ECONOMIA

METODI STATISTICI PER L'ANALISI ECONOMICA E AZIENDALE

GLI INDICATORI STATISTICI

FABRIZIO ANTOLINI
fantolini@unite.it

INDICATORI STATISTICI

Gli **indicatori statistici** sono strumenti di valutazione complessa di un *fenomeno* a cui si riferisce un certo legame con un referente concettuale che fa parte di un modello di ricerca e di uno schema interpretativo.

Un *indicatore rappresenta una statistica*: un indicatore statistico è sostanzialmente un numero che si ottiene in modo opportuno operando mediante differenze e rapporti tra i dati statistici raccolti e moltiplicando eventualmente i rapporti per convenienti potenze di 10.

Esso fornisce informazioni sul comportamento reciproco dei dati medesimi e, di conseguenza, sul fenomeno statistico che si vuole studiare.

Un **indice diviene indicatore** quando la sua definizione e misurazione è collegata ad un obiettivo definito.

In considerazione della loro natura, gli indicatori (o indici) si distinguono in:

- **indici assoluti** quando sono espressi nella *stessa unità di misura del fenomeno*
- **indici relativi** quando *non dipendono dall'unità di misura del fenomeno* e si ottengono o rapportando due misure assolute oppure rapportando un indice assoluto al suo massimo.

Infine si hanno gli **indici normalizzati**, che sono indici relativi, la cui caratteristica è quella di assumere valori all'interno di un intervallo finito: quasi sempre [0,1] oppure [-1,+1].

Le caratteristiche di un indicatore sono:

Esclusività. Dovrebbe essere non sostituibile con un altro indice per indicare lo stesso fenomeno sociale. Nel caso si abbiano due indici che ci danno la medesima informazione bisogna vedere se c'è un indice comune che è possibile utilizzare.

Sensibilità. Dovrebbe essere in grado di esprimere l'intero fenomeno osservato e registrare tutte le variazioni del fenomeno stesso. Ci deve essere un *legame lineare rispetto a ciò che viene misurato*.

Univocità. Interpretabile in modo non ambiguo rispetto alle direzioni delle variazioni. Ad esempio, se abbiamo una variazione negativa del reddito pro-capite è legittimo aspettarci che corrisponda una variazione positiva della povertà. Non dobbiamo avere indici che abbiano una variazione non lineare.

Fedeltà. Dovrebbe imputare le variazione lungo il tempo a cambiamenti della realtà e non alla qualità della misurazione. È evidente che l'indice non dovrebbe essere soggetto a variazioni non imputabili ad altro che alla variazione del fenomeno sottostante.

Sintesi. Dovrebbe essere la sintesi di aggregazione di osservazioni parziali.

Finalizzazione. Dovrebbe essere correlato agli obiettivi dello studio.

Esaustività. Dovrebbe descrivere tutte le dimensioni considerate.

Significatività. Dovrebbe essere significativo relativamente alle unità (anche territoriali) considerate.

INDICATORI OTTENUTI PER DIFFERENZA

Esempio Matrice dei dati

Anno scolastico	2012	2013	2014	2015	2016
Maschi	97	109	109	98	105
Femmine	80	87	84	91	92
Totale	177	196	193	189	197

Un indicatore ottenuto per differenza si dice **assoluto** quando è ricavato eseguendo la differenza pura e semplice tra la dimensione o consistenza di un fenomeno ad una certa data o ad un certo periodo e la dimensione o consistenza dello stesso fenomeno ad un tempo antecedente a quello considerato.

In formula:

$$\boldsymbol{d = x_t - x_0}$$

Con riferimento all'esempio della tabella, un indicatore assoluto è quello che si ottiene effettuando la differenza tra i numeri degli alunni iscritti nell'anno 2016 e nell'anno 2012:

$$\boldsymbol{d = 197 - 177 = 20}$$

Questo indicatore ci dice banalmente che il numero degli iscritti nell'anno 2016 è cresciuto di 20 unità

rispetto all'anno 2012.

INDICATORI OTTENUTI PER DIFFERENZA

Un indicatore ottenuto per differenza è **relativo** quando è ricavato eseguendo il rapporto tra un indicatore assoluto e la dimensione o consistenza che il fenomeno osservato aveva al tempo antecedente.

In formula:

$$\boldsymbol{d = (x_t - x_0)/\, x_0}$$

Riprendendo, con riferimento alla stessa tabella, il numero degli alunni iscritti nell'anno 2016 (=197) e quello degli iscritti nell'anno 2012 (=177), si ha:

$$\boldsymbol{d = (197 - 177)/\, 177 = 0,1130}$$

N.B: In genere questi indicatori vengono moltiplicati per 100 e il risultato diviene una percentuale.

INDICATORI OTTENUTI PER RAPPORTO: I RAPPORTI STATISTICI

Si definisce **rapporto statistico** un numero che si può rappresentare nel seguente modo:

$$\boldsymbol{R = \frac{A}{B}}$$

I rapporti statistici sono misure statistiche elementari finalizzate al confronto tra i dati stessi. In un rapporto statistico si mettono a confronto due termini, frequenze o quantità, di cui uno almeno è di natura statistica e tale che tra i due termini sussiste un qualche legame logico.

Un rapporto statistico indica **quanta parte** dell'intensità del fenomeno posta a numeratore compete, **in media**, ad **ogni unità** di intensità del fenomeno posta a denominatore. I rapporti così costruiti permettono di confrontare l'intensità di un fenomeno misurato su un collettivo, in tempi o luoghi diversi, e sono **largamente impiegati** nella descrizione di fenomeni di tipo socio-economico.

Riassumendo:

- Almeno una delle due intensità considerate deve avere **natura statistica**, cioè riferirsi a un fenomeno collettivo.
- Tra i fenomeni le cui intensità sono poste a confronto deve intercorrere un **nesso logico**.
- Assumono un **valore sempre positivo** e *non dipendono (quasi sempre) dall'unità di misura* consentendo così un confronto tra fenomeni diversi, ovvero vengono calcolati per eliminare l'influenza di circostanze che altrimenti non renderebbero confrontabili i dati stessi.

I RAPPORTI STATISTICI: RAPPORTO DI COMPOSIZIONE

A seconda di quale relazione sussiste tra il numeratore e il denominatore della frazione, si hanno diversi **tipi** di rapporti statistici. Ne prendiamo in considerazione quattro, in base ad altrettanti tipi di rapporti:

- Rapporto di composizione
- Rapporto di coesistenza
- Rapporto di derivazione
- Rapporto di densità

Mentre i primi tre si ottengono dividendo grandezze omogenee e sono pertanto numeri puri. Il rapporto di densità è confronta due grandezze eterogenee ed è pertanto caratterizzato da un misura.

Il **rapporto di composizione** è una misura ottenuta rapportando il valore di una parte di un determinato collettivo o popolazione a quello dell'intero collettivo o popolazione. Un rapporto di questo tipo, dunque, la quantità al numeratore è una parte della quantità a denominatore.

Il rapporto di composizione è detto anche **rapporto di parte al tutto**.

$$\boldsymbol{Rapporto\ di\ composizione = parte\ del\ fenomeno/totale\ fenomeno}$$

I RAPPORTI STATISTICI: RAPPORTO DI COMPOSIZIONE

Riprendendo la tabella e ponendo l'attenzione sull'anno 2016, si può constatare che i maschi (=105) costituiscono la seguente frazione del totale degli iscritti (=197):

$$\boldsymbol{Rapporto\ di\ composizione = \frac{105}{197} = 0,53299}$$

Il rapporto di composizione può assumere valori compresi tra 0 e 1 (oppure tra 0 e 100, 0 e 1000, …, se il rapporto è moltiplicato per comodità di lettura rispettivamente per 100, 1000, …).

Alcuni rapporti di composizione noti:

- **Tasso di occupazione**: su una popolazione di persone che hanno compiuto 15 anni, è la percentuale di coloro che hanno un lavoro.
- **Tasso di disoccupazione**: su una popolazione di persone che hanno compiuto 15 anni, è la percentuale di coloro che non hanno un lavoro.
- **Tasso di scolarità:** sugli alunni di una determinata fascia di età, è la percentuale di coloro che frequentano la scuola. Per esempio, in Italia, il tasso di scolarità della secondaria di 2° grado è la percentuale di coloro che frequentano quest'ordine di scuola rispetto al totale dei giovani della fascia di età 15-18 anni.

I RAPPORTI STATISTICI: RAPPORTO DI COESISTENZA

Se l'ammontare complessivo di una quantità (frequenza o intensità) viene classificato in più modalità o classi, **il rapporto di coesistenza** è il rapporto tra la frequenza di una modalità rispetto alla frequenza corrispondente di un'altra modalità. Si può anche definire come un rapporto fra le intensità o frequenze riferite a fenomeni fra loro antitetici ma che coesistono. Evidenziano l'eventuale squilibrio fra le due grandezze coesistenti.

Si tratta, in pratica, di **un rapporto tra due "parti"** che, se unite insieme, darebbero luogo alla totalità dei casi.

$$\boldsymbol{Rapporto\ di\ coesistenza = frequenza\ modalit\grave{a}/frequenza\ modalit\grave{a}\ B}$$

Questo rapporto, come si può capire facilmente, è un numero puro non negativo. Generalmente, i rapporti di coesistenza vengono moltiplicati per 100 o per 1000.

Un rapporto di coesistenza molto utilizzato è l'indice di vecchiaia, il quale indica il grado di invecchiamento di una popolazione. Lo si calcola moltiplicando per 100 il rapporto fra il numero di persone anziane e quello dei giovani di una certa comunità.

RESIDENTI PER GRANDI CLASSI DI ETA'					**Indice di vecchiaia (65 e oltre *100)/0-14**
ANNO	**0-14**	**15-64**	**65 e oltre**	**totale**	
2020	13.702	71.938	27.684	113.324	**202**
2021	13.828	69.865	28.535	112.228	**206,4**

Questo indice mostra che per l'anno 2021 ad esempio ci sono 206,4 anziani ogni 100 giovani.

I RAPPORTI STATISTICI: RAPPORTO DI DERIVAZIONE

Un **rapporto di derivazione** è ottenuto dividendo la frequenza o intensità di un fenomeno per la frequenza o intensità di un altro fenomeno che, sul piano logico o temporale, ne costituisce l'antecedente, la causa o il presupposto.

Com'è facile notare, il numeratore non è più una parte del denominatore, ma anzi **il numeratore e il denominatore vanno a contare due quantità tra loro diverse**, seppure legate logicamente tra loro.

$$\boldsymbol{Rapporto\ di\ derivazione = modalit\grave{a}\ susseguente/modalit\grave{a}\ antecedente}$$

Anche in questo caso assume valori sempre positivi e per comodità di lettura viene moltiplicato per cento o per 1000.

Un esempio per comprendere meglio può essere il tasso di mortalità $\mathbf{T_m}$ dato da $\mathbf{T_m = m*1000/N}$ dove **m** indica il numero dei morti (modalità susseguente) della popolazione **N** (modalità antecedente) nel periodo o nella ripartizione territoriale (o entrambi).

Ripartizione geografica	nati vivi	morti	popolazione media	quoziente di natalità	tasso di mortalità
Nord	249.677	259.406	26.211.512	9,53	9,9
Centro	104.740	112.846	11.124.059	9,42	10,14
Mezzogiorno	208.182	174.406	20.663.632	10,07	8,44
Totale	562.599	546.658	58.175.310	9,67	9,4

Va detto che la "popolazione" di cui si parla va intesa in senso statistico, potendo essere un insieme di persone, di aziende , di beni in senso lato, eccetera.

I RAPPORTI STATISTICI: RAPPORTO DI DENSITA'

Questo particolare tipo di rapporto statistico si ha quando la relazione logica tra numeratore e denominatore ha a che fare con l'"*addensarsi*" di qualcosa rispetto a qualcos'altro: si tratta dei cosiddetti **rapporti di densità.**

Un rapporto di densità è definito mediante il confronto tra la dimensione globale di una popolazione o di un fenomeno e la dimensione spaziale o temporale cui il fenomeno stesso fa riferimento.

$$\boldsymbol{Rapporto\ di\ densit\grave{a} = dimensione\ globale\ /\ dimensione\ temporale\ o\ spaziale}$$

Il rapporto di densità per antonomasia è la densità della popolazione, data dalla popolazione residente ad un certo istante rapportata alla superficie del territorio:

$$\boldsymbol{densit\grave{a}\ di\ popolazione = \frac{popolazione}{superficie\ (km2)}}$$

Alcuni rapporti di densità noti:

- **Prodotto interno lordo pro capite**: rapporto tra il prodotto interno lordo e la popolazione.
- **Numero medio componenti per famiglia**: rapporto tra popolazione e numero di famiglie residenti nello stesso territorio.
- **Indice di dotazione di posti letto negli istituti di cura**: rapporto tra il numero di posti letto degli istituti di cura e la popolazione.

I TASSI: TASSO DI INCREMENTO

Nel caso della popolazioni, le statistiche ad essa riferite possono essere distinte in:

- **statistiche di stato**, se riferite ad un dato istante
- **statistiche di flusso**, se riferite alla dinamica o alle variazioni intervenute nel corso di un periodo di tempo.

Un esempio della seconda tipologia di rapporti statistici è il **tasso di incremento**: esso si prefigge lo scopo di valutare il cambiamento intercorso tra due dati statistici che **fanno riferimento al medesimo aggregato**, nel momento in cui tra i due dati è trascorso un intervallo di tempo, ovvero quando i due dati danno conto di un fenomeno registrato in due momenti diversi.

In termini formali, detti x_i e x_{i-1} , rispettivamente, i valori della variabile considerata (***X***) al tempo i e al tempo ***i-1*** (un anno prima se parliamo in termini di anni, un mese prima se parliamo in termini di mesi, ecc.), si dice tasso di incremento (o di decremento nel caso il tasso sia negativo) relativo al tempo ***i*** il rapporto:

$$\boldsymbol{TI_I = \frac{X_I - X_{I-1}}{X_{I-1}}}$$

In generale, i confronti nel tempo e nello spazio dei fenomeni vengono effettuati mediante particolari rapporti statistici denominati **tassi**, dove il dato di flusso viene rapportato all'ammontare medio della popolazione.

I tassi consentono di studiare il manifestarsi di un fenomeno di interesse nel tempo e di procedere a confronti temporali e spaziali anche con livelli di approfondimento diversi (a riguardo si parla di tassi generici o grezzi, tassi specifici e tassi standardizzati).

Ad esempio, utilizziamo il **tasso di mortalità.**

Supponiamo di stare indagando il tasso di mortalità della provincia di Ascoli Piceno, pari a 400 per 100mila (in quell’anno, sono morte 400 persone su 100mila). Questo tasso è certamente utile per dare un misura immediata del fenomeno di studio. Discorso diverso se volessimo confrontare questo tasso allo stesso anno con un altra provincia, ad esempio Teramo.

E’ evidente che il tasso di mortalità è fortemente dipendente dalla struttura per età della popolazione che generato il tasso stesso. In genere, una popolazione anziana, tenderà ad avere tassi di mortalità più elevati rispetto ad una popolazione giovane. Anche il sesso può essere una variabile importante, così come anche le diverse condizioni socio-economiche. Queste possono essere chiamate **variabili confondenti**.

Un primo passo consiste nel calcolare, al posto del **tasso generale (tasso grezzo),** una serie di tassi, ciascuno per ogni classe di età: sono **i tassi età-specifici;** **a**d esempio:

$$\boldsymbol{TM_{(25-34)} = \frac{n.decessi, per\ una\ certa\ causea, nell'anno\ X, in\ età\ 25-34\ anni}{n.anni-persona\ nell'anno\ X, in\ età\ 25-34\ anni} * (10^n)}$$

Il passo successivo consiste nel trovare un metodo che consenta di eliminare (o quasi) l’effetto della struttura per età (e, eventualmente, per sesso) della popolazione generante, sia per la popolazione A che per la popolazione B.

I TASSI GREZZI, SPECIFICI E STANDARDIZZATI

Questo metodo consiste nel calcolare tutti i tassi età-specifici delle popolazioni, rispettivamente, A e B e, successivamente, applicare tali tassi specifici non più alle reali strutture per età delle popolazioni per età, rispettivamente, A e B, bensì un'altra struttura per età di una popolazione qualunque (detta POPOLAZIONE STANDARD), la cui scelta è arbitraria.

In questo modo, si calcola un tasso di mortalità, sia per A che per B, "come se" fosse quello sperimentato da tale ipotetica, comune POPOLAZIONE STANDARD; così facendo i due tassi, per A e per B, diventano confrontabili perché non hanno più nulla a che fare con le strutture per età, né di A, né di B.

Questo metodo è detto di **STANDARDIZZAZIONE DIRETTA.**

Date due serie di tassi di mortalità età-specifici (pop.A e pop.B):

$$_ATM_{1;}\ _ATM_{2;}\ldots\ldots\ldots;\ _ATM_{i;}\ldots\ldots\ldots\ldots;\ _ATM_{k;}$$

$$_BTM_{1;}\ _BTM_{2;}\ldots\ldots\ldots;\ _BTM_{i;}\ldots\ldots\ldots\ldots;\ _BTM_{k;}$$

E data la struttura per età della popolazione-standard che è stata scelta nell'intervallo di tempo:

$$\boldsymbol{P_i = \frac{Popolazione\ nella\ classe\ di\ età\ i - esima\ (Italia\ anno\ x)}{Popolazione\ totale\ (Italia\ anno\ x)}}$$

I TASSI GREZZI, SPECIFICI E STANDARDIZZATI

Il tasso standardizzato di mortalità, per la Popolazione A, e B è dato da:

$$_{A}TM_{STAND\text{-}ITALIA\text{-}ANNO\ x} = {}_{A}TM_{1} * P_{1}\ ;\ {}_{A}TM_{2;} * P_{2}, \ldots\ldots\ldots ;\ {}_{A}TM_{i;} * P_{i}, \ldots\ldots\ldots\ldots ;\ {}_{A}TM_{k;} * P_{K}$$

$$_{B}TM_{STAND\text{-}ITALIA\text{-}ANNO\ x} = {}_{B}TM_{1} * P_{1}\ ;\ {}_{B}TM_{2;} * P_{2}, \ldots\ldots\ldots ;\ {}_{B}TM_{i;} * P_{i}, \ldots\ldots\ldots\ldots ;\ {}_{B}TM_{k;} * P_{K}$$

<u>ATTENZIONE</u>: naturalmente, per potere fare confronti, la popolazione-standard deve essere sempre la stessa!

I NUMERI INDICI

Una serie di dati statistici relativi ad un insieme ordinato di istanti (o intervalli) di tempo in successione si dice **serie storica.** Uno dei modi per rappresentare adeguatamente una serie storica (e non solo) è costituito dalla costruzione di un particolare tipo di rapporti statistici: i **numeri indici**.

I numeri indice, di grande importanza soprattutto in campo economico:

- assumono sempre un **valore positivo**
- si configurano come "**numeri puti**", in quanto non dipendono dall'unità di misura nella quale è espresso il fenomeno originario
- servono per **facilitare la comprensione** delle variazioni intervenute nel tempo o nello spazio nei dati statistici.

In generale si distingue tra:

- **numeri indici semplici:** qualora l'obiettivo è il confronto tra due singole "situazioni" differenti
- **numeri indici complessi:** se invece l'obiettivo è quello di descrivere in modo sintetico la variazione, ad esempio, di un gruppo di *n* beni e/o di *n* servizi simultaneamente, elencati in una serie storica multipla, in due "situazioni" differenti.

I NUMERI INDICI

Dato un certo tempo ***t***, un **numero indice (semplice)** relativo al tempo ***t***, è un rapporto del tipo:

$$I_1^t = \frac{y_t}{y_1} * 100$$

La quantità a numeratore di questa frazione è detta **base** del numero indice. Se, nella serie dei numeri indici, si considera sempre come **base** il valore y_1 , siamo in presenza di **NUMERI INDICI A BASE FISSA**.

Se, invece, nella serie dei numeri indici si considera come base il <u>valore precedente</u> (ad esempio) della serie storica, e si ha quindi un rapporto del tipo:

$$I_{t-1}^t = \frac{y_t}{y_{t-1}} * 100$$

siamo in presenza di **NUMERI INDICI A BASE MOBILE (o CONCATENATI)**.

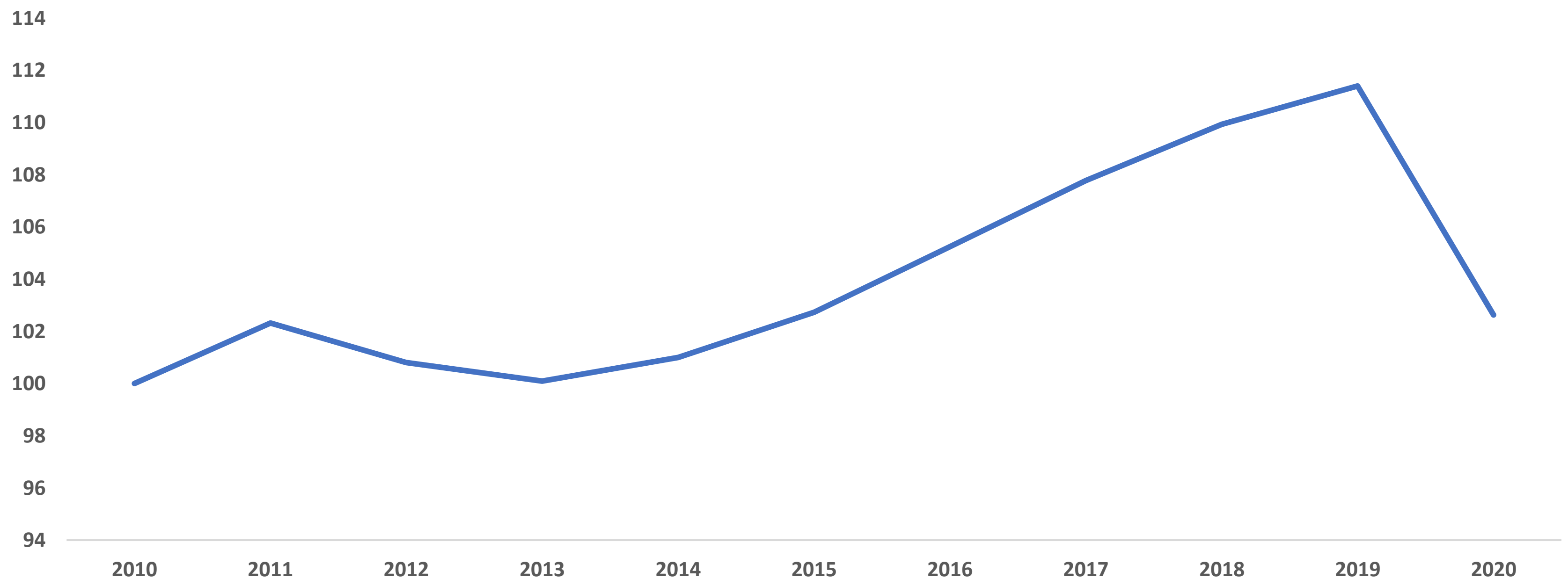
Facciamo un esempio:

anno	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
PIL_{ITALIA} (milioni di euro)	1.611.279,4	1.648.755,8	1.624.358,7	1.612.751,3	1.627.405,6	1.655.355,0	1.695.786,8	1.736.592,8	1.771.391,2	1.794.934,9	1.653.577,2

I NUMERI INDICI

anno	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
PIL_{ITALIA} (milioni di euro)	1.611.279,4	1.648.755,8	1.624.358,7	1.612.751,3	1.627.405,6	1.655.355,0	1.695.786,8	1.736.592,8	1.771.391,2	1.794.934,9	1.653.577,2
INDICE A BASE FISSA	100	102,33	100,81	100,09	101,00	102,74	105,24	107,78	109,94	111,40	102,63

$$I_1^t = \frac{y_t}{y_{2010}} * 100$$



I NUMERI INDICI

anno	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
PIL_{ITALIA} (milioni di euro)	1.611.279,4	1.648.755,8	1.624.358,7	1.612.751,3	1.627.405,6	1.655.355,0	1.695.786,8	1.736.592,8	1.771.391,2	1.794.934,9	1.653.577,2
INDICE A MOBILE	0	102,33	98,52	99,29	100,91	101,72	102,44	102,41	102,00	101,33	92,12

$$I_{t-1}^{t} = \frac{y_t}{y_{t-1}} * 100$$