Models In Epidemiology And Biostatistics
Gordon Hilton Fick

The Logarithmic Link

aka :
Log Binomial Models
Binomial Regression With The Log Link

$$\log(p) = \sum_{i=0}^{k} \beta_i x_i$$

A 2x2 Table from a cohort study

p_1 = P(D | E)

p_0 = P(D | not E)

$$\log(p) = \beta_0 + \beta_1 E$$

$$E = 0 : \log(p_0) = \beta_0$$

$$E = 1 : \log(p_1) = \beta_0 + \beta_1$$

$$\beta_1 = \log(p_1) - \log(p_0) = \log \frac{p_1}{p_0}$$

$$RR = \frac{p_1}{p_0} = \exp(\beta_1)$$

2 2x2 Tables from a cohort study

p_{1j} = P(D | E and Strata j)

p_{0j} = P(D | not E and Strata j)

$$\log(p)=\beta_0+\beta_1 E+\beta_2 S+\beta_3 ES$$

$$S=0:\log(p)=\beta_0+\beta_1 E \quad S=1:\log(p)=\beta_0+\beta_2+(\beta_1+\beta_3)E$$

$$\beta_1=\log(p_{10})-\log(p_{00})=\log\frac{p_{10}}{p_{00}}$$

$$\beta_1+\beta_3=\log(p_{11})-\log(p_{01})=\log\frac{p_{11}}{p_{01}}$$

$$RR_0=\frac{p_{10}}{p_{00}}=\exp(\beta_1) \quad RR_1=\frac{p_{11}}{p_{01}}=\exp(\beta_1+\beta_3)$$

Compare with Logistic Regression

When constructing models with the log link, one uses the same processes as with Logistic Regression. Now, with log link, log odds are replaced with log probabilities and odds ratios are replaced with probability ratios. All of the interpretations are the same except the changes noted above.

Back to the Kalbfleisch data

```
. cs suc tr, by(surg)

           surg |       RR       [95% Conf. Interval]   M-H Weight
----------------+-------------------------------------------------
              1 |            2     .8342841    4.79453     4.545455
              2 |          1.9     1.759944   2.051202     45.45455
----------------+-------------------------------------------------
          Crude |    .3861386     .3348359   .4453018
    M-H combined |    1.909091      1.71543   2.124615
----------------------------------------------------------------
Test of homogeneity (M-H)      chi2(1) =    0.026  Pr>chi2 = 0.8724

. cs fail tr, by(surg)

           surg |       RR       [95% Conf. Interval]   M-H Weight
----------------+-------------------------------------------------
              1 |    .9473684       .90163   .9954271     86.36364
              2 |          .1     .0424614   .2355078     45.45455
----------------+-------------------------------------------------
          Crude |    1.521008     1.431053   1.616618
    M-H combined |    .6551724      .580039   .7400379
----------------------------------------------------------------
Test of homogeneity (M-H)      chi2(1) =  231.866  Pr>chi2 = 0.0000
```

```
gen s=surg-1
binreg suc tr s ts, rr coef
```

fail	EIM Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
tr	.6931472	.4460745	1.55	0.120	-.1811428	1.567437
s	2.302585	.4370155	5.27	0.000	1.44605	3.15912
ts	-.0512933	.4477821	-0.11	0.909	-.92893	.8263435
_cons	-2.995732	.4358699	-6.87	0.000	-3.850022	-2.141443

```
binreg fail tr s ts, rr coef
```

fail	EIM Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
tr	-.0540672	.0252473	-2.14	0.032	-.103551	-.0045834
s	-.6418539	.0390681	-16.43	0.000	-.7184259	-.5652818
ts	-2.248518	.4377442	-5.14	0.000	-3.106481	-1.390555
_cons	-.0512933	.0229416	-2.24	0.025	-.096258	-.0063286

binreg suc tr s, rr coef

suc	EIM Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
tr	.6422541	.0389	16.51	0.000	.5660114	.7184968
s	2.253795	.0952433	23.66	0.000	2.067121	2.440468
_cons	-2.947204	.0998327	-29.52	0.000	-3.142873	-2.751536

An example that illustrates trouble

```
. binreg dis e1 e2 e3,rr coeff

Variance function: V(u) = u*(1-u)                  [Bernoulli]

Link function    : g(u) = ln(u)                    [Log]

------------------------------------------------------------------------------
             |                 EIM
         dis |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          e1 |   1.066394    .212288     5.02   0.000     .6503173    1.482471
          e2 |   1.323152   .2086871     6.34   0.000     .9141332    1.732172
          e3 |   .9215181   .2401699     3.84   0.000     .4507938    1.392242
       _cons |    -2.8006   .1996354   -14.03   0.000    -3.191879   -2.409322
------------------------------------------------------------------------------
```

The fitted values must be negative or zero but ...

```
. predict lr,xb

. tab lr

     Linear |
 prediction |      Freq.     Percent        Cum.
------------+-----------------------------------
    -2.8006 |        210       62.31       62.31
  -1.879082 |         50       14.84       77.15
  -1.734206 |         36       10.68       87.83
  -1.477448 |         17        5.04       92.88
  -.8126882 |         14        4.15       97.03
  -.5559298 |          7        2.08       99.11
  -.4110539 |          2        0.59       99.70
   .5104642 |          1        0.30      100.00  !! nonsense !!
------------+-----------------------------------
      Total |        337      100.00
```

The boundaries on the parameters to ensure that the fitted values are negative or zero

There are 8 fitted values corresponding to the possible values of the explanatory variables.
The crucial boundary here is:

$$\beta_0+\beta_1+\beta_2+\beta_3=0$$

Maybe the MLE is very close to this boundary or maybe even on this boundary

If the MLE is on this boundary ...

.... then we must have that :

$$\beta_0 = -\beta_1 - \beta_2 - \beta_3$$

and so

$$\log(p) = \beta_1(e_1 - 1) + \beta_2(e_2 - 1) + \beta_3(e_3 - 1)$$

We then get a 'tentative' set of estimated regression coefficients

```
. gen v1=e1-1

. gen v2=e2-1

. gen v3=e3-1

. binreg dis v1 v2 v3,rr coeff nocons

Variance function: V(u) = u*(1-u)                  [Bernoulli]

Link function    : g(u) = ln(u)                    [Log]


------------------------------------------------------------------------------
             |                 EIM
         dis |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          v1 |   .8112575   .2407559     3.37   0.001     .3393846     1.28313
          v2 |   1.051051    .237187     4.43   0.000     .5861735    1.515929
          v3 |   .7476093   .2644474     2.83   0.005     .2293018    1.265917
------------------------------------------------------------------------------
```

The fitted values are now OK

```
. predict lrc,xb

.  table e1 e2 e3,c(mean lr mean lrc)

----------------------------------------------------
          |                   e3 and e2
          | --------- 0 --------    --------- 1 --------
       e1 |         0          1            0          1
----------+-----------------------------------------
        0 |   -2.8006  -1.477448    -1.879082  -.5559298
          | -2.609918  -1.558867    -1.862309  -.8112575
          |
        1 | -1.734206  -.4110539    -.8126882   .5104642
          | -1.798661  -.7476093    -1.051051          0
----------------------------------------------------
```

We now refit the original model with 'starting values'

```
. gen elrc=exp(lrc)

. binreg dis e1 e2 e3,rr coeff mu(elrc)

Variance function: V(u) = u*(1-u)                  [Bernoulli]

Link function    : g(u) = ln(u)                    [Log]


------------------------------------------------------------------------------
             |                 EIM
         dis |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          e1 |   .8112557     .24076     3.37   0.001     .3393748    1.283137
          e2 |   1.051049   .2371934     4.43   0.000     .5861583    1.515939
          e3 |   .7476083   .2644499     2.83   0.005      .229296    1.265921
       _cons |  -2.609917   .1942437   -13.44   0.000    -2.990628   -2.229206
------------------------------------------------------------------------------
```

R packages : lbreg & logbin

- Very recently [2018], a methodology for correctly fitting log binomial models has been released.
- Presumably, Stata, SAS & SPSS will catch up 'soon'.
- No need to use fake 'approximations'.

The Identity Link

Aka:

Identity Binomial Models
Binomial Regression With The Identity Link

$$p = \sum_{i=0}^{k} \beta_i x_i$$

A 2x2 Table from a cohort study

p_1 = P(D | E)

p_0 = P(D | not E)

$$p = \beta_0 + \beta_1 E$$

$$E = 0 : \; p_0 = \beta_0$$

$$E = 1 : \; p_1 = \beta_0 + \beta_1$$

$$\beta_1 = p_1 - p_0$$

2 2x2 Tables from a cohort study

p_{1j} = P(D | E and Strata j)

p_{0j} = P(D | not E and Strata j)

$$p = \beta_0 + \beta_1 E + \beta_2 S + \beta_3 ES$$

$$S = 0 : \; p = \beta_0 + \beta_1 E \quad S = 1 : \; p = \beta_0 + \beta_2 + (\beta_1 + \beta_3) E$$

$$\beta_1 = p_{10} - p_{00}$$

$$\beta_1 + \beta_3 = p_{11} - p_{01}$$

$$RD_0 = p_{10} - p_{00} = \beta_1 \quad RD_1 = p_{11} - p_{01} = \beta_1 + \beta_3$$

$$RD_1 - RD_0 = \beta_3$$

Comparisons

Now, with Rate Difference Regression, there are no logarithms and there are no ratios.

Interpretations now involve rates [instead of log odds or log rates] and rate differences [instead of odds ratios or rate ratios]

```
. binreg fail tr s ts,rd

------------------------------------------------------------------------------
             |                 EIM
        fail | Risk Diff.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          tr |       -.05   .0237697    -2.10   0.035    -.0965878   -.0034122
           s |       -.45   .0269258   -16.71   0.000    -.5027736   -.3972264
          ts |        -.4   .0359166   -11.14   0.000    -.4703952   -.3296048
       _cons |        .95   .0217945    43.59   0.000     .9072836    .9927164
------------------------------------------------------------------------------

. binreg suc tr s ts,rd

------------------------------------------------------------------------------
             |                 EIM
         suc | Risk Diff.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          tr |        .05   .0237697     2.10   0.035     .0034122    .0965878
           s |        .45   .0269258    16.71   0.000     .3972264    .5027736
          ts |         .4   .0359166    11.14   0.000     .3296048    .4703952
       _cons |        .05   .0217945     2.29   0.022     .0072836    .0927164
------------------------------------------------------------------------------
```

binreg suc tr s,rd mu(muinit)

...needed 'very good starting values' to 'converge' and get the 'right' estimates

suc	EIM Risk Diff.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
tr	.0926905	.0166177	5.58	0.000	.0601204	.1252605
s	.5131088	.0196188	26.15	0.000	.4746567	.551561
_cons	.0212516	.0139003	1.53	0.126	-.0059925	.0484957