CAMPIONAMENTO STATISTICO E DISTRIBUZIONI CAMPIONARIE

PIANIFICAZIONE DI UNA RICERCA O INDAGINE STATISTICA

Piano della Rilevazione

A partire dalle ipotesi di ricerca, bisogna:

1. individuare con precisione la popolazione, l'unità statistica e l'unità di rilevazione;
2. stabilire i caratteri quantitativi e/o qualitativi da rilevare per ogni unità;
3. indicare i mezzi tecnici per raccogliere le informazioni su tali caratteri;
4. fissare l'estensione della rilevazione in ordine al territorio, al periodo, alle disponibilità finanziarie.

Rilevazioni Statistiche

Due tipi di rilevazioni statistiche:

1. RILEVAZIONI TOTALI o CENSIMENTI
   vengono osservate tutte le unità della popolazione in merito al fenomeno in esame;
2. RILEVAZIONI PARZIALI o CAMPIONARIE
   viene osservato un sottoinsieme delle unità che costituiscono la popolazione.

Fasi dello studio di un fenomeno collettivo

1. *Rilevazioni totali o censimenti*: la collettività di individui è interamente da osservare; si ha pertanto:
   - schematizzazione del fenomeno: precisazione delle caratteristiche che si intendono studiare;
   - osservazione: raccolta di dati, ordinamento, sistemazione in tabelle;
   - descrizione con metodi analitici.

2. *Rilevazioni parziali o campionarie*: è possibile (o conviene) osservare solo una parte della popolazione di riferimento; si ha pertanto:
   - schematizzazione del fenomeno: precisazione delle caratteristiche che si intendono studiare;
   - formulazione delle ipotesi;
   - osservazione: raccolta di dati, ordinamento, sistemazione in tabelle;
   - descrizione con metodi matematici;
   - inferenza: si cercano informazioni su caratteristiche della popolazione a partire dai valori rilevati nel campione.

Nello studio si possono distinguere anche una:

1. fase metodologica in cui si individuano gli scopi dell'indagine, la popolazione da indagare e le informazioni (di natura quantitative e/o qualitativa) da raccogliere;
2. fase tecnica che riguarda la raccolta delle informazioni e la loro verifica; tali dati vengono in genere riportati in forma di matrici;
3. fase di pubblicazione che comprende l'analisi dei dati ottenuti, la loro presentazione e discussione.

RILEVAZIONI CAMPIONARIE

Per campione si intende quel gruppo di unità elementari, sottoinsieme particolare della popolazione, individuato in essa in modo da consentire, con un rischio definito di errore, la generalizzazione DI risultati di analisi all'intera popolazione. Le conclusioni a cui si arriva possono, non "con certezza" ma con "una certa probabilità", attraverso procedimenti statistici di inferenza, essere attribuite all'intera popolazione.

I principali vantaggi di un'analisi su dati campionari, che a prima vista potrebbe apparire limitata e non esaustiva, possono essere sintetizzati nei seguenti quattro punti.

- *Costi ridotti.*

Se si osservano le manifestazioni di un fenomeno analizzando un sottoinsieme della popolazione i costi complessivi per l'acquisizione dei dati risultano, evidentemente, inferiori rispetto a quelli che si sosterrebbero se si effettuasse il censimento di tutte le unità della popolazione. Di fatto, oggigiorno, il ricorso ai censimenti di tutta la popolazione italiana viene fatto quasi esclusivamente dall'ISTAT, ogni 10 anni, per ottenere un quadro delle principali caratteristiche socio-economico-demografiche dell'intera popolazione. Tutte le altre indagini in ambito sociosanitario sono effettuate quasi sempre su campioni di popolazione.

- *Maggiore rapidità di acquisizione dati.*

I dati e le informazioni che si intendono raccogliere sono più rapidamente accessibili con rilevazioni parziali piuttosto che con quelle totali. La tempestività nel raccogliere i dati risulta di notevole rilevanza quando le informazioni e i risultati sono necessari nel più breve tempo possibile.

- *Maggiore possibilità di azione*.

In alcune indagini poi, la raccolta dei dati, va affidata a gruppi di esperti, a persone altamente qualificate e specializzate che sono frequentemente di difficile reperibilità. In questi casi quindi, l'attuazione di un censimento risulta assolutamente impraticabile e pertanto l'indagine campionaria appare come l'unica via per ottenere le informazioni sul fenomeno che s'intende analizzare.

- *Maggiore accuratezza.*

In conseguenza di quanto detto sopra si evidenzia che in talune situazioni l'analisi risulta più approfondita in presenza di una numerosità limitata. Il campione permette allora lo svolgimento dell'indagine in maniera più accurata di quanto non lo permetterebbe uno studio complessivo di tutte le unità della popolazione in studio.

L'elencazione dei vantaggi delle indagini campionarie non facciano pensare che si possa individuare un campione che sia il "migliore" (il più rappresentativo) di tutti nel senso che permetta di ottenere il più basso costo, la maggiore rapidità di esecuzione e la massima accuratezza.

Le indagini campionarie possono, sommariamente, essere classificate in due tipi: descrittive e analitiche. Le prime mirano semplicemente ad ottenere informazioni su ampi gruppi di unità: ad esempio il numero di donne, uomini e bambini che ricorrono ai servizi offerti dalle U.S.L.; le seconde hanno come obiettivo quello di effettuare confronti tra sottogruppi di una popolazione al fine di scoprire eventuali differenze tra loro e di verificare alcune ipotesi o formularne delle altre circa le ragioni di tali differenze.

Indipendentemente dallo scopo che l'indagine si prefigge va sempre elaborato un "piano di campionamento" che costituisce una delle principali fasi di una indagine campionaria.

Piano di campionamento

Si definisce piano (o anche disegno) di campionamento un metodo attraverso il quale si selezionano gli elementi che entrano a far parte del campione. I metodi per selezionare le unità da campionare possono essere diversi e più o meno complessi; la scelta di uno di essi dipende da vari fattori: oltre ai problemi di costo, di tempestività e di precisione, va considerato anche la disponibilità o meno di una lista dell'unità da campionare, la presenza di informazioni sui caratteri della popolazione ecc.

Nel piano di campionamento si stabilisce sia il metodo attraverso cui si estraggono gli elementi che entreranno a far parte del campione, sia la dimensione dello stesso.

Metodi di campionamento

Il procedimento di campionamento delle unità da sottoporre ad analisi consiste nel definire la metodologia che sta alla base della scelta di tali unità.

Esistono diversi tipi di campioni che possono essere usati in dipendenza del tipo di indagine che si deve effettuare; la loro differenziazione consiste nel modo secondo il quale sono scelti gli elementi che comporranno il campione stesso.

Una prima grande distinzione che occorre tener presente è quella relativa a:

- ***campioni probabilistici***
- ***campioni non probabilistici***.

CAMPIONAMENTI PROBABILISTICI

Sono probabilistici quei campioni le cui unità vengono estratte dalla popolazione in modo tale che ogni elemento abbia una probabilità nota di entrare a far parte del campione stesso.

Dal punto di vista statistico i campioni probabilistici sono gli unici che giustificano il ricorso all'inferenza statistica in quanto i campioni non probabilistici, pur potendo essere campioni rappresentativi, non permettono di valutare il grado di errore in cui si incorre, in quanto per essi non sono applicabili gli schemi del Calcolo delle Probabilità. L'utilizzo di campioni probabilistici avviene anche allo scopo di escludere la possibilità che la scelta del campione risulti "influenzata" in qualche modo dal ricercatore stesso. Una scelta distorta potrebbe portare ad analizzare un campione affetto da "errore sistematico" (bias).

I metodi di selezione di campioni probabilistici possono essere diversi. Si distinguono**: campione casuale semplice, campione stratificato, campione a grappoli, campione sistematico**, **campione a più stadi** ecc. a seconda del modo con il quale vengono selezionate le unità campionarie.

Campionamento casuale semplice

Il campione casuale prende questo nome proprio dal fatto che tutti gli elementi della popolazione vengono presi in considerazione ed hanno tutti uguale probabilità di essere selezionati; ognuno di essi cioè può "casualmente" costituire una delle unità del campione.

Il campionamento avviene estraendo unità per unità gli **N** elementi della popolazione fino ad ottenere le **n** unità del campione. Necessita disporre di un elenco, numerato da 1 a N, degli elementi della popolazione tra i quali vengono presi quelli i cui numeri corrispondono ad una successione di **n** numeri casuali compresi tra 1 ed N (estratta mediante l'uso delle tavole dei numeri casuali o con altri procedimenti che assicurino l'imprevedibilità di un risultato fra i tanti possibili e garantiscano uguale probabilità per tutte le unità che possono essere scelte); le n unità così "casualmente" identificate costituiscono il campione. Se all'estrazione di ogni numero, l'unità corrispondente viene escluso dalla estrazione successiva, cioè la probabilità di entrare a far parte del campione per ogni elemento (*probabilità di inclusione del primo ordine*) è pari a n/N, il metodo viene chiamato ***campionamento senza ripetizione (o esaustivo)***. E' possibile anche estrarre dei ***campioni con ripetizione (bernoulliani)***, quando ad ogni estrazione vengono riconsiderati tutti gli elementi della popolazione talché la probabilità di inclusione è $1-\left(1-1/N\right)^n$ e ciascun elemento pertanto può entrare a far parte delle unità campionate una, due o più volte.

Vantaggi:

1. E' il tipo di campionamento più semplice che consente di fare riferimento ai modelli più elementari di calcolo delle probabilità;
2. Richiede una minima conoscenza a priori delle caratteristiche della popolazione;
3. Garantisce una scelta obiettiva delle unità da rilevare e tale da escludere qualunque distorsione nei risultati;
4. Le stime dei parametri della popolazione e dei loro errori medi sono molto facili;
5. E' particolarmente conveniente quando la popolazione non è molto grande.

Svantaggi:

1. Comporta spesso costi di rilevazione più elevati rispetto ad altri procedimenti di scelta delle unità campionarie;
2. A parità di dimensione, ossia del numero di osservazioni, può condurre a stime dei parametri della popolazione meno precise di quelle ottenute con altri tipi di campionamento;
3. Nella pratica non viene molto impiegato, specie nelle indagini di grandi dimensioni;
4. In generale non vengono utilizzate tutte le informazioni che si posseggono sulla popolazione.

Campionamento casuale stratificato

Il campionamento stratificato differisce fondamentalmente dal precedente. Un tale metodo può essere usato, sotto certe circostanze, per migliorare l'efficienza del piano di campionamento. In altre parole, il campione stratificato permette di raggiungere una più grande accuratezza ad uno stesso costo o, analogamente, con un minor costo la medesima accuratezza.

Quale prima fase gli N elementi della popolazione vengono suddivisi in **k** gruppi o ***strati*** il più possibile omogenei fra di loro rispetto ad un opportuno criterio, di numerosità N_1, N_2,...,N_k, tale che $N_1+N_2+...+N_k = N$. Ciò significa che ogni strato è formato in modo tale che non ci sia sovrapposizione, cioè ogni elemento compare in un solo strato. Quindi il campione viene formato estraendo da ogni strato n_1, n_2,...,n_k unità in modo indipendente.

La scelta delle unità all'interno di ogni strato può essere effettuata o con un campionamento casuale semplice o con un campionamento sistematico (di seguito descritto).

I motivi che rendono frequente il ricorso a questa tecnica campionaria possono essere sintetizzati nei seguenti punti.

1. Se si desidera un certo grado di precisione per certe suddivisioni della popolazione è consigliabile trattare ogni "suddivisione" come una popolazione a se stante.

2. L'analisi campionaria può riguardare diversi tipi di popolazioni che potrebbero essere difficilmente accorpati in una unica popolazione ed inoltre lo svolgimento dell'indagine può essere effettuato in tempi diversi da differenti operatori.

3. E' possibile dividere una popolazione altamente eterogenea in sottopopolazioni ognuna delle quali al suo interno sia omogenea. Ciò, oltre a consentire analisi di popolazioni i cui elementi sono molto diversi tra loro, permette anche di ridurre l'ampiezza del campione nei singoli strati composti da elementi omogenei tra loro, così che presentano una piccola variabilità interna.

Nel considerare i singoli strati la frazione campionaria relativa a ciascuno può essere uguale, nel qualcaso si parla di ***campionamento stratificato proporzionale***, oppure diversa allora si ha il ***campionamento stratificato "ottimale"***.

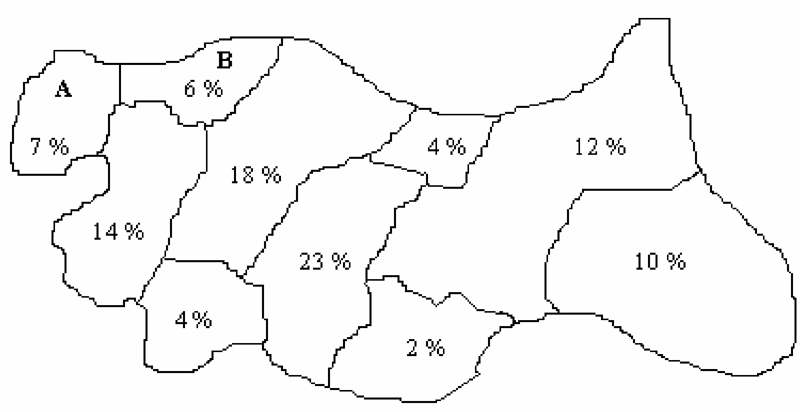
Il ***campionamento stratificato proporzionale*** viene usato per ottenere un campione più rappresentativo di quanto non si otterrebbe con un campione casuale semplice o sistematico. Dopo aver effettuato la stratificazione degli elementi della popolazione secondo i caratteri che maggiormente influenzano il fenomeno che si intende studiare, si estrae da ogni strato una certa quantità di unità in proporzione alla numerosità dello strato, tale che: $n_1/N_1 = n_2/N_2 = ... = n_k/N_k = n/N$ e cioè ogni strato contribuisce alla formazione del campione totale nella stessa misura in cui ogni sotto popolazione contribuisce a formare l'intera popolazione.

Per ottenere un campione stratificato proporzionale occorre, naturalmente, conoscere la parte di popolazione che afferisce ad ogni strato ed è possibile effettuare la stratificazione solo per quelle variabili di cui si hanno informazioni al momento del campionamento.

Ad esempio, per accertare l'influenza dell'età sull'incidenza di una certa patologia ed evitare che, mediante un estrazione casuale semplice, il campione risulti prevalentemente rappresentato o da soggetti giovani o anziani si procede alla stratificazione per età (secondo k classi) della popolazione, poi si calcola una frazione costante per ogni strato.

Esempio. Nell'ambito di una campagna per la prevenzione delle malattie cardiovascolari è interessante conoscere il livello della colesterolemia in una popolazione al fine di indirizzare verso un più corretto comportamento alimentare. Si concorda di eseguire l'indagine solo sui soggetti di sesso maschile e di età compresa tra i 45 e i 65 anni, cioè la popolazione più a rischio.

La figura mostra la composizione dei comuni del territorio scelto per la ricerca con relativa percentuale di residenti maschi nati tra il 1937 e il 1957.



Composizione dei comuni ai fini della ricerca

Si decide di campionare 4482 unità corrispondenti all' 1% della popolazione composta da 448208 soggetti. Quindi, dal Comune A campioniamo 314 sogg., pari al 7% di 4482; dal Comune B 269 soggetti (6% di 4482) ecc.

Il ***campionamento stratificato ottimale*** tende a migliorare l'efficienza del piano di campionamento ricorrendo a frazioni di campionamento diverse per ogni singolo strato. In alcune situazioni un tale metodo offre maggiore precisione rispetto al campione proporzionale. Così quando la variabilità interna degli strati è elevata è più conveniente ricorrere ad un campione ottimale. Allora la frazione di campionamento varierà da strato a strato, sia in funzione della caratteristica che si intende studiare che della variabilità dello strato, applicando quindi il criterio della "ottima ripartizione".

Esempio Alcuni anni fa, in seguito alla recrudescenza dell'infezione tubercolare registrata, una USL volle studiare la situazione sul proprio territorio attraverso un indagine campionaria. Si intendeva stimare la prevalenza della reattività alla tubercolina con particolare riguardo ai giovani, indicatori più fedeli di una recrudescenza infettiva (si considerava, infatti, che la popolazione più anziana era vissuta nel periodo in cui la TBC era fortemente endemica e molto probabilmente era portatrice di memoria di remoti contatti). La tabella seguente mostra la composizione per classi di età della popolazione in studio.

Per le suddette ragioni, si procedette ad un campionamento stratificato ottimale, con frazione di campionamento maggiore per le classi di età più basse e precisamente: 5% per la prima classe, 4% per la seconda, 3% per la terza, 2% per la quarta e 1% per la quinta.

Piano di campionamento

Classi età	N Popolazione	n Campione
< 15	27315	5 % = 1366
15 – 24	18751	4 % = 750
25 – 39	24890	3 % = 747
40 – 64	36847	2 % = 737
> 65	16298	1 % = 163
Tot.	124101	3763

Vantaggi:

1. La stratificazione consente, in generale, di aumentare la precisione delle stime senza accrescere la dimensione totale del campione;
2. La stratificazione è molto conveniente quando la distribuzione statistica della variabile da rilevare è fortemente asimmetrica.

Svantaggi:

Se non si hanno sufficienti informazioni a priori, la costruzione degli strati può risultare alquanto costosa; inoltre se la stratificazione è errata si possono ottenere risultati fuorvianti.

Campionamento casuale a grappoli

Mentre nel campionamento stratificato si suddivide la popolazione in sottogruppi detti strati, a volte può essere utile dividerla in un gran numero di sottoinsiemi detti grappoli (clusters) ed effettuare il campionamento casuale tra i grappoli. Ad esempio si selezionano 20 quartieri di una città avente 100 quartieri ed si inseriscono nel campione tutti gli abitanti dei 20 quartieri.

Il metodo non prevede quindi il campionamento diretto degli elementi, ma vengono campionati grappoli di elementi.

Viene spesso fatto ricorso ad un tale campionamento per ridurre il costo della raccolta dei dati. Si evita infatti il costo relativo alla elencazione complessiva di tutti gli elementi della popolazione ed anche le spese di viaggio per gli intervistatori che invece vengono concentrati su un grappolo.

Nel più semplice campionamento a grappoli si può usare la scelta casuale per selezionare i grappoli formati e quindi indagare su tutte le unità contenute in quelli campionati. Un tale metodo viene usualmente definito campionamento ad uno stadio, poiché il processo campionario viene effettuato un'unica volta. Per contrapposto si definisce campionamento a più stadi quello in cui la scelta delle unità campionarie è molto più complessa ed il procedimento campionario interviene più di una volta nella scelta delle unità. La frazione di campionamento deve, in questo caso, essere calcolata in modo tale che gli elementi della popolazione abbiano la stessa probabilità di essere campionati.

In generale il campionamento a grappoli al vantaggio di una riduzione di costi non associa un'alta precisione, in quanto per ridurre troppo i costi si rischia di non ottenere un campione rappresentativo di tutta la popolazione, ma solo di una parte, quella relativa ai grappoli selezionati.

Per ovviare in parte a ciò si deve cercare di riferirsi a grappoli di numerosità piccola, rispetto alla numerosità totale del campione, per poter ricorrere ad un numero di grappoli piuttosto grande. Ad esempio in una indagine sulla educazione sanitaria nelle scuole di una provincia, non disponendo di un elenco generale di tutti gli alunni delle scuole, si possono estrarre alcune classi (grappoli) di cui si dispone dell'elenco completo e di esse si intervistano poi tutti gli alunni.

Esempio. Si vuole studiare l'incidenza di reazioni avverse da vaccinazione antimorbillosa, in particolare l'iperpiressia post vaccinica.

Si stabilisce di svolgere l'indagine in una grande città. Il campione selezionato sarà successivamente parte integrante di una coorte da seguire per eventuale controllo di efficacia del vaccino. La tabella che segue è fornita dall'USL locale e mostra tutti i medici pediatrici convenzionati con il relativo numero di assistiti. Non essendoci alcun motivo per ipotizzare una differenza di reattività tra bambini iscritti a pediatri diversi, si decide di effettuare un campionamento "cluster", inserendo cioè nel campione tutti i bambini o ragazzi seguiti da quei pediatri scelti con un campionamento randomizzato semplice.

Numero bambini assistiti per pediatra

Ped.	Ass.	Ped.	Ass.	Ped.	Ass.	Ped.	Ass.	Ped.	Ass.
1	240	7	200	13	344	**19**	**310**	25	464
2	441	8	374	14	480	20	290	**26**	**440**
3	**420**	**9**	**333**	15	424	21	426	27	287
4	312	10	420	**16**	**380**	22	413	28	292
5	346	**11**	**412**	17	490	23	398		
6	480	12	307	18	282	**24**	**380**		

Stimando in 1000 la numerosità campionaria richiesta, e sapendo che circa il 50% dei soggetti in età pediatrica viene sottoposta a vaccinazione, possiamo campionare 7 pediatra (vedi tabella), es. 3, 9, 11, 16, 19, 24 e 26, per un totale di 2675 bambini, tra i quali identificare quelli vaccinati.

Campionamento sistematico

A volte può risultare molto arduo dover numerare, come richiesto nel campionamento casuale semplice, tutti gli elementi della popolazione, specie se questa è molto numerosa. Qualora si disponga di un elenco delle N unità di una popolazione numerate da 1 a N secondo un ordine specifico, per individuare le n unità del campione si sceglie una unità ogni **k**, essendo **k** *il passo di campionamento* N/n (se N è multiplo di n) . Si prendono quindi le unità secondo i termini di una progressione aritmetica di ragione **k** a partire da una qualunque **i** di esse scelta a caso fra le prime **k**. Il campione, quindi, risulta determinato dall'insieme delle unità individuate dall'indice: **i+(j-1)k** (per j=1,2,…n). Questo sistema appare molto conveniente quando la popolazione di riferimento è riportata in elenchi, come ad esempio gli elenchi relativi ai degenti dei diversi reparti di un ospedale, l'elenco del personale medico, paramedico, amministrativo ecc. di una U.S.L.. Per cui anziché comporre un unica lista da cui estrarre le unità da campionare si prelevano da ogni singolo registro le unità secondo un campionamento sistematico.

Esempio. Se si debbono estrarre 50 gestanti da 500 che hanno partorito durante l'ultimo anno presso una divisione di Ostetricia e Ginecologia di un ospedale, il rapporto **k**= 500/50=10 denota che si deve scegliere, dall'elenco presso il Reparto, una donna ogni dieci. La prima donna del campione si individua scegliendo casualmente (ad esempio tramite le tavole dei numeri casuali) un numero compreso tra le prime dieci donne dell'elenco e le altre si sceglieranno, partendo da questa, intervallate di dieci unità.

Vantaggi:

1. Le operazioni di estrazione, rilevazione e controllo del campione sono, in generale, più facili e rapide rispetto a quelle dei campionamenti casuale semplice e stratificato;

2. La stime sono più precise di quelle di un analogo campione casuale se la variabilità entro il campione è maggiore di quella dell'intera popolazione.

Svantaggi:

1. Se la popolazione ordinata varia con tendenza lineare, il campionamento sistematico è più efficiente del campionamento casuale ma meno di quello stratificato;
2. Se la popolazione ordinata ha un andamento periodico, l'efficienza di un campione sistematico dipende dal valore k e dalla relazione fra k ed il periodo dell'oscillazione;
3. Se N non è un multiplo intero di k, le stime, a rigore, si considerano corrette soltanto per campioni di dimensione maggiore di 50.

Campionamento casuale a più stadi

E' una tecnica di campionamento che risulta molto vantaggiosa quando la popolazione da studiare è molto numerosa e gli elementi possono essere raggruppati in diversi sottoinsiemi.

Dopo aver suddiviso la popolazione di partenza in successive sottoclassi o stadi (es. province, comuni, scuole e così via), si estrae un campione di unità di primo stadio (province) e nell'ambito delle unità ottenute si procede alla scelta dei campione di secondo stadio (comuni) e cosi via. Il campione è costituito dalle unità estratte dall'ultimo stadio.

Questa tecnica viene spesso utilizzata, ad esempio, nei sondaggi di opinione, allorquando si procede al successivo campionamento delle città, quindi dei rioni ed in ultimo dei soggetti da intervistare.

Il campionamento viene definito a **n** stadi in riferimento al numero dei campionamenti successivi (nel caso città-rioni-persone si parla di campionamento a 3 stadi).

Esempio L'Assessorato alla Sanità di una Regione è interessato ad indagare sui comportamenti sessuali di ragazzi e ragazze iscritti alle scuole medie superiori. Si decide di condurre lo studio attraverso la compilazione di un questionario e un colloquio con uno psicologo.

Nella seguente tabella è riportata la distribuzione per tipo di istituto delle 133 scuole statali e private della Regione

Distribuzione scuole di una Regione

SCUOLE	N
Liceo Classico (L.C.)	9
Liceo Scientifico (L.S.)	18
Ist. Tecnico Commerciale (RAG.)	21
Ist. Tecnico per Geometri (GEO.)	16
Ist. Tecnico Industriale (I.T.I.)	21
Scuola d'Arte	11
Ist. Professionale (PROF.)	17
Altre	20
Totale	133

Il ricercatore decide che un campionamento a più stadi si adatta a questo tipo di studio.

I Fase: si procede al campionamento delle 133 scuole attraverso l'uso di numeri casuali o con estrazione da bussolotti; vengono campionate 10 scuole: ad es. 1 L.C., 2 L.S., 2 RAG., 1 GEO., 3 I..T.I. e 1 PROF.

II Fase: si procede al campionamento dei ragazzi all'interno delle 10 scuole scelte, e poi, per ognuna di queste, ad un campionamento sistematico fino al raggiungimento del numero prestabilito.

CAMPIONI NON PROBABILISTICI

In molte indagini epidemiologiche risulta impossibile o comunque impraticabile ricorrere al campionamento casuale in quanto non risultano valide alcune implicazioni procedurali statistiche. Di fatto si è costretti a campionare una parte della popolazione che risulta quella realmente accessibile. Negli esperimenti di laboratorio su animali, ad esempio, il ricercatore frequentemente è costretto ad "usare" quegli animali di cui dispone; altrimenti, se dovesse selezionare casualmente le cavie, difficilmente riuscirebbe a fare della ricerca. Naturalmente anche in queste situazioni è lecito supporre che il campione sia equivalente ad un campione casuale, non essendoci grossi motivi per ritenere che gli animali di cui dispone non siano rappresentativi.

In molti progetti di ricerca in campo sanitario, si è costretti a ricorrere a dei campioni ad hoc, i cui elementi sono dei "volontari" o comunque soggetti "disponibili" a sottoporsi alla ricerca. Ciò accade quando, ad esempio, si indaga su situazioni delicate, oppure quando questionari vengono inviati per posta e la percentuale dei non rispondenti potrebbe comunque alterare il campione (in quanto coloro che non accettano di rispondere hanno delle caratteristiche diverse dai rispondenti).

In altre situazioni è possibile introdurre la casualità (randomizzazione) nell'esperimento anziché nei soggetti; nel confronto di due trattamenti, ad esempio, i pazienti selezionati vengono casualmente attribuiti ad uno o all'altro trattamento; in tal caso le implicazioni statistiche (inferenza), saranno applicate ai trattamenti anziché ai soggetti.

Sotto le giuste condizioni, quindi, i campioni non probabilistici possono dare utili risultati. Essi però non sono trattabili con la teoria campionaria perché non si basano su alcun principio di casualità e pertanto la loro validità è strettamente legata alla situazione cui si riferiscono e non si ha nessuna garanzia della loro validità in circostanze diverse.

Questi metodi "non corretti" sono meno impegnativi da applicare e, sebbene l'accuratezza è discutibile, trovano frequentemente pratica applicazione. Sono i metodi molto usati da organizzazioni per le ricerche di mercato, per sondare gli orientamenti elettorali nella popolazione ecc.

Un esempio in ambito sanitario può essere il seguente. Si vuole valutare l'efficacia di un nuovo vaccino contro l'AIDS. Tutti i soggetti a rischio che volontariamente si presentano ai centri sieroprofilattici e che risultano sieronegativi vengono sottoposti alla vaccinazione. Avendo stimato in 1000 la numerosità campionaria richiesta, si procede al reclutamento di tutti i soggetti in campo nazionale fino al raggiungimento del numero previsto.

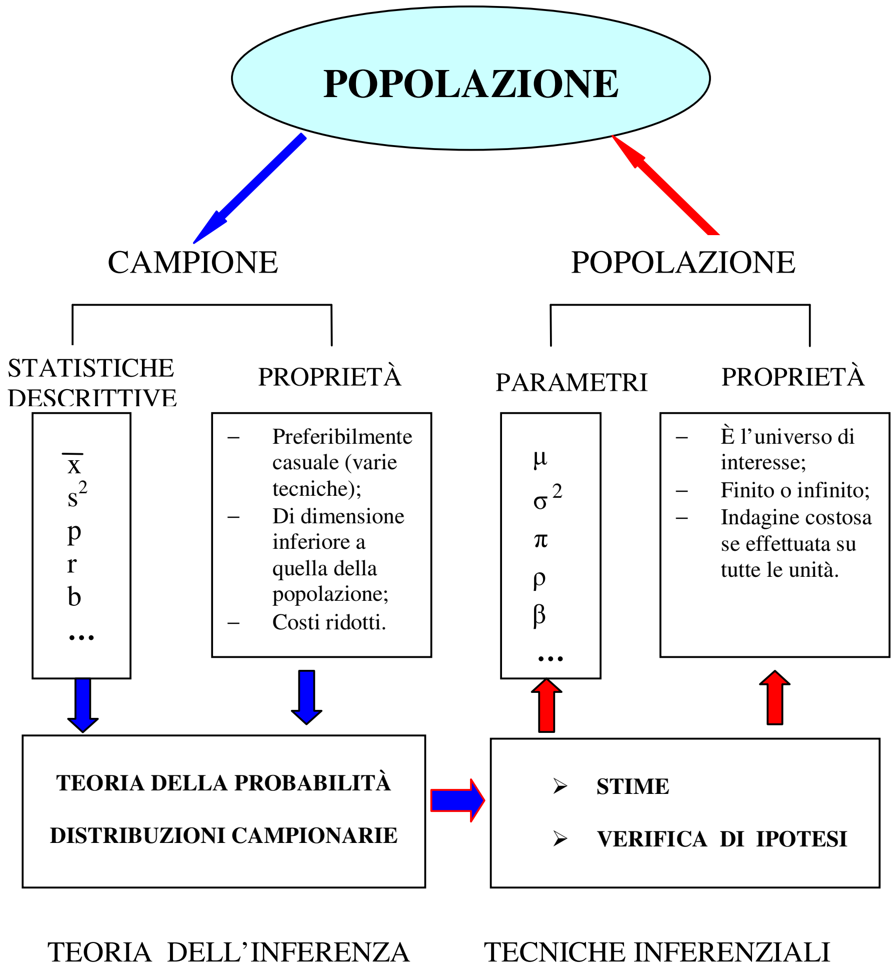
STATISTICHE CAMPIONARIE E DISTRIBUZIONI

Inferenza statistica

L'inferenza statistica è un insieme di metodi, messi a punto con l'ausilio del *Calcolo delle Probabilità*, attraverso le quali si possono trarre alcune conclusioni su una data popolazione in base ai risultati ottenuti sperimentalmente su un campione. Generalmente le conclusioni a cui si giunge riguardano la *stima* di alcuni parametri caratteristici della popolazione e la *verifica di ipotesi*.

In primo luogo, è necessario analizzare alcune funzioni dei dati campionari, le ***statistiche campionarie***, e le rispettive ***distribuzioni campionarie***. Va precisato fin d'ora che le funzioni dei dati campionari vengono generalmente indicate con le lettere dell'alfabeto latino, mentre i parametri o altre caratteristiche della popolazione si indicano con le lettere dell'alfabeto greco. Così μ, σ^2, π, ρ, β stanno ad indicare, rispettivamente, i parametri media, varianza, proporzione, coefficiente di correlazione, coefficiente di regressione di una popolazione, mentre $\bar{x}$, s^2, p, r, b denotano le *stime* degli stessi parametri ottenute sul campione (valori ricavati dal calcolo di *statistiche* campionarie). Va osservato che i valori caratteristici della popolazione, generalmente sconosciuti, sono fissi, mentre le statistiche variano da campione a campione.

Di seguito è riportato un diagramma che schematizza le metodiche inferenziali.



<u>Distribuzioni Campionarie</u>

Scopo principale della statistica inferenziale è quello di stimare i parametri di una popolazione o di sottoporre ad esame delle ipotesi di una popolazione, sulla base dell'osservazione di un numero ridotto di elementi appartenenti ad un campione.

A partire dai dati campionari si può costruire una quantità che permetta di ottenere indicazioni sul parametro che si intende studiare sulla popolazione. Tale quantità è detta ***statistica***. Si tratta di una variabile casuale (numeri reali con valori di probabilità "associati"), funzione delle **n** variabili (X_1, X_2, …,X_n) rilevate sul campione e relative ad un carattere statistico X oggetto di studio.

Nel caso lo studio sia relativo ad un carattere statistico quantitativo e il parametro di interesse sia la media della popolazione, si considera la *statistica* ***media campionaria:***

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \,.$$

Accanto alla media campionaria si può considerare la *statistica* ***varianza campionaria:***

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \ .$$

Se l'oggetto di studio riguarda la proporzione di un dato evento ("successo") nella popolazione di riferimento, si può considerare la *statistica* ***proporzione campionaria P*** dei successi rilevati su campioni estratti dalla popolazione.

Distribuzione della media

Si consideri su una popolazione il carattere quantitativo X avente media μ e deviazione standard σ e sia $\overline{x}$ la media aritmetica di un campione casuale di dimensione ***n*** estratto dalla popolazione, risulta:

- la distribuzione campionaria della media $\overline{X}$ ha media uguale alla media della popolazione di riferimento:
$$E(\overline{X}) = \mu \ ;$$
- la distribuzione campionaria di $\overline{X}$ ha deviazione standard uguale alla deviazione standard della popolazione diviso la radice quadrata di ***n***;
$$DS(\overline{X}) = \frac{\sigma}{\sqrt{n}} \ ;$$
- se la dimensione campionaria è "abbastanza" grande (in genere $n \geq 30$), la distribuzione campionaria di $\overline{X}$ è *"approssimativamente" normale*, indipendentemente dalla distribuzione della variabile X nella popolazione oggetto di studio (Teorema Centrale del Limite).

Esempio1. Popolazione con X = $\{1, 2, 3, 4, 5, ..., 86, 87, 88, 89, 90\}$ distribuita uniformemente, si estraggono campioni casuali di numerosità **n = 3** e, successivamente, di numerosità **n = 6.** La media aritmetica nella popolazione è $\mu =$ 45.5 e la deviazione standard $\sigma = 26.1$.

Campioni con n = 3

Campione	Unità campionarie			media
1°	70	83	20	57.7
2°	44	11	70	41.7
3°	74	45	38	52.3
4°	78	1	30	36.3
5°	50	12	2	21.3
Media della medie campionarie				41.9
Dev. Stand. medie campionarie				14.2

Campioni con n = 6

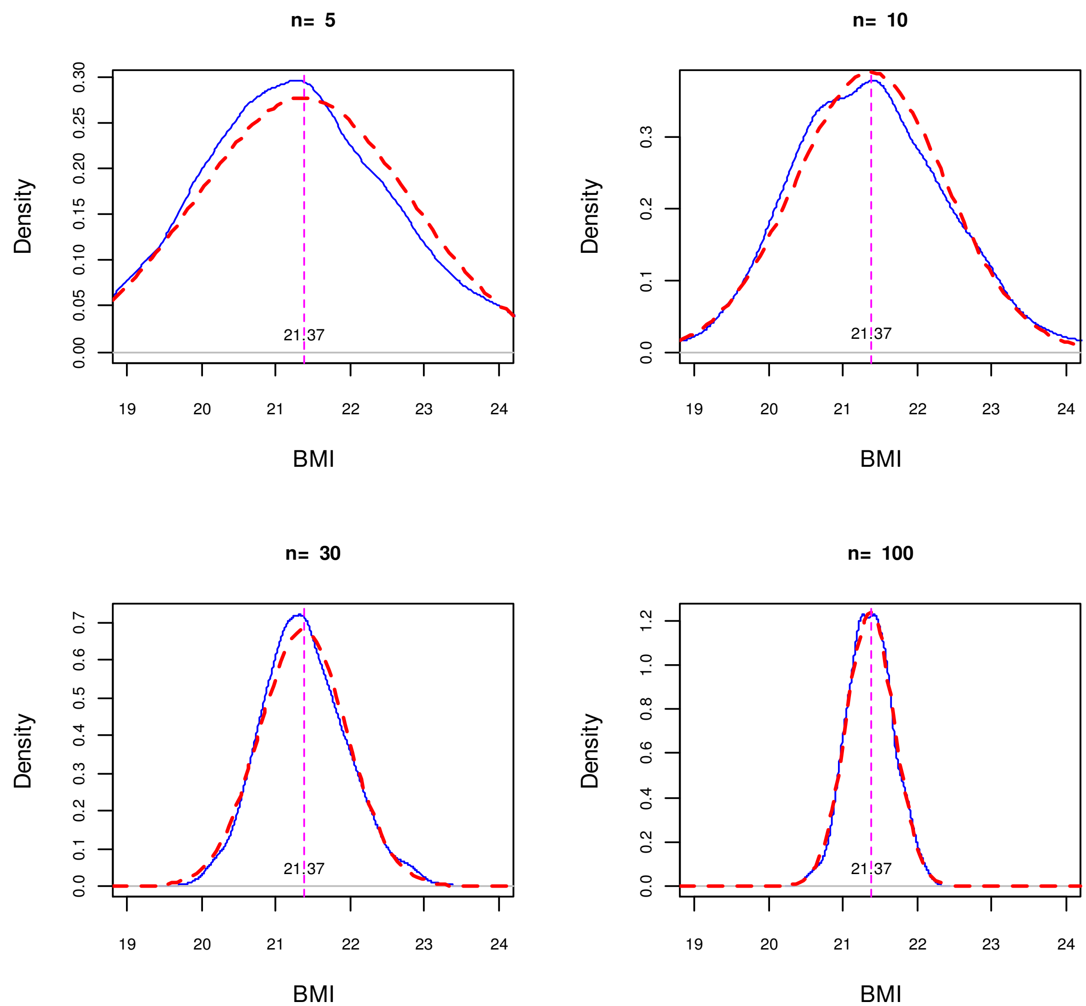
Campione	Unità campionarie						media
1°	16	2	48	65	41	33	34.2
2°	20	48	36	13	89	2	34.7
3°	40	52	65	72	17	82	54.7
4°	29	52	19	2	9	76	31.2
5°	78	16	9	47	78	66	49.0
Media della medie campionarie							40.7
Dev. Stand. medie campionarie							10.4

La media delle medie nei due campioni (41.9 per n=3 e 40.7 per n=6) è “vicina” alla media della popolazione $\mu =$ 45.5.

La variabilità delle medie campionarie è più piccola della variabilità nella popolazione e il suo valore atteso è dato da $\sigma/\sqrt{n}$, nel caso dei campioni considerati:

campioni	Dev. Stand. medie	Valore atteso $\sigma/\sqrt{n}$
n = 3	14.2	$26.1/\sqrt{3} = 15.1$
n = 6	10.4	$26.1/\sqrt{6} = 10.6$

Esempio2. Si è presa in esame la distribuzione del BMI in una popolazione di 1039 ragazzi, con μ=21.37 kg/cm^2 e deviazione standard $\sigma =$ 3.22 kg/cm^2. Le figure seguenti riportano le distribuzioni delle medie campionarie per campioni casuali di numerosità n = 5, n = 10, n = 30 e n = 100 (linea continua in blue) e, sovrapposti i grafici delle distribuzioni normali con media μ e deviazione standard $\sigma/\sqrt{n}$ (linea tratteggiata in rosso).



Si osserva che, al crescere di ***n***, la media aritmetica dei dati campionari tende a distribuirsi secondo una distribuzione normale, anche se la variabile considerata non segue tale distribuzione.

Alcuni risultati importanti si ottengono se il carattere X preso in esame si distribuisce secondo una ***distribuzione normale***. Si supponga che il campionamento consista in **n** osservazioni indipendenti e che la media μ sia incognita.

Se la **deviazione standard σ è nota**, la variabile

$$Z = \frac{\overline{X} - \mu}{DS(\overline{X})} = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$$

si distribuisce secondo la ***normale standard.***

Se, oltre alla media, anche la deviazione standard σ della popolazione **è incognita**, la deviazione standard s del campione rappresenta la stima più logica ed attendibile della deviazione standard della popolazione.

Con σ ignota, la distribuzione di probabilità non è fornita dalla distribuzione normale, bensì dalla

<u>DISTRIBUZIONE T DI STUDENT</u>.

In tale contesto si considera la variabile casuale:

$$t = \frac{\overline{X} - \mu}{S / \sqrt{n}}$$

e la distribuzione a cui essa obbedisce prende il nome di ***t di Student*** con $\nu = (n\text{-}1)$ ***gradi di libertà*** [1] (Figura 1). Essa ha valor medio 0 e *varianza* $\frac{\nu}{\nu - 2}$ per $\nu \geq 3$.

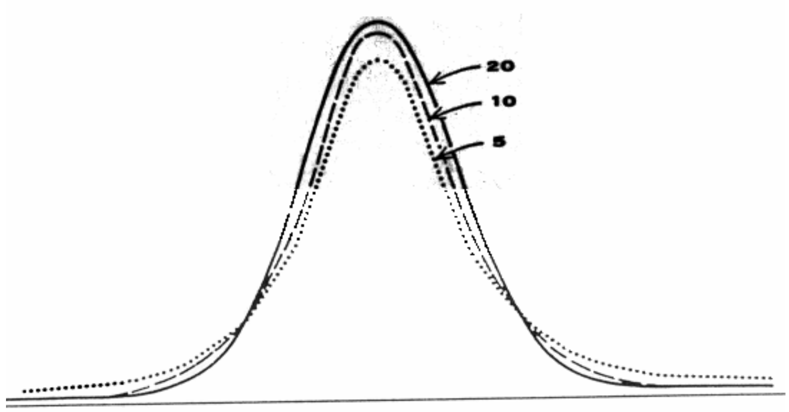
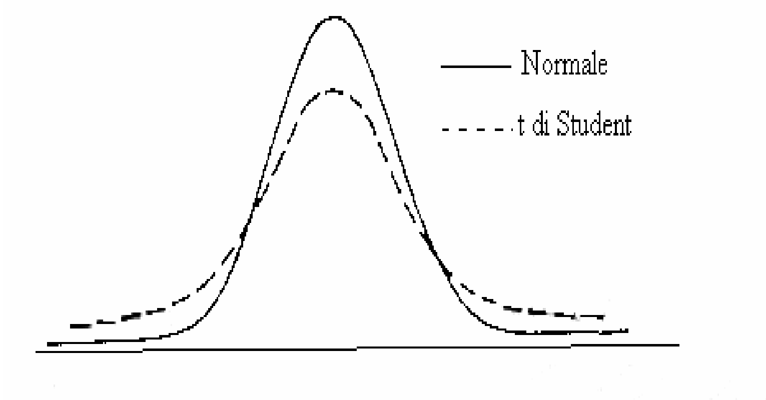


Figura 1 – Distribuzione t di Student al variare dei gradi di libertà

All’aumentare dei gradi di libertà, la distribuzione **t** converge alla distribuzione normale standard e viene generalmente confusa con essa se $\nu > 30$.

[1] Con il ***termine di libertà*** si intende “libertà di variare”. Per esempio se si devono scegliere 2 numeri si ha, per ognuno di essi, completa libertà di scelta. Ma se si impone che la somma dei 2 numeri sia 8, una volta scelto il primo, il secondo è determinato dalla scelta effettuata: si ha quindi un solo grado di libertà. Se dobbiamo scegliere 5 numeri e non vi è alcuna restrizione, i 5 numeri sono ognuno indipendente dall'altro. Basta tuttavia dare una semplice restrizione, ad esempio che la loro somma debba essere 40, che già la scelta viene regolata. L'ultimo numero, infatti, scelti i primi 4 è già fissato. In questo caso si hanno quattro gradi di libertà. In molti problemi è necessario determinare il numero delle grandezze che sono libere di variare: prima si fissano il numero delle variabili e, successivamente, si segnano le restrizioni.



Fig, 2 – Confronto tra la distribuzione t di Student e la Normale Standad

Le principale differenze tra la distribuzione normale e la distribuzione t sono:
- la distribuzione normale considera la variazione di campionamento solo della media;
- la distribuzione t considera anche la variazione di campionamento della deviazione standard.

Le caratteristiche della distribuzione t sono:
- distribuzione dei dati normale;
- osservazioni raccolte in modo indipendente;
- area unitaria al di sotto della curva di densità e di forma simmetrica;
- si ottiene una famiglia di distribuzioni (una distribuzione per ogni grado di libertà) a differenza di quanto avviene per la gaussiana;
- sempre più "dispersa" al diminuire dei gradi di libertà;
- valida anche per distribuzioni di dati con marcate deviazione dalla normalità (*robusta)***.**

Nel caso che l'ipotesi di normalità per la popolazione non possa essere ritenuta valida, bisogna ancora distinguere il caso in cui la deviazione standard della popolazione sia nota da quello in cui essa è incognita.

Se σ **è nota,** la variabile casuale:

$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$$

si distribuisce, per *" n grande"* (vedi anche sopra) secondo la normale standard. Se la distribuzione da cui si effettua il campionamento è simmetrica e unimodale o è una distribuzione uniforme su un intervallo finito, la Z può pensarsi distribuita normalmente per valori di n già relativamente piccolo (es. n = 10); in generale per n≥30 si accetta nella pratica di considerare Z distribuita normalmente.

Se σ **non è nota**, si considera ancora che la variabile casuale:

$$Z = \frac{\overline{X} - \mu}{s / \sqrt{n}}$$

si distribuisce secondo una ***t di Student*** con $\nu = (n\text{-}1)$ ***gradi di libertà***.

Corre l'obbligo di puntualizzare, infine, che quanto detto è valido per osservazioni campionarie indipendenti, in sostanza per *popolazioni* (potenzialmente) *infinite.* Per *popolazioni finite,* l'ipotesi di indipendenza presuppone che dopo ogni osservazione la situazione sia riportata allo stato iniziale, in tal caso una stessa unità statistica potrebbe essere presa in esame più volte. Ciò ovviamente non corrisponde alla pratica reale, dove il campionamento è effettuato "senza reimbussolamento". La differenza non è apprezzabile se la dimensione N della popolazione è grande. In generale il problema viene risolto "correggendo" la deviazione standard della media campionaria $\overline{X}$ con il fattore $k = \sqrt{(N - n)/(N - 1)}$;

così, ad esempio, la standardizzazione della media campionaria diventa $Z = \dfrac{\overline{X} - \mu}{k \cdot \sigma / \sqrt{n}}$. Il fattore correttivo k tende a 1 se N cresce ed è "praticamente" 1 se N è molto più grande di n.

In Appendice è riportata la Tavola (Tavola 2) della ***distribuzione t di Student***. In essa, in corrispondenza di un fissato grado di libertà ν, è riportato il valore $t_{\alpha,\nu}$ della t per il quale l'area indicata nella figura (misura la probabilità che t sia maggiore o uguale $t_{\alpha,\nu}$) è pari ad α.

Distribuzione di una proporzione

Se l'oggetto di studio riguarda la proporzione di un dato evento ("successo") in una popolazione, il carattere X da prendere in esame è una variabile *bernoulliana* con media π e deviazione standard $\sqrt{\pi(1-\pi)}$.

Sia ***p*** la proporzione di successi in un campione di dimensione ***n***, risultano valide le seguenti:

- la distribuzione campionarie della proporzione **P** ha media uguale alla media della popolazione da cui è stato estratto il campione:
$$E(P) = \pi;$$
- la distribuzione campionarie della proporzione ha deviazione standard:
$$DS(P) = \sqrt{\pi(1-\pi)};$$
- se la dimensione campionaria è "abbastanza" grande ($n \geq 30$), la distribuzione campionaria di **P** è *"approssimativamente" normale*.

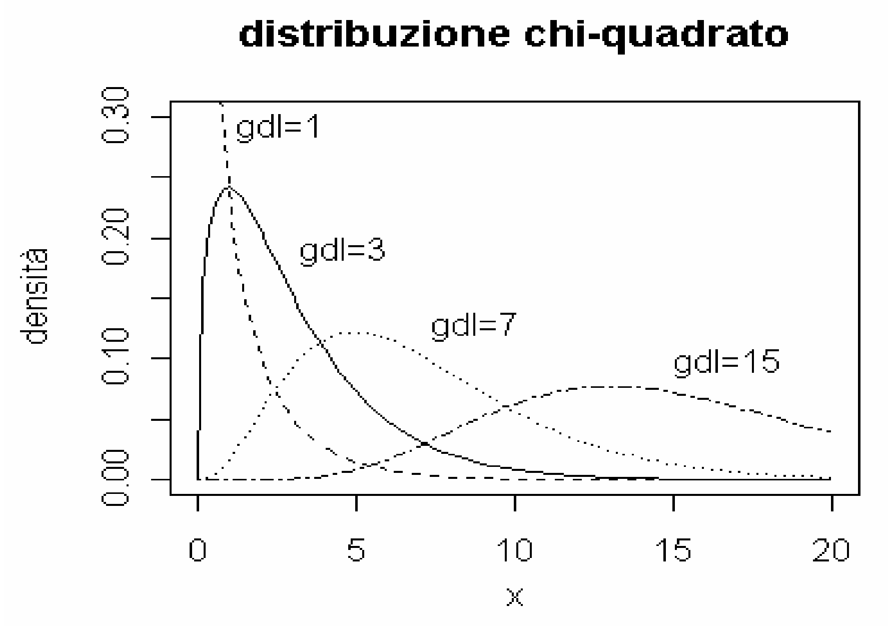
Distribuzione Chi-quadrato

Considerata una variabile X, relativa ad una caratteristica di una data popolazione, che si distribuisce *normalmente* nella popolazione con media μ e deviazione standard σ e dato un campione casuale $(X_1, X_2, \ldots, X_n)$ avente varianza S^2, la ***statistica***:

$$\chi^2 = \frac{S^2(n-1)}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2$$

si distribuisce secondo una ***DISTRIBUZIONE CHI-QUADRATO*** con **n-1** ***gradi di libertà (gdl).***

Il grafico che segue riporta le curve del chi-quadrato per alcuni valori dei gradi di libertà.



Si può provare che se $f_1, f_2, \ldots, f_k$ è la distribuzione di frequenze osservate di un carattere statistico suddiviso in k categorie e $f_1^*, f_2^*, \ldots, f_k^*$ è la corrispondente distribuzione di frequenze attese o teoriche, la variabile

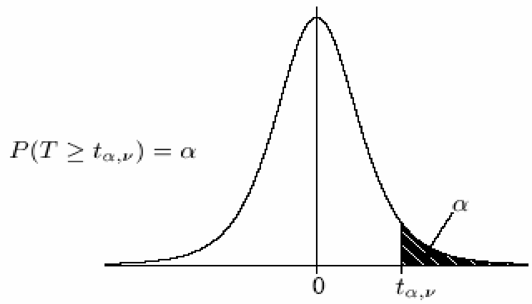
$$\chi^2 = \sum_i \frac{(f_i - f_i^*)^2}{f_i^*}$$

per valori della dimensione campionaria n sufficientemente grandi, si distribuisce secondo una distribuzione chi-quadrato con **k-1 gradi di libertà**.

In Appendice è riportata la Tavola (Tavola 3) della ***distribuzione chi-quadrato***. Per essa va data una interpretazione analoga a quella della Tavola 2.

APPENDICE

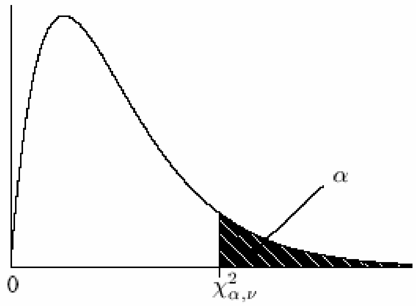
Tavola 2: Valori critici della Distribuzione t



ν	0.20	0.10	0.05	0.025	α 0.01	0.005	0.001	0.0005	0.0001
1	1.3764	3.0777	6.3137	12.7062	31.8210	63.6559	318.2888	636.5776	3185.2722
2	1.0607	1.8856	2.9200	4.3027	6.9645	9.9250	22.3285	31.5998	70.7060
3	0.9785	1.6377	2.3534	3.1824	4.5407	5.8408	10.2143	12.9244	22.2027
4	0.9410	1.5332	2.1318	2.7765	3.7469	4.6041	7.1729	8.6101	13.0385
5	0.9195	1.4759	2.0150	2.5706	3.3649	4.0321	5.8935	6.8685	9.6764
6	0.9057	1.4398	1.9432	2.4469	3.1427	3.7074	5.2075	5.9587	8.0233
7	0.8960	1.4149	1.8946	2.3646	2.9979	3.4995	4.7853	5.4081	7.0641
8	0.8889	1.3968	1.8595	2.3060	2.8965	3.3554	4.5008	5.0414	6.4424
9	0.8834	1.3830	1.8331	2.2622	2.8214	3.2498	4.2969	4.7809	6.0094
10	0.8791	1.3722	1.8125	2.2281	2.7638	3.1693	4.1437	4.5868	5.6939
11	0.8755	1.3634	1.7959	2.2010	2.7181	3.1058	4.0248	4.4369	5.4529
12	0.8726	1.3562	1.7823	2.1788	2.6810	3.0545	3.9296	4.3178	5.2631
13	0.8702	1.3502	1.7709	2.1604	2.6503	3.0123	3.8520	4.2209	5.1106
14	0.8681	1.3450	1.7613	2.1448	2.6245	2.9768	3.7874	4.1403	4.9849
15	0.8662	1.3406	1.7531	2.1315	2.6025	2.9467	3.7329	4.0728	4.8801
16	0.8647	1.3368	1.7459	2.1199	2.5835	2.9208	3.6861	4.0149	4.7905
17	0.8633	1.3334	1.7396	2.1098	2.5669	2.8982	3.6458	3.9651	4.7148
18	0.8620	1.3304	1.7341	2.1009	2.5524	2.8784	3.6105	3.9217	4.6485
19	0.8610	1.3277	1.7291	2.0930	2.5395	2.8609	3.5793	3.8833	4.5903
20	0.8600	1.3253	1.7247	2.0860	2.5280	2.8453	3.5518	3.8496	4.5390
21	0.8591	1.3232	1.7207	2.0796	2.5176	2.8314	3.5271	3.8193	4.4925
22	0.8583	1.3212	1.7171	2.0739	2.5083	2.8188	3.5050	3.7922	4.4517
23	0.8575	1.3195	1.7139	2.0687	2.4999	2.8073	3.4850	3.7676	4.4156
24	0.8569	1.3178	1.7109	2.0639	2.4922	2.7970	3.4668	3.7454	4.3819
25	0.8562	1.3163	1.7081	2.0595	2.4851	2.7874	3.4502	3.7251	4.3516
26	0.8557	1.3150	1.7056	2.0555	2.4786	2.7787	3.4350	3.7067	4.3237
27	0.8551	1.3137	1.7033	2.0518	2.4727	2.7707	3.4210	3.6895	4.2992
28	0.8546	1.3125	1.7011	2.0484	2.4671	2.7633	3.4082	3.6739	4.2759
29	0.8542	1.3114	1.6991	2.0452	2.4620	2.7564	3.3963	3.6595	4.2538
30	0.8538	1.3104	1.6973	2.0423	2.4573	2.7500	3.3852	3.6460	4.2340
35	0.8520	1.3062	1.6896	2.0301	2.4377	2.7238	3.3400	3.5911	4.1531
40	0.8507	1.3031	1.6839	2.0211	2.4233	2.7045	3.3069	3.5510	4.0943
45	0.8497	1.3007	1.6794	2.0141	2.4121	2.6896	3.2815	3.5203	4.0489
50	0.8489	1.2987	1.6759	2.0086	2.4033	2.6778	3.2614	3.4960	4.0140
55	0.8482	1.2971	1.6730	2.0040	2.3961	2.6682	3.2451	3.4765	3.9855
60	0.8477	1.2958	1.6706	2.0003	2.3901	2.6603	3.2317	3.4602	3.9622
70	0.8468	1.2938	1.6669	1.9944	2.3808	2.6479	3.2108	3.4350	3.9255
80	0.8461	1.2922	1.6641	1.9901	2.3739	2.6387	3.1952	3.4164	3.8987
90	0.8456	1.2910	1.6620	1.9867	2.3685	2.6316	3.1832	3.4019	3.8778
100	0.8452	1.2901	1.6602	1.9840	2.3642	2.6259	3.1738	3.3905	3.8615
120	0.8446	1.2886	1.6576	1.9799	2.3578	2.6174	3.1595	3.3734	3.8370
∞	0.8416	1.2816	1.6449	1.9600	2.3264	2.5758	3.0902	3.2905	3.7189

Tavola 3: Valori critici della Distribuzione Chi-Quadrato

$$P(\chi^2 \geq \chi^2_{\alpha,\nu}) = \alpha$$



ν	α 0.9999	0.9995	0.999	0.995	0.99	0.975	0.95	0.90	0.80
1	1.57E-8	3.93E-7	1.57E-6	3.93E-5	0.0002	0.0010	0.0039	0.0158	0.0642
2	0.0002	0.0010	0.0020	0.0100	0.0201	0.0506	0.1026	0.2107	0.4463
3	0.0052	0.0153	0.0243	0.0717	0.1148	0.2158	0.3518	0.5844	1.0052
4	0.0284	0.0639	0.0908	0.2070	0.2971	0.4844	0.7107	1.0636	1.6488
5	0.0821	0.1581	0.2102	0.4118	0.5543	0.8312	1.1455	1.6103	2.3425
6	0.1723	0.2994	0.3810	0.6757	0.8721	1.2373	1.6354	2.2041	3.0701
7	0.2998	0.4849	0.5985	0.9893	1.2390	1.6899	2.1673	2.8331	3.8223
8	0.4634	0.7104	0.8571	1.3444	1.6465	2.1797	2.7326	3.4895	4.5936
9	0.6611	0.9718	1.1519	1.7349	2.0879	2.7004	3.3251	4.1682	5.3801
10	0.8890	1.2651	1.4787	2.1558	2.5582	3.2470	3.9403	4.8652	6.1791
11	1.1449	1.5870	1.8338	2.6032	3.0535	3.8157	4.5748	5.5778	6.9887
12	1.4281	1.9345	2.2141	3.0738	3.5706	4.4038	5.2260	6.3038	7.8073
13	1.7341	2.3049	2.6172	3.5650	4.1069	5.0087	5.8919	7.0415	8.6339
14	2.0601	2.6966	3.0407	4.0747	4.6604	5.6287	6.5706	7.7895	9.4673
15	2.4084	3.1073	3.4825	4.6009	5.2294	6.2621	7.2609	8.5468	10.3070
16	2.7736	3.5357	3.9417	5.1422	5.8122	6.9077	7.9616	9.3122	11.1521
17	3.1561	3.9800	4.4162	5.6973	6.4077	7.5642	8.6718	10.0852	12.0023
18	3.5559	4.4391	4.9048	6.2648	7.0149	8.2307	9.3904	10.8649	12.8570
19	3.9687	4.9125	5.4067	6.8439	7.6327	8.9065	10.1170	11.6509	13.7158
20	4.3950	5.3978	5.9210	7.4338	8.2604	9.5908	10.8508	12.4426	14.5784
21	4.8342	5.8954	6.4467	8.0336	8.8972	10.2829	11.5913	13.2396	15.4446
22	5.2862	6.4041	6.9829	8.6427	9.5425	10.9823	12.3380	14.0415	16.3140
23	5.7482	6.9240	7.5291	9.2604	10.1957	11.6885	13.0905	14.8480	17.1865
24	6.2231	7.4528	8.0847	9.8862	10.8563	12.4011	13.8484	15.6587	18.0618
25	6.7087	7.9905	8.6494	10.5196	11.5240	13.1197	14.6114	16.4734	18.9397
26	7.1980	8.5374	9.2222	11.1602	12.1982	13.8439	15.3792	17.2919	19.8202
27	7.6997	9.0929	9.8029	11.8077	12.8785	14.5734	16.1514	18.1139	20.7030
28	8.2115	9.6558	10.3907	12.4613	13.5647	15.3079	16.9279	18.9392	21.5880
29	8.7303	10.2266	10.9861	13.1211	14.2564	16.0471	17.7084	19.7677	22.4751
30	9.2559	10.8040	11.5876	13.7867	14.9535	16.7908	18.4927	20.5992	23.3641
35	11.9929	13.7879	14.6881	17.1917	18.5089	20.5694	22.4650	24.7966	27.8359
40	14.8820	16.9058	17.9166	20.7066	22.1642	24.4331	26.5093	29.0505	32.3449
45	17.8922	20.1361	21.2509	24.3110	25.9012	28.3662	30.6123	33.3504	36.8844
50	21.0077	23.4611	24.6736	27.9908	29.7067	32.3574	34.7642	37.6886	41.4492
55	24.2133	26.8650	28.1731	31.7349	33.5705	36.3981	38.9581	42.0596	46.0356
60	27.5006	30.3393	31.7381	35.5344	37.4848	40.4817	43.1880	46.4589	50.6406
70	34.2581	37.4671	39.0358	43.2753	45.4417	48.7575	51.7393	55.3289	59.8978
80	41.2407	44.7917	46.5197	51.1719	53.5400	57.1532	60.3915	64.2778	69.2070
90	48.4095	52.2768	54.1559	59.1963	61.7540	65.6466	69.1260	73.2911	78.5584
100	55.7202	59.8946	61.9182	67.3275	70.0650	74.2219	77.9294	82.3581	87.9453
150	93.9492	99.4617	102.1127	109.1423	112.6676	117.9846	122.6918	128.2750	135.2625
200	134.0154	140.6591	143.8420	152.2408	156.4321	162.7280	168.2785	174.8353	183.0028

Tavola 3 $_{(segue)}$: Valori critici della Distribuzione Chi-Quadrato

ν	α 0.20	0.10	0.05	0.025	0.01	0.005	0.001	0.0005	0.0001
1	1.6424	2.7055	3.8415	5.0239	6.6349	7.8794	10.8274	12.1153	15.1343
2	3.2189	4.6052	5.9915	7.3778	9.2104	10.5965	13.8150	15.2014	18.4247
3	4.6416	6.2514	7.8147	9.3484	11.3449	12.8381	16.2660	17.7311	21.1040
4	5.9886	7.7794	9.4877	11.1433	13.2767	14.8602	18.4662	19.9977	23.5064
5	7.2893	9.2363	11.0705	12.8325	15.0863	16.7496	20.5147	22.1057	25.7507
6	8.5581	10.6446	12.5916	14.4494	16.8119	18.5475	22.4575	24.1016	27.8527
7	9.8032	12.0170	14.0671	16.0128	18.4753	20.2777	24.3213	26.0179	29.8814
8	11.0301	13.3616	15.5073	17.5345	20.0902	21.9549	26.1239	27.8674	31.8268
9	12.2421	14.6837	16.9190	19.0228	21.6660	23.5893	27.8767	29.6669	33.7247
10	13.4420	15.9872	18.3070	20.4832	23.2093	25.1881	29.5879	31.4195	35.5572
11	14.6314	17.2750	19.6752	21.9200	24.7250	26.7569	31.2635	33.1382	37.3647
12	15.8120	18.5493	21.0261	23.3367	26.2170	28.2997	32.9092	34.8211	39.1306
13	16.9848	19.8119	22.3620	24.7356	27.6882	29.8193	34.5274	36.4768	40.8735
14	18.1508	21.0641	23.6848	26.1189	29.1412	31.3194	36.1239	38.1085	42.5752
15	19.3107	22.3071	24.9958	27.4884	30.5780	32.8015	37.6978	39.7173	44.2596
16	20.4651	23.5418	26.2962	28.8453	31.9999	34.2671	39.2518	41.3077	45.9255
17	21.6146	24.7690	27.5871	30.1910	33.4087	35.7184	40.7911	42.8808	47.5591
18	22.7595	25.9894	28.8693	31.5264	34.8052	37.1564	42.3119	44.4337	49.1853
19	23.9004	27.2036	30.1435	32.8523	36.1908	38.5821	43.8194	45.9738	50.7873
20	25.0375	28.4120	31.4104	34.1696	37.5663	39.9969	45.3142	47.4977	52.3832
21	26.1711	29.6151	32.6706	35.4789	38.9322	41.4009	46.7963	49.0096	53.9599
22	27.3015	30.8133	33.9245	36.7807	40.2894	42.7957	48.2676	50.5105	55.5244
23	28.4288	32.0069	35.1725	38.0756	41.6383	44.1814	49.7276	51.9995	57.0668
24	29.5533	33.1962	36.4150	39.3641	42.9798	45.5584	51.1790	53.4776	58.6071
25	30.6752	34.3816	37.6525	40.6465	44.3140	46.9280	52.6187	54.9475	60.1360
26	31.7946	35.5632	38.8851	41.9231	45.6416	48.2898	54.0511	56.4068	61.6666
27	32.9117	36.7412	40.1133	43.1945	46.9628	49.6450	55.4751	57.8556	63.1660
28	34.0266	37.9159	41.3372	44.4608	48.2782	50.9936	56.8918	59.2990	64.6561
29	35.1394	39.0875	42.5569	45.7223	49.5878	52.3355	58.3006	60.7342	66.1524
30	36.2502	40.2560	43.7730	46.9792	50.8922	53.6719	59.7022	62.1600	67.6230
35	41.7780	46.0588	49.8018	53.2033	57.3420	60.2746	66.6192	69.1975	74.9253
40	47.2685	51.8050	55.7585	59.3417	63.6908	66.7660	73.4029	76.0963	82.0551
45	52.7288	57.5053	61.6562	65.4101	69.9569	73.1660	80.0776	82.8734	89.0704
50	58.1638	63.1671	67.5048	71.4202	76.1538	79.4898	86.6603	89.5597	95.9713
55	63.5772	68.7962	73.3115	77.3804	82.2920	85.7491	93.1671	96.1607	102.7735
60	68.9721	74.3970	79.0820	83.2977	88.3794	91.9518	99.6078	102.6971	109.4967
70	79.7147	85.5270	90.5313	95.0231	100.4251	104.2148	112.3167	115.5766	122.7443
80	90.4053	96.5782	101.8795	106.6285	112.3288	116.3209	124.8389	128.2636	135.7728
90	101.0537	107.5650	113.1452	118.1359	124.1162	128.2987	137.2082	140.7804	148.6198
100	111.6667	118.4980	124.3421	129.5613	135.8069	140.1697	149.4488	153.1638	161.3297
150	164.3492	172.5812	179.5806	185.8004	193.2075	198.3599	209.2652	213.6135	223.1209
200	216.6088	226.0210	233.9942	241.0578	249.4452	255.2638	267.5388	272.4220	283.0448