SERIE STORICHE ECONOMICHE

1

STATISTICA ECONOMICA: serve a sviluppare concetti, definizioni, classificazioni e metodi per PRODURRE INFORMAZIONI STATISTICHE che descrivono lo stato e l'andamento, nel tempo e nello spazio, dei fenomeni economici.
- ANALIZZARE COMPORTAMENTO OPERATORI ECONOMICI
- PREVISIONE SU DINAMICA AGGREGATI ECONOMICI
- PRENDERE DECISIONI DI POLITICA ECONOMICA E AZIENDALE
- VALUTARE PRO E CONTRO DI INVESTIMENTI ALTERNATIVI
- ETC.

SNA (System of National Accounts)
SEC (Sistema Europeo dei Conti)
(SEC deriva da SNA)
} SISTEMI DI CONTI: regole e sistemi di classificazione universalmente riconosciuti
→ utili ad esempio per CONFRONTARE PIL di più paesi.

CONTABILITÀ NAZIONALE: insieme di INFORMAZIONI STATISTICHE attraverso il quale viene ~~[illegible]~~ DESCRITTO IN TERMINI QUANTITATIVI L'ATTIVITÀ ECONOMICA di un paese, regione, ambito territoriale.
→ RIASSUME le OPERAZIONI FATTE DA SOGGETTI ECONOMICI in un SISTEMA INTEGRATO DI EQUAZIONI CONTABILI.

SISTEMA ECONOMICO: SOGGETTI di un TERRITORIO ECONOMICO e MODALITÀ con le quali essi interagiscono con altri soggetti economici.
→ ATTIVITÀ di produzione, distribuzione e utilizzo di beni e servizi.
→ TERRITORIO GEOGRAFICO dove avvengono liberamente gli scambi
→ OPERATORI ECONOMICI (chi fa scelte economiche)

OPERATORI ECONOMICI:
- IMPRESE: PRODUCONO BENI e SERVIZI
- FAMIGLIE: PRESTANO LAVORO per ottenere CAPITALE per ACQUISTARE BENI E SERVIZI prodotti dalle imprese
- PUBBLICA AMMINISTRAZIONE (PA): producono SERVIZI NON DESTINATI ALLA VENDITA, fornisce la REDISTRIBUZIONE della ricchezza.
- RESTO DEL MONDO: tutti gli operatori che NON APPARTENGONO AL SISTEMA ECONOMICO ma hanno RAPPORTI con quelli del paese di riferimento

TRANSAZIONI ECONOMICHE:
- BILATERALI: si realizzano come ~~[illegible]~~ SCAMBIO SUL MERCATO
- UNILATERALI (trasferimenti): NON si realizzano sul mercato ma avvengono SENZA CONTROPARTITA
→ per rappresentarle quantitativamente è necessario FISSARE UN PERIODO DI RIFERIMENTO
→ AGGREGATO: insieme di TRANSAZIONI dello STESSO TIPO, nello STESSO PERIODO CONTABILE (es. aggregato insieme dei consumi)

AGGREGATI ECONOMICI: rappresentano le grandezze economiche che MISURANO IL RISULTATO D'INSIEME delle operazioni ~~[illegible]~~ svolte da tutte le unità economiche del sistema economico.

CONTABILITÀ NAZIONALE DESCRIVE:
- → FLUSSI
- → STOCK (inizio e fine periodi contabili)

SEC:
- criteri di VALUTAZIONE degli AGGREGATI
- DELIMITAZIONE dell'economia nazionale
- CLASSIFICAZIONI delle unità produttive in BRANCHE di attività economica
- CLASSIFICAZIONE delle OPERAZIONI e degli AGGREGATI

DELIMITAZIONE DELL'ECONOMIA NAZIONALE: operazioni rilevanti per la contabilità nazionale sono quelle compiute in un dato TERRITORIO ECONOMICO dagli operatori residenti RESIDENTI
→ operatori che esercitano l'attività per ALMENO 1 ANNO (esclusi quindi i TURISTI)

CLASSIFICAZIONE OPERATORI (UNITA' ISTITUZIONALI) → per SETTORI ISTITUZIONALI / per BRANCHE

UNITA' ISTITUZIONALE: centro elementare di DECISIONE ECONOMICA AUTONOMA

SETTORI ISTITUZIONALI: classificazione sulla base della FUNZIONE PRINCIPALE ESERCITATA e del TIPO DI RISORSE prevalentemente impiegate.

Il SEC raggruppa le UNITA' ISTITUZIONALI in 5 SETTORI ISTITUZIONALI residenti:

- SOCIETA' NON FINANZIARIE (+ di 5 addetti)
- SOCIETA' FINANZIARIE (Es. Banche)
- AMMINISTRAZIONI PUBBLICHE (comuni, ...)
- FAMIGLIE (5 addetti o meno)
- ~~...~~ ISTITUZIONI SENZA SCOPO DI LUCRO AL SERVIZIO DELLE FAMIGLIE
- RESTO DEL MONDO

BRANCA ECONOMICA: insieme di unità produttive con PRODUZIONE OMOGENEA ristretta ad un prodotto o a un gruppo di prodotti.

UNITA' PRODUTTIVE: vengono classificate in BRANCHE di attività economica

STATISTICA UFFICIALE: statistiche su cui si basano le scelte della Nazione

ATECO (ATtività ECOnomiche): raggruppa attività economiche che producono beni simili.

ISIC (mondiale) → NACE (europea) → ATECO (nazionale)

SEZIONI > DIVISIONI > GRUPPI > CLASSI > CATEGORIE > SOTTOCATEGORIE

Attenzione all'anno di riferimento ATECO 2002 ≠ ATECO 2007

→ Es. $\underbrace{C}_{\text{SEZIONE}}\ \underbrace{25}_{\text{DIVISIONE}}.\underbrace{9}_{\text{GRUPPO}}\underbrace{9}_{\text{CLASSE}}.\underbrace{11}_{\text{CATEGORIA}}$ Fabbricazione di candele e bottiglie isolate in metallo

AGGREGATI: FLUSSO AGGREGATO DI TRANSAZIONI di un certo tipo in un determinato PERIODO CONTABILE
(Es. PRODUZIONE, CONSUMO, INVESTIMENTI, ETC.)

AGGREGATI → INTERNI: relativi a flussi nel territorio economico che non per forza costituiscono la residenza degli operatori.
AGGREGATI → NAZIONALI: relativi a flussi di operatori RESIDENTI nel territorio economico del paese.

AGGREGATI ECONOMICI: valori compositi che MISURANO IL RISULTATO DELL'ATTIVITA' NEL COMPLESSO DELL'ECONOMIA.

→ aggregati che si riferiscono direttamente alle OPERAZIONI NEL SISTEMA SEC2010: es. $\boxed{\text{PRODUZIONE BENI E SERVIZI}}$
→ aggregati che rappresentano SALDI CONTABILI: es. $\boxed{\text{PIL}}$

PRODUZIONE: è costituita dal TOTALE DEI PRODOTTI risultanti dall'ATTIVITA' DI PRODUZIONE nel corso del PERIODO CONTABILE

→ deriva dall'economia REGOLARE, ILLEGALE (droga, contrabbando), INFORMALE (attività senza addetti/occulti), SOMMERSA (attività regolari non dichiarate (evasione fiscale))

CONSUMI INTERMEDI: costituiti da beni e servizi consumati quali input di un processo di produzione

CONSUMI FINALI:
- SPESA PER CONSUMI FINALI: ~~[illegible]~~ SPESA SOSTENUTA per beni e servizi.
- CONSUMI FINALI EFFETTIVI: beni e servizi ACQUISITI

PRODUZIONE di un'impresa è data dal VALORE DEI BENI E SERVIZI PRODOTTI in un dato INTERVALLO DI TEMPO che sono stati VENDUTI o che avrebbero potuto esserlo.

$$P = V + r + g_p + g_c + K_i$$

- K_i → PARTE CHE VA AD INCREMENTARE GLI INVESTIMENTI
- g_c → GIACENZA MAGAZZINI PRODOTTI NON FINITI
- g_p → PRODUZIONE INVENDUTA (MAGAZZINO/SCORTE)
- r → PRODUZIONE REIMPIEGATA
- V → PRODUZIONE VENDUTA
- P → PRODUZIONE IMPRESA

VALORE AGGIUNTO: rappresenta la remunerazione dei FATTORI PRODUTTIVI PRINCIPALI ed è dato da:

$$Y = P - C_x$$ (VALORE AGGIUNTO = PRODUZIONE − CONSUMO INTERMEDIO)

VALUTAZIONE AGGREGATI:
- PREZZI DI MERCATO (prezzo di acquisto)
- PREZZI BASE (prezzi netti incassati dal produttore, TASSE ESCLUSE, CONTRIBUTI INCLUSI)
- COSTO DEI FATTORI (prezzi netti incassati dal produttore, TASSE ESCLUSE, CONTRIBUTI ESCLUSI)

PIL: misura della PRODUZIONE FINALE del paese

$$Y_p = C + I + I_s + (E - M)$$

Y_p → PIL; C → CONSUMI FINALI; I → INVESTIMENTI FISSI; I_s → VARIAZIONI DELLE SCORTE; E → ESPORTAZIONI; M → IMPORTAZIONI

NUMERI INDICE

NUMERO INDICE: rapporto fra numeri ${}_0I_t$ ← PERIODO DI RIFERIMENTO (0 ↑ BASE) (ELIMINO L'ORDINE DI GRANDEZZA DEL FENOMENO E L'UNITÀ DI MISURA)

DIFFERENZA ASSOLUTA: $X_1 - X_0$

DIFFERENZA RELATIVA: $\frac{X_1 - X_0}{X_0} = \frac{X_1}{X_0} - 1$ ← N° INDICE

NUMERI INDICE:
- → A BASE FISSA (divido per uno stesso valore fisso) ${}_0I_t = \frac{X_t}{X_0}$ $t = 0, 1, 2, \dots, T$
- → A BASE MOBILE (divido il valore per il valore del periodo precedente) ${}_{t-1}I_t = \frac{X_t}{X_{t-1}}$

VARIAZIONE PERCENTUALE = (n° indice × 100) − 100

PASSAGGIO DA UNA BASE ALL'ALTRA:
- DA FISSA A MOBILE [DIVIDO]: $\frac{{}_0I_t}{{}_0I_{t-1}} = \frac{\frac{X_t}{\cancel{X_0}}}{\frac{X_{t-1}}{\cancel{X_0}}} = \frac{X_t}{X_{t-1}} = {}_{t-1}I_t$
- DA MOBILE A FISSA [MOLTIPLICO]: ${}_0I_1 \cdot {}_1I_2 \cdot {}_2I_3 \cdots {}_{t-1}I_t = \frac{\cancel{X_1}}{X_0} \cdot \frac{\cancel{X_2}}{\cancel{X_1}} \cdot \frac{\cancel{X_3}}{\cancel{X_2}} \cdots \frac{X_t}{\cancel{X_{t-1}}} = {}_0I_t$

PROPRIETÀ DEI N.I. ELEMENTARI:
- IDENTITÀ: ${}_tI_t = 1$ $\frac{\cancel{X_t}}{\cancel{X_t}} = 1$
- REVERSIBILITÀ DELLE BASI: ${}_rI_t \cdot {}_tI_r = 1$ $\frac{\cancel{X_t}}{\cancel{X_r}} \cdot \frac{\cancel{X_r}}{\cancel{X_t}} = 1$
- IMPORTANTE × CAMBIO DI BASE → CIRCOLARITÀ O TRANSITIVITÀ: ${}_rI_s \cdot {}_sI_t = {}_rI_t$ $\frac{\cancel{X_s}}{X_r} \cdot \frac{X_t}{\cancel{X_s}} = \frac{X_t}{X_r}$
- UTILE × SCOMPORRE LE VARIAZIONI (POSSO ELIMINARE L'EFFETTO PREZZO) → DECOMPOSIZIONE DELLE CAUSE: ${}^V_0I_t = {}^P_0I_t \cdot {}^Q_0I_t$ VALORE = PREZZO × QUANTITÀ
- EXTRA: COMMENSURABILITÀ — Il N.I. non varia se cambia unità di misura

→ PIM ${}^V_0I_t = \frac{V_t}{V_0} = \frac{p_t \cdot q_t}{p_0 \cdot q_0} = {}^P_0I_t \cdot {}^Q_0I_t$

MEDIE N.I. ELEMENTARI:

	SEMPLICI	PONDERATE
• M. ARITMETICA	${}_0I_t(M_1) = \dfrac{\sum_{k=1}^{K} \frac{{}_kX_t}{{}_kX_0}}{K}$	${}_0I_t(M_1, {}_kW) = \dfrac{\sum_{k=1}^{K} \frac{{}_kX_t}{{}_kX_0} \cdot {}_kW}{\sum_{k=1}^{K} {}_kW}$
• M. GEOMETRICA	${}_0I_t(M_0) = \sqrt[K]{\prod_{k=1}^{K} \frac{{}_kX_t}{{}_kX_0}}$	${}_0I_t(M_0, {}_kW) = \sqrt[\sum_{k=1}^{K} {}_kW]{\prod_{k=1}^{K} \left(\frac{{}_kX_t}{{}_kX_0}\right)^{{}_kW}}$
• M. ARMONICA	${}_0I_t(M_{-1}) = \dfrac{K}{\sum_{k=1}^{K} \frac{{}_kX_0}{{}_kX_t}}$	${}_0I_t(M_{-1}, {}_kW) = \dfrac{\sum_{k=1}^{K} {}_kW}{\sum_{k=1}^{K} \frac{{}_kX_0}{{}_kX_t} \cdot {}_kW}$

NUMERI INDICI SINTETICI $K \geq 2$ SERIE TEMPORALI ${}_kX_t$ $t = 0, 1, \dots, T$

$k = 1, 2, \dots, K$ ← BENE

Metodi per costruirli:

a) RAPPORTO TRA SOMME ${}_0I_t = \dfrac{\sum_{k=1}^{K} {}_kX_t}{\sum_{k=1}^{K} {}_kX_0}$

b) MEDIE SEMPLICI DI INDICI ELEMENTARI (LIMITATIVA PERCHÉ DÀ LO STESSO PESO AI VARI FENOMENI)

c) MEDIE PONDERATE DI INDICI ELEMENTARI (PESI DEVONO ESSERE DI TIPO ADDITIVO, devono potersi sommare)

NUMERI INDICE SINTETICI DI GRANDEZZE ECONOMICHE:

$$
{}_kX_t \begin{cases} \to {}_kp_t & \text{PREZZI} \\ \to {}_kq_t & \text{QUANTITÀ} \\ \to {}_kv_t & \text{VALORE (PREZZO × QUANTITÀ)} \end{cases}
$$

NUMERI INDICI SINTETICI DEI PREZZI

- PROBLEMI PRELIMINARI:
  - SCELTA NUMERO E SPECIE DI BENI (K)
  - SCELTA DELLA BASE (X_0)
- SCELTA DEL TIPO DI MEDIA
- SCELTA DEL SISTEMA DI PONDERAZIONE:
  - QUANTITÀ
  - VALORE (QUESTA È PIÙ UTILIZZATA)

PONDERAZIONE MEDIANTE VALORI: VALORE = QUANTITÀ × PREZZO

${}_kp_0 \cdot {}_kq_0$ LASPEYRES (PESI ALL'ANNO BASE)

${}_kp_t \cdot {}_kq_t$ FITTIZIA

${}_kp_0 \cdot {}_kq_t$ PAASCHE (VARIO LE QUANTITÀ)

${}_kp_t \cdot {}_kq_0$ FITTIZIA

LASPEYRES:
$$
{}_0^pI_t^L = \frac{\sum_{k=1}^{K} {}_0^kI_t \, {}_kV_0}{\sum_{k=1}^{K} {}_kV_0} = \frac{\sum_{k=1}^{K} \frac{{}_kp_t}{\cancel{{}_kp_0}} \cdot \overbrace{\cancel{{}_kp_0} \cdot {}_kQ_0}^{\text{PESO}}}{\sum_{k=1}^{K} {}_kp_0 \cdot {}_kQ_0} = \frac{\sum_{k=1}^{K} {}_kp_t \cdot {}_kQ_0}{\sum_{k=1}^{K} {}_kp_0 \cdot {}_kQ_0} \quad \text{(LASPEYRES × PREZZI)}
$$

$$
{}_0^qI_t^L = \frac{\sum_{k=1}^{K} {}_kQ_t \cdot {}_kp_0}{\sum_{k=1}^{K} {}_kp_0 \cdot {}_kQ_0} \quad \text{(LASPEYRES × QUANTITÀ)}
$$

Problemi: Laspeyres tende a SOPRAVVALUTARE la variazione dei prezzi (infatti osservando il numeratore all'aumentare del prezzo la quantità dovrebbe scendere invece per l'indice resta FISSA).

Problema del LOGORAMENTO DELLA BASE (base vecchia non affidabile per descrivere il fenomeno)

PAASCHE: $${}^p_0I^P_t = \frac{\sum_{k=1}^{K} \frac{{}_kP_t}{{}_kP_0} \cdot \overbrace{{}_kP_0 \; {}_kQ_t}^{\text{PESO}}}{\sum_{k=1}^{K} {}_kP_0 \; {}_kQ_t} = \frac{\sum_{k=1}^{K} {}_kP_t \cdot {}_kQ_t}{\sum_{k=1}^{K} {}_kP_0 \cdot {}_kQ_t}$$ (PAASCHE x PREZZI)

$${}^q_0I^P_t = \frac{\sum_{k=1}^{K} {}_kP_t \cdot {}_kQ_t}{\sum_{k=1}^{K} {}_kP_t \cdot {}_kQ_0}$$ (PAASCHE x QUANTITÀ)

Problemi: Per l'indice di Paasche SERVONO PIÙ INFORMAZIONI.
Tende a SOTTOVALUTARE le variazioni dei prezzi.

CONSIDERAZIONI:

$${}^p_0I^L_t \neq {}^p_0I^P_t \longrightarrow {}^p_0I^P_t = {}^p_0I^L_t + \underbrace{\overset{-}{r} \cdot \frac{\overset{+}{\sigma_p} \, \overset{+}{\sigma_q}}{{}^q_0I^L_t \; +}}_{-} \longrightarrow \boxed{{}^p_0I^P_t > {}^p_0I^L_t}$$

PAASCHE > LASPEYRES

DIMOSTRAZIONE SU DISPENSE

INDICE DI FISHER: Media geometrica LASPEYRES e PAASCHE $${}^p_0I^F_t = \sqrt{{}^p_0I^L_t \cdot {}^p_0I^P_t}$$

PROPRIETÀ (DIM SU DISPENSA)

	LASPEYRES	PAASCHE	FISHER	
IDENTITÀ	SI	SI	SI	
REVERSIBILITÀ DELLE BASI	NO	NO	SI	
CIRCOLARITÀ	NO	NO	NO	← PROBLEMATICO (COEFFICIENTI DI RACCORDO ISTAT)
DECOMPOSIZIONE DELLE CAUSE	NO	NO	SI	← PROBLEMATICO (VALE IN SENSO DEBOLE)
COMMENSURABILITÀ	SI	SI	SI	
PROPORZIONALITÀ	SI	SI	SI	
DETERMINATEZZA	SI	SI	SI	

- PROPORZIONALITÀ: se quantità e prezzi variano di una proporzione alla quale il n.i. varia nella stessa proporzione
- DETERMINATEZZA: indice non deve annullarsi, tendere ad infinito se si annulla o tende ad infinito un termine della formula.

AGGREGATI REALI $\sum_{k=1}^{K} {}_kP_t \, {}_kQ_t$ $\quad \sum_{k=1}^{K} {}_kP_0 \, {}_kQ_0$

AGGREGATI FITTIZI $\sum_{k=1}^{K} {}_kP_t \, {}_kQ_0$ $\quad \sum_{k=1}^{K} {}_kP_0 \, {}_kQ_t$

SI COMBINANO PER OTTENERE:
- INDICI DI PREZZO
- INDICI DI QUANTITÀ
- INDICI DI VALORE

LASPEYRES MODIFICATO:

$$\frac{{}^p_0I^L_t}{{}^p_0I^L_{t-1}} = \frac{\frac{\sum {}_kP_t \cdot {}_kQ_0}{\cancel{\sum {}_kP_0 \cdot {}_kQ_0}}}{\frac{\sum {}_kP_{t-1} \cdot {}_kQ_0}{\cancel{\sum {}_kP_0 \cdot {}_kQ_0}}} = \frac{\sum_k {}_kP_t \cdot {}_kQ_0}{\sum_k {}_kP_{t-1} \cdot {}_kQ_0}$$

UTILE PER VEDERE LA VARIAZIONE DA $t-1$ A t (PAASCHE NON GODE DI QUESTA PROPRIETÀ)

NUMERI INDICI A CATENA: per risolvere problemi LOGORAMENTO DELLA BASE di LASPEYRES

$${}^C_0I^*_t = \prod_{s=1}^{t} {}_{s-1}I_s = {}_0I_1 \cdot {}_1I_2 \cdot {}_2I_3 \cdot \ldots \cdot {}_{t-1}I_t$$

(N.I. SEMPLICI TRANSITIVITÀ RISPETTATA)

Esempio con Laspeyres:

$${}^C_0C^L_2 = \underbrace{\frac{\sum {}_kP_1 \cdot {}_kQ_0}{\sum {}_kP_0 \cdot {}_kQ_0}}_{{}^p_0I^L_1} \cdot \underbrace{\frac{\sum {}_kP_2 \cdot {}_kQ_1}{\sum {}_kP_1 \cdot {}_kQ_1}}_{{}^p_1I^L_2}$$

(N.I. SINTETICI TRANSITIVITÀ NON RISPETTATA → USO LASPEYRES PERCHÉ PIÙ SEMPLICE DA GESTIRE)

I PRINCIPALI N.I. PRODOTTI DALL'ISTAT

NUMERI INDICI
- DEI PREZZI
  - PREZZI AL CONSUMO:
    - NIC (PER L'INTERA COLLETTIVITÀ NAZIONALE)
    - FOI (PER LE FAMIGLIE DI OPERAI E IMPIEGATI)
    - IPCA (SIMILE AL NIC MA COSTRUITO PER CONFRONTI CON N.I. UE)
  - PREZZI ALLA PRODUZIONE:
    - prodotti industriali
    - dei fabbricati residenziali
    - prodotti agricoli
- DELLE QUANTITÀ
  - produzione agricola e forestale
  - produzione industriale
- DEI VALORI
  - del fatturato e degli ordinativi dell'industria
  - vendite al dettaglio
  - retribuzioni contrattuali
  - dei compensi al fattore lavoro nelle grandi imprese

N.I. DEI PREZZI A CONSUMO tengono conto solo delle FAMIGLIE RESIDENTI e dei beni ACQUISTABILI ATTRAVERSO RISORSE MONETARIE

NIC. utilizzato come misura dell'INFLAZIONE al livello dell'intero sistema economico

FOI. utilizzato per adeguare periodicamente i VALORI MONETARI (es. affitti, ...) (erroneamente INDICE DEL COSTO DELLA VITA)

IPCA: indice armonizzato utilizzato per assicurare una misura dell'INFLAZIONE COMPARABILE a LIVELLO EUROPEO

(Questi tre indici vengono elaborati dall'Istat MENSILMENTE, dati raccolti entro FINE MESE, pubblicati ENTRO 15, dati raccolti CAPOLUOGHI PROV. SCELTI)

DIVERSITÀ TRA I 3 INDICI:

- CONCETTO DI PREZZO
  - NIC, FOI (PREZZO PIENO)
  - IPCA (PREZZO PAGATO DAL CONSUMATORE)
- POPOLAZIONE DI RIFERIMENTO
  - NIC, IPCA (INTERA POPOLAZIONE)
  - FOI (FAMIGLIE OPERAI IMPIEGATI)
- SISTEMI DI PONDERAZIONE (PESI PROPORZIONALI AI CONSUMI DELLE RISPETTIVE POPOLAZIONI DI RIFERIMENTO) (≈ PANIERI ≈ ?)

CONCATENAMENTO:
- BASE DI CALCOLO (cambia OGNI ANNO)
- BASE DI RIFERIMENTO (cambia OGNI 5 ANNI)

(concatenare gli indici)

$${}_{2015}I_{2017,\,gennaio} = {}_{2016}I_{2017,\,gennaio} \cdot {}_{2015}I_{2016,\,dicembre}$$

RIFERIMENTO AL 2015 (BASE DI RIFERIMENTO)

VARIAZIONE CONGIUNTURALE: $\frac{\text{MESE}}{\text{MESE PRECEDENTE}}$ $\left(\text{es. } \frac{\text{GEN-17}}{\text{DIC-16}}\right)$

VARIAZIONE TENDENZIALE: $\frac{\text{MESE ANNO}}{\text{STESSO MESE - ANNO PRECEDENTE}}$ $\left(\text{es. } \frac{\text{GEN-17}}{\text{GEN-16}}\right)$

COSTRUZIONE NIC:
1) RILEVAZIONE PREZZI
2) CALCOLO INDICI ELEMENTARI (media geometrica indici semplici)
3) CALCOLO INDICI ELEMENTARI REGIONALI (media ponderata con PESI la POPOLAZIONE RESIDENTE)
4) CALCOLO INDICI ELEMENTARI NAZIONALI (media ponderata con PESI i CONSUMI DELLE DIVERSE AREE)
5) CALCOLO INDICI NAZIONALI (media ponderata con PESI i CONSUMI di CIASCUN BENE)

LA MISURA DELL'INFLAZIONE

INFLAZIONE: processo generalizzato di aumento dei prezzi

TASSO TENDENZIALE DI INFLAZIONE: $\dfrac{{}_0I_{m,t}}{{}_0I_{m,t-1}} - 1$

TASSO DI INFLAZIONE MEDIA: $\dfrac{M_{12,t}}{M_{12,t-1}} - 1$

POSSO CONFRONTARE I DUE TASSI PER CAPIRE SE SONO IN UN PERIODO DI INFLAZIONE CRESCENTE/DECRESCENTE [N.B TUTTI GLI INDICI CALCOLATI SULLA STESSA BASE]

$\hookrightarrow$ ES. $\dfrac{{}_0I_{12,t}}{{}_0I_{12,t-1}} > \dfrac{M_{12,t}}{M_{12,t-1}} \Longrightarrow$ INFLAZIONE CRESCENTE

$$\frac{M_{12,t}}{M_{12,t-1}} = \underbrace{\frac{{}_0I_{12,t-1}}{{}_0M_{12,t-1}}}_{\text{INFL. EREDITATA DALL'ANNO } t} \cdot \underbrace{\frac{M_{12,t}}{{}_0I_{12,t-1}}}_{\text{INFL. PROPRIA DELL'ANNO } t} \qquad \frac{{}_0I_{12,t}}{{}_0I_{12,t-1}} = \frac{M_{12,t}}{{}_0I_{12,t-1}} \cdot \underbrace{\frac{{}_0I_{12,t}}{M_{12,t}}}_{\text{INFL. LASCIATA IN EREDITÀ DALL'ANNO } t}$$

CONFRONTO DEGLI AGGREGATI NEL TEMPO

Per volumi bisogna depurare dall'effetto prezzo ...
- FLUSSI DI BENI E SERVIZI: flussi monetari dati da QUANTITÀ × PREZZO
- FLUSSI MONETARI PURI: flussi già espressi come quantità monetarie

- **METODO ANALITICO (DIRETTO)** (torno all'ANNO BASE) ${}_tX_t = \sum_{k=1}^{K} {}_kp_t\,{}_kq_t \Longrightarrow {}_0X_t = \sum_{k=1}^{K} {}_kp_0 \cdot {}_kq_t$
- **DEFLAZIONE (METODO INDIRETTO)** (divido per un indice dei prezzi PAASCHE) ${}_0X_t = \dfrac{{}_tX_t}{{}^P_0I^P_t}$ ($\uparrow$ PAASCHE) $\left(\sum_k {}_kp_t \cdot {}_kq_t \cdot \dfrac{\sum_k {}_kp_0 \cdot {}_kq_t}{\sum_k {}_kp_t \cdot {}_kq_t} = \sum_k {}_kp_0 \cdot {}_kq_t\right)$ ($\uparrow$ RECIPROCO PAASCHE)
- **ESTRAPOLAZIONE (METODO INDIRETTO)** (moltiplico per indice quantità LASPEYRES) ${}_0X_t = {}_0X_0 \cdot {}^Q_0I^L_t$ ($\uparrow$ QUANTITÀ LASPEYRES) $\left(\sum_k {}_kp_0\,{}_kq_0 \cdot \dfrac{\sum_k {}_kp_0\,{}_kq_t}{\sum_k {}_kp_0 \cdot {}_kq_0} = \sum_k {}_kp_0 \cdot {}_kq_t\right)$ ($\uparrow$ LASPEYRES)

DEFLATORE IMPLICITO < IND. PREZZI PAASCHE / IND. QUANTITÀ LASPEYRES } li ottengo mettendo a RAPPORTO AGGREGATI

SCOMPOSIZIONE INDICE DI VALORE: ${}^V_{t-1}I_t = \underbrace{{}^P_{t-1}I^P_t}_{\text{PAASCHE}} \cdot \underbrace{{}^Q_{t-1}I^L_t}_{\text{LASPEYRES}}$

INDICE CONCATENATO DELLE VARIAZIONI IN VOLUME ${}^{C(Q)}_0I^L_t = \dfrac{{}^C_0X_t}{{}_0X_0} \rightarrow {}^C_0X_t = {}_0X_0 \cdot {}^{C(Q)}_0I^L_t$

INDICE CONCATENATO DEI PREZZI IMPLICITI ${}^{C(P)}_0I^P_t = \dfrac{{}_tX_t}{{}^C_0X_t} \rightarrow \dfrac{{}_tX_t}{{}^C_0X_t} = {}^{C(P)}_0I^P_t$

INDICE CONCATENATO DEI VALORI ${}^V_0I_t = {}^{C(P)}_0I^P_t \cdot {}^{C(Q)}_0I^L_t$

ANALISI SERIE STORICHE

SERIE STORICA: successione di osservazioni ORDINATE secondo il TEMPO

↳ serve a PREVEDERE L'ANDAMENTO (problema: PREVEDERE LE REGOLARITA')

SERIE CONTINUA: es. serie incidenti automobilistici, ha senso UNIRE I PUNTI, incidenti avvengono lungo tutto il periodo annuale.

INTERRUZIONE SERIE: salti dovuti a mancanza delle misure.

ANDAMENTI CICLICI: (fase VARIABILE) ≠

MOVIMENTI STAGIONALI: (fase COSTANTE solitamente inferiore ai 12 mesi)

TREND DI FONDO

SERIE STORICHE:
- VARIABILI DI STOCK (istanti temporali definiti)
- VARIABILI DI FLUSSO (aggregato valori intervallo temporale) CUMULO

SERIE STORICA:
- UNIVARIATA (osservo 1 fenomeno)
- MULTIVARIATA (osservo più fenomeni)

TEMPO: parametro DISCRETO
- INTERVALLI EQUISPAZIATI
- ISTANTI EQUISPAZIATI

SERIE STORICHE:
- DETERMINISTICHE (previste sulla base della propria storia passata)
- STOCASTICHE (determinate solo in parte dal passato e le previsioni NON SONO PRIVE DI ERRORE)

OBIETTIVI:
- DESCRIZIONE (osservo la serie per trovare dati anomali)
- SPIEGAZIONE
- PREVISIONE
- FILTRAGGIO (estraggo qualcosa dalla serie che non è immediatamente visibile, es. TREND)
- CONTROLLO

APPROCCIO:
- CLASSICO: serie scomponibile in 4 fattori: T C S E (TREND, CICLO, STAGIONALITA', ERRORE)
- MODERNO: fenomeni generati da PROCESSO STOCASTICO, studio il meccanismo che li genera per fare inferenza.

TREND: movimento di LUNGO PERIODO, tendenza di fondo del fenomeno considerato

CICLO: fluttuazioni attribuibili al succedersi nel fenomeno considerato di fasi crescenti e di fasi decrescenti.
NON HA DURATA FISSA!

STAGIONALITA': movimenti del fenomeno nel corso dell'anno che si RIPETONO NEL MEDESIMO PERIODO.

ERRORE (COMPONENTE ACCIDENTALE): tiene conto del comportamento non perfettamente prevedibile degli agenti economici.

$$\varepsilon_t \begin{cases} E(\varepsilon_t)=0 \\ Var(\varepsilon_t)=\sigma \text{ COSTANTE} \\ Cov(\varepsilon_t, \varepsilon_{t+\tau})=0 \end{cases}$$

+ NORMALITA' (condizione + debole)

$$Y_t = \underbrace{f(t)}_{\text{PARTE DETERMINISTICA}} + \underbrace{U_t}_{\text{PARTE STOCASTICA}}$$

← approccio moderno studia questa parte

↑ approccio classico si concentra a studiare questa parte

PUNTO DI VISTA DECOMPOSITIVO: voglio decomporre la serie nelle sue 4 componenti T, C, S, E

- TREND: costante come $\beta_0+\beta_1 t$ (retta) o curva purché abbia ANDAMENTO DOLCE
- CICLO: insieme delle variabili con ciclicità costante (o meno)
- ERRORE

MODELLI

- ADDITIVO $Y_t = T_t + C_t + S_t + \varepsilon_t$ COMPONENTI INDIPENDENTI
- MOLTIPLICATIVO $Y_t = T_t \cdot C_t \cdot S_t \cdot \varepsilon_t$ COMPONENTI NON INDIPENDENTI (componenti evolvono legate alle altre componenti)
- MISTO $Y_t = T_t \cdot C_t \cdot S_t + \varepsilon_t$ (o altre combinazioni)

Modello SPESSO UTILIZZATO: $Y_t = (T_t \cdot C_t) + S_t + \varepsilon_t$

↓ SOPRATTUTTO IN CASO DI SERIE STORICHE NON MOLTO LUNGHE

→ TREND E CICLO ACCORPATI (lentamente oscillante)

Anche classe abbastanza RIGIDA: presenta SOLO I MODELLI SOVRACITATI.

TRASFORMAZIONE: MOLTIPLICATIVO $\xrightarrow{\text{LOGARITMO}}$ ADDITIVO [$\log Y_t = \log T_t + \log C_t + \log S_t + \log \varepsilon_t$]

AMPIEZZA AUMENTA NEL TEMPO:  ⇒ MODELLO MOLTIPLICATIVO

STAZIONARIETÀ IN MEDIA:  ⇒ MODELLO NON HA TREND

OPERAZIONI PRELIMINARI:

- VAGLIO CONTINUITÀ DELLA SERIE: la serie deve essere CONTINUA, DATI DI TIPO OMOGENEO, STESSA BASE
- VAGLIO DI LUNGHEZZA DELLA SERIE: voglio avere un NUMERO SIGNIFICATIVO DI DATI, specie per eliminare stagionalità
- DEPURAZIONE DELLE VARIAZIONI DI CALENDARIO: voglio pulire: MESI DELLA STESSA DURATA per poterli confrontare
- TRASFORMAZIONI DEI DATI:
  - NUMERI INDICE
  - DIFFERENZE
  - TRASFORMAZIONI MATEMATICHE (reciproci, radici, logaritmi)

Dopo aver svolto le operazioni preliminari posso continuare a lavorare con i tre passi della DECOMPOSIZIONE (T/C/S/E)

DETERMINAZIONE DEL TREND

Per semplificare lavoro con $Y_t = T_t + \varepsilon_t$ $\varepsilon_t \approx WN(0, \sigma^2_\varepsilon)$

- METODO GRAFICO  Determino retta tale che le aree si bilancino.
- METODO ANALITICO: Determino la funzione che approssima la tendenza di fondo
- METODO DELLE MEDIE MOBILI: uso medie al posto dei dati per stimare il trend di fondo.

METODO ANALITICO

- TREND LINEARE O LINEARIZZABILE NEI PARAMETRI
  - TREND POLINOMIALE (GRADO q) $f(t) = a_0 + a_1 t + a_2 t^2 + \dots + a_q t^q$
  - TREND ESPONENZIALE $f(t) = a_0 e^{a_1 t}$ $a_0 > 0$
  - ATTRAVERSO METODO DEI MINIMI QUADRATI
- TREND NON LINEARE NEI PARAMETRI: CURVE DI CRESCITA
  - CURVA ESPONENZIALE MODIFICATA
  - CURVA LOGISTICA
  - CURVA GOMPERTZ

TREND POLINOMIALE $f(t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \dots + \alpha_q t^q \xrightarrow[\text{LINEARE}]{\text{MODELLO REGRESSIONE}} Y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \dots + \alpha_q t^q + \varepsilon_t$

↳ COME SCEGLIERE q? < CRITERIO DIFFERENZE SUCCESSIVE / R^2 CORRETTO

- **R^2 CORRETTO**: provo tanti q e vedo qual è il modello migliore scegliendo quello con R^2 corretto più alto

- **CR. DIFF. SUCCESSIVE**:

OPERATORE RITARDO $\boxed{B^h Y_t = Y_{t-h}}$

OPERATORE DIFFERENZA $\boxed{(1-B)\, Y_t = Y_t - B Y_t}$

Per scegliere q differenzio finchè la differenza osservata non è nulla:

$$\begin{cases} (1-B)^q f(t) = \text{COSTANTE} & \longleftarrow \text{scelgo } q \\ (1-B)^r f(t) = 0 & \text{con } r = q+1 \end{cases}$$

Esempi: RETTA $f(t) = \alpha_0 + \alpha_1 t \quad (1-B)^2 f(t) = 0 \quad q = 1$

PARABOLA $f(t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2 \quad (1-B)^3 f(t) = 0 \quad q = 2$

ATTENZIONE PERCHE' ABBIAMO ANCHE UNA PARTE ALEATORIA: $Y_t = f(t) + \varepsilon_t$

$$E(Y_t) = f(t) + \underbrace{E(\varepsilon_t)}_{=0 \times \text{ipotesi}} \qquad \text{se differenzio } E(\varepsilon_t) \text{ se ne va!}$$

PIU' DIFFERENZIO PIU' AUMENTA LA VARIABILITA' DEI MIEI ERRORI:

$$Var\left[(1-B)^q Y_t\right] = Var\left[(1-B)^q \varepsilon_t\right]$$ [SONO OMOSCHEDASTICI: STESSA VAR σ^2 / SONO INCORRELATI

è somma lineare di varianze quindi più q è grande più aumenta!

⚠ ATTENZIONE! Metodo analitico può dare previsioni ERRATE perchè tiene conto di tutte le osservazioni passate, meglio usare PREVISIONE LOCALE (METODO MEDIE MOBILI)

TREND ESPONENZIALE $f(t) = \alpha_0 e^{\alpha_1 t} \quad \alpha_0 > 0$

$Y_t = f(t)$ è LINEARIZZABILE passando al LOGARITMO $\log f(t) = \log \alpha_0 e^{\alpha_1 t} = \log \alpha_0 + \alpha_1 t$

$Y_t = f(t) + \varepsilon_t$ (MODELLO ADDITIVO) NON è LINEARIZZABILE

$Y_t = f(t) \cdot \varepsilon_t$ (MODELLO MOLTIPLICATIVO) È LINEARIZZABILE (PASSO AI LOGARITMI)

↳ $\log Y_t = \log \alpha_0 + \alpha_1 t + \log \varepsilon_t$

$(1-B)\, Y_t = \text{cost} \Rightarrow q = 1$

$(1-B) \log Y_t = \text{cost} \longrightarrow$ allora va bene il modello esponenziale

METODO DELLE MEDIE MOBILI (TREND LOCALE): da usare quando i dati presentano ANDAMENTO IRREGOLARE

MEDIE MOBILI: medie aritmetiche SEMPLICI o PONDERATE di K successive osservazioni

- K DISPARI $\rightarrow \dfrac{Y_1+Y_2+Y_3}{3} = Y_2^*, \quad \dfrac{Y_2+Y_3+Y_4}{3} = Y_3^* \quad \dots$
- K PARI $\rightarrow \dfrac{Y_1+Y_2}{2} \quad \dfrac{Y_2+Y_3}{2} \longrightarrow$ MEDIA $\dfrac{Y_1+2Y_2+Y_3}{4} = Y_2^*$ TERMINI ESTREMI PESANO LA METÀ

- ORDINE DELLA MEDIA (n° valori considerati per la media)
- MEDIE MOBILI SEMPLICI o PONDERATE

DEVO SCEGLIERE IL GIUSTO ORDINE!

⚠ Più aumento l'ordine delle MM più liscio la serie ma contemporaneamente perdo ADERENZA DELLE OSCILLAZIONI che possono servire!

Considero una serie senza errore $Y_t = T_t + C_t$

Se il movimento ciclico si ripete ogni 5 anni $\Rightarrow$ MM di 5 mi dà il trend

(se vado oltre MA5 creo movimenti anniciclici)

Nella realtà abbiamo TREND + CICLO + ERRORE

EFFETTI SULLA COMPONENTE DISTURBO

$$\varepsilon_t^* = \sum_{i=-m}^{m} \theta_i \, \varepsilon_{t+i}$$

MEDIA MOBILE FATTA SUI RESIDUI

peso, residui

$\boxed{E(\varepsilon_t^*) = 0}$ $\boxed{Var(\varepsilon_t^*) = \sigma^2 \sum_{i=-m}^{m} \theta_i^2}$ $\boxed{Cov(\varepsilon_t^*, \varepsilon_{t+h}^*) = 0}$ ← Quando $h > 2m$, altrimenti ho valori non nulli

(ANDIAMO OLTRE NO TERMINI CHE COMPONGONO LA MEDIA)

EFFETTO SLUTZKY-YULE: creazione di oscillazioni inesistenti

VANTAGGI MM
- SEMPLICITÀ DI CALCOLO E STIMA TREND PIÙ FLESSIBILE RISPETTO A QUELLA ANALITICA
- L'AGGIUNTA DI NUOVI DATI NON MODIFICA I VALORI GIÀ STIMATI
- MEDIA MOBILE DI K TERMINI ELIMINA LE FLUTTUAZIONI DI PERIODO (COSTANTE) K

SVANTAGGI MM
- PERDITA VALORI INIZIALI E FINALI SERIE STORICA
- DEFORMA FLUTTUAZIONI CICLICHE CHE NON HANNO PERIODO K O SUO SOTTOMULTIPLO
- POSSIBILITÀ DI INTRODURRE DISTORSIONI NEL CASO IN CUI L'ANDAMENTO TENDENZIALE NON SIA LINEARE
- NEL CASO DI SERIE PURAMENTE ALEATORIE POSSONO SCATURIRE FLUTTUAZIONI CICLICHE COMPLETAMENTE ARTIFICIALI → (SLUTZKY-YULE)
- LA SOMMA DEI VALORI PEREQUATI NON COINCIDE CON LA SOMMA DEI VALORI ORIGINALI
- ARBITRARIETÀ NELLA SCELTA DEL NUMERO DEI TERMINI

Se elimino TREND posso ottenere:

CICLO LORDO $\quad Y_t - T_t^* = C_t + \varepsilon_t$ $\quad$ se elimino anche ε_t ottengo

CICLO NETTO $\quad Y_t - T_t^* - \varepsilon_t^* = C_t$

STAGIONALITÀ: viene eliminata "spalmandola" su tutto l'anno

MODUS OPERANDI: 1) VERIFICARE SE C'È STAGIONALITÀ
2) DESTAGIONALIZZARE

INDIVIDUARE STAGIONALITÀ:
- USO FUNZIONI ANALITICHE $\quad f(t) = \alpha_0 + \alpha_1 t$

D < 0 (NON SI VERIFICA L'EVENTO) / 1 (SI VERIFICA L'EVENTO)

VARIABILE DUMMY

Es. voglio vedere come 3 tipi di lavoro influenzano il modello: uso 3 DUMMY D_1, D_2, D_3

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_1 + \beta_3 D_2 + \varepsilon_i$$

→ NON HO INSERITO D_3 PERCHÉ COMB. LIN. DELLE ALTRE 2 $\quad D_3 = 1 - D_2 - D_1$

INFLUENZA
- TIPO 1 $\quad Y_i = \beta_0 + \beta_1 X_i + \beta_2 + \varepsilon_i$
- TIPO 2 $\quad Y_i = \beta_0 + \beta_1 X_i + \beta_3 + \varepsilon_i$
- TIPO 3 $\quad Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$

POSSO USARE LE VARIABILI DUMMY PER DESTAGIONALIZZARE I DATI (UNA DUMMY PER OGNI PERIODO)

COSTRUIRE DUMMY IN EVIEWS $\quad S_1 = @SEAS(1)$

← PARI IN CORRISPONDENZA DEL PRIMO MESE

POSIZIONE

STAGIONALITÀ

• USO MEDIE MOBILI

ADDITIVO $Y_t = T_t + C_t + S_t + \varepsilon_t \longrightarrow Y_t - S_t^* = T_t + C_t + \varepsilon_t$

MOLTIPLICATIVO $Y_t = T_t \cdot C_t \cdot S_t \cdot \varepsilon_t \longrightarrow \frac{Y_t}{S_t^*} = T_t \cdot C_t \cdot \varepsilon_t = Y_t^* \leftarrow$ Y DESTAGIONALIZZATO

Nel modello moltiplicativo $Y_t = (T_t \cdot C_t) \cdot \boxed{S_t \cdot \varepsilon_t} \longrightarrow$ STAGIONALITÀ LORDA

STIMA TREND-CICLO $(TC)_t^*$

COSTRUZIONE MM con n° di termini pari al periodo con cui si ripete la stagionalità e stimo $(TC)_t^*$

TRATTAMENTO COMPONENTE STAGIONALE

- MODELLO ADDITIVO $\longrightarrow$ SCARTI STAGIONALITÀ $\quad Y_t - (TC)_t^* = S_t + \varepsilon_t$
- MODELLO MOLTIPLICATIVO $\rightarrow$ COEFFICIENTI STAGIONALITÀ $\quad Y_t / (TC)_t^* = S_t \cdot \varepsilon_t$

STAGIONALITÀ:
- COSTANTE S_t uguale ogni volta che si ripete il fenomeno
- VARIABILE S_t no

Se STAGIONALITÀ COSTANTE faccio la media ed elimino l'errore aleatorio $E(\varepsilon) = 0$

Ge	Fe	Ma	Ap	Ma	...
Me_1	Me_2	Me_3	Me_4	Me_5	...
σ_1^2	σ_2^2	σ_3^2	σ_4^2	σ_5^2	...

MEDIE TUTTE UGUALI $\Rightarrow$ NON HO STAGIONALITÀ

1 MEDIA DIFFERISCE $\Rightarrow$ HO STAGIONALITÀ IN QUEL PERIODO

$\leftarrow \sigma_i^2$ devono essere uguali se σ_i^2 è più grande HO QUALCHE VALORE ANOMALO

Se STAGIONALITÀ CRESCENTE (NON COSTANTE) potrei passare alla TRASFORMAZIONE LOGARITMICA (CASO PIÙ COMPLICATO)

STIMA STAGIONALITÀ $\boxed{S_t^* = \frac{1}{12} \sum_{t=1}^{12} S_t \varepsilon_t}$ $= 0$ (M. ADDITIVO) $\Rightarrow$ NO STAGIONALITÀ; $= 1$ (M. MOLTIPL.) $\Rightarrow$ NO STAGIONALITÀ

Metodi per DESTAGIONALIZZARE uso la procedura $\boxed{\text{TRAMO-SEATS}}$, procedura vecchia non più usata $\boxed{\text{X11}}$

Studio i RESIDUI: valori AUTOCORRELAZIONE devono comportarsi come un WHITE-NOISE.

LISCIAMENTO ESPONENZIALE: stima nuovo valore come media pesata dei valori vecchi con pesi decrescenti.

$$\boxed{\hat{Y}_{t+1} = C_0 Y_t + C_1 Y_{t-1} + C_2 Y_{t-2} + \dots} \quad \text{con } \sum C_i = 1 \quad C_0 > C_1 > C_2 > \dots$$

$$C_j = C_0 \delta^j, \quad j = 0, 1, \dots, n-1 \quad 0 < \delta \leq 1$$

$$\boxed{\hat{Y}_{t+1} = (1-\delta) Y_t + \delta \hat{Y}_t}$$

VISCHIOSITA' DEL SISTEMA : δ (la scelgo io! oppure computer)

- $\delta \to 0$ PIU' PESO AI NUOVI DATI, effetto paragonabile quasi nullo
- $\delta \to 1$ PESO PRESSOCHE' NULLO AI NUOVI DATI, nuova previsione tende a coincidere con la precedente

COME SCELGO δ ?

- A PIACERE (solitamente $\delta = 0{,}05$)
- MINIMIZZO $\sum_t (\hat{Y}_t - Y_t)^2$ SOMMATORIA SCARTI AL QUADRATO (DATO PREVISTO − DATO OSSERVATO)
- MINIMIZZO INDICE ERRORE MEDIO PREVISIONE

Se non abbiamo un trend stazionario → HOLT-WINTERS (cambio di livello)

VALUTAZIONE QUALITA' PREVISIONI

- PREVISIONE PUNTI DI SVOLTA

PREVISTI \ REALIZZATI	P.S.	N.P.S.	
P.S.	m_{11}	m_{12}	$m_{1\cdot}$
N.P.S.	m_{21}	m_{22}	
	$m_{\cdot 1}$		m

In m_{11}, m_{12} ho azzeccato i PUNTI DI SVOLTA

m_{12} pt di svolta PREVISTI MA NON REALIZZATI

↳ $\boxed{E_1 = \frac{m_{12}}{m_{1\cdot}} \quad \text{\% ERR. PRIMA SPECIE}}$

m_{21} pt di svolta REALIZZATI ma NON PREVISTI

↳ $\boxed{E_2 = \frac{m_{21}}{m_{\cdot 1}} \quad \text{\% ERR. SECONDA SPECIE}}$

- INDICI DELL'ERRORE MEDIO DI PREVISIONE

? (x oppure − ?)

$$MQE = \sqrt{\frac{\sum_{t=1}^{T} (P_t - R_t)^2}{T}}$$ ← COMPUTER SCEGLIE δ MINIMIZZANDO QUESTA

(ERRORE QUADRATICO MEDIO)

PROCESSI STOCASTICI (ANALISI MODERNA)

PARTE DETERMINISTICA

Analisi classica prevede modelli rigidi (moltipl., additivo, misto) $Y_t = f(t) + U_t$

COMPONENTE ACCIDENTALE (ALEATORIA)

ANALISI MODERNA si concentra maggiormente sulla parte aleatoria $Y_t = \cancel{f(t)} + U_t$

↳ ogni serie storica è un campione di osservazioni che sarebbero potute essere diverse (problema An. Moderna)

PROCESSO STOCASTICO è una famiglia di variabili casuali descritte da un parametro $t \in T$ (INSIEME PARAMETRICO)

← SOLITAMENTE IL TEMPO

PROC. STOC. < CONTINUO / DISCRETO — a seconda delle variab. casuali

PROC. STOC < A TEMPO CONTINUO / A TEMPO DISCRETO

Noi consideriamo PROC. STOCASTICI A TEMPO DISCRETO con variabili potenzialmente CONTINUE

SERIE STORICA vista come parte finita di un processo stocastico (da campione voglio risalire al processo stocastico)

↳ numero finito di osservazioni

Dovremmo studiare tutte le distribuzioni di densità delle osservazioni e le relative correlazioni ma è MOLTO COMPLICATO

perciò STUDIO I PRIMI MOMENTI:

- FUNZIONE MEDIA $\mu_t = E(Y_t)$
- FUNZIONE VARIANZA $\sigma_t^2 = Var(Y_t) = E(Y_t - \mu_t)^2$
- FUNZIONE AUTOCOVARIANZA $\gamma_{t,s} = E[(Y_t - \mu_t)(Y_s - \mu_s)]$

tempi, scarti $\rightsquigarrow \gamma_{t,t+k} = E[(Y_t - \mu_t)(Y_{t+k} - \mu_{t+k})]$

salto temporale

PRESENTA UNITÀ DI MISURA PER DEPURARE PASSO A:

- FUNZIONE DI AUTOCORRELAZIONE $\rho_{t,s} = \dfrac{\gamma_{t,s}}{\sigma_t \sigma_s}$ (NUMERO PURO)

$\rightsquigarrow \rho_{t,t+k} = \dfrac{\gamma_{t,t+k}}{\sigma_t \sigma_{t+k}}$

Se ho noti processi stocastici posso CONFRONTARLI con l'INDICE DI AUTOCORRELAZIONE (n° puro).

PROCESSO WHITE NOISE ε_t è un disturbo (non è correlato con le altre variabili)

$E(\varepsilon_t) = 0 \qquad Var(\varepsilon_t) = E(\varepsilon_t)^2 = \sigma_\varepsilon^2 \qquad Cov(\varepsilon_t, \varepsilon_s) = E(\varepsilon_t, \varepsilon_s) = 0 \quad \forall t \neq s$

Se $\forall t$ ε_t è anche una v.c. NORMALE $\longrightarrow$ PROCESSO WHITE NOISE GAUSSIANO

PROCESSO GAUSSIANO se le variabili componenti del processo sono $\forall t$ delle variabili casuali normali.

È sufficiente conoscere μ_t (VETTORE VALORI MEDI), MATRICE DELLE VARIANZE, MATRICE DELLE COVARIANZE per conoscere interamente il processo.

PROCESSO STOCASTICO STAZIONARIO (non dipende dal tempo)

- IN SENSO FORTE: distrib. v.c. non dipende da t

  se considero le distribuzioni congiunte $(Y_{t_1}, Y_{t_2}, \dots, Y_{t_n})$ e $(Y_{t_1+\tau}, Y_{t_2+\tau}, \dots, Y_{t_n+\tau})$ sono UGUALI

  per ogni insieme $t_1, t_2, \dots, t_n$ e per ogni τ

  MEDIA E VARIANZA NON CAMBIANO: $\mu_t = \mu \quad \sigma_t^2 = \sigma^2$

AUTOCOVARIANZA $\gamma_{t,t+k} = E[(Y_t - \mu)(Y_{t+k} - \mu)] = \gamma_k \leftarrow$ DIPENDE SOLO DA K (SFASAMENTO TEMPORALE)

1. è un indice delle relazioni lineari esistenti tra coppie di v.c. componenti il processo prese con sfasamento K.
2. è una FUNZIONE PARI di K nel senso che $\gamma_K = \gamma_{-K}$

AUTOCORRELAZIONE $\rho_{t,t+k} = \dfrac{E[(Y_t - \mu)(Y_{t+k} - \mu)]}{\sigma_t \, \sigma_{t+k}} = \dfrac{\gamma_k}{\sigma^2}$ $\sigma^2 = \gamma_0$ (con $K=0$ $\gamma_K = \sigma^2$)

$\longrightarrow \boxed{\rho_K = \dfrac{\gamma_K}{\gamma_0}}$ dipende solo da K

$\boxed{\rho_0 = 1} \qquad \boxed{\rho_K = \rho_{-K}} \qquad \boxed{|\rho_K| \leq 1}$

PARI

NOTAZIONE: ACF (autocorrelation function) PACF (partial autocorrelation function)

CORRELOGRAMMA: per rappresentare il valore che assume la funzione di autocorrelazione

TRA -1 e 1

K ← RITARDI

Da PROCESSO risalgo al CORRELOGRAMMA, NON VALE il viceversa (rumore delle costruzioni fatte).

PACF: Correlazione parziale usata per escludere l'eventuale influenza delle variabili comprese tra Y_t e Y_{t+k} che vengono considerate FISSATE. Stima correlaz. tra Y_t e Y_{t+k} escludendo l'influenza dei valori intermedi.

- IN SENSO DEBOLE se la media è COSTANTE e la funzione di AUTOCOVARIANZA DIPENDE SOLO DA K

$$\boxed{\mu_t=\mu \quad \forall t} \qquad \boxed{\gamma_{t,t+k}=\gamma_k \quad \forall t, \forall k} \qquad \boxed{\sigma_t^2=\sigma^2}$$

TEOREMA DI WOLD Ogni processo aleatorio stazionario IN SENSO DEBOLE può essere SCOMPOSTO in una componente DETERMINISTICA e in una STOCASTICA, ~~data~~ data dalla somma ponderata di un processo WHITE NOISE

PROCESSO ERGODICO quando i valori dell'ACF TENDONO A 0 all'aumentare dei valori.

PROCESSO STOCASTICO INVERTIBILE se può essere ESPRESSO come una funzione lineare delle variabili aleatorie precedenti Y_{t-K} e di un processo White Noise ε_t in modo tale che per ogni t:

$$Y_t = h(\underbrace{Y_{t-1}, Y_{t-2}, \dots}_{\text{VARIAB. ALEATORIE PRECEDENTI}}) + \varepsilon_t$$

← WHITE NOISE

Da un processo STOCASTICO (STAZIONARIO) INVERTIBILE ho la possibilità di RISALIRE AL PROCESSO a partire dalla FUNZIONE DI AUTOCORRELAZIONE.

PROCESSO STAZIONARIO e INVERTIBILE → STIME

$$\boxed{\hat{\mu} = \frac{1}{n}\sum_{t=1}^{n} Y_t} \qquad \boxed{\hat{\gamma}_k = \frac{1}{n}\sum_{t=1}^{n-k}(Y_t-\hat{\mu})(Y_{t+k}-\hat{\mu})} \qquad \boxed{\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0}}$$

Ho bisogno di un NUMERO GRANDE DI OSSERVAZIONI (in genere $K < \frac{1}{4} n$)

PROCESSO A MEDIA MOBILE MA(q)

$$\boxed{Y_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \dots - \theta_q\varepsilon_{t-q}}$$

è una SOMMA DI IMPULSI CASUALI (VARIABILI WHITE NOISE)

riscrivendola in TERMINI DI RITARDO $\boxed{Y_t = \Theta(B)\,\varepsilon_t}$

$\boxed{E(Y_t)=0}$ perché variabili W.N. hanno tutte valore atteso nullo

AUTOCOVARIANZA $\gamma_K = E(Y_t \cdot Y_{t-K})$

- Se $K=0$ $\quad \gamma_0 = \ldots = \sigma_\varepsilon^2(1+\theta_1^2+\theta_2^2+\ldots+\theta_q^2) = \sigma_\varepsilon^2\left(1+\sum_{i=1}^{q}\theta_i^2\right)$
- Se $K=1,\ldots,q$ $\quad \gamma_K = \ldots = \sigma_\varepsilon^2(-\theta_K+\theta_1\theta_{K+1}+\theta_2\theta_{K+2}+\ldots+\theta_{q-K}\theta_q)$
- Se $K>q$ $\quad \gamma_K = 0$

ACF (AUTOCORRELAZIONE) $\longrightarrow \rho_K = \frac{\gamma_K}{\gamma_0}$

- $K=0$ $\quad \rho_K = 1$
- $K=1,\ldots,q$ $\quad \rho_K = \dfrac{-\theta_K+\theta_1\theta_{K+1}+\theta_2\theta_{K+2}+\ldots+\theta_{q-K}\theta_q}{1+\theta_1^2+\theta_2^2+\theta_3^2+\ldots+\theta_q^2}$
- $K>q$ $\quad \rho_K = 0$

↳ UTILE PER CAPIRE QUAL E' L'ORDINE q ; ACF si annulla

PACF (AUTOCORRELAZIONE PARZIALE)

Tende a 0 al divergere di K

MA(q) E' SEMPRE STAZIONARIO

MA(q) E' INVERTIBILE QUANDO: tutte le radici B_i dell'equazione $\theta(B) = 1-\theta_1 B-\theta_2 B^2-\ldots-\theta_q B^q$ sono in modulo >1 $\quad \boxed{|B_i|>1}$

Esempi MA:

- $\boxed{MA(1)}$ $\quad Y_t = \varepsilon_t - \theta_1\varepsilon_{t-1} \quad Y_t = (1-\theta_1 B)\varepsilon_t \quad \theta(B) = (1-\theta_1 B)$

  $|\theta_1|<1 \leftarrow$ CONDIZIONE INVERTIBILITA' $\qquad$ PROCESSO SEMPRE STAZIONARIO

  $E(Y_t)=0$ $\qquad$ AUTOCOV:
  - $K=0$ $\quad \gamma_K = \sigma_\varepsilon^2(1+\theta_1)^2$
  - $K=1$ $\quad \gamma_K = \sigma_\varepsilon^2(-\theta_1)$
  - $K>1$ $\quad \gamma_K = 0$

  ACF:
  - $K=0$ $\quad \rho_K=1$
  - $K=1$ $\quad \rho_K = \dfrac{-\theta_1}{1+\theta_1^2}$
  - $K>1$ $\quad \rho_K=0$

  PACF: Tende a 0 al divergere di K

- $\boxed{MA(2)}$ $\quad Y_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} \quad Y = (1-\theta_1 B-\theta_2 B^2)\varepsilon_t \quad \theta(B) = (1-\theta_1 B-\theta_2 B^2)$

  $$\left.\begin{array}{l}\theta_2+\theta_1<1\\ \theta_2-\theta_1<1\\ -1<\theta_2<1\end{array}\right\}$$ CONDIZIONI DI INVERTIBILITA' $\qquad$ PROCESSO SEMPRE STAZIONARIO

  $E(Y_t)=0$

  AUTOCOV:
  - $K=0$ $\quad \gamma_0 = \sigma_\varepsilon^2(1+\theta_1^2+\theta_2^2)$
  - $K=1$ $\quad \gamma_1 = \sigma_\varepsilon^2(-\theta_1+\theta_1\theta_2)$
  - $K=2$ $\quad \gamma_2 = -\theta_2\sigma_\varepsilon^2$
  - $K>2$ $\quad \gamma_K = 0$

  ACF:
  - $K=0$ $\quad \rho_0=1$
  - $K=1$ $\quad \rho_1 = \dfrac{-\theta_1+\theta_1\theta_2}{1+\theta_1^2+\theta_2^2}$
  - $K=2$ $\quad \rho_2 = \dfrac{-\theta_2}{1+\theta_1^2+\theta_2^2}$
  - $K>2$ $\quad \rho_K=0$

  PACF Tende a 0 al divergere di K

PROCESSO AUTOREGRESSIVO AR(p)

$$\boxed{Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \varepsilon_t}$$ è combinazione variab. al tempo precedente più una variabile aleatoria

scrivendolo in termini di ritardo $\boxed{\phi(B) Y_t = \phi_0 + \varepsilon_t}$

1) $\boxed{E(Y_t) = \dfrac{\phi_0}{1-\phi_1-\phi_2-\ldots-\phi_p}} \longrightarrow \phi_0 = \mu(1-\phi_1-\phi_2-\ldots-\phi_p)$

TRASFORMAZIONE: $\boxed{Z_t = Y_t - \mu}$ $\quad Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \ldots + \phi_p Z_{t-p} + \varepsilon_t$

$\Rightarrow \boxed{E(Z_t) = 0}$

AUTOCOVARIANZA 2) • $\gamma_0 = \phi_1 \gamma_1 + \phi_2 \gamma_2 + \ldots + \phi_p \gamma_p + \sigma^2_\varepsilon$

3) • $\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \ldots + \phi_p \gamma_{k-p}$ — 1), 2), 3) valgono nel caso di stazionarietà

ACF (AUTOCORRELAZIONE) $\longrightarrow$ $\rho_k = \dfrac{\gamma_k}{\gamma_0}$

- $\rho_0 = 1$
- $\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \ldots + \phi_p \rho_{k-p}$

Per un AR(p) stazionario ACF tende a 0 al divergere di K

PACF (AUTOCORR. PARZIALE)

$K \leq p \longrightarrow PACF \neq 0$

$K > p \longrightarrow PACF = 0$

↳ utile per capire qual è l'ordine p : PACF si annulla.

AR(P) È SEMPRE INVERTIBILE

AR(P) È STAZIONARIO QUANDO: tutte le radici dell'equazione $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p$ sono in modulo > 1 $\quad \boxed{|B_i| > 1}$

Esempi AR:

- $\boxed{AR(1)}$ $\quad Y_t = \phi_0 + \phi_1 Y_{t-1} + \varepsilon_t \quad \phi(B) Y_t = \phi_0 + \varepsilon_t \quad E(Y_t) = \dfrac{\phi_0}{1-\phi_1} \rightarrow \phi_0 = \mu(1-\phi_1)$

  $Z_t = Y_t - \mu \quad Z_t = \phi_1 Z_{t-1} + \varepsilon_t \quad \phi(B) = (1-\phi_1 B) \quad |\phi_1| < 1 \leftarrow$ CONDIZ. DI STAZIONARIETÀ

  AUTOCOV: $\gamma_0 = \phi_1 \gamma_1 + \sigma^2_\varepsilon$
  $\gamma_k = \phi_1 \gamma_{k-1} \quad k > 0$

  ACF: $K=0 \quad \rho_k = 1$; $K=1 \quad \rho_k = \phi_1$; $K>1 \quad \rho_k = \phi_1^k$

  PACF: $K=1 \rightarrow PACF = \phi_1$; $K>1 \rightarrow PACF = 0$

- $\boxed{AR(2)}$ $\quad Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t \quad \phi(B) = (1 - \phi_1 B - \phi_2 B^2)$

  $$\left.\begin{array}{l} \phi_2 + \phi_1 < 1 \\ \phi_2 - \phi_1 < 1 \\ -1 < \phi_2 < 1 \end{array}\right\} \text{CONDIZIONE DI STAZIONARIETÀ}$$

  ACF: $K=0 \quad \rho_k = 1$; $K=1 \quad \rho_1 = \dfrac{\phi_1}{1-\phi_2}$; $K>1 \quad \rho_2 = \phi_2 + \phi_1 \rho_1$

  PACF: $K > 2 \rightarrow PACF = 0$

PROCESSO ARMA (p, q)

$$\boxed{Y_t = \phi_0 + \sum_{i=1}^{p} \phi_i Y_{t-i} + \varepsilon_t - \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}}$$ è COMBINAZIONE PROCESSO AR e MA

scrivendolo in TERMINI DI RITARDO. $\boxed{\phi(B) Y_t = \phi_0 + \theta(B)\varepsilon_t}$

STAZIONARIO quando lo è la parte autoregressiva e quindi valgono per i parametri ϕ_i le condizioni che garantiscono la stazionarietà per i processi AR(p)

INVERTIBILE quando lo è la parte a media mobile quindi valgono per i parametri θ_j le condizioni che garantiscono l'invertibilità dei processi MA(q)

ACF (AUTOCORRELAZIONE)

- $K = 0, 1, 2, \dots q$ La ACF dipende sia dalla parte AR che dalla MA
- $K > q$ La ACF si comporta come quella di un processo AR e tende a 0 più o meno velocemente a seconda del valore dei parametri

PACF (AUTOCORR. PARZIALE)

- $K > p$ → La PACF tende a 0 come per un modello MA

Es. $\boxed{\text{ARMA } (1,1)}$ $Y_t = \phi_0 + \phi_1 Y_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1}$ $\quad \phi(B) Y_t = \phi_0 + \theta(B)\varepsilon_t$

CONDIZIONE STAZIONARIETÀ $|\phi_i| < 1$

CONDIZIONE INVERTIBILITÀ $|\theta_j| < 1$

	ACF	PACF
AR(1)	DISCESA ESPONENZIALE	1 VALORE ≠ 0
MA(1)	1 VALORE ≠ 0	DISCESA ESPONENZIALE
AR(p)	DISCESA ESPONENZIALE	P VALORI ≠ 0
MA(q)	Q VALORI ≠ 0	DISCESA ESPONENZIALE
ARMA(p,q)	DISCESA ESPONENZIALE	DISCESA ESPONENZIALE

PROCEDURA DI BOX - JENKINS

① ANALISI PRELIMINARE DELLA SERIE

Serie STAZIONARIA IN MEDIA e IN VARIANZA?

NON STAZIONARIETÀ :
- ANALISI GRAFICA
- ESAME FUNZIONE AUTOCORRELAZIONE

SOLUZIONE:
- IN MEDIA → DIFFERENZIABILITÀ $(1-B)^d Y_t$
  - TREND LINEARE → $d=1$
  - TREND QUADRATICO → $d=2$
- IN VARIANZA → TRASFORMAZIONI : $\log Y_t$, $1/Y_t$, $\sqrt{Y_t}$, $\sqrt[3]{Y_t}$

PRIMA SI SISTEMA LA STAZIONARIETÀ IN VARIANZA E POI QUELLA IN MEDIA!

2) IDENTIFICAZIONE (p e q) Che modello è? Quali p e q?

- ANALISI ACF PACF
- VERIFICA DI CONDIZIONI DI STAZIONARIETÀ E ~~[illegible]~~ INVERTIBILITÀ ??! (invertibilità?)

(eventualmente criteri AKAIKE e SCHWARZ)

3) STIMA DEI PARAMETRI uso STIMATORI MASSIMA VEROSIMIGLIANZA

(ALTERNATIVA METODO DEI MINIMI QUADRATI NON LINEARI)

4) CONTROLLO DIAGNOSTICO ¿ RESIDUI si comportano come un WHITE-NOISE?

- ANALISI GRAFICHE: non ci devono essere regolarità, SCATTER PLOT a "nuvola"
- AUTOCORRELAZIONE DEI RESIDUI: correlogramma come W-N $\left[-1{,}96/\sqrt{n} \div 1{,}96/\sqrt{n}\right]$
- TEST DI PORTMANTEAU: osservo complessivamente le correlazioni uso statistica LJUNG-BOX
- VERIFICA CASUALITÀ RESIDUI
- TEST NORMALITÀ RESIDUI: uso TEST JARQUE-BERA (H_0: normalità)
- DURBIN-WATSON: autocorrelazione di primo ordine (D-W ~ 2)

Q DI LJUNG-BOX

$$Q = n(n+2)\sum_{j=1}^{K}\frac{\hat{\rho}_j^{\,2}}{n-j} \sim \chi^2_{K,p,q}$$

H_0: non c'è AUTOCORRELAZIONE nel gruppo considerato

PARAMETRI SIGNIFICATIVI $|t^*| < 1{,}96$ accetto H_0: t non è SIGNIFICATIVO

JARQUE-BERA

$$JB = \frac{n}{6}\left[\underset{\text{SIMMETRIA}}{S^2} + \frac{1}{4}(\underset{\text{CURTOSI}}{K^2}-3)\right] \sim \chi^2_2$$

H_0: distribuzione NORMALE

AKAIKE: $AIC(K) = -\frac{2}{n}\left(\log \underbrace{L(\hat{\delta})}_{\text{VEROSIMIGLIANZA}} - k\right)$

(SCHWARZ) per evitare SOVRAPARAMETRIZZAZIONE [PRINCIPIO PARSIMONIA]

AR come INFINITI TERMINI MA

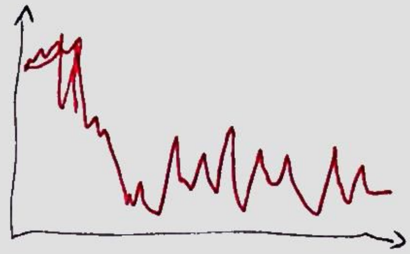
$Y_t = \phi_1 Y_{t-1} + \varepsilon_t \qquad Y_{t-1} = \phi_2 Y_{t-2} + \varepsilon_t$ e ITERO

MA come INFINITI TERMINI AR

$Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} \qquad Y_{t-1} = \varepsilon_{t-1} - \theta_1 \varepsilon_{t-2}$ e ITERO

PROCESSO RANDOM WALK (NON è STAZIONARIO)

$$Y_t = Y_{t-1} + \varepsilon_t \qquad Y_0 = \mu$$



$E(Y_t) = \mu$

$Var(Y_t) = t\,\sigma^2_\varepsilon$ ⟶ dipende da t

$\rho_k = \sqrt{\frac{t}{t+k}}$ ⟶ dipende da t e tende a 1 al crescere di t (il processo ha MEMORIA INFINITA)

POSSIBILI CAUSE DELLA NON STAZIONARIETÀ

- TREND DETERMINISTICO $Y_t = \underbrace{\beta_0 + \beta_1 t}_{\text{PARTE DETERMINISTICA}} + \underbrace{\phi_1 Y_{t-1} + \varepsilon_t}_{\text{PARTE ALEATORIA}}$ ⟶ ELIMINO CON DETRENDIZZAZIONE o DIFFERENZIAZIONE

- TREND STOCASTICO $Y_t = \underbrace{\beta + Y_{t-1} + \varepsilon_t}_{\text{RANDOM WALK CON DRIFT (RANDOM WALK + }\beta)}$ ⟶ ELIMINO SOLO CON DIFFERENZIAZIONE

UNITROOT TEST

mira a testare che $\phi_1 = 1$ come IPOTESI ALTERNATIVA

REGRESSIONE

A

$Y = \beta_0 + \beta_1 X$ Relazione funzionale
(intercetta, coeff. angolare)

$Y = f(x) + \varepsilon$ Relazione statistica
(var. casuale, componente deterministica, componente stocastica)

CR. MINIMI QUADRATI per stimare β_0, β_1

R^2: Bontà approssimazione

Y_i si distribuisce come una Normale rispetto ad X_i

IPOTESI SU ε:
1) ε si distribuisce come una Normale
2) ε_i sono variab. casuali indipendenti con $E(\varepsilon_i)=0$ e varianza costante $V(\varepsilon_i)=\sigma^2$ → omoschedasticità
3) $Y_i = \beta_0 + \beta_1 X_i + \varepsilon \quad \forall i=1,\dots,n$
4) I valori X_i della variabile esplicativa X sono senza errore

$Y_i \sim N(\beta_0 + \beta_1 X_i \,;\, \sigma^2)$

CR. MINIMI QUADRATI trovo stime di β_0, β_1: $\boxed{\hat{\beta}_1 = \dfrac{\sigma_{XY}}{\sigma^2_X}}$ $\boxed{\hat{\beta}_0 = Y - \hat{\beta}_1 X}$

↳ stimatori puntuali o intervallari

DECOMPOSIZIONE DELLA VARIANZA (devianza è più semplice): (ANOVA: analisi della varianza)

Devianza di regressione: parte di varianza spiegata dal modello di regressione
Devianza residua: parte di varianza non spiegata dal modello di regressione

$$SQT = SQR + SQE \implies \frac{\cancel{SQT}}{\cancel{SQT}}_{\,1} = \frac{SQR}{SQT} + \frac{SQE}{SQT}$$

(Dev. totale, Dev. regress., Dev. errori)

$\frac{SQR}{SQT} \to R^2 \to [0, 1]$ (0: non spiega nulla; 1: spiega tutta la varianza)

STIMA INTERVALLARE

- β_0, β_1 stimatori corretti $\implies E(B_1)=\beta_1 \quad E(B_0)=\beta_0$
- B_0, B_1 si distribuiscono normalmente $\implies B_0 \sim N(\beta_0, Var(B_0)),\; B_1 \sim N(\beta_1, Var(B_1))$
- B_0, B_1 sono gli stimatori più efficienti (varianza minima tra tutti gli stimatori corretti)

$$V(B_1) = \frac{\sigma^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \qquad V(B_0) = \sigma^2\left[\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n}(X_i-\bar{X})^2}\right]$$

$$Cov(B_0, B_1) = -\sigma^2 \cdot \frac{X}{\sum_{i=1}^{n}(X_i-\bar{X})^2}$$

Valgono in caso di omoschedasticità [σ^2 costante]

Stima σ^2 — residui

$$S^2 = \frac{\sum_{i=1}^{n}\hat{e}_i^2}{n-2}$$

↳ serve ad stimare σ^2

$\sqrt{S^2}$ è l'errore standard di regressione

Se sfrutto S^2 posso trovare $V(B_1), V(B_0), Cov(B_0, B_1)$

$B_0 \sim N$, $B_1 \sim N$ $\Big\}$ $B_j \sim N(\beta_j, Var(B_j))$ (generalizzo)

$\boxed{BLUE}$ → Best Linear Unbiased Estimator (migliori lineari corretti stimatori)

Posso ~~...~~ ottenere la normale std B_1: $\dfrac{B_1-\beta_1}{\sqrt{Var(B_1)}} \sim N(0,1)$ — da stimare!

$$\frac{B_1-\beta_1}{S(B_1)} \sim t_{n-2}$$ t di Student con $n-2$ gr. di libertà

Se ho più variabili: $\hat{y} = \hat\beta_0 + \hat\beta_1 x_1 + \hat\beta_2 x_2 + \hat\beta_3 x_3 + \dots$

QUALI VARIABILI SPIEGANO Y E QUALI NO?

$\begin{cases} \text{IP. NULLA } H_0: \beta_1 = 0 \\ \text{IP. ALTER. } H_1: \beta_1 \neq 0 \text{ (BILATERALE)} \end{cases}$ H_0 bisogna dare una STATISTICA TEST e di α LIVELLO DI CONFIDENZA ($\alpha = 0{,}05$)

$\frac{B_1 - \beta_1}{S(B_1)} \sim t_{n-2}$ funzione TEST

Sotto H_0 ho la DISTRIBUZIONE $\frac{B_1}{S(B_1)} \sim t_{n-2}$ $H_0: \beta_1 = 0$

$t^* = \frac{\hat{B}_1}{S(B_1)}$ 

USO P-VALUE

Voglio che $\boxed{\text{inserire un NUOVO CARATTERE}}$ $X_{k,i}$ nel modello (NUOVA VARIABILE)

IPOTESI DI ASSENZA DI MULTICOLLINEARITÀ (tutte le variabili sono l.i.)

$B_j \sim N(\beta_j, Var(B_j))$ ovvero K PARAMETRI $\boxed{n >> K}$

OSSERVAZIONI PARAMETRI

PRINCIPIO DELLA PARSIMONIA: Usare meno VARIABILI POSSIBILI per aumentare la precisione delle stime.

↳ USEREMO R^2 CORRETTO

$R^2 = \frac{SQR}{SQT} = 1 - \frac{SQE}{SQT}$ → in inferenza consideriamo le VARIANZE CORRETTE

$\boxed{\text{VAR. CORRETTA} = \frac{SQT}{n-1}}$ $\boxed{\text{VAR. DEI RESID. CORR} = \frac{SQE}{n-K}}$ $\boxed{\text{VAR. REGR. CORR.} = \frac{SQR}{K-1}}$

$\Rightarrow \bar{R}^2$ (R^2 corretto) $= 1 - \frac{\left(\frac{SQE}{n-K}\right)}{\left(\frac{SQT}{n-1}\right)} = 1 - \frac{n-1}{n-K}(1 - R^2)$ $\bar{R}^2 = \begin{cases} \text{MAX} = 1 \\ \text{MIN} = \text{NEGATIVO} \end{cases}$

TEST F: va a vedere se tutti parametri sono contemporaneamente NON SIGNIFICATIVI PER IL MODELLO

$\begin{cases} H_0: \beta_1 = \beta_2 = \dots = \beta_{K-1} = 0 \\ H_1: \text{almeno un } \beta_j \neq 0 \text{ con } j = 1, \dots, K \end{cases}$

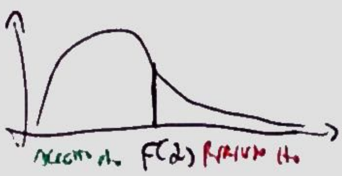
GR. LIBERTÀ

Con $\begin{cases} H_0: \beta_1 = 0 \\ H_1: \beta_1 \neq 0 \end{cases} \longrightarrow \begin{cases} SQT = \sum (y_i - \bar{y})^2 & n-1 \\ SQE = \sum e_i^2 & n-2 \\ SQR = \sum (\hat{y}_i - \bar{y})^2 & 1 \end{cases}$ $\longrightarrow SQT/(n-1)$, $\longrightarrow SQE/(n-2)$, $\longrightarrow SQR/1$

VAR. CORRETTE

$\boxed{F = \frac{SQR/1}{SQE/(n-2)}}$ $\boxed{F \sim F_{1,(n-2)}}$ GRADI DENOMINATORE

GRADI NUMERATORE



$F^* > F(\alpha) \Longrightarrow$ RIFIUTO H_0 e ACCETTO H_1

$F = \frac{SQR(n-2)}{SQE} \cdot \frac{SQT}{SQT} = \underbrace{\frac{SQR}{SQT}}_{R^2} \cdot \underbrace{\frac{SQT}{SQE}}_{1/(1-R^2)} \cdot (n-2) = R^2 \cdot \frac{1}{1-R^2} \cdot (n-2)$

↳ F è un test unilat. di R^2

B

Prima di poter usare un modello lineare verifico la validità delle ipotesi del modello:

ANALISI DEI RESIDUI $\varepsilon_i = Y_i - (\beta_0 + \beta_1 x_i)$

DISTRIB. CASUALE → BUONO!

DISTRIB. PARABOLA

→ NO REGR. LINEARE, PROVO AD AGGIUNGERE UN PARAMETRO $Y = \beta_0 + \beta_1 X + \beta_2 X^2$ DI II° GRADO

OMOSCHEDASTICITÀ (non residui costanti)

← MANTIENE STESSA VARIANZA

ETEROSCHEDASTICITÀ → PROBLEMA! provo a TRASFORMARE il modello (uso LOGARITMO!)

ASSUNZIONE DI INDIPENDENZA → TEST DI DURBIN-WATSON $\begin{cases} H_0: \rho = 0 & \text{INDIP} \\ H_1: \rho \neq 0 & \text{AUTOCORRELAZIONE} \end{cases}$

$$\boxed{d = \frac{\sum_{t=2}^{n} (\hat{e}_t - \hat{e}_{t-1})^2}{\sum_{t=2}^{n} \hat{e}_t^2}}$$

$$\boxed{d \simeq 2(1-\rho)}$$

SI DISTRIBUISCE STIMA

se $\rho = 1 \rightarrow d = 0$ AUTOCORR. POSITIVA PRIMO ORDINE

$\rho = -1 \rightarrow d = 4$ AUTOCORR. NEGATIVA PRIMO ORDINE

$\rho = 0 \rightarrow d = 2$ NON C'È AUTOCORR. DEL PRIMO ORDINE

Se calcolo d abbiamo una prima grossolana stima di ρ ("rò")

ASSUNZIONE DI NORMALITÀ

$\boxed{96\%}$ dei RESIDUI deve stare tra -1 e 1 affinché sia NORMALE

NORM. STANDARDIZZATA

Jarque-Bera : TEST DI NORMALITÀ basato su Skewness e Curtosi