



## VI Scuola Estiva Residenziale

# STATISTICA PER ECONOMIA, SCIENZE SOCIALI, EPIDEMIOLOGIA CLINICA E SANITA' PUBBLICA

Firenze, 9-14 Settembre 2018

La Scuola Estiva in Statistica per le Scienze Economiche, Scienze Sociali, Epidemiologia Clinica e Sanità Pubblica è studiata per offrire agli studenti di dottorato/specializzazione una combinazione ideale di fondamenti teorici ed elementi applicati. In ciascun corso, le sessioni teoriche sono affiancate ad applicazioni pratiche ed esempi, in cui il docente chiarisce le limitazioni e i punti di forza di ogni metodologia. Enfasi è posta anche sui criteri per la scelta dello strumento di analisi statistica più appropriato per il problema oggetto di studio. Il fine ultimo dei corsi è assicurare ai partecipanti sia la piena comprensione dei temi teorici, sia familiarità con i problemi pratici che si possono incontrare nelle applicazioni sui dati, così da poter implementare autonomamente le metodologie oggetto dei singoli corsi. Alla fine della scuola, i partecipanti saranno in grado, con l'aiuto delle routine di Stata sviluppato durante le sessioni, di implementare autonomamente le metodologie utilizzate durante il corso nel proprio specifico contesto di ricerca.

La scuola si apre con il corso **Introduzione a Stata** (obbligatorio per tutti coloro che non hanno maturato una precedente esperienza con il software Stata) che fornisce ai partecipanti gli strumenti necessari per ottimizzare al meglio le sessioni applicate durante la settimana e di eseguire la gestione e l'analisi dei dati autonomamente in Stata in un modo efficace e riproducibile. Per chi invece vanta già una buona conoscenza di Stata, viene proposto il modulo relativo alla **Programmazione**, utile per chi desidera personalizzare i comandi di Stata o scrivere i propri *do file* per rendere il lavoro più efficace, permettendoli di automatizzare le fasi più ripetitive del lavoro per quanto riguarda la gestione, l'analisi e la presentazione dei dati in Stata.

### CODICE SCUOLA ESTIVA

I-SE10

### DATA E LUOGO

Firenze, 9-14 Settembre 2018  
CISL Studium Center  
Via Della Piazzola, 71  
I-50123 Firenze  
<http://www.centrostudi.cisl.it>

Domenica: 10.00 - 17.30  
Lunedì-Venerdì: 9.00-17.00

La sessione di **Scienze Economiche** offre ai partecipanti un'introduzione esaustiva alle principali metodologie utilizzate nell'analisi dei dati microeconomici, illustrando, sia dal punto di vista teorico, sia da quello applicato, le seguenti metodologie: analisi di regressione lineare, modelli con variabili strumentali, modelli con variabili dipendenti binarie e categoriche, modelli multinomiali e l'analisi dei dati panel.

La sessione di **Epidemiologia Clinica e Sanità Pubblica** si concentra invece sui metodi per il trattamento di dati amministrativi e ricerca biomedica, illustrando: le diverse metodologie disponibili per calcolare i principi, indici di sintesi e di elaborare statistiche inferenziali in ambito sanitario; gli strumenti principali per l'analisi dei dati continui e binomiale attraverso i modelli di regressione; e offrendo un'introduzione all'analisi di sopravvivenza utilizzata negli studi di coorte e sperimentali.

# SCUOLA ESTIVA IN STATISTICA PER ECONOMIA, SCIENZE SOCIALI, EPIDEMIOLOGIA CLINICA E SANITA' PUBBLICA

## DESTINATARI

La scuola è rivolta a studenti, dottorandi, specializzandi e assegnisti di ricerca in Demografia, Scienze Economiche, Scienze Politiche, Scienze Sociali ed Epidemiologia Clinica, Sanità Pubblica e Medicina, Psicologia e Sociologia.

## REQUISITI RICHIESTI

**Scienze Economiche:** è richiesta la conoscenza di base di statistica o di econometria.  
**Epidemiologia Clinica e Sanità Pubblica:** è richiesta la conoscenza della statistica di base.

Per tutte le sessioni è gradita un minimo di dimestichezza nell'uso del PC.

## PROGRAMMA

### MODULO A

(9 settembre - 10.00-17.30)

#### SESSIONE I: INTRODUZIONE A STATA

### INTRODUZIONE A STATA

1. Interfaccia utente: le finestre di Stata
2. Il lavoro interattivo - organizzazione del lavoro in Stata
3. Risorse disponibili in Stata
4. I file e le cartelle – tipi ed estensioni
5. La sintassi di Stata
6. I filtri: *by, bysort, if, e in*
7. Il registro del lavoro – il così detto “file log”

#### SESSIONE II: GLI ELEMENTI FONDAMENTALI DI STATA

1. Caricare/Importare ed esportare banche dati in diversi formati in Stata: *use, insheet, infile, import delimited, import excel, export excel, outsheet*
2. Esplorare i dati: *describe, list, codebook, browse/edit*
3. Visione di sintesi dei dati – creare le statistiche sommarie e le tabelle: *summarize, table, tabulate, tabstat, epitab*

#### SESSIONE III: GESTIONE DEI DATI IN STATA

1. Il “Data housekeeping” in Stata: *label, recode, replace, gen, egen*
2. Come creare, eliminare e trasformare dati
3. Alcuni grattacapi: valori mancanti (*mvencode/mvdecode*), gestione delle date (*date()*)
4. Gestione delle variabili stringhe
5. Fusione delle banche dati in Stata: *append e merge, reshape*

#### SESSIONE IV: I GRAFICI CON STATA UN PRIMO SGUARDO

1. Logica e sintassi dei comandi *graph*
2. Anatomia e fisiologia dei grafici in Stata
3. Tipologie di grafici: *pie, bar, hist, box, twoway scatter, twoway line, twoway connect, twoway rarea, ecc.*
4. Personalizzazione di un grafico

#### SESSIONE V: I FILE DO (TEMPO PERMETTENDO)

1. Sviluppo di un *do file*

<https://www.tstat.it/formazione/scuola-estiva-statistica/>



# SCUOLA ESTIVA IN STATISTICA PER ECONOMIA, SCIENZE SOCIALI, EPIDEMIOLOGIA CLINICA E SANITA' PUBBLICA

## MODULO B

(9 settembre - 10.00-17.30)

### SESSIONE I: ACQUISIRE I FONDAMENTI DELLA PROGRAMMAZIONE IN STATA

### SESSIONE II: PROGRAMMAZIONE DI BASE IN STATA

### SESSIONE III: POTENZIARE LA CAPACITA' DI PROGRAMMARE IN STATA

## INTRODUZIONE ALLA PROGRAMMAZIONE IN STATA

1. La sintassi di Stata
  2. Macro globali e locali
  3. Scalari e matrici
  4. Riutilizzare i risultati dei comandi in Stata: comandi *e-class* e *r-class*
- 
1. Scrivere e modificare un programma
    - Programmi con argomenti posizionali
    - Programmi con argomenti denominati
    - Programmi con argomenti utilizzando il costrutto *syntax*
    - I file *do* e *ado*
  2. Il *debugging* in Stata
  3. Schemi di riferimento per i comandi scritti dagli utenti: alcuni suggerimenti
- 
1. Le costruzioni di un loop in Stata
    - Il comando *foreach*
    - Il comando *forvalues*
    - Il comando *while*
  2. Utilizzo del qualificatore *if*
  3. Lavorare con le matrici in Stata
    - Inserire una matrice "a mano"
    - Creare un matrice dai risultati di una stima
    - Ottenere sotto-matrici e matrici combinate
    - Operatori e funzioni matriciali
    - *Looping* con le matrici

## SESSIONE SCIENZE ECONOMICHE

## MODULO C

(10-14 settembre - 9.00-17.00)

### SESSIONE I: ANALISI DI REGRESSIONE LINEARE IN STATA

### SESSIONE II: I TEST

## ANALISI DI REGRESSIONE IN MICROECONOMETRIA

1. Un semplice esempio
  2. Un primo esame dei dati
  3. Ottenere le "predizioni"
  4. Regressione multipla e interpretazione dei coefficienti
  5. Tipologie di coefficienti standardizzati
  6. Testare ipotesi lineari sui coefficienti
  7. Ottenere le "predizioni"
  8. Regressione multipla e interpretazione dei coefficienti
  9. Tipologie di coefficienti standardizzati
  10. Testare ipotesi lineari sui coefficienti
- 
1. Identificazione e trattamento di dati anomali ed influenti
  2. Verifica e trattamento della multicollinearità
  3. Verifica della normalità dei residui
  4. Controllo e trattamento dell'omoschedasticità dei residui
  5. Verifica della linearità
  6. Test di corretta specificazione del modello
  7. Test di autocorrelazione dei residui (cenni)
  8. Predizioni marginali

<https://www.tstat.it/formazione/scuola-estiva-statistica/>



# SCUOLA ESTIVA IN STATISTICA PER ECONOMIA, SCIENZE SOCIALI, EPIDEMIOLOGIA CLINICA E SANITA' PUBBLICA

## SESSIONE III: STIMATORI IV PER MODELLI LINEARI

1. Stimatori IV in Stata: *ivregress*, *ivreg2*
  - Stimatore IV nel caso esattamente identificato
  - Stimatori per il caso sovra identificato: 2SLS e GMM
  - Stimatori per modelli con regressori binari endogeni: *treatreg*
  - *Limited Information Maximum Likelihood*: LIML
  - Stimatore per sistemi di equazioni simultanee: 3SLS
2. Test di validità delle restrizioni di sovra-identificazione, test di rilevanza degli strumenti (*weak instruments*)

## SESSIONE IV: MODELLI A VARIABILE DIPENDENTE BINARIA E CATEGORICA

1. Stimatori per modelli a variabile dipendente binaria in Stata: *probit*, *logit*, *hetprobit*, *ivprobit*, *regress*
  - Test di specificazione e test delle ipotesi
  - Stima degli effetti marginali
2. Stimatore probit con eteroschedasticità: *hetprobit*
3. Modelli binari con regressori endogeni: *ivprobit*
4. Stimatori per modelli multinomial logit: *mlogit*, *clogit*, *asclogit*, *nlogit*
5. Stimatori per modelli con categorie ordinate: *oprobit*, *ologit*

## SESSIONE V: I FONDAMENTI, GESTIONE E ANALISI ECONOMETRICA DEI DATI PANEL

1. Cenni preliminari:
  - Stata
  - Il modello classico di regressione lineare multivariata
2. *Data-set* in formato *panel*:
  - Gestione dei dati
3. Gli operatori *Time Series* in Stata
4. Benefici dei dati *panel* per l'analisi econometria

## SESSIONE VI: MODELLI STATISTICI PER DATI PANEL

1. Il modello di regressione ad effetti "fissi"
  - Un metodo di stima semplice: lo stimatore in differenze prime (FD).
  - Stimatori più "precisi" nel caso di errori idiosincratrici, *white-noise*: lo stimatore *Least Squares dummy variable* (LSDV) e lo stimatore *Within*. Equivalenza tra gli stimatori *Within* e LSDV.
  - Ma se l'errore idiosincratrico è persistente?
  - Allora FD è più preciso di LSDV
  - Una trasformazione utile nei modelli *panel*: *Forward Orthogonal Deviations (FOD)*
  - Cautele da seguire per l'implementazione in Stata: significato della costante nella stima FD; significato alla costante nella stima LSDV; correzione degli standard errors nella stima *Within*
  - Eterogeneità individuale: Test di significatività congiunta degli effetti fissi
2. Il modello di regressione ad effetti "random"
  - Stimatore *Pooled Ordinary Least Squares (OLS)*
  - Stimatore *Within*
  - Stimatore *Between*
  - Stimatore *Generalised Least Squares (GLS)*
  - Stimatore *Feasible Generalised Least Squares (FGLS)*
  - Eterogeneità individuale: test di *Breusch e Pagan*
3. Effetti "fissi" o effetti "random"?
  - Test di *Hausman*
  - Un test robusto per eteroschedasticità e autocorrelazione: l'approccio della regressione ausiliaria a la *Mundlak*



## SCUOLA ESTIVA IN STATISTICA PER ECONOMIA, SCIENZE SOCIALI, EPIDEMIOLOGIA CLINICA E SANITA' PUBBLICA

### SESSIONE VII: APPROFONDIMENTI

1. Test di autocorrelazione
2. Test di eteroschedasticità
3. Correzione degli standard errors per autocorrelazione e eteroschedasticità
  - La correzione di *White* per autocorrelazione ed eteroschedasticità suggerita da Arellano
4. Il risultato di *Stock & Watson* sulla inconsistenza della correzione di *White* per sola eteroschedasticità nei modelli con effetti individuali
5. Sbilanciamento nei dati
6. Modelli per dati *multi-level*
  - La critica di *Moulton* ai modelli che non specificano adeguatamente le componenti dell'errore con dati *multi-level*
  - Stimatori GLS per modelli con componenti multiple dell'errore
  - Test di specificazione
7. Modelli con variabili esplicative predeterminate e endogene
  - Stimatori LSDV e *Random effects* a variabili strumentali
  - Stimatore di *Hausman-Taylor*
  - Stimatore FD a variabili strumentali
  - Stimatore FOD a variabili strumentali
  - Cenni di stima per i modelli dinamici
  - Analisi di corretta specificazione: test di validità e rilevanza degli strumenti, test di autocorrelazione
8. Considerazioni sugli sviluppi futuri

## SESSIONE EPIDEMIOLOGIA CLINICA E SANITA' PUBBLICA (METODI PER IL TRATTAMENTO DI DATI AMMINISTRATIVI E DI RICERCA BIOMEDICA)

### MODULO D

(10-11 settembre - 9.00-17.00)

#### SESSIONE I: NATURA E FONTI DEI DATI

1. Natura dei dati e loro archiviazione in Stata: dati categorici dicotomici, nominali, ordinali, dati numerici
2. Interpretare le variabili: Misure di Esposizione, Misure di Esito, Confondenti
3. Fonti dei dati: Database istituzionali e raccolte ad hoc

#### SESSIONE II: LA CLASSIFICAZIONE DELLE MALATTIE

1. Sistemi di classificazione in ambito sanitario
2. Classificazione delle malattie e delle procedure. ICD-9, ICD-9CM, ICD-10
  - La suite di comandi icd: *icd9*, *icd9p*, *icd10*

#### SESSIONE III: ANALISI DI DATI OSPEDALIERI

1. Il sistema DRG
2. La scheda di dimissione ospedaliera ed il dialogo *groupes-Stata*
3. Indici di case-mix. Indici di qualità della codifica

#### SESSIONE IV: TECNICHE DI STATISTICA DESCRITTIVA

1. Indici di tendenza centrale ed indici di dispersione: *summarize*, *means*
2. Tabelle a singola e doppia entrata: *tabulate*, *table*
3. Rappresentazioni grafiche: diagrammi a barre, distribuzioni di frequenza, serie temporali, mappe

#### SESSIONE V: TEORIA DELLA PROBABILITA' ED INFERENZA STATISTICA

1. Incertezza e probabilità
2. Distribuzioni di probabilità: distribuzione normale e binomiale
3. Introduzione all'inferenza statistica
4. La distribuzione campionaria

<https://www.tstat.it/formazione/scuola-estiva-statistica/>





# SCUOLA ESTIVA IN STATISTICA PER ECONOMIA, SCIENZE SOCIALI, EPIDEMIOLOGIA CLINICA E SANITA' PUBBLICA

5. Limiti di confidenza: *ci, cij*
6. Test di ipotesi parametrici e non parametrici: *ttest, ranksum, kwallis*, le opzioni *chi2* ed *exact* di *tabulate*
7. Potenza e dimensioni del campione

## MODULO E (12-13 settembre - 9.00-17.00)

### SESSIONE I: REGRESSIONE LINEARE

## REGRESSIONE LINEARE E LOGISTICA

1. Correlazione: i comandi *correlate, pwcorr, spearman, graph matrix*
2. Regressione lineare in Stata: i comandi *regress, predict, margins*
3. La Tabella ANOVA
4. Regressione lineare semplice
5. Interpretazione dei coefficienti
6. Regressione lineare multivariabile
7. Variabili *dummy*
8. Interazione e confondimento
9. Valutazione del modello

### SESSIONE II: REGRESSIONE LOGISTICA

1. Il modello di regressione logistica
2. Regressione logistica in Stata: i comandi *logit, logistic, predict, margins*
3. Maximum Likelihood Estimation
4. Interpretazione dei coefficienti del modello logistico ed *Odds Ratio*
5. Regressione logistica multivariabile
6. Test del Likelihood-ratio
7. Test di Hosmer-Lemeshow

### SESSIONE III: APPLICARE LE TECNICHE DI REGRESSIONE AI DATI AMMINISTRATIVI

1. Analizzare la durata di degenza e la sopravvivenza dopo interventi chirurgici
2. Regressione logistica e *risk adjustment* sui database amministrativi

## MODULO F (14 settembre - 9.00-17.00)

### SESSIONE I: PRINCIPI GENERALI

## ANALISI DI SOPRAVVIVENZA

1. Motivazioni dell'analisi di sopravvivenza: il comando *stset*
2. Osservazioni *censored*
3. Stima della funzione di sopravvivenza secondo il metodo di Kaplan-Meier: *sts list*
4. Funzioni di rischio (rischio istantaneo e rischio cumulativo)

### SESSIONE II: INFERENZA SULLA SOPRAVVIVENZA

1. Confronto tra funzioni di sopravvivenza: *sts graph*
2. Test del Log-Rank e di Wilcoxon: *sts test*
3. Numerosità campionaria: *power logrank*

### SESSIONE III: STUDIARE LA SOPRAVVIVENZA CON MODELLI DI REGRESSIONE

1. Modello di regressione di Cox: *stcox*
2. Interpretazione dei coefficienti del modello di regressione
3. Valutazione dell'assunzione di proporzionalità del rischio: *stphplot, estat phtest*

### SESSIONE IV: RISK ADJUSTMENT CON DATI DI SOPRAVVIVENZA

1. La sopravvivenza intraospedaliera come "time-to-event"



# SCUOLA ESTIVA IN STATISTICA PER ECONOMIA, SCIENZE SOCIALI, EPIDEMIOLOGIA CLINICA E SANITA' PUBBLICA

## DOCENTI

**Una-Louise Bell** TStat Training

**Giovanni Bruno**

Università Commerciale L. Bocconi

**Giovanni Capelli**

Università di Cassino e del Lazio Meridionale

**Bruno Federico**

Università di Cassino e del Lazio Meridionale

**Roberto Gabriele**

Università degli Studi di Trento

## REGISTRAZIONE

La domanda di partecipazione dovrà avvenire tramite il modulo di registrazione, disponibile presso la segreteria, unitamente al curriculum vitae **entro il 5 Agosto 2018**. L'ente organizzatore vaglierà le domande e sulla base dell'ordine d'iscrizione e del curriculum deciderà l'ammissione. Un acconto, non rimborsabile, pari a € 150.00 sarà richiesto come deposito a seguito dell'avvenuta ammissione, a garanzia del posto. La Scuola si attiverà al raggiungimento del numero minimo previsto di 10 iscritti per sessione e, per assicurare una proficua interazione con il docente, il numero di partecipanti verrà limitato a 15 per ciascuna sessione.

## CONTATTI

**Monica Gianni**

TStat S.r.l. | Via Rettangolo, 12-14

I-67039 Sulmona (AQ)

T. +39 0864 210101

TStat Training | Kleebergstraße, 8

D-60322 Frankfurt am Main

[formazione@tstat.it](mailto:formazione@tstat.it)

[www.tstattraining.eu](http://www.tstattraining.eu)

[www.tstat.it](http://www.tstat.it)

## QUOTA DI ISCRIZIONE

A beneficio dei partecipanti delle scuole estive precedenti e/o dei Workshop TStat, abbiamo introdotto un sistema di moduli a scelta, per permettere loro che hanno già preso parte a precedenti edizioni di seguire solo i moduli di interesse.

### INTERA SETTIMANA

Studenti\*: € 950.00

Assegnisti/Specializzandi: € 1450.00

### MODULO A, B oppure F (costo per ogni singolo modulo da 1 giorno)

Studenti\*: € 210.00

Assegnisti/Specializzandi: € 310.00

### MODULO D oppure E (costo per ogni singolo modulo da 2 giorni)

Studenti\*: € 395.00

Assegnisti/Specializzandi: € 495.00

### MODULO C (modulo da 5 giorni)

Studenti\*: € 820.00

Assegnisti/Specializzandi: € 1245.00

\*Per usufruire dello status "studente" è necessario presentare copia del libretto universitario o un certificato di iscrizione (in carta semplice) all'Università ed essere **studenti a tempo pieno**.

I prezzi si intendono IVA 22% esclusa. L'aliquota IVA non sarà applicata per Enti Pubblici soggetti ad esenzione a norma dell'art. 14 c. 10 della L. 537/93 per la partecipazione a corsi di formazione dei propri dipendenti.

### La quota di iscrizione include:

Il materiale didattico (dispense, *do file* e banche dati utilizzate). Grazie al progetto TStat "Investire nei Giovani Oggi", a tutti i dottorandi sarà rilasciata una licenza semestrale di Stata nella versione StataIC/15. Agli iscritti non riferibili allo status di studente/dottorando verrà rilasciata una licenza short course di Stata 15 della durata di 30 giorni.

Pernottamento in camera singola presso il Centro Studi CISL, a partire dal giorno precedente la data di inizio corso fino al giorno di chiusura, con trattamento di mezza pensione (colazione, pranzo e pause caffè). Aperitivo di benvenuto e visita guidata della Città di Firenze.

Ulteriori informazioni sulla modalità di iscrizione, incluso termini e condizioni di partecipazione sono disponibili nel nostro sito alla pagina <https://www.tstat.it/formazione/scuola-estiva-statistica/>

