LIUC
Università Cattaneo

ECONOMETRICS FOR TIME SERIES

1

Laboratorio di Competizione, mercati e politiche economiche

17 maggio 2017

Fausto Pacicco - E-mail fpacicco@liuc.it

AGENDA

Lezione 1

- Dati
- Statistica univariata e bivariata
- Modelli di regressione
- Diagnostica delle regressioni

Lezione 2

- Trend, stagionalità e stazionarietà
- Modelli AR per serie storiche
- Forecasting
- Assignment

LEZIONE 1

STATISTICA ED ECONOMETRIA

La statistica si occupa di **reperire, analizzare, interpretare, presentare e organizzare dati**.

È alla **base** di qualunque **processo decisionale**, dato la non perfetta conoscenza degli eventi

In **campo economico**, alcune sue applicazioni rientrano sotto la definizione di **econometria**

Durante queste lezioni vedremo alcune nozioni di econometria, per poi applicarle a serie storiche di dati

ANALISI UNIVARIATA

Variabili casuali

Si definiscono variabili casuali le **funzioni** che **associano** ad ogni **evento** uno ed un solo **numero reale**.

Per descrivere in **maniera sintetica** «l'andamento» delle variabili, si può inizialmente ricorrere a:

- **Media**
- **Varianza**
- **Moda**
- **Mediana**

ANALISI UNIVARIATA

Esercizio su statistica univariata - Comandi Eviews

- File→Open→Foreign data as workfile (file Dati.xlsx)
- Predefined Range, foglio UNIVAR→Next→Next→Undated Panel, inserite country_name sotto identifier series→Finish→No*
- Aprite la serie GDP_2011→View→Descriptive statistics & tests→Histogram and stats

Questa finestra contiene le statistiche univariate e l'istogramma della serie; ripetete la procedura per GDP_CAP_2011

* Controllate sempre se le impostazioni di default siano corrette rispetto l'analisi da effettuare

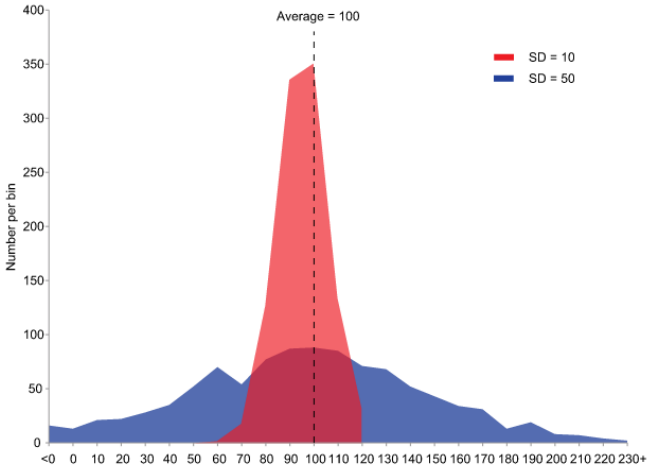
ANALISI UNIVARIATA - LIMITI

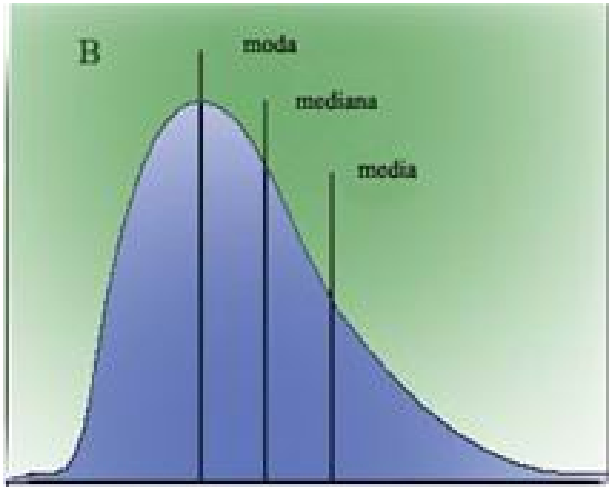
Le statistiche **univariate** (ed anche i grafici) permettono un'analisi dei dati iniziale, ma assolutamente **non definitiva, né esaustiva.**

- La **media** risente degli **outliers** e **non** dà **informazione** circa il resto della **distribuzione**
- Il secondo problema affligge anche la **mediana**
- La **moda** può **non esistere** o essere **non univoca**, e **nascondere** i fenomeni «**rari**» (che però potrebbero avere un forte impatto)

Una **prima miglioria** si ottiene calcolando i **quantili** e le **deviazioni standard**

ANALISI UNIVARIATA





- Formule

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$mediana = X_i + (X_{i+1} - X_i)\frac{0.5 - F_{i-1}}{F_i - F_{i-1}}$$

$$moda = X_i \; con \max freq.\%$$

$$\sigma_X = \sqrt{\frac{\sum_{i=1}^{N}(X_i - \bar{X})^2}{N-1}}$$

ANALISI BIVARIATA

Le analisi **bivariate** permettono di stimare la **dipendenza lineare** tra **2** serie di dati:

- Se A sale e B sale, si ha una relazione **concorde**;
- Se A sale e B scende, si ha una relazione **discorde**;
- Se al variare di A non si ha alcuna variazione di B si ha **un'assenza di relazione lineare**

Per semplicità, valuteremo solo il **coefficiente di correlazione di Pearson**, definito come:

$$\rho_{XY} = \frac{E[(X-\bar{X})(Y-\bar{Y})}{\sigma_X \sigma_Y} \qquad \text{con } -1 \leq \rho_{XY} \leq 1$$

Ovviamente, tale misura è stocastica, per cui deve essere sottoposta ad un test di significatività

ANALISI BIVARIATA

Esercizio su statistica bivariata - Comandi Eviews

- File → Open→ Foreign data as workfile (file Dati.xlsx)
- Predefined Range, foglio BIVAR, Next→Next→Dated – specified by date series, e inserite *year* come Date Series →Finish → No
- Selezionate tutte le serie (tranne c, residuals e year), apritele in gruppo → View → Covariance analysis → selezionate Correlation e Probability |t| = 0

Il primo valore indica il coefficiente di correlazione di Pearson, ed il secondo il p-value del test, con la seguente ipotesi nulla:

H_0: *Correlazione statisticamente non diversa da zero*

Se il **p-value** è **inferiore** alle soglie canoniche (1%, 5% e 10%) **rigettiamo** l'ipotesi con correlazione statiticamente non diversa da zero

ANALISI BIVARIATA

Pur se molto importante come step, nell'analisi bivariata si deve fare attenzione a:

- **Correlazione non implica causalità** - è necessario identificare diversamente la catena causale (teoria, analisi diverse, etc.)
- **Assenza di correlazione lineare non implica l'assenza di correlazione non lineare**
- **Correlazioni spurie** – «l'aumento nel consumo di gelati è correlato con l'aumento dei decessi per annegamento»
- **«By chance»** - negli US si è registrata una correlazione significativa pari a .9979 tra **spesa in tecnologia, scienza, e esplorazione dello spazio** con numero di **suicidi per soffocamento**

Aprite il file *mult_corr.wf1* e ripetete l'analisi della correlazione per tutte le serie

REGRESSIONE LINEARE SEMPLICE

Abbiamo i **voti di un esame** e le **ore di studio** ad esso dedicate, e vogliamo indagare quanto l'antecedente (ore di studio) ha influenzato il conseguente (voto)

In nostro soccorso interviene il **modello di regressione lineare**:

$$voto = \beta_0 + \beta_1 studio + u$$

Il voto è la ***y***, (variabile dipendente), mentre le ore di studio rappresentano la ***x*** (variabile indipendente)

Questo modello ci permette di stabilire l'impatto delle ore di studio sul voto finale, **ceteris paribus**

REGRESSIONE LINEARE SEMPLICE

Esercizio regressione lineare Eviews

- Importate il foglio VOTI del file Dati.xlsx, come unstructured/undated
- Quick → Estimate equation → inserite *voto c ore_studio*

Osserviamo l'output, in particolare:

- **R^2 Adjusted**
- **Significatività dei coefficienti** (singole e congiunta)
- Criteri informativi (**AIC** e **SC**, migliori se più bassi)

TEST DI SIGNIFICATIVITÀ

- L'output contiene i test di **significatività dei coefficienti**, singolarmente e congiuntamente, e i p-values
- Le H0 sono, rispettivamente

Test t → *H_0: Coefficiente statisticamente non diverso da zero*

Test F → *H_0: I coefficienti non sono statisticamente diversi tra di loro e diversi da zero*

- Se il **p-value** è **inferiore** alle soglie canoniche (1%, 5% e 10%) **rigettiamo** le H0
- **Attenzione**, le soglie sono del tipo «**dentro o fuori**»: è un grave errore dire che, ad esempio, un p-value di 0.105 è vicino a 0.1

Dependent Variable: GFR
Method: Least Squares
Date: 05/19/17 Time: 11:04
Sample: 1 72
Included observations: 72

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	103.5600	2.240069	46.23071	0.0000
PILL	-25.94636	4.052438	-6.402656	0.0000

R-squared	0.369335	Mean dependent var	95.63194
Adjusted R-squared	0.360326	S.D. dependent var	19.80464
S.E. of regression	15.83968	Akaike info criterion	8.390298
Sum squared resid	17562.69	Schwarz criterion	8.453539
Log likelihood	-300.0507	Hannan-Quinn criter.	8.415475
F-statistic	40.99400	Durbin-Watson stat	0.102508
Prob(F-statistic)	0.000000		

P-value del test-F di significatività congiunta

P-value del test-t di significatività singola

REGRESSIONE LINEARE MULTIVARIATA

Una costante significativa può ricordarci che ci sono **variabili omesse**

Esercizio Eviews con un modello macroeconomico

- Importate il foglio GDP (da Dati.xlsx) e stimate l'equazione con *gdp c cn i g nx*

Possiamo valutare l'impatto di più X su una Y, ancora con l'assunto di ceteris paribus; i.e. l'effetto stimato di **una** X sulla Y, assume **fissi** gli **altri** regressori

Eviews esegue di **default** una regressione ***Ordinary Least Squares*** (OLS)

Per gli scopi di queste lezioni, limitatevi a modelli multivariati con 4 o 5 regressori (al massimo); modelli più complessi presentano problemi, ivi non trattabili

CHECKS - 1

È necessario svolgere alcuni **controlli addizionali** sulle serie (per eventuali problemi di data collection):

- **Discontinuità** → dovute a cambi di metodologie nel calcolo (e.g. criteri deflazione o valutazione, cambiamento anno base)
- **Effetti di calendario** → dovuti alle diverse lunghezze dei mesi; si risolve «normalizzando» il dato per il numero di giorni lavorativi (per comodità, 5 alla settimana), o prendendo dati trimestrali/quadrimestrali
- **Outliers** → dati con valori non in linea con il resto; aprite il file ouliers.wf1 e selezionate l'equazione E1 → View → Stability Diagnostic → Influence Statistics

Visualizzate grafico e tabella di Hat Matrix, DFFITS, COVRATIO; chiaramente, l'osservazione 10 è un outlier*

Sostituite i valori degli eventuali **outlier** con la **media**, ma solo se il valore non ha una spiegazione (e.g. spiegazione storica)

* Anche altre risultano essere outliers secondo uno o più di questi criteri, ma è meglio essere «prudenti»

DIAGNOSTICA - 1

Un modello di regressione deve essere rispettare gli assunti di base degli OLS. Occupiamoci quindi dell'**analisi dei residui** verificando se i dati presentano

- **Normalità**
- **Omoschedasticità**
- **Assenza di correlazione seriale**

Come «diagnosi **preliminare**» possiamo visualizzare il loro **grafico**:

Comandi Eviews (dopo la regressione precedente, del GDP)

- Proc → Make residual series → Ok → View (serie) → Graph → Ok

Tuttavia, il grafico **non rappresenta** un **valido** strumento da **solo**

DIAGNOSTICA - 2

Dopo la stima di **un'equazione multivariata**, eseguite View → Residual Diagnostic

- **Normalità** (comando *Histogram – Normality Test)*

Oltre al grafico, osserviamo il valore *Probability*: è il **p-value** dell'ipotesi nulla:

H_0: I dati provengono da una distribuzione normale

Se il **p-value** è **inferiore** alla soglia prefissata, **rigettiamo** l'ipotesi della distribuzione normale

In caso di **non-normalità**, i **modelli** sono ***biased*** in campioni di dimensione finita.

Per «risolvere» questo problema, si possono **prendere** le **serie** in **differenze** o i **logaritmi** dei **livelli** (o simili)

DIAGNOSTICA - 3

- **Omoschedasticità** (comando *Heteroskedasticity test → White → Ok)*

I risultati da osservare sono i **tre p-value** degli F-test, che testano la seguente ipotesi nulla:

H_0: Errori omoschedastici, (i.e. errori con stessa e finita varianza)

Se il **p-value** è **inferiore** alla soglia prefissata, **rigettiamo** l'ipotesi di errori omoschedastici

La presenza di **eteroschedasticità** non rende inconsistenti le stime, ma crea **problemi** nella **stima** degli errori standard

DIAGNOSTICA - 4

- **Assenza di correlazione seriale** (comando *Serial correlation LM test*)

L'esito di questo test valuta la seguente ipotesi nulla:

H_0: Assenza di correlazione seriale

Se il **p-value** è **inferiore** alla soglia prefissata, **rigettiamo** l'ipotesi di assenza di correlazione

Poiché ci concentriamo su analisi di **serie storiche**, questa è una delle condizioni che **verrà violata spesso**

DIAGNOSTICA - 5

L'equazione del **GDP** stimata in precedenza, contiene problemi di **eteroschedasticità** e **correlazione**; cosa fare?

- **Ripetiamo** la stessa equazione, ma dalla tab ***Options*** selezioniamo:

Covariance method: HAC (Newey-West) → deselezionate *d.f. Adjustments* → *Hac options* → *Lag specification: None* → *Kernel: Bartlett* → *Bandwidth method: Newey-West Automatic*

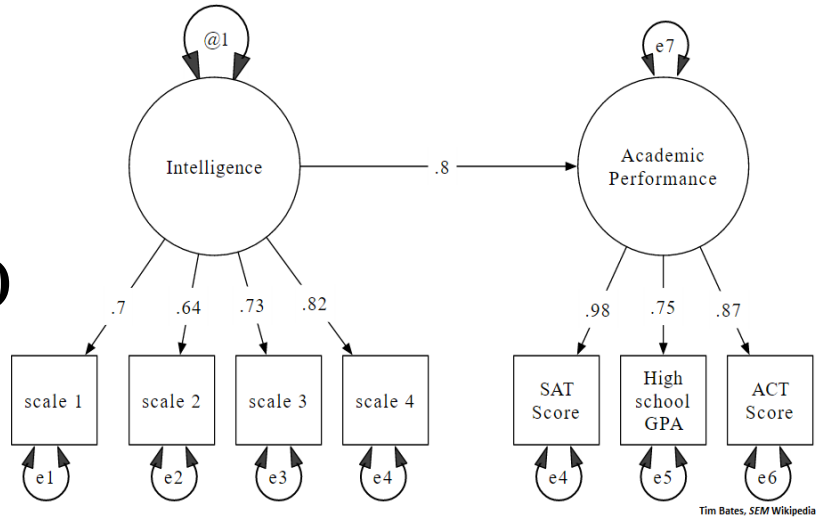
Valutate nell'output l'esito del **Wald statistics** invece del F-test, poiché l'ultimo non è ora affidabile; inoltre, potete vedere che questa metodologia presenta **errori standard diversi**

Nel caso ci sia solo eteroschedasticità, potete limitarvi all'altra opzione, *Huber-White* senza *d.f. adjustment*

SVILUPPI DEI MODELLI DI REGRESSIONE

L'evoluzione dei modelli di regressione ha portato alla creazione di **modelli con molti regressori**, in grado di stimare impatti su costrutti latenti ed effetti simultanei:

Structural Equation Models (SEM)



Pur essendo strumenti **molto potenti**, **richiedono** numerose **restrizioni a priori** e altre condizioni **raramente soddisfatte**; non li analizzeremo qui, ma vedremo alternative più parsimoniose

WORKGROUP

Scegliete un modello economico tra quelli visti nel corso, di cui dovrete affrontare un intero processo di analisi dati (come visto nelle lezioni):

- Trovate un tema di vostro interesse, per il quale abbiate sia riferimenti accademici*, sia disponibilità dei dati
- Effettuate analisi descrittive/esplorative dei dati a vostra disposizione, evidenziando similarità/discontinuità rispetto la teoria
- Dopodiché, stimate dei modelli di regressione lineare multivariata, evidenziandone le diverse capacità esplicative (sia in termini statistici che economici)
- Occupatevi delle procedure di controllo e di pulizia dei dati per le serie (sia per le serie cross, ristimando le regressioni, sia per le serie storiche)
- Dovrete anche inserire la parte spiegata nella prossima lezione

* Generalmente parlando, potete anche utilizzare Wikipedia per ottenere le informazioni iniziali, ma dopo DOVETE verificare se le fonti di Wikipedia siano affidabili, ed in caso negativo supplire

WORKGROUP

Il workgroup dovrà essere condensato in un documento contenente solo gli elementi che ritenete fondamentali al supporto della vostra tesi; eventuali elementi complementari dovranno essere inseriti in un'appendice

Il documento dovrà essere massimo 10 pagine, in italiano, (escludendo copertina, references e appendice), Times New Roman 12, margine superiore 2,5 cm, margini destro, sinistro ed inferiore 2 cm, spaziatura 6 punti, interlinea singola.

I gruppi da formare dovranno essere di minimo 4 e massimo 6 membri. Eventuali eccezioni dovranno essere concordate tramite mail.

Consegna entro il 19/06/2017 (12/06/2017 per i laureandi)

CONSIGLI PER IL WORKGROUP

Il workgroup sarà incentrato su analisi di serie storiche; iniziate ad indagare la disponibilità dei dati circa uno dei modelli economici affrontati nel corso di Competizione .

Ricordatevi sempre di ragionare su un «doppio» binario: le vostre idee devono essere **sia** supportata da una (o più) teoria economica, **sia** dalla disponibilità dei dati in serie storica (minimo 150 osservazioni temporali).

Consiglio di pensare prima a uno/più temi plausibili (da un punto di vista teorico) e poi verificarne la disponibilità dei dati

Qualora abbiate bisogno dei file richiamati nelle slide, potete trovarli nella cartella in condivisione nell'aula computer, come durante la lezione.

FONTI DATI

Per mettere in pratica tutto quello che abbiamo visto, caricate delle serie di dati da siti con dati macroeconomici

http://www.imf.org/en/Data

http://data.worldbank.org

http://www.nber.org/