



Corso Luigi Einaudi, 55 - Torino

Appunti universitari

Tesi di laurea

Cartoleria e cancelleria

Stampa file e fotocopie

Print on demand

Rilegature

NUMERO : 454

DATA : 18/02/2013

A P P U N T I

STUDENTE : Parrino

MATERIA : Logistica di Distribuzione + schemi + Eserc.

Prof. Alfieri

Il presente lavoro nasce dall'impegno dell'autore ed è distribuito in accordo con il Centro Appunti.

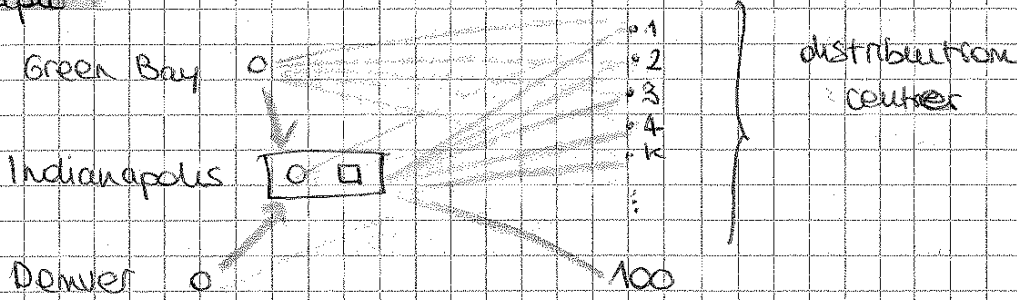
Tutti i diritti sono riservati. È vietata qualsiasi riproduzione, copia totale o parziale, dei contenuti inseriti nel presente volume, ivi inclusa la memorizzazione, rielaborazione, diffusione o distribuzione dei contenuti stessi mediante qualunque supporto magnetico o cartaceo, piattaforma tecnologica o rete telematica, senza previa autorizzazione scritta dell'autore.

**ATTENZIONE: QUESTI APPUNTI SONO FATTI DA STUDENTIE NON SONO STATI VISIONATI DAL DOCENTE.
IL NOME DEL PROFESSORE, SERVE SOLO PER IDENTIFICARE IL CORSO.**

LOGISTICA DI DISTRIBUZIONE

4 OTTOBRE

esempio



le strategie vengono valutate in termini di costi, servizi ai clienti, and time e' tutto in funzione del costo espliciti / impliciti

L'impatto della strategia in termini di efficienza e' legato ai costi.

es: 3 IMPIANTI distribuiti in 3 punti diversi che producono TV, televisore oggetti di elettronica che devono essere distribuiti in 100 centri di distribuz.

Assemblaggio → nel mag. centrale (warehouse) → **INDIRETTO**
 → nei centri di distribuzione → **DIRETTO**

Mag. centrale e' considerato come a distanza 0 dall'impianto

① Ogni impianto distribuisce a tutti gli impianti e l'assemblaggio al centro di distribuzione (diretto)

② Oppure Denver e Green Bay trasferiscono i prodotti al mag. centrale di Indianapolis e poi distribuzione nei centri (indiretto)

Data la capienza dei trasporti, come conciliamo la capacita'? Pieno carico o no?

4 STRATEGIE

FTL = FULL TRUCK LOAD (dir) a pieno carico - trasporto diretto tra impianti e centri di distribuzione a pieno carico

FTL = trasporto a pieno carico (indir) indiretto - trasporto indiretto tra un impianto (mag. intermedio) e centri di distribuz.

LFTL = Less than FULL TRUCK LOAD (diretto) - assemblaggio nei centri di distribuzione non a pieno carico

LFTL indiretto - non carico completamente e passa da Indianapolis

Ogni DC ha domanda ANNUALE di 2500 (10.250)

2500	PC	} u/anno
2500	monitor + k	
2500	TV	
2500	console	

I COSTI SONO:

• COSTO DI TRASPORTO

Aggregare le domande, collezionando la domanda in un macrocentro e commettere grosse approssim.



E' errato far partire fisicamente un camion da un DC all'altro perché non abbiamo tra i dati le distanze tra i centri.

Per fare un'analisi semplificata o considero distanze nulla* → AGGREGARE IN 1 UNICO CENTRO I DC introduce un grande errore
 distanza infinita → considero i centri in maniera indipendente

FTL diretto

GB	→ 2500 u/anno
D	→ 2500 u/anno
I	→ 2500 TV/anno
	→ 2500 monitor + k/anno

GB → $\frac{2500 \cdot 5100}{30'000} = 0,41667$ (camion che partono da GB all'anno)

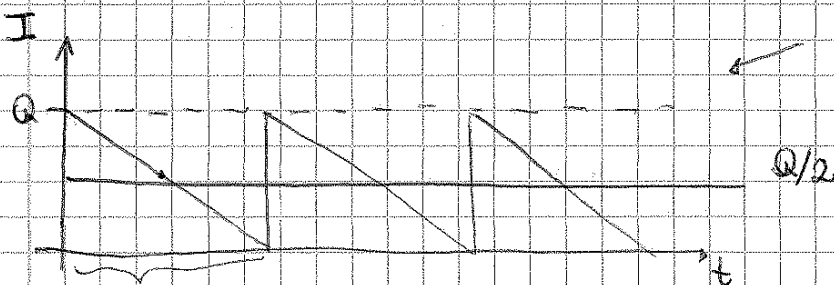
D → $\frac{2500 \cdot 30}{30'000} = 2,5$ (camion/anno)

I → $\frac{2500 \cdot 2 \cdot 10}{30'000} = 1,667$ (camion/anno)

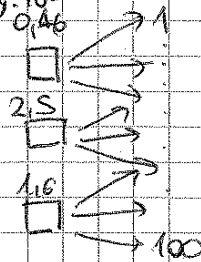
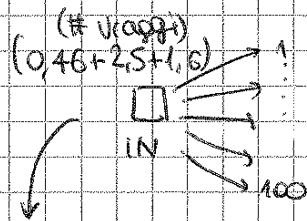
camion = # viaggi

Costo trasporto = $0,4166 \cdot 100 \cdot 1000 \cdot 1 + 2,5 \cdot 100 \cdot 1000 \cdot 1 + 1,667 \cdot 100 \cdot 1000 \cdot 1$
 (DC) (me) (\$/me)

= $\frac{t}{\text{anno}} \cdot DC \cdot me \cdot \frac{\$}{me} = \$/\text{anno} = 4,6 \cdot 10^5 \$/\text{anno}$
 ↓
 frequenza



		C. trasp.	C _{NAB}	C_{trasp}	C _{TOT}
①	FTL dir	$4,6 \cdot 10^5$	F $4,65 \cdot 10^4$ SF $2,35 \cdot 10^7$		F $4,7 \cdot 10^4$ SF $2,41 \cdot 10^7$
②	FTL indir	$7,5 \cdot 10^5$	SF $5,25 \cdot 10^6$ F $10,1 \cdot 10^6$		SF $6 \cdot 10^6$ F $10,86 \cdot 10^6$
③	LFTL dir	F $3,41 \cdot 10^6$ SF $2,41 \cdot 10^6$	F $3,41 \cdot 10^6$ SF $2,42 \cdot 10^6$		F $6,8 \cdot 10^6$ SF $4,84 \cdot 10^6$
④	LFTL indir	F $2,46 \cdot 10^6$ SF $1,85 \cdot 10^6$	F $2,2 \cdot 10^6$ SF $1,35 \cdot 10^6$ 0,46		F $4,66 \cdot 10^6$ SF $3,44 \cdot 10^6$



lavoro a camion pieni per questo faccio sempre lo stesso # viaggi all'anno

② si utilizza il mag. centrale. → ci aspettiamo un aumento dei costi di trasporto
 Conoscendo i camion con oggetti differenti non fanno i costi di trasporto ma i costi di magazzino

Lo il # di camion è uguale al primo caso → camion omogenei
 Mi aspetto che tale strategia abbia un impatto positivo sui costi di magazzino

I camion ibridi mi consentono di seguire la domanda anche se non si ~~considerano~~ modificano # trasporti

Il mix di item trasportati a camion pieni deve essere tale da avere basso valore (meno costo di magazzino) totale di modo da avere max carico con minor valore immobilizzato

COSTI DI TRASPORTO

$$GB \rightarrow IND \quad \frac{\text{domande } DC \times \# DC \times \text{peso } DC}{\text{capacità camion}} \cdot ME \cdot \$/ME = 16640 \text{ \$/anno}$$

$$\frac{2500 \cdot 100 \cdot 5}{30000} \cdot 400 \cdot 1 = 16640$$

$$D \rightarrow IND \quad \frac{2500 \cdot 100 \cdot 30}{30000} \cdot 1000 \cdot 1 = 275000 \text{ \$/anno}$$

$$IND \rightarrow DC \quad 4,6 \cdot 10^5 \quad (\text{come nel caso ①} - \text{è necessaria la stessa merce in ogni centro all'anno})$$

$$C_{TOT} = 7,5 \cdot 10^5 \rightarrow \text{ci si può arrivare anche con il bundle} \quad \left[\begin{array}{l} \text{peso} = 55 \\ \text{costo} = 1200 \end{array} \right]$$

$$IND = \# \text{ viaggi } \cdot DC \cdot 1000 \cdot 1 \cdot 100 = \frac{2500 \cdot 55}{30000} \cdot 1000 \cdot 1 \cdot 100 =$$

bundle / anno × DC

ho senso creare un BUNDLE → considerare come insieme un mix di oggetti diversi
 che in qst caso è: tv + monitor + console + PC

perché tutto la domanda è omogenea

ragioniamo nel continuo

peso BUNDLE = 10 + 30 + 5 + 10 = 55 lbs

quanti BUNDLE stanno su un camion? $\frac{30000}{55} = 545,5$ bundle

~~TRASPORTI SPASATI~~

value/costo BUNDLE = 400 + 100 + 300 + 400 = 1200 (TV)

TRASPORTI SPASATI

CM (IN) = $\frac{545,5 \cdot 1200}{2}$

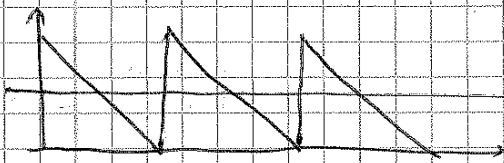
CM (DC) = $\frac{545,5 \cdot 1200}{2} \cdot 100 \cdot 0,15$

CM (TR) = $\frac{545,5 \cdot 1200}{2} \cdot (100+1) \cdot 0,15$
 = $33 \cdot 10^7 \cdot 0,15$

\$ che si accumulano in MS un ottavo prima che il camion parta

il costo è dovuto alle "attese" di carico e quel al nodo del nodo come origine

CMA6 = $(1,9 \cdot 10^6 + 3,3 \cdot 10^7) \cdot 0,15 = 5,25 \cdot 10^6$ \$/anno



se il costo del mag. è all'anno di co perché abbiamo calcolato il max carico di magazzino, la prossima è arriva a "nulla"

con Q/2 considero valore medio x tutto l'anno immobilizzato

TRASPORTI SPASATI

IND $\frac{545,5 \cdot 1200 \cdot 100}{2} \cdot 0,15 =$

DC $\frac{545,5 \cdot 1200}{2} \cdot 100 \cdot 0,15$

CM = $\frac{545,5 \cdot 1200}{2} \cdot (100+100) \cdot 0,15 = 6,55 \cdot 10^7$ \$/anno

CMA6 = $(1,9 \cdot 10^6 + 6,55 \cdot 10^7) \cdot 0,15 = 10,4 \cdot 10^6$ \$/anno

la ② abbate i costi alle sorgenti ed è vincita della ①

TRASPORTI SFASATI

è uguale al caso fasato tranne che per i costi totali

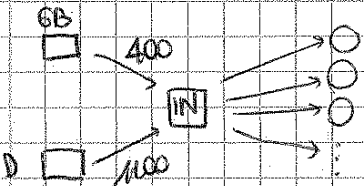
$$\text{Costo TOT} = \text{costo unitario } \frac{Q}{2} (100+1) + \frac{1}{2} Q \cdot 101$$

$$Q^* = \sqrt{\frac{2 \cdot 100 \cdot AD}{101}} =$$

⊙ (a) In il caso fasato $h_0 + \text{magazzino in partenza mentre nel caso sfasato riempio un camion alla volta e quindi } h_0 \text{ minor costo di magazzino che mi dà + margine nel magazzino dei clienti}$

LFT INDIRETTO

perché nei casi di LFT i costi transp e mag sono dello stesso ordine di grandezza non mi posso aspettare che ad una diminuzione del cmg e ad un aumento del CRT h_0 CRT minore



lo sfasato e il fasato non importa sulla prima parte (distribuzione punto a punto) [da origine a destino]

Aumento di costo $\rightarrow A \cdot \frac{D}{Q}$ \rightarrow non moltiplica i costi $\times 100$ ma $\times 100$ D significa la domanda aggregata di tutti i DC

$A \rightarrow$ costo fisso di trasporto

I PARTE

$$A \cdot \frac{D}{Q} + \frac{1}{2} Q \cdot 2$$

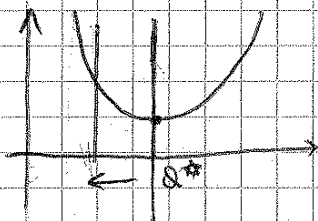
perché a INDIANAPOLIS, pago anche parte del mg per cercare i camion per DC

$$GB \rightarrow Q^* = \sqrt{\frac{(400 \cdot 1) \cdot (2500 \cdot 100)}{300 \cdot 0,15}} = 1490,7$$

$$D \rightarrow Q^* = \sqrt{\frac{(100 \cdot 1) \cdot (2500 \cdot 100)}{100 \cdot 0,15}} = 4281,7$$

non si crea un problema di sincronizzazione perché lavoro

penz. di costo



A D (la qto' ottima) supera il costo max per cui prendo la qto' max che mi dà minor costo $\rightarrow \frac{30000}{30}$

$$\text{COT} = \frac{(400 \cdot 1) \cdot (2500 \cdot 100)}{1490,7} + (300 \cdot 0,15) 1490,7 + \frac{(100 \cdot 1) \cdot (2500 \cdot 100)}{30000/30} + (100 \cdot 0,15) (30000/30) = 0,42 \cdot 10^6 \text{ \$/anno}$$

CAPITOLO 1

10 ottobre

LOGISTICA

Nasce in contesto militare

NAPOLEONE → Spostamento e sostentamento efficiente dell'esercito

TRASFERIMENTO DI RISORSE

focus on RELAZIONE TRA Nodi

le tecniche de noi studiate II GUERRA MOND (lotto economico)

Gestione di un prodotto dall'estrazione delle materie prime al consumatore

Descrive l'integrazione di due o più attività con lo scopo di controllare il flusso di materie prime e prod. finiti dall'origine al consumo finale

Che forma dovrà avere la nostra rete?

CONFIGURAZIONE LINEARE



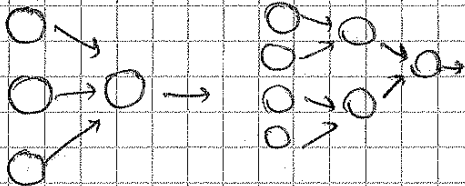
MAGAZINI per trasferire i prod. di prod. al consumo

ogni nodo ha un solo successore e un solo predecessore

Spesso il magazzino intermedio ha i suoi vantaggi. ← hanno un certo costo

I Nodi saranno per noi delle BLACKBOX, non abbiamo idea di che tipo possano essere però dobbiamo rappr. la natura interna di un nodo che impatta sulla chain. Il CT di produzione impatta sulla catena, ed è importante per avere un'idea di quale sia il LEAD TIME (variabile decisionale) che la catena vede.

la config. lineare non consente assemblaggio e per facilitare dovrebbe essere CONVERGENTE

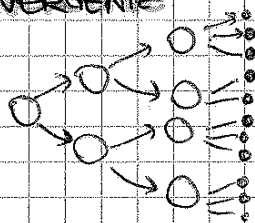


CONFIGURAZIONE CONVERGENTE (consente l'assemblaggio)

Non obbliga a fare assemblaggi

CONFIGURAZIONE DIVERGENTE

consente di distribuire in un territorio piuttosto vasto



I nodi intermedi → sembra che siano solo un costo e che non diano nessun valore aggiunto

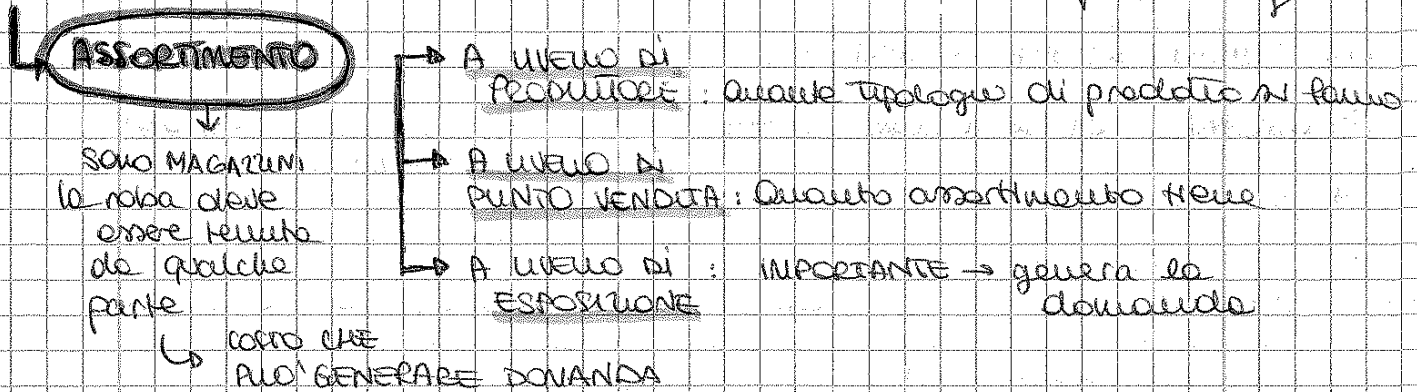
IMAG. INTERMEDI
Riduciamo l'incertezza sulla domanda

← hanno un ruolo importante: aggregare la merce da distribuire
← tale struttura mi permette di trasportare di meno in ognuno dei centri (meno magazzini)

Devo capire bene qual'è la natura dei Nodi INTERMEDI:

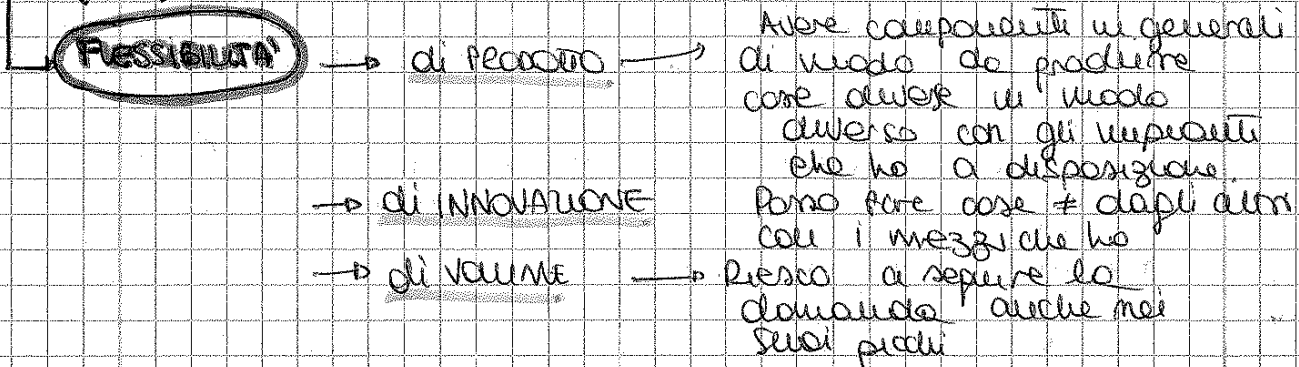
I prodotti si fermano o transitano? → che condizioni i costi

la **VARIABILITÀ** in azienda ~~si~~ va considerata → Quello che porta il cliente e l'incertezza e la variabilità dei tempi di consegna
 (per essere competitivi nel tempo)
 → SIA PER L'ECESSO, SIA PER IL DIFETTO.



INNOVAZIONE → lanciare sul mercato qualcosa di nuovo richiede molto ricerca

PERSONALIZZAZIONE



DRIVER DI COSTO → ogni costo è legato alla scelta → COSTIFICARE LE POLITICHE

NATURA DEL LEGAME TRA COSTO & DECISIONE

FORMA DEL LEGAME TRA COSTO & DECISIONE

In funzione della decisione metto o elimino alcuni costi

I costi fissi = sunk cost (impedite) → costo di normale apprendimento rispetto alla nostra variabile decisionale non dipende dalla decisione (F)

COSTO MEDIO
 $\frac{C(x)}{x}$

COSTO MARGINALE

$C'(x)$
 se il costo è lineare
 $C'(x) = \frac{d(Cx)}{dx}$

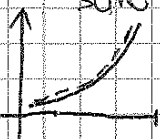
$$C_{tot}(x) = \begin{cases} F + cx & x > 0 \\ 0 & x = 0 \end{cases}$$

da cui: $C(x) = c \cdot x$ $\frac{C(x)}{x} = c$

$\frac{d(Cx)}{dx} = c$

nella realtà i costi fissi sono variabili a tratti
 nella realtà i costi non sono lineari

↳ se lineare x piccoli valori di x lo rapp. a tratti

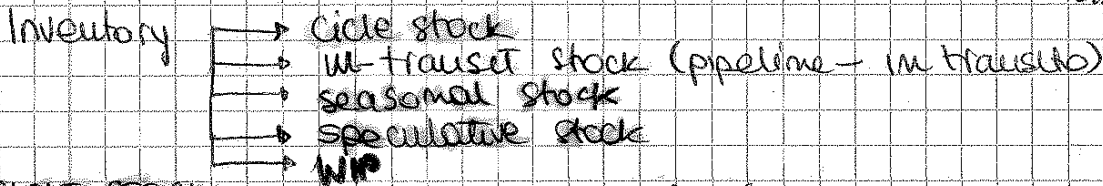


devo quindi confrontare i 2 costi totali dei 2 fornitori

$$C_1 = \sqrt{2A_1 d h_1} + C_1 d \rightarrow \text{anche la II parte e' rilevante per il confronto tra 2 fornitori}$$

$$C_{TOT2} = \sqrt{2A_2 d h_2} + C_2 d$$

12 OTTOBRE



CICLE STOCK

→ riguardano i vincoli di capacità
 Per rispondere a domanda costante siamo obbligati a produrre lotti a (lotto economico) non possiamo avere una produzione continua perché troppo costoso (A) per cui siamo obbligati a tenere delle scorte

Dal lato del mercato abbiamo vincoli di capacità momentanea. Non possiamo essere sempre allineati, sia nella produzione che nel trasporto. X fare prod. e domanda

IN-TRANSIT STOCK - PIPELINE

Le scorte in transito non dipendono dalla capacità ma dal fatto che esiste una distanza. Se ci fosse il teletrasporto non ci sarebbero. Se ci sono tempi di trasporto queste scorte devono essere prese in considerazione.

Le scorte di trasporto → NON DIPENDONO DALLA CAPACITÀ (dei vincoli di capacità)

→ considerate se ho due punti lontani temporalmente

WIP

→ dovuto alla presenza di un sistema di produzione per gestire bene la produzione per fare in modo efficiente: setup, x proteggere dai guasti

→ AL DI SOTTO DEL WIP CRITICO IL SISTEMA FUNZIONA MALE

SPECULATIVE STOCK

→ si vedono x ottenere opportunità economica

Per gli approvvigionamenti di materie prime

→ scorte che nascono dal fatto che abbiamo un vantaggio economico (offerta supermercato)

SEASONAL STOCK

→ servono x accordare una produzione continua ad una domanda discontinua

→ inizio primo a produrre e poi conservo in magazzino

→ domanda discontinua (pavettoni - gelati)

→ Anche la produzione può essere discontinua e la domanda continua (olio - pomodoro) → non si solo fatto l'anno

STAGIONATA

→ PRODUZIONE

→ disallinea la produzione dalla domanda

l'incertezza sull' INNOVAZIONE

↳ di fiducia dei paesi degli ESPERTI delle PREVISIONI DI MERCATO

PIU' RISCHIOSA ← descrizione dell'incertezza SOBGETTIVA

dipende dall' ORIZZONTE TEMPORALE

IN cui ci poniamo
BREVE PERIODO
LUNGO PERIODO

riguarda il quanto durerà il mio prodotto (determinare quando sarà il fine-vita)

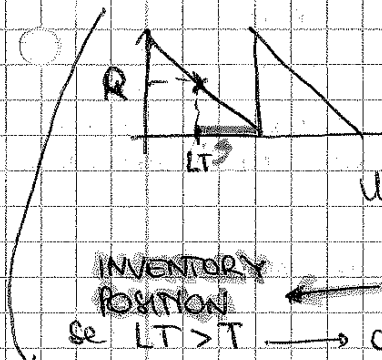
incertezza di livello

SCORTE DI SICUREZZA

x attitudine
↳ l'oggetto dell'incertezza
x quanto riguarda la domanda (x una particolare tipologia di incertezza)

↳ tali SCORTE generano notevoli COSTI

ABBIAMO 2 INCERTEZZE:
• domanda variabile
• LT variabile



ordino quando il magazzino è ad un livello R

↳ la produzione non è Istantanea
devo ordinarla un LEAD TIME PRIMA

$R = D \cdot LT = pz/c \cdot t = pt$
se $LT < T$

dato la variabilità del LT e della domanda lo ordinerò prima (Scorta di Sicurezza) → ordinerò ad una quantità > R
nei casi reali sarà più piccola se D e LT saranno + alti

Se la DOMANDA è VARIABILE → devo capire che distribuzione ha durante il LT

ASSUNZIONE:
 $D_{LT} \sim N(\mu_{LT}, \sigma_{LT}^2)$

livello di servizio al cliente come probabilità di soddisfare l'ordine

PER IL TEO DEL LIMITE CENTRALE è una NORMALE perché SOMMA di tutte le domande dei centri

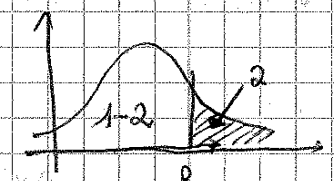
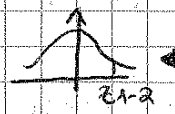
(la domanda durante il lead time)

$P = \{ D_{LT} \leq R \} = 1 - \alpha$

α = probabilità di stockout
probabilità di soddisfare totalmente la domanda

$R = \mu_{LT} + z_{1-\alpha} \sigma_{LT}$

↳ quantile della NORMALE STANDARDIZZATA che poi viene DESTANDARDIZZATA



μ_{LT} = domanda media durante il LT

PROBLEMA SUL USUO

COMP.	C1	C2	C3	C4	C5
A1	1	1	1	/	/
A2	1	1	/	1	1
A3	1	1	/	/	1

PRODOTTI DAL NOSTRO STABILIMENTO
 S_1, S_2, S_3
 y_1, y_2, y_3

i \ m	M1	M2	M3
C1	1	2	1
C2	1	2	2
C3	2	2	0
C4	1	2	0
C5	3	2	0

RISORSE M
 che sono necessarie x la produzione

DISPOSIZ. DELLE RISORSE → L_m → Qtà risorse disponibili.
 M1 800
 M2 700
 M3 600

COSTO DI PRODUZIONE c_i

C1	C2	C3	C4	C5
20	30	10	10	10

PREZZO & DOMANDA P.F

P _j	prezzo	S ₁	S ₂	S ₃	d
A1	80	100	50	120	90
A2	70	50	25	60	45
A3	90	100	100	60	90

tre scenari equiprobabili prob. 1/3 →

Imposto un semplice modello di PROGRAMMAZIONE LINEARE

↳ MODELLO A DUE STADI CON RICORSO *

demande media

j = prod. finiti
 i = componenti

(ricavi - costi) / pro' venduta

↳ le variabili decise x sono quelle

fo → max $\sum_j P_j y_j - \sum_i c_i x_i$

implementate nel I stadio (subto) l'assemblaggio (y_j) e' nell II stadio in cui la dom. e' certa ma vincolata su disp. dei compon.

s.t.

$\sum_i \text{tim} \cdot x_i \leq L_m \quad \forall m \rightarrow$ Capacità sulle risorse

$\sum_j a_{ij} y_j \leq x_i \quad \forall i \rightarrow$ α_{ij} = COEFFICIENTE

$y_j \leq \bar{d}_j \quad \forall j$

$x_i, y_j > 0 \quad \forall i, j$

↳ INTERE → possiamo farlo con il PL per i grandi numeri

$x_1^* = 116,67 \quad x_4^* = 0$
 $x_2^* = 116,67 \quad x_5^* = 90$
 $x_3^* = 28,67$
 $y_1^* = 26,67 \quad y_2^* = 0 \quad y_3^* = 90$

non posizionarsi su posizioni estreme e si ha perdere nel profitto.

Il profitto è un valore atteso e non ha senso confrontarlo con un valore noto che è 3233,33.

Poiché valuto un valore atteso posso confrontarlo solo con un valore atteso che ottengo usere i valori delle x in ogni scenario

↳ RISOLVO TRE PROBLEMI DEL TIPO

$$\max \sum_j p_j y_j^s \rightarrow \text{I costi non li pago perché ho già prodotto } x_i^*$$

s.t. $\sum_j a_{ij} \cdot y_j^s \leq x_i^* \quad \forall i$ le x diventano dei parametri

$$y_j^s \leq d_j^s \quad \forall j$$

$$y_j^s \geq 0 \quad \forall j$$

$$R_j^s = \sum_j p_j \cdot y_j^s$$

profitto per ogni scenario

è più basso di -3233,3 e anche del valore di -2885,71 (sol + BASSA DEL PREVISTO) ↑

PROFITTO COME MEDIA PONDERATA $\rightarrow \sum_s \pi^s R_j^s - \sum_i x_i^* c_i = 2333,3$

come si comporta in media la soluzione negli scenari futuri

Quando il vincolo di capacità è molto stringente si può non tener conto dell'incertezza perché posso produrre talmente poco che produco qnt poco → gli scenari non possono

* MODELLO STOCASTICO A DUE STADI CON RICORSO

- x → VARIABILI I STADIO → decisioni di qnt prod. Adesso
- y → VARIABILI II STADIO → non poi produrre adesso.

2 stadi: 2 livelli decisionali ↓
 GUIDA A PRENDERE la DECISIONE quando lo scenario si è realizz.
 (in base ai valori di domanda che si sono realizzati)

produco i componenti e quando il cliente ordina assemblo, prendo la miglior combiaz. di componenti già fatti.

A Ricorso perché → prendo delle decisioni e quando l'incertezza non c'è + (i valori di domanda si sono realizzati) rivedo le mie decisioni.

l'effetto dell'incertezza è qll di AUMENTARE le variabili decisionali in gioco. All'aumentare degli scenari → aumentano le variabili di II stadio

molti mezzi di trasporto hanno orari vincolati. (freno, aereo, nave)

Pensando al trasporto su strada → non è limitato e non abbiamo vincoli

I trasporti possono essere

PUNTO A PUNTO

partono da un origine e arrivano ad una destinazione

quando si calcola il costo di tale politica saltano fuori approcci simili al lotto economico

distruzione → fissato × trasport. punto a punto

NON PUNTO A PUNTO

Ricerca un routing tra vari punti della rete che non hanno approccio EQ

↓
VEICOLI ROUTING
Instradamento di veicoli

↓
VARI APPROCCI di soluzione in qst problemi molto complessi

I trasporti sono

FULL TRUCK

Se non sono nostri non è + compito nostro gestire la capacità ma del terzo

↓
PIENO CARICO O NO rientra nelle nostre decisioni di costi in base alle tariffe del trasportatore

NON FULL TRUCK

Se sono nostri i comuni generali. cerchiamo di usare una logica a pieno carico

MEZZI DI TRASPORTO (freno, aereo, nave)

→ usando tali mezzi siamo vincolati da aeroporti, porti etc. e dagli orari

FUSSI INFORMATIVI IN UNA RETE

↳ generalmente vanno da valle a monte → si parte dalla previsione della domanda.

INEFFICIENZE

avere info aggiornate a monte è molto difficile

↓
• chi le possiede non le vuole dare perché si pensa che qualcuno potrebbe usarle contro di lui. (es. fornitore - competitors)

• tutti vogliono tenere le info e tutti vogliono prendere la decisione

le distorsioni della catena logistica si propagano da valle a monte creando scorte esagerate

→ Per il problema di non voler condividere le informazioni

↓
si creano situazioni di inefficienze

("adesso 30 anziché 10 pr al fornitore perché voglio fare una promozione e aumento artificialmente la domanda ma

In base alle mie disponibilità dei mezzi di trasporto variano i costi e anche la strategia migliore → non posso decidere quale è il punto migliore solo in base alle distanze.

Quando scelgo dove posizionare, oltre ai costi che conosco

cerco di anticipare anche i costi futuri della scelta che sto implementando (ANTICIPAZIONE)

di modo da arrivare a livello operativo con tutti i costi noti

(decisione strategica in funzione della decisione operativa)

APPROCCI PULL & PUSH → sono CARATTERISTICHE di una soluzione

Al di là dei parametri che uso, come funziona il mio sistema, la mia politica?

Non la soluzione stessa!

PULL? PUSH? MAKE TO STOCK o MAKE TO ORDER?

I sistemi reali sono ibridi, soprattutto quelli logistici. la stessa soluzione, a seconda dei contesti può assumere caratteristiche di pull e push.

la vera differenza tra pull e push → Non esistono sistemi "pur" push o pull

Pull → i cambiamenti di stato del sistema sono dovuti ad 1 cambiamento di stato del sistema

tutte le conseguenze di cambiamenti di stato nel sistema

→ c'è un'azione e' innescata dall'interno del sistema

PUSH → Qualunque cosa accade perché ho un piano da seguire

es. ho previsto la domanda e produco in base a questo. → azioni generate dall'esterno

EOO → succede quando arrivo ad 1 certo livello → PULL (in funzione dello stato del sistema)

→ previsione della domanda → PUSH (pianificazione)

MPP → push → determino gli ordini in funzione del tempo di domanda atteso

→ pull → nel funzionamento

KANBAN → pull → si basa su 1 cambiamento interno

→ push → determiniamo il n° di kanban cercando di prevedere la domanda

MPS → push → se mi serve implemento il piano

→ pull → se non mi serve mi adatto al cambiamento di stato interno

Qui sistema reale è ibrido ad esempio: il contratto è in funzione della domanda, la consegna dipende dallo stato del sistema. (la consegna segue la domanda)

PROGETTO DI RETI & DI PROBLEMI DI TRASPORTO

CAP. 2

(NO2.42)

Progetto di reti → decisione strategica: dove posizionare i nodi?
 problemi di trasporto → quali collegare (i nodi), come farlo.

In termini di riprogettazione delle reti si cerca di dividere (separare) i problemi.

- ↳ si considerano i punti vendita come dati → FISSO
- ↳ si considerano IRRILEVANTI le DISTANZE (Solo x i costi)
- ↳ si vede la domanda finale come indep dalle scelte prese (ESOGENA)
- ↳ si vede la domanda finale come AGGREGATO

RETAIL MANAGEMENT → si considerano fissati nodi, magazzini centrali

↳ la domanda dipende dalle scelte che faccio → non è + ESOGENA
 ↳ conseguenze delle scelte che faccio

↳ le distanze non sono + IRRILEVANTI → le distanze influenzano sul livello di servizio al cliente (influenzano i tempi di rifornimento)

IL PROGETTO DELLA RETE → tipico strategico anche se è un progetto.

PARCOI logistici (agglomerati di capannoni affittati) → specializzazioni dei sistemi logistici (flessibilità attuale) → decisioni meno immediate che mi permettono di ridurre l'orizzonte di pianificazione → MAGGIORE FLESSIBILITÀ

I PROBLEMATICA: Voglio o no i centri di stoccaggio intermedi?

UTILITÀ di un CENTRO

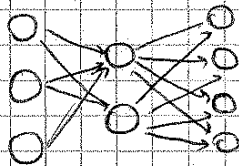
di primo acconto lo vedo come un costo (il prod. non acquista altro valore) e genera un costo di mag.

- ↳ sfruttamento ottimale di scale → RISPARMIO
- ↳ abbattere l'incertezza della domanda finale

↳ Trasferisco dirett. o metto strada intermedi? Quanti?

Come Valuto i pro e i contro??
 A priori imposto un problema

esempio: RETE A 3 livelli → gestisco con una logica (Q,R) Ordino le quantità di ogni volta che il livello di scorte scende a R e ho un costo di ordine A e di mag. h.



suppongo di avere m retail $j=1 \dots m$

domanda indipend. ← e una domanda d_j per ogni centro $d_j \sim N(\mu, \sigma^2)$

$d_j \sim N(\mu, \sigma^2)$ → domanda durante il LT del centro j

In questa soluzione abbiamo assunto che:

- i lead time sono tutti uguali
- è uguale il trasferimento dal mag. centrale o dal mag. intermedi (in termini di costo)

NELLA REALTÀ → **ECONOMIA sui costi**

La soluzione → permette di sfruttare l'economia di scala con MAG. CENTRALE e di abbattere l'incertezza aggregando le varianze → **RISK POOLING**

tutto funziona bene se il cliente accetta di attendere x i tempi di consegna, x noi è considerato soddisfatto, in realtà attende
 ↳ ALLUNGO IL LEAD TIME

ma ho un **EFFETTO NEGATIVO** sul livello di servizio al cliente

se le varianze non sono indipendenti → uso il coefficiente di **CORRELAZIONE** ρ

$$\sum_j \sigma_j \geq \sqrt{\sum_j (\sigma_j^2 + 2\rho\sigma_1\sigma_2 \dots)}$$

se la correlazione è negativa ho

↳ correlazione
 • positiva
 • negativa

$$\sum_j \sigma_j \geq \sqrt{\sum_j (\sigma_j^2 - 2\rho\sigma_1\sigma_2 \dots)}$$

che è ancora minore del fattore precedente → ok

se la correlazione è positiva: al massimo (caso peggiore) $\rho = 1$

quindi

$$\sigma_1 + \sigma_2 \geq \sqrt{\sigma_1^2 + 2\sigma_1\sigma_2 + \sigma_2^2}$$

$$\sigma_1 + \sigma_2 \geq \sqrt{(\sigma_1 + \sigma_2)^2} \rightarrow \sigma_1 + \sigma_2 \geq \sigma_1 + \sigma_2$$

GRAFO: Nodi, archi → ha un'origine e una destinazione gli archi collegano i nodi

RETE → Nodi, Archi ↳ **GRAFI DIREZIONATI** (VERTICI, COSTI)

↳ grafo che può avere i nodi aggiuntivi, sui link o sui nodi

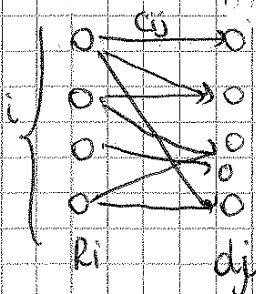
gestiscono molto male ↳ ecco perché parliamo di **RETI LOGISTICHE** & non di grafi

Il problema dell'incertezza della domanda - lavoreremo → per trattare l'incertezza con i valori medi o è tutto deterministico ↳ **MODELLI A LINEE CON SCENARI**

↳ I MODELLI CHE TRATTIAMO

- NON TRATTO INCERTEZZA
- COSTI NON LINEARI

↳ semplifichiamo i costi non lineari



ogni collegamento ha un suo costo che dipende dall'origine e dalla destinazione: deve soddisfare la domanda d_j di ogni centro con un determinato massimo (capacità + semplice) sapendo che ha capacità R_i produttiva

PROBL. STATICO

IGNORO LA DINAMICA TEMPORALE DELLA DOMANDA

è lo stesso come $x_{ij} \leq |R_i| z_{ij}$

PROBABE A FARE prob. trasporto con costi di produg, con dimamente,
 con costo fisso di trasporto e nel caso di rete multiprodotto
 se transportation (vincoli) tattici \rightarrow plant location } è un problema strategico
 min cost flow

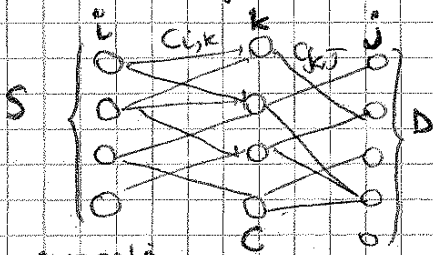
19 OTTOBRE 2011

PROBLEMA DI FLUSSO A COSTO MINIMO

\rightarrow voglio sempre min il costo di un flusso ma poi avere più livelli della rete di distribuzione ma i collegamenti sono soggetti a dei vincoli

DIFFERENZA CON PROBLEMA di TRASPORTO

- nodi di transshipment (nodi tra sorgenti e destinazioni)



$c_{i,k}, g_{k,j} \rightarrow$ costi dovuti ai nodi intermedi (punti di transshipment)

- S: insieme delle sorgenti
- D: insieme delle destinaz.
- C: insieme dei nodi intermedi

\rightarrow usati da variabili decisionali \neq

sorgenti con capacità R_i e costo p_i di produzione

nodi destinazione \rightarrow domanda d_j

\rightarrow hanno limite di capacità $H_k = \max$ volume movimentabile

i $h_{i,k}$ hanno dei vincoli \rightarrow supponiamo che tali vincoli siano di volume $V_{i,k} =$ volume che può essere portato da k a i $W_{k,j} =$ volume da k a j

MULTIPRODOTTO

\rightarrow $c_{i,k}$ e $c_{i,j}$ diventano costi per unità di volume } dipendono non solo dalla distanza

chiamo L insieme di prodotti e chiamo l l'indice di prodotto e poiché i prodotti sono diversi indico con

- $r_{i,l}$ = quantità di risorsa i per produrre l
- $p_{i,l}$ = costi di produzione
- v_{l} = volume di prod l

\rightarrow introdotto il multiprodotto questo si ribatte sui costi di produzione e di conseguenza sulla DOMANDA

MODELLO \rightarrow minimo costo
 \rightarrow max soddisfazione della domanda
 \rightarrow vincoli su movim, trasporto, capacità

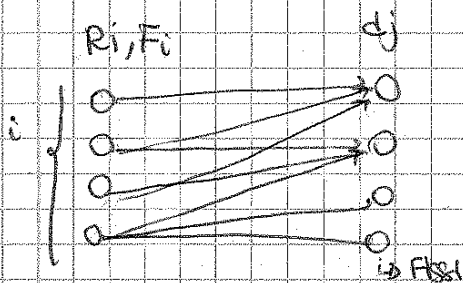
$x_{i,k,l}$

\rightarrow sono 2 set di variabili diverse

$y_{k,j,l}$

\downarrow
 quantità di prodotto l realizzato nello stabilimento i , trasportata verso il centro di distribuz k

\rightarrow quantità di prodotto l trasportata nel centro di distribuzione k al punto vendite j



VARIABILI DI PROGETTO

Se avessi solo problema strategico aprire i siti che costano meno e che producono a sufficienza * soddisf. la domanda

→ esistono poi però degli altri costi (trasporti) che dipendono dal link che utilizza e che dipendono dai che siti ho aperto

→ devo ANTICIPARE quindi i costi di trasporto * non fare scelte avventate (il costo di apertura del sito avviene poi un SUNK COST)

↑
NODI SU CUI DECIDO (Quali aprire?)

poi devo decidere in base ai costi di trasporto come soddisfare la domanda di tutti i centri

Qui sito ha un costo di apertura F_i e una capacità R_i

↳ costo nell'unità di tempo dello stesso materiale di c_{ij}

MONOPRODOTTO

x_{ij} = qte di prodotto (continuo) da i a j

y_i → 1 se apro stabilim. i → 0 altrimenti

VARIABILE DECISION. LOGICA

NON CI SONO PERICOLI e' un problema statico (modello per unita' di tempo)

$$\min \sum_i F_i y_i + \sum_i \sum_j c_{ij} x_{ij}$$

s.t. parte strategica di costo

↳ se il link si attivera' ci parra' un costo (non strategico) ma tutto il costo di APERTURA e' un costo una tantum (investimento)

$$\sum_i x_{ij} = d_j \quad \forall j \rightarrow \text{soddisf. delle domande}$$

$$\sum_j x_{ij} \leq R_i \quad \forall i \quad (2)$$

$$x_{ij} \geq 0 \quad \forall i, j$$

$$y_i \in \{0, 1\}$$

$$x_{ij} \leq R_i y_i \rightarrow \text{vincolo che funziona ma e' + debole che e' dominato da (2)}$$

↳ Usandolo singolarmente non include il problema tattico. il costo di apertura deve diventare un costo per unita' di tempo (AMMORTIZZATO) COSTO AMMORTIZZATO la natura di tale parametro deve essere tattica.

se $y_i = 1$ → qui x può essere max par alla sua capacità

Non pagherò i costi di trasporto con qte modello e' solo una stima a Spaulle

↳ pagamenti dall'esterno

LA DOMANDA e' ASSUNTA ESOGENA

In ambito RETAIL la domanda e' endogene (interna, derivata alla prod.)

MONOPRODOTTO con DOMANDA NON SODDISFATTA

$$\min \sum_i f_i y_i + \sum_s u^s \sum_i \sum_j c_{ij} x_{ij}^s + \sum_s \pi^s \sum_j \beta_j z_j^s$$

penalite' del centro j
(costo di NON SODDISFACIMENTO)

voglio spendere x ad essere necessariamente $= d_j$

$$\sum_i x_{ij}^s + z_j^s = d_j^s \quad \forall j, s$$

$$\sum_j x_{ij}^s \leq R_j y_j \quad \forall s, i$$

VARIABLE DI SLACK:
voglio comunque il soddisfacimento delle domanda in diversi modi in modo diverso
voglio che il cap x_{ij} per essere anche $< d_j$ ma voglio soddisfarlo

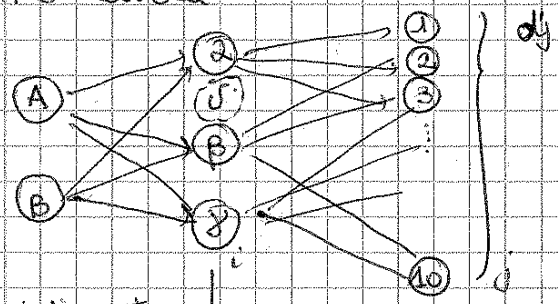
IPOTESI e che ogni periodo sia uguale al successivo
↳ MODELLO STATICO

Minimizzando fluttuazioni della domanda nei periodi devo introdurre t

ATTENZIONE agli stadi decisionali e a come questi siano rapportati con i periodi

Mettiamo insieme le decisioni del trasporto e di riprogettazione delle rete - **RILOCALIZZAZIONE & ESPANSIONE CAPACITA' CENTRI**

RETE A 3 livelli



dato d_j vogliamo sapere se tenere α, β, γ e se voglio aumentare γ o aprire un δ

• se apro δ , α o β devono essere chiusi (MAX 3 centri attivi)

• Capacita' di massimizzazione
↳ H_i
N.B. = i-link sono POTENZIALI e legati alla nostra decisione strategica

OPZIONI

la decisione e' vincolo: max 3 map. centrali

- γ → aumento la capacita'
- δ → lo apro? con quale capacita'?
- α, β → li chiudiamo?

LOWER
UPPER

↳ tale decisione a livello centrale mi fara' cambiare anche i costi dello stato di trasporto che saranno condizionati da come ho configurato la rete.

↳ imponiamo il soddisfacimento della domanda

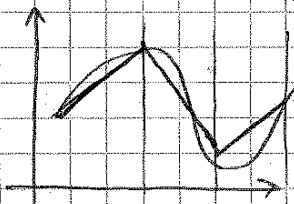
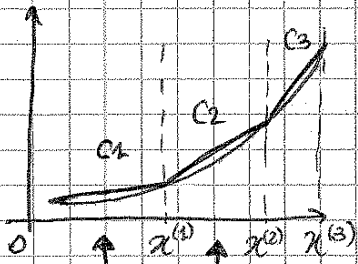
COSTI DI TRASPORTO NON LINEARE → questo è la realtà!

Lineare a tratti
LA FUNZIONE

costi dipendono dalla q.tà

non ha più funzioni
CONVESSE

→ minimizzarle
diventa difficile



potrebbe avere degli OTTIMI LOCALI

perché se
funz è CONCAVA?

→ però usare dei
metodi di
risoluzione
con uno sguardo
globale

In ognuno di questi
tratti definisce
un COSTO
DIVERSO

$$\begin{aligned}
 0 \leq x \leq x^{(1)} & \quad c_1 x \\
 x^{(1)} \leq x \leq x^{(2)} & \quad c_1 x^{(1)} + c_2 (x - x^{(1)}) \\
 x^{(2)} \leq x \leq x^{(3)} & \quad c_1 x^{(1)} + c_2 x^{(2)} + c_3 (x - x^{(2)})
 \end{aligned}$$

Come faccio ad
includerli in un
MODELLO?

→ dipende da come sono c

dipende se sono la
linearizzazione di una funzione

CONCAVA

CONVEXA

$$c_1 < c_2 < c_3$$

definisce una variabile
CONTINUA per ogni intervallo

CON CONCAVITA'
VERSO
L'ALTO

$$z_1 \in (0, x^{(1)})$$

$$z_2 \in (0, x^{(2)} - x^{(1)})$$

$$z_3 \in (0, x^{(3)} - x^{(2)})$$

definisce una
funzione
di COSTO

combinazione lineare
delle variabili
e dei costi (locali)

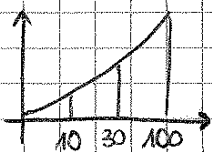
$$f(z) = c_1 z_1 + c_2 z_2 + c_3 z_3$$

$$x = z_1 + z_2 + z_3$$

se minimizzo questa funzione
otterrò la x che mi fa avere
il costo minimo

Questo funziona perché
la minimizzazione di
una funzione convessa
è un problema convesso!

esempio:



z_1 è tra 0 e 10 → e infatti ci sono 10x
che posso pagare c_1

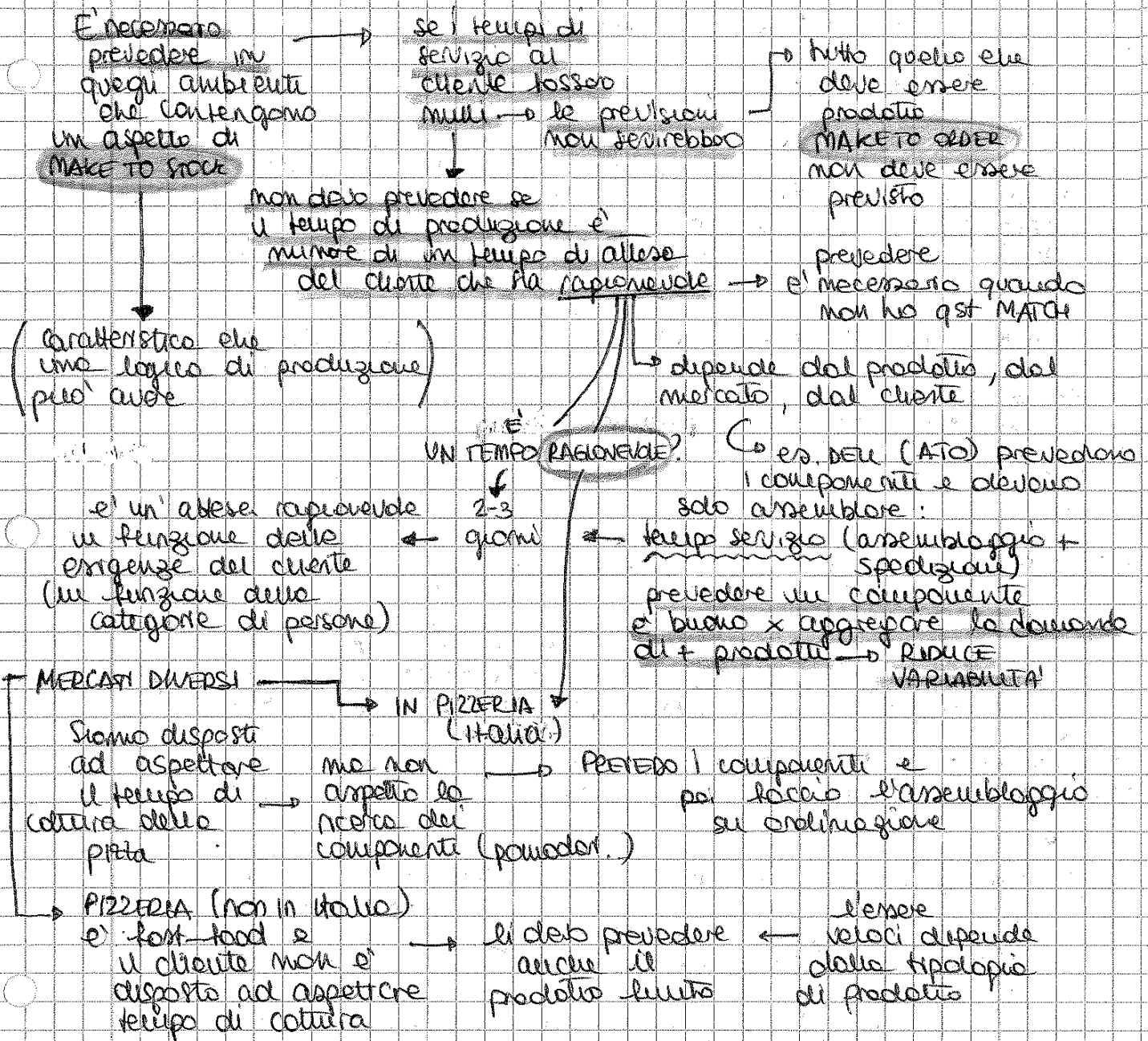
z_2 è tra 0 e $(30-10)=20$ → e infatti ci sono
20 punti tra $x^{(1)}$ e
 $x^{(2)}$ che posso pagare
 c_2

z_3 è tra 0 e $(100-30)=70$ → ...

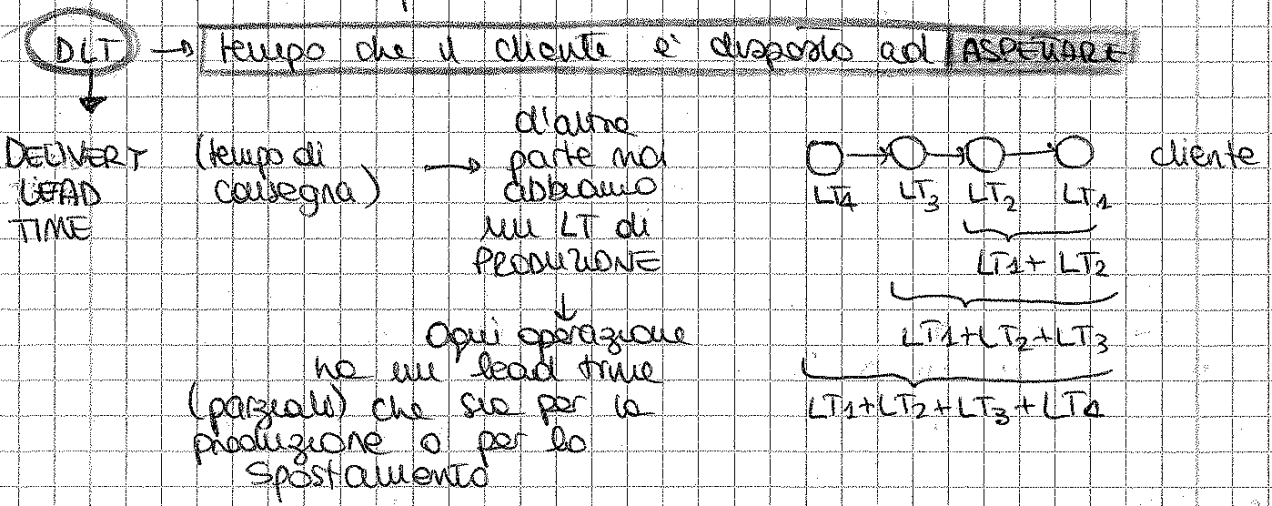
PREVISIONI

CAP 3

25 OTTOBRE



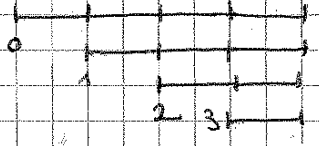
Come facciamo a sapere cosa prevedere?



③ **FREQUENCY** → Ogni quanto RIVEDO le mie previsioni
 frequenza di AGGIORNAMENTO
 RICALCO LE PREVISIONI

ORIZZONTE ROTOLANTE: (aggiorno continuamente le previsioni)
 Ogni 4 settimane rivedo le previsioni per 6 mesi

ORIZZONTE CHE SI ACCORCIA: mantengo sempre l'orizzonte iniziale (orizzonte fisso)



All'equilibrio ogni previsione lo rivedo lo stesso n° di volte

Rivedo la prima settimana 2 volte la 3° 3 volte e così via

se non ho nemmeno info AGGIUNTIVA non ho senso

le PREVISIONI VANNO RIVISTE SOLO IN CASO DI INFO ADDIZIONALI

- LA CONFIGURAZIONE CHE GARANTISCE ERRORI + CONTENUTI e':**
- TIME BUCKET Ampio
 - ORIZZONTI brevi
 - AGGREGAZIONI di prodotti
 - GRANDI MERCATI

④ **PRODUCTS** → i livelli di AGGREGAZIONE possono essere infiniti → lo previsionale può avvenire in AGGREGATO

+ AGGREGO + e' facile prevedere perché' aggrego l'incertezza

+ vado nel dettaglio + e' difficile prevedere

+ le nostre previsioni sono rischiose

⑤ **MARKET** → decido come aggregare nei punti vendita se su base NAZIONALE, REGIONALE, LOCALE

Aggregazione geografica

Aggregando alcuni parametri + precisi rispetto alla disaggregazione dei punti vendita o dei prodotti

E' NECESSARIO CAPIRE QUAL'E' IL PROCESSO DECISIONALE

E' IMPORTANTE sapere a COSA CI SERVONO LE PREVISIONI?

facendo un'AGGREGAZIONE tempo, item, geografico x le PROMOZIONI

↳ qualitativamente miste

nel caso delle PROMOZIONI deve sapere in quale punto vendita di quel determ. prodotto in quell'area geografica

Il cliente x le promozioni viene nel punto vendita e x quel prodotto

Il ritardo del lancio → vendita 0, domanda 0 → inaffetto

in una regione
x smaltire le
scorte

mapas non era nulla
perché la gente conosceva
il prodotto e lo chiedeva
infatti VENDITA ≠ DOMANDA

promozioni
possibile
distorcere i dati

VENDITA ≠
DOMANDA

ANALISI DELLA DOMANDA

ANDAMENTI PARTICOLARI

alcuni prodotti hanno
domanda stagionale

dinamiche della domanda
particolari

Assegni domanda
Stazionaria: il presente
è uguale al passato

in base a quello che ho in
magazzino prevedo la
domanda

**SOLUZIONE METODO E TURING PARAMETRI
DI PREVISIONE**

Selezionare un
metodo di
previsione che si
adatti alla situazione

rendere quanto + possibili
accurati i parametri
(aumento di un E anche non
è il + possibile metodo)

COMPRESIONE
GENERICA del METODO
DI PREVISIONE NON
CONSENTE USO NEI
SISTEMI REALI

CALCOLO ERRORE DI PREVISIONE

Incoerenze
tra modello
e comportamenti
della domanda

lo calcolo sempre ex post
confrontando previsione
e realizzazione del
parametro previsto

tipicamente
→ è sempre piccolo
↓
sottoscrivere la previsione
aggiornare
maschera il reale
errore: dovuto alla
previsione iniziale

**MISURE ERRORE DI
PREVISIONE**

Indicheremo con

Y_t = la realizzazione effettiva di quello che prevediamo
al tempo t (va confrontato con la previsione)

$F_{t,h}$ = previsione fatta in t (momento in cui faccio la
previsione) per il tempo t+h

F_t = previsione per t (della domanda x il periodo t)

con cui confronto la realizzazione

**ERRORE
DI
PREVISIONE**

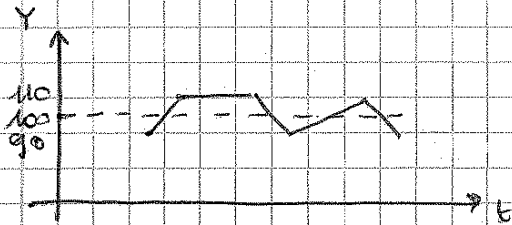
$e_t = Y_t - F_t$ (REALIZZAZIONE - PREVISIONE) → Valore a posteriori

> 0 se la previsione aveva sottostimato la domanda
 $Y_t < F_t$

< 0 la previsione è + alta della
realizz.

nell'errore
non uso $F_{t,h}$ perché non mi interessa quando ho fatto
la previsione, mi interessa la previsione

indipendentemente da quando è stata fatta



Il suo metodo non è accurato → il BIAS è un INDICATORE di DEVIANZA della PREVISIONE e si limita a MISURARE se la PREVISIONE è SOTTOSTIMA (BIAS positivo) o SOVRASTIMA (BIAS negativo) la domanda.

② MEAN ABSOLUTE DEVIATION MAD → indica quanto il metodo non è accurato
 INDICATORE di ACCURATEZZA
 non consente + ad errori negativi di compensare gli errori positivi
 → gli errori si sommano

$$MAD = \frac{1}{n} \sum_{t=1}^n |e_t|$$

Il MAD × l'es. precedente 1

$F_t(1) \rightarrow 10$
 $F_t(2) \rightarrow \phi$

60/10
 che sia grande o piccolo come errore (10 moltip. su una media di 100) non possiamo saperlo

es 2

	1	2	3	4	5	6	ME BIAS	MAD
Y_t	7	13	9	12	8	11		
$F_t(1)$	10	10	10	10	10	10	ϕ	2
$F_t(2)$	6	12	8	11	7	10	(bias) 1	1
$e_t(1)$	-3	+3	-1	+2	-2	+1	ϕ	$12/6=2$
$e_t(2)$	1	1	1	1	1	1	1	1

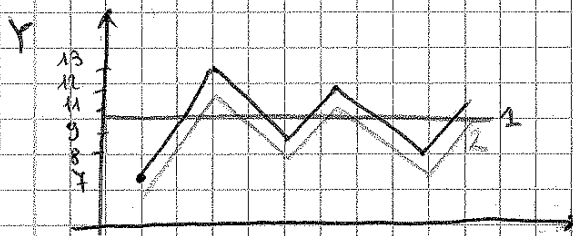
Somma tutti gli errori
 5/3 segno

Il I metodo è meno accurato ma senza BIAS
 Il II è + accurato e ha + BIAS

Quale problema penso di risolvere + facilmente?
 Uno di BIAS o uno di accuratezza?

CORREGGERE devianza e + facile

IL METODO 1



- ACCURATEZZA → segue l'andamento di 1
- ha BIAS → sovrastima o sottoestima la domanda

Il 2° metodo è + accurato e ha BIAS medio di 1
 → risolvo il problema aggiungendo un BIAS medio ad ogni previsione

Compensare l'accuratezza del 1° metodo è + difficile

→ è molto + DIFFICILE COMPENSARE l'ACCURATEZZA pg 75

t	1	2	3	4	5	6	ME	MAD	RSME	MDE	MAPE
Y_t	7	13	9	12	8	11					
F_t	8	12	10	11	7	12					
Y_t	70	130	90	120	80	110					
F_t	75	125	95	115	75	115					
$e_t(1)$	-1	+1	-1	+1	+1	-1	0	1	1	-1%	-0,57%
$e_t(2)$	-5	+5	-5	+5	+5	-5	0	5	5	10,5%	5,3%

Entrambi sono senza BAYAS ma d'impulso sceglierei il I perché ha meno MAD e RSME

6) tutte queste misure non sono ADIMENSIONALI → sono dimensionate → DEVO RAPPORTARE quelle MISURE a
 ↳ errore percentuale sulla domanda è diversa
 ↳ Spazio 1 su 10 (nel I)
 ↳ Spazio 5 su 100 (nel II)
 ↳ non posso confrontarle senza rapportarle → introduco degli indici PERCENTUALI
 ↳ INDICATORI DI TIPO RELATIVO

④ $MPE = \frac{1}{n} \sum_{t=1}^m \frac{e_t}{Y_t}$
 MEAN PERCENTAGE ERROR

INDICATORE DI DEVIAZIONE
 ↳ quanto sbaglia IN RAPPORTO a quello che si è realizzato

⑤ $MAPE = \frac{1}{n} \sum_{t=1}^m \frac{|e_t|}{Y_t}$
 MEAN ABSOLUTE PERCENTAGE ERROR

questi coefficienti sono necessariamente ADIMENSIONALI → SONO NUMERI PURI
 ↳ INDICATORE DI ACCURATEZZA

$MPE_1 = \left[-\frac{1}{7} + \frac{1}{13} - \frac{1}{9} + \frac{1}{12} + \frac{1}{8} - \frac{1}{11} \right] \frac{1}{6} = -1\%$

$MPE_2 = \left[\frac{1}{7} + \frac{1}{13} + \frac{1}{9} + \frac{1}{12} + \frac{1}{8} + \frac{1}{11} \right] \frac{1}{6} = 10,54\%$

$MPE_2 = \left[-\frac{5}{75} + \frac{5}{125} - \frac{5}{95} + \frac{5}{115} + \frac{5}{75} - \frac{5}{115} \right] \frac{1}{6} = -0,57\%$

2) è + ACCURATO del I metodo

ANCHE QUESTI METODI HANNO DEBOLEZZE

* Posso usare questi metodi solo se $Y_t \neq 0$ → non posso dividere per 0

AGGREGANDO SARA' + FACILE che $Y_t \neq 0$
 ↳ quanto + disaggrego tanto + aumenta la probabilità che $Y_t = 0$

Questi indicatori possono essere gestiti / manipolati artificialmente
 che i previsioni scegliendo MPE e MAPE concentrandoci su periodi di domanda basse.

ME = 0 → indica che l'errore non è affetto da bias

Con questi metodi è difficile confrontare diversi settori di domanda stream di dati diversi.

Un singolo metodo non deve dare risultati diversi in relazione alle previsioni con la difficoltà del processo.

Introduciamo

ME %, MAD %, RMSE %

che siano adimensionali
 che simulano il modello della domanda medio adimensionali

$$\%ME = \frac{ME}{\bar{Y}}$$

domanda medio - mi permette di simulare gli errori della domanda dividendo x \bar{Y}

$$\%MAD = \frac{MAD}{\bar{Y}}$$

valuto indip dal volume di domanda

$$RMSE = \frac{RMSE}{\bar{Y}}$$

se la MEDIA \bar{Y} è nulla non posso avere previsioni % Ma il previsionista non deve preoccuparsi maggiormente dei problemi a domanda basse

\bar{Y}	1	2	3	4	5	6	ME	MAD %	MAPE	MPE	RMSE
Y_{LA}	10	9	10	11	10	10					
F_{LA}	9	10	11	10	9	11					
etA	1	-1	-1	1	+1	-1	0	1/10			1
Y_{LB}	15	8	5	12	13	7					
F_{LB}	14	9	7	10	12	8					
etB	1	-1	-2	2	1	-1	0	8/10			1,4

$$ME_A = \frac{1}{6} (1 - 1 - 1 + 1 + 1 - 1) = 0$$

$$MPE = \frac{1}{6} \left(\frac{1}{10} - \frac{1}{9} - \frac{1}{10} + \frac{1}{11} + \frac{1}{10} - \frac{1}{10} \right) = -0,0033$$

$$ME_B = \frac{1}{6} (1 - 1 - 2 + 2 + 1 - 1) = 0$$

$$MPE_B = \frac{1}{6} \left(\frac{1}{15} - \frac{1}{8} - \frac{2}{5} + \frac{2}{12} + \frac{1}{13} - \frac{1}{7} \right) = -0,059$$

$$MAD_A = \frac{1}{6} (1 + 1 + 1 + 1 + 1 + 1) = 1$$

$$MAPE_A = \frac{1}{6} \left(\frac{1}{10} + \frac{1}{9} + \frac{1}{10} + \frac{1}{11} + \frac{1}{10} + \frac{1}{10} \right) = 0,1$$

$$MAD_B = \frac{1}{6} (1 + 1 + 2 + 2 + 1 + 1) = \frac{8}{6} = 1,33$$

$$MAPE_B = \frac{1}{6} \left(\frac{1}{15} + \frac{1}{8} + \frac{2}{5} + \frac{2}{12} + \frac{1}{13} + \frac{1}{7} \right) = 0,163$$

$$RMSE_A = \sqrt{\frac{1}{6} (1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2)} = 1$$

$$RMSE_B = \sqrt{\frac{1}{6} (1^2 + 1^2 + 4^2 + 4^2 + 1^2 + 1^2)} = \sqrt{2} = 1,4$$

Se $Y_t = 0$ la STATISTICA si traduce in

$$MAD \% = \frac{\sum_{t=1}^{M-1} |Y_t - Y_{t+1}|}{Y}$$

e lo si rapporta al metodo "stupido"

Ogni azienda si crea la sua misura di ERRORE che dipende dai costi che l'azienda paga nel caso in cui sbaglia la previsione.

UTILIZZI DEGLI INDICATORI DI PERFORMANCE (indici)

- uso misura di barto ON-OFF → se l'errore è inferiore al 30% va bene altrimenti la previsione non è sostenibile
 ↓
 percentuale di volte che la differenza tra domanda e previsione è < 30%
 ↓
 su n periodi che testa cento su quanti sono neutri nell'errore del 30%
 ad esempio perché ricalcoliamo i costi dovuti agli errori con le promozioni che ci facciamo 30% in + dalle vendite

↳ non ho info su sottostima o sovrastima ma solo le distorsioni

- **ME** Calcolare l'errore di serie x capire quanto il nostro metodo funziona e alle variazioni della domanda deve essere flessibile → può capire che a variazioni di domanda il metodo non si adatta

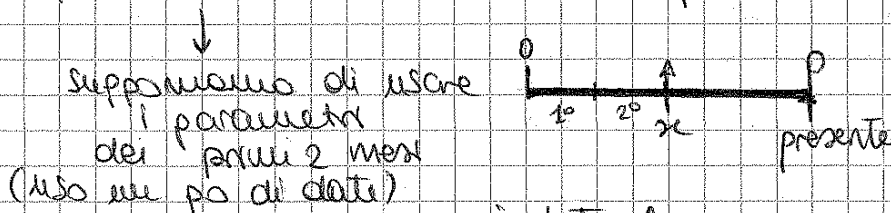
- I metodi di ordine nientono della variabilità della domanda e quindi l'errore o il calcolo dell'errore mi dà una stima di tale variabilità (incertezza)

- Gli errori vengono usati x premiare / punire qualcuno

↳ attenzione ad usare incertezze giuste

COME MISURARE LE PRESTAZIONI DI UN METODO DI PREVISIONE

- 1) Scelgo dei parametri, a tentativi, per le mie previsioni e misuro le prestazioni del metodo durante un periodo di prova
- 2) faccio una stima su i dati passati e uso questi dati x settare i nuovi parametri → **USO LA STORIA PASSATA PER VERIFICARE LE PRESTAZIONI DEL METODO**



↳ uso i dati fino a x e li uso x vedere le prestazioni i

metodi QUANTITATIVI

CHIEDONO LA FORMAZIONE DI IPOTESI ESPLICITE CIRCA IL COMPORTAMENTO DELLA DOMANDA

- Sono CONSISTENTI TEMPORALMENTE
- Non influenzati
- trattano ugualmente nel dettaglio e nell'aggregato
- meno flessibili

l'ideale sarebbe estrarre le tipologie di metodi in parallelo (qualitativo e quantitativo che si integrano)

↳ uso un metodo quantitativo per dare all'esperto una base di partenze (CONSISTENZA di un Algoritmo)

metodo l'esperto nel metodo: il suo giudizio diventa una variabile del metodo (FLESSIBILITA' MENTE UMANA). Lo es: METODO DELLA REGRESSIONE

METODI QUANTITATIVI

MEDIA MOBILE

→ l'unica cosa che va scelta è il k

ANALISI dell'ANDAMENTO STORICO della domanda

→ tale metodo assume che la domanda sia stazionaria

→ nessun trend
nessuna stagionalità

→ il passato = futuro

PROCESSO di DOMANDA

$$Y_t = E(d_t) + \epsilon_t$$

→ la media più o meno sul E_t tale per cui il suo valore atteso $E(\epsilon_t) = 0$ è nullo

Parlando di domanda STAZIONARIA

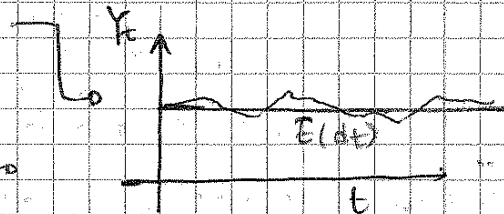
La domanda attesa

termine di errore

↳ in media non ho nessun aumento e nemmeno diminuzione

↳ aspettiamo lievi fluttuazioni rispetto al valore medio

→ che in media non si allontanano dalla linea centrale



ragionando per periodi

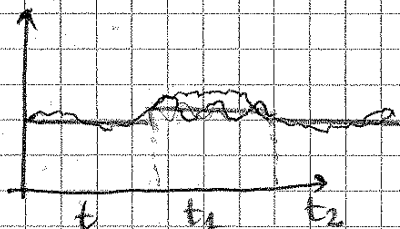
il mio valore atteso in t potrebbe essere diverso dal valore atteso di t+1

→ richiediamo allora che le variazioni della media siano in media nulle

STAZIONARIETA':

la domanda media tra i periodi non deve essere molto diversa, non è dovuto essere basata

E le variazioni della media devono essere in media nulle



→ il VALORE atteso di tutto che oscilla, deve essere nullo

$$E(d_t) = E(d_{t-1}) + \epsilon_t$$

$$E(\epsilon_t) = 0$$

NO BIAS né tra i periodi né all'interno del periodo

la media delle MEDIE è costante

con $k=6$ → la previsione è + stazionaria → la stabilizza
 Ci aiuta a mantenerla

k piccolo aggiunge l'attenuazione su fluttuazione → aggiunge alle fluttuazioni della domanda, quello della previsione

con $k=2$ → mezzo la previsione da 3 → per un piccolo alto di domanda registra un piccolo di prev.

per k è grande + il piccolo è basso anche se dura di +.
 su $k=6$ il filtro meglio l'errore su 6 il piccolo si vede meno.
 (il piccolo lo media su 2 periodi)

ma sappiamo che parlo a prevedere da + avanti

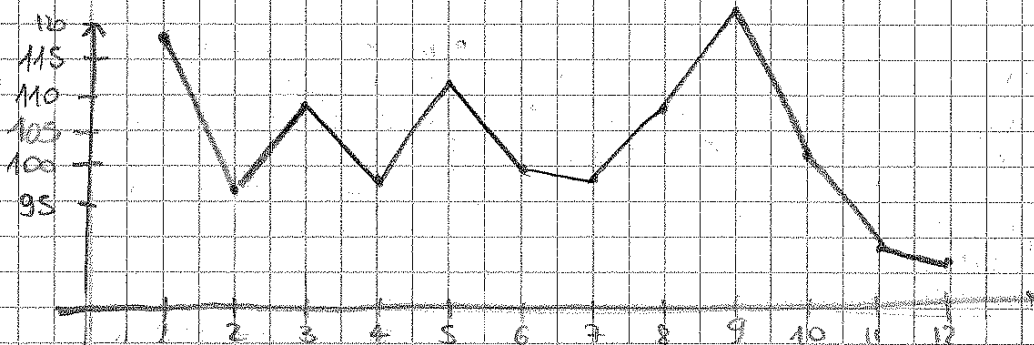
In entrambi i casi la risposta al fenomeno è con ritardo.

esempio

t	1	2	3	4	5	6	7	8	9	10	11	12
Y_t	116.36			99.9	110.31	99.88	98.07	107.38		121.21	89.63	
t	1	2	3	4	5	6	7	8	9	10	11	12
Y_t	116.36	96.30	109.65	110.31	99.88				102.99			98.43

$k=3$?
 $k=6$?

$k=3$ → 3 sono



$$B_3 = \frac{116.36 + 96.30 + 109.65}{3} = 77.43$$

se l' α è basso il mio punto iniziale è + basso curvi un poco elevatissimo

$$B_t = F_{t,h} = B_{t-1} + \alpha (Y_t - B_{t-1})$$

↳ Bene dice che non vuol dire un altro! \Rightarrow

Assumo $h=1$ siamo int

$$F_{t,1} = F_{t-1,1} + \alpha (Y_t - F_{t-1,1}) \Rightarrow \boxed{F_{t+1} = F_t + \alpha (Y_t - F_t)}$$

la previsione che faccio per domani è = alla previsione che ho fatto x oggi + l'errore

↳ se avevo sottovalutato la domanda in periodo $Y_t - F_t < 0$ → la mia previsione attuale

(Vedi ing)

α = PARAMETRO DI SMORZAMENTO

di quanto aumento/diminuisce me lo dice α ← diminuisce

(filtro del rumore)

smorza event: BASSO → AUTO: REATTIVITÀ

Questo VA BENE nell'ottica di processo stocastico ma non perfetto

Oltre all' α devo determinare il punto da cui partiamo

↳ LA MEDIA MOBILE NON È RECURSIVA

↳ LO SMORZAMENTO ESPONENZIALE È RECURSIVO

2 NOVEMBRE 2011

Abbiamo esplicitato

$$B_t = \alpha Y_t + (1-\alpha) (\alpha Y_{t-1} + (1-\alpha) B_{t-2}) - \alpha Y_t + \alpha (1-\alpha) Y_{t-1} + (1-\alpha)^2 B_{t-2}$$

dei B ma ce ne possiamo sbarazzare → un valore di B iniziale deve essere fornito B_0

Recursivo → quando leggiamo in modo progressivo un indice α al tempo

B_0 e α sono i parametri di questo metodo → sono IMPORTANTI x il funzionamento

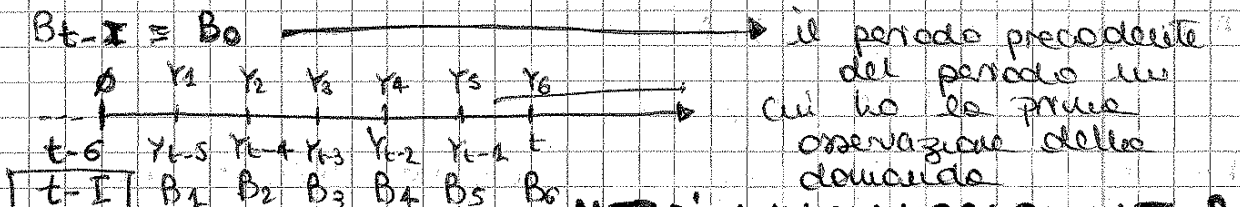
↳ settare: α scelgo un valore che mi piace
 ↳ inizializzare: scelgo B_0 , valore iniziale

2 GRANDE segue bene le oscillazioni
 2 PICCOLO filtro bene eventuali errori

↳ posso vedere 2 periodo nei periodi in cui la domanda è stazionaria

↳ quando la domanda varia MASSIMAMENTE

segue meglio le piccole oscillazioni



METODI di INIZIALIZZAZIONE B_0

① $B_{t-I} = \emptyset = B_0$

VANTAGGIO: di fatto non uso nessuno dei dati storici, nemmeno Y del periodo x settore B

↳ NO FIT SAMPLE: non ho bisogno di dati x fissare i parametri
 TEST SAMPLE: grande

obiettivo è prevedere

$F_{t+1} = F_{t+1} = B_t = B_0$ - funzione di B_0

partendo da \emptyset calcoleremo il B anche x questi periodi

↳ SVANTAGGIO: partire dal livello \emptyset di domanda significa inserire in BAYES. Tale scelta avrà un grosso impatto.

livello iniz. deviato ← con I piccolo e I grande può essere l'osservaz. aggregata che pesa di +, ha grande impatto

② $B_{t-I} = Y_{t-I+1}$

($B_0 = Y_1$)

valore dello I osservazione disponibile

non posso dire se ho un bias

↳ poco errore di I ma tale valore potrebbe essere:
 - esattamente il valore medio
 - può essere scostato di poco dalla media
 - " " molto sbagliato.

↳ la valutazione di accuratezza possa farlo dal periodo 2

se $B_{t-I} = F_{t-I}, h = F_{t-I+h} = F_{t-I+1}$

la previsione del I periodo è esatta. la domanda che ho avuto nel I periodo

↳ I domanda non può essere usata per la valutazione

e. $t+I+1 = \emptyset$ → è' avio

③ $B_{t-I} = \sum_{i=t-I+1}^{t-I+I} \frac{Y_i}{I}$

I osserv. disponibile

↳ scelgo i primi I valori per inizializzare la prima previsione
 ↳ come media di valori "grandi" (e valori)

non c'è B_t e non c'è B_{t-I+1}

↳ se li uso entrambi → NON STO USANDO SMORZZAMENTO ESPONENZIALE

il valore BAYES UN'E B_{t-I+1} viene smorzato meno volte

↳ RAPPRESENTA MEGLIO LA MEDIA DELLA DOMANDA

il valore iniziale peserà tanto di meno quanto più I è grande $(1-\alpha)^I B_{t-1}$

3 NOVEMBRE 2011

$$B_t = \alpha Y_t + \alpha(1-\alpha)Y_{t-1} + \alpha(1-\alpha)^2 Y_{t-2} + \dots + (1-\alpha)^{t-1} B_{t-1}$$

○ Nel grafico 3.16 di uno ipesi di qst osservazioni
dimostriamo che la sommatoria di tutti questi pesi è uguale a 1: SERIE GEOMETRICA

↓
 B_t previsione fatta in t

$$\begin{aligned} & \rightarrow \sum_{i=0}^{t-1} \alpha(1-\alpha)^i + (1-\alpha)^{t-1} = \\ & = \alpha \left[\frac{1-(1-\alpha)^t}{1-(1-\alpha)} \right] + (1-\alpha)^{t-1} \end{aligned}$$

↑ aumentando α , aumento il peso dell'osservazione

con $\alpha = 0,2 \rightarrow$ le y via via + vicine \rightarrow l'altro meno rispetto a y_0 ma sepuò + la domanda
 con $\alpha = 0,5 \rightarrow B_0$ peserà pochissimo

○ La scelta dell' α impatta sul peso che avrà l'inizializzazione B_0 , il variare degli α e dei metodi con cui scegliamo B_0 varia la previsione.

A seconda di come inizializzo $B_0 \rightarrow$ fig 3.17 **I METODO**

↳ con α grande bias ~~metodo~~ ancora tollerabile
 con α piccolo bias enorme
 ↳ $B_0 = 0$ affetto da Bias

II METODO 3.18

$B_0 = Y_1 \rightarrow$ non possiamo parlare di BIAS

(stiamo supponendo che le osservazioni delle domande sono casuali)

○ α piccolo: dimentico lentamente il valore iniziale
 Scegliendo un valore distante dalla media le previsioni possono essere molto distanti

III METODO 3.19

↳ $B_0 =$ uno i primi 10 valori disponibili \rightarrow il modello si comporta meglio
 ↳ i 3 valori di α non si distinguono molto.

Se voglio calcolare l'accuratezza i primi 10 periodi con $l=10$ non posso considerarli
 ↳ l'errore sui primi 10 periodi sarà nullo

○ Il limite di questo metodo è nell'assunzione di stazionarietà delle domande
Se la domanda non è stazionaria qst 2 metodi non vanno bene.

Come scegliamo il β ? Discorso analogo.

- ↳ quanto α è grande tanto α importance viene data al salto di livello dato
- ↳ quanto α è piccolo tanto α peso viene dato al trend calcolato al passo precedente.

LEGAME TRA α e β

Per controllare quanto sono reattivo (variazione del trend) allora controllare sia il β che il α (anche α ripetuto nel trend)

I se α è piccolo (no) non c'è β grande che possa usare α avere un trend molto reattivo.
 ↳ la scelta di β è influente

$$T_t = \beta (B_t - B_{t-1}) + (1-\beta) T_{t-1}$$

Sostituisco B_t con la formula in termini di α .

$$T_t = \beta [\alpha Y_t + (1-\alpha)(B_{t-1} + T_{t-1}) - B_{t-1}] + (1-\beta) T_{t-1}$$

$$= \beta \cdot \alpha Y_t - \beta \cdot \alpha B_{t-1} + \beta T_{t-1} - \beta \alpha T_{t-1} + T_{t-1} + \beta T_{t-1}$$

$$= \alpha \cdot \beta Y_t + 2\beta B_{t-1} + (1-2\beta) T_{t-1}$$

↳ se α è 0 i primi 2 termini sono nulli

↓
 il trend al passo t è molto vicino al trend al passo $t-1$
 il trend è molto **CONSERVATIVO**
 (il tasso di crescita rimane costante)

Quello che abbiamo detto sull' α sarà anche β ma gettando l'ultimo da solo non basta

Se scelgo coerentemente α (abbast. basso) implicitamente assumo crescita costante - non voglio cambiare il livello del trend → do molto peso a

↳ se α è piccolo l'errore (piccolo) viene filtrato e cerco di mantenere sempre lo stesso trend

$(B_{t-1} + T_{t-1})$ è un valore di β è influente

II se α è grande peso tutte quelle componenti (pesano) allora il cambiamento del trend dipende dal valore di β

↳ controllo la velocità con cui cambio il trend

2° modello

$$T_t = \beta (Y_t - Y_{t-1}) + (1-\beta) T_{t-1}$$

↳ uso osservazioni Y_t invece del livello base
 ↳ il trend dipende parzialmente da β senza includere α

se α è grande il valore del trend dipende da β ~~meno~~

↳ ha un grado di libertà in t e include il I modello

In questo caso il trend e il livello base sono indipendenti → POSSO SCEGLIERE IN MODO INDIPENDENTE α e β

supponendo time bucket settimanali

$$n = 58 \quad F_{t_0, 58} = B_{t_0} + 58 T_{t_0}$$

PROBLEMI / LIMITI

$t_0 = \text{adesso!}$ ☺

① Il trend T_{t_0} viene amplificato tanto + n è grande - se il trend non contiene errori va tutto bene

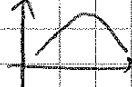
② se il trend è negativo la previsione è negativa e ciò non ha senso!

→ se ho commesso errori nel trend questi vengono amplificati per n grande

③ TREND "COSTANTE"

↳ sensibilità del modello per eventuali errori del trend

Questo assunto mi dà molti problemi quando il trend si inverte



Supponendo di essere in t_0 e di avere B_{t_0}

→ $T_{t_0} = 10$ dopo che il trend decende
 $\tilde{t}_0 = -7$

decido di prevedere con $h = 4$

$$F_{t_0, 4} = B_{t_0} + 4 \cdot 10$$

Supponiamo che non ci siano oscillazioni casuali

$$Y_{t_0+4} = B_{t_0} + 4 \cdot 7$$

cometto un ERRORE GRANDISSIMO

la domanda che si verifica in t_0+4

$$e_t = Y_t - F_t = 4(10+7)$$

Tale modello NON GESTISCE BENE le INVERSIONI del Trend!

METODO + ROBUSTO PER CALCOLARE T_0 e B_0

Quando abbiamo trend la domanda dipende dal tempo: la mia domanda Y_t

$$Y_t = a + bt$$

↑ tempo

→ come la domanda, variabile dipendente, il linguaggio del tempo, variabile indep.

per ogni osservazione conosci Y_t

← coppie di punti

① posso stimare a e b mediante la regressione lineare statistica

METODO CHE INTERPOLA I DATI DI DOMANDA E EVIDENZIA UNA TENDENZA LINEARE nell'ANDAM. dei DATI

$a = B_0 = \text{intercetta}$
 $b = T_0 = \text{coefficiente angolare}$

tale metodo è abbastanza costoso → servono molti dati a disposizione:

grande t_0 sample affinché i dati siano statisticamente significativi

② INIZIAZIONE LINEARE DI TREND = medie di $l-1$ incrementi osservati - (vedi pg precedente)

stagione → verrà divisa in tutte bucket che permette di tornare al punto di partenza (periodo)

$$B_t = \alpha \left(\frac{Y_t}{S_{t-s}} \right) + (1-\alpha) B_{t-1}$$

$S = \#$ periodi stagionali
($\#$ time bucket nella stagione)

LEVEL
MEDIO
DOMANDA

Y_t sarà nuova che è + alto o + basso del base level

dove S è un fattore di stagionalità → Aggiorniamo il base level

$$S_t = \gamma \left(\frac{Y_t}{B_t} \right) + (1-\gamma) S_{t-s}$$

confronto lo domanda con il livello base

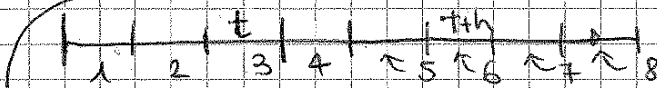
AGGIORNAM

ci saranno tante fattori di stagionalità quante sono gli S periodi che costituiscono una stagione

← come calcolo la previsione per $t+h$

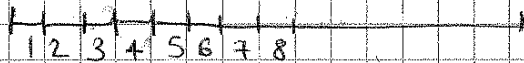
$$F_{t,h} = B_t \cdot S_{t+h-s}$$

quando $h \leq S \rightarrow$ PREVEDO ENTRO UNA STAGIONE



$t=3$
 $h=3$ $S=4$

→ prendo le stagionalità + recente ma non è indice corretto



$t=3$
 $h=7$
 $S_{3+7-4} = S_6$

$$F_{t,h} = B_t \cdot S_{t+h-s} \left\lfloor \frac{h-1}{S} + 1 \right\rfloor$$

→ se voglio prevedere andando OLTRE la STAGIONE UORIZZONTE eccede la STAGIONE

$$3+7-4 \left\lfloor \frac{7-1}{4} + 1 \right\rfloor$$

FORMULA GENERALE

→ vale anche se $h < S$

COEFFICIENTE di SMORZAMENTO

Dobbiamo definire i parametri α , γ , B_0 e S

fatto + γ è alto + è veloce il cambiamento della stagionalità

Valore di regressione di B ed $S \rightarrow$ in funzione di α e γ

ipotizzando che mai si influenzano come scegliamo α e γ ?

si $0 \leq \alpha \leq 1$
 $0 \leq \gamma \leq 1$

legame tra α e γ ?

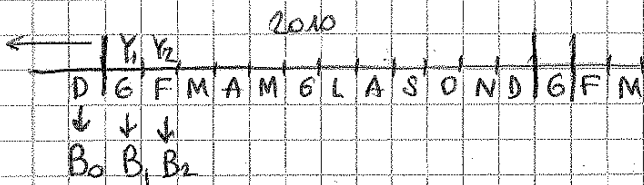
Non ho S_0 da inizializzare, la S di gennaio è \neq dalla S di febbraio i coefficienti di

stagionalità li aggiorniamo una volta all'anno

$B_t \rightarrow$ li aggiorniamo una volta al mese. \rightarrow \times altre nel modello fortemente realista

i coefficienti di stagionalità sono tutti 1 per ogni periodo delle stagionalità

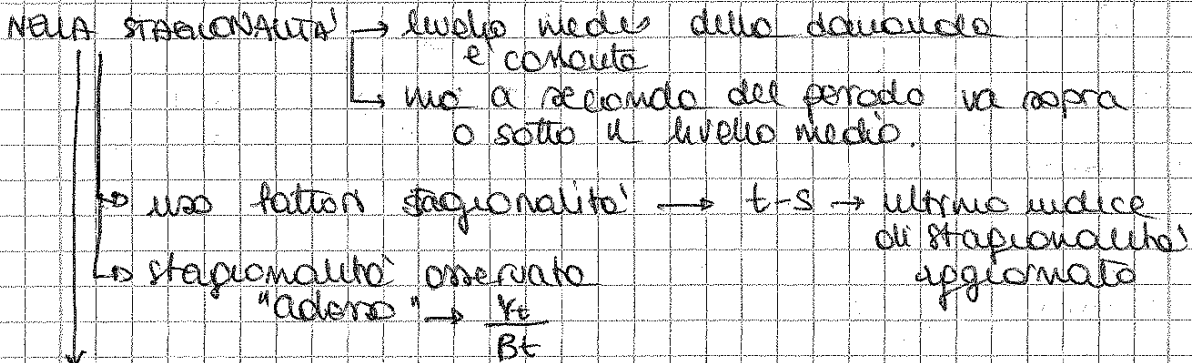
$\gamma > \alpha$



(100) l'inizializz. potrebbe portare ad alcuni errori i fattori di stagion. sono calcolati in un particolare time bucket

I parametri che inizializza li mette da B_0 indietro

uso domande del 2010 per calcolare B_0 e x calcolare gli indici di stagionalità del 2009 e per calcolare gli indici di stagionalità del 2010



$$S_t = \gamma \frac{Y_t}{B_t} + (1-\gamma) S_{t-s} \quad \text{AGGIORNAMENTO FATTORE} \quad 8/11$$

L'inizializzazione come descritto non sarà di buona qualità perché uso solo pochi dati

→ possiamo usare un n° superiore di dati x inizializzare B_0

Abbiamo 2 alternative

- ① # dati multiplo di s (i dati con l'multiplo di s)
- ② # dati arbitrario (non è multiplo s)

la mia scelta dipende dalle disponibilità dei dati e dal TRADEOFF tra fit sample e test sample

"MEDIA SEMPLICE DELLE OSSERVAZIONI"

① i dati multiplo di s

$$S_{j-s} = \frac{\sum_{k=0}^{(s/(s-1))} (Y_{j+k} + k s)}{B_0} \cdot \frac{1}{s}$$

$s=12 \quad l=36 \quad j=1 \dots m \text{ (periodi)}$

$$B_0 = \frac{\sum_{i=1}^l Y_i}{l}$$

una volta inizializz. B_0 posso ricavare S

quando $k=0 \rightarrow Y_{12}$ = gennaio
 $k=1 \rightarrow Y_{13}$ = " dell'anno dopo
 $k=2 \rightarrow Y_{25}$

→ Ogni indice di stagionalità lo calcolo con 1+ di dati maggiore.

ESERCIZIO : Educazione : voglio prevedere la base giornaliera

Weekday	W1	W2	W3	W4	W5	W6
M	46	57	23	36	29	x?
ME	37	43	24	35	34	
G	19	35	34	43	32	
V	50	50	60	50	52	
S	66	79	92	63	72	
D	95	81	81	110	91	
L	121	114	123	116	113	

↳ i quotidiani si vendono solo un giorno!

↳ le settimane inizia da martedì

↳ voglio prevedere la domanda per martedì della sett. 6

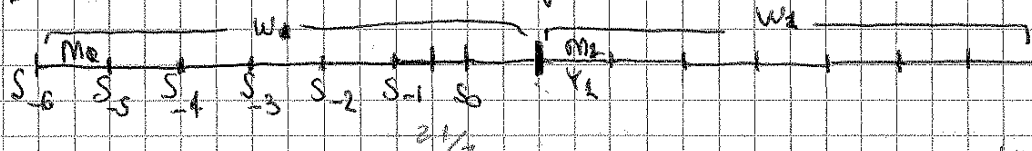
$t=1$
 $\alpha=0,1$
 $\delta=0,2$

Volevo info sulle distribuzioni delle domande (misure di errore di previsione)

Qual'è la variabile della domanda di giornali?

- ↳ FIT SAMPLE = W1, W2, W3
- ↳ TEST SAMPLE = W4, W5

HO 7 INDICI DI STAGIONALITÀ: VMO al giorno!



$S_{5-5} = S_{-6} = \frac{(46+57+23)}{63,3} = 0,663$

$S_{4-5} = S_{-5} = \frac{(37+43+24)}{63,3} = 0,442$

$S_{3-5} = \frac{(19+35+34)}{63,3} = 0,442$

$S_{2-5} = \frac{(50+50+60)}{63,3} = 0,84$

$S_{1-5} = \frac{(66+79+92)}{63,3} = 3,74/3 = 1,24$

$S_{0-5} = \frac{(95+81+81)}{63,3} = 1,35$

$S_{-1} = \frac{(121+114+123)}{63,3} = 5,65/3 = 1,88$

data $B_0 = \frac{\sum_{i=1}^{43} Y_i}{21} = 63,3$
 LIVELLO BASE CALCOLATO USANDO LE 3 SETTIMANE DEL FIT SAMPLE

NON SAPPIAMO CHE la previsione

$F_t = B_t \cdot S_{t-1}$ e

Volevo la previsione per il giorno 35

$F_{35} = B_{35} \cdot S_{35-1} = B_{35} \cdot S_{29}$

MISURE CALCOLARE

B_{35} : la base fine a 35

← S_{29} : il fattore di stagionalità aggiornato.

Indice di stagionalità del martedì della settimana 5

Se voglio una MISURA DELLA DEVIAZIONE STANDARD?

↳ MI SERVE RMSE : per questo mi servono le previsioni dal periodo 22

$$F_{22} = F_{21,1} = B_{21} \cdot S_{21+1-2}$$

↳ Y_{22} → attraverso qst calcolo RMSE = 10.05

Se rivedevo info sulla distribuzione della domanda avrei fatto 35 iterazioni tutte uguali → avrei potuto usare tutte le 5 settimane come fit sample

NEW RMSE → non posso mettere i primi 21 valori → sono fit sample

ATTENZIONE: i pedici sono i t → Adesso e' t

↳ le iterazioni sono fuori dall'orizzonte dei dati disponibili

↳ se mi serve B35 mi devo fare tutte le iterazioni !!

DEVIATIONE STANDARD della domanda = 10,05

Calcolando una deviazione standard campionaria con 35 componenti di t → ho 34,48

USARE L'UNO O L'ALTRO DATO DIPENDE DALLO SCOPO

variabilità non controllabile dovuta all'incertezza

RMSE misura l'errore di precisione: la variabilità della domanda che non so prevedere

deviaz. st. Campionaria

↳ se si prepara e' dovuta alla STAGIONALITA'

prevedo + e' prevedo rispetto alla varianza campionaria significa che gran parte di quella varianza la so controllare

può essere controllata

sono grane o stagioni

↳ LA MIA PREVISIONE sarà accurata

+ e' vicino alla deviazione campionaria + la variabilità sarà non controllabile

↳ le mie previsioni saranno INESATE

→ RIMUOVO IL FATTORE DI TREND DALLE OSSERVAZIONI

$$B_0 = \frac{\sum_{i=1}^{s+1} (Y_i - iT_0)}{s+1}$$

→ NON VA BENE: i dati non sono omogenei (2 GENNAIO)

$$= \frac{(Y_1 - iT_0) + (Y_{s+1} - (s+1)T_0)}{2} + \sum_{i=2}^s (Y_i - iT_0)$$

prendo le domande depurate dalle stagionalità

$$S_{j-s} = \left(\frac{Y_j}{B_0 + jT_0} \right) \quad j=2 \dots s, j \neq 1$$

$$\sum_{i=2}^s \frac{(Y_i - iT_0)}{s} = \text{DOMANDA MEDIA IN CIASCUN CI di } s$$

quando $j=1 \rightarrow S_{j-s} = \left(\frac{Y_1}{B_0 - 1T_0} + \frac{Y_{s+1}}{B_0 + (s+1)T_0} \right) \frac{1}{2}$

domanda "de-trendizzata"

usiamo un numero minimale di dati → non ci possiamo aspettare grande accuratezza

Questo T_0 è affetto da STAGIONALITÀ

Come facciamo ad avere un'inizializzazione del trend depurato dalle stagionalità?

↳ se che la STAGIONALITÀ media sullo stagione è 1 → il problema è che non posso usare s dati

NO DI DATI **2s**

con 2 stagioni complete posso fare la media ma le differenze di tutti i periodi della stagione → $(Y_{i+s} - Y_i)$

CON 2s OSSERVAZIONI

$$T_0 = \frac{\sum_{i=1}^s (Y_{i+s} - Y_i)}{s^2}$$

"differenze" tra GEN10 e GEN11 diviso 12 "di quanto aumento di mese x aumento GEN"

↳ trovato lui i ho il salto tra GEN 11 e GEN 12 e quest salto lo divido $\times 12 \rightarrow$ TREND GEN sommo quest trend (di tutti i mesi) e per trovare il trend mensile devo ancora dividere per s

$$T_0 = \left[\frac{\sum_{i=1}^s (Y_{i+s} - Y_i)}{s} \right] \frac{1}{s}$$

PER OGNI TIME BUCKET HO IL TREND CONSIDERANDO IL SALTO NET 12 MESI e poi faccio la media

se volessi inizializzare il trend con n° multiplo di s dati? 30?

posso sfruttare la formula precedente per B_0 tutti i, j a partire da s saranno uguali nel B_0

$$B_0 = \frac{\sum_{i=1}^{2s} (Y_i - iT_0)}{2s}$$

ho i DATI omogenei (vedi libro)!!!

↳ non mi creo problema perché ragguaglio se un numero intero di stagioni

per un'estensione del n° dei dati basta sostituire $2s$ con n

TRATTEREMO LA REGRESSIONE LINEARE SEMPLICE

andamento
lineare
(ideale)

↳ tratteremo 1 var
indipendente

Sicò in termini statistici

si parte dicendo che: IL MODELLO DI TIPO LINEARE che
LEGA le 2 VARIABILI

$$y_i = \alpha + \beta x_i + \epsilon_i$$

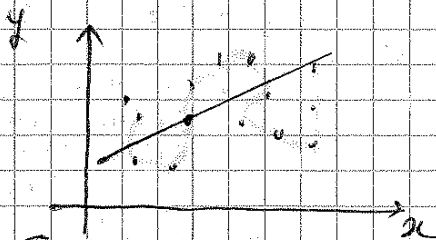
PARAMETRI INCONNITI
CHE CARATTERIZZANO
IL PROCESSO

var.
dipendente

var. indipendente

↳ nome: la parte che non
meso ad appoggiare
alla variabile

Dimentichiamo il discorso statistico →



Note delle coppie
di punti io voglio
cercare una funz.
matematica che
approssima meglio
l'andamento dei
punti

↳ se avessi una
funzione che lega
tutti qst punti
non avrebbe utilità!
se aggiungessi un
punto! ciao!!

se i punti sono 4
di 2 non troveremo
mai una retta

↳ cerco una RETTA
che approssima
qst punti

Cerchiamo la
retta che
minimizza la
somma dei
quadrati

BONTÀ DELL'APPROSSIMAZIONE
calcolata attraverso l'errore
(la distanza tra la retta
e la y del punto)

per evitare
che errori negativi
e errori positivi
si elidano

MINIMIZZARE I QUADRATI
porta ad una soluzione
in forma chiusa (ANALITICA)

↳ mi parte ad
un'equazione

usando la
MINIMIZZI dei valori
assoluti mi
do un MODELLO
matematico assoluto
(dipendente dai
dati)

↳ ho una formula
per a e per b
per cui se

Cambiamo i dati
abbiamo flessibilità
e possiamo ancora
calcolare a e b

• Misura l'accuratezza
con cui il modello
lineare rappresenta
dati empirici

• è + facilmente gestibile
a livello matematico
rispetto a misure in valore assoluto

INTERPRETAZIONI INTERESSANTE DI b

Per definizione di media campionaria

$$\sum_{i=1}^n (x_i - \bar{x}) = 0 \quad \sum_{i=1}^n (y_i - \bar{y}) = 0 \quad \rightarrow \text{le usiamo al numeratore}$$

$$b = \frac{\sum_{i=1}^n x_i (y_i - \bar{y}) - \left(\sum_{i=1}^n \bar{x} (y_i - \bar{y}) \right)}{\sum_{i=1}^n x_i (x_i - \bar{x}) - \left(\sum_{i=1}^n \bar{x} (x_i - \bar{x}) \right)}$$

Riscriviamo il b
AGGIUNGO 2 termini nulli

attenzione! a 2 errori \rightarrow se qst 2 sommatorie sono nulle non significa che i primi 2 termini del numer. e denom. sono nulli e gli altri 2 si può essere portati fuori dalle sommatorie

RACCOGLIO $\bar{x} \sum_{i=1}^n (y_i - \bar{y}) = 0$ uso qst termini per raccogliere

$$b = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{S_{xy}}{S_x^2}$$

(covarianza) / (varianza campionaria)

se prendo la covarianza / deviazione st. x, y

$$b = \frac{r_{xy} S_x S_y}{S_x^2} = \frac{r_{xy} S_y}{S_x}$$

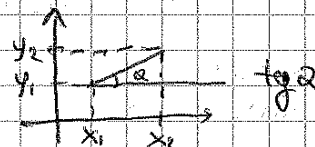
$\frac{S_{xy}}{S_x S_y} = r_{xy}$

perché è + utile?

Sappiamo che a è l'intercetta mentre b è il coefficiente angolare per cui per definizione

SI CONFIGURA COME RAPPORTO INCREMENTALE

$$b = \frac{y_2 - y_1}{x_2 - x_1} = \frac{\Delta y}{\Delta x}$$

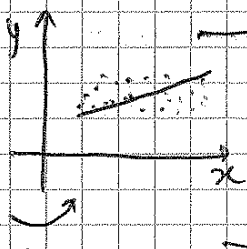


il rapporto incrementale sulle y \rightarrow è sostituito dalla deviazione standard sulle y

vale anche per le x

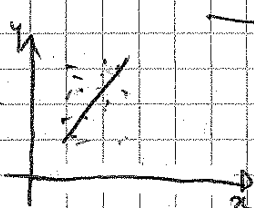
la deviazione standard misura la dispersione dei dati

Questi dati sono molto dispersi sulle x ma poco sulle y



Quanto + la variabile delle ordinate è maggiore rispetto a quella delle x, ne lo numerico più verticale

qst sono + dispersi sulle y



Quanto + la variabile delle ascisse > variabile ordinate, in aspetto la retta di REGRESSIONE + orizzontale

i punti \hat{y} approssimati si distanziano dalle y

$$\hat{y}_i = a + b x_i + \varepsilon$$

$\rightarrow \varepsilon$

variazione "spiegata" della y in funzione della x

non so spiegare: è inclusa nella \hat{y}_i variabile non spiegabile

dipendono dalle osservazioni generate dal processo stocastico

a e b sono stimatori di α e β per cui per noi sono delle variabili casuali

contengono quella variabile casuale che è l'errore

discrepanza tra la y stimata e la y effettiva

lo colpa è dell'errore $\varepsilon \rightarrow$ se non ci fosse $\alpha = a, \beta = b$

dato una $x_i \rightarrow$ la previsione y_i conterrà a, β e ε

ma quando faccio l'approssimazione ho una \hat{y}_i che è dovuta ai fattori a e b

una volta che ho già modello

$$\hat{y} = a + b x$$

$\hat{y}_0 = ?$ se ho x_0

stimo puntuale della variabile y

$$\hat{y}_0 = a + b x_0$$

dovrei dare anche una stima di quant \hat{y}_0 è affidabile.

BUONO STIME: il valore atteso dello stimatore è uguale al parametro stimato: in media il valore di a non è né superiore né inferiore a α

quanto sono affidabili a e b ?
DEVONO ESSERE delle buone stime

ACCURATEZZA / POCA DISPERSIONE (2)

NO BIAS (1) NON DEVIATEZZA

(1) **ASSENZA DI BIAS:** i miei stimatori non devono avere bias, non devono essere devianti: $E[a] = \alpha, E[b] = \beta$

$$E[y_i] = \alpha + \beta x_i$$

$$var(y_i) = \sigma_\varepsilon^2 \quad (\text{a prescindere dalle } i)$$

ho delle informazioni sulla distribuzione

α, β e $\sigma_\varepsilon^2 \rightarrow$ non le conosco e vanno stimate

$$E[a] = \alpha$$

$$E[b] = \beta$$

NON DEVIATEZZA di a o b

$$(I) b = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

scriviamo x_i come $y_i = \alpha + \beta x_i + \varepsilon_i$

$$b = \frac{\sum_{i=1}^n (x_i - \bar{x})(\alpha + \beta x_i + \varepsilon_i - (\alpha + \beta \bar{x} + \bar{\varepsilon}))}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

10/11 ② DEVIAZIONE STANDARD DELL'ERRORE DI STIMA DI UN PARAMETRO

residuo $e_i = y_i - \hat{y}_i = y_i - (a + bx_i)$

\uparrow Valore reale
 \uparrow Valore stimato

ACCURATEZZA / POCA DISPERSIONE

STANDARD ERROR OF ESTIMATE See \rightarrow See (a)
See (b)

$See(b) = \sqrt{E[(b-\beta)^2]}$
 $See(a) = \sqrt{E[(a-\alpha)^2]}$

oppure

distribuzione delle variabili casuali $a-\alpha, b-\beta$

A cosa è legato il valore atteso del quadrato?

Dalla statistica sappiamo che: dato una variabile z

$var(z) = E[z^2] - E[z]^2$

$E[z^2] = var(z) + E[z]^2$

il quadrato di una variabile casuale e' un'altra variabile casuale

$See b = \sqrt{var(b-\beta) + E[(b-\beta)^2]}$

$E[b-\beta] = 0$
 $0^2 = 0$

$See a = \sqrt{var(a-\alpha) + E[(a-\alpha)^2]} = 0$

dimostrazione statistica

perche' $E[b-\beta] = E[b] - E[\beta] = \beta - \beta = 0$

$E[a-\alpha] = 0$

$See b = \sqrt{var(b-\beta)}$
 $See a = \sqrt{var(a-\alpha)}$

Quanto piu' See b e See a sono piccoli quanto + i parametri sono vicini agli stimatori tanto + sono vicini del modello che ho costruito

A seconda delle x_i e delle y_i vengono determinate a e β CALCOLO See a e See b

perche' sappiamo che a e' funzione di $b \rightarrow$ partiamo da b

① $var(b-\beta) = var\left(\frac{\sum_{i=1}^n (x_i - \bar{x})(e_i - \bar{e})}{\sum_{i=1}^n (x_i - \bar{x})^2} + \beta - \beta\right) =$

$= var\left(\frac{\sum_{i=1}^n (x_i - \bar{x})(e_i - \bar{e})}{\sum_{i=1}^n (x_i - \bar{x})^2}\right)$

il denominatore e' un numero (le x sono numeri) posso tirare fuori da var

esempio gelati: intervallo delle x (temperature) (tra 24°C e 25°C) molto piccolo

NON HO ELEMENTI PER VALUTARE QUAL'E' L'ANDAMENTO DEL PROCESSO!
 ↳ i valori della domanda non possono essere così diversi. Al massimo ci sarà un aumento piccolissimo
 ↳ la parte casuale del processo si potrebbe confondere i dati

Nei casi reali i modelli non sono lineari \rightarrow approssimiamo noi il legame!

TRADEOFF SUL NUMERO DEI DATI x_i
 ↳ da un lato voglio + dati x diminuire l'errore (*)
 ↳ con + x se il fenomeno non è lineare è meno efficacemente linearizzato

esattamente come per $b \rightarrow$ calcoliamo la varianza $\rightarrow x$ See (a)

$\text{var}(a - \hat{a})$ con $a = \bar{y} - b\bar{x} = a + (\beta - b)\bar{x} + \bar{\epsilon}$

$\text{var}(a - \hat{a}) = \text{var}(a + (\beta - b)\bar{x} + \bar{\epsilon} - \hat{a}) =$ (la variable è b)

$\text{var}((\beta - b)\bar{x} + \bar{\epsilon}) =$ $\bar{\epsilon} + \bar{x}$ = somma di 2 variabili

$= \text{var}((\beta - b)\bar{x}) + \text{var}(\bar{\epsilon}) + 2\text{cov}((\beta - b)\bar{x}, \bar{\epsilon}) =$ $\rightarrow \bar{x} = \text{numero}$

$= \bar{x}^2 \text{var}((\beta - b)) + \text{var}\left(\frac{\sum \epsilon_i}{n}\right) + 2\text{cov}((\beta - b)\bar{x}, \bar{\epsilon}) =$

$= \bar{x}^2 \text{See } b^2 + \frac{1}{n^2} \sum \text{var}(\epsilon_i) + 2\text{cov}(\dots) =$ $\text{var}(\beta - b) = \text{var}(b - \beta)$
 ↳ se $\text{See } b = \sqrt{\text{var}(b - \beta)}$

$= \bar{x}^2 \text{See } b^2 + \frac{m \sigma^2 \epsilon^2}{n^2} + 2\text{cov}((\beta - b)\bar{x}, \bar{\epsilon})$ $\text{var}(b - \beta) = \text{See } b^2$

possiamo scrivere \rightarrow dimostrano che qst termine è nullo!!

$\text{cov}((\beta - b)\bar{x}, \frac{\sum \epsilon_i}{n}) = -\frac{\bar{x}}{n} \text{cov}(b - \beta, \sum \epsilon_i) =$

$= -\frac{\bar{x}}{n} \text{cov}\left(\frac{\sum (x_i - \bar{x}) \epsilon_i}{\sum (x_i - \bar{x})^2}, \sum \epsilon_i\right)$ $\text{cov}(\beta - b) = -\text{cov}(b - \beta)$

↳ possiamo portarlo fuori

$= -\frac{\bar{x}}{n \sum (x_i - \bar{x})^2} \text{cov}\left(\sum (x_i - \bar{x}) \epsilon_i, \sum \epsilon_i\right) =$ perché

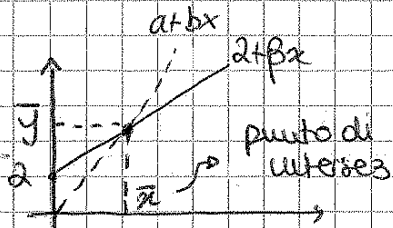
$= \text{cov}\left(\sum (x_i - \bar{x}) \epsilon_i, \sum \epsilon_i\right) =$

$\text{cov}(x+y, z+k)$
 \downarrow
 $\text{cov}(x, z) + \text{cov}(x, k) + \text{cov}(y, z) + \text{cov}(y, k)$

$= \sum_i \sum_j \text{cov}((x_i - \bar{x}) \epsilon_i, \epsilon_j) =$

molti di qst termini saranno nulli

passo portarlo fuori
 fuori d'un numero



per vedere qual'è il legame tra l'errore su β e l'errore su a andiamo che

$$\sum e_i = 0$$

→ gli errori positivi compensano quelli negativi

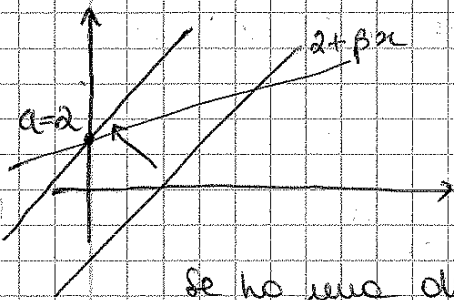
se $b > \beta$ → la retta sarà + pendente un errore su b da una pendenza diversa $b \neq \beta$ e le 2 rette si incontrano, errore su a

NO BIAS

la retta di regressione parte sempre per

Per definizione di baricentro dei dati (\bar{x}, \bar{y}) → \bar{x} sta sulla retta

$$a = \bar{y} - b\bar{x}$$

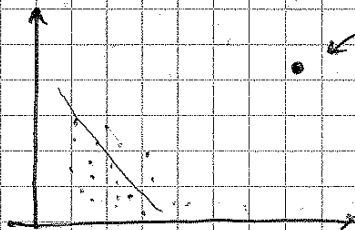


$$\text{se } \bar{x} = 0$$

→ la retta si incontrano sull'asse y per cui l'effetto sull'errore viene completamente meno

Se ho una distribuzione di qst tipo

EFFETTO KING-KONG / BIG APPLE



← punto molto grande attiner

↳ ci fa saltare fuori una relazione molto diversa

Assumendo regressione lineare b dovrebbe essere negativa ma se quel punto è molto pesante potrebbe essere positivo

RIASSUMENDO

Se b

- dipende da:
 - VARIABILITÀ INTRINSECA DEL PROCESSO (+VAR. + DIFFICOLTÀ)
 - NUMERO DI OSSERVAZIONI (> NUM. > ACCURATEZZA)
 - DISPERSIONE DELLE OSSERVAZIONI (PIÙ È AMPIO L'INTERVALLO SU CUI SONO LE x_i PIÙ È ACCURATA LA STIMA)

Se a

dipende da errori su b

- se $\beta = b$
 - $\sum e_i > 0$ → LA RETTA STIMATA SOPRA QUELLA REALE $a > 2$
 - $\sum e_i < 0$ → LA RETTA STIMATA SOTTO QUELLA REALE $a < 2$

maggior o minor errore (+ AFFIDABILITÀ)

- se $\sum e_i = 0$
 - se $b > \beta$ LA STIMATA È PIÙ PENDENTE DI QUELLA REALE
 - se $b < \beta$ LA STIMATA È + ORIZZONTALE DI QUELLA REALE

l'effetto degli errori di stima è tanto + pesante quanto è elevato il baricentro \bar{x} il legame si manifesta su un braccio + elevato!

PRIMO SET

$$b = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} + \beta$$

$$\bar{x}_1 = \bar{x}_2 = 5$$

$$\bar{y}_1 = 125 \quad \bar{y}_2 = 115,7$$

$$b = \frac{\sum_{i=1}^m (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^m (x_i - \bar{x})^2}$$

$$a_1 = \bar{y}_1 - b\bar{x} = 125 - 5,07 \cdot 5 = 99,64$$

$$\textcircled{2} \quad b_1 = \frac{-5(-25) + (-4)(-21) + (-3)(-14) + (-2)(-11) + (-1)(-4) + (0) + 4 \cdot 1 + \dots}{[(-5)^2 + 4^2 + 3^2 + 2^2 + 1^2 + 4^2] \cdot 2}$$

$$= \frac{8 + 16 \cdot 3 + 22 \cdot 4 + 25 \cdot 5}{\dots}$$

$$\textcircled{4} \quad \text{See } b = \frac{\hat{\sigma}_E}{\sqrt{\sum (x_i - \bar{x})^2}}$$

$$\textcircled{5} \quad \text{See } a_1 = \sigma_1^E \sqrt{\frac{\bar{x}^2}{\sum (x_i - \bar{x})^2} + \frac{1}{m}} = 0,688$$

$$\textcircled{3} \quad \hat{\sigma}_E = \sqrt{\frac{(100 - (99,64 + 5,07 \cdot 0))^2 + (104 - (99,64 + 5,07 \cdot 1))^2 + \dots}{11 - 2}}$$

$\hat{y}_i = a + bx_i$

$$\textcircled{4} \quad \text{See } b = \frac{\hat{\sigma}_E}{\sqrt{(0 - \bar{x}_0)}}$$

VALORI DEGLI STIMATORI

$$a_1 = 99,64$$

$$b_1 = 5,07$$

$$a_2 = 99,32$$

$$b_2 = 5,06$$

See $a_1 = 0,688$
See $b_1 = 0,116$

See $a_2 = 19,53$
See $b_2 = 3,30$

1 campionari sono 11, non sono tutti il range delle x non è piccolo ma dipende da cosa sono le x

forte correlazione lineare tra x e y e riesco a trovare bene una retta approssimativa molto vicina alle nuvole dei dati

Ci sono molte + rette possibili, da un vicinissimo intorno ai dati

↳ il legame tra le x e le y non è così bene rappresentato da una retta. Prima di usare questa retta x la previsione di y può essere

σ_E deve essere piccola

↳ poca variabilità in gioco

LA RELAZIONE EMERGE IN MANIERA STATISTICAMENTE NON MOLTO SIGNIFICATIVA

$$\alpha \Rightarrow a \pm S_{ee} a \cdot t_{1-(p/2), n-2}$$

Sto considerando un intervallo bilaterale: encapsulo α tra 2 valori

definito
 $p =$ livello di signif.
 $(1-\alpha)$
 $\frac{\alpha}{2} = \frac{(1-p)}{2}$
 $1-\alpha$

$$\beta \Rightarrow b \pm S_{ee} b \cdot t_{1-(1-p)/2, n-2}$$

ALLAUMENTARE di S_{ee} AUMENTA LA DIMENSIONE DELL'INTERVALLO di CONFIDENZA: CONSCUENZA - PRECISIA DI α e β

se $S_{ee} a$ e $S_{ee} b$ sono grandi $\rightarrow \alpha$ potrebbe essere molto diverso da a

Lo nel secondo set di dati, lo $\alpha = 99,32$ veniva encapsulato in un intervallo molto ampio \rightarrow ho + incertezza.

NEL II SET DI DATI

$$-2,9 \leq \beta \leq 12,93$$

\rightarrow il fenomeno non è lineare

possiamo anche avere un β nullo

È IMPORTANTE PER IL β

o negativo

con tale confidenza \rightarrow GRAVE! \rightarrow uso il mio modello lineare con rischio

il β è UNA PENDENZA!

Lo se α è il prezzo lo domanda aumenta o diminuisce?

Nei precedenti che aumenta e invece poi diminuisce ad esempio.

per vedere se $S_{ee} a$ o $S_{ee} b$ sono grandi lo vediamo dagli intervalli di confidenza (se ho solo 1 set di dati)

se l'intervallo di confidenza su α è molto grande \rightarrow ho gli stessi problemi ma l'intervallo di β

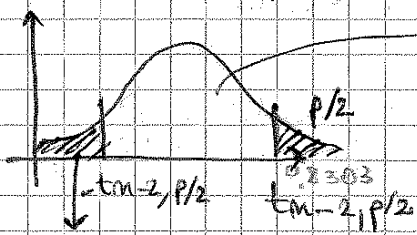
che contiene lo 0 è + grave perché non so qual'è la pendenza

Cosa significa livello di confidenza del 95%?

\rightarrow non significa che al 95% di si trova in quell'intervallo

ma se estraggo 95 campioni su 100 α si trova dentro quell'intervallo

$p =$ livello di significatività



zona di accettazione: se TS sta in uno di qst punti non possiamo accettare l'ipotesi nulla

zona di rifiuto

se l'intervallo di confidenza contiene lo θ non posso

$$\begin{matrix} H_0 & \alpha = 5 \\ H_1 & \alpha \neq 5 \end{matrix}$$

$$p = 5\%$$

$$\text{liv } l = 1 - 0,05 = 95\%$$

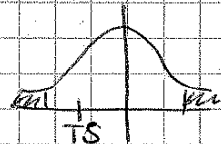
STO TESTANDO SE S E' INCLUSO NELL'INT.

Lo se qst intervallo contiene 5 non posso rifiutare l'ipotesi nulla

E' importante capire se $\alpha \neq \beta \neq 0$

$\rightarrow = 0$ lo x e lo y non sono legate

I TEST DI IPOTESI SI USANO QUANDO C'E' DA STIMARE I p -VALUE fissando $p = 1\%$.



TS passa nella zona di accettazione dell'Hp nulla

pe averon

$$\text{cullto } p = 9\%$$

\rightarrow e TS rientra nella zona di rifiuto non posso accettare l'ipotesi nulla

L'accettazione dell'Hp nulla dipende dal livello di significatività

p -value \rightarrow valore di probabilità corrispondente al valore di t di Student che coincide con la nostra statistica del test

$$TS = \frac{a - a_0}{S_{eaa}} = t_{n-2, p\text{-value}}$$

dato il valore t trovo il p -value

TS \rightarrow analizzo e vado a leggere l'area che mi rappresenta la probabilità p -value

Nel iii set di dati x rifiutare l'ipotesi che $B_{a0} = 0$ bisogna avere un livello di significatività basso

TEST DI IPOTESI \rightarrow li usiamo a capire se i coefficienti che abbiamo trovato sono significativi

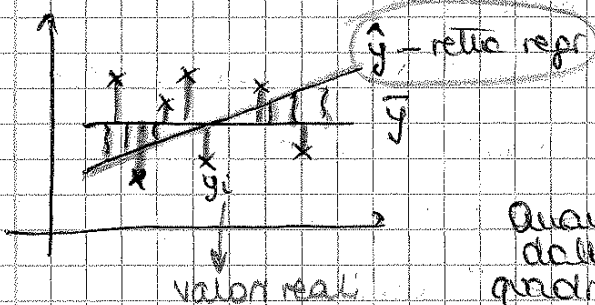
e poiché $\sum_{i=1}^{m-1} (x_i - \bar{x})(y_i - \bar{y}) = S_{xy}$

$= b((m-1)S_{xy} - b(m-1)S_x^2) = b(m-1)(S_{xy} - bS_x^2) =$
 $= b(m-1)(S_{xy} - \frac{S_{xy}}{S_x^2} S_x^2)$

Questo termine è nullo invece solo π al num.

$R^2 = \frac{(\sum (y_i - \hat{y})^2)^2}{\sum (y_i - \bar{y})^2 \sum (y_i - \hat{y})^2} = \frac{\sum (y_i - \hat{y})^2}{\sum (y_i - \bar{y})^2} \rightarrow R^2 = \frac{S_{\hat{y}}^2}{S^2_y}$
 VARIANZA

Se \hat{y} sono diverse quando \rightarrow le variabile al numeratore
 le x sono diverse e quella che è alla
 retta di regressione



Lo spiegato dalla
 retta di
 REGRESSIONE

Quanto + i punti sono lontani dalla retta tanto + \rightarrow il quadrato daranno un errore maggiore rispetto a \rightarrow il numeratore sarà piccolo e R^2 scenderà

Quello che x non spiega di variabile y , può dipendere dalle variabili volatilità oppure non sto considerando altre variabili significative

R^2 alto spiega \rightarrow che ad un aumento x del prezzo fatto l'aumento y che vedo non dipende solo da x

- \rightarrow NO \rightarrow o grande errore
- \rightarrow o modello sbagliato

R^2 = RAPPORTO TRA VARIANZA delle \hat{y}_i e la VARIANZA delle osserv. y_i

Quale percentuale della variabile osservata nelle y_i è possibile spiegare attraverso la retta di regressione lineare? $y = a + bx$

Quale parte del processo stocastico siamo in grado di comprendere?

Al crescere di R^2 la retta ha una maggior capacità di spiegare le osservazioni del fenomeno a nostra disposizione