For Stata version 8.0 or higher

The commands as ado files for making prediction are:

for the model case(i,t) = popu(i,t)*(alpha(i) + beta(i)*t), [$c_{it} = n_{it}(\alpha_i + \beta_i t)$] predaaap.ado
for the model case(i,t) = popu(i,t)*exp(alpha(i) + beta(i)*t) [$c_{it} = n_{it}\exp(\alpha_i + \beta_i t)$] predmaap.ado
for the model case(i,t) = popu(i,t)*exp(alpha(i) + beta*t) [$c_{it} = n_{it}\exp(\alpha_i + \beta t)$] predmap.ado
for the model case(i,t) = popu(i,t)*alpha(i)*(1 + beta*t) [$c_{it} = n_{it}\alpha_i(1+\beta t)$] prednap.ado and _prednap.ado

all files come with exhaustive Stata help files, which are the main source of information on how to use properly the commands. You can view the help window in Stata by typing "help the_name_of_file", for example `help predaaap`. The `if` expression and `in` range qualifiers can be used with the commands.
In order to use the commands load the above 5 files(don't forget about _prednap.ado) and Stata help files(`predaaap.hlp`, `predmaap.hlp`, `predmap.hlp`, `prednap.hlp`) somewhere where STATA can see them (e.g. c:\ado\personal). You can use the STATA command `sysdir` to get the directories where the files can be loaded. Each command can be used to fit the model, make predictions based on the fitted model, or make graphs based on the fitted model.

- to fit the model only, for example:
```
predmap stom popu, age(age_grou) peri(year)
```
where the first variable, stom, contains the number of incident/death cases and the second, popu, person-years for each cell of data. The option `age()` contains the name of age variable and the option `peri()` contains the name of time variable which may be period or cohort.

- to make prediction based on the fitted model for a given time period with accompanying 95% prediction intervals for number of cases and age-adjusted rate and adjusting the results for a possible overdispersion:
```
predmap stom popu, age(age_grou) peri(year) ppn(MPPN) pps(MPPS) tp(1999)
```
In that case before using the command one must prepare a column matrix with future age-specific population, for example
```
matrix MPPN = (12349 \ 53442\ 463635\...\67903)
```
and a column matrix with standard population, for example
```
matrix MPPS = (3000 \ 4000 \ 2000...\1000)
```
The options `ppn()` and `pps()`, respectively, should contain the names of these matrices (note that the number of rows in MPPN and MPPS has to match the number of age groups in the data, otherwise you get an error). The option `tp()` should contain values of the future period (note that the value of tp has to match the code of period existing in your data and stored in the period variable).
The command takes into account an overdispersion if it appears for the data and adjusts the covariance matrix of the fitted model for it by applying the dispersion factor (McCullagh P, Nelder JA. Generalized Linear Models. London: Chapman and Hall, 1989, page 174) and the value of the factor is displayed by the command. The command displays also the p-value of the Pearson goodness of fit statistic of the fitted model. The results of prediction are displayed and stored in the form of matrix. For example after executing the command:
```
predmap stom popu, age(age_code) peri(peri_code) ppn(MPPN) pps(MPPS) tp(9)
```
, from the example file, the part of the ouput is
```
OUTMAP[11,6]
                inci/10^5        no-cases  se(expe-value)  se(pred-value)  Low95%PredInte              Up
        age1         0.48            1.62            0.42            1.34            0.00            4.25
        age2         1.25            4.11            0.75            2.16            0.00            8.35
        age3         1.80            5.36            0.89            2.48            0.50           10.22
        age4         2.68            7.68            1.14            3.00            1.80           13.55
        age5         4.09           12.89            1.73            3.99            5.08           20.70
        age6         6.35           18.59            2.31            4.89            9.01           28.18
        age7        10.70           21.99            2.55            5.34           11.52           32.45
        age8        17.08           29.05            3.25            6.29           16.72           41.39
        age9        27.07           44.86            4.88            8.28           28.62           61.10
ages-combined        6.10          146.16           15.30           19.50          107.94          184.38
adj-rate/10^5        5.51               .            0.58            0.74            4.05            6.96
```
Based on the above matrix OUTMAP for the age group 6: the predicted crude incidence rate is 6.35 per 100 000 person-years, 146 is the predicted number of cases(after rounding as 146.16 refers to the expected value of predicted number of cases which is a real number), 15.30 is the standard error of the expected value of the predicted number of cases, all ages combined, and expresses the variability due to using the model to estimate

this value, 17.24 is the standard error of the predicted number of cases and expresses the variability due to using the model to estimate this value plus the variabilty due to the fact that the future observation has its own variability, 107.94 is the lower bound of the 95% prediction interval and 184.38 is the upper bound of the 95% prediction interval Thus the age-specific prediction for age group 6 can be presented, after rounding, in the form of 19 (9 ; 28) that is the predicted number of cases is 19 with accompanning 95% prediction interval: (9 ; 28). Similarly for all ages combined the result of prediction is: 146 (108 ; 184) and for age-adjusted rate is: 5.51 (4.05 ; 6.96) per 100 000 person years. The matrix OUTMAP can next be used by a user. In order to save all results of prediction in a separate Stata file the option `file(`file name`)` can be applied:

```
predmap stom popu, age(age_code) peri(peri_code) ppn(MPPN) pps(MPPS) tp(9) tp(9),file(male2006p)
```

the new created file `male2006p.dta` will replace without warning the file with the same name if it exists.

- to get additional graphs of historical trend and predicted numbers. In that case extra option `graph` , without paranthesis, has to be added, indicating that plots for cases and rates are to be produced and saved and the option `title(`user_text`)`, if applied, adds text to the graphs:

```
predmap stom popu, age(age_code) peri(peri_code) ppn(MPPN) pps(MPPS) tp(9), graph
```

Automatically three graphs are produced:

`*_case.gph` ← graph of historical and predicted cases, ages combined, with prediction interval
`*_adju.gph` ← as above but for the result expressed in the form of age-adjusted rate
`*_age.gph` ← historical and predicted crude incidence rates, per 10^5 person-years, with prediction intervals for each age group separately

where * states for the name of the command which produces a given graph. For example: `predaaap_age.gph, prednap_age.gph`, and so on. Each graph can be opened using the command "`graph use`". There is also possible to get a separate graph representing historical and predicted crude incidence rates, per 10^5 person-years, with 95% prediction intervals for one age group only by using the option `agegr()`:

```
predmap stom popu, age(age_code) peri(peri_code) ppn(MPPN) pps(MPPS) tp(9), graph agegr(3)
```

Age code 3 in the above example refers to the third consecutive age group regardless of its actual code in the data.

Remarks:

- use the `if` expression and/or `in` range qualifiers to quickly subset the data for which the models are to be fitted. In that way the definition of the base of prediction and/or elimination of age groups without cases can easilly be performed. For example:

```
predmap stom popu if year>=1981 & year<=1990 & age_grou>=4 & site=="stom" & sex=="male"
, age(age_grou) peri(year) ppn(MPPN) pps(MPPS) tp(1999)
```

- remember that change of the base of prediction can significantly change the fit of the models
- too many empty cells in data can result in non-convergence. Solution: elimination of age groups with empty cells or combining some age groups in order to avoid too many empty cells. After combining age groups necessary use the command `collapse` to get one unique age group per one record for a given period.
- any errors during fitting or executing the commands do not make any harm to the data

The best way to get familiar with the commands is to run the do file: example_pred.do which includes the data and practical examples.

Information about statistical theory concerning the commands:

- Hakulinen T, Dyba T. Precision of incidence predictions based on Poisson distributed observations. Statistics in Medicine 13: 1513--23, 1994.
- Dyba T, Hakulinen T, Päivärinta L. A simple non-linear model in incidence prediction. Statistics in Medicine 16: 2297--309, 1997.
- Dyba T, Hakulinen T. Comparison of different approaches to incidence prediction based on simple interpolation techniques. Statistics in Medicine 19: 1741-52, 2000.
- Correction, Statistics in Medicine 19: 1251, 2000.

tadek.dyba@ec.europa.eu