Analysing Duration Data

Professor Vernon Gayle
AQMEN
University of Edinburgh

March 2019

Alternative terminology

- Duration models
- Survival models
- Cox regression
- Cox models
- Failure time analysis
- Hazard models
- Event history analysis

Models for duration data allow the data analyst to assess the relative influence of a number of explanatory factors upon how long it takes for an event to occur

Original paper Cox (1972)

Applications

- Study the lifetimes of machine components in engineering
- Duration of unemployment in economics
- Time taken to complete cognitive tasks in psychology
- Lengths of tracks on a photographic plate in particle physics
- Survival times of patients in clinical trials

Research Examples

Heckman and Borjas (1980) used duration modelling approaches to study unemployment

Blossfeld and Hakim (1997) studied female part-time employment

Mulder and Smits (1999) investigated first time home ownership

Lillard et al. (1995) studied premarital cohabitation and subsequent marital dissolution

Research Examples

Kiernan and Mueller (1998) undertook an analysis of divorce using the BHPS and the NCDS

Boyle et al. (2008) examined union dissolution using the Austrian Family and Fertility Survey (FFS)

Chan and Halpin (2002) used BHPS to examine gender role attitudes and the domestic division of labour on divorce

Pevalin and Ermisch (2004) investigated mental health, union dissolution and re-partnering

Measuring a Duration

Three requirements for correctly determining a duration

1. A starting time must be unambiguously defined
2. Time must have a defined unit of measurement
3. The event must be clearly defined

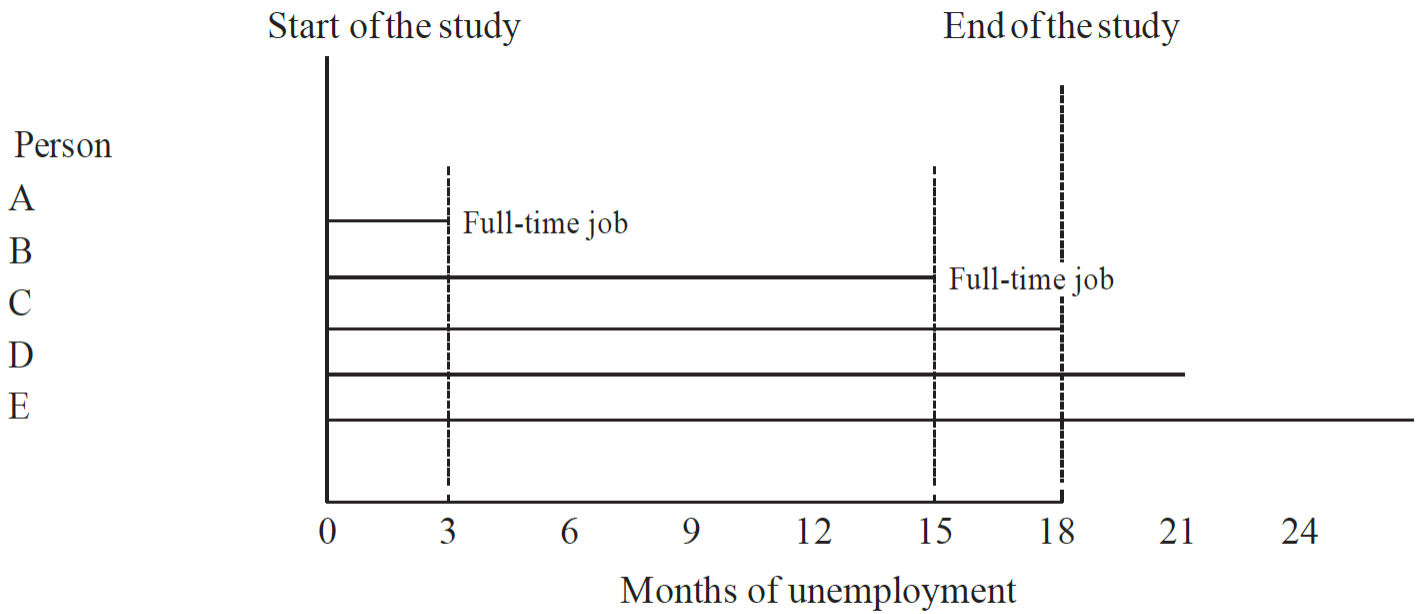


Figure 4 A diagram of a hypothetical study of unemployment

The Accelerated Life Model

Regression models can be estimated with duration data

Historically the log of the duration has been modelled

Censored Observations

- Censored observations affect regression model results
- The impact on the results may sometimes be negligible
- Plewis (1997) states that when there is a very small proportion of censored cases they will have little effect, and an accelerated life model might still be suitable
- Supervisors, examiners and referees may not be convinced

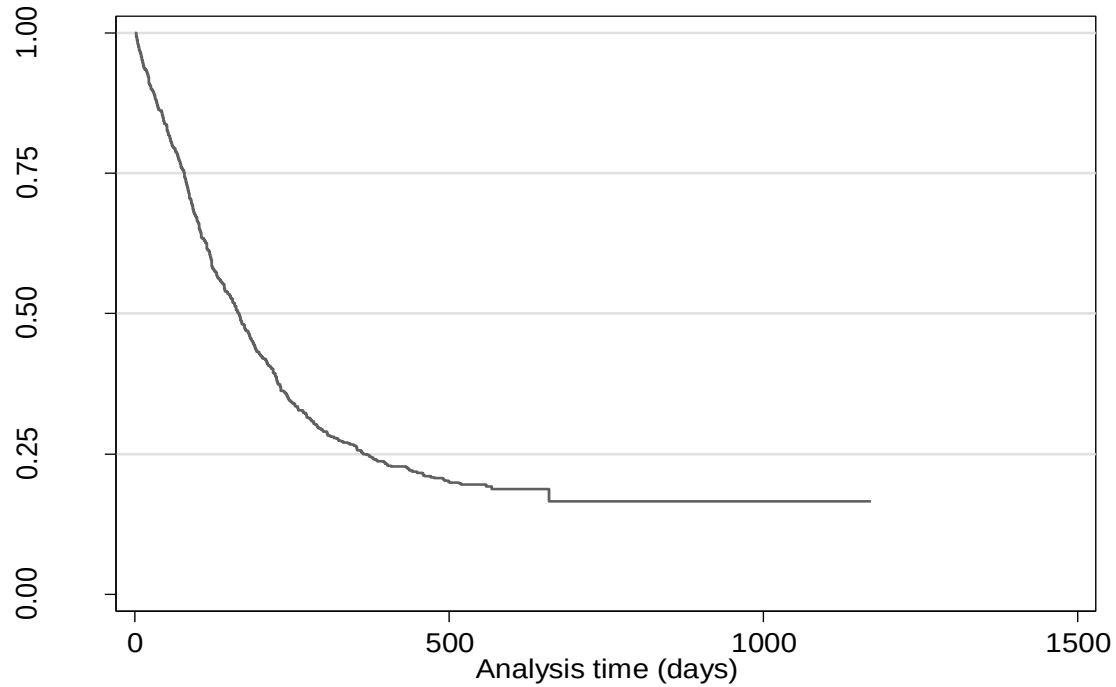
Duration Modelling

- No longer directly modelling the duration
- The focus is on modelling the probability that an event occurs at time t , conditional on it not having occurred before t

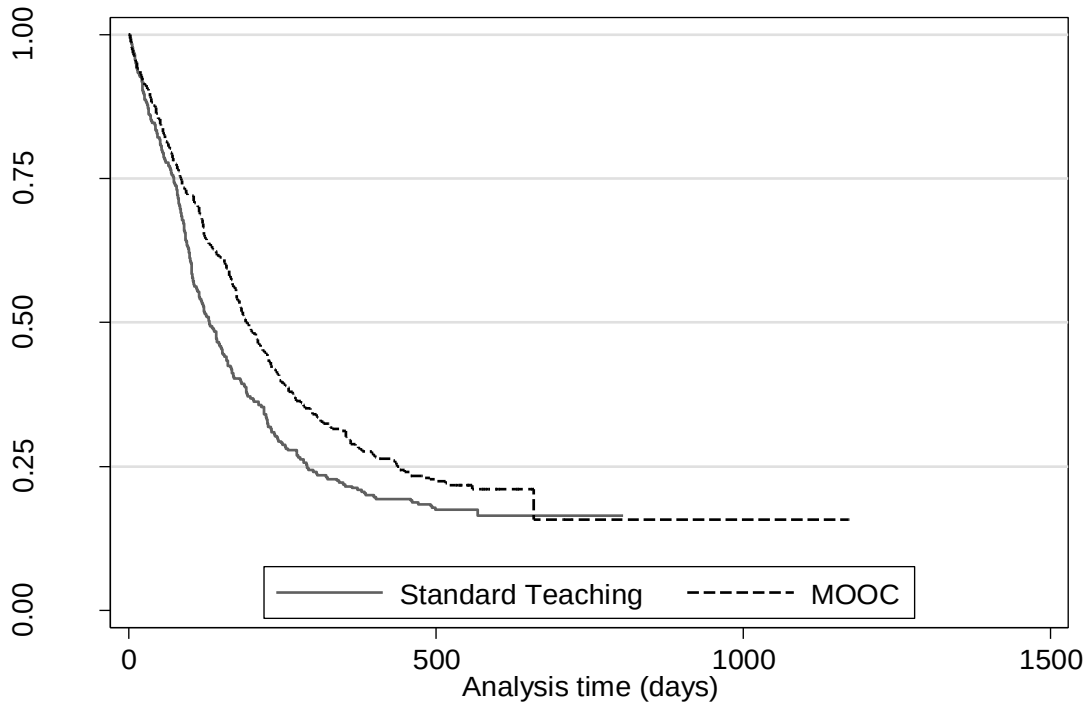
Codebook for the College Skills Program Dataset

Variable	Obs	Unique	Mean	Min	Max	Label
id	628	628	314.5	1	628	student id
time	628	338	234.7038	2	1172	number of days until test passed
test	628	2	.8089172	0	1	test passed (or censored)
age	623	31	32.36918	20	56	age at enrolment
no_jobs	611	28	4.574468	0	40	number of previous jobs
mooc	628	2	.4904459	0	1	taught by massive open online course
campus	628	2	.2929936	0	1	college campus
quals1	628	2	.4601911	0	1	no qualifications
quals2	628	2	.1815287	0	1	lower qualifications (below A'level)
quals3	628	2	.3582803	0	1	higher qualifications(above A'level)

Output: Kaplan-Meier Plot of Time to Passing the Test (College Skills Program Data)



Output: Kaplan-Meier Plot of Time to Passing the Test (College Skills Program Data)



Output: Log-Rank Test for Equality of Survivor Functions

```
 failure _d:  test
   analysis time _t:  time


Log-rank test for equality of survivor
functions

      |   Events          Events
mooc  |  observed        expected
------+-------------------------
0     |       265          235.80
1     |       243          272.20
------+-------------------------
Total |       508          508.00

           chi2(1) =        6.80
           Pr>chi2 =      0.0091
```

Output: Cox Regression Model Time to Passing the Test (College Skills Program Data)

```
 failure _d:  test

   analysis time _t:  time


Iteration 0:   log likelihood =  -2868.555
Iteration 1:   log likelihood = -2851.6989
Iteration 2:   log likelihood = -2851.0884
Iteration 3:   log likelihood = -2851.0863
Refining estimates:
Iteration 0:   log likelihood = -2851.0863

Cox regression -- Breslow method for ties


No. of subjects =          610                 Number of obs    =         610

No. of failures =          495

Time at risk    =       142994

                                               LR chi2(6)       =       34.94

Log likelihood  =   -2851.0863                 Prob > chi2      =      0.0000


------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -.0237543   .0075611    -3.14   0.002    -.0385737   -.0089349
     no_jobs |    .034745   .0077538     4.48   0.000     .0195478    .0499422
        mooc |  -.2540169    .091005    -2.79   0.005    -.4323834   -.0756504
      campus |  -.1723881   .1020981    -1.69   0.091    -.3724966    .0277205
      quals2 |   .2467753   .1227597     2.01   0.044     .0061706    .4873799
      quals3 |    .125668   .1030729     1.22   0.223    -.0763513    .3276873
------------------------------------------------------------------------------
```

Output: Test of the Effects of Previous Education in Cox Regression Model of Time to Passing the Test (College Skills Program Data)

```
( 1)  quals2 = 0
( 2)  quals3 = 0

           chi2(  2) =    4.36
         Prob > chi2 =    0.1130
```

Output: Hazard Ratios Cox Regression Model Time to Passing the Test (College Skills Program Data)

```
Cox regression -- Breslow method for ties

No. of subjects =          610                 Number of obs    =         610
No. of failures =          495
Time at risk    =       142994
                                               LR chi2(3)       =       27.76
Log likelihood  =   -2854.6735                 Prob > chi2      =      0.0000

------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .9794475   .0072674    -2.80   0.005     .9653067    .9937955
     no_jobs |   1.036128   .0078949     4.66   0.000     1.020769    1.051718
        mooc |   .7940896   .0716076    -2.56   0.011     .6654445    .9476047
------------------------------------------------------------------------------
```

Output: Time to Passing the Test - Survival Functions Comparing Women Aged 30 with 5 Previous Jobs by Teaching Methods (College Skills Program Data)