

# Analisi statistica di dati biomedici

## Analysis of biological signals

**II Parte – Verifica delle ipotesi (a)**

**Agostino Accardo ([accardo@units.it](mailto:accardo@units.it))**

Master in Ingegneria Clinica – LM in Neuroscienze  
2013-2014 e segg.

# VERIFICA DELLE IPOTESI

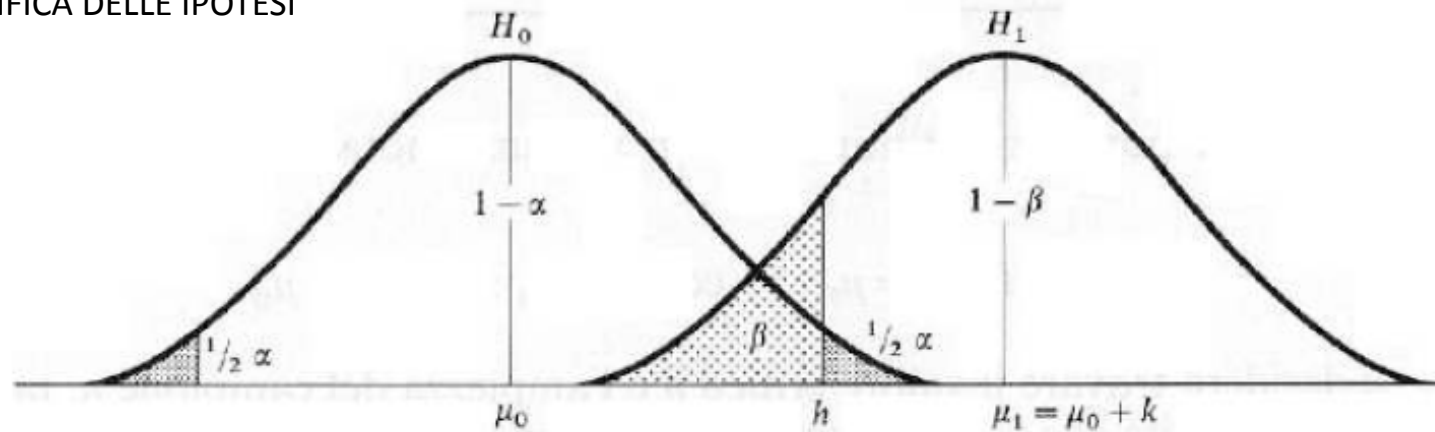
L'ipotesi da verificare riguarda in genere la generalizzazione di un risultato ottenuto su un campione, a tutta la popolazione.

La verifica avviene tramite varie fasi che si concludono con un test.

## FASI:

1. Formulazione dell'**ipotesi NULLA (H0)**, nulla perché viene formulata allo scopo di rifiutarla. H1 rappresenta l'ipotesi alternativa (generalmente contraria ad H0). H0 di solito pone l'assenza di relazioni significative tra variabili o tra campioni (p.es. delle differenze presenti tra campioni), per cui le eventuali differenze sono dovute al caso
2. Individuazione della distribuzione campionaria che, a seconda del test usato, dovrà soddisfare ad opportuni requisiti: la casualità dell'estrazione è sempre richiesta. Possono inoltre essere richieste indipendenza dei campioni, distribuzione nota, ecc. **SEMPRE RICHIESTA: CASUALITA' 'ESTRAZIONE'**
3. Scelta del livello  $\alpha$  di significatività (in genere tra 0.01 e 0.05) che si può ricavare anche a posteriori (p-value = valore di probabilità di ottenere la differenza osservata, se H0 è vera)
4. Selezione del test e sua applicazione con relativa decisione di accettare o rifiutare l'H0 con un livello di probabilità  $\alpha$  di sbagliare

## VERIFICA DELLE IPOTESI



Distribuzioni legate alle due ipotesi  $H_0$  e  $H_1$ , mutuamente esclusive

Il test statistico confronta la stima campionaria con le distribuzioni  $H_0$  e  $H_1$  e la associa ad una delle due

- A priori non sono mai sicuro che il valore misurato appartenga di diritto più all'una che all'altra distribuzione
- Esiste un'area di sovrapposizione che dipende dalla distanza dei valori medi e dalle variabilità
- Scelto  $\alpha$  ed il numero di 'code' viene determinato  $h$  (=soglia di significatività) e quindi  $\beta$

## VERIFICA DELLE IPOTESI

Ogni test è associato a 4 probabilità interdipendenti che misurano il rischio che si corre, ovvero della sicurezza che si ha, nel formulare una conclusione:

- **Errore di I tipo** (rischio  $\alpha$  o livello di significatività a cui corrisponde il p-value): probabilità che esprime il rischio di RIFIUTARE  $H_0$  quando è VERA (falso negativo)
- **Errore di II tipo** (rischio  $\beta$ ): probabilità del rischio di ACCETTARE  $H_0$  quando è FALSA (falso positivo)
- **Protezione del test** ( $1 - \alpha$ ): probabilità di ACCETTARE  $H_0$  quando è VERA
- **Potenza del test** ( $1 - \beta$ ): probabilità di RIFIUTARE  $H_0$  quando è FALSA

CONCLUSIONE DEL TEST	REALTA'	
	$H_0$ vera	$H_0$ falsa
<b>accetto <math>H_0</math></b> statisticamente non significativo	<b>Esatto</b> $p = 1 - \alpha$ PROTEZIONE	<b>Errore <math>\beta</math></b> di II <sup>^</sup> tipo $p = \beta$
<b>rifiuto <math>H_0</math></b> statisticamente significativo	<b>Errore <math>\alpha</math></b> di I <sup>^</sup> tipo $p = \alpha$	<b>Esatto</b> $p = 1 - \beta$ POTENZA

## VERIFICA DELLE IPOTESI

Poiché è impossibile diminuire un errore senza aumentare l'altro:

- scelgo tra i test a disposizione quello più 'potente' (che minimizza  $\beta$ )
- aumento la numerosità dei dati nel campione => cala la varianza delle distribuzioni campionarie (si 'stringono') e di conseguenza il  $\beta$

Note:

- Di solito la distribuzione  $H_1$  non è nota e ci si basa solo sui dati per l' $H_p$ .  $H_0$ , di conseguenza non si valutano bene gli errori  $\beta$
- Si può determinare la dimensione ottima  $N$  del campione imponendo un certo  $\beta$ , utilizzando il Nomogramma precedente (che parte da  $1-\beta$ , da  $\alpha$ , da 's' = dev.std nella popolazione e dalla differenza significativa accettata,  $\delta$ , da cui si ricava  $\delta/s$  = differenza standardizzata), ovvero, dato  $N$ , dal medesimo Nomogramma posso stimare il  $\beta$

# VIA NOMOGRAMMA

456 Clinical trials

can be, but greater power requires a larger sample, as we will see. It is common to require a power of between 80% and 90%. In effect, we try to make clinical importance and statistical significance agree, and thus reduce problems of interpretation.

The necessary sample size is usually obtained from complicated formulae or there are extensive tables available (Machin and Campbell, 1987), but it is much simpler to use a graphical method. Figure 15.2 shows a nomogram that can be used to calculate the appropriate sample size for all the situations considered in this chapter. It is simple to use and has the added

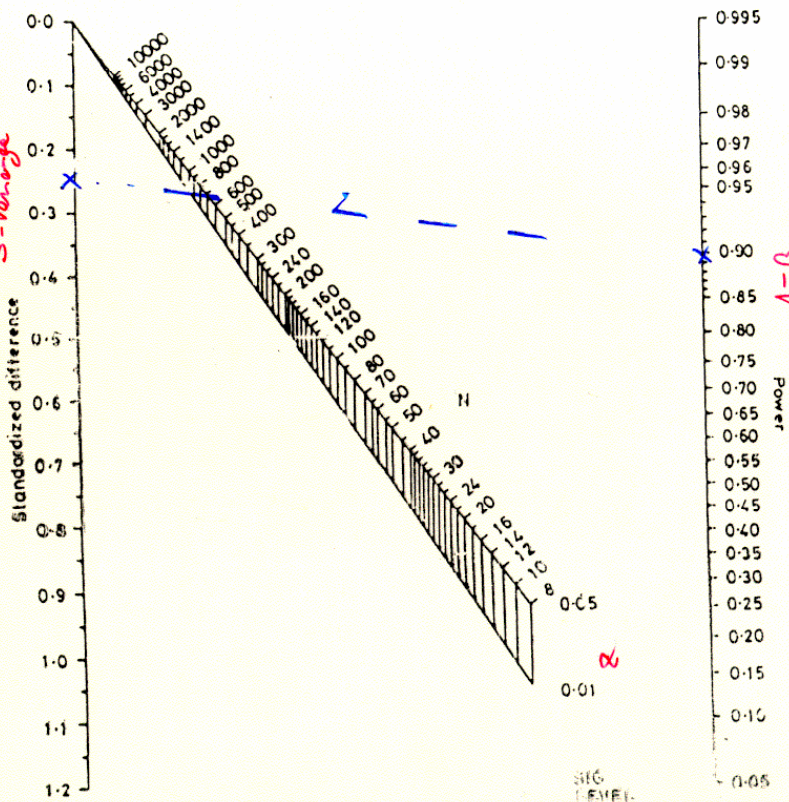


Figure 15.2 Nomogram for calculating sample size by power (adapted from Altman, 1982b, with permission)

146's Sample size 457

advantage of being equally easy to use in reverse for determining the power of a study of given sample size.

I shall first consider the case where we intend to have two groups of equal size. The nomogram can be used, however, for unequal sample sizes, as I shall show later. All of the sample size calculations are based on the quantity known as the **standardized difference**. This is calculated in a different way for continuous or categorical outcome variables, but in principle it is based in each case on the ratio of the difference of interest to the standard deviation of the observations. In other words, we express the difference of interest as a multiple of the standard deviation. As we would expect, the smaller this ratio is the larger the required size of the trial.

### (a) Continuous data - two independent groups

For studies of two independent groups of patients with a continuous outcome measure we need to specify the following quantities:

1. standard deviation of the variable (in each group) ( $s$ );
2. clinically relevant difference ( $\delta$ );
3. the significance level ( $\alpha$  - two-sided);
4. the power ( $1 - \beta$ );

and it is assumed that the variable has a Normal distribution in the population. The total sample size is  $N$ .

The standardized difference is calculated simply as the ratio of the difference of interest to the standard deviation, that is  $\delta/s$ . We can use Figure 15.2 to calculate the necessary sample size from the standardized difference for any desired power, choosing either a 5% or 1% level of significance.

For example, suppose that we are planning a milk-feeding trial in five-year-old children, to see if a daily supplement of milk for a year will lead to an increased gain in height compared with a control group. (Such a study would in fact be difficult to carry out, for practical and ethical reasons.) We know from published data that at this age children grow on average about 6 cm in a year, with a standard deviation of 2 cm. Suppose that the effect of the milk on height gain will be considered important if it is at least 0.5 cm. We want a high probability of detecting such a difference, so we set the power to be 0.9 (90%) and choose a 1% significance level. The standardized difference is  $0.5/2.0 = 0.25$ . We can now use Figure 15.2 to calculate the necessary sample size. We 'draw' a straight line from the value 0.25 on the scale for the standardized difference to the value 0.90 on the scale for power and read off the value for  $N$  on the line corresponding to  $\alpha = 0.01$ , which gives a total sample size of 900, i.e. 450 in each group.

There are several possible approaches if no estimate of the standard deviation is available. One way is to start the trial and use the data for the

# TEST DI SIGNIFICATIVITA'

**Parametrici**: richiedono assunzioni sul tipo di distribuzione della popolazione (Normale, Binomiale, t-Student, Fisher, ...)

se:           campioni poco numerosi e/o  
              forma della distribuzione sulla popolazione non certa  
=> EVITARLI!

Sono i più POTENTI e consentono, fissato  $\alpha$ , di minimizzare gli errori di II tipo

**Non Parametrici** ('distribution free'): richiedono solo l'ordinabilità della variabile (quindi vanno bene anche con variabili non numeriche purchè ordinabili).

Sono meno POTENTI/EFFICIENTI (ma non troppo); si basano sui RANGHI

Ok per basse numerosità, dati fortemente asimmetrici (sui quali si può eventualmente agire mediante trasformazione con funzioni –p.es.log), ....

- I TEST SI APPLICANO PER INFERIRE SULLA MEDIA, SU DIFFERENZE DI MEDIE, SU PROPORZIONI, SU UN CAMPIONE, SU PIU' CAMPIONI.
- DIETRO TUTTI I TEST C'E' L'IDEA DI UN MODELLO STATISTICO, OVVERO DI UN LEGAME/RELAZIONE ESISTENTE TRA 2 O PIU' VARIABILI, CHE GIUSTIFICHI I RISULTATI OSSERVATI SUL CAMPIONE

# VERIFICA DELLE IPOTESI

1 CAMPIONE vs 1 POPOLAZIONE DI RIFERIMENTO =>  $H_0$ : il campione appartiene alla popolazione

2 CAMPIONI (non serve conoscere la popolazione) =>  $H_0$ : i due campioni appartengono alla stessa popolazione

PIU' DI 2 CAMPIONI: considero coppie di campioni oppure =>  $H_0$ : tutti i campioni appartengono alla medesima popolazione; se  $H_0$  è falsa allora almeno 1 campione è 'estraneo' e posso individuarlo/i ripetendo il test su sottoinsiemi di campioni



## VERIFICA - TEST SU 1 CAMPIONE

Test sulla media: confronto il valore medio di un campione con quello **NOTO** della popolazione.

$H_0$ : il campione appartiene alla popolazione, ovvero l'eventuale differenza tra i valori medi  $\bar{x}$  e  $\mu$  è dovuta al caso e non è ***statisticamente significativa***

$H_1$ :  $\mu_0 > \mu$  (e/o  $\mu_0 < \mu$ , dipende dal caso in questione, 1 o 2 code)

### Z - Test (parametrico):

- Popolazione con distribuzione normale di note media  $\mu$  e  $\sigma$
- Campione di numerosità  $n$  di media  $\bar{x}$

Si calcola  $Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$  e si confronta con i valori di  $Z_\alpha$  di una Normale

Se  $Z > Z_\alpha$  rifiuterò l'ipotesi nulla (test ad 1 coda) con un livello di significatività pari ad  $\alpha$

ovvero (test a 2 code)

se  $Z > Z_{\alpha/2}$  oppure se  $Z < -Z_{\alpha/2}$ , allora rifiuterò l'ipotesi nulla con un livello di significatività pari ad  $\alpha$

ovvero concluderò con una probabilità pari a  $1 - \alpha$  che il campione non proviene dalla stessa popolazione che aveva media  $\mu$  e deviazione standard  $\sigma$

Se il valore di  $\sigma$  della popolazione **non è noto o la numerosità del campione è bassa (<100)** allora posso stimare  $\sigma$  attraverso la deviazione standard del campione  $s$  e, per tener conto dell'incertezza sulla stima di  $\sigma$ , si utilizza il

### T-Test o t-Student (parametrico):

- Popolazione con distribuzione normale di nota media  $\mu$
- Campione di numerosità  $n$  di media  $\bar{x}$  e deviazione standard  $s$

Si calcola  $t = \frac{\bar{x} - \mu}{s / \sqrt{n-1}}$  con  $s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$  (stima della varianza del campione)

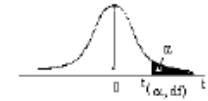
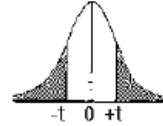
Si confronterà il valore di  $t$  con le tabelle della t-Student (a 1 o 2 code) con livello di significatività  $\alpha$  e  $n - 1$  gradi di libertà  $\Rightarrow t_{\alpha, n-1}$  ovvero  $t_{\alpha/2, n-1}$

# TABELLE t-Student

## Valori critici della distribuzione t di Student per un test unilaterale

(prima parte)

### Valori critici della distribuzione t di Student per un test bilaterale



Gradi di libertà	$\alpha$								
	0,500	0,400	0,200	0,100	0,050	0,025	0,010	0,005	0,001
1	1.000	1.376	3.078	6.314	12.706	25.457	63.657	14.089	31.598
2	.816	1.061	1.886	2.920	4.303	6.205	9.925	14.089	31.598
3	.765	0.978	1.638	2.353	3.182	4.176	5.841	7.453	12.941
4	.741	.941	1.533	2.132	2.776	3.495	4.604	5.598	8.610
5	.727	.920	1.476	2.015	2.571	3.163	4.032	4.773	6.859
6	.718	.906	1.440	1.943	2.447	2.969	3.707	4.317	5.959
7	.711	.896	1.415	1.895	2.365	2.841	3.499	4.029	5.405
8	.706	.889	1.397	1.860	2.306	2.752	3.355	3.832	5.041
9	.703	.883	1.383	1.833	2.262	2.685	3.250	3.690	4.781
10	.700	.879	1.372	1.812	2.228	2.634	3.169	3.581	4.587
11	.697	.876	1.363	1.796	2.201	2.593	3.106	3.497	4.437
12	.695	.873	1.356	1.782	2.179	2.560	3.055	3.428	4.318
13	.694	.870	1.350	1.771	2.160	2.533	3.012	3.372	4.221
14	.692	.868	1.345	1.761	2.145	2.510	2.977	3.326	4.140
15	.691	.866	1.341	1.753	2.131	2.490	2.947	3.286	4.073
16	.690	.865	1.337	1.746	2.120	2.473	2.921	3.252	4.015
17	.689	.863	1.333	1.740	2.110	2.458	2.898	3.222	3.965
18	.688	.862	.330	1.734	2.101	2.445	2.878	3.197	3.922
19	.688	.861	1.328	1.729	2.093	2.433	2.861	3.174	3.883
20	.687	.860	1.325	1.725	2.086	2.423	2.845	3.153	3.850
21	.686	.859	1.323	1.721	2.080	2.414	2.831	3.135	3.819
22	.686	.858	1.321	1.717	2.074	2.406	2.819	3.119	3.792
23	.685	.858	1.319	1.714	2.069	2.398	2.807	3.104	3.767
24	.685	.857	1.318	1.711	2.064	2.391	2.797	3.090	3.745
25	.684	.856	1.316	1.708	2.060	2.385	2.787	3.078	3.725
26	.684	.856	1.315	1.706	2.056	2.379	2.779	3.067	3.707
27	.684	.855	1.314	1.703	2.052	2.373	2.771	3.056	3.690
28	.683	.855	1.313	1.701	2.048	2.368	2.763	3.047	3.674
29	.683	.854	1.311	1.699	2.045	2.364	2.756	3.038	3.659
30	.683	.854	1.310	1.697	2.042	2.360	2.750	3.030	3.646
35	.682	.852	1.306	1.690	2.030	2.342	2.724	2.996	3.591
40	.681	.851	1.303	1.684	2.021	2.329	2.704	2.971	3.551
45	.680	.850	1.301	1.680	2.014	2.319	2.690	2.952	3.520
50	.680	.849	1.299	1.676	2.008	2.310	2.678	2.937	3.496
55	.679	.849	1.297	1.673	2.004	2.304	2.669	2.925	3.476
60	.679	.848	1.296	1.671	2.000	2.299	2.660	2.915	3.460
70	.678	.847	1.294	1.667	1.994	2.290	2.648	2.899	3.435
80	.678	.847	1.293	1.665	1.989	2.284	2.638	2.887	3.416
90	.678	.846	1.291	1.662	1.986	2.279	2.631	2.878	3.402
100	.677	.846	1.290	1.661	1.982	2.276	2.625	2.871	3.390
120	.677	.845	1.289	1.658	1.980	2.270	2.617	2.860	3.373
$\infty$	.6745	.841	1.28161	1.6448	1.9600	2.2414	2.5758	2.8070	3.2905

Gradi Di Libertà	Aree della coda superiore					
	0.25	0.10	0.05	0.025	0.01	0.005
1	1.0000	3.07	6.313	12.706	31.8207	63.6574
2	0.8163	1.88	2.920	4.302	6.9646	9.9248
3	0.7649	1.63	2.353	3.182	4.5407	5.8409
4	0.7407	1.53	2.131	2.776	3.7469	4.6041
5	0.7267	1.47	2.015	2.570	3.3649	4.0322
6	0.7176	1.43	1.943	2.447	3.1427	3.7074
7	0.7111	1.41	1.894	2.364	2.9980	3.4995
8	0.7064	1.39	1.859	2.306	2.8965	3.3554
9	0.7027	1.38	1.833	2.262	2.8214	3.2498
10	0.6998	1.37	1.812	2.228	2.7638	3.1693
11	0.6974	1.36	1.795	2.201	2.7181	3.1058
12	0.6955	1.35	1.782	2.178	2.6810	3.0545
13	0.6938	1.35	1.770	2.160	2.6503	3.0123
14	0.6924	1.34	1.761	2.144	2.6245	2.9768
15	0.6912	1.34	1.753	2.131	2.6025	2.9467
16	0.6901	1.33	1.745	2.119	2.5835	2.9208
17	0.6892	1.33	1.739	2.109	2.5669	2.8982
18	0.6884	1.33	1.734	2.100	2.5524	2.8784
19	0.6876	1.32	1.729	2.093	2.5395	2.8609
20	0.6870	1.32	1.724	2.084	2.5280	2.8453
21	0.6864	1.32	1.720	2.079	2.5177	2.8314
22	0.6858	1.32	1.717	2.073	2.5083	2.8188
23	0.6853	1.31	1.713	2.068	2.4999	2.8073
24	0.6848	1.31	1.710	2.063	2.4922	2.7969
25	0.6844	1.31	1.708	2.059	2.4851	2.7874
26	0.6840	1.31	1.705	2.055	2.4786	2.7787
27	0.6837	1.31	1.703	2.051	2.4727	2.7707
28	0.6834	1.31	1.701	2.048	2.4671	2.7633
29	0.6830	1.31	1.699	2.045	2.4620	2.7564
30	0.6828	1.31	1.697	2.042	2.4573	2.7500
31	0.6825	1.30	1.695	2.039	2.4528	2.7440
32	0.6822	1.30	1.693	2.036	2.4487	2.7385
33	0.6820	1.30	1.692	2.034	2.4448	2.7333
34	0.6818	1.30	1.690	2.032	2.4411	2.7284
35	0.6816	1.30	1.689	2.030	2.4377	2.7238
36	0.6814	1.30	1.688	2.028	2.4345	2.7195
37	0.6812	1.30	1.687	2.026	2.4314	2.7154
38	0.6810	1.30	1.686	2.024	2.4286	2.7116
39	0.6808	1.30	1.684	2.022	2.4258	2.7079
40	0.6807	1.30	1.683	2.021	2.4233	2.7045
41	0.6805	1.30	1.682	2.019	2.4208	2.7012
42	0.6804	1.30	1.682	2.018	2.4185	2.6981
43	0.6802	1.30	1.681	2.016	2.4163	2.6951
44	0.6801	1.30	1.680	2.015	2.4141	2.6923
45	0.6800	1.30	1.679	2.014	2.4121	2.6896
46	0.6799	1.30	1.678	2.012	2.4102	2.6870
47	0.6797	1.29	1.677	2.011	2.4083	2.6846
48	0.6796	1.29	1.677	2.010	2.4066	2.6822
49	0.6795	1.29	1.676	2.009	2.4049	2.6800
50	0.6794	1.29	1.675	2.008	2.4033	2.6778
51	0.6793	1.29	1.675	2.007	2.4017	2.6757
52	0.6792	1.29	1.674	2.006	2.4002	2.6737
53	0.6791	1.29	1.674	2.005	2.3988	2.6718
54	0.6791	1.29	1.673	2.004	2.3974	2.6700
55	0.6790	1.29	1.673	2.004	2.3961	2.6682
56	0.6789	1.29	1.672	2.003	2.3948	2.6665
57	0.6788	1.29	1.672	2.002	2.3936	2.6649
58	0.6787	1.29	1.671	2.001	2.3924	2.6633
59	0.6787	1.29	1.671	2.001	2.3912	2.6618
60	0.6786	1.29	1.670	2.000	2.3901	2.6603

**Esempio:** verificare se il tasso di glicemia, riscontrato nel sangue su un campione casuale di 26 individui, sia significativamente diverso da quello medio in soggetti normali ( $\mu=90\text{mg}/100\text{ml}$ ) con distribuzione normale.

Nel campione  $\bar{x} = 140\text{mg}/100\text{ml}$  e  $s = 52.5$

$H_0: \mu_0$  (della popolazione da cui proviene il campione) =  $\mu$

$H_1: \mu_0 > \mu$  (basta 1 coda in quanto  $\bar{x} > \mu$ )

Siccome  $\sigma$  è ignota si applicherà la t-Student. Si sceglie  $\alpha$  (p.es.=0.01), si calcola

$t_{\alpha, n-1} = t_{0.01, 25} = 2.48$  e poi

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n-1}} = 50 / (52.5/5) = 4.76 > \text{valore tabulato}$$

=> con probabilità pari all'1% di commettere errore, possiamo respingere  $H_0$   
Ovvero il campione proviene da una popolazione con media  $90\text{mg}/100\text{ml}$  e  
concludere che **PRESUMIBILMENTE** i soggetti del campione appartengono ad  
una popolazione con ridotta tolleranza glucidica.

NOTE:

- Se la distribuz. ha alta skewness, si può usare la mediana al posto della media
- Per  $n \Rightarrow \infty$  (ma basta  $>100$ ) la t-Student tende alla Z, altrimenti ci vuole la correzione che aumenta il coefficiente tenendo conto dell'ulteriore incertezza

Se i dati sono qualitativi/numerabili classificabili in 2 gruppi e la numerosità  $n$  è bassa, si utilizza un test binomiale

**Test Binomiale (parametrico):**

Si confronta  $p_{n,k}$  (probabilità dell'evento) con  $\alpha$ , dove:

$$p_{n,k} = \binom{n}{k} p^k \cdot q^{n-k}$$

$k$  = nr elementi con un certo carattere

$n$  = numerosità del campione

$$\binom{n}{k} = n! / (k! (n-k)!)$$

oppure si utilizza l'approssimazione Normale:

$$Z = \frac{(|k - np| - 1/2)}{\sqrt{np(1-p)}}$$

( $k$ =valore osservato -  $np$ =valore ipotizzato - 'continuity correction')/errore standard

e la si confronta con una Normale ( $Z_{\alpha/2}$ )

**Esempio:** dalla letteratura è noto che un certo trattamento di una data malattia ottiene risultati positivi nel 50% dei casi. Si vuole valutare se l'introduzione di un farmaco in quel trattamento comporta aumento dei casi positivi. Supponiamo di avere  $n=10$  e che su 8 si è avuto miglioramento. Vogliamo valutare se il risultato è significativamente diverso da quello noto.

$H_0$ : terapia nuova non ha effetto sulla malattia e quindi la differenza riscontrata è dovuta al caso.

Poiché la variabile è dicotomica (risponde/non risponde alla terapia) e il campione è piccolo ( $n=10$ ) si può usare la binomiale.

$p=q=50\%=0.5$  (i soggetti erano già in una terapia che dava risultati nel 50% dei casi)

Per verificare l' $H_0$  calcoliamo la probabilità  $p_{n,k} = \binom{n}{k} p^k \cdot q^{n-k}$  dove  $k$  può valere 8, 9 o 10 perché la prob di ottenere in un gruppo di 10 soggetti 8 risposte positive alla terapia solo per caso equivale a dire 'almeno' 8. La prob che il campione provenga da una popolazione in cui il 50% risponde positivamente vale la somma delle  $p_{n,k}$  per  $k=8,9,10$  che, con  $n=10$ , vale:  $0.055 = \text{prob di ottenere 8 risposte alla terapia soltanto per caso}$ . Scegliendo un livello  $\alpha$  pari al 5% allora bisognerà accettare l' $H_0$ ! Sostenendo che la terapia supplementare non migliora significativamente il quadro clinico (nella popolazione da cui proviene il campione)

**SIGN Test** (non parametrico): si ordinano le differenze tra le osservazioni ed il riferimento; N1 conta le differenze negative e N2 quelle positive. H0= Differenze non significative.

Si calcola  $\chi^2 = (N1-N2)^2 / (N1+N2)$  e si confronta con  $\chi^2_{N1+N2-1, \alpha}$

meglio:

**WILCOXON-Signed Rank Sum Test** (non parametrico): si usa quando oltre a considerare se ogni osservazione è sotto o sopra un certo valore di interesse, si vuole tener conto anche dell'ampiezza delle osservazioni.

1. Calcolo la differenza tra ciascuna osservazione e il valore d'interesse
2. Si ordinano ignorando il segno e si assegna il Rango a partire dal più piccolo. NOTA: se 2 o più Ranghi sono uguali, si prende quello intermedio
3. Si calcola la somma dei ranghi delle osservazioni che sono negative rispetto al valore ipotetico quella delle osservazioni positive;

4. Si confronta l'una o l'altra somma con l'opportuna tabella di WILCOXON

(per  $n \leq 25$ ), altrimenti si procede con la normale, assumendo  $\mu = \frac{n \cdot (n+1)}{4}$

e  $\sigma^2 = \frac{n \cdot (n+1) \cdot (2n+1)}{24}$  e confrontando con la  $Z_{\alpha/2}$ .

## TEST SULLA FREQUENZA

Si utilizzano quando le frequenze osservate di una modalità (valore) di un carattere (variabile), su un campione, differiscono (in modo signif. statistico) da quelle teoriche (attese o note a priori). Si parla anche di conteggi e tavole di contingenza.

**Test  $\chi^2$  (chi quadrato)** (test non parametrico, in genere la numerosità è elevata)

*ESEMPIO: Si ha un campione di 800 famiglie con 2 figli, aventi distribuzione rispetto al sesso:*

MM	MF (o FM)	FF
160	440( $n_i^2$ )	200

Poiché la probabilità che nasca un Maschio o una Femmina è 1/2 anche dopo il primo figlio (eventi indipendenti), la distribuzione teorica dei sessi dovrebbe essere:

MM	MF (o FM)	FF
200	400 ( $n_i^{2*}$ )	200

*$H_0$ : i dati del campione seguono la distribuzione teorica e le diversità sono imputabili solo al caso.*

Il test  $\chi^2$  è adatto per confrontare frequenze empiriche ( $n_i^2$ ) con frequenze teoriche  $n_i^{2*}$

$$\chi^2 = \sum_{i=1}^n \frac{(n_i^2 - n_i^{2*})^2}{n_i^{2*}}$$

Il risultato si confronterà con il  $\chi_{n-1, \alpha}^2$  essendo  $\alpha$  il livello di significatività.



## TEST SULLA FREQUENZA

*Nell'esempio il test sarà bilaterale perché non c'è motivo di supporre che le diversità rispetto ai valori teorici abbiano una propensione verso un segno piuttosto che l'altro.*

$$\chi^2 = \frac{(160-200)^2}{200} + \frac{(440-400)^2}{400} + \frac{(200-200)^2}{200} = 12, \text{ mentre} \quad \chi_{2,0.01}^2 = 7.38$$

*quindi l'ipotesi nulla va rifiutata.*

### Test di Kolmogorov (non parametrico):

indicato per  $n$  piccoli e se i caratteri sono continui.

Si propone di verificare l'ipotesi che un campione provenga da un dato universo continuo con distribuzione nota (verifica la forma della distribuzione).

Si calcola:  $D_{calc} = \max_i |F_i^* - F_i|$  dove  $F_i^*$  è la distribuzione teorica delle frequenze cumulate e  $F_i$  quella empirica delle freq cumulate => si valuta quanto le due distribuzioni siano 'simili' mediante una misura di distanza (il max su tutti i valori considerati)

Dato  $\alpha$  ed  $n$ , dalle apposite tabelle si ricava il  $D_{teorico}$  e lo si confronta con il precedente.

Se  $D_{calc} > D_{teorico}$  si rifiuta l'ipotesi nulla.

## TEST SU DUE CAMPIONI

### Test sulla differenza tra le medie

Servono per stabilire con quale probabilità i 2 campioni (con medie diverse) provengono da popolazioni con la stessa media.

Si estendono tutti i test precedenti e anziché parlare di una sola media, si tratteranno DIFFERENZE DI MEDIE.

### Z-test a 2 campioni (parametrico):

$H_0$  = i 2 campioni appartengono alla stessa popolazione.

La variabile differenza di due distribuzioni campionarie ha:

$$\text{media} = \mu_1 - \mu_2 \text{ e varianza} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

Si calcola 
$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$

Se le due popolazioni sono le stesse, si avrà  $\mu_1 = \mu_2$  e  $\sigma_1 = \sigma_2$  per cui la formula applicata sarà:

$$Z = \frac{(\bar{x}_1 - \bar{x}_2)}{\sigma \cdot \sqrt{1/n_1 + 1/n_2}}$$

La Z si confronterà con la  $Z_{\alpha/2}$

### t-Test (t-Student) a 2 campioni (test parametrico):

Se i valori di  $\sigma$  non sono noti allora: se si ipotizza che tra i 2 campioni le varianze siano diverse, si dovranno utilizzare test complessi (test di Behrens-Fisher); altrimenti se si ipotizzano uguali, (attraverso il test di Fisher si verificherà quanto questa supposizione sia valida) si valuta:

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{s \cdot \sqrt{1/n_1 + 1/n_2}}$$

con

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)} = \frac{\sum(x_{i_1} - \bar{x}_1)^2 + \sum(x_{i_2} - \bar{x}_2)^2}{n_1 + n_2 - 2}$$

*'pooled variance'*, varianza ponderata

Con  $s_1$  e  $s_2$  calcolate sui due campioni. Si confronta con la  $t_{\alpha, n_1+n_2-2}$ .  
Se  $t$  è  $< t_{\alpha, n_1+n_2-2}$ , prima di accettare  $H_0$  devo verificare l'omogeneità delle varianze mediante il test di Fisher.

**Esempio:**

*Si confrontano le medie di 2 campioni casuali, indipendenti estratti da due popolazioni di forma Normale, relativamente ai giorni di degenza degli operati per un motivo xx in due diverse strutture ospedaliere (non possiedo  $\sigma$  della popolazione)*

*Ospedale A:  $n_1=16$ ,  $\bar{x}_1=13$ ,  $s_1=4$*

*Ospedale B:  $n_2=16$ ,  $\bar{x}_2=9$ ,  $s_2=6$*

*Prima di concludere che  $\bar{x}_2$  è significativamente  $< \bar{x}_1$  (e la diff non sia dovuta al caso) eseguo il test*

*H0: le due medie provengono dalla medesima popolazione e quindi anche le due dev standard devono essere comparabili*

*Stimo  $\sigma$  a partire dalla  $s$  ponderata e ricavo  $t \Rightarrow 2.22$  che confronto con  $t_{0.01, 30} = 2.46$  e ..... concludo che H0 è vera!*

*Con prob. 1% di commettere un errore di tipo I potremo accettare l'hp che non esiste differenza significativa nella degenza media relativa all'intervento xx nei 2 ospedali.*

A questo punto eseguo il test di Fisher per verificare l'omogeneità delle varianze

## Test di Fisher

Utile per verificare se ( $H_0$ ) due varianze possano considerarsi provenire da una medesima popolazione, con un certo grado  $\alpha$  di significatività

Dati  $n_1, n_2$  (dei 2 campioni) e  $s_1^2, s_2^2$  (con  $s_1^2 > s_2^2$ ) si confronta se:

$$F = \frac{s_1^2}{s_2^2} < F_{\alpha, n_1-1, n_2-1}$$

Nell'eventualità che questo si verifichi, si accetterà l'ipotesi  $H_0$  e si confermerà il risultato della t-Student che i due campioni provengono dalla medesima popolazione

In caso contrario **NULLA** si può concludere e bisogna utilizzare altri test, p.es. quello di Welch (che stima diversamente la varianza complessiva ed il nr di gradi di libertà)

## TABELLA di Fisher per $\alpha=0.01$

### Valori critici della distribuzione F di Fisher-Snedecor

I gradi di libertà del numeratore (o varianza maggiore) sono riportati in orizzontale (prima riga)

I gradi di libertà del denominatore (o varianza minore) sono riportati in verticale (prima colonna)

$$\alpha = 0.01$$

NUMERATORE

DEN.	1	2	3	4	5	6	7	8	12	24	$\infty$
1	4052	5000	5403	5625	5764	5859	5928	5981	6106	6235	6366
2	98,50	99,00	99,17	99,25	99,30	99,33	99,36	99,37	99,41	99,46	99,50
3	34,12	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,05	26,60	26,13
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,37	13,93	13,46
5	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	9,89	9,47	9,02
6	13,75	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,72	7,31	6,88
7	12,25	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,47	6,07	5,65
8	11,26	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,67	5,28	4,86
9	10,56	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,11	4,73	4,31
10	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,71	4,33	3,91
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,16	3,78	3,36
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	3,80	3,43	3,00
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,55	3,18	2,75
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,37	3,00	2,57
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,23	2,86	2,42
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	2,84	2,47	2,01
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,66	2,29	1,80
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,50	2,12	1,60
120	6,85	4,79	3,95	3,48	3,17	2,96	2,79	2,66	2,34	1,95	1,38
$\infty$	6,63	4,61	3,78	3,32	3,02	2,80	2,64	2,51	2,18	1,79	1,00

Se i campioni **NON sono indipendenti**, p.es. misure ripetute sugli stessi soggetti in due tempi diversi, si può utilizzare la

**t-Student per campioni 'accoppiati' (paired)**

che parte dalle differenze tra le misure nei medesimi soggetti nei due momenti =  $D_i$

Si calcolano la media  $\bar{D}$  e la  $\sigma$  delle differenze  $D_i$  e quindi si stima la

$$t = \frac{\bar{D}}{\sqrt{\sigma^2/n}}$$

che confronterò con  $t_{\alpha, n-1}$

In pratica valuto quanto  $\bar{D}$  si discosta dallo zero....

## Test di Wilcoxon-Mann-Whitney o U-test (non parametrico) (*Wilcoxon Rank Sum test*)

$H_0$  = i 2 campioni (indipendenti) appartengono alla stessa popolazione.

Si ordinano le osservazioni dei 2 gruppi come se fossero di uno unico e si associa il rango (*se ci sono troppe osservazioni uguali ci sono correzioni da apportare*)

Si calcola il rango complessivo per ciascun gruppo ( $R_1$  e  $R_2$ ) e si valutano:

$$U = n_1 n_2 + \frac{1}{2} n_1 (n_1 + 1) - R_1$$
$$U' = n_1 n_2 + \frac{1}{2} n_2 (n_2 + 1) - R_2$$

Mediante apposita tabella si entra con  $\min(U, U')$  e si confronta il valore relativo con un  $\alpha$  prefissato: se il valore è maggiore di  $\alpha$ , si accetterà  $H_0$ .

Per  $n_1, n_2 > 10$ , il valore di  $R_1$  (o  $R_2$ , il minimo dei due) =  $T$  tende ad una Normale con media e SD:

$$\mu_T = \frac{n_s(n_s + n_l + 1)}{2} \quad \text{e} \quad \sigma_T = \sqrt{\frac{n_l * \mu_T}{6}}$$

dove  $n_s$  è il campione meno numeroso e  $n_l$  il più numeroso tra i 2.

A questo punto si utilizza la statistica  $Z = \frac{T - \mu_T}{\sigma_T}$  e si confronta con la  $Z_{\alpha/2}$



Se i due campioni sono appaiati si utilizza il (equivalente al t-Student appaiato)

## Test di Wilcoxon per dati appaiati (non parametrico)

*(Wilcoxon Signed Rank test)*

$H_0$  = i 2 campioni (accoppiati) appartengono alla stessa popolazione.

Si ordinano le **differenze** (in valore assoluto) dei 2 gruppi e si associa il rango

Si calcola il rango (p.es.) delle differenze positive (**T**) e si valutano (per  $n > 20$ ):

$$\mu_T = \frac{n_1(n_1+1)}{4} \quad \text{e} \quad \sigma_T = \sqrt{\frac{(2n_1+1) * \mu_T}{6}}$$

dove  $n_1$  è la numerosità del singolo campione.

A questo punto si utilizza la statistica  $Z = \frac{T - \mu_T}{\sigma_T}$  e si confronta con la  $Z_{\alpha/2}$ .

Se  $n < 20$  si utilizza apposita tabella entrando col parametro  $W$  che tiene conto della somma dei quadrati dei ranghi:

$$W = \frac{T}{\sqrt{\sum R^2}}$$