

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

DESCRIZIONE GENERALE

La Scuola Estiva in Statistica per le Scienze Economiche, Scienze Sociali, Epidemiologia Clinica e Sanità Pubblica è studiata per offrire una combinazione ideale di fondamenti teorici ed elementi applicati. In ciascun corso, le sessioni teoriche (in cui sono illustrati i fondamenti statistici di alcuni strumenti di analisi) 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 del corso.

Per assicurare una proficua interazione con il docente, il numero di partecipanti a ogni sessione viene limitato.

A beneficio dei partecipanti delle scuole estive precedenti e/o dei Workshop TStat, abbiamo introdotto un sistema di moduli a scelta. Per coloro che hanno già preso parte a precedenti edizioni, il corso di Introduzione a Stata non è obbligatorio e sarà possibile scegliere fra i moduli a disposizione così da poter seguire corsi diversi da quelli già frequentati in precedenza.

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 e 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 SESSIONE EPIDEMIOLOGIA CLINICA E SANITÀ PUBBLICA

MODULO A: INTRODUZIONE A STATA (4 GIUGNO)

L'obiettivo del modulo è quello di fornire all'utente le nozioni introduttive che consentono di lavorare autonomamente in Stata oltre a una panoramica completa delle funzioni di base, che sono illustrate attraverso una miscela di esempi concreti.

SESSIONE I: INTRODUZIONE A STATA

1. Interfaccia utente: le finestre di Stata
2. Logica e sintassi: Come "pensa" Stata?
3. Piccoli ma importanti: *if, by, bsort, in*
4. Usare Stata in forma interattiva e in programmazione: riga di comando e menù VS .do files
5. File e cartelle in Stata:
 - tipi ed estensioni (*.dta, .gph, .log, .smcl, .dct, ...*)
 - cartelle di lavoro (*pwd, cd, mkdir, ls*)

SESSIONE II: GLI ELEMENTI FONDAMENTALI DI STATA

1. Caricare dati di diversi formati in Stata: *use, insheet, infile, import delimited, import excel*
2. Esplorare i dati in Stata: *describe, list, codebook, browse/edit, summarize*
3. Le Tabelle in Stata: *tabulate, table, epitab*

SESSIONE III: STATA COME GESTIRE DI DATI

1. Il "Data housekeeping" in Stata: *label, recode, replace, gen, egen*
2. Alcuni grattacapi: valori missing (*mvencode/mvdecode*), gestione delle date (*date()*)
3. Come fondere basi di dati con Stata: *append* e *merge, reshape*

SESSIONE IV: GRAFICI CON STATA

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.

MODULO B: EPIDEMIOLOGIA CLINICA E SANITÀ PUBBLICA**STATISTICA DESCRITTIVA E INFERENZIALE SU DATI EPIDEMIOLOGICI, AMMINISTRATIVI E OSPEDALIERI CON STATA (5 E 6 GIUGNO)**

Il modulo introduce lo studente alla conoscenza delle diverse tipologie di dati analizzabili in ambito di epidemiologia e sanità pubblica e dei principali sistemi di classificazione in ambito sanitario, sia ospedalieri che extra-ospedalieri, ed alla loro gestione e manipolazione con Stata. Inoltre, si propone di fornire allo studente gli strumenti utili a creare tabelle, grafici, a calcolare i principali indici di sintesi e ad elaborare statistiche inferenziali.

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 *grouper*-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 PROBABILITÀ ED INFERENZA STATISTICA

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

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

REGRESSIONE LINEARE E LOGISTICA (7 E 8 GIUGNO)

Il corso intende fornire allo studente i principali strumenti per l'analisi di dati continui e binomiali attraverso la costruzione di modelli di regressione univariabili e multivariabili, e proporre l'applicazione a dati provenienti da raccolte ad hoc o da database amministrativo-epidemiologici.

SESSIONE I: REGRESSIONE LINEARE

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

ANALISI DELLA SOPRAVVIVENZA (9 GIUGNO)

Il corso intende fornire allo studente le conoscenze e gli strumenti per l'analisi di variabili del tipo "time-to-event". Tale tecnica è comunemente utilizzata negli studi di coorte ed in quelli sperimentali.

SESSIONE I: PRINCIPI GENERALI

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"

PROGRAMMA**SESSIONE SCIENZE ECONOMICHE****MODULO A: INTRODUZIONE A STATA (4 GIUGNO)**

L'obiettivo del corso è quello di fornire all'utente le nozioni introduttive che consentono di lavorare autonomamente in Stata oltre a una panoramica completa delle funzioni di base, che sono illustrate attraverso una miscela di esempi concreti.

SESSIONE I: INTRODUZIONE A STATA

1. Organizzazione dei files di Stata: *pwd, cd, mkdir*
2. Interfaccia utente: le finestre di Stata
3. I file di Stata – tipi ed estensioni
4. Il lavoro interattivo
5. Organizzazione del lavoro in Stata
6. *Help*
7. *Web resources* in Stata - caricare *updates* e nuovi comandi tramite internet
8. Come interrompere un'esecuzione in Stata
 - Caricamento di banche dati in formato Stata
 - La sintassi di Stata
 - Il file *log*
 - L'uso dei commenti in Stata

SESSIONE II: ELEMENTI FONDAMENTALI DI STATA

1. Visione di sintesi dei dati: *describe, summarize, table*
2. Tipi di variabili
3. Il prefisso *by*
4. Etichette di valore (*Value Labels*)
5. Altri tipi di etichette

6. Visualizzazione dei dati – diverse modalità
7. Come creare, eliminare e trasformare dati
8. Il comando *count*
9. Il comando *sort*
10. Il comando *assert*
11. Il comando *foreach*
12. Le variabili di tipo categorico
13. Come lavorare con valori mancanti

SESSIONE III: IL FILE "DO" – UN PRIMO SGUARDO**SESSIONE IV: COME CARICARE I DATI IN STATA**

1. Importare ed esportare banche dati ASCII create da *spreadsheet: insheet – outsheet*
2. Accenno al comando *infile*
3. Visualizzazione dei dati: *edit, browse, list, describe, codebook*
4. Caricamento di dati in formato string
5. Trattamento di numeri interi di grande dimensione
6. Il software *Stat/Transfer*

SESSIONE V: GESTIONE DEI DATI – FUSIONE DI BANCHE DATI

1. Il comando *append*
2. L'importanza delle banche dati *master* e *using*
 - Unione *Match*
 - Unione assicurandosi che gli elementi siano unici
 - Errori di unione
3. *Updates*
4. I dati in formato *wide* invece di *long*

SESSIONE VI: GRAFICI

1. Aspetti di base del comando *graph* (*matrix, box, bar, pie, twoway*)
2. Personalizzazione di un grafico

MODULO B: INTRODUZIONE ALLA MICROECONOMETRIA (DAL 5 AL 9 GIUGNO)**SESSIONE I: ANALISI DI REGRESSIONE LINEARE IN STATA**

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

SESSIONE II: I TEST

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

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 econometrica

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 idiosincratichi, *white-noise*: lo stimatore *Least Squares dummy variable* (LSDV) e lo stimatore *Within*. Equivalenza tra gli stimatori *Within* e LSDV.

- Ma se l'errore idiosincratico 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* (POLS)
 - 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*

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

DOCENTI

BELL Una-Louise, TStat S.r.l.
 BRUNO Giovanni, Università Commerciale L. Bocconi
 CAPELLI Giovanni, Università di Cassino e del Lazio Meridionale
 FEDERICO Bruno, Università di Cassino e del Lazio Meridionale
 GABRIELE Roberto, Università degli Studi di Trento

DATA E ISCRIZIONE

La scuola estiva è prevista a Firenze dal 4 al 9 Giugno 2017. La frequenza alla scuola è subordinata al pagamento del seguente contributo di partecipazione. Il costo è riferito ad una sessione che dovrà essere indicata nello specifico modulo di registrazione.

QUOTA DI ISCRIZIONE STANDARD

INTERA SETTIMANA (Moduli A + B, 6 giorni)

Studenti*: € 450.00
 Assegnisti di Ricerca / Specializzandi: € 650,00
 Ricercatori: € 950.00

MODULO A (1 giorno)

Studenti*: € 151.00
 Assegnisti di Ricerca / Specializzandi: € 206.00
 Ricercatori: € 256.00

MODULO B (5 giorni)

Studenti*: € 383.00
 Assegnisti di Ricerca / Specializzandi: € 553.00
 Ricercatori: € 808.00

QUOTA DI ISCRIZIONE RESIDENZIALE

INTERA SETTIMANA (Moduli A + B, 6 giorni)

Studenti*: € 894.00
 Assegnisti di Ricerca / Specializzandi: € 1094,00
 Ricercatori: € 1394.00

MODULO A (1 giorno)

Studenti*: € 151.00
 Assegnisti di Ricerca / Specializzandi: € 206.00
 Ricercatori: € 256.00

MODULO B (5 giorni)

Studenti*: € 759.00
 Assegnisti di Ricerca / Specializzandi: € 929.00
 Ricercatori: € 1184.00

*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 "standard" include:

Il materiale didattico e, grazie al progetto TStat "Investire nei Giovani Oggi", a tutti i dottorandi sarà rilasciata una licenza semestrale di StatalC/14 GradPlan. Iscritti non referibili allo status di studente / dottorando verrà rilasciata una licenza gratuita di Stata 14 della durata di 30 giorni.

La quota di iscrizione "residenziale" include:

Il materiale didattico e, grazie al progetto TStat "Investire nei Giovani Oggi", a tutti i dottorandi sarà rilasciata una licenza semestrale di StatalC/14 GradPlan. Iscritti non referibili allo status di studente / dottorando verrà rilasciata una licenza gratuita di Stata 14 della durata di 30 giorni.

Prenottamento in camera singola presso il Centro Studi CISL, a partire dal giorno di inizio corso fino al giorno di chiusura, con trattamento di mezza pensione (colazione, pranzo e pause caffè). Partecipanti che desiderano pernottare il giorno precedente il corso o il successivo, è pregato di contattare la segreteria per informazioni e disponibilità.

La domanda di partecipazione dovrà avvenire tramite il modulo di registrazione, disponibile presso la segreteria, unitamente al curriculum vitae a formazione@tstat.it entro il 15 Maggio 2017. 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 € 100.00 sarà richiesto come deposito, a seguito dell'avvenuta ammissione, a garanzia del posto. Il numero dei partecipanti è limitato a 15 persone per sessione.



PER ULTERIORI INFORMAZIONI RIVOLGERSI A:

Monica Gianni

Via Rettangolo, 12-14 | 67039 Sulmona (AQ)

Tel. 0864 210101 | Fax 0864 206014

corsi@tstat.it | www.tstat.it