Analysis of Hurricane Data

The number of hurricanes with winds greater than 110 mph has declined since the 1950s. Major hurricanes to hit the USA:

1900s	'10s	'20s	'30s	'40s	'50s	'60s	'70s	'80s	'90s
6	8	5	8	8	9	6	4	6	2

Source: National Hurricane Center; Story by: Julie Stacey, USA TODAY, August 1995.

Estimating the rate [assuming no temporal trend]

Model

$E(\text{count}_i) = E(c_i) = T_i\lambda$, where T_i is the amount of experience in period i (10 years in each of first 9 periods; 5 in last one) and λ is the rate (dimension of year^{-1})

Least Squares (LS) Approach (Homoskedastic errors)

Choose λ which minimizes $\sum[c_i - E(\text{count}_i)]^2$ i.e., $\sum[c_i - T_i\lambda]^2$.
We have a regression with a zero intercept $E(c_i) = T_i\lambda$, where T_i is fixed/known and λ is the regression coefficient. One can check that the LS solution is

$$\hat{\lambda} = \sum T_i c_i / \sum T_i^2 \quad [=610/925 = 0.6595 \text{ year}^{-1}]$$

Note it has the form $\hat{\lambda} = \dfrac{\sum (c_i/T_i) \bullet T_i^2}{\sum T_i^2}$ i.e. a weighted average of the period-specific rates c_i/T_i, with weights proportional to T_i^2, i.e., the 9 estimates 6/10, 8/10, ..., 6/10 get weights of $10^2/925$ each, and the rate of 2/5 in the last period gets a weight of 25/925 i.e., each of the rates from periods 1-9 has 4 times the weight given to the last period.

Poisson Model: Maximum Likelihood Estimation

$c_i \sim \text{Poisson}(T_i)$ i.e., $E(c_i) = T_i\lambda = \mu_i$

$L(\lambda \mid c_1 \ldots C10) = e^{-\sum y_i} \bullet \prod(\mu_i^{y_i} / y_i!)$; $\log L = -\sum\mu_i + \sum c_i \log\mu_i$ - constant

$d \log L / d\lambda = -\sum T_i + \sum c_i/\lambda \implies \hat{\lambda} = \sum c_i/\sum T_i = \dfrac{\sum(c_i/T_i) \bullet T_i}{\sum T_i}$,

a **different** weighted average of the rates.

In our example,
$$\hat{\lambda} = \frac{\frac{10}{95}\bullet\frac{6}{10} + \frac{10}{95}\bullet\frac{8}{10} + \ldots + \frac{10}{95}\bullet\frac{6}{10} + \frac{5}{95}\bullet\frac{2}{5}}{\frac{10}{95} + \ldots\ldots + \frac{10}{95} + \frac{5}{95}}$$

i.e. $\hat{\lambda} = \dfrac{62}{95} = 0.6526/\text{year}^{-1}$

This is slightly lower than the model that considered the variation to be Poisson around each mean. The lower than average 10th rate 2/5 or 0.4/year gets relatively more weight in the Poisson model (5/95 or 5.2%, vs 25/925 or 2.7%). The Poisson model assumes that variation is related to the mean [var(c)   E(c)], whereas the least squares approach assumes equal variation around E(c) no matter what the E is.

Poisson Model: Iteratively Reweighted Least Squares

In fact, the approach used by the program GLIM is not based on directly solving the equations suggested by maximizing the likelihood, but rather on iteratively reweighted least squares, i.e., it finds the λ which minimizes the weighted sum of squares

$$\sum[c_i - T_i\lambda]^2 w_i$$

The weights w_i are proportional to $1/\text{var}(c_i) = \dfrac{1}{T_i\lambda}$ so that c's with larger var's get relatively less weight.

In this simple example, with no covariates, the estimation requires only one iteration. The weighted least squares estimate, based on a trial value $\lambda_{(0)}$, is

$$\hat{\lambda}_{(1)} = \frac{\sum c \bullet T \bullet w}{\sum T^2 \bullet w} = \frac{\sum c_i \bullet T \bullet \frac{1}{\lambda_{(0)}T}}{\sum T^2 \bullet \frac{1}{\lambda_{(0)}T}} = \frac{\sum c}{\sum T}$$

In practice, with covariates, the procedure requires further refinement of the w's, leading to refinement of $\hat{\lambda}$ and vice versa. In this particular example, the value of λ cancels out in the weights.

Analysis of Hurricane Data

Estimating the rate [assuming NO trend] ...with GLIM

```
$units 10 $data decade time count $read
0 10 6
1 10 8
2 10 5
3 10 8
4 10 8
5 10 9
6 10 6
7 10 4
8 10 6
9  5 2
$yvar count
```

Hurricanes striking US this century
Source USA Today Aug 1995

All computations carried out with GLIM

*E(count) = λ * time*

Normal Errors, Identity Link

```
$fit time -%gm $dis e
 estimate      s.e.  parameter
   0.6595   0.05339  TIME
```

$\Rightarrow \hat{\lambda} = 0.6595\,yr^{-1}$

*E(count) = λ * time*

Poisson Errors, Identity Link

```
$err poisson $link I
$fit time - %gm $dis e $
 estimate    s.e.  parameter
   0.6526 0.08288  TIME
```

$\Rightarrow \hat{\lambda} = \frac{62}{95} = 0.6526\,yr^{-1}$

*E(count) = λ * time => ln[E]=ln[λ] + ln[time]*

offset

Poisson Errors, Log Link

```
$link l$ $calc logt=%log(time)
$offset logt $
$fit $dis e$
 estimate   s.e. parameter
  -0.4267 0.1270 1
```

$\Rightarrow ln[\hat{\lambda}] = ln[\frac{62}{95}] = -0.4267$

Modeling a trend in the rate ...with GLIM

*E(count) = (λ_0 + ß *Decade) * time*

*= λ_0 * time + ß * Decade*time*

Normal Errors, Identity Link

```
$calc dt=decade*time

$fit time + dt -%gm $dis e
      estimate       s.e.     parameter
```

	estimate	s.e.	parameter
$\Rightarrow \hat{\lambda}_0 =$	0.7361	0.09921	TIME
$\Rightarrow \hat{ß}_0 =$	-0.01853	0.02015	DT

*E(count) = (λ_0 + ß*decade) * time*

Poisson Errors, Identity Link

```
$err poisson $link I

$fit time + dt - %gm $dis e $
```

	estimate	s.e.	parameter
$\Rightarrow \hat{\lambda}_0 =$	0.7551	0.1592	TIME
$\Rightarrow \hat{ß} =$	-0.02404	0.02997	DT

*E(count) = (λ_0 $e^{ß*decade}$) * time*

*Ln[E] = ln[λ_0] + ß * decade + ln[time]*

offset

Poisson Errors, Log Link

```
$offset logt
$fit decade $dis e$
```

	estimate	s.e.	parameter
$\Rightarrow ln[\hat{\lambda}_0] =$	-0.2911	0.2253	1
$\Rightarrow \hat{ß} =$	-0.03276	0.04634	DECA

Estimating the rate [assuming NO trend]
...with SAS Proc GENMOD in SAS version ≥6.10

```
data a;
input  decade time count;
x0=1;
dt=decade*time;
lntime=log(time);
lines;
```

see page 2 for 3 statistical models and interpretation of fitted coefficients

```
proc genmod;
model count = time / dist=normal noint;

Distribution: NORMAL   Link Function: IDENTITY

 Parameter DF  Estimate  Std Err ChiSquare  Pr>Chi

 INTERCEPT  0    0.0000   0.0000         .       .
 TIME       1    0.6595   0.0506  169.5216  0.0001
 SCALE      1    1.5404   0.3445         .       .
```

```
proc genmod;
model count = time /
       link = identity dist=Poisson noint;

Distribution: POISSON  Link Function: IDENTITY

Parameter  DF  Estimate  Std Err   ChiSquare  Pr>Chi

 INTERCEPT  0    0.0000   0.0000          .      .
 TIME       1    0.6526   0.0829    62.0000  0.0001
 SCALE      0    1.0000   0.0000          .      .
```

```
proc genmod;
model count = x0 / dist=Poisson
                      offset =lntime;

Distrn:POISSON; Link: LOG; Offset Variable: LNTIME

Parameter  DF  Estimate  Std Err ChiSquare Pr>Chi

INTERCEPT   1  -0.4267    0.1270   11.2908 0.0008
X0          0   0.0000    0.0000     .      .
SCALE       0   1.0000    0.0000     .      .
```

Modeling a trend in the rate
....with SAS Proc GENMOD in SAS version ≥6.10

see page 2 for 3 statistical models of trend

```
proc genmod;
model count =time dt / dist=Normal noint;

 Distrn: NORMAL;  Link IDENTITY

 Parameter  DF Estimate   Std Err ChiSquare  Pr>Chi

 INTERCEPT  0   0.0000    0.0000         .       .
 TIME       1   0.7361    0.0887  68.8132  0.0001
 DT         1  -0.0185    0.0180   1.0570  0.3039
 SCALE      1   1.4650    0.3276         .       .
```

```
proc genmod;
model count = time dt / dist=Poisson noint
link=Identity;

Distrn: POISSON; Link:IDENTITY

 Parameter  DF Estimate  Std Err ChiSquare Pr>Chi

  INTERCEPT  0   0.0000  0.0000         .       .
  TIME       1   0.7551  0.1656   20.7929  0.0001
  DT         1  -0.0240  0.0318    0.5708  0.4499
  SCALE      0   1.0000  0.0000         .       .
```

```
proc genmod;
model count = decade / dist=poisson
offset=lntime;

Distrn:POISSON; Link:LOG;  Offset Variable: LNTIME

Parameter   DF Estimate Std Err ChiSquare Pr>Chi

INTERCEPT    1  -0.2911  0.2254  1.6690  0.1964
DECADE       1  -0.0328  0.0464  0.4995  0.4797
SCALE        0   1.0000  0.0000       .       .
```