

01 – Un mare di informazioni...

Unità n° 01

Oggi giorno aprendo un qualsiasi quotidiano o ascoltando il telegiornale ci imbattiamo in centinaia di affermazioni di questo tipo:



«oggi l'indice Mibtel è sceso di 3 punti»

«l'indice dei prezzi al consumo è salito di 0,8 punti»

«negli ultimi anni il tasso di natalità è in diminuzione»

«i sondaggi indicano che il consenso del Premier è al 31%»

«il tasso di disoccupazione è pari al 32%»

Nell'era moderna siamo continuamente sommersi da dati: per poter capire i vari aspetti della società occorre assorbire e interpretare una grande mole di informazioni

Una vasta gamma di segnali, cifre, opinioni, frammenti di pensiero, potrebbero però non solo non arricchire il panorama informativo rispetto ad un qualche fenomeno di interesse, ma al contrario aumentare l'incertezza sulle conoscenze oggettive che su di esso si acquisiscono



ABBIAMO BISOGNO DI UNA "BUSSOLA" CHE CI GUIDI...

02 – Che cosa è la Statistica

Unità n° 01

Il termine **Statistica** deriva dal latino **status** e originariamente indicava la scienza degli avvenimenti notevoli dello Stato

➔ l'esigenza di conoscere quantitativamente alcuni fenomeni fu sentita fin dall'antichità perché tale conoscenza serviva a meglio organizzare l'amministrazione della società: nel XVII secolo in Inghilterra W. Petty parlava di **Aritmetica Politica**

Molti studiosi hanno cercato di dare una definizione sintetica della Statistica:

(1768) - è *l'esposizione delle cose più notevoli di uno Stato* (G. Achenwall)

(1847) - è *la scienza dei fatti sociali espressi in cifre* (A. Moreau de Joannes)

(1895) - è *l'osservazione dei fenomeni che possono essere contati o espressi in cifre* (R. Mayo-Smith)

(1943) - è *la branca del metodo scientifico che concerne il trattamento dei dati ottenuti enumerando e misurando le proprietà di popolazioni naturali* (M. Kendall)

DEFINIZIONE (G. Leti, 1986):

La statistica è la scienza che analizza in termini quantitativi i fenomeni collettivi

03 – Perché studiare la Statistica

Unità n° 01

Il dato in sé non possiede valore informativo, o meglio lo possiede solo in forma potenziale

È grazie al processo di analisi, presentazione e diffusione dei risultati che dal **dato** si arriva alla **informazione**, e dal confronto di diverse informazioni, anche provenienti da fonti diverse, si passa poi alla **conoscenza**, attraverso la quale si può rispondere a domande sul *come* e *perché* dei fenomeni studiati

La Statistica può essere vista come un insieme di tecniche che hanno come scopo la conoscenza quantitativa di fenomeni della realtà che ci circonda

**La conoscenza quantitativa dei fenomeni è preliminare
e necessaria a qualsiasi processo decisionale**

04 – L'approccio statistico

Unità n° 01

La mente umana non è in grado di sintetizzare numericamente fenomeni che coinvolgono un numero elevato di osservazioni, riesce soltanto a percepirla singolarmente:

la Statistica aiuta a superare questo limite

Operazioni tipiche delle analisi statistiche sono:

- 1) il **conteggio** (quante famiglie hanno più di 5 componenti)
- 2) la **misurazione** (qual è il livello di produttività industriale conseguito nel 2007 dall'Italia)
- 3) la **classificazione** (suddivisione della forza lavoro per qualifica)
- 4) la **sintesi**, tramite modelli esplicativi, dei fenomeni reali (relazione sul risultato economico raggiunto dalla XYZ Spa nel 2007 e previsione per il 2008)

È importante definire innanzi tutto quali sono gli obiettivi della ricerca (cosa vogliamo studiare) e quindi delineare tutte le fasi da dover seguire per raggiungere tali obiettivi

05 – Il piano della ricerca statistica

Unità n° 01

Approccio “esplorativo” o “confermativo”



06 – Concetti preliminari: cosa studiamo e su chi

Unità n° 01

Supponiamo di voler studiare come un certo fenomeno si manifesta in un insieme definito di entità (ES. le caratteristiche socio-demografiche degli studenti iscritti a questo c. di laurea)



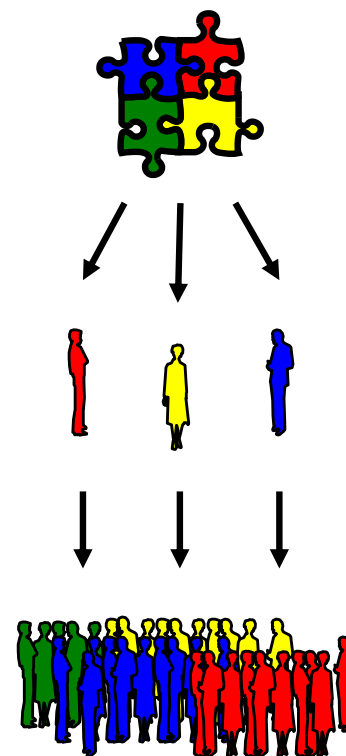
FENOMENO.....



UNITA' STATISTICA.....

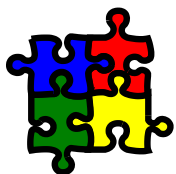


POPOLAZIONE.....



07 – I fenomeni e la realtà che ci circonda: *cosa studiamo?*

Unità n° 01



Un evento, naturale o provocato dall'uomo, può essere osservato direttamente (es. un evento demografico) o indirettamente (es. qualità della vita di una popolazione)

La statistica si interessa dei **fenomeni collettivi**:

- 1) Fenomeni che si riferiscono ad una collettività di casi singoli
(il fenomeno del consumo è collettivo in quanto la sua conoscenza avviene tramite l'osservazione di un collettivo, quello dei consumatori)
- 1) Fenomeni originati dalla ripetizione dell'osservazione di un unico fenomeno
(i prezzi di un certo prodotto rilevati in un certo arco temporale)

La **statistica descrittiva** è l'insieme dei metodi utilizzati per rilevare, sintetizzare e interpretare i dati espressione di un certo fenomeno d'interesse in un insieme di entità

Unità statistica → Popolazione



- 1 Unità elementare su cui viene studiato il fenomeno oggetto di studio
- 2 Insieme di unità omogenee rispetto ad una o più caratteristiche d'interesse

Una popolazione è diversa dalle singole unità che la compongono: la nascita e il decesso sono fatti individuali, la *natalità* e la *mortalità* sono fatti collettivi e quindi fenomeni di interesse per uno studio di tipo statistico

Una popolazione può essere:

un insieme di unità amministrative



i Comuni

un insieme di stabilimenti



le imprese manifatturiere

una superficie



il territorio di una regione

un insieme di eventi



i fatti criminosi in un periodo

...

...

un insieme di soggetti



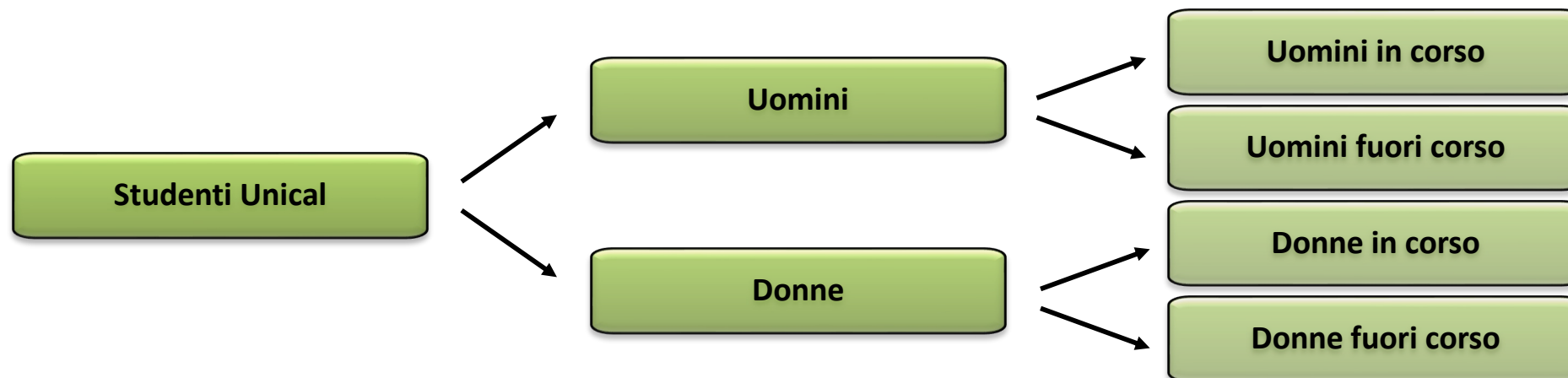
i componenti di una famiglia

09 – Popolazione e sottopopolazioni

Unità n° 01

Una popolazione statistica, definita anche universo o più comunemente **collettivo statistico** è un insieme di unità statistiche che hanno almeno una caratteristica in comune

È possibile che una popolazione omogenea rispetto ad una caratteristica possa poi essere suddivisa sulla base di altre informazioni, ad esempio il sesso, il titolo di studio...

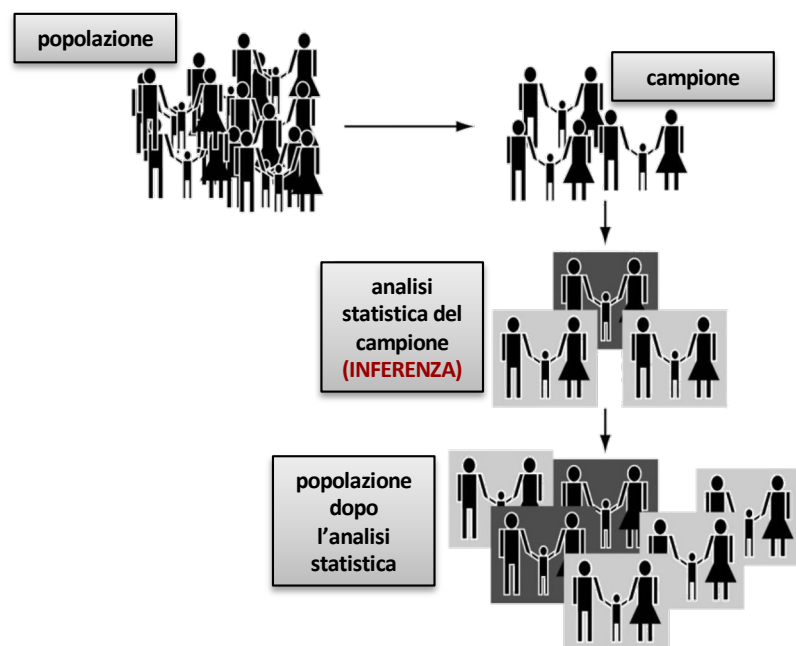


Ogni gruppo di unità di una popolazione, a sua volta omogeneo rispetto ad un altro aspetto, è detto **sotto-popolazione**: gli italiani di sesso maschile e gli italiani di sesso femminile sono due sottopopolazioni del collettivo formato da tutti i cittadini italiani

10 – Popolazione e campione

Unità n° 01

Se consideriamo un gruppo selezionato dalla popolazione in base a certi criteri prestabiliti parliamo di **campione**: alcuni consumatori scelti per un sondaggio telefonico su un prodotto



La **statistica inferenziale** è l'insieme dei metodi che utilizzano la probabilità per ottenere una stima di una caratteristica relativa a tutta la popolazione a partire dai risultati osservati su campione

Campionamento probabilistico (casuale):

insieme di tutte quelle tecniche di formazione del campione in cui la selezione delle unità è affidata a regole probabilistiche

Campionamento non probabilistico:

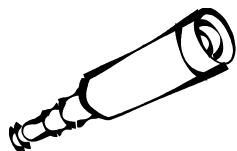
insieme di tutte quelle tecniche di formazione del campione in cui la selezione delle unità è affidata a specifiche esigenze di ricerca e/o alla mancanza di informazioni esatte sulla popolazione

11 – Come si raccolgono i dati

Unità n° 01

Esiste un'ampia gamma di possibilità tecniche per l'acquisizione dei dati, la cui scelta dipende da numerosi parametri:

- ◆ estensione degli scopi perseguiti
- ◆ ampiezza del collettivo di riferimento
- ◆ costo e budget disponibili per la raccolta dei dati
- ◆ grado di precisione desiderato



osservare...

è la via più diretta e immediata per studiare i comportamenti manifesti



domandare...

è la via obbligata per esplorare motivazioni, aspettative, credenze, sentimenti, percezioni, atteggiamenti

12 – Rilevazione dei dati

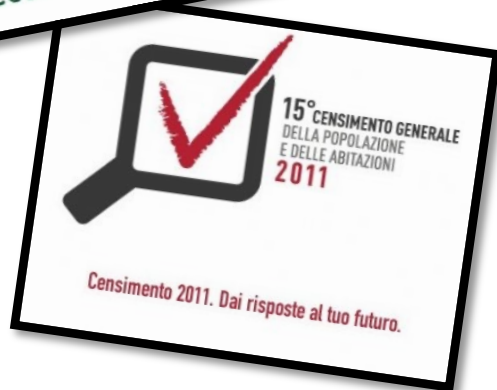
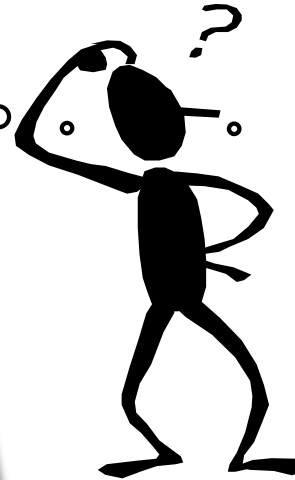
Unità n° 01

Rilevazione totale

Nella rilevazione totale si
ottiene la conoscenza esatta
del fenomeno analizzato

Rilevazione parziale

Nella rilevazione parziale
si ottiene una stima del
fenomeno analizzato



D'altra parte, bisogna anche considerare:



I tempi della rilevazione



I costi della rilevazione



La ricchezza di dettagli della rilevazione



Gli errori associati alla rilevazione



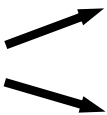
I tempi e i costi della rilevazione

Le indagini campionarie consentono di ottenere notevoli vantaggi competitivi fornendo risultati in tempi più brevi e con costi nettamente ridotti rispetto ad una indagine censuaria

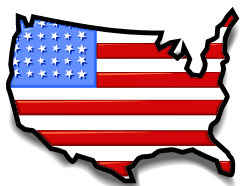


La ricchezza di dettagli della rilevazione

Le indagini campionarie si distinguono dalle indagini censuarie per la possibilità di andare in profondità nella ricerca dell'informazione

Rapidità  nel raccogliere e trattare i dati
nel pubblicare i risultati delle analisi

Analisi di eventi stagionali o periodici che richiedano interventi immediati (Occupazione, Attività produttive, Malattie, Migrazioni, ...)



1936: Elezioni Presidenziali U.S.A.

Candidati: F.D. Roosevelt, A. Landon



**Indagine
Literary Digest**

10 milioni di fac-simile di schede elettorali inviate a nominativi estratti dagli elenchi telefonici e dai registri automobilistici (2 milioni di risposte)

<i>Risultato previsto:</i>	<i>Roosevelt 41%</i>	<i>Landon 59%</i>
----------------------------	----------------------	-------------------

**Indagine
Gallup**

300.000 interviste ad elettori estratti in modo casuale dall'intera popolazione

<i>Risultato previsto:</i>	<i>Roosevelt 60%</i>	<i>Landon 40%</i>
----------------------------	----------------------	-------------------

Esito... **Roosevelt 61%**
Cosa ha sbagliato Literary Digest?



15 – I caratteri: come identificare e descrivere un fenomeno

Unità n° 01

Ciascun fenomeno può essere rappresentato per mezzo di una o più caratteristiche che lo identifichino (ES. il reddito per la ricchezza). In statistica ogni caratteristica è detta **carattere** (**X**) e può assumere **modalità** (**x**) differenti in corrispondenza delle diverse unità del collettivo

caratteri qualitativi:

le modalità sono espresse da attributi

caratteri quantitativi:

le modalità sono espresse numericamente

	Qualitativo		Quantitativo	
	Sconnesso	Ordinabile	Discreto	Continuo
<i>esempio</i>	<i>Genere Nazionalità</i>	<i>Qualifica prof. Titolo di studio</i>	<i>N. figli N. dipendenti</i>	<i>Peso Statura</i>

N.B. - Le modalità di un carattere devono essere esaustive e non sovrapposte

16 – Natura dei caratteri

Unità n° 01

CARATTERE QUALITATIVO o MUTABILE

sconnesso (non ordinabile)

(religione, luogo di nascita...)

non ammettono alcun ordine logico

ordinabile (rettilineo o ciclico)

(grado di soddisfazione, titolo di studio...)

hanno un ordine (logico) crescente

CARATTERE QUANTITATIVO o VARIABILE

discreto -> *enumerabile*

(n. di figli, età in anni...)

le modalità sono espresse da numeri interi

continuo -> *misurabile*

(peso, altezza...)

le modalità sono espresse da numeri reali

Possiamo anche classificarli come:

trasferibili (es. reddito, N. di dipendenti...)

non trasferibili (es. peso, età...)

Per semplicità d'ora in avanti non si parlerà di mutabili e variabili ma solo di **variabili qualitative** o **variabili quantitative** (tenendo ben presente i diversi tipi)

17 – Esempio

Unità n° 01

Analizzando la tabella riportata di seguito rispondere alle seguenti domande:

ID	nome azienda	tipo	settore	numero addetti
1	Adecco	Spa	Attività professionali	2123
2	Aeneas	Srl	Attività professionali	765
3	Air Liquide Welding Italia	Srl	Fabbricazione apparecchiature meccaniche	325
4	Andersen	Spa	Attività professionali	2711
5	Archimede Informatica	Srl	Produzione software	131
6	Asg	Srl	Informatica e attività connesse	24
7	Atos Origin	Spa	Informatica e attività connesse	58
8	Axiom Cambridge Limited	Spa	Informatica e attività connesse	234
9	Axioma	Spa	Produzione software	98
10	Azure	Spa	Informatica e attività connesse	12
11	Banksiel	Spa	Informatica e attività connesse	34
13	Barilla	Spa	Industrie alimentari	688
14	Bilab	Sas	Fabbricazione apparecchiature meccaniche	1187

- 1) Da chi è composto il collettivo e che dimensione ha
- 2) Quanti e quali sono i caratteri studiati
- 3) Che natura hanno i caratteri studiati
- 4) Quali sono le modalità dei caratteri studiati

- **Da chi è composto il collettivo e che dimensione ha:**

Il collettivo è costituito dalle diverse aziende ed ha una numerosità pari a $N=14$

- **Quanti e quali sono i caratteri studiati:**

Sono stati osservati tre caratteri, il *tipo di azienda* (assetto societario), il *settore di attività* e il *numero di addetti*

- **Che natura hanno i caratteri studiati:**

I primi due sono variabili qualitative sconnesse, il terzo è una variabile quantitativa discreta

- **Quali sono le modalità dei caratteri studiati:**

Le modalità del tipo di azienda sono i diversi assetti societari (Spa, Srl, e così via); le modalità del settore di attività sono i diversi settori in cui operano le aziende (ad es. Informatica e Attività Connesse); le modalità del numero di addetti sono i conteggi delle diverse unità lavorative impiegate dalle aziende

19 – Cosa è quantitativo, cosa è qualitativo

Unità n° 01

Identificare la natura del carattere in talune circostanze è un problema non banale: i numeri sono segni a cui semanticamente si attribuisce il significato di quantità o posizione in un insieme

Esempio:

*Franco è nato nel 1985, Luca nel 1990 e Gianni nel 1995: conto/misuro qualcosa?
Franco ha 37 anni, Luca 32 anni e Gianni 27 anni: conto/misuro qualcosa?*

Nel primo caso possiamo leggere l'anno di nascita come un attributo: di fatto può essere visto come un carattere qualitativo; nel secondo caso abbiamo l'età, carattere quantitativo

➔ **l'informazione apparentemente è la stessa, ma cambia la natura del carattere**

Non tutto i numeri sono espressione di un carattere quantitativo, ma non tutti gli attributi sono espressione di un carattere qualitativo

Esempio:

Franco ha trentasette anni, Luca ha trentadue anni, Gianni ha ventisette anni...

20 – Notazione

Unità n° 02

Popolazione di N elementi $\rightarrow P = \{u_1, u_2, \dots, u_i, \dots, u_N\}$

$u_i =$ i -esima unità statistica (per $i=1,2,\dots,N$)
indica una generica unità della popolazione

i = può assumere qualsiasi valore intero compreso tra 1 e N

N = numerosità (dimensione) della popolazione

Per i caratteri utilizzeremo le lettere maiuscole, mentre per le modalità quelle minuscole:

Carattere $\rightarrow X$ (qualitativo o quantitativo)

Modalità $\rightarrow x_1, x_2, \dots, x_i, \dots, x_N$

modalità del carattere X osservata sulla i -esima unità statistica

Una stessa modalità può essere osservata/misurata su più unità contemporaneamente

21 – L'osservazione di un fenomeno su un collettivo statistico

Unità n° 01

Dato statistico



osservazione del carattere
su una unità del collettivo

I **dati qualitativi** sono espressi in forma:
- di aggettivi (biondo, bruno, castano, ...)
- di locuzioni (addetti all'agricoltura, ...)

I **dati quantitativi** sono espressi da numeri:
- da un conteggio (numero di figli in una famiglia)
- da una misurazione (peso, temperatura, reddito)



22 – Operazioni su dati quantitativi

Unità n° 01

Consideriamo un carattere (ES. prezzo in €) osservato in due momenti diversi per un bene



2019	2022
€ 15	€ 25
€ 45	€ 55



La differenza tra il prezzo del 2022 e il prezzo del 2019 prende il nome di **variazione assoluta**

$$x_j - x_i$$

Il rapporto tra il prezzo del 2022 e il prezzo del 2019 prende il nome di **variazione relativa**

$$\frac{x_j}{x_i}$$

23 – Variazioni assolute e relative

Unità n° 01

La **variazione assoluta** ci dice come il fenomeno si è modificato in termini assoluti tra due momenti t e $t+1$ ed è espressa nella stessa unità di misura di quanto osservato



il prezzo del noleggio di un ombrellone e due lettini e di una cena al ristorante tra il 2019 e il 2022 è aumentato di 10 €

Ma possiamo dire che i due caratteri studiati si sono modificati nello stesso modo tra il 2019 e il 2022? Con la variazione assoluta no!

La **variazione relativa** ci dice come il fenomeno si è modificato in $t+1$ in relazione a come lo stesso si presentava nel momento iniziale t : ci dice quindi quanta parte del numeratore spetta ad ogni quantità unitaria espressa al denominatore del rapporto



il prezzo dei due servizi, in termini relativi, ha mostrato una variazione rispettivamente di 1.67 e 1.20 (per ciascun Euro speso nel 2019) nel periodo di tempo considerato

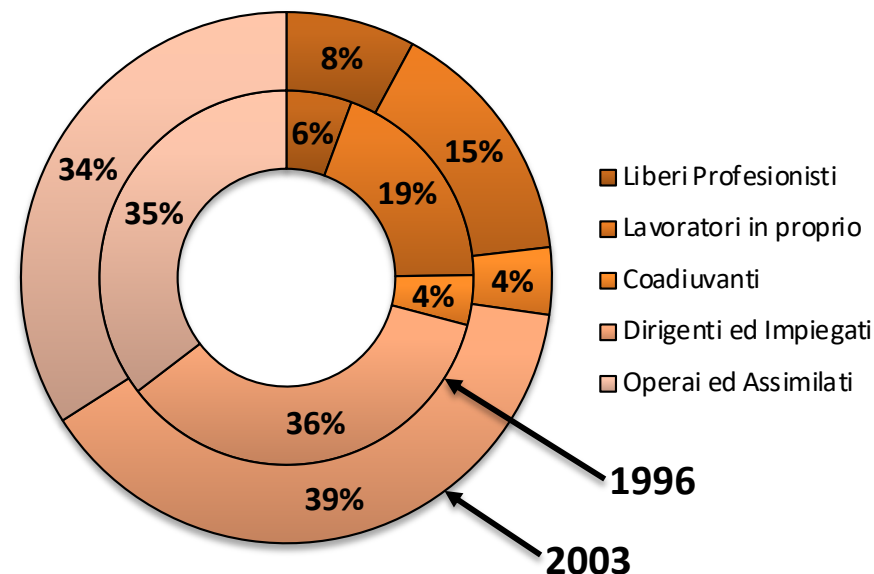
24 – Tassi di variazione

Unità n° 01

Se alla var. relativa sottraiamo 1, ed esprimiamo in percentuale il risultato, si ottiene il cosiddetto **tasso di variazione**: possiamo leggerlo come la *velocità* di cambiamento del fenomeno d'interesse

Occupati per posizione professionale (1996-2003)

	1996	2003	var. ass	var. rel
Liberi Professionisti	1138	1737	599	1.53
Lavoratori in proprio	3850	3371	-479	0.88
Coadiuvanti	866	901	35	1.04
Dirigenti ed Impiegati	7153	8540	1387	1.19
Operai ed Assimilati	7119	7506	387	1.05
Totale	22122	24058		



Dall'analisi dei dati sull'occupazione per posizione professionale risulta che c'è stata in termini assoluti una maggior variazione per dirigenti e impiegati, anche se in termini relativi il maggior incremento è stato osservato tra i liberi professionisti. Osservando la struttura dell'occupazione con i rapporti di composizione possiamo rilevare come ad es. il "peso" dei lavoratori in proprio dal 1996 al 2003 sia diminuito, mentre quello dei coadiuvati è rimasto uguale

Individuare negli esempi riportati di seguito il collettivo, le unità statistiche, il carattere oggetto di studio e la sua natura, quali potrebbero essere le modalità del carattere:

- 1) Gli studenti iscritti al primo anno delle diverse facoltà dell'UNICAL
- 2) I premi pagati dagli assicurati di una certa Compagnia
- 3) Il titolo di studio conseguito dagli impiegati di un certo stabilimento
- 4) Il ritardo accumulato dai treni in arrivo in una certa stazione
- 5) I consumi alimentari delle famiglie residenti a Milano
- 6) La percentuale di disoccupati rispetto alla popolazione attiva delle regioni italiane

26 – Soluzione

Unità n° 01

- **Gli studenti iscritti al primo anno delle diverse facoltà dell'UNICAL**

Il collettivo è costituito da tutte le facoltà, quindi ogni facoltà rappresenta una unità statistica; il carattere osservato è il numero di studenti iscritto al primo anno, ed è una variabile quantitativa discreta; le modalità sono date dal conteggio degli studenti

- **I premi pagati dagli assicurati di una certa Compagnia**

Il collettivo è costituito dagli assicurati che hanno sottoscritto una polizza con la Compagnia; il carattere osservato è l'ammontare del premio pagato, ed è una variabile quantitativa continua; le modalità sono date da tutti i possibili premi pagati

- **Il titolo di studio conseguito dagli impiegati di un certo stabilimento**

Il collettivo è costituito dagli impiegati dello stabilimento; il carattere osservato è il titolo di studio, ed è una variabile qualitativa rettilinea; le modalità sono i diversi titoli di studio che è possibile conseguire

- **Il ritardo accumulato dai treni in arrivo in una certa stazione**

Il collettivo è dato da tutti i treni che arrivano nella stazione; il carattere studiato è il ritardo accumulato da ciascun treno, ed è una variabile quantitativa continua; le modalità del carattere sono i diversi ritardi che ogni treno può accumulare (da 0 in su)

- **La spesa per consumi alimentari delle famiglie residenti a Milano**

Il collettivo è composto da tutte le famiglie residenti nel comune di Milano, e l'unità statistica è la famiglia (anche se costituita da più individui); il carattere studiato è l'ammontare di spesa per consumi alimentari, ed è una variabile quantitativa continua; le modalità sono tutti i possibili livelli di spesa

- **La percentuale di disoccupati rispetto alla popolazione attiva delle regioni italiane**

Il collettivo è composto da tutte le regioni italiane; il carattere studiato è la percentuale di disoccupati, ed è una variabile quantitativa continua; le modalità sono i diversi livelli di disoccupazione tra 0% e 100% (rispetto alla popolazione attiva)