



# XV CONVEGNO ITALIANO DEGLI UTENTI DI STATA

Bologna, 15-16 Novembre 2018



9.00 - 9.30 Registrazione dei partecipanti

## 9.30 - 10.30 I SESSIONE - INVITED SPEAKER

**Joint Modelling of Longitudinal and Survival Data** • Michael J. Crowther, Department of Health Sciences University of Leicester

Multi-state models are increasingly being used to model complex disease profiles. By modelling transitions between disease states, accounting for competing events at each transition, we can gain a much richer understanding of patient trajectories and how risk factors impact over the entire disease pathway. In this talk, I'll introduce some new Stata commands for the analysis of multi-state survival data. This includes **-msset-**, a data preparation tool which converts a dataset from wide (one observation per subject, multiple time and status variables) to long (one observation for each transition for which a subject is at risk for). **-msaj-** calculates the non-parametric Aalen-Johansen estimates of transition probabilities. **-msboxes-** creates a descriptive plot of the multi-state process through the transition matrix and numbers at risk. **-stms-** fits joint transition-specific survival models, allowing each transition to have a different parametric model, yet maximised jointly to enable sharing of covariate effects across transitions. **-predictms-** calculates a variety of predictions from a multi-state survival model, including transition probabilities, length of stay (restricted mean time in each state), the probability of ever visiting each state and more. Predictions are made at user-specified covariate patterns. Differences and ratios of predictions across covariate patterns can also be calculated. Standardised (population-averaged) predictions can be obtained. Confidence intervals for all quantities are available. Simulation or the Aalen-Johansen estimator are used to calculate all quantities. User-defined predictions can also be calculated by providing a user-written Mata function, to provide complete flexibility. **-predictms-** can be used with a general transition matrix (cyclic or acyclic) and allows the use of transition-specific timescales. I will illustrate the software using a dataset of patients with primary breast cancer.

10.30 - 10.50 Pausa caffè

## 10.50 - 12.15 II SESSIONE - COMMUNITY CONTRIBUTED, I

**The Stata module CUB for fitting mixture models for ordinal data** • Christopher F. Baum, Boston College; Giovanni Cerulli IRCrES-CNR; Francesca di Iorio, Domenico Piccolo e Rosaria Simone, Università degli Studi di Napoli Federico II

## CODICE

I-SUG

## LUOGO E DATA

Il Convegno si terrà a Bologna  
il 15-16 Novembre 2018 presso

**I portici Hotel**

Via Indipendenza, 69

## CONTATTI

**Monica Gianni**

Via Rettangolo, 12-14

67039 Sulmona (AQ)

T. +39 0864 210101

[www.tstat.it](http://www.tstat.it) | [www.tstattraining.eu](http://www.tstattraining.eu)  
[formazione@tstat.it](mailto:formazione@tstat.it) | [training@tstat.eu](mailto:training@tstat.eu)

We present CUB, a Stata routine for modelling ordinal data via a class of finite mixture distributions accounting for both uncertainty and feeling components of an ordered decisional process. This routine allows for modelling also over-dispersion, inflated categories, and large heterogeneity occurrences. Model parameters are estimated by maximum likelihood. We explore various features of the package CUB, including simulation routines.

**Two-part models, hurdle models and zero-inflated are robust to endogeneity among the parts** • David M. Drukker, StataCorp.

There are many models for an outcome that a mass point at a boundary value and is continuously or discrete distributed over a large number of off-boundary values. Two-part models (TPMs), hurdle models (HMs), and zero-inflated models (ZIMs) use different approaches to combine distinct models for boundary and off-boundary values. Except for a few “cake debate” papers whose assertions were not accepted, the vast majority of the literature has either assumed that the process determining when the outcome is on or off the boundary must be exogenous or that any endogeneity must be modeled. Drukker (2017) showed that, contrary to conventional belief, TPMs are robust to endogeneity of the on-off-boundary process in that they identify the mean of the outcome conditional on covariates. This talk covers the following points: i) It reviews this literature and the argument that shows that TPMs are robust to the endogeneity of on-off-boundary process. ii) It discusses new results showing that HMs and ZIMs are also robust to the endogeneity of on-off-boundary process. iii) It presents a new Stata command that estimates the parameters of interest in endogenous TPMS and HMs and shows how to subsequently use `-margins-` to estimate the effects of interest.

**Risk-adjustment procedures and graphical representations of outcome rates for institutional comparisons** • Jacopo Lenzi, Università degli Studi di Bologna

An overriding goal of outcomes research is measuring and comparing hospital performance using readily available administrative data. Risk-adjustment techniques develop from conventional logistic regression analysis, but some precautions need to be taken to account for the positive correlation between observations from within the same hospital. These include the use of generalized estimating equations, which are available in Stata as `xtgee`. Two effective graphs that illustrate outcome measures across different providers and incorporate sample size information are the caterpillar plot and the funnel plot, which can be obtained using the `ecplot` and `funnelcompar` packages, respectively.

## 12.15 - 13.00 III SESSIONE - EXPLOITING THE POTENTIAL OF STATA 15, I

**Efficient dynamic documents using Stata** • Bill Rising, StataCorp.

Stata 15 includes three new commands for producing dynamic documents: `-dyndoc-`, `-putdocx-`, and `-putpdf-`. These commands have generated much interest in the user community; this has led to a large amount of community-contributed software. In this presentation, will give some tips about how to use the commands efficiently both with official Stata software and with some of these community-contributed tools.

13.00 - 14.00 Pranzo

## 14.00 - 14.50 IV SESSIONE - COMMUNITY CONTRIBUTED, II

**Recurrent-event analysis with Stata: methods and applications** • Francesca Ghilotti, Rino Bellocco, Karolinska Institutet, Università degli Studi di Milano-Bicocca

Often in medical studies the event of interest can recur in the same patient over time. Even if time-to-first event analysis or the use of Poisson regression are still possible, they prevent the use of data to its full potential. If the outcome can occur more than once, failure times are correlated within subject and methods accounting for lack of independence are needed. Different statistical models are available for analyzing such data, among which the Andersen and Gill (AG) model, the Prentice, Williams and Peterson Total Time (PWP-TT) model and frailty models. The aim of this presentation is to review statistical techniques for multiple-failure survival data, and to show how to implement them in Stata.

**Le determinanti del voto referendario del 4 Dicembre 2016** • Mariano Bella e Giovanni Graziano, Ufficio Studi Confcommercio

This paper offers new evidence on the socio-economic-demographic determinants of the referendum of December 4, 2016 through the analysis of the vote in Italian municipalities. The results indicate a strong ideology of the vote, in the sense that the political orientation expressed in the 2014 elections significantly influences its orientation. Moreover, it is not so much the youth vote that has determined the rejection of the reform as the social unease, summarized through the unemployment rate and the share of commuters in the municipalities. This also involves a reduced explanatory capacity of the genuine territorial variables. Overall, the vote was determined mainly by political affiliation, then by the self-assessment of one's socio-economic status and, only residually, by personal opinion on the contents of the reform. If democracy is a result built up by the exercise of “electoral



profession”, then the revival of civic education, updated in the direction of enriching it with the economic implications of good institutions, can be an incentive to actually express preferences on the basic rules.

#### 14.50 - 15.50 V SESSIONE - TRICKS AND TIPS

**Simple Tools for Saving Time** • Bill Rising, StataCorp.

This brief talk will show some simple tools for saving time when working with Stata. This will be a hodgepodge of items whose goal is to reduce the amount of thought, coordination, and human memory required of common tasks in a complex work environment while speeding up such tasks greatly.

**Calling External Routines in Stata** • Giovanni Cerulli e Antonio Zinilli, IRCrES-CNR

One of less well known functions in Stata, is the possibility to call external routines, written in other software, to perform specific tasks within Stata. This talk offers some insights on how to develop a Stata ADO file embedding an external software routine to execute in Stata using the Stata module **stree**, written to allow users to run regression trees (a Machine Learning technique currently unavailable in Stata).

**Finding data embedded in text files: using -fileread()- and basic string functions to extract spatial coordinates from google map or counts in preformatted documents** •

Giovanni Capelli, Università degli Studi di Cassino e del Lazio Meridionale

From Stata 13 on, Stata supports the long string format strL. The programming function **fileread()** can be used to upload in a Stata long string field an entire text or binary file, found in a local directory or uploaded from a webpage. These long string fields can then be searched to extract specific numeric or categorical data. We illustrate the use of the **fileread()** programming function, coupled with the string functions **strpos()** and **substr()**, to solve the following issues: i) Extract spatial coordinates from a database of individual addresses using google map APIs calls; and ii) Extract count data from a non-anonymous version of multiple Word structured files and automatically rebuild an anonymous PDF version of the file through LaTeX.

15.50 - 16.00 Pausa caffè

#### 16.00 - 17.00 VI SESSIONE - EXPLOITING THE POTENTIAL OF STATA 15, II

**Estimating the average causal effect on an ordinal outcome of an endogenously assigned treatment from an endogenously selected sample** • David M. Drukker, StataCorp.

This talk discusses the average causal effect (ACE) of an endogenous binary treatment on an ordinal outcome when the sample is subject to endogenous selection. It shows how to estimate the ACE using an extended regression model (ERM) command in Stata. It illustrates how to do regression adjustment in Stata and discusses standard errors for sample-averaged treatment effects and population-averaged treatment effects.

#### 17.00 - 17.15 VII SESSIONE - REPORT TO USERS WISHES AND GRUMBLES • BILL RISING AND DAVID M. DRUKKER, STATA CORP

La sessione “Wishes and Grumbles” offre ai partecipanti la possibilità di interagire direttamente con la StataCorp: sarà possibile evidenziare problemi o limitazioni del software nonché suggerire eventuali miglioramenti o comandi che potrebbero essere inclusi in Stata.

20.15 *Cena Sociale (facoltativa)*

Come ogni anno proponiamo una serata insieme al termine del Convegno. Quest’anno la cena sociale si terrà, la sera del 15 Novembre, presso il ristorante **C’ERA UNA VOLTA...**, Via Massimo D’Azeglio, 9.

Il ritrovo sarà direttamente al ristorante alle ore 20.15 (previa conferma di adesione al momento dell’iscrizione).



## CORSO DI FORMAZIONE

# Joint Modelling of Longitudinal and Survival Data

Michael J. Crowther • Department of Health Sciences University of Leicester

## DESCRIZIONE DEL CORSO

The joint modelling of longitudinal and survival data has been an area of growing interest in recent years, with the benefits of the approach becoming recognised in ever widening fields of study. The models can provide both an effective way of conducting an analysis of a survival endpoint (e.g. time to death), influenced by a time-varying covariate measured with error, or alternatively correct for non-random dropout in the analysis of a longitudinal outcome (e.g. a biomarker such as blood pressure). This one-day course will provide an introduction to joint modelling through real applications to both clinical trial data and electronic health records, using examples in cancer and liver cirrhosis. We will study the methodological framework, underlying assumptions, estimation, model building and predictions. We will also consider current developments in the field, looking at some of the many extensions of the standard framework, such as the ability to model multiple biomarkers and competing risks. The course will consist of lectures and computing exercises making use of the **-stjm-** and **-megenreg-** packages in Stata, written by the course lecturer.

**DESTINATARI:** This one day workshop is of particular interest to biostatisticians, epidemiologists, applied statisticians and researchers or professionals working in economics, the social sciences or public health wishing to carry out survival analysis on longitudinal/panel data in their applied research.

**REQUISITI RICHIESTI:** Participants should be familiar with Stata. A working knowledge of survival analysis and an introductory knowledge of panel data is required.

## PROGRAMMA

- Introduction
- Survival analysis, longitudinal analysis and their combination
- Joint modelling of longitudinal and survival data
- Extended association structures and predictions
- Further topics in joint modelling

## INFORMAZIONI GENERALI

Il materiale didattico distribuito include le dispense con la parte teorica, i do file e le banche dati per l'implementazione di tutte le applicazioni empiriche e una licenza temporanea del Software Stata 15 valida per 30 giorni dall'inizio del corso. Data la natura applicata del corso si consiglia l'utilizzo del proprio personal computer per seguire autonomamente le sessioni applicate.

Il numero massimo di iscritti ammessi al Corso di Formazione è 15, ed il termine per presentare la propria richiesta di ammissione è il 5 Novembre 2018.

Per ulteriori informazioni consultare la pagina del convegno <https://www.tstat.it/utenti/xv-convegno-italiano-degli-utenti-di-stata/> oppure contattare la segreteria organizzativa a [formazione@tstat.it](mailto:formazione@tstat.it).

