UNIVERSITA' DEGLI STUDI DI BRESCIA-FACOLTA' DI MEDICINA E CHIRURGIA
CORSO DI LAUREA IN INFERMIERISTICA SEDE DI DESENZANO dG
STATISTICA MEDICA

LEZIONE 5

Ing. Andrea Ghedi AA 2008/2009

La distribuzione "t-Student"

Quando la varianza è “stimata dal campione”

Si è visto in precedenza che se ad una variabile gaussiana (x) sottraiamo la media (μ) e dividiamo tale differenza per la deviazione standard (s) otteniamo una **deviata gaussiana standard** (z) con media 0 e varianza 1:

se $x \sim N(\mu, \sigma^2)$, $z \sim N(0, 1)$ dove $z = \dfrac{x-\mu}{\sigma}$

Poiché le medie campionarie ($\overline{x}$), calcolate su campioni tratti dalla variabile $x \sim N(\mu, \sigma^2)$, hanno distribuzione gaussiana con media μ e varianza σ^2/n, **se standardizziamo la variabile media campionaria** otteniamo una deviata gaussiana standard z con media 0 e varianza 1:

$$\boxed{\text{se} \quad \overline{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right), \quad z \sim N(0,1) \quad \text{dove} \quad z = \frac{\overline{x}-\mu}{\sigma/\sqrt{n}}}$$

Quando il parametro σ^2 è ignoto, possiamo sostituirlo con la sua stima campionaria s^2, ed ottenere il rapporto

$$\boxed{\text{Gaussiana} \quad z = \frac{(x-\mu)}{\sigma}}$$

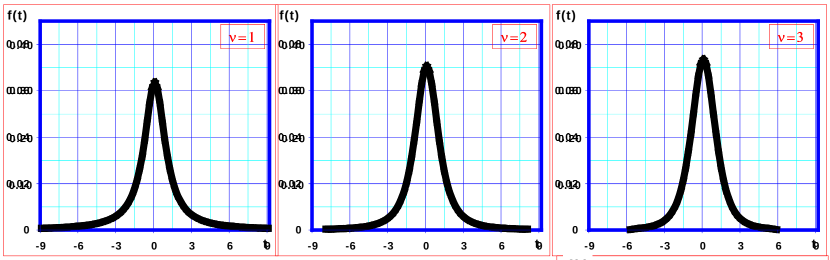
$$t = \frac{\overline{x}-\mu}{s/\sqrt{n}}$$

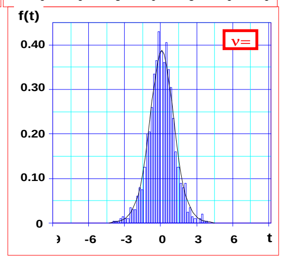
Qual è la distribuzione di tale rapporto ?

Distribuzione dei rapporti "t" calcolati su **4 serie di 1000 campioni**

(rispettivamente di dimensione **n=2, 3, 4 e 10**), tratti dalla distribuzione dei livelli ematici di Calcio in una popolazione di riferimento.

Si può dimostrare che, per campioni tratti da una variabile gaussiana, il rapporto "t" è una variabile casuale la cui distribuzione è descritta da una funzione simmetrica la cui forma dipende da (i gradi di libertà della stima campio-naria della varianza) e che è nota con il nome di "t" di Student.





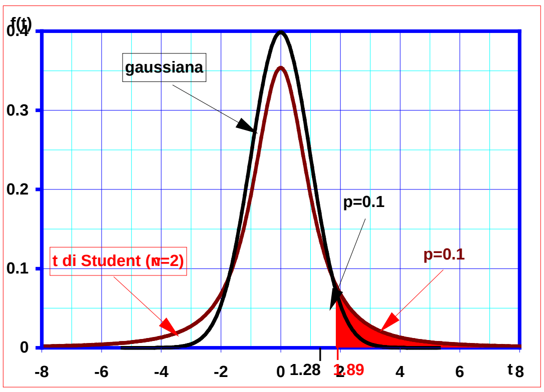
all' aumentare dei gradi di libertà
la distribuzione "t" di Student
tende rapidamente alla Gaussiana standard.

Cioè la distribuzione t di Student ha un area maggiore sotto le code rispetto alla normale

Questa ha **code più alte, fianchi più stretti e varianza maggiore** rispetto alla Gaussiana standard:

all' aumentare dei gradi di libertà la distribuzione "t" di Student **tende rapidamente alla Gaussiana standard**.

$$\frac{\bar{x}-\mu}{s/\sqrt{n}} \sim \text{t di Student (con } \nu = n-1 \text{ g.d.l.)}$$



UNIVERSITA' DEGLI STUDI DI BRESCIA-FACOLTA' DI MEDICINA E CHIRURGIA
CORSO DI LAUREA IN INFERMIERISTICA SEDE DI DESENZANO dG

STATISTICA MEDICA

Probabilità

ν	.7500	.8000	.8500	.9000	.9500	.9750	.9900	.9950	.9990	.9995
1	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.22	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850

Ing. Andrea Ghedi AA 2008/2009

UNIVERSITA' DEGLI STUDI DI BRESCIA-FACOLTA' DI MEDICINA E CHIRURGIA
CORSO DI LAUREA IN INFERMIERISTICA SEDE DI DESENZANO dG

STATISTICA MEDICA

Probabilità

ν	.7500	.8000	.8500	.9000	.9500	.9750	.9900	.9950	.9990	.9995
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
120	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617	3.160	3.373
250	0.675	0.843	1.039	1.285	1.651	1.969	2.341	2.596	3.123	3.330
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
INF	0.675	0.842	1.036	1.282	1.645	1.960	2.327	2.576	3.091	3.291

Ing. Andrea Ghedi AA 2008/2009

Distribuzione T di Student

E' la distribuzione di densità di probabilità della variabile:

$$t = \frac{\overline{x} - \mu}{s_m}$$

Che corrisponde all'errore commesso utilizzando una stima (POPOLAZIONE) della media x in luogo della media vera (CAMPIONE).

A prima vista potrebbe sembrare una gaussiana ma al denominatore c'è S_m che è a sua volta una stima e quindi soggetta ad errore.

La stima S_m si usa quando non si conosce σ, quindi la t di Student si utilizza quando è ignota la deviazione standard, dove l'unica σ^2 noto è quello del campione.

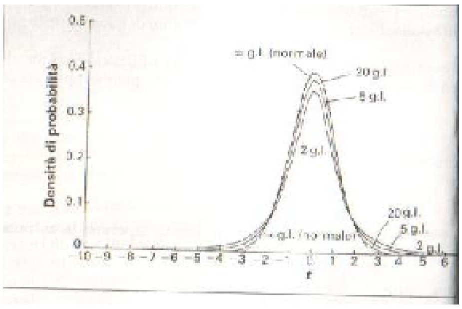
UNIVERSITA' DEGLI STUDI DI BRESCIA-FACOLTA' DI MEDICINA E CHIRURGIA
CORSO DI LAUREA IN INFERMIERISTICA SEDE DI DESENZANO dG
STATISTICA MEDICA

Distribuzione t di Student

La t segue la distribuzione nota come "distribuzione della variabile continua t" con (n-1) gradi di libertà.

Le distribuzioni t formano una famiglia di distribuzioni contraddistinte da un indice che sono i gradi di libertà, essi sono pari alla dimensione del campione meno 1.

All'aumentare dei gradi di libertà la distribuzione t tende alla distribuzione normale standardizzata.



Ing. Andrea Ghedi AA 2008/2009

Quando si usa la distribuzione t?

La t di student si usa in due casi:

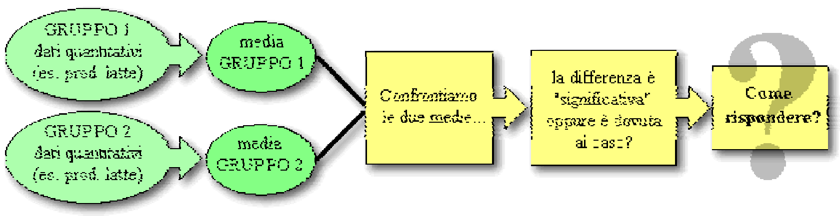
-Quando non è nota la varianza della popolazione (e la si stima con la varianza campionaria. Se la varianza è nota, si usa la normale)

- Quando il campione è piccolo (se è grande si può usare per approssimazione la normale, alla quale la t tende quando il numero di gradi di libertà è elevato)

-Quando si confrontano due medie

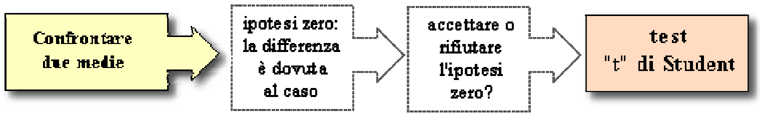
UNIVERSITA' DEGLI STUDI DI BRESCIA-FACOLTA' DI MEDICINA E CHIRURGIA
CORSO DI LAUREA IN INFERMIERISTICA SEDE DI DESENZANO dG
STATISTICA MEDICA

Una circostanza comune è quella in cui si sono esaminati due campioni di pazienti, in ciascuno dei quali è stata misurata una variabile *numerica* (es. altezza, peso, frequenza cardiaca, ecc.) di cui è stata poi calcolata la media. Ci si chiede: *la differenza fra le due medie è significativa?*
Ossia: si può affermare che *la differenza non è dovuta al caso ma che, invece, esiste una reale diversità tra le medie delle due popolazioni da cui i campioni stessi derivano?*



Ing. Andrea Ghedi AA 2008/2009

Allora, si deve ricorrere ad un altro test: la t di Student
Lo schema di ragionamento è:



Una volta trovato il valore *t*, esso va confrontato con quelli tabulati (le tabelle si trovano in tutti i libri di statistica) al fine di stabilire se la differenza fra le due medie non sia dovuta al caso.

Ovviamente oggi il test *t* si esegue al computer con l'aiuto di apposito software. Più che il metodo di calcolo, è importante che tu conosca almeno l'esistenza del test *t* ed il contesto in cui esso si applica.
Importantissimo è che ci si renda conto che, anche nel confronto fra due medie, non si possono trarre conclusioni «ad occhio», ma è indispensabile ricorrere ad un test statistico.

Ipotesi:

Supponiamo di avere un campione di numerosità n estratto da una popolazione con media μ e deviazione standard s entrambe NON CONOSCIUTE.

Ipotizziamo che la media del campione sia μ_0 e quindi formuliamo un ipotesi H_0: $\mu = \mu_0$. calcoliamo la statistica t

$$\frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

Se il valore che ottengo dal test è all'interno del mio "intervallo di confidenza" allora accetto l'ipotesi nulla altrimenti la rifiuto.

STATISTICA MEDICA

Un esempio di applicazione del test *t*

Si sperimenta l'effetto sul suino dell'aggiunta alla razione di una miscela «probiotica» costituita da batteri normalmente presenti nella flora intestinale del suino. L'ipotesi da verificare è che il probiotico incrementi l'accrescimento degli animali.
Allo scopo di verificare l'ipotesi, si disegna uno studio sperimentale preliminare su due piccoli gruppi di suini. I gruppi sono fra loro omogenei (stessa razza, età, provenienza ecc.) e sono mantenuti nelle stesse condizioni di allevamento (alimentazione, temperatura ambiente ecc.).
L'unica differenza è che alla razione del Gruppo 1 (10 suini) viene aggiunto il probiotico, mentre al gruppo 2 (11 suini) no.
All'inizio dell'esperimento ciascun suino viene pesato; dopo 21 giorni di trattamento i suini vengono pesati di nuovo e per ogni animale si calcola l'incremento giornaliero medio. I dati sono riportati nella tabella:

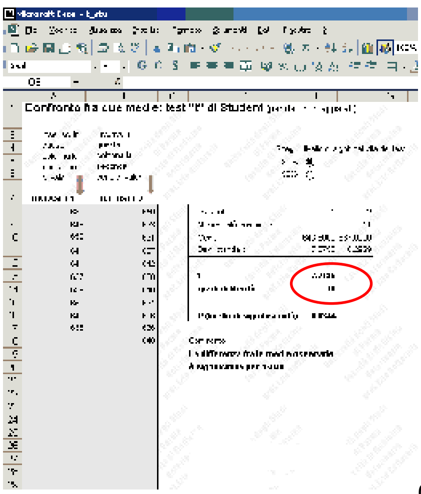
Incremento medio giornaliero (grammi)

n	Gruppo 1 (trattati)	Gruppo 2 (controlli)
1	639	650
2	646	633
3	650	631
4	641	637
5	641	642
6	637	638
7	659	640
8	650	634
9	640	626
10	635	636
11		640
media	643.8	637.0

UNIVERSITA' DEGLI STUDI DI BRESCIA-FACOLTA' DI MEDICINA E CHIRURGIA
CORSO DI LAUREA IN INFERMIERISTICA SEDE DI DESENZANO dG

STATISTICA MEDICA

Confrontando le medie degli accrescimenti, puoi notare che il valore del Gruppo 1 è superiore a quello del Gruppo 2 (643.8 g/giorno contro 637.0 g/giorno). La domanda è: questa differenza è dovuta al probiotico oppure al caso?
L'ipotesi zero dice che la differenza è dovuta al caso... accetti o rifiuti questa ipotesi? Per rispondere, applichiamo il test *t* ai dati



UNIVERSITA' DEGLI STUDI DI BRESCIA-FACOLTA' DI MEDICINA E CHIRURGIA
CORSO DI LAUREA IN INFERMIERISTICA SEDE DI DESENZANO dG

STATISTICA MEDICA

In pratica, si deve confrontare il valore che si è ottenuto (ossia 2.2796) con quelli presenti nella Tabella dei valori t, per 19 gradi di libertà (gradi di libertà=numero osservazioni-numero gruppi; nel nostro esempio 21-2=19). Il valore è *superiore* a quello della colonna p=5% (ma inferiore a quello della colonna p=1%). Perciò si puo rifiutare l'ipotesi zero e concludere che la differenza è significativa per p<0.05 (ma non per p<0.01).
Ciò significa che c'è una probabilità inferiore al 5% (ma non all'1%) che la differenza di accrescimento tra il gruppo trattato e quello di controllo è dovuta al caso.

Notare che i due valori p (0.05 e 0.01) sono valori convenzionalmente utilizzati nel modo scientifico (questo concetto verrà spiegato meglio nel proseguo).
Nel foglio di calcolo, così come in tutti i pacchetti statistici, il confronto del valore t che si ottiene con i valori della Tabella per 'n' gradi di libertà è inutile: infatti il programma restituisce direttamente il valore p. Con i tuoi dati, p=0.0344; quindi che c'è una probabilità inferiore a 3.44% che la differenza di accrescimento tra il gruppo trattato e quello di controllo sia dovuta al caso. In altri termini, si puo affermare che la differenza fra il gruppo dei trattati ed il gruppo dei controlli è significativa per p=0.344.

Valori t di Student

Gradi di libertà	p=0.05*	p=0.01
1	12.706	63.657
2	4.303	9.925
3	3.182	5.841
4	2.776	4.604
5	2.571	4.032
6	2.447	3.707
7	2.365	3.499
8	2.306	3.355
9	2.262	3.250
10	2.228	3.169
11	2.201	3.106
12	2.179	3.055
13	2.160	3.012
14	2.145	2.977
15	2.131	2.947
16	2.120	2.921
17	2.110	2.898
18	2.101	2.878
19	2.093	2.861
20	2.086	2.845
...	...	...
...	...	...

* La probabilità 0.05 è quella di osservare un valore t superiore a quello tabulato per i gradi di libertà desiderati

Ing. Andrea Ghedi AA 2008/2009

UNIVERSITA' DEGLI STUDI DI BRESCIA-FACOLTA' DI MEDICINA E CHIRURGIA
CORSO DI LAUREA IN INFERMIERISTICA SEDE DI DESENZANO dG

STATISTICA MEDICA

Distribuzione ?² (chi-quadro)

Rappresenta una famiglia di distribuzioni (in funzione dei gradi di libertà) continue ed asimmetriche (l'asimmetria diminuisce con l'aumentare dei gradi di libertà), che possono assumere solo valori positivi (da 0 a +¥). E' la distribuzione assunta da una variabile continua casuale elevata al quadrato o dalla somma dei quadrati di *n* variabili continue, pertanto un esempio di questa distribuzione e' la seguente:

$$\chi^2 = \sum_{i=1}^{k} \frac{(x_i - \mu)^2}{\sigma^2} = \frac{(n-1)s^2}{\sigma^2} = \frac{dev(x)}{\sigma^2}$$

dove:

k = gradi di liberta'

μ = media della popolazione

σ^2 = varianza della popolazione

n = numero dei casi

La varianza campionaria e' distribuita nel seguente modo: $s^2 \approx \frac{\chi^2 \cdot \sigma^2}{n-1}$

Il chi quadro e' un numero che dovrebbe dire quanto i dati sperimentali sono vicini a una distribuzione teorica.

Ing. Andrea Ghedi AA 2008/2009

UNIVERSITA' DEGLI STUDI DI BRESCIA-FACOLTA' DI MEDICINA E CHIRURGIA
CORSO DI LAUREA IN INFERMIERISTICA SEDE DI DESENZANO dG
STATISTICA MEDICA

Distribuzione ?² (chi-quadro)

1.PRENDIAMO UNA POPOLAZIONE CON UNA DISTRIBUZIONE NORMALE CON LA MEDIA E LA DEVIAZIONE STANDARD NOTE.

2. ESTRAIAMO DEI CAMPIONI DI DIMENSIONE n DA QUESTA POPOLAZIONE.

3. CONVERTIAMO CIASCUN VALORE IN UNA VARIABILE STANDARDIZZATA SOTTRAENDO LA MEDIA PARAMETRICA E DIVIDENDO PER LA DEVIAZIONE STANDARD PARAMETRICA.

4.FACCIAMO IL QUADRATO DI CIASCUNA VARIABILE STANDARDIZZATA E SOMMIAMO I QUADRATI. LO FACCIAMO UN NUMERO INFINITO DI VOLTE.

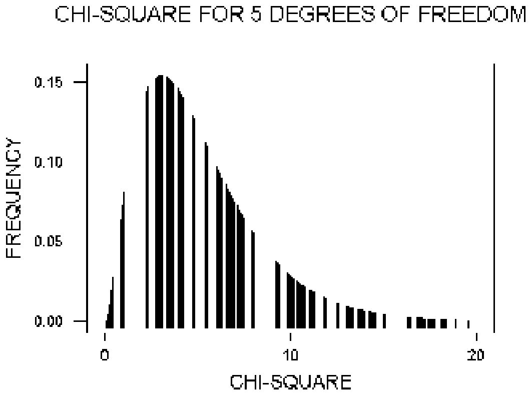
5. LA SOMMA DELLE VARIABILI STANDARDIZZATE ELEVATE AL QUADRATO HA UNA DISTRIBUZION CHI-QUADRO CON n-1 GRADI DI LIBERTA'.

$$\chi^2 - \sum\left[\frac{(X-\mu)}{\sigma}\right]^2$$

UNIVERSITA' DEGLI STUDI DI BRESCIA-FACOLTA' DI MEDICINA E CHIRURGIA
CORSO DI LAUREA IN INFERMIERISTICA SEDE DI DESENZANO dG
STATISTICA MEDICA

Distribuzione ?² (chi-quadro)

LA DISTRIBUZIONE CHI-QUADRO PER 5 GRADI DI LIBERTA' (n = 6) SI PRESENTA COSI':



Intervallo di confidenza

I parametri della popolazione (quale media e varianza) sono delle costanti, ma spesso, date le caratteristiche della popolazione, non è dato conoscerli, siamo soltanto in grado di **stimarli** studiando dei campioni estratti casualmente dalla popolazione. Tanto maggiore è la numerosità campionaria tanto migliore sarà la statistica stimata, ovvero questa si avvicinerà sempre di più al valore reale del parametro.

Potendo estrarre differenti campioni ed ottenendo altrettante stime, risulta interessante individuare un modo che ci consenta di esprimere l'incertezza relativa alla nostra stima campionaria.

In statistica viene solitamente utilizzata come riferimento la probabilità del 95% per definire un **evento probabile** e la probabilità del 5% (indicata come a) per definirlo **improbabile**, possiamo allora definire un intervallo (appunto del 95%) per il quale un evento ha una probabilità accettabile di verificarsi. In particolare è interessante definire tale intervallo relativamente ad una data statistica (media o varianza) campionaria: tale intervallo si chiama **intervallo fiduciale al 95%**.

Nell'intervallo sarà compreso il reale parametro della popolazione nel 95% dei casi, per una serie di intervalli calcolati su altrettanti campioni, di identica numerosità, estratti dalla popolazione. Non sapremo però mai quale di questi intervalli conterranno effettivamente il parametro.

Intervallo di confidenza

L'**intervallo di confidenza** esprime il margine statistico d'errore. Ad esempio, in un campione il 47% ha risposto "sì" a una certa domanda. Con un intervallo di confidenza 4 (cioè del 4%) la percentuale di persone che risponderebbero "sì", nell'eventualità di un'intervista a tappeto a tutta la popolazione, sarebbe compresa fra il 43% (47%-4%) e il 51% (47%+4%).

Fattori che determinano gli intervalli di confidenza

Sono tre i fattori che determinano la grandezza dell'intervallo di confidenza per un certo livello di confidenza: grandezza del campione, percentuale e grandezza della popolazione.

Grandezza del campione

Più il campione è ampio, più è alto il grado di attendibilità delle risposte. Ciò significa che, dato un certo livello di confidenza, più è ampio il campione, più è piccolo l'intervallo di confidenza. Si tratta comunque di una relazione non lineare: per esempio, raddoppiando il campione non si dimezza l'intervallo di confidenza.

Percentuale

L'accuratezza dipende anche dalla distribuzione percentuale delle risposte. Per esempio se il 99% del campione risponde "sì" e l'1% "no", le probabilità di errore sono remote, indipendentemente dalla grandezza del campione. Nel caso invece di risposte che oscillano intorno al 50%, c'è una maggiore possibilità di ottenere dati non verosimili.

Grandezza della popolazione

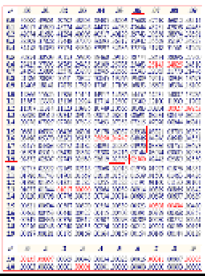
Spesso non si conosce l'esatta grandezza della popolazione, ma questo non è un problema. La statistica dimostra infatti che la grandezza della popolazione è irrilevante, a meno che questo parametro non ecceda di una piccola percentuale la popolazione totale in esame.

UNIVERSITA' DEGLI STUDI DI BRESCIA-FACOLTA' DI MEDICINA E CHIRURGIA
CORSO DI LAUREA IN INFERMIERISTICA SEDE DI DESENZANO dG
STATISTICA MEDICA

Per calcolare l'intervallo di confidenza è necessario un modello di probabilità che tenga conto dei diversi *possibili* risultati di uno studio. Di tali modelli ne esistono, in statistica, diversi tipi (ad es. sulla distr. gaussiana). In genere, quando il numero di osservazioni è abbastanza ampio, si utilizza proprio il modello di distribuzione gaussiana (detta anche "normale").

Dato un certo valore di prevalenza P, l'intervallo di confidenza 95% (1.96 lo trovo nella tabella facendo 1-(0.025-0.025)) si ottiene con il seguente calcolo, dove il segno +/- permette il calcolo del limite superiore (+) e del limite inferiore (-) dell'intervallo:

$$P \pm 1.96\sqrt{\frac{P(1-P)}{N}}$$



Intervallo di confidenza:esempio

ESEMPIO. Supponiamo che in uno studio sulla displasia dell'anca di cani di razza "pastore tedesco" siano risultati affetti dalla malattia 18 cani su 180 esaminati. La prevalenza è: 18/180 = 0.1, cioè 10%.
Quindi calcoliamo:

Px(1-P)=0.1(1-0.1)=0.10x0.9

$$0.10 \pm 1.96\sqrt{\frac{0.10*0.9}{180}} = 0.056\,;\,0.145$$

Pertanto, il limite inferiore dell'intervallo di confidenza 95% è 0.056 (5.6%) ed il limite superiore 0.145 (14.5%). Ciò significa che, in media, il 95% di tali intervalli derivanti da studi privi di errori sistematici contiene il parametro vero della popolazione. In altre parole, possiamo essere abbastanza sicuri che la percentuale di cani con displasia dell'anca nella *intera popolazione* da cui è stato tratto il campione di 180 cani sia compresa fra 5.6 e 14.5%.

Livello di significatività

Nell'operare un test, la massima probabilità con la quale siamo disposti a rischiare un errore è detta "**Livello di Significatività**" del test. Questa probabilità è in generale specificata prima di aver estratto il campione, di guisa che i risultati ottenuti non influenzeranno la nostra scelta.
In pratica si usano livelli di significatività dello 0,05 (5%) o dello 0,01 (1%), sebbene sia possibile usare anche altri valori. Se ad esempio si sceglie un livello di significatività dello 0,05 ossia 5%, ci saranno allora 5 possibilità su 100 che si rigettino ipotesi che avrebbero dovuto essere accettate, cioè siamo "*fiduciosi*" circa al 95% di aver preso la decisione giusta. In questi casi diremo che l'ipotesi è stata rigettata al livello di significatività dello 0,05, che significa che potremmo aver sbagliato con una probabilità pari allo 005.

La verifica delle ipotesi

Può capitare di dover decidere sulla base di dati campionari (prelevati da un campione)

Se un nuovo vaccino abbia veramente un effetto positivo.

E' utile per raggiungere una decisione certa (non potendo analizzare tutta la popolazione) oppure per comparare i risultati di due gruppi di campioni diversi (campione 1 e campione 2) su cui analizzo la stessa caratteristica.

Fare delle **IPOTESI STATISTICHE.**

Le ipotesi statistiche si basano sull' ipotesi zero o ipotesi nulla.

Ipotesi nulla

H_0: $\mu_1=\mu_2$ Cioè non esiste nessuna differenza tra la media del campione 1 e quella del campione 2

Ipotesi alternativa

H_a: $\mu_1\neq\mu_2$ Tra i campioni confrontati esiste una differenza non casuale per quanto riguarda il caratterre esaminato

La verifica delle ipotesi

Alla fine si giunge all'accettazione o al rifiuto dell'ipotesi nulla (ipotesi principale) e la regola che, sulla base dell'informazione campionaria, porta all'accettazione o al rifiuto di tale ipotesi è conosciuta con il nome di "test statistico".

La scelta tra **le due ipotesi (H0 e H1)** è fondata sulla **probabilità di ottenere per caso il risultato osservato nel campione o un risultato ancor più distante da quanto atteso, nella condizione che l'ipotesi nulla H0 sia vera**. Quanto più tale probabilità (indicata con a) è piccola, tanto più è improbabile che l'ipotesi nulla H0 sia vera.

STATISTICA MEDICA

La verifica delle ipotesi

L'insieme di valori ottenibili con il test formano la **distribuzione campionaria dell'indice statistico**.

Essa può essere divisa in due zone:

1 - la **zona di rifiuto** dell'ipotesi nulla, detta anche **regione critica**, che **corrisponde ai valori collocati**

agli estremi della distribuzione secondo la direzione dell'ipotesi alternativa Ha; sono quei valori che hanno una probabilità piccola di verificarsi per caso, **quando l'ipotesi nulla H0 è vera**;

2 - la **zona di accettazione** dell'ipotesi nulla H0, che comprende **i restanti valori**, quelli che si possono trovare abitualmente per effetto della variabilità casuale.

Se il valore dell'indice statistico calcolato cade nella zona di rifiuto, si respinge l'ipotesi nulla H0.

Con un test statistico è posta in discussione la credibilità dell'ipotesi nulla. Occorre sempre accettarla, a meno di dimostrare che quanto effettivamente trovato abbia una probabilità piccola di essere avvenuto per caso.

UNIVERSITA' DEGLI STUDI DI BRESCIA-FACOLTA' DI MEDICINA E CHIRURGIA
CORSO DI LAUREA IN INFERMIERISTICA SEDE DI DESENZANO dG

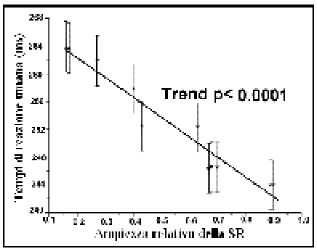
STATISTICA MEDICA

La verifica delle ipotesi

Per consolidata **convenzione internazionale, i livelli di soglia delle probabilità** a ai quali di norma si

ricorre sono tre: 0.05 (5%); 0.01 (1%); 0.001 (0.1%). Nella presentazione sintetica dei risultati e nella discussione conclusiva dei test, quando è possibile solo l'uso di tabelle sinottiche (riassuntive) con i valori critici, i differenti livelli di significatività sono indicati con una **simbologia e con parole chiave**.

LIVELLO DI PROBABILITA'	RISULTATO DEL TEST	SIMBOLO
P < 0.05 (livello 5%)	significativo	*
P < 0.01 (livello 1%)	molto significativo	**
P < 0.001 (livello 0,1%)	altamente significativo	***



FINE