Ch. 3 Generalized Linear Models

Components of GLM

1. Random Component : Identify response variable Y

Assume independent observations $y_1,\cdots,y_n$ from particular form of distributions such as Poisson or Binomial. Model how $\mu = E(Y_i)$ depends on explanatory variables.

2. Systematic component (linear predictor)

Pick explanatory variables $x_1,\cdots,x_k$ for linear predictor

$$\alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$$

3. Link function

Model function $g(\mu)$ of $\mu = E(Y)$ using

$$g(\mu) = \alpha + \beta_1 x_1 + \cdots + \beta_k x_k$$

where g is the link function.

The link function g connects the random component with the linear predictor function of the explanatory variables.

Example

● $\log(\mu) = \alpha + \beta_1 x_1 + \cdots + \beta_k x_k$ uses $g(\mu) = \log(\mu)$,

log link often used for a "count" random component for which $\mu > 0$

● $\log\left(\frac{\mu}{1-\mu}\right) = \alpha + \beta_1 x_1 + \cdots + \beta_k x_k$ uses $g(\mu) = \log\left(\frac{\mu}{1-\mu}\right)$,

the logit link (logit=log of odds)

often used for binomial, with $\mu = \pi$ between 0 and 1

● $\mu = \alpha + \beta_1 x_1 + \cdots + \beta_k x_k$ uses $g(\mu) = \mu$, identity link

eg) ordinary regression for normal response.

Note :

● A GLM generalizes ordinary regression by

1. permitting Y to have a distribution other than normal.
2. permitting modeling of $g(\mu)$ rather than μ

● The same ML (max. likelihood) fitting procedure applies to all GLMs. This is

basis of software such as PROC GENMOD in SAS (Nelder and Weddernurn, 1972)

GLMs for Binary Data

Suppose $Y = 1$ or 0

Let $P(Y=1) = \pi = 1 - P(Y=0)$ "Bernoulli Trial"

This is binomial for $n = 1$ trial.

$$E(Y) = \pi, \quad Var(Y) = \pi(1-\pi)$$

For explanatory variable x, $\pi = \pi(x)$ varies as x varies.

Linear probability model

$$\pi(x) = \alpha + \beta x$$

This is a GLM for binomial random component and identity link function.

$Var(Y) = \pi(x)(1-\pi(x))$ varies as x varies, so least squares method is not optimal. Use ML to fit this and other GLMs.

Example) Y = infant sex organ malformation (1=present, 0=absent)

x = mother's alcohol consumption (average drinks per day)

Alcohol Consumption	Malformation Preset	Malformation Absent	Total	Proportion Present
0	48	17,066	17,114	0.0028
< 1	38	14,464	14,502	0.0026
1-2	5	788	793	0.0063
3-5	1	126	127	0.0079
≥ 6	1	37	38	0.0262

```
DATA infants;
  INPUT alcohol malform total @@;
CARDS;
0 48 17114  0.5 38 14502  1.5 5 793  4.0 1 127  7.0  1 38
;
 PROC genmod descending;
  MODEL malform/total=alcohol / dist=bin link=identity;
 RUN;

 PROC genmod descending;
  MODEL malform/total=alcohol / dist=bin link=logit;
 RUN;
```

Output

```
                    The GENMOD Procedure
                   Model Information

Data Set                        WORK.INFANTS
```

Distribution	Binomial
Link Function	Identity
Response Variable (Events)	malform
Response Variable (Trials)	total

Number of Observations Read	5
Number of Observations Used	5
Number of Events	93
Number of Trials	32574

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	3	2.9795	0.9932
Scaled Deviance	3	2.9795	0.9932
Pearson Chi-Square	3	3.3551	1.1184
Scaled Pearson X2	3	3.3551	1.1184
Log Likelihood		-636.1122	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	0.0025	0.0003	0.0019	0.0032	58.52	<.0001
alcohol	1	0.0011	0.0007	-0.0003	0.0025	2.24	0.1348
Scale	0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

The GENMOD Procedure
Model Information

Data Set	WORK.INFANTS
Distribution	Binomial
Link Function	Logit
Response Variable (Events)	malform
Response Variable (Trials)	total

Number of Observations Read	5
Number of Observations Used	5
Number of Events	93
Number of Trials	32574

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	3	1.9487	0.6496
Scaled Deviance	3	1.9487	0.6496
Pearson Chi-Square	3	2.0523	0.6841
Scaled Pearson X2	3	2.0523	0.6841
Log Likelihood		-635.5968	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	-5.9605	0.1154	-6.1867	-5.7342	2666.41	<.0001
alcohol	1	0.3166	0.1254	0.0707	0.5624	6.37	0.0116
Scale	0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

R output

```
> alcohol <- c('0','<1','1-2','3-5','>=6')
> pre <- c(48,38,5,1,1)
> abs <- c(17066,14464,788,126,37)
> tot <- mal.pre+mal.abs
> Malf <- data.frame(alcohol,pre,abs,tot)
> Malf
  alcohol pre   abs   tot
1       0  48 17066 17114
2      <1  38 14464 14502
3     1-2   5   788   793
4     3-5   1   126   127
5     >=6   1    37    38
>
> Malf$alc <- c(0, 0.5,1.5,4.0,7.0)
>
> # Identity link with binomial random component
> fit1 <- glm(pre/tot ~ alc, family=quasi (link=identity, variance="mu(1-mu)"),
  weights=tot,data=Malf)
> summary(fit1, dispersion=1)
Call:
glm(formula = pre/tot ~ alc, family = quasi(link = identity,
    variance = "mu(1-mu)"), data = Malf, weights = tot)
Deviance Residuals:
       1         2         3         4         5
 0.6564   -1.0492    0.8631    0.1302    0.8282
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) 0.0025476  0.0003523   7.232 4.76e-13 ***
alc         0.0010872  0.0008324   1.306    0.192
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for quasi family taken to be 1)

    Null deviance: 6.2020  on 4  degrees of freedom
Residual deviance: 2.9795  on 3  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 10
> # Logit link with binomial random component
> fit2 <- glm(pre/tot ~ alc, family=binomial (link=logit), weights=tot,data=Malf)
```

```
> summary(fit2)
Call:
glm(formula = pre/tot ~ alc, family = binomial(link = logit),
    data = Malf, weights = tot)
Deviance Residuals:
       1         2         3         4         5
 0.5921   -0.8801    0.8865   -0.1449    0.1291
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  -5.9605     0.1154 -51.637   <2e-16 ***
alc           0.3166     0.1254   2.523   0.0116 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 6.2020  on 4  degrees of freedom
Residual deviance: 1.9487  on 3  degrees of freedom
AIC: 24.576
Number of Fisher Scoring iterations: 4
> fit2
Call:  glm(formula = pre/tot ~ alc, family = binomial(link = logit),
    data = Malf, weights = tot)
Coefficients:
(Intercept)          alc
    -5.9605       0.3166
Degrees of Freedom: 4 Total (i.e. Null);  3 Residual
Null Deviance:      6.202
Residual Deviance: 1.949        AIC: 24.58
> fitted(fit2)
          1           2           3           4           5
0.002572090 0.003011861 0.004128844 0.009065077 0.023100302
> confint(fit2)
Waiting for profiling to be done...
                  2.5 %      97.5 %
(Intercept) -6.19302366 -5.7396968
alc          0.01868149  0.5234947
>
> library(car)
> Anova(fit2)  # likelihood-ratio tests for effect parameters in a GLM
Analysis of Deviance Table (Type II tests)
```

```
Response: pre/tot
     LR Chisq Df Pr(>Chisq)
alc    4.2533  1     0.03917 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Using x scores (0, 0.5, 1.5, 4.0, 7.0), linear prob. model for π=prob. malformation present has ML fit

$$\hat{\pi} = \hat{\alpha} + \hat{\beta}x = 0.025 + 0.0011x$$

At $x = 0$, $\hat{\pi} = \hat{\alpha} = 0.0025$

$\hat{\pi}$ increases by $\hat{\beta} = 0.0011$ for each 1-unit increase in alcohol consumption.

Note

● ML estimates $\hat{\alpha}$, $\hat{\beta}$ obtained by iterative numerical optimization.

● To test $H_0 : \beta = 0$ (independence), can use

$$Z = \frac{\hat{\beta} - 0}{SE(\hat{\beta})}$$

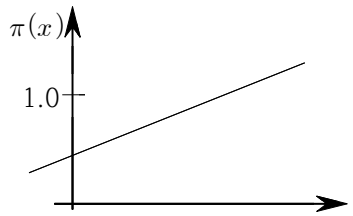
(for large n, it has approximate standard normal dist. under null hypothesis)

ex) $Z = \frac{0.0011}{0.0007} = 1.50$, $p - value = 0.13$ $(H_a : \beta \neq 0)$

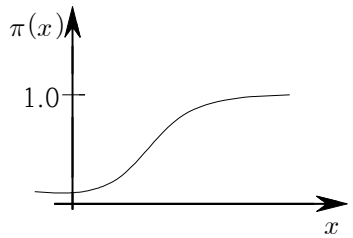
● Could use Pearson X^2 (or G^2) to test independence, but ignores ordering of row.

● Alternative way to apply X^2 (or deviance G^2) is to test fit of model

● Model $\pi(x) = \alpha + \beta x$ can give $\hat{\pi} > 1$ or $\hat{\pi} < 0$



● More realistic models are nonlin in shape of $\pi(x)$



Logistic Regression Model

$$\log\left(\frac{\pi}{1-\pi}\right) = \alpha + \beta x$$

is GLM for binomial Y with logit link.

$\hat{\pi} \uparrow \ as \ x \uparrow$, and $p-$ value=0.12 for $H_0: \ \beta = 0$

Note

- For contingency table, one can test $H_0: \ model \ fits$, using estimated expected frequency that satisfy the model, with $X^{2,}$, G^2 test statistics.

ex) $X^2 = 2.05, \ G^2 = 1.95$ for $H_0:$ logistic regression model

$df = 3 =$ (5 binomial observation) - (2 parameter)

$p-$ value is large $\Rightarrow$ no evidence against H_0

- Odds $\frac{\pi(x)}{1-\pi(x)} = e^{\alpha+\beta x} = e^{\alpha}(e^{\beta})^x$. So odds increase multiplicatively by e^{β} for every 1-unit increase in x
- Model generalizes to

$$\log\left[\frac{\pi(x)}{1-\pi(x)}\right] = \alpha + \beta_1 x_1 + \cdots + \beta_p x_p$$

- Model can be fitted using ML.

GLMs for count data

When Y is a count $(0,1,2,\cdots)$, it is traditional to assume Poisson dist.

$$P(y) = \frac{e^{-\mu}\mu^y}{y!}, \ y = 0,1,2,\cdots$$

- $E(Y) = \mu = Var(Y), \ \sigma = \sqrt{\mu}$
- In practice, often $\sigma^2 > \mu$
- Negative binomial dist. permits overdispersion

Poisson regression for count data

Suppose we assume Y has Poisson dist. and let x be an explanatory variable

Model

$$\mu = \alpha + \beta x: \ Identity \ link$$

or

$$\log\mu = \alpha + \beta x: \ \log \ link$$

Ex) Wafer defects (problem 3.11)

$Y =$ no. defects on silicon wafer

$x =$ treatment (1=B, 0=A) dummy (indicator) variable

10 wafers for each

A: 8, 7, 6, ... $\bar{y}_A = 5.0$

B: 9, 9, 8, ... $\bar{y}_B = 9.0$

```
data silicon;
input trt defect @@;
cards;
1 9 1 9 1 8 1 14 1 8 1 13 1 11 1 5 1 7 1 6
2 8 2 7 2 6 2  6 2 3 2  4 2  7 2 2 2 3 2 4
;
proc format;
 value trt 2='Treatment A' 1='Treatment B'
run;
❶ proc genmod order=data;
 format trt trt.;
 class trt;
 model defect = trt / dist=poi link=identity;
❷ proc genmod order=data;
 format trt trt.;
 class trt;
 model defect = trt / dist=poi link=log;
❸ proc genmod order=data;
 format trt trt.;
 model defect = / dist=poi link=log;
run;
```

Output

❶

The GENMOD Procedure
Model Information

Data Set	WORK.SILICON
Distribution	Poisson
Link Function	Identity
Dependent Variable	defect

Number of Observations Read	20
Number of Observations Used	20

Class Level Information

Class	Levels	Values
trt	2	Treatment B run Treatment A

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	18	16.2676	0.9038
Scaled Deviance	18	16.2676	0.9038
Pearson Chi-Square	18	16.0444	0.8914
Scaled Pearson X2	18	16.0444	0.8914
Log Likelihood		138.2221	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	5.0000	0.7071	3.6141	6.3859	50.00	<.0001
trt Treatment B run	1	4.0000	1.1832	1.6809	6.3191	11.43	0.0007
trt Treatment A	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale	0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

❷

The GENMOD Procedure

Model Information

Data Set	WORK.SILICON
Distribution	Poisson
Link Function	Log
Dependent Variable	defect
Number of Observations Read	20
Number of Observations Used	20

Class Level Information

Class	Levels	Values
trt	2	Treatment B run Treatment A

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	18	16.2676	0.9038
Scaled Deviance	18	16.2676	0.9038
Pearson Chi-Square	18	16.0444	0.8914
Scaled Pearson X2	18	16.0444	0.8914
Log Likelihood		138.2221	

L_1

Algorithm converged.

Analysis Of Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	1.6094	0.1414	1.3323	1.8866	129.51	<.0001
trt	Treatment B run	1	0.5878	0.1764	0.2421	0.9335	11.11	0.0009
trt	Treatment A	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale		0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

❸

The GENMOD Procedure

Model Information

Data Set	WORK.SILICON
Distribution	Poisson
Link Function	Log
Dependent Variable	defect
Number of Observations Read	20
Number of Observations Used	20

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	19	27.8570	1.4662
Scaled Deviance	19	27.8570	1.4662
Pearson Chi-Square	19	27.7143	1.4586
Scaled Pearson X2	19	27.7143	1.4586
Log Likelihood		132.4274	

L_0

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	1.9459	0.0845	1.7803	2.1116	530.12	<.0001
Scale	0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

R output

```
> trt <- c(rep('B',10),rep('A',10))
> defect <- c(9,9,8,14,8,13,11,5,7,6,8,7,6,6,3,4,7,2,3,4)
> silicon <- data.frame(trt,defect)
> # identity link
> pois.fit1 <- glm(defect ~ factor(trt), family=quasi (link=identity, variance="mu"),
 data=silicon)
> summary(pois.fit1, dispersion=1)
Call:
glm(formula = defect ~ factor(trt), family = quasi(link = identity,
    variance = "mu"), data = silicon)
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.5280  -0.7622  -0.1699   0.6938   1.5399
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)    5.0000     0.7071   7.071 1.54e-12 ***
factor(trt)B   4.0000     1.1832   3.381 0.000723 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
(Dispersion parameter for quasi family taken to be 1)

    Null deviance: 27.857  on 19  degrees of freedom
Residual deviance: 16.268  on 18  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 3

> # log link
> pois.fit2 <- glm(defect ~ factor(trt), family=poisson (link=log), data=silicon)
> summary(pois.fit2)
Call:
glm(formula = defect ~ factor(trt), family = poisson(link = log),
    data = silicon)
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.5280  -0.7622  -0.1699   0.6938   1.5399
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)    1.6094     0.1414  11.380  < 2e-16 ***
factor(trt)B   0.5878     0.1764   3.332 0.000861 ***
```

```
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 27.857  on 19  degrees of freedom
Residual deviance: 16.268  on 18  degrees of freedom
AIC: 94.349
Number of Fisher Scoring iterations: 4

> pois.fit3 <-  glm(defect ~ 1, family=poisson (link=log), data=silicon)
> #pois.fit3 <- update(pois.fit2,.~.-factor(trt))
> summary(pois.fit3)
Call:
glm(formula = defect ~ 1, family = poisson(link = log), data = silicon)
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.2336  -0.9063   0.0000   0.4580   2.3255
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.94591    0.08451   23.02   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 27.857  on 19  degrees of freedom
Residual deviance: 27.857  on 19  degrees of freedom
AIC: 103.94
Number of Fisher Scoring iterations: 4
```

For model $\mu = \alpha + \beta x$ (identity link)

$$\hat{\mu} = 5.0 + 4.0x$$

$x = 0$: $\hat{\mu}_A = 5.0 (= \bar{y}_A)$

$x = 1$: $\hat{\mu}_B = 9.0 (= \bar{y}_B)$

$\hat{\beta} = 4.0 = \hat{\mu}_B - \hat{\mu}_A$ has $S.E = 1.18$, 95% C.I. for β is

$$4.0 \pm 1.96(1.18)$$

For log-linear model $\log(\mu) = \alpha + \beta x$

$$\log(\hat{\mu}) = 1.609 + 0.588x$$

$x = 0$: $\log\hat{\mu}_A = 1.609$, $\hat{\mu} = e^{1.609} = 5.0$

$x = 1$: $\log\hat{\mu}_B = 1.609 + 0.588 = 2.197$, $\hat{\mu}_B = 9.0$

Inference for GLM parameters

C.I. : $\hat{\beta} \pm Z_{\alpha/2}(SE)$.

Test : H_0 : $\beta = 0$

1. Wald test

$$Z = \frac{\hat{\beta}}{SE} \text{ has approx. } N(0, 1) \text{ dist.}$$

For H_a : $\beta \neq 0$, can also use wald stat.

$$Z^2 = \left(\frac{\hat{\beta}}{SE}\right)^2 \text{ is approx. } \chi_1^2$$

C.I. = Set of β_0 value for H_0 : $\beta = \beta_0$ Such that

$$\frac{|\hat{\beta} - \beta_0|}{SE} < Z_{\alpha/2}$$

2. Likelihood-ratio test

l_0 = maximized likelihood when $\beta = 0$

l_1 = maximized likelihood for arbitrary β

Test stat. = $-2\log\left(\frac{l_0}{l_1}\right)$

= $-2\log l_0 - (-2\log l_1)$

$= -2(L_0 - L_1)$

where $L =$ maximized log likelihood.

ex) Wafer defects (Revisited)

Log-linear model : $\log(\mu) = \alpha + \beta x$

$\beta = \log\mu_B - \log\mu_A$

$$H_0 \ : \ \mu_A = \mu_B \Leftrightarrow \beta = 0$$

Wald test

$$Z = \frac{\hat{\beta}}{SE} = \frac{0.588}{0.176} = 3.33$$

$Z^2 = 11.1, \ df = 1, \ p = 0.0009$ for $H_a \ : \beta \neq 0$

Likelihood-ratio test

$$L_1 = 138.2, \ L_0 = 132.4$$

Test statistic$= -2(L_0 - L_1) = 11.6, \ df = 1, \ p = 0.007$

Proc GENMOD reports LR test result with 'type 3' option.

⇒ In SAS code : Model defect=trt/dist=poi link=log type3;

Note

● For very large n, Wald test and likelihood ratio test are approx. equivalent, but for small to moderate n the LR test is more reliable and powerful.

● LR stat. also equals difference in "deviances", goodness of fit stat.

ex) $27.86 - 16.27 = 11.59$

● LR method also extends to C.I.s :

$100(1-\alpha)\%$ C.I. = Set of β_0 in $H_0 : \beta = \beta_0$ for which $p-$ value $> \alpha$ in LR test (i.e., do not reject H_0 at $\alpha-$ level.)

ex) $\beta = \log\mu_B - \log\mu_A = \log\left(\frac{\mu_B}{\mu_A}\right) \Rightarrow e^{\beta} = \frac{\mu_A}{\mu_B}$

$$e^{\hat{\beta}} = e^{0.5878} = 1.8 = \frac{\hat{\mu}_B}{\hat{\mu}_A}$$

95% C.I. for β is $0.588 \pm 1.96(0.176) = (0.242, \ 0.934)$

Thus 95% C.I. for $e^{\beta} = \frac{\mu_B}{\mu_A}$ is

$$(e^{0.242}, \ e^{0.934}) = (1.27, 2.54)$$

We are 95% confident that μ_B is between 1.27 and 2.54 times μ_A

Deviance of a GLM

The saturated model has a separate paramter for each observation and has the perfect fit $\hat{\mu}_i = y_i$.

For a model M with maximized log-likelihood L_M,

Deviance = $-2[L_M - Ls]$

= LR stat. for testing that all parameters that are in S but not in M equal 0

where S=saturated model.

i.e., for H_0 : model holds vs. H_a : saturated model

ex) Poisson model for counts

Deviance= $G^2 = 2\sum_i y_i \log\left(\frac{y_i}{\hat{\mu}_i}\right)$ for M

When $\hat{\mu}_i$ are large and no. of predictor setting fixed,

$$G^2 \text{ and } X^2 = \sum_i \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i} \quad \text{(Pearson)}$$

are used to test goodness-of-fit of model (i.e. H_0: model holds)

G^2 and $X^2 \sim approx.\ \chi^2_{df}$, with df=no. observations - no. model parameters

ex) Wafer defects (Revisited)

$\hat{\mu}_i = 5$ for 10 observations in treatment A

$\hat{\mu}_i = 9$ for 10 observations in treatment B

For log-linear model : $\log(\mu) = \alpha + \beta x$

Deviance $G^2 = 16.3,\ df = 20 - 2 = 18$

Pearson $X^2 = 16.0$

These do not contradict H_0 : model holds, but their use with chi-square dist. is questionable.

- $\hat{\mu}_i$ not that large
- theory applies for fixed df as $n \uparrow$ (happens with contingency tables)

Note

● For GLMs one can study lack of fit using residuals (later chapter)

● Count data often show overdispersion relative to Poisson GLMs. i.e., at fixed x, sample variance > mean. (often caused by subject heterogeneity)

ex) Y= no. times attended religious services in past year.

Suppose $\mu = 25$. Is $\sigma^2 = 25 (\sigma = 5)$?

Negative binomial GLM

More flexible model for count data that has

$$E(Y) = \mu, \quad Var(Y) = \mu + D\mu^2$$

where D is called a dispersion parameter

As $D \to 0$, neg. bin $\to$ Poisson (Can derive as "gamma dist. mixture" of Poissons, where the Poisson mean varies according to a gamma dist.)

ex) GSS data "In past 12 months, how many people have you known personally that were victims of homicide ?

Y	Black	White
0	119	1070
1	16	60
2	12	14
3	7	4
4	3	0
5	2	0
6	0	1

```
data new;
input white black other response;
datalines;
1070 119 55 0
  60  16  5 1
  14  12  1 2
   4   7  0 3
   0   3  1 4
   0   2  0 5
   1   0  0 6
; run;
data new;
   set new;
   count = white; race = 0; output;
   count = black; race = 1; output;
   drop white black other; run;
data new2;
   set new;
   do i = 1 to count;
    output;
   end;
   drop i; run;
❶ proc genmod data=new2;
   model response = race / dist=negbin link=log; /*Negative binomial*/
❷ proc genmod data=new2;
   model response = race / dist=poi link=log scale=pearson;/*(Poisson)*/

 data new;
  set new;
 case = _n_; run;
❸ proc nlmixed  data = new  qpoints=400;
   parms alpha=-3.7 beta=1.90 sigma=1.6;
   eta = alpha + beta*race + u;  mu = exp(eta);
   model response ~ poisson(mu);
   random  u ~ normal(0, sigma*sigma) subject=case;
   replicate count;
run;/*(GLMM)*/
```

Output

The GENMOD Procedure
Model Information

Data Set	WORK.NEW2
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	response
Number of Observations Read	1308
Number of Observations Used	1308

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	1306	412.5964	0.3159
Scaled Deviance	1306	412.5964	0.3159
Pearson Chi-Square	1306	1424.0269	1.0904
Scaled Pearson X2	1306	1424.0269	1.0904
Log Likelihood		-434.4794	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	-2.3832	0.1172	-2.6129	-2.1535	413.49	<.0001
race	1	1.7331	0.2385	1.2657	2.2006	52.82	<.0001
Dispersion	1	4.9429	1.0005	2.9820	6.9038		

NOTE: The negative binomial dispersion parameter was estimated by maximum likelihood.

The GENMOD Procedure
Model Information

Data Set	WORK.NEW2
Distribution	Poisson
Link Function	Log
Dependent Variable	response

Number of Observations Read	1308
Number of Observations Used	1308

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	1306	844.7073	0.6468
Scaled Deviance	1306	483.8821	0.3705
Pearson Chi-Square	1306	2279.8690	1.7457
Scaled Pearson X2	1306	1306.0000	1.0000
Log Likelihood		-283.8854	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	-2.3832	0.1283	-2.6347	-2.1317	344.88	<.0001
race	1	1.7331	0.1937	1.3536	2.1127	80.10	<.0001
Scale	0	1.3212	0.0000	1.3212	1.3212		

NOTE: The scale parameter was estimated by the square root of Pearson's Chi-Square/DOF.

The NLMIXED Procedure
Specifications

Data Set	WORK.NEW
Dependent Variable	response
Distribution for Dependent Variable	Poisson
Random Effects	u
Distribution for Random Effects	Normal
Subject Variable	case
Replicate Variable	count
Optimization Technique	Dual Quasi-Newton
Integration Method	Adaptive Gaussian Quadrature

Dimensions

Observations Used	11
Observations Not Used	3
Total Observations	14
Subjects	1308
Max Obs Per Subject	1
Parameters	3
Quadrature Points	400

Parameters

alpha	beta	sigma	NegLogLike
-3.7	1.9	1.6	500.798926

Iteration History

Iter	Calls	NegLogLike	Diff	MaxGrad	Slope
1	4	500.690098	0.108828	0.178335	-44.7633
2	7	500.689121	0.000977	0.021172	-0.03966
3	9	500.68912	9.399E-7	0.01887	-0.00001
4	10	500.689119	1.008E-6	0.002686	-2.54E-6

NOTE: GCONV convergence criterion satisfied.

Fit Statistics

-2 Log Likelihood	1001.4
AIC (smaller is better)	1007.4
AICC (smaller is better)	1010.8
BIC (smaller is better)	1022.9

The NLMIXED Procedure
Parameter Estimates

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper	Gradient
alpha	-3.6887	0.2435	1307	-15.15	<.0001	0.05	-4.1664	-3.2110	-0.00163
beta	1.8968	0.2458	1307	7.72	<.0001	0.05	1.4145	2.3791	-0.00048
sigma	1.6284	0.1548	1307	10.52	<.0001	0.05	1.3246	1.9321	-0.00269

Model $\log(\mu) = \alpha + \beta x$

Black : $\bar{y} = 0.52,\ s^2 = 1.15$

White : $\bar{y} = 0.09,\ s^2 = 0.16$

$\log\hat{\mu} = -2.38 + 1.73x$ for Poisson or negative binomial.

$$e^{1.73} = 5.7 = \frac{.522}{0.092} = \frac{\overline{Y}_B}{\overline{Y}_W}$$

However, SE for $\hat{\beta} = 1.73$ is 0.194 for Poisson, 0.238 for negative binomial.

Wald 95% C.I. for $e^{\beta} = \dfrac{\mu_B}{\mu_A}$ is

Poisson : $e^{1.73 \pm 1.96(0.194)} = (4.2,\ 7.5)$

Neg. bin. : $e^{1.73 \pm 1.96(0.238)} = (3.5,\ 9.0)$

In accounting for overdispersion, neg. bin. model has wider C.I.'s

LR C.I.'s are $((e^{1.444},\ e^{2.019}) = (4.2,\ 4.7)$ for Poisson, $(3.6,\ 9.2)$ for Neg. bin.

For Neg. bin. model, estimated overdispersion parameter

$\hat{D} = 4.94 (SE = 1.0)$ (Strang evidence of overdispersion)

$$(\widehat{Var}(Y) = \hat{\mu} + \hat{D}\,\hat{\mu}^2 = \hat{\mu} + 4.94\hat{\mu})$$

When Y is a count, safest strategy is to use negative binomial GLM, especially when dispersion parameter is significantly >0

Models for Rates

When y_i have different bases (eg., no. murders for cities with different population sizes), it is more relevant to model rate at which events occur

Let y = count with index t

Sample rate $\frac{y}{t}$,

$$E\left(\frac{y}{t}\right)=\frac{\mu}{t}$$

log-linear model $\log\left(\frac{\mu}{t}\right)=\alpha+\beta x$ or $\log\mu-\log(t)=\alpha+\beta x$

See Text pp82~84 for discussion