PROGRAMMA

La scuola estiva intende fornire ai partecipanti la strumentazione teorica e applicata necessaria per poter svolgere autonomamente analisi empiriche in Stata. Le lezioni saranno di tipo interattivo ed i partecipanti potranno sperimentare le tecniche apprese attraverso numerose applicazioni empiriche su dati reali effettuate dalle proprie postazioni di calcolo sotto la guida del docente. Inoltre, i corsi si soffermeranno sugli aspetti applicati dell'analisi, enfatizzando l'interpretazione dei risultati piuttosto che la parte computazionale.

SESSIONE SCIENZE ECONOMICHE

INTRODUZIONE A STATA

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

- Organizzazione dei *files* di Stata: ***pwd, cd, mkdir***
- Interfaccia utente: le finestre di Stata
- I *file* di Stata – tipi ed estensioni
- Il lavoro interattivo
- Organizzazione del lavoro in Stata
- *Help*
- *Web resources* in Stata - caricare *updates* e nuovi comandi tramite internet
- 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

- Visione di sintesi dei dati: ***describe, summarize, table***
- Tipi di variabili
- Il prefisso ***by***
- Etichette di valore (*Value Labels*)
- Altri tipi di etichette
- Visualizzazione dei dati – diverse modalità
- Come creare, eliminare e trasformare dati
- Il comando ***count***
- Il comando ***sort***
- Il comando ***assert***
- Il comando ***foreach***
- Le variabili di tipo categorico
- Come lavorare con valori mancanti

SESSIONE III: IL FILE "DO" – UN PRIMO SGUARDO

SESSIONE IV: COME CARICARE I DATI IN STATA

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

SESSIONE V: GESTIONE DEI DATI – FUSIONE DI BANCHE DATI

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

SESSIONE VI: GRAFICI

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

ANALISI DELLA MICROECONOMETRIA

SESSIONE I: ANALISI DI REGRESSIONE LINEARE IN STATA

- Un semplice esempio
- Un primo esame dei dati
- Ottenere le "predizioni"
- Regressione multipla e interpretazione dei coefficienti
- Tipologie di coefficienti standardizzati
- Testare ipotesi lineari sui coefficienti

SESSIONE II: I TEST

- Identificazione e trattamento di dati anomali ed influenti
- Verifica e trattamento della multicollinearità
- Verifica della normalità dei residui
- Controllo e trattamento dell'omoschedasticità dei residui
- Verifica della linearità
- Test di corretta specificazione del modello
- Test di autocorrelazione dei residui (cenni)
- Predizioni marginali

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

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

SESSIONE IV: MODELLI STATISTICI PER DATI *PANEL*

- 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 idiosincratici, *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 della costante nella stima LSDV; correzione degli standard *errors* nella stima *Within*
  - Eterogenità individuale: Test di significatività congiunta degli effetti fissi
- 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*
- Effetti "fissi" o effetti "random"?
  - Test di *Hausman*
  - Un test robusto per eteroschedasticità e autocorrelazione: l'approccio della regressione ausiliaria a la *Mundlak*

SESSIONE V: APPROFONDIMENTI

- Test di autocorrelazione
- Test di eteroschedasticità
- Correzione degli *standard errors* per autocorrelazione e eteroschedasticità
  - La correzione di *White* per autocorrelazione ed eteroschedasticità suggerita da Arellano
- Il risultato di *Stock & Watson* sulla inconsistenza della correzione di *White* per sola eteroschedasticità nei modelli con effetti individuali
- Sbilanciamento nei dati
- 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
- 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
- Considerazioni sugli sviluppi futuri

SESSIONE VI: STIMATORI IV PER MODELLI LINEARI

- 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
- Test di validità delle restrizioni di sovra-identificazione, test di rilevanza degli strumenti (*weak instruments*)

SESSIONE VII: QUANTILE REGRESSION

- Quantile regression in *Stata*: ***qreg***, ***bsqreg*** e ***sqreg***.
- Interpretazione dei coefficienti stimati
- Visualizzazione grafica dei coefficienti stimati per i vari quantili
- Test di specificazione (eteroschedasticità) e test delle ipotesi

SESSIONE VIII: MODELLI A VARIABILE DIPENDENTE BINARIA E CATEGORICA

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

SESSIONE SCIENZE SOCIALI

INTRODUZIONE A STATA

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

- Organizzazione dei *files* di Stata: ***pwd***, ***cd***, ***mkdir***
- Interfaccia utente: le finestre di Stata
- I *file* di Stata – tipi ed estensioni
- Il lavoro interattivo
- Organizzazione del lavoro in Stata
- *Help*
- *Web resources* in Stata - caricare *updates* e nuovi comandi tramite internet
- 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

- Visione di sintesi dei dati: ***describe***, ***summarize***, ***table***
- Tipi di variabili
- Il prefisso ***by***
- Etichette di valore (*Value Labels*)
- Altri tipi di etichette
- Visualizzazione dei dati – diverse modalità
- Come creare, eliminare e trasformare dati
- Il comando ***count***
- Il comando ***sort***
- Il comando ***assert***
- Il comando ***foreach***
- Le variabili di tipo categorico
- Come lavorare con valori mancanti

SESSIONE III: IL FILE "DO" – UN PRIMO SGUARDO

SESSIONE IV: COME CARICARE I DATI IN STATA

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

SESSIONE V: GESTIONE DEI DATI – FUSIONE DI BANCHE DATI

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

SESSIONE VI: GRAFICI

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

ANALISI QUANTITATIVA DEI FENOMENI SOCIALI: LA REGRESSIONE

Questo corso offre un'introduzione alla regressione come strumento per l'analisi quantitativa della variazione dei fenomeni sociali misurati a livello individuale. Data una proprietà individuale di interesse rappresentata da una certa variabile Y, l'analisi di variazione è qui intesa in una duplice accezione:

1) Descrizione dei modi e della misura in cui le manifestazioni di Y variano entro l'intera popolazione oggetto di studio – o, in termini più formali, descrizione della distribuzione di probabilità incondizionata di Y, cioè p(Y).

2) Descrizione dei modi e della misura in cui la distribuzione di probabilità di Y (o qualche sua caratteristica specifica come il valore atteso) varia all'interno di un determinato spazio delle *covariate* V. In questo caso, l'oggetto di interesse è la distribuzione di probabilità condizionata p(Y|V) o, più tipicamente, il valore atteso condizionato E(Y|V). Sia p(Y|V) che E(Y|V) rappresentano particolari esempi della funzione di regressione. La stima dei possibili valori di questa funzione e il loro utilizzo per rispondere a specifici interrogativi di ricerca costituiscono l'analisi di regressione e rappresentano il tema fondamentale di questo corso.

Il corso si divide in sei parti. La prima è dedicata a una breve illustrazione dei modi per descrivere le distribuzioni di probabilità incondizionate p(Y). La seconda parte illustra la logica generale dell'analisi di regressione. La terza parte presenta l'uso dell'analisi di regressione a scopi predittivi, mentre la quarta è dedicata all'analisi di regressione come strumento per la stima di effetti causali. La quinta parte discute l'utilizzo del modello di regressione lineare generalizzato per la stima dei valori della funzione di regressione E(Y|V). Infine, la sesta parte offre un'introduzione all'inferenza statistica applicata all'analisi di regressione.

MODULO A

ANALISI QUANTITATIVA DEI FENOMENI SOCIALI: TECNICHE MULTIVARIATE

Il corso si propone di fornire ai partecipanti un'introduzione ai metodi per l'analisi di dati multivariati attraverso l'impiego di Stata. Grazie all'enorme quantità di dati ormai disponibili in ogni settore industriale e commerciale, le tecniche di analisi statistica multivariata ricoprono oggi più che mai un ruolo fondamentale per l'estrazione di utili informazioni dai dati stessi. Durante il corso saranno illustrate le principali metodologie di analisi multivariata (analisi dei cluster, analisi delle componenti principali, analisi fattoriale) attraverso esempi e casi concreti.

GIORNO I – I DATI MULTIVARIATI: PRIMI INDICATORI DI SINTESI E RAPPRESENTAZIONI GRAFICHE

- Tipi di variabili e il problema dei dati mancanti
- Covarianze, correlazioni e misure di distanza.
- La distribuzione normale multivariata
- Grafici per la visualizzazione di dati multivariati

GIORNO II – ANALISI DELLE COMPONENTI PRINCIPALI E ANALISI FATTORIALE

- Introduzione
- Calcolo delle componenti principali

- Calcolo degli *scores* delle componenti principali
- Scelta del numero di componenti

GIORNO III – ANALISI DEI CLUSTER

- Introduzione agli algoritmi di clustering
- Principali metodi agglomerativi di clustering
    - Il dendrogramma
    - Single linkage
    - Complete linkage
    - Average linkare
    - Metodo di Ward
- Principali metodi non-gerarchici di clustering
    - K-means
- Profilazione dei cluster
- Altre tecniche di analisi multivariata (analisi delle corrispondenze, *scaling* multidimensionale, etc.)
- Profilazione dei cluster

SESSIONE EPIDEMIOLOGIA CLINICA E SANITA' PUBBLICA

INTRODUZIONE A STATA

SESSIONE I: INTRODUZIONE A STATA

- Organizzazione dei *files* di Stata: ***pwd, cd, mkdir***
- Interfaccia utente: le finestre di Stata
- I *files* di Stata – tipi ed estensioni
- Il lavoro interattivo
- Organizzazione del lavoro in Stata
- *Help*
- *Web resources* in Stata - caricare *updates* e nuovi comandi tramite internet
- 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

- Visione di sintesi dei dati: ***describe, summarize, table***
- Tipi di variabili
- Il prefisso ***by***
- Etichette di valore (*Value Labels*)
- Altri tipi di etichette
- Visualizzazione dei dati – diverse modalità
- Come creare, eliminare e trasformare dati
- Il comando ***count***
- Il comando ***sort***
- Il comando ***assert***
- Il comando ***foreach***
- Come lavorare con valori mancanti
- Le variabili di tipo categorico

SESSIONE III: IL FILE "DO" – UN PRIMO SGUARDO

SESSIONE IV: COME CARICARE I DATI IN STATA

- Importare ed esportare banche dati ASCII create da spreadsheet oppure file .txt: insheet – outsheet, infile
- Visualizzazione dei dati: ***edit, browse, list, describe, codebook***
- Gestione della memoria – una nota
- Caricamento di dati in formato *string*
- Trattamento di numeri interi di grande dimensione
- Il software Stat/Transfer

SESSIONE V: VISUALIZZAZONE E DESCRIZIONE DEI DATI

- Il comando append
- L'importanza delle banche dati *master* e *using*
- Unione dei dati
- Unione uno a uno
- Unione **Match**
- Unione assicurandosi che gli elementi siano unici
- *Spreads*
- Errori di unione

- *Updates*
- L'uso dei comandi ***append*** e *merge*
- I dati in formato *wide* invece di *long*

STATISTICA DESCRITTIVA E INFERENZIALE

Il corso 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 – STATISTICA DESCRITTIVA

- Natura dei dati
- Indici di tendenza centrale ed indici di dispersione
- Tabelle a singola e doppia entrata
- Rappresentazioni grafiche

SESSIONE II – TEORIA DELLA PROBABILITA' ED INFERENZA STATISTICA

- Incertezza e probabilità
- Distribuzioni di probabilità: distribuzione normale e binomiale
- Introduzione all'inferenza statistica
- La distribuzione campionaria
- Limiti di confidenza
- Test di ipotesi parametrici e non parametrici
- Potenza e dimensioni del campione

REGRESSIONE LINEARE E LOGISTICA

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.

SESSIONE I – REGRESSIONE LINEARE

- Correlazione
- La Tabella ANOVA
- Regressione lineare semplice
- Interpretazione dei coefficienti
- Regressione lineare *multivariabile*
- Variabili *dummy*
- Interazione e confondimento

SESSIONE VI: GRAFICI

- Aspetti di base del comando *graph* ***(box, bar, pie, histogram, scatter)***
- Personalizzazione di un grafico

- Valutazione del modello

SESSIONE II – REGRESSIONE LOGISTICA

- Il modello di regressione logistica
- *Maximum Likelihood Estimation*

- Interpretazione dei coefficienti del modello logistico ed *Odds Ratio*
- Regressione logistica *multivariabile*
- Test del *Likelihood-ratio*
- Test di *Hosmer-Lemeshow*

MODULO B

ANALISI DEI DATI AMMINISTRATIVI E OSPEDALIERI

Il corso introduce lo studente alla conoscenza dei principali sistemi di classificazione in ambito sanitario, sia ospedalieri che extra-ospedalieri, ed alla loro gestione e manipolazione con Stata.

SESSIONE I – LA CLASSIFICAZIONE DELLE MALATTIE

- Sistemi di classificazione in ambito sanitario
- Classificazione delle malattie e delle procedure. ICD-9, ICD-9CM, ICD-10

SESSIONE II – ANALISI DI DATI OSPEDALIERI

- Il sistema DRG
- La scheda di dimissione ospedaliera ed il dialogo *grouper*-Stata
- Indici di case-mix. Indici di qualità della codifica
- Rappresentazioni grafiche, tabulazioni e confronti statistici basati su database amministrativi: mappe, diagrammi a barre, distribuzioni di frequenza

SESSIONE III – APPLICARE LE TECNICHE INFERENZIALI AI DATI AMMINISTRATIVI

- Analizzare la durata di degenza e la sopravvivenza dopo interventi chirurgici
- Regressione logistica e *risk adjustment* sui *database* amministrativi
- La durata di degenza preoperatoria e totale e la sopravvivenza intraospedaliera come "*time- to event*": introduzione alla analisi di sopravvivenza

MODULO C

META-ANALISI

Il Corso fornisce ai partecipanti le tecniche fondamentali per portare a termine la maggior parte delle meta-analisi richieste nella ricerca clinica: meta-analisi di studi randomizzati e di studi osservazionali, con outcome categorici e continui, a braccio singolo e con pochi eventi. Vengono inoltre fornite conoscenze di base per lo svolgimento di meta-analisi di test diagnostici, e per il monitoraggio e valutazione critica dei risultati, con analisi del publication bias e meta-regressione. Nonostante la durata molto limitata, grazie a continue esercitazioni pratiche su database esistenti, e ad una selezione rigorosa delle informazioni fornite, il Corso è in grado di trasmettere le conoscenze di statistica sufficienti per poter svolgere le analisi elencate in modo autonomo, dal giorno seguente, con dati reali. Sebbene sia fornita una spiegazione chiara del significato di ciascuna azione, per poter realizzare tale ambizioso obiettivo è tuttavia necessario che i partecipanti possiedano una conoscenza di base delle principali misure e disegni di studio epidemiologici.

SESSIONE I: FONDAMENTI E META-ANALISI DI STUDI RANDOMIZZATI

- I comandi principali per la meta-analisi in Stata (metan, metacum, metafunnel, metabias, metareg, metandi, metandiplot, metamiss): come scaricarli e installarli
- Meta-analisi di studi randomizzati con outcome categorico (risk ratio, odds ratio, hazards ratio, rate ratio, risk difference)
- Eterogeneità, effetti fissi o casuali, analisi di sensibilità e per sotto-gruppi
- Forest plot, funnel plot, e test per la valutazione del publication bias

SESSIONE II: ALTRE TIPOLOGIE DI META-ANALISI (DI STUDI OSSERVAZIONALI O TEST DIAGNOSTICI, A BRACCIO SINGOLO O CON POCHI EVENTI)

- Meta-analisi di studi randomizzati con outcome continui (mean difference e standardized mean difference o effect size)
- Meta-analisi di studi osservazionali con outcome aggiustati tramite analisi multivariate: problematiche nella scelta della misura da estrarre
- Meta-analisi di studi con pochi eventi
- Proportion meta-analysis (meta-analisi a braccio singolo)
- Meta-analisi di test diagnostici: cenni
- Concetto di meta-regression, e limiti metodologici delle meta-analisi
- Problematiche nella combinazione di outcome diversi e scelta del metodo di analisi