values of the regression variables
- predicted mean, $\hat{\mu} = g^{-1}(\eta)$, where $\eta = \mathbf{x}_i' \hat{\beta}$ is the linear predictor and g is the link function. If there is an offset, it is included in $\mathbf{x}_i' \hat{\beta}$.
- estimate of the linear predictor $\mathbf{x}_i' \hat{\beta}$. If there is an offset, it is included in $\mathbf{x}_i' \hat{\beta}$.
- standard error of the linear predictor $\mathbf{x}_i' \hat{\beta}$
- the value of the Hessian weight at the final iteration
- lower confidence limit of the predicted value of the mean. The confidence coefficient is specified with the ALPHA= option. See the section "[Confidence Intervals on Predicted Values](#)" for the computational method.
- upper confidence limit of the predicted value of the mean
- raw residual, defined as $Y - \mu$
- Pearson, or chi residual, defined as the square root of the contribution for the observation to the Pearson chi-square, that is

$$\frac{Y - \mu}{\sqrt{V(\mu)/w}}$$

where Y is the response, μ is the predicted mean, w is the value of the prior weight variable specified in a WEIGHT statement, and $V(\mu)$ is the variance function evaluated at μ .

- the standardized Pearson residual
- deviance residual, defined as the square root of the deviance contribution for the observation, with sign equal to the sign of the raw residual
- the standardized deviance residual
- the likelihood residual

```
title1 " ";
options nodate nonumber ls=80 nocenter;
data Challenger;
input Temp Damage $ @@;
cards;
66 NO 70 NO
75 NO 70 YES
57 YES 70 NO
69 NO 63 YES
81 NO 70 YES
76 NO 68 NO
78 NO 79 NO
67 NO 67 NO
75 YES 72 NO
53 YES 76 NO
73 NO 67 NO
58 YES
;

proc genmod descending;
model damage = temp / dist = bin link = logit obstats;
run;
```

The GENMOD Procedure

Model Information

Data Set	WORK.CHALLENGER
Distribution	Binomial
Link Function	Logit
Dependent Variable	Damage

Number of Observations Read	23
Number of Observations Used	23
Number of Events	7
Number of Trials	23

Response Profile

Ordered Value	Damage	Total Frequency
1	YES	7
2	NO	16

PROC GENMOD is modeling the probability that Damage='YES'.

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	21	20.3152	0.9674
Scaled Deviance	21	20.3152	0.9674
Pearson Chi-Square	21	23.1691	1.1033
Scaled Pearson X2	21	23.1691	1.1033
Log Likelihood		-10.1576	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits	Chi-Square	Pr > ChiSq
Intercept	1	15.0429	7.3786	0.5810 29.5048	4.16	0.0415
Temp	1	-0.2322	0.1082	-0.4443 -0.0200	4.60	0.0320
Scale	0	1.0000	0.0000	1.0000 1.0000		

NOTE: The scale parameter was held fixed.

The GENMOD Procedure

Observation Statistics

Observation	Damage	Temp Lower StResdev	Pred Upper StReschi	Xbeta Resraw Reslik	Std Reschi	HessWgt Resdev
1	NO	66	0.4304931	-0.279839	0.5936193	0.2451688
		0.191033	0.7075745	-0.430493	-0.869428	-1.061117
		-1.110155	-0.909608	-1.09428		
2	NO	70	0.2299683	-1.20849	0.5952546	0.1770829
		0.085086	0.4895483	-0.229968	-0.546487	-0.722943
		-0.74675	-0.564483	-0.736641		
3	NO	75	0.0855436	-2.369304	0.9401806	0.0782259
		0.0146	0.371314	-0.085544	-0.305853	-0.422908
		-0.438334	-0.317009	-0.431045		
4	YES	70	0.2299683	-1.20849	0.5952546	0.1770829
		0.085086	0.4895483	0.7700317	1.8298705	1.7145343
		1.7709953	1.8901296	1.778705		
5	YES	57	0.8593166	1.8096252	1.310985	0.1208916
		0.3186771	0.9876187	0.1406834	0.4046177	0.5506685
		0.6186795	0.4545905	0.5883645		
6	NO	70	0.2299683	-1.20849	0.5952546	0.1770829
		0.085086	0.4895483	-0.229968	-0.546487	-0.722943
		-0.74675	-0.564483	-0.736641		
7	NO	69	0.2736211	-0.976328	0.5645321	0.1987526
		0.1107808	0.5324879	-0.273621	-0.613752	-0.799604
		-0.826199	-0.634165	-0.815378		
8	YES	63	0.6026811	0.4166488	0.7726214	0.2394566
		0.2501783	0.8733555	0.3973189	0.8119436	1.006347
		1.0870329	0.8770428	1.0595688		
9	NO	81	0.0227033	-3.762281	1.5141575	0.0221878
		0.0011931	0.3111847	-0.022703	-0.152416	-0.214313
		-0.219981	-0.156447	-0.217198		
10	YES	70	0.2299683	-1.20849	0.5952546	0.1770829
		0.085086	0.4895483	0.7700317	1.8298705	1.7145343
		1.7709953	1.8901296	1.778705		
11	NO	76	0.0690441	-2.601467	1.0296372	0.064277
		0.0097613	0.3581472	-0.069044	-0.272332	-0.378268
		-0.391855	-0.282114	-0.38537		
12	NO	68	0.3220941	-0.744165	0.5536202	0.2183495
		0.1383288	0.5844117	-0.322094	-0.689297	-0.881756
		-0.912829	-0.713589	-0.900873		
13	NO	78	0.0445405	-3.065792	1.2177315	0.0425567
		0.0042673	0.3364661	-0.044541	-0.215909	-0.301871
		-0.311872	-0.223062	-0.307027		
14	NO	79	0.0356414	-3.297955	1.3150567	0.0343711
		0.0027998	0.3272848	-0.035641	-0.192246	-0.269414
		-0.277797	-0.198228	-0.273715		
15	NO	67	0.3747243	-0.512002	0.5636706	0.234306
		0.1656499	0.6440009	-0.374724	-0.774141	-0.969085
		-1.007304	-0.804672	-0.993644		
16	NO	67	0.3747243	-0.512002	0.5636706	0.234306
		0.1656499	0.6440009	-0.374724	-0.774141	-0.969085
		-1.007304	-0.804672	-0.993644		
17	YES	75	0.0855436	-2.369304	0.9401806	0.0782259
		0.0146	0.371314	0.9144564	3.2695491	2.2175345
		2.2984222	3.3888106	2.3898839		
18	NO	72	0.1580491	-1.672816	0.7041887	0.1330696
		0.0450881	0.4273598	-0.158049	-0.433264	-0.586572
		-0.606939	-0.448308	-0.59777		
19	YES	53	0.9392478	2.7382762	1.713217	0.0570614
		0.3498798	0.9977535	0.0607522	0.2543261	0.3540506
		0.3880328	0.2787366	0.3719734		
20	NO	76	0.0690441	-2.601467	1.0296372	0.064277
		0.0097613	0.3581472	-0.069044	-0.272332	-0.378268
		-0.391855	-0.282114	-0.38537		
21	NO	73	0.129546	-1.904979	0.7757603	0.1127639
		0.0315098	0.4050381	-0.129546	-0.385779	-0.526764
		-0.545602	-0.399576	-0.53695		
22	NO	67	0.3747243	-0.512002	0.5636706	0.234306
		0.1656499	0.6440009	-0.374724	-0.774141	-0.969085
		-1.007304	-0.804672	-0.993644		
23	YES	58	0.8288448	1.5774625	1.2137338	0.1418611
		0.3097236	0.9812261	0.1711552	0.454421	0.6127353
		0.6889376	0.5109346	0.6557441		

```

title1 " ";
options nodate nonumber ls=80 nocenter;

data ch12example22;
input CK YesHA      NoHA  Total;
cards;
20      2      88      90
60      13     26      39
100     30      8      38
140     30      5      35
180     21      0      21
220     19      1      20
260     18      1      19
300     13      1      14
340     19      0      19
380     15      0      15
420      7      0      7
460      8      0      8
500     35      0      35
;

proc genmod;
model YesHa/Total = CK / dist = bin link = logit obstats;
run;

```

The GENMOD Procedure

Model Information

Data Set	WORK.CH12EXAMPLE22
Distribution	Binomial
Link Function	Logit
Response Variable (Events)	YesHA
Response Variable (Trials)	Total

Number of Observations Read	13
Number of Observations Used	13
Number of Events	230
Number of Trials	360

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	11	28.1402	2.5582
Scaled Deviance	11	28.1402	2.5582
Pearson Chi-Square	11	163.9379	14.9034
Scaled Pearson X2	11	163.9379	14.9034
Log Likelihood		-93.8864	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits	Chi-Square	Pr > ChiSq
Intercept	1	-3.0284	0.3670	-3.7476 -2.3091	68.09	<.0001
CK	1	0.0351	0.0041	0.0271 0.0431	73.98	<.0001
Scale	0	1.0000	0.0000	1.0000 1.0000		

NOTE: The scale parameter was held fixed.

The GENMOD Procedure

Observation Statistics

Observation	YesHA	Total	CK HessWgt Reschi Reslik	Pred Lower Resdev	Xbeta Upper StResdev	Std Resraw StReschi
1	2	90	20 7.2949357 -2.224193 -4.029396	0.0889706 0.0515144 -2.625952	-2.326269 0.1493723 -4.462637	0.2993608 -6.007355 -3.779874
2	13	39	60 7.9393534 0.6754785 0.8064336	0.2845311 0.2122404 0.6649186	-0.922095 0.3698832 0.8024187	0.1986657 1.9032862 0.8151623
3	30	38	100 8.9687471 2.172755 2.8403687	0.6182385 0.5193533 2.2686045	0.4820782 0.7082096 2.8863655	0.2064458 6.5069377 2.7644153
4	30	35	140 4.0017181 -0.19569 -0.250137	0.8683276 0.7806294 -0.193425	1.8862518 0.9243634 -0.248977	0.3147576 -0.391465 -0.251893
5	21	21	180 0.7268568 0.8843066 1.2944174	0.9640989 0.9164278 1.239185	3.2904254 0.985022 1.3454993	0.4569742 0.7539235 0.9601745
6	19	20	220 0.1795911 -1.932047 -1.442168	0.9909383 0.97067 -1.346368	4.694599 0.9972402 -1.393713	0.609819 -0.818767 -1.999987
7	18	19	260 0.0424758 -4.645532 -2.236184	0.9977594 0.9900043 -2.108906	6.0987726 0.9995008 -2.135757	0.7669639 -0.957429 -4.704679
8	13	14	300 0.0077119 -11.2994 -2.943932	0.9994488 0.9966237 -2.795784	7.5029462 0.9999102 -2.805078	0.9262228 -0.992284 -11.33697
9	19	19	340 0.0025723 0.0507248 0.0717879	0.9998646 0.9988619 0.0717333	8.9071198 0.9999839 0.0718425	1.0866666 0.0025727 0.050802
10	15	15	380 0.0004988 0.0223345 0.0315917	0.9999667 0.9996164 0.0315855	10.311293 0.9999971 0.0315978	1.2478384 0.0004988 0.0223432
11	7	7	420 0.0000572 0.0075608 0.0106929	0.9999918 0.9998707 0.0106926	11.715467 0.9999995 0.0106932	1.4094884 0.0000572 0.0075612
12	8	8	460 0.000016 0.0040054 0.0056646	0.9999998 0.9999564 0.0056646	13.119641 0.9999999 0.0056647	1.5714691 0.000016 0.0040055
13	35	35	500 0.0000172 0.0041517 0.0058715	0.9999995 0.9999853 0.0058714	14.523814 1 0.0058716	1.7336878 0.0000172 0.0041518