

Percorso introduttivo alla statistica ufficiale

SAPER LEGGERE I DATI

INDICE

- 1) La statistica nella vita quotidiana
- 2) I metadati: le informazioni per capire i dati
- 3) Esempi di uso “improprio” della statistica
- 4) Grafici “fuorvianti”

La statistica nella vita quotidiana

La statistica è

“... un insieme di regole con le quali è possibile pervenire ad una percezione – o ad una più accurata percezione – di fenomeni/caratteri osservabili che, per loro natura, si manifestano con un certo grado di variabilità

Di fatto, laddove la realtà si esprime in modo complesso, al punto che l’occhio fatica a leggerne con chiarezza i molteplici aspetti e le interrelazioni, la statistica offre gli strumenti per mettere a punto un appropriato “paio di occhiali”: un utile supporto che, come è buona norma, va comunque adottato solo dopo aver letto e compreso le appropriate prescrizioni d’uso...”

Blangiardo G.C. (2001), “Statistica, sondaggi e buon senso”

Statistica e vita quotidiana

Le statistiche fanno parte della nostra vita quotidiana:

Tabelle, grafici, indicatori, medie, ... ci aiutano a rappresentare in maniera sintetica il mondo in cui viviamo e a prendere decisioni in condizioni di incertezza

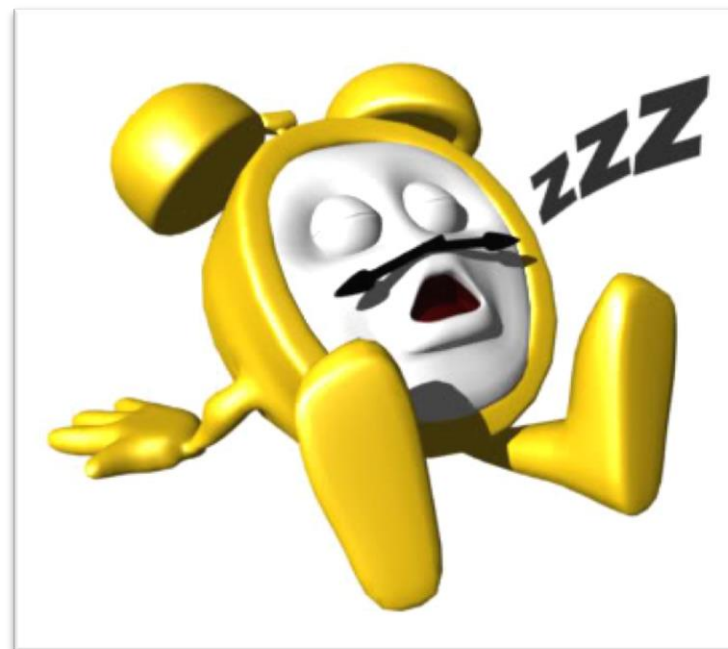
Concorrono a determinare la nostra visione del mondo, a formare il senso comune

» Senso comune: maniera ordinaria e semplice, propria della maggior parte della gente, di intendere e giudicare (Il nuovo Zingarelli)

Statistica e vita quotidiana

La statistica fa parte della nostra vita, a partire dalla sveglia al mattino. Ognuno la regola sulla base della propria esperienza, per arrivare in orario, senza doversi affrettare per non arrivare in ritardo ma senza rinunciare al sonno "più del necessario".

Insomma, ognuno di noi ha "fatto una statistica" per puntare la sveglia ... anche senza rendersene conto ...



Statistica e vita quotidiana

I **biscotti che mangiamo a colazione** sono il frutto di studi statistici sui gusti dei consumatori

Anche la loro disposizione sugli scaffali del supermercato nasce da analisi statistiche sui comportamenti in fase di acquisto



La statistica è alla base della **programmazione televisiva**: le fasce orarie dei programmi e la loro eventuale riproposizione sono determinati sui dati di audience

Come viene percepita la statistica?

Spesso la **statistica** viene **percepita** in base a **due atteggiamenti opposti**:

- come **strumento di manipolazione della realtà**, la cui varietà di manifestazioni non può essere ricondotta entro la "gabbia" di rappresentazioni sintetiche e semplicistiche

*siconno le statistiche d'adesso
risulta che te tocca un pollo all'anno
e, se nun entra nelle spese tue
t'entra nelle statistiche lo stesso
perché c'è un antro che se magna due*

(Trilussa)

- come **verità assoluta ed inconfutabile**
Il dato è di per sé eloquente; il dato è la realtà oggettiva

Questo perché la sintesi produce:

oltre a un GUADAGNO

Le misure di sintesi (es. le medie) permettono di avere informazioni sintetiche su collettivi numerosi



Numero medio di libri letti in un anno = 4

... anche una PERDITA

Le misure di sintesi fanno perdere informazioni sulle singole unità osservate



Studente	Libri letti in un anno
Giacomo	6
Mirella	6
Luca	0
Andrea	5
Valerio	3
Martina	6
Anna	2

I Metadati:

le informazioni per capire i dati

I metadati

- Per una corretta lettura dei dati statistici è necessaria la **metainformazione**, che consiste in informazioni sulle procedure seguite per la raccolta ed il trattamento dei dati, sulle definizioni e le classificazioni utilizzate, sulla tipologia dell'indagine (totale o campionaria), etc.
- Al di là dei numeri e delle sintesi che la statistica produce, assumono fondamentale importanza le informazioni "aggiuntive" che raccontano "come" si è arrivati a tali conclusioni, ad es.:
 - quanti casi abbiamo osservato? Tutta la popolazione o solo una sua parte?
 - Che elaborazioni? Quali calcoli sono stati effettuati?

Uno dei principi fondamentali della statistica pubblica

Le istituzioni statistiche devono avere il diritto di fornire commenti sulle interpretazioni erranee e sulle utilizzazioni non corrette

È il IV dei principi fondamentali della statistica ufficiale adottati dall'ONU e implica un duplice diritto-dovere:

- 1) la statistica pubblica deve opporsi a qualsiasi tipo di interpretazione errata
- 2) gli organi della statistica pubblica sono autorizzati a commentare interpretazioni fuorvianti e utilizzazioni abusive dei loro risultati statistici

Esempi di uso "improprio" della statistica

Ciononostante a volte i dati statistici vengono usati impropriamente

- **per (di)mostrare una certa verità, quando magari non è supportata da alcun dato e anzi sembra essere smentita**

ES.: una casa farmaceutica vuole commercializzare un nuovo collutorio contro il mal di gola e afferma che da analisi di laboratorio risulta che 10 mg di sostanza attiva uccidono 30.000 microbi

È corretto? Forse sì, ma viene spontaneo chiedersi:

- a) quanto descritto accade in laboratorio, ma in una gola umana? Sarebbe ugualmente efficace?
- b) una volta che la sostanza è diluita, per evitare che mi bruci i tessuti della gola, risulta ancora efficace?
- c) ma sono proprio quelli i microbi che causano il mal di gola?

Ciononostante a volte i dati statistici vengono usati impropriamente

- per “fare sensazione”

Es.:

Quando, molti anni fa, la John Hopkins University (USA) iniziò ad accettare anche le donne come studenti, qualcuno pensò di riportare la notizia secondo la quale il **33,3%** delle studentesse aveva sposato un insegnante

... Ma a quell'epoca le donne iscritte erano solo tre ed una aveva sposato un professore

Ciononostante a volte i dati statistici vengono usati impropriamente

- a causa di banali errori

ES.:

“Un giorno, in occasione di una eclissi, in un telegiornale nazionale italiano fu spiegato che il fenomeno avviene quando il sole si frappone fra la terra e la luna. Quell’ipotesi di catastrofe cosmica fu ripetuta due volte – e a nessuno venne in mente di correggerla. [...] è un esempio di come un errore clamoroso possa passare inosservato.”

Giancarlo Livraghi – Capire con le statistiche

I numeri che “mancano”

Quando le statistiche “condensano” le informazioni e sintetizzano i dati di un collettivo, è fondamentale conoscere:

- sia il **valore assoluto** (l’aspetto dimensionale) del fenomeno;
- sia le **percentuali**, ovvero il “contributo” che ciascuna parte apporta al collettivo nel suo complesso

Esempio:

Studentesse	Frequenze assolute
Hanno sposato un insegnante	1
NON hanno sposato un insegnante	2
TOTALE	3



$$\frac{1}{3} \cdot 100 = 33,3\%$$

In un certo anno le studentesse erano 3 ed una di loro ha sposato un insegnante

Valore ASSOLUTO e PERCENTUALE: l’informazione è corretta quando sono forniti entrambi

Se il carattere è “troppo concentrato”

Di **16 donne** di successo americane selezionate dalla Camera di Commercio di Boston intorno alla metà del XX secolo, fu detto che esse **possedevano 60 lauree ed avevano 18 figli**

... Ma nel gruppo erano presenti due donne “particolari”, Virginia Gildersleeve, presidente del Barnard College, e Lillian M. Gilbreth, nota, insieme al marito, nello sviluppo delle tecnologie industriali

Le due, insieme, presentavano un terzo di tutti i titoli accademici e 12 dei “18 figli” erano della signora Gilbreth

Quale è il numero medio di figli per donna?

Riprendendo l'esempio delle 16 donne di successo e dei loro 18 figli:

- Se manteniamo nel collettivo tutte le donne con la rispettiva prole:

$$\frac{18}{16} = 1,12 \quad \text{figli per donna}$$

- Se dal gruppo escludiamo la signora Gilbreth con i suoi 12 figli:

$$\frac{6}{15} = 0,4 \quad \text{figli per donna}$$



**Nemmeno
"mezzo" figlio per
donna!!!**

Cosa pensereste se vi dicessi che ...

Tavola 1.1.1 - Attività svolte in un giorno medio settimanale dalla popolazione di 15 anni e più per tipo di attività e alcune caratteristiche

Anni 2002-2003 - (durata media generica in ore e minuti e quota percentuale di tempo sulle 24 ore)

CARATTERISTICHE	Dormire, mangiare e altra cura della persona		Lavoro retribuito		Istruzione e formazione	
	Mg	% sulle 24 ore	Mg	% sulle 24 ore	Mg	% sulle 24 ore
TOTALE	12:01	50,0	2:36	10,8	0:25	1,8

... nel 2002-2003 la popolazione italiana di 15 anni e più lavorava – in modo retribuito – in media, 2 ore e 36 minuti al giorno?

Fonte: Istat, 2007, *L'uso del tempo. Indagine multiscopo sulle famiglie "Uso del tempo"* - Anni 2002-2003, Istat, Roma (Informazioni, n.2).

Osservare i risultati con maggiore attenzione

Tavola 1.1.10 - Attività svolte nei giorni feriali (lun.-ven.) dalla popolazione di 15 anni e più per tipo di attività e alcune caratteristiche

Anni 2002-2003 (*durata media specifica in ore e minuti e frequenza di partecipazione in percentuale*)

CARATTERISTICHE	Dormire, mangiare e altra cura della persona		Lavoro retribuito		Istruzione e formazione	
	Ms	%	Ms	%	Ms	%
TOTALE	11:43	100,0	7:41	41,5	6:04	8,2
CARATTERISTICHE DEMOGRAFICHE						
CLASSI DI ETÀ						
15-24	11:31	100,0	7:28	26,4	6:33	50,3
25-44	10:58	100,0	7:48	66,6	4:49	4,3
45-64	11:23	100,0	7:38	45,5	2:35	0,9
65 e più	13:30	100,0	6:31	3,6	1:48	0,2

Fonte: ISTITUTO NAZIONALE DI STATISTICA, 2007a. *L'uso del tempo. Indagine multiscopo sulle famiglie "Uso del tempo"* - Anni 2002-2003. Roma: Istat. (Informazioni, n.2). Disponibile su <http://www.istat.it/dati/catalogo/20070301_00/> [25 ottobre 2010]

Come si spiegano queste differenze?

I dati apparentemente **contrastanti** delle due tabelle poste a confronto si spiegano prendendo in considerazione i seguenti fattori:

- **Giorno settimanale di riferimento**
 - tutti i giorni della settimana
 - dal lunedì al venerdì
- **“tipo” di media utilizzata:**
 - media **generica**
 - media **specificata**

Giorno medio settimanale e media generica

Tavola 1.1.1 - Attività svolte in un giorno medio settimanale dalla popolazione di 15 anni e più per tipo di attività e alcune caratteristiche



**Durata media
giornaliera
dell'attività
lavorativa retribuita
= 2h 36 min**



- E' riferita al **giorno medio settimanale**, "alla cui costruzione concorrono tutti i giorni della settimana", compresi i sabati e le domeniche
- **La media è generica**

Giorno feriale e media specifica

Tavola 1.1.10 - Attività svolte nei giorni feriali (lun.-ven.) dalla popolazione di 15 anni e più per tipo di attività e alcune caratteristiche



Durata media giornaliera dell'attività lavorativa retribuita = 7h 41 min



- E' riferita esclusivamente ai **giorni feriali**, ovvero considera solo i giorni dal lunedì al venerdì
- **La media è specifica**

Media generica e media specifica

➤ media generica

“Nel calcolo delle medie generiche, le durate sono riferite al totale della popolazione (...). Per esempio, la durata media generica di un'attività indica il tempo mediamente dedicato a tale attività **da tutta la popolazione, compresi quanti non l'hanno svolta**”

➤ media specifica

“Tale indicatore è calcolato **solo sull'insieme** della popolazione **che effettivamente ha svolto un'attività**”

Grafici “fuorvianti”

Le rappresentazioni grafiche

Sono **strumenti statistici** che permettono di:

- interpretare **più velocemente le informazioni** raccolte sul fenomeno osservato,
- di coglierne **immediatamente** alcune **caratteristiche**

Ciò **non** significa che i grafici possano **sostituire** i numeri presenti nelle tabelle: devono solo fornire un ulteriore ma utile supporto per l'analisi statistica

Rappresentazione grafica o tabellare?

Alcuni vantaggi che i grafici presentano rispetto alle tabelle che corredano sono:

- **visualizzazione immediata dell'andamento** del fenomeno (es.: è in crescita oppure è in diminuzione?) e **della struttura** della distribuzione (es.: sono più i maschi o le femmine?), che consente **una globale descrizione** dei dati
- sintesi e quindi possibilità, in poco spazio, di confrontare più distribuzioni (curve, spezzate, ecc.)
- **forma più divulgativa** per i dati statistici rispetto a quanto consentito dalla forma tabellare

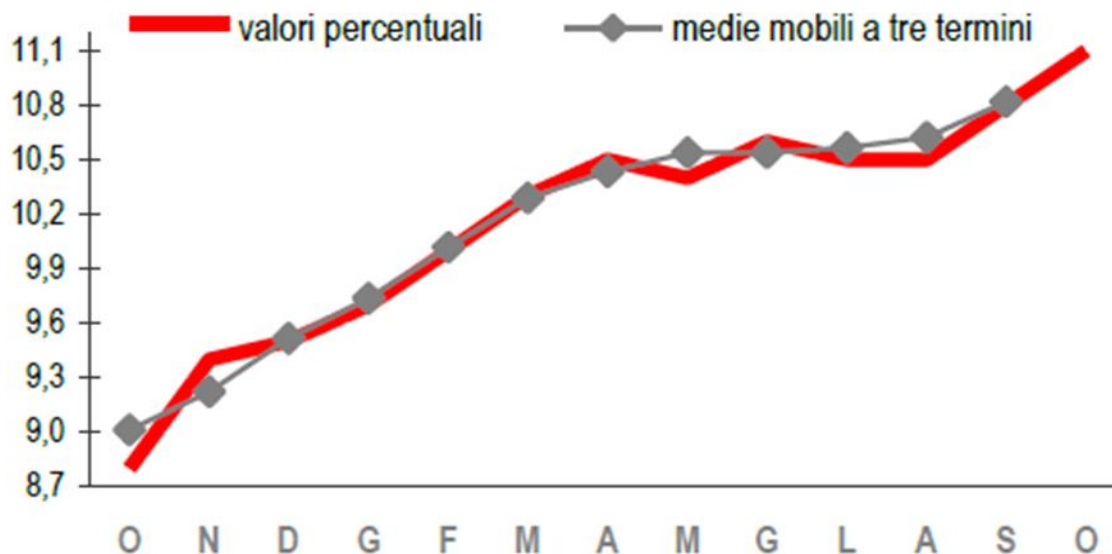
Ad ogni modo è bene ricordare che ...

... affinché una rappresentazione grafica sia utile ed efficace deve contenere tutte le informazioni necessarie alla comprensione dei dati in essa rappresentati, ovvero:

- il **titolo**, che deve indicare l'**oggetto**, il **luogo** e l'**epoca** a cui i dati si riferiscono
- il **carattere con le rispettive modalità** (es.: "maschi" e "femmine" per la variabile "sesso"), in funzione delle quali sono classificate le unità statistiche
- l'**unità di misura** impiegata per graduare gli assi
- la **fonte di provenienza** dei dati

Ma anche quando un grafico è "ben fatto"...

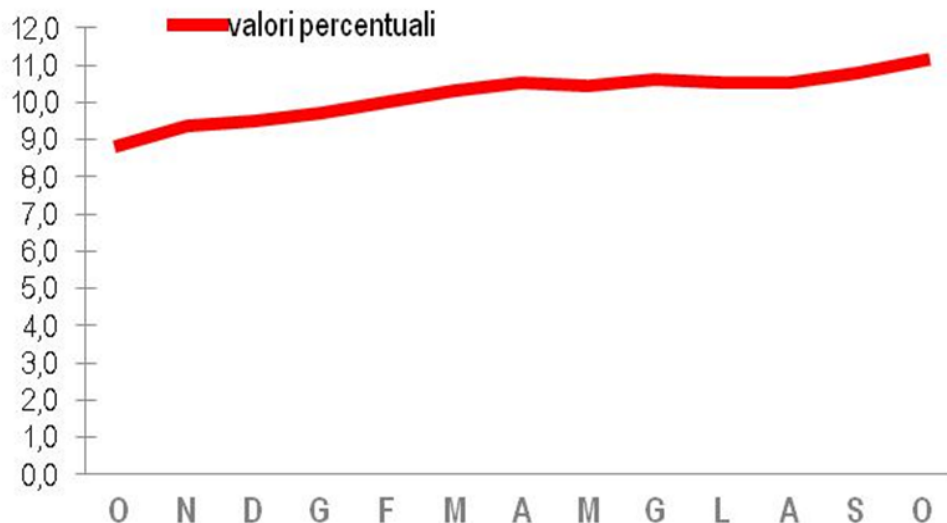
TASSO DI DISOCCUPAZIONE. Ottobre 2011- ottobre 2012, dati destagionalizzati, valori percentuali



... se non lo si osserva attentamente se ne può ricavare una percezione del fenomeno distorta!

Il fenomeno è in aumento: quanto?

I due grafici mostrano **entrambi l'ascesa** del tasso di disoccupazione da ottobre 2011 a ottobre 2012



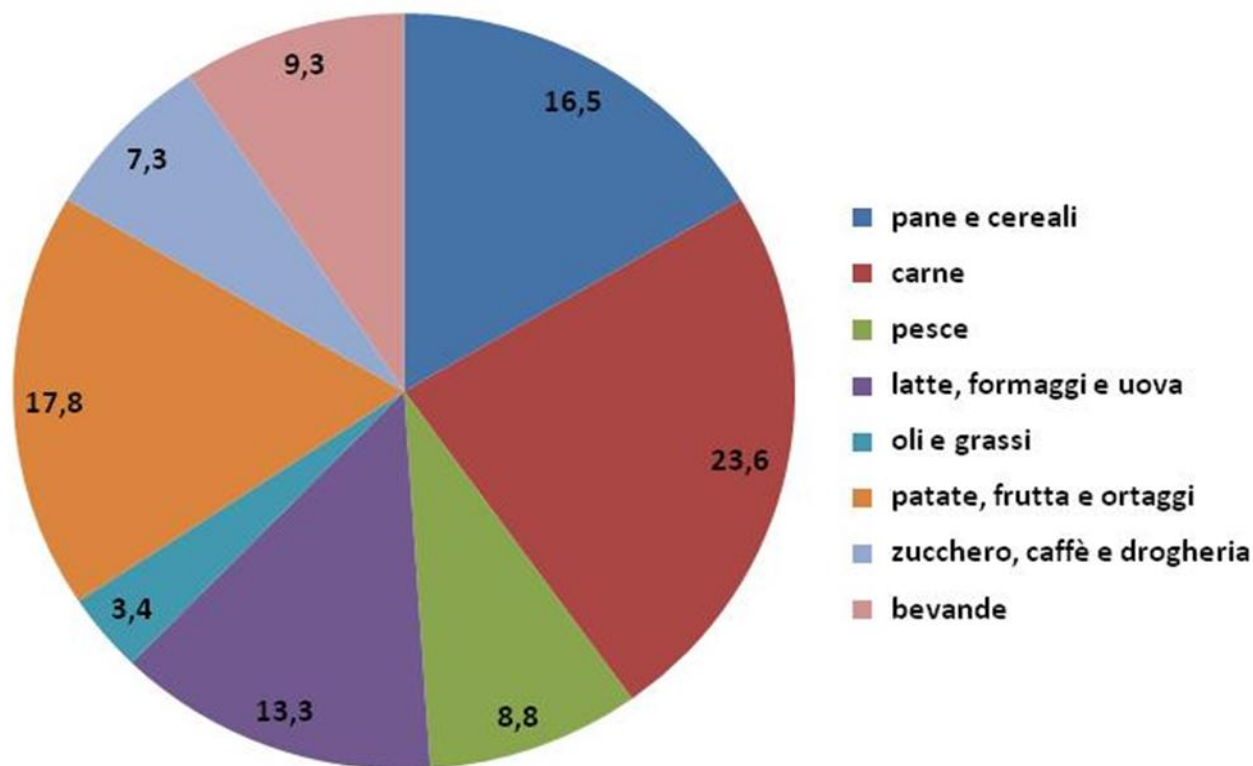
L'unica differenza tra i due è nel valore di partenza dell'asse delle ordinate: al grafico in alto (quello della pubblicazione) manca la parte tra l'origine (zero) e il valore 8,7

La tabella "sottostante"

anno	mese	tasso di disoccupazione (%)
2011	Ottobre	8,8
	Novembre	9,4
	Dicembre	9,5
2012	Gennaio	9,7
	Febbraio	10,0
	Marzo	10,3
	Aprile	10,5
	Maggio	10,5
	Giugno	10,6
	Luglio	10,5
	Agosto	10,5
	Settembre	10,8
	Ottobre	11,1

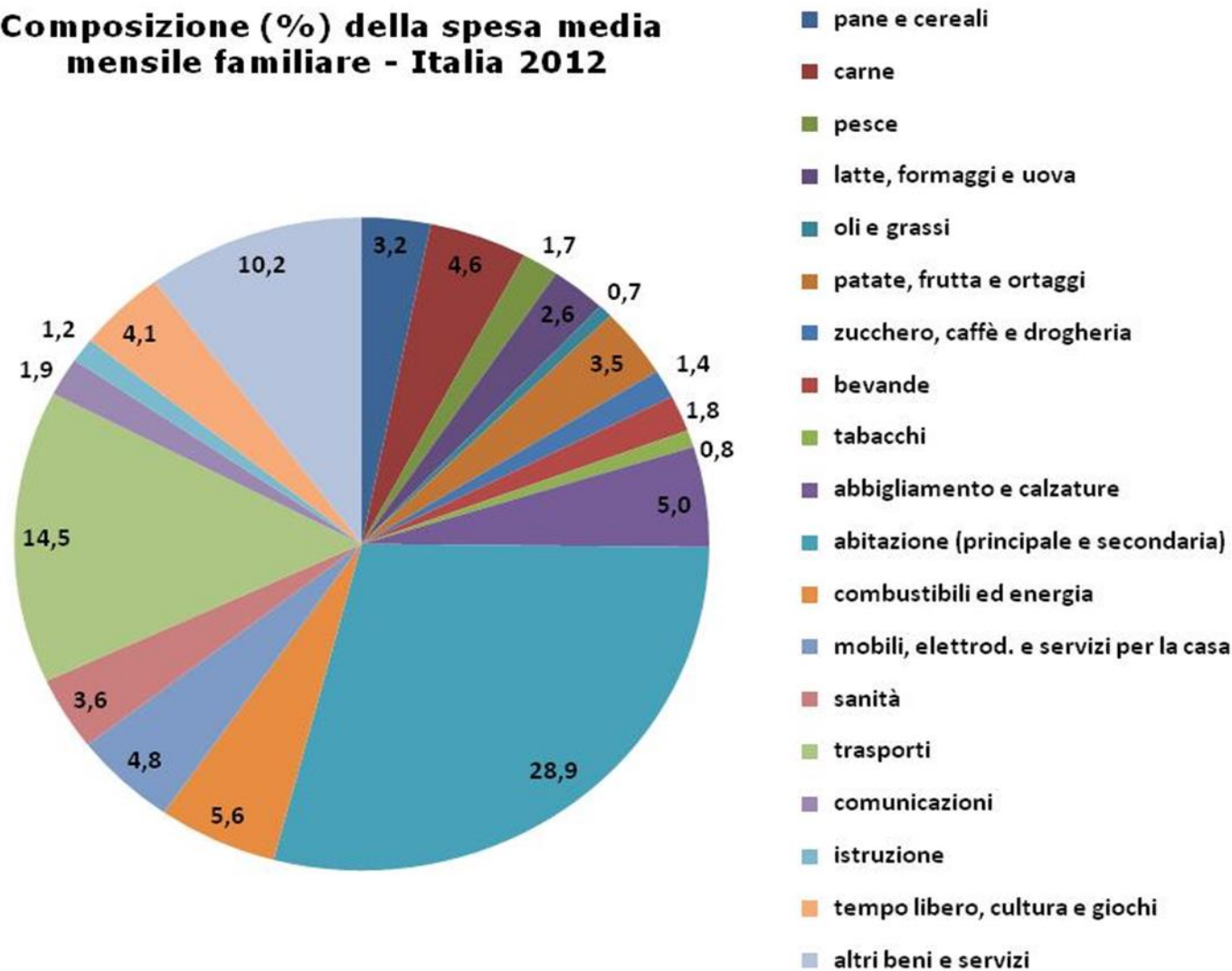
I grafici a torta sono utili per rappresentare il contributo di ogni parte alla formazione del totale

Composizione (%) della spesa media mensile familiare per alimentari e bevande - Italia 2012



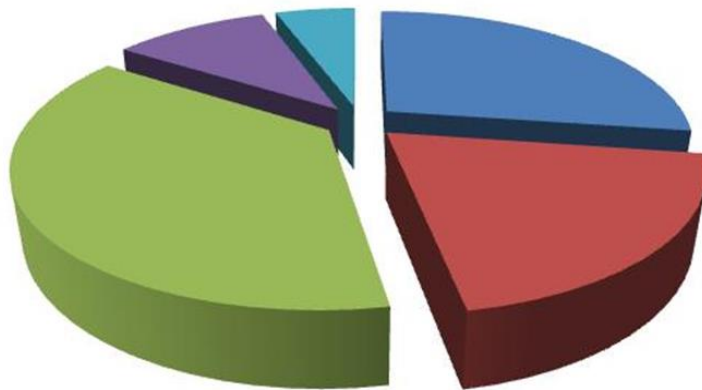
Troppe categorie!

Composizione (%) della spesa media mensile familiare - Italia 2012



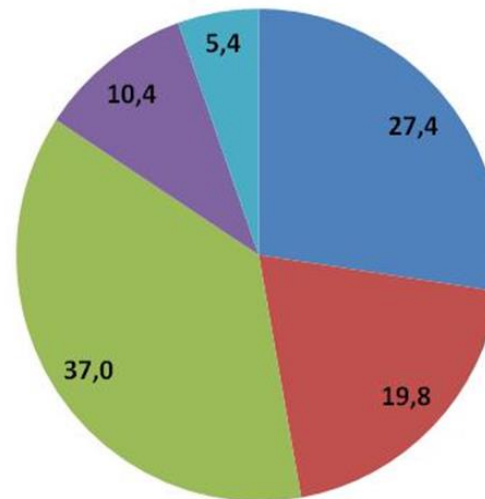
Le “torte esplose” possono ingannare la percezione del fenomeno!

Numero di prestiti (% sul totale) effettuato dalle biblioteche pubbliche statali - Italia 2010



■ Nord-ovest ■ Nord-est ■ Centro ■ Sud ■ Isole

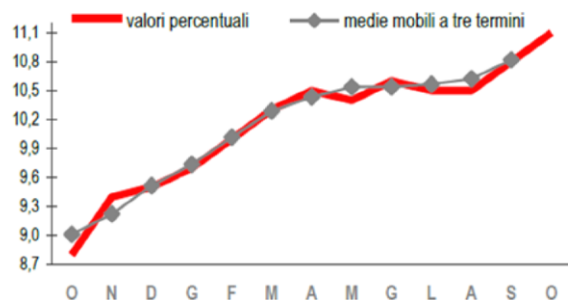
Numero di prestiti (% sul totale) effettuato dalle biblioteche pubbliche statali - Italia 2010



■ Nord-ovest
■ Nord-est
■ Centro
■ Sud
■ Isole

Conclusioni...

TASSO DI DISOCCUPAZIONE. Ottobre 2011- ottobre 2012,
dati destagionalizzati, valori percentuali



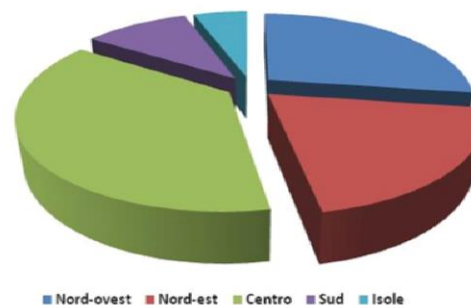
Studente	Libri letti in un anno
Giacomo	6
Mirella	6
Luca	0
Andrea	5
Valerio	3
Martina	6
Anna	2



Numero medio di libri
letti in un anno = 4

... affinché la statistica e i suoi strumenti possano rivelarsi utili ed essere impiegati in modo proficuo, è bene imparare a dare ai numeri "una seconda occhiata"!!!

Numero di prestiti (% sul totale) effettuato
dalle biblioteche pubbliche statali - Italia 2010



...e adesso...
buon lavoro!

Rete per la promozione
della cultura statistica

Per ulteriori moduli didattici relativi alle scuole secondarie di secondo grado si prega di consultare anche la piattaforma Scuola di statistica – Lab (accessibile dal link <http://scuoladistatistica-lab.istat.it/>)