市販後薬剤副作用調査
Person-time & Poisson Regression を用いた疫学研究の解析

薬剤 A は既に市販されています。最近この薬剤 A の使用者で胃腸障害が多いのではないかと指摘されています。そこで我々は入院するほどの重篤な胃腸障害の発生状況を retrospective cohort study の形で調査してみました。重篤な胃腸障害は大きな問題ですが、頻度は少ないため person-time を用いて解析します。このような希な発生（分母に比べて分子が小さい）を解析する場合には poisson regression analysis を用います。

1982-1986 年までの間、228,392 人に対して薬剤が処方されていました。服薬期間も考慮にいれると、679,075 person-years となります。この中から 2,302 人の胃腸障害による入院を認めました。極僅かの発生のため、Poisson でデータを解析することにします。感染症ではないので independence assumption を満たし、平均およそ 3 年間の服用ですが、発生時期に関しては、非使用者(>151 日)、現在の使用者(0 – 30 日)、最近の使用者(31 – 60 日)、過去の使用者(61 – 150 日)の４つに分けて考えることにしました。このことにより Cox Hazard model でなくとも時間的要素も大まかに検討することができます。それぞれのカテゴリー内では発生頻度は同じと考えています。更に、性別、年齢、などで副作用が違う傾向にないかどうかも合わせて調べてみようと思います。

それでは実際のデータを用いて解析してみましょう。

<u>変数の説明</u>

status 0: 非使用者(>151 日)
1: 現在の使用者(0 – 30 日)
2: 最近の使用者(31 – 60 日)
3: 過去の使用者(61 – 150 日)

暦年 1: 1982
2: 1983
3: 1984
4: 1985
5: 1986

年齢 1: 0 – 4
2: 5 – 9
……
20:

性 1: 男性
2: 女性

gi: gastrointestinal
0: 胃腸障害による入院なし
1: 胃腸障害による入院あり

ここで示した数値はただのシンボルです。その他の副作用も入力してありますが、ここでは使用しません。

上の変数(variables)の組み合わせは全部で何通りありますか？ 4 x 5 x 20 x 2 = 800 です。胃腸障害を起こして入院した 2,302 人は、800 種類のどれかの組み合わせにあてはまります。実際. list age year status sex pyr gi をみてもらえば、０人のところもあれば 6 人のところもあります。ただ 800 種類の組み合わせうち副作用の全く無かった組み合わせが１つだけあったので、その行ははずしてあります。そのため解析は 799 で行ないます。もしも空のカラムが多ければ、解析数は減ることになります。

データの構造

どのような変数があるか確認してみましょう。

```
. describe

Contains data from G:\shared\EPI204D\Sask.dta
  obs:           799 (max=        1,089)
 vars:            12 (max=           99)      11 Apr 1997 13:04
width:            14 (max=          200)
-------------------------------------------------------------------------------
   1. status    byte   %9.0g
   2. year      byte   %9.0g
   3. age       byte   %9.0g
   4. sex       byte   %9.0g
   5. pyr       int    %9.0g
   6. gi        byte   %9.0g
   7. renal     byte   %9.0g
   8. skin      byte   %9.0g
   9. liver     byte   %9.0g
  10. allergic  byte   %9.0g
  11. blood     byte   %9.0g
  12. other     int    %9.0g
-------------------------------------------------------------------------------
Sorted by:  status
```

まずは indicator variable を作っていきます。Indicator variable とは 0 か 1 の 2 進法で表される変数です。男性は sex1, 女性は sex2 と最初に決めましたが、男性を 0, 女性を 1, にするか、男性を 1、女性を 0 にする２つのオプションがあります。

単純な解析をしてみましょう。男性、女性の人数とその比はどの程度ですか？

```
.tabulate sex, generate(sex)

        sex |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |        400       50.06       50.06
          2 |        399       49.94      100.00
------------+-----------------------------------
      Total |        799      100.00
```

ほぼ半々のようです。STATA 内部では sex1 と sex2 を異なる変数として捉え、男性であれば sex1=1, sex2=0, 女性であれば sex1=0, sex2=1 となっています。ちょっと確認してみましょう。

```
. list sex sex1 sex2 in 1/4, nodisplay

           sex      sex1      sex2
  1.         1         1         0
  2.         2         0         1
  3.         1         1         0
  4.         2         0         1
```

薬剤の使用状況、すなわち status についてはどうでしょうか？

```
. tab status, gen(status)

     status |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |        200       25.03       25.03
          1 |        200       25.03       50.06
          2 |        199       24.91       74.97
          3 |        200       25.03      100.00
------------+-----------------------------------
      Total |        799      100.00
```

ほぼ同じ人数が分布しているようです。それでは暦年はどうでしょう？

```
. tab year, gen(year)

       year |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |        160       20.03       20.03
          2 |        160       20.03       40.05
          3 |        160       20.03       60.08
          4 |        159       19.90       79.97
          5 |        160       20.03      100.00
------------+-----------------------------------
      Total |        799      100.00
```

これについてもほぼ同数です。

年齢は 5 年毎、0 歳から 100 歳までを 20 に区分してあります。

```
. tab age, gen(age)

        age |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |         39        4.88        4.88
          2 |         40        5.01        9.89
          3 |         40        5.01       14.89
          4 |         40        5.01       19.90
          5 |         40        5.01       24.91
          6 |         40        5.01       29.91
          7 |         40        5.01       34.92
          8 |         40        5.01       39.92
          9 |         40        5.01       44.93
         10 |         40        5.01       49.94
         11 |         40        5.01       54.94
         12 |         40        5.01       59.95
         13 |         40        5.01       64.96
         14 |         40        5.01       69.96
         15 |         40        5.01       74.97
         16 |         40        5.01       79.97
         17 |         40        5.01       84.98
         18 |         40        5.01       89.99
         19 |         40        5.01       94.99
         20 |         40        5.01      100.00
------------+-----------------------------------
      Total |        799      100.00
```

ほぼ均等です。年齢は 5 年で区切ってあるので均等になるはずです。何年で区切るかは自由であり、5 年に拘る必要はありません。もしも小児が少ないとき、1-15 歳とまとめてしまうことも可能です。

このように tab を用いることにより STATA は age1, age2, age3 などのように age それぞれのカテゴリーを１つの変数として認識してきれます。その証拠に左の variables の中には age1, age2 といった変数が増えているはずです。

疫学テーブルを用いたIncidence Rates の解析

先にこの解析では分子が小さいので poisson で解析する方針をたてましたが、その前に通常の binomial で解析し、Poisson がどの程度違うかあとで比べてみたいと思います。
女性を exposure と考え（変な話ですが）、女性の方が男性より胃腸障害を起こしやすいかどうか検討します。ir は incidence rates の略で、その次に結果、この場合 gi 胃腸障害による入院の有無を置きます。そして、検討したい変数を書いて、STATA に person-year に基いたデータですよと教えるため pyr を最後に加えます。

```
. ir gi sex2 pyr

                 | sex== 2.0000           |
                 |   Exposed   Unexposed  |      Total
-----------------+------------------------+----------
              gi |      1183        1119  |       2302
             pyr |    376938      301548  |     678486
-----------------+------------------------+----------
                 |                        |
  Incidence Rate |  .0031384    .0037109  |   .0033928
                 |                        |
                 |      Point estimate    |  [95% Conf. Interval]
                 |------------------------+----------------------
 Inc. rate diff. |        -.0005724       | -.0008539   -.0002909
 Inc. rate ratio |         .8457484       |   .778707    .9185976  (exact)
 Prev. frac. ex. |         .1542516       |  .0814024     .221293  (exact)
 Prev. frac. pop |         .0856956       |
                 +-----------------------------------------------
                   (midp)   Pr(k<=1183) =                  0.0000  (exact)
                   (midp) 2*Pr(k<=1183) =                  0.0001  (exact)
```

男女を合わせた crude incidence rate は 0.003393 でした。男性、女性の Incidence rate はそれぞれ 0.003711, 0.003138 であり、Rate Ratio は 0.8457 であり、95%CI は１を含んでいませんから女性の方が薬の副作用による胃腸炎で入院しにくいことになります。また rate difference をみても 0 を含んでいないことから、女性の方が薬剤 A を服用した際、胃腸障害を起こしにくいことがわかります。

Poisson Regression 法を用いた解析

　それではいよいよ Poisson regression を用いた解析に入ります。まずは男女合わせた crude data を算出してみましょう。

```
. poisson gi, exposure(pyr)

Iteration 0:   log likelihood = -2592.4007
Iteration 1:   log likelihood = -2592.4007

Poisson regression                                Number of obs   =        799
                                                  LR chi2(0)      =       0.00
                                                  Prob > chi2     =          .
Log likelihood = -2592.4007                       Pseudo R2       =     0.0000

------------------------------------------------------------------------------
      gi |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
   _cons |  -5.686086   .0208424   -272.814   0.000      -5.726936   -5.645235
     pyr | (exposure)
------------------------------------------------------------------------------
```

Crude incidence = exp (-5.686) = 0.003393

Crude data は前に行った incidence rate の値と同じになりました。

それでは男女分けて行なってみます。S
ex2 でなければ必然的に sex1 となるので、sex1 と sex2 を並べる必要はありません
(`poisson gi sex1 sex2, exposure(pyr)`とする必要はありません)。

```
. poisson gi sex2, exposure(pyr)

Iteration 0:   log likelihood = -2584.3545
Iteration 1:   log likelihood = -2584.3524
Iteration 2:   log likelihood = -2584.3524

Poisson regression                                Number of obs   =        799
                                                  LR chi2(1)      =      16.10
                                                  Prob > chi2     =     0.0001
Log likelihood = -2584.3524                       Pseudo R2       =     0.0031

------------------------------------------------------------------------------
      gi |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    sex2 |  -.1675334   .0417009     -4.018   0.000      -.2492656   -.0858011
   _cons |  -5.596494   .0298941   -187.211   0.000      -5.655085   -5.537902
     pyr | (exposure)
------------------------------------------------------------------------------
```

Rate Ratio = exp (- 0.1675) = 0.8458
Incidence rate (men) = exp (- 5.5965) = 0.003711
Incidence rate (women) = exp (- 5.5965 – 0.1675) = 0.003139
「疫学テーブルを用いた incidence rate の解析」と同じ結果を得ました。

更に以下のようなコマンドをすると incidence rate ratio を計算してくれます。

```
. poisson, irr

Poisson regression                                Number of obs   =        799
                                                  LR chi2(1)      =      16.10
                                                  Prob > chi2     =     0.0001
```

```
Log likelihood = -2584.3524                       Pseudo R2       =     0.0031

------------------------------------------------------------------------------
      gi |        IRR   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    sex2 |   .8457484   .0352685     -4.018   0.000        .779373    .9177767
     pyr | (exposure)
------------------------------------------------------------------------------
```

先の「疫学テーブルを用いた incidence rate の解析」と同じ結果を得ました。このことは、分子が分母に比べて小さいため Poisson regression model で近似できるということです。つまり rare assumption (分母が分子に比べ極端に大きい)が成立するということです。

最初のモデルでは男女差がないものと考えて計算しており、2 番目のモデルでは男女差が存在するものとして計算しています。最初のモデル の log likelihood は-2592.401(L0)であり、2 番目におけるそれは-2584.352 (L1)なので,

PseudoR2：1 – L1/L0 = 1 – (-2584.4)/(-2592.4) = 0.0031

となります。更に、変数を加えることによってモデルの log likelihood の値が十分 0 に近付くと、「その変数を加えることに価値があった」、すなわち「その変数を加えて同時解析することにより、より真実に近い結論を下すことができる」と解釈します。

-2592.401 - (- 2584.352) = - 8.049

これは自由度 1 のχ^2 を上回っていますので有意です。すなわち、女性は有意に薬剤Aによる胃腸症状で入院しにくいと言えます。

非使用者を reference group (=1)としてそれぞれの薬剤使用歴(status)における胃腸障害入院歴の Incidence Rate Ratio (IRR)を疫学テーブルより算出してください。

. sort status

sort し直すことによって status 0 (非使用者)は status 1 に、status 1 (現在の使用者)は status 2 に、status 2 (最近の使用者)は status 3 に、status 3 (過去お使用者)は status 4 になります。まずは status 3 と status 4 は無視して status 1（非使用者）に対して status 2 （現在の使用者）がより多く（あるいは少なく）胃腸障害を発生しているかどうかみます。

```
. ir gi status2 pyr if status3==0&status4==0

                 | status== 1.0000        |
                 |   Exposed   Unexposed  |     Total
-----------------+------------------------+----------
              gi |      1052         685  |      1737
             pyr |    107367      406751  |    514118
-----------------+------------------------+----------
                 |                        |
  Incidence Rate |  .0097982    .0016841  |  .0033786
                 |                        |
                 |      Point estimate    |  [95% Conf. Interval]
                 |------------------------+----------------------
 Inc. rate diff. |         .0081141       |  .0075087    .0087195
 Inc. rate ratio |         5.818124       |  5.279347    6.415215  (exact)
 Attr. frac. ex. |         .8281233       |  .8105826    .8441206  (exact)
 Attr. frac. pop |         .5015462       |
                 +-----------------------------------------------
                   (midp)   Pr(k>=1052) =                 0.0000  (exact)
                   (midp) 2*Pr(k>=1052) =                 0.0000  (exact)
```

status2(現在の使用者)は、そうでない人と比較して 5.8 倍まで胃腸障害による入院の incidence rate が増えています。しかも 95%CI も比較的狭く、ｐも十分小さく、統計学的にも有意です。以下同じように非使用者と最近の使用者とを比較してみましょう。

```
. ir gi status3 pyr if status2==0&status4==0

                 | status== 2.0000        |
                 |   Exposed   Unexposed  |     Total
-----------------+------------------------+----------
              gi |       315         685  |      1000
             pyr |     62386      406751  |    469137
-----------------+------------------------+----------
                 |                        |
  Incidence Rate |  .0050492    .0016841  |  .0021316
                 |                        |
                 |      Point estimate    |  [95% Conf. Interval]
                 |------------------------+----------------------
 Inc. rate diff. |         .0033651       |  .0027935    .0039368
 Inc. rate ratio |         2.998206       |  2.615335    3.431182  (exact)
 Attr. frac. ex. |         .6664672       |  .6176398    .7085553  (exact)
 Attr. frac. pop |         .2099372       |
                 +-----------------------------------------------
                   (midp)   Pr(k>=315) =                  0.0000  (exact)
                   (midp) 2*Pr(k>=315) =                  0.0000  (exact)
```

3 倍まで増えています。しかも 95%CI も比較的狭く、ｐも十分小さく、最近の薬剤 A 使用者では、非使用者と比較して明らかに胃腸障害による入院が多くなっているのがわかります。今度は非使用者と過去の使用者と比較してみましょう。

```
. ir gi status4 pyr if status2==0&status3==0
```

```
                 | status== 3.0000         |
                 |   Exposed   Unexposed   |     Total
-----------------+-------------------------+----------
              gi |       250         685   |       935
             pyr |    101982      406751   |    508733
-----------------+-------------------------+----------
                 |                         |
  Incidence Rate |  .0024514    .0016841   |  .0018379
                 |                         |
                 |      Point estimate     |  [95% Conf. Interval]
                 |-------------------------+------------------------
 Inc. rate diff. |         .0007673        |  .0004383    .0010963
 Inc. rate ratio |         1.455642        |  1.254303     1.68492  (exact)
 Attr. frac. ex. |         .3130178        |  .2027447    .4064999  (exact)
 Attr. frac. pop |         .0836946        |
                 +--------------------------------------------------
                   (midp)   Pr(k>=250) =                   0.0000  (exact)
                   (midp) 2*Pr(k>=250) =                   0.0000  (exact)
```

1.5 倍まで増えています。しかも 95%CI も比較的狭く、ｐも十分小さく、最近の薬剤 A 使用者では、非使用者と比較して明らかに胃腸障害による入院が多くなっているのがわかります。薬剤 A 服用開始から時間が経てば経つほど胃腸障害を起こすリスクは減る傾向にあります。逆に「服用中止後も日が浅ければ入院するほどの胃腸障害を起こし得るので注意が必要である」とも言え、注目すべきデータです。

非使用者を reference group としてそれぞれの薬剤使用歴(status)における胃腸障害入院歴の Incidence Rate Ratios (IRR)を Poisson Regression Model より算出してください。

Status1 を reference とするときは、status2-status4 というふうにコマンドします。

```
. poisson gi status2-status4, exposure(pyr)

Iteration 0:   log likelihood = -1912.5658
Iteration 1:   log likelihood = -1912.5565
Iteration 2:   log likelihood = -1912.5565

Poisson regression                                Number of obs   =        799
                                                  LR chi2(3)      =    1359.69
                                                  Prob > chi2     =     0.0000
Log likelihood = -1912.5565                       Pseudo R2       =     0.2622

------------------------------------------------------------------------------
      gi |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
 status2 |   1.760978   .0490961     35.868   0.000       1.664751    1.857204
 status3 |   1.098014   .0680768     16.129   0.000       .9645859    1.231442
 status4 |    .375447   .0738908      5.081   0.000       .2306236    .5202703
   _cons |  -6.386538    .038208   -167.152   0.000      -6.461424   -6.311651
     pyr | (exposure)
------------------------------------------------------------------------------

. poisson, ir

Poisson regression                                Number of obs   =        799
                                                  LR chi2(3)      =    1359.69
                                                  Prob > chi2     =     0.0000
Log likelihood = -1912.5565                       Pseudo R2       =     0.2622

------------------------------------------------------------------------------
      gi |        IRR   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
 status2 |   5.818124    .285647     35.868   0.000       5.284359    6.405804
 status3 |   2.998206   .2041084     16.129   0.000       2.623701    3.426167
 status4 |   1.455642   .1075586      5.081   0.000       1.259385    1.682482
     pyr | (exposure)
------------------------------------------------------------------------------

. poisson gi status2-status4, irr exposure(pyr)

poisson regression, normalized by pyr                            Number of obs        =
799
Goodness-of-fit chi2(795) = 2340.462                               Model chi2 (3)
= 1359.688
Prob > chi2                          =      0.0000                           Prob >
chi2            =       0.0000
Log Likelihood                   = -1912.566                            Pseudo  R2
=       0.2622


------------------------------------------------------------------------------
      gi |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
 status2 |   5.818124   .285647     35.686   0.000       5.284091    6.406128
 status3 |   2.998206  .2041084     16.129   0.000       2.623517    3.426408
 status4 |  1.455642   .1075586      5.081   0.000       1.259289    1.682611
------------------------------------------------------------------------------
```

status 1 (> 150): - 6.387	exp (-6.387) = 1.68×10^{-3}	1.00
status 2 (1 – 30): - 6.387 + 1.761	exp (- 6.387 + 1.761) = 9.79×10^{-3}	5.82
status 3 (31 – 60): - 6.387 + 1.098	exp (- 6.387 + 1.098) = 5.05×10^{-3}	3.00

status 4 (61 – 150): - 6.387 + 0.375　exp (- 6.387 + 0.375) = 2.45 x 10^{-3}　1.45

疫学的表より計算したそれぞれの薬剤A使用状況のリスクと一致しました。表の上左にはGoodness-of-fit chi2(795) があり、右上にはmodel chi2(3) があります。左は 799 の組み合わせ全てを考慮したモデルとどれくらい異なるかを示しており、右は単純なintercept のみのモデルとどれくらい違うかを示しています。それぞれχ^2 で示していますから、数値が大きい程異なることになります。そしてそれぞれの下にはｐ値が示してあり、ここで用いた４つのパラメータによりintercept よりはましだが、799 のパラメータを用いた場合(saturated model, maximum log likelihood)より劣ると評価できます。パラメータを少しずつ加えていって 799 のパラメータを用いた場合と統計学的に変わらなくなった時点（数値が一致する必要はない）でとめる方法と、逆に 799 のパラメータから始めて、パラメータを減らしていく方法があります。

1986 年のデータをreference として、暦歴と胃腸障害の関係を、Poisson を用いて算出してください。

1986 年は year5 です。Year1-year4 とコマンドすると STATA は year1, year2, year3, year4 それぞれとコマンドしなかった year5 とを自動的に比較してくれます。

```
. poisson gi year1-year4, irr exposure(pyr)

Iteration 0:   log likelihood = -2545.5093
Iteration 1:   log likelihood = -2545.4201
Iteration 2:   log likelihood = -2545.4201

Poisson regression                                Number of obs   =        799
                                                  LR chi2(4)      =      93.96
                                                  Prob > chi2     =     0.0000
Log likelihood = -2545.4201                       Pseudo R2       =     0.0181

------------------------------------------------------------------------------
      gi |        IRR   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
   year1 |   2.033095   .1468276      9.825   0.000       1.764757    2.342235
   year2 |   1.187425   .0785729      2.596   0.009       1.042994    1.351858
   year3 |   1.193224   .0718253      2.935   0.003       1.060436    1.342639
   year4 |    1.04834   .0617503      0.801   0.423       .9340364    1.176631
     pyr | (exposure)
------------------------------------------------------------------------------
```

1986 年の胃腸障害発生と比較して、過去にさかのぼるにつれ胃腸障害による入院のリスクが高くなっています。どういうことでしょうか？医師が薬剤 A 投与によりひどい胃腸障害を合併しうる点を知り、入院するほどひどくなる前に休薬するようになったのでしょうか？しかし、1982 年は胃腸炎を合併しやすい男性や高齢者に多く処方されていたかもしれません。実際のところはどうなのでしょうか？

理由はともかくとして、治療副作用はこのように市販後年数を経るに従って除々に減少する傾向にあります。

暦年と性別という 2 つの変数を含めて Poisson model で解析してください。

今度は暦年と性別という２つの変数を組み合わせてみます。まずは男性、女性に分けて、前ページで行なった操作を繰り返してみます。

```
. poisson gi year1-year4 if sex==1, exposure(pyr)

Iteration 0:   log likelihood = -1199.3345
Iteration 1:   log likelihood = -1199.2286
Iteration 2:   log likelihood = -1199.2286

Poisson regression                                Number of obs   =        400
                                                  LR chi2(4)      =      33.30
                                                  Prob > chi2     =     0.0000
Log likelihood = -1199.2286                       Pseudo R2       =     0.0137

------------------------------------------------------------------------------
      gi |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
   year1 |   .6184259   .1066758      5.797   0.000       .4093451    .8275066
   year2 |   .1207337   .0946743      1.275   0.202      -.0648244    .3062919
   year3 |   .1000961    .085747      1.167   0.243       -.067965    .2681572
   year4 |    .007148    .082577      0.087   0.931         -.1547    .1689961
   _cons |   -5.68977   .0558146   -101.941   0.000      -5.799164   -5.580375
     pyr | (exposure)
------------------------------------------------------------------------------
```

Incidence of 1986 (year5) = exp (-5.68977) = 0.00338
Incidence of 1982 (year1) = exp (-5.68977 + .6184259) = 0.00627
Incidence of 1983 (year2) = exp (-5.68977 + .1207337) = 0.00381
Incidence of 1984 (year3) = exp (-5.68977 + .007148) = 0.00374
Incidence of 1985 (year4) = exp (-5.68977 + .0472078) = 0.00340

1986 年を reference (=1)としてそれぞれの年の Incidence rate ratio を求めると、

Incidence of 1986 (year5) = 1
Incidence of 1982 (year1) = 0.00598 / 0.00338 = 1.86
Incidence of 1983 (year2) = 0.00349 / 0.00338 = 1.13
Incidence of 1984 (year3) = 0.00351 / 0.00338 = 1.11
Incidence of 1985 (year4) = 0.00308 / 0.00338 = 1.01

女性についても同じように検討してみます。

```
. poisson gi year1-year4 if sex==2, exposure(pyr)

Iteration 0:   log likelihood = -1335.9789
Iteration 1:   log likelihood = -1335.9579
Iteration 2:   log likelihood = -1335.9579

Poisson regression                                Number of obs   =        399
                                                  LR chi2(4)      =      65.03
                                                  Prob > chi2     =     0.0000
Log likelihood = -1335.9579                       Pseudo R2       =     0.0238

------------------------------------------------------------------------------
      gi |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
   year1 |   .8077172   .0987612      8.178   0.000       .6141488    1.001286
   year2 |   .2331836   .0927686      2.514   0.012       .0513606    .4150066
   year3 |   .2576967   .0847588      3.040   0.002       .0915725    .4238209
   year4 |    .091793   .0841088      1.091   0.275      -.0730572    .2566433
   _cons |  -5.961667   .0585206   -101.873   0.000      -6.076365   -5.846969
     pyr | (exposure)
------------------------------------------------------------------------------
```

Incidence of 1986 (year5) = exp (-5.961667) = 0.00257
Incidence of 1982 (year1) = exp (-5.961667 + .8077172) = 0.00578
Incidence of 1983 (year2) = exp (-5.961667 + .2331836) = 0.00325
Incidence of 1984 (year3) = exp (-5.961667 + .2576967) = 0.00333
Incidence of 1985 (year4) = exp (-5.961667 + .091793) = 0.00282

1986 年を reference (=1)としてそれぞれの年の Incidence rate ratio を求めると、

Incidence of 1986 (year5) = 1
Incidence of 1982 (year1) = 0.00578 / 0.00257 = 2.24
Incidence of 1983 (year2) = 0.00325 / 0.00257 = 1.26
Incidence of 1984 (year3) = 0.00333 / 0.00257 = 1.29
Incidence of 1985 (year4) = 0.00282 / 0.00257 = 1.10

前の解法では暦年と性別を別々に計算しました。そんな面倒なことはしないでも、今度は１つの表で一度に計算してみましょう。そのためには性と暦年を組み合わせた新しい変数を創ります。1986 年は reference year としますので year1 – year4 まで、sex1 or 2 なので８つの組み合わせができます。

```
. gen myear1 = sex1 * year1

. gen myear2 = sex1 * year2

. gen myear3 = sex1 * year3

. gen myear4 = sex1 * year4

. gen fyear1 = sex2 * year1

. gen fyear2 = sex2 * year2

. gen fyear3 = sex2 * year3

. gen fyear4 = sex2 * year4
```

最初に言葉を仕込んでおいて、同じように STATA に計算してもらいましょう。

```
. poisson gi sex2 myear1-myear4 fyear1-fyear4, exposure(pyr)

Iteration 0:   log likelihood = -2535.3575
Iteration 1:   log likelihood = -2535.1865
Iteration 2:   log likelihood = -2535.1865

Poisson regression                                Number of obs   =        799
                                                  LR chi2(9)      =     114.43
                                                  Prob > chi2     =     0.0000
Log likelihood = -2535.1865                       Pseudo R2       =     0.0221

------------------------------------------------------------------------------
      gi |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    sex2 |  -.2718972   .0808698     -3.362   0.001      -.4303991   -.1133953
  myear1 |   .6184257   .1066758      5.797   0.000        .409345    .8275065
  myear2 |   .1207337   .0946743      1.275   0.202      -.0648245    .3062919
  myear3 |   .1000961    .085747      1.167   0.243       -.067965    .2681572
  myear4 |    .007148    .082577      0.087   0.931         -.1547     .168996
  fyear1 |   .8077173   .0987612      8.178   0.000       .6141489    1.001286
  fyear2 |   .2331836   .0927686      2.514   0.012       .0513606    .4150066
  fyear3 |   .2576967   .0847588      3.040   0.002       .0915725    .4238209
  fyear4 |    .091793   .0841088      1.091   0.275      -.0730572    .2566433
   _cons |   -5.68977   .0558146   -101.941   0.000      -5.799164   -5.580375
     pyr | (exposure)
------------------------------------------------------------------------------
```

前と同様にまず reference を計算し、それぞれの incidence を計算します。

Incidence of 1986 (year5) = exp (-5.68977) = 0.00338

Incidence of 1982 in male (myear1) = exp (-5.68977 + .6184257) = 0.00627
Incidence of 1983 in male (myear2) = exp (-5.68977 + .1207337) = 0.00381
Incidence of 1984 in male (myear3) = exp (-5.68977 + .1000961) = 0.00374
Incidence of 1985 in male (myear4) = exp (-5.68977 + .007148) = 0.00340
Incidence of 1982 in female (fyear1) = exp (-5.68977 + .8077173-.2718972) = 0.00578
Incidence of 1983 in female (fyear2) = exp (-5.68977 + .2331836-.2718972) = 0.00325
Incidence of 1984 in female (fyear3) = exp (-5.68977 + .2576967-.2718972) = 0.00333
Incidence of 1985 in female (fyear4) = exp (-5.68977 + .091793-.2718972) = 0.00282

Incidence of 1986 (year5) = 0.00338 → 1
Incidence of 1986 (year5) = exp (-5.68977 - .2718972) = 0.00258 → 1

Incidence of 1982 in male (myear1) = 0.00627 / 0.00338 = 1.86
Incidence of 1983 in male (myear2) = 0.00381 / 0.00338 = 1.13
Incidence of 1984 in male (myear3) = 0.00374 / 0.00338 = 1.11
Incidence of 1985 in male (myear4) = 0.00340 / 0.00338 = 1.01
Incidence of 1982 in female (fyear1) = 0.00578 / 0.00258 = 2.24
Incidence of 1983 in female (fyear2) = 0.00325 / 0.00258 = 1.26
Incidence of 1984 in female (fyear3) = 0.00333 / 0.00258 = 1.29
Incidence of 1985 in female (fyear4) = 0.00282 / 0.00258 = 1.10

分けて計算しても一緒に計算しても同じ値を得ることができます。

Reference は必ずしも端のものを選ぶ必要はありません。例えば age16 を reference としてみてもよいのです。そして年齢と胃腸副作用の関係を Poisson model で調べてください。

今までは性別と暦年について調べてきましたが、今度は年齢との相関について調べてみます。今までと違ってカテゴリーが多くなっています。年齢カテゴリー16 を reference として胃腸障害の頻度を比較してみてください。

```
. glm gi age1-age15 age17-age20, family(poisson) lnoffset(pyr)

Iteration 1 : deviance = 2021.6941
Iteration 2 : deviance = 1575.2470
Iteration 3 : deviance = 1503.3002
Iteration 4 : deviance = 1482.3611
Iteration 5 : deviance = 1474.9683
Iteration 6 : deviance = 1472.2541
Iteration 7 : deviance = 1471.2556
Iteration 8 : deviance = 1470.8883
Iteration 9 : deviance = 1470.7531
Iteration 10 : deviance = 1470.7034
Iteration 11 : deviance = 1470.6851
Iteration 12 : deviance = 1470.6784
Iteration 13 : deviance = 1470.6759
Iteration 14 : deviance = 1470.6750
Iteration 15 : deviance = 1470.6747
Iteration 16 : deviance = 1470.6746
Iteration 17 : deviance = 1470.6745
Iteration 18 : deviance = 1470.6745
Iteration 19 : deviance = 1470.6745
Iteration 20 : deviance = 1470.6745
Iteration 21 : deviance = 1470.6745

Residual df  =       779                             No. of obs =       799
Pearson X2   =  1716.472                             Deviance   =  1470.674
Dispersion   =  2.203431                             Dispersion =    1.8879
```

Poisson distribution, log link, offset ln(pyr)

gi	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
age1	-16.68646	2124.719	-0.008	0.994	-4181.06	4147.687
age2	-17.37593	2041.478	-0.009	0.993	-4018.599	3983.848
age3	-20.13446	2262.872	-0.009	0.993	-4455.281	4415.012
age4	-3.242593	.3579592	-9.059	0.000	-3.944181	-2.541006
age5	-2.588632	.2076892	-12.464	0.000	-2.995695	-2.181568
age6	-2.397525	.1828601	-13.111	0.000	-2.755924	-2.039126
age7	-2.25436	.1696344	-13.290	0.000	-2.586838	-1.921883
age8	-2.133222	.1624519	-13.131	0.000	-2.451622	-1.814822
age9	-1.857211	.1483331	-12.521	0.000	-2.147939	-1.566484
age10	-1.607635	.1304748	-12.321	0.000	-1.863361	-1.351909
age11	-1.447309	.1146071	-12.628	0.000	-1.671934	-1.222683
age12	-1.14108	.0983399	-11.603	0.000	-1.333823	-.9483377
age13	-.867791	.0891206	-9.737	0.000	-1.042464	-.6931177
age14	-.6655787	.0856532	-7.771	0.000	-.833456	-.4977015
age15	-.280872	.0797534	-3.522	0.000	-.4371857	-.1245582
age17	.4220233	.079243	5.326	0.000	.2667099	.5773367
age18	.7530481	.0874031	8.616	0.000	.5817412	.924355
age19	1.127377	.1055215	10.684	0.000	.9205582	1.334195
age20	1.306467	.1758197	7.431	0.000	.9618668	1.651067
_cons	-4.845815	.0559893	-86.549	0.000	-4.955552	-4.736078
pyr	(exposure)					

Incidence of age16 = exp (-4.845815) = 0.00786

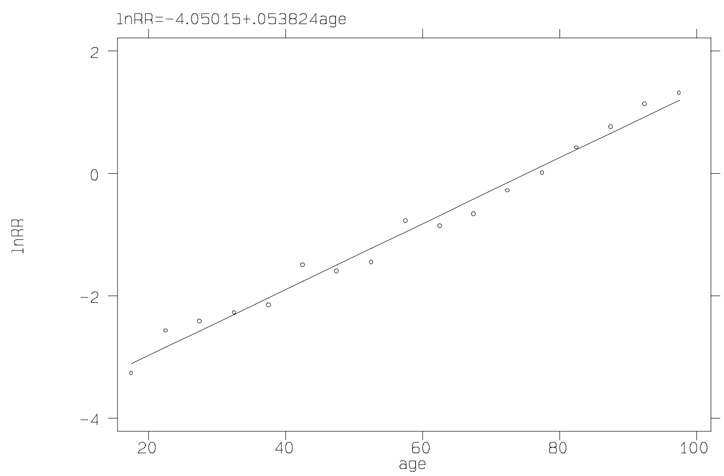
Incidence of age1 = exp (-4.845815 -16.68646) = 4.45 x 10^{-10}
Incidence of age2 = exp (-4.845815 -17.37593) = 3.21 x 10^{-10}
Incidence of age3 = exp (-4.845815 -20.13446) = 0.14 x 10^{-10}
Incidence of age4 = exp (-4.845815 -3.242593) = 0.0003
Incidence of age5 = exp (-4.845815 -2.588632) = 0.0006
Incidence of age6 = exp (-4.845815 -2.397525) = 0.0007
Incidence of age7 = exp (-4.845815 -2.25436) = 0.0008
Incidence of age8 = exp (-4.845815 -2.133222) = 0.0009
Incidence of age9 = exp (-4.845815 -1.857211) = 0.00176
Incidence of age10 = exp (-4.845815 -1.607635) = 0.001.58
Incidence of age11 = exp (-4.845815 -1.447309) = 0.00185
Incidence of age12 = exp (-4.845815 -1.14108) = 0.0036
Incidence of age13 = exp (-4.845815 -.867791) = 0.0033
Incidence of age14 = exp (-4.845815 -.6655787) = 0.00404
Incidence of age15 = exp (-4.845815 -.280872) = 0.00594

Incidence of age17 = exp (-4.845815 + .4220233) = 0.0120
Incidence of age18 = exp (-4.845815 + .7530481) = 0.0167
Incidence of age19 = exp (-4.845815 + 1.127377) = 0.0243
Incidence of age20 = exp (-4.845815 + 1.306467) = 0.0290

Incidence of age16 = exp (-4.845815) = 0.00786 ➔ 1 ln(RR) = 0

IRR of age1 4.45×10^{-10} / 0.00786 = 5.7×10^{-8}
IRR of age2 3.21×10^{-10} / 0.00786 = 4.1×10^{-8}
IRR of age3 0.14×10^{-10} / 0.00786 = 5.7×10^{-8}
IRR of age4 0.0003 / 0.00786 = 0.038 ln(RR) = - 3.27
IRR of age5 0.0006 / 0.00786 = 0.0763 ln(RR) = - 2.57
IRR of age6 0.0007 / 0.00786 = 0.0890 ln(RR) = - 2.42
IRR of age7 0.0008 / 0.00786 = 0.102 ln(RR) = - 2.28
IRR of age8 0.0009 / 0.00786 = 0.115 ln(RR) = - 2.16
IRR of age9 0.00176 / 0.00786 = 0.224 ln(RR) = - 1.50
IRR of age10 0.00158 / 0.00786 = 0.201 ln(RR) = - 1.60
IRR of age11 0.00185 / 0.00786 = 0.235 ln(RR) = - 1.45
IRR of age12 0.0036 / 0.00786 = 0.458 ln(RR) = - 0.78
IRR of age13 0.0033 / 0.00786 = 0.420 ln(RR) = - 0.87
IRR of age14 0.00404 / 0.00786 = 0.514 ln(RR) = - 0.67
IRR of age15 0.00594 / 0.00786 = 0.756 ln(RR) = - 0.28

IRR of age17 0.0120 / 0.00786 = 1.53 ln(RR) = 0.42
IRR of age18 0.0167 / 0.00786 = 2.12 ln(RR) = 0.75
IRR of age19 0.0243 / 0.00786 = 3.09 ln(RR) = 1.13
IRR of age20 0.0290 / 0.00786 = 3.69 ln(RR) = 1.31



年齢が上がると薬により胃腸障害をきたしやすい傾向にあることがよくわかります。特に胃腸障害の incidence の log は年齢とほぼ比例関係にあります。

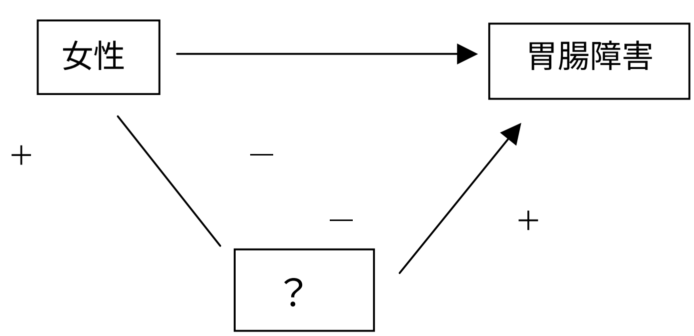
それでは 3 変数（性別、暦年、年齢）を同時に Poisson model で検討してみてください。

それでは、変数を３つ統合してみましょう。各変数で reference になるものは含めずコマンドを出します。

```
. poisson gi sex2 year1-year4 age1-age15 age17-age20, exposure(pyr)

Iteration 0:   log likelihood = -1457.3074
Iteration 1:   log likelihood = -1420.7944
Iteration 2:   log likelihood = -1413.8825
Iteration 3:   log likelihood = -1412.2734
Iteration 4:   log likelihood = -1411.9907
Iteration 5:   log likelihood = -1411.9259
Iteration 6:   log likelihood = -1411.9101
Iteration 7:   log likelihood = -1411.9069
Iteration 8:   log likelihood = -1411.9062
Iteration 9:   log likelihood =  -1411.906
Iteration 10:  log likelihood =  -1411.906

Poisson regression                                Number of obs   =        799
                                                  LR chi2(24)     =    2360.99
                                                  Prob > chi2     =     0.0000
Log likelihood =  -1411.906                       Pseudo R2       =     0.4554

------------------------------------------------------------------------------
          gi |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        sex2 |   -.371472    .041925    -8.860  0.000    -.4536435   -.2893005
       year1 |   .5363477   .0723159     7.417  0.000     .3946111    .6780844
       year2 |   .0873811   .0662095     1.320  0.187    -.0423872    .2171494
       year3 |   .1302761   .0602094     2.164  0.030     .0122678    .2482845
       year4 |   .0276945    .058906     0.470  0.638    -.0877592    .1431481
        age1 |   -13.1255   574.0952    -0.023  0.982    -1138.331     1112.08
        age2 |  -13.79954   551.9906    -0.025  0.980    -1095.681    1068.082
        age3 |  -16.55463   620.9668    -0.027  0.979    -1233.627    1200.518
        age4 |  -3.247426   .3579563    -9.072  0.000    -3.949007   -2.545844
        age5 |  -2.608346   .2078012   -12.552  0.000    -3.015629   -2.201063
        age6 |    -2.4309     .18303   -13.281  0.000    -2.789632   -2.072168
        age7 |  -2.282254   .1697966   -13.441  0.000    -2.615049   -1.949459
        age8 |  -2.149279   .1625666   -13.221  0.000    -2.467904   -1.830654
        age9 |  -1.873387   .1484116   -12.623  0.000    -2.164268   -1.582506
       age10 |  -1.620927   .1305177   -12.419  0.000    -1.876737   -1.365117
       age11 |  -1.466343   .1146423   -12.791  0.000    -1.691037   -1.241648
       age12 |  -1.159022   .0983726   -11.782  0.000    -1.351828   -.9662148
       age13 |   -.884214   .0891493    -9.918  0.000    -1.058943   -.7094845
       age14 |  -.6743553   .0856625    -7.872  0.000    -.8422508   -.5064598
       age15 |  -.2841135    .079757    -3.562  0.000    -.4404343   -.1277927
       age17 |   .4270917   .0792459     5.389  0.000     .2717727    .5824108
       age18 |   .7677081   .0874262     8.781  0.000     .5963559    .9390603
       age19 |   1.155303   .1055688    10.944  0.000     .9483922    1.362214
       age20 |   1.335737   .1758427     7.596  0.000     .9910919    1.680383
       _cons |  -4.737086   .0694933   -68.166  0.000    -4.873291   -4.600882
         pyr | (exposure)
------------------------------------------------------------------------------
```

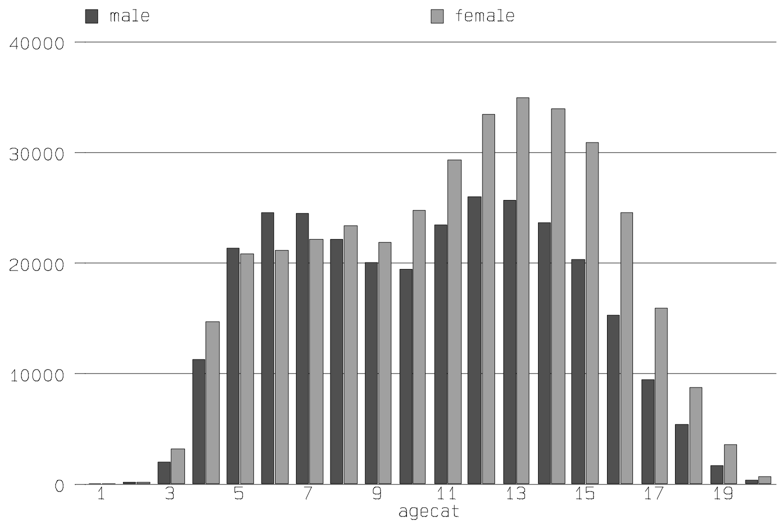
gi	adjusted	crude
sex2	-.371472	-0.168
year1	.5363477	0.710
year2	.0873811	0.172
year3	.1302761	0.177
year4	.0276945	0.047
age1	-13.1255	-16.7
age2	-13.79954	-17.0
age3	-16.55463	-16.7
age4	-3.247426	-3.27
age5	-2.608346	-2.57
age6	-2.4309	-2.42
age7	-2.282254	-2.28
age8	-2.149279	-2.16
age9	-1.873387	-1.50
age10	-1.620927	-1.60
age11	-1.466343	-1.45
age12	-1.159022	-0.78
age13	-.884214	-0.87
age14	-.6743553	-0.67
age15	-.2841135	-0.28
age17	.4270917	0
age18	.7677081	0.42
age19	1.155303	0.75
age20	1.335737	1.13
_cons	-4.737086	1.31
pyr	(exposure)	

上の表で adjusted と crude data の間に大きな開きのある変数が存在します。例えば女性は crude data において exp(- 0.168)、すなわち女性胃腸障害合併率は男性の 0.845 倍と少なくなっていますが、adjustment 後 exp(-0.371), すなわち 0.69 倍と更に男女差が顕著となっています。これは confounder の存在を示唆しています。



adjustment により女性の胃腸障害を起こす Incidence rate は小さくなっていますから、confounder はプラスの方向に作用しなくてはなりません。すなわち confounder と女性の関係がプラスであれば（正比例）、confounder と胃腸障害の関係もプラスとなるはずですし、逆の場合もあり得ます。ここで考慮した因子は年齢と暦年ですから、この２つを合わせたものが女性と胃腸障害と関係を持てばよいわけです。年齢が高くなればなる程、

胃腸障害を起こしやすいことは確認しました。また暦年が早い程胃腸障害が多い傾向にあります。実際薬を処方された女性は男性と比較して高い年齢に多いようでした（下図）。男女比は薬を処方された暦年で一定でした。



更に薬の使用状況(status)を入れ検討してください。

Status2-4 を追加します。

```
. poisson gi sex2 year1-year4 age1-age15 age17-age20 status2-status4, irr

Iteration 0:   log likelihood = -1253.4946
Iteration 1:   log likelihood = -1225.9165
Iteration 2:   log likelihood =  -1219.856
Iteration 3:   log likelihood =  -1218.693
Iteration 4:   log likelihood = -1218.4458
Iteration 5:   log likelihood = -1218.3854
Iteration 6:   log likelihood = -1218.3727
Iteration 7:   log likelihood =   -1218.37
Iteration 8:   log likelihood = -1218.3694
Iteration 9:   log likelihood = -1218.3693
Iteration 10:  log likelihood = -1218.3692

Poisson regression                                Number of obs   =        799
                                                  LR chi2(27)     =    3146.26
                                                  Prob > chi2     =     0.0000
Log likelihood = -1218.3692                       Pseudo R2       =     0.5635

------------------------------------------------------------------------------
      gi |        IRR   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    sex2 |   1.057194   .0440852      1.334   0.182       .9742252    1.147229
   year1 |   .4551393   .0328691    -10.900   0.000       .3950686    .5243437
   year2 |    .593801   .0392917     -7.877   0.000       .5215754    .6760281
   year3 |   .8189235   .0492938     -3.319   0.001       .7277909    .9214675
   year4 |   .8874389    .052272     -2.027   0.043       .7906804    .9960382
    age1 |   6.97e-09   4.71e-06     -0.028   0.978              0           .
    age2 |   7.01e-09   4.69e-06     -0.028   0.978              0           .
    age3 |   7.01e-09   4.69e-06     -0.028   0.978              0           .
    age4 |   .0250796   .0089772    -10.297   0.000       .0124347    .0505831
    age5 |   .0783699   .0162764    -12.260   0.000       .0521633    .1177423
    age6 |   .1034465   .0189162    -12.407   0.000       .0722879    .1480357
    age7 |   .1222553   .0207386    -12.389   0.000       .0876749    .1704748
    age8 |   .1347951   .0218976    -12.336   0.000       .0980383    .1853329
    age9 |   .1661458   .0246446    -12.101   0.000       .1242309    .2222025
   age10 |   .2257143   .0294493    -11.409   0.000       .1747839    .2914852
   age11 |    .313498   .0359279    -10.122   0.000       .2504287    .3924509
   age12 |   .4796483   .0471673     -7.471   0.000       .3955652    .5816045
   age13 |   .6520466     .05811     -4.799   0.000       .5475454    .7764923
   age14 |   .7460767   .0639033     -3.420   0.001       .6307771    .8824519
   age15 |    .971781   .0775021     -0.359   0.720       .8311565    1.136198
   age17 |   .9968645   .0789937     -0.040   0.968       .8534637     1.16436
   age18 |   .6959273   .0608255     -4.148   0.000       .5863637    .8259631
   age19 |    .391873   .0413497     -8.878   0.000       .3186604    .4819063
   age20 |   .1128508   .0198413    -12.409   0.000       .0799553    .1592804
 status2 |   1.535759   .0753985      8.739   0.000       1.394868    1.690881
 status3 |   .4598514   .0313048    -11.412   0.000       .4024123    .5254891
 status4 |   .3649625   .0269669    -13.641   0.000       .3157572    .4218356
------------------------------------------------------------------------------
. poisson, irr

Poisson regression                                Number of obs   =        799
                                                  LR chi2(27)     =    3077.31
                                                  Prob > chi2     =     0.0000
Log likelihood = -1053.7449                       Pseudo R2       =     0.5935

------------------------------------------------------------------------------
      gi |        IRR   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    sex2 |    .651293   .0273479    -10.212   0.000       .5998384    .7071614
   year1 |    1.11034   .0821419      1.415   0.157       .9604721    1.283592
   year2 |   .9268268   .0617373     -1.141   0.254       .8133899    1.056084
   year3 |   1.048734   .0632694      0.789   0.430       .9317793    1.180369
```

```
      year4 |   .9958866   .0586827     -0.070   0.944      .8872638    1.117808
       age1 |   1.48e-06   .0011922     -0.017   0.987             0           .
       age2 |   7.38e-07   .0005717     -0.018   0.985             0           .
       age3 |   4.95e-08    .000041     -0.020   0.984             0           .
       age4 |    .054537   .0195386     -8.119   0.000      .0270232    .1100642
       age5 |   .1038367   .0216358    -10.870   0.000      .0690225    .1562106
       age6 |   .1213201   .0222718    -11.490   0.000      .0846584    .1738584
       age7 |    .137944   .0234923    -11.632   0.000      .0987962    .1926039
       age8 |   .1538072   .0250698    -11.485   0.000      .1117469    .2116986
       age9 |   .1962314   .0291907    -10.947   0.000      .1466043    .2626579
      age10 |    .244538   .0319849    -10.768   0.000      .1892394    .3159957
      age11 |   .2736232   .0314216    -11.286   0.000      .2184765    .3426896
      age12 |   .3585125   .0353165    -10.413   0.000      .2955655    .4348653
      age13 |   .4535351   .0404647     -8.862   0.000       .380773    .5402015
      age14 |    .543529    .046578     -7.114   0.000      .4594926    .6429347
      age15 |   .7772319   .0619948     -3.160   0.002       .664746    .9087523
      age17 |   1.474781    .116889      4.902   0.000      1.262591    1.722632
      age18 |   2.053975   .1796279      8.230   0.000      1.730431    2.438012
      age19 |   2.931316   .3096529     10.181   0.000      2.383115    3.605623
      age20 |   3.559705   .6261138      7.219   0.000       2.52172    5.024943
    status2 |   3.762875    .193625     25.753   0.000      3.401887     4.16217
    status3 |   2.241911   .1553654     11.650   0.000      1.957176    2.568071
    status4 |   1.309797   .0976457      3.620   0.000       1.13174    1.515868
        pyr | (exposure)
------------------------------------------------------------------------------
```

.

	Crude	Adjusted
0- 30	5.82(0.162)	3.76 (0.194)
31 - 60	3.00 (0.186)	2.24 (0.155)
61 - 150	1.46(0.286)	1.31 (0.098)

上の表をみると crude data と比較して adjusted data は小さくなる傾向にあります。性別、年齢、暦年の不等分布が confounder として働き、status crude data を不必要に押し上げていたものと考えられます。Adjusted data は crude data より confounders を相殺したものですから、これを最終値として結論に用います。

年齢を連続性の数値としてモデルを作り直してください。

年齢カテゴリー番号に 5 を掛けて 2.5 を引いてやると、そのカテゴリーの中間値になります。個々の年齢がここでは判らないので、この方法で代用します。まず、上の計算に従い新しい変数 agemid を作ります。

```
. gen agemid=age*5-2.5

. summarize agemid

Variable      Obs            Mean          Std. Dev.          Min            Max
  agemid       799        50.05945         28.81844          2.5             97.5
```

平均年齢は 50 歳です。

```
. poisson gi agemid, irr exposure(pyr)

       gi              IRR        Std. Err.              T             p>|t|
[95% conf. Interval]
 agemid     1.058991    .0014554          41.706       0.000             1.056141
1.061849

. poisson gi sex2 year1-year4 agemid status2-status4, irr exposure(pyr)

Iteration 0:   log likelihood = -1161.3504
Iteration 1:   log likelihood = -1074.7727
Iteration 2:   log likelihood = -1074.3459
Iteration 3:   log likelihood = -1074.3458

Poisson regression                                Number of obs   =        799
                                                  LR chi2(9)      =    3036.11
                                                  Prob > chi2     =     0.0000
Log likelihood = -1074.3458                       Pseudo R2       =     0.5856

------------------------------------------------------------------------------
      gi |        IRR   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    sex2 |   .6529403    .027388    -10.162   0.000        .601408    .7088881
   year1 |   1.102986   .0815724      1.325   0.185       .9541537    1.275033
   year2 |   .9216466   .0613717     -1.225   0.220       .8088791    1.050135
   year3 |   1.045637    .063071      0.740   0.459        .929048    1.176858
   year4 |   .9946654   .0586087     -0.091   0.928       .8861794    1.116432
  agemid |   1.053119   .0014884     36.620   0.000       1.050206    1.056041
 status2 |   3.736329   .1915978     25.704   0.000       3.379059    4.131373
 status3 |   2.209196   .1528354     11.457   0.000       1.929065    2.530007
 status4 |   1.295804   .0965537      3.478   0.001       1.119733    1.499563
     pyr | (exposure)
------------------------------------------------------------------------------
```

Change in model fit: 664.041 (linear) – 622.839 (catogorical) = 41.2 ≈ χ_{18}^2 $p<0.005$
18 はパラメーターの違いであり、19 の age category から agemid,の 1 つを差し引いたものです。年齢を連続変数としてとらえてた方がカテゴリーとしてとらえるよりも、モデルとして有意に改善しています。つまり、年齢をカテゴリーとしてではなくて連続変数としてとらえた方が、より正確な判断ができることになります。

このように最も適切なモデルをまず設定します。そうすると、最初は暦年が早い程、胃腸障害が多かったのですが、性別、薬剤使用歴、年齢により相殺されてしまいました。有意差が無くなってしまった点に注目してください。

年齢、性別、処方暦年で調整したところ、現在薬剤 A を服用している人が胃腸障害のため入院するリスクは、そうでない人と比較して 3.7 倍（95%CI 3.4-4.1）にまで上がっていました。

このようにたった１つの結論を導くのに、裏ではいろいろな操作を行っているのです。一方、論文の結果のみからはどのように数値を導き出したかの詳細を知ることはできないのが普通です。

それでは(5 年の中点)^2 を創ってみたらどうなるでしょうか？

```
. gen agemid2 = agemid * agemid

. poisson gi sex2 year1-year4 agemid agemid2 status2-status4, irr exposure(pyr)

Iteration 0:   log likelihood = -1150.1314
Iteration 1:   log likelihood = -1061.3713
Iteration 2:   log likelihood = -1059.6352
Iteration 3:   log likelihood = -1059.6338
Iteration 4:   log likelihood = -1059.6338

Poisson regression                                Number of obs   =        799
                                                  LR chi2(10)     =    3065.53
                                                  Prob > chi2     =     0.0000
Log likelihood = -1059.6338                       Pseudo R2       =     0.5913

------------------------------------------------------------------------------
      gi |        IRR   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    sex2 |   .6508506   .0273164    -10.233   0.000       .5994544    .7066535
   year1 |   1.109282   .0820509      1.402   0.161       .9595789    1.282341
   year2 |   .9259846   .0616701     -1.155   0.248       .8126699    1.055099
   year3 |   1.048375   .0632403      0.783   0.434       .9314729    1.179948
   year4 |   .9958086    .058677     -0.071   0.943       .8871963    1.117718
  agemid |   1.011036   .0073817      1.503   0.133       .9966712    1.025608
 agemid2 |   1.000323   .0000574      5.627   0.000       1.000211    1.000436
 status2 |   3.756466   .1931628     25.738   0.000       3.396327    4.154794
 status3 |   2.236388   .1549105     11.620   0.000       1.952478    2.561583
 status4 |    1.30712   .0974303      3.593   0.000       1.129454    1.512734
     pyr | (exposure)
------------------------------------------------------------------------------
```

先にも述べましたが、β / se(β) = z です。

Change in model fit: 634.616 (Quadratic) – 622.839 (categorical) = 41.2 ≈ χ_{17}^2 p>0.80
17 はパラメーターの違いであり、19 の age category から agemid, agemid2 の 2 つを差し引いたものです。Quadratic と Categorical の間には統計学的に有意差がありません。つまり、Quadratic にしてもあまり意味がないことになります。

ここで年齢を 5 年毎に区切らず、1 年毎に区切ることができるでしょうか？その場合、年齢は 20 のカテゴリーから、（99 歳までとすると）100 のカテゴリーにまで増えます。更に性、暦年、薬の使用歴の条件と組み合わせると、800 通りの組み合わせから 4000 通りの組み合わせに増えてしまいます。

まとめ

薬剤 A の副作用を調査しました。女性は男性と比較して入院するほどの胃腸障害を起こすリスクは 3 割低く、高齢になるほどリスクが上がる傾向にありました（1 年歳をとる毎に 5%リスクが上がる）。そして、性別、年齢、暦年で adjust したところ、胃腸障害入院のリスクは使用後 151 日以上経っている患者さんと比較して、使用後 30 日以内で 3.7 倍、31－60 日で 2.2 倍、61－150 日で 1.3 倍に上昇していました。薬剤 A を最近使用したものでは胃腸障害に注意する必要がありますが、過去の使用であっても胃腸障害を合併するリスクがある点を念頭に診療する必要があります。