

# **Statistica inferenziale**

## **Test di ipotesi**

Bari, 10 Dicembre 2007

## **Test per il confronto di due medie**

In questo test vogliamo confrontare i valori  $\mu_X$  e  $\mu_Y$  delle medie di due v.a.  $X$  e  $Y$  per stabilire se esse sono uguali o diverse. Distingueremo i casi di campioni accoppiati e campioni indipendenti.

## **Campione accoppiato**

Un campione accoppiato é costituito da un insieme di coppie di osservazioni

$$(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$$

delle v.a.  $X$  e  $Y$ , in cui ciascuna coppia  $(X_k, Y_k)$  si riferisce alla stessa unità statistica.

Ad esempio, per capire l'effetto di un farmaco, misuriamo una grandezza su  $n$  soggetti prima ( $X$ ) e dopo ( $Y$ ) della somministrazione del farmaco e confrontiamo  $\mu_X$  e  $\mu_Y$ . Quindi  $(X_k, Y_k)$  sono le misure eseguite sul soggetto  $k$ -esimo prima e dopo la somministrazione del farmaco.

## **Campioni indipendenti**

Parleremo di campioni indipendenti quando non vi é nessuna relazione tra gli elementi  $X_1, X_2, \dots, X_n$  del primo campione e gli elementi  $Y_1, Y_2, \dots, Y_m$  del secondo campione. Quindi  $X_k$  ed  $Y_k$  non sono misurati sullo stesso individuo.

Osserva che  $n$  ed  $m$  possono essere diversi.

Un esempio di campioni indipendenti si ha quando si vuole confrontare l'effetto di un farmaco con quello di un placebo. A due gruppi di pazienti affetti dalla stessa patologia somministro ad uno il farmaco e all'altro il placebo, e poi confronto i valori medi di una grandezza misurata nei due gruppi per valutare l'effetto del farmaco.

## **Ipotesi**

Nel confronto di due medie avremo un test bilaterale se le ipotesi sono:

$$H_0 : \mu_X = \mu_Y, \quad H_1 : \mu_X \neq \mu_Y,$$

ed un test unilaterale destro se le ipotesi sono:

$$H_0 : \mu_X \leq \mu_Y, \quad H_1 : \mu_X > \mu_Y.$$

Un test unilaterale sinistro si ottiene scambiando il ruolo di  $X$  ed  $Y$ .

## **Confronto nel caso di campioni accoppiati**

In questo caso costruiamo il campione  $Z_1, Z_2, \dots, Z_n$ , dove  $Z_k = X_k - Y_k$  e poi verifichiamo se la media  $\mu_Z$  della v.a.  $Z = X - Y$  si può considerare nulla o diversa da zero.

Quindi il confronto tra due medie nel caso di campioni accoppiati si riduce al test sulla media con  $\mu_0 = 0$ .

## **Confronto nel caso di campioni accoppiati ( $\sigma^2$ nota)**

Se le v.a.  $Z_k$  sono  $N(\mu, \sigma^2)$  e la varianza  $\sigma^2$  é nota allora

$$\bar{Z} = \frac{1}{n} \sum_{k=1}^n Z_k$$

é  $N(0, \frac{\sigma^2}{n})$  nell'ipotesi  $H_0$  vera, e quindi la v.a.

$$U_0 = \frac{\bar{Z}}{\sigma/\sqrt{n}}$$

é  $N(0, 1)$ .

## **Confronto nel caso di campioni accoppiati ( $\sigma^2$ nota)**

Allora, per un dato livello  $\alpha$ , la regione critica di un test bilaterale é

$$D = \{|U_0| > \varphi_{1-\frac{\alpha}{2}}\}$$

con significativit   $\alpha_s = 2[1 - \Phi(|U_0|)]$ , mentre la regione critica di un test unilaterale  

$$D = \{U_0 > \varphi_{1-\alpha}\}$$

con significativit   $\alpha_s = 1 - \Phi(U_0)$ .



## **Confronto nel caso di campioni accoppiati ( $\sigma^2$ non nota)**

Se le v.a.  $Z_k$  sono  $N(\mu, \sigma^2)$  e la varianza  $\sigma^2$  non é nota allora stimiamo  $\sigma^2$  utilizzando la varianza corretta

$$S^2 = \frac{1}{n-1} \sum_{k=1}^n (Z_k - \bar{Z})^2,$$

ed osserviamo che, nell'ipotesi  $H_0$  vera, la v.a.

$$T_0 = \frac{\bar{Z}}{S/\sqrt{n}}$$

é Student con  $n - 1$  gradi di libertà  $t(n - 1)$ .

## **Confronto nel caso di campioni accoppiati ( $\sigma^2$ non nota)**

Allora, per un dato livello  $\alpha$ , la regione critica di un test bilaterale é

$$D = \{|T_0| > t_{1-\frac{\alpha}{2}}(n-1)\}$$

con significativit   $\alpha_s = 2[1 - F_{n-1}(|T_0|)]$ , mentre la regione critica di un test unilaterale é

$$D = \{T_0 > t_{1-\alpha}(n-1)\}$$

con significativit   $\alpha_s = 1 - F_{n-1}(T_0)$ .

## Esempio

In tabella riportiamo il numero di pulsazioni al minuto misurate in  $n = 30$  pazienti prima ( $X$ ) e dopo ( $Y$ ) l'assunzione di un farmaco.

X	Y	Z	X	Y	Z	X	Y	Z
80	85	5	70	82	12	78	70	-8
80	84	4	65	73	8	75	77	2
82	87	5	83	89	6	76	76	0
75	81	6	74	85	11	78	82	4
80	79	-1	81	86	5	77	83	6
74	85	11	68	72	4	75	80	5
80	87	7	69	74	5	72	80	8
72	78	6	71	77	6	71	81	10
91	86	-5	70	75	5	75	76	1
88	80	-8	73	81	8	78	77	-1

Possiamo affermare, ad un livello  $\alpha = 0.05$ , che la somministrazione del farmaco ha provocato un aumento del numero di pulsazioni?

## **Esempio (cont.)**

Si tratta di un campione accoppiato perché le misure  $(X_k, Y_k)$  sono relative allo stesso soggetto. Le ipotesi sono:

$$H_0 : \mu_Y \leq \mu_X, \quad H_1 : \mu_Y > \mu_X.$$

Definendo  $Z = Y - X$ , le ipotesi diventano:

$$H_0 : \mu_Z \leq 0, \quad H_1 : \mu_Z > 0.$$

Per il campione  $Z_1, \dots, Z_n$  calcoliamo:

$$\bar{Z} = 4.23, \quad S = 5.01, \quad T_0 = \frac{\bar{Z}}{S/\sqrt{n}} = 4.63.$$

## **Esempio (cont.)**

Ora dobbiamo confrontare  $T_0$  con il quantile  $t_{0.95}(29) = 1.699$ . Poiché  $T_0 > 1.699$ , allora  $T_0$  cade nella regione critica. Quindi rigettiamo  $H_0$  e concludiamo che il farmaco ha prodotto un aumento della frequenza delle pulsazioni. La significatività statistica (p-value) del test é:  $\alpha_s = 0.00003$ .

Calcoliamo la potenza del test di livello  $\alpha = 0.05$  nell'ipotesi alternativa che  $\mu_Z = 2$ . Ossia vogliamo calcolare la probabilità di rigettare  $H_0$ , e quindi di accettare  $H_1$ , quando  $H_1$  é vera:

$$\pi(2) = \mathbb{P}\{D\} \qquad H_1 : \mu_Z = 2,$$

dove  $D = \{T_0 > 1.699\}$ .

**Esempio (cont.)**

Osserva che se  $H_1$  é vera allora la v.a.  $T_1 = \frac{\bar{Z} - \mu_Z}{S/\sqrt{n}}$  é una Student  $t(n - 1)$ . Calcoliamo:

$$\pi(2) = \mathbb{P}\{D\} = \mathbb{P}\{T_0 > 1.699\} = \mathbb{P}\left\{\frac{\bar{Z}}{S/\sqrt{n}} > 1.699\right\} =$$

$$\mathbb{P}\left\{\frac{\bar{Z}}{S/\sqrt{n}} - \frac{\mu_Z}{S/\sqrt{n}} > 1.699 - \frac{\mu_Z}{S/\sqrt{n}}\right\} = \mathbb{P}\left\{\frac{\bar{Z} - \mu_Z}{S/\sqrt{n}} > 1.699 - \frac{\mu_Z}{S/\sqrt{n}}\right\} =$$

$$\mathbb{P}\left\{T_1 > 1.699 - \frac{2}{5.01/\sqrt{30}}\right\} = \mathbb{P}\{T_1 > -0.49\} = 0.686.$$

Quindi, se la media  $\mu_z$  vale 2, il test lo rileva nel 68.6% dei casi.

## Esempio (cont.)

Per aumentare la potenza di un test di livello  $\alpha$  dobbiamo aumentare il numero  $n$  delle osservazioni.

Vogliamo determinare il numero  $n$  delle osservazioni tale che, nelle stesse condizioni sperimentali, otteniamo  $\pi(\mu_Z) = 0.95$ :

$$\begin{aligned} 0.95 = \pi(\mu_Z) &= \mathbb{P}\{T_0 > t_{1-\alpha}(n-1)\} = \mathbb{P}\left\{\frac{\bar{Z}}{S/\sqrt{n}} > t_{1-\alpha}(n-1)\right\} = \\ \mathbb{P}\left\{\frac{\bar{Z}}{S/\sqrt{n}} - \frac{\mu_Z}{S/\sqrt{n}} > t_{1-\alpha}(n-1) - \frac{\mu_Z}{S/\sqrt{n}}\right\} &= \mathbb{P}\left\{T_1 > t_{1-\alpha}(n-1) - \frac{\mu_Z}{S/\sqrt{n}}\right\} = \\ 1 - \mathbb{P}\left\{T_1 \leq t_{1-\alpha}(n-1) - \frac{\mu_Z}{S/\sqrt{n}}\right\}. \end{aligned}$$

Quindi:

## Esempio (cont.)

$$\mathbb{P} \left\{ T_1 \leq t_{0.95}(n-1) - \frac{\mu_Z}{S/\sqrt{n}} \right\} = 0.05$$

Per  $n$  grande, sappiamo che la distribuzione di  $T_1$  può essere approssimata da  $N(0, 1)$  e quindi:

$$\varphi_{0.05} \approx t_{0.95}(n-1) - \frac{\mu_Z}{S/\sqrt{n}}.$$

Inoltre  $t_{0.95}(n-1) \approx \varphi_{0.95}$ . Sapendo che  $\varphi_{0.05} = -\varphi_{0.95}$ :

$$\varphi_{0.05} \approx \varphi_{0.95} - \frac{\mu_Z}{S/\sqrt{n}} \Rightarrow \frac{\mu_Z}{S/\sqrt{n}} \approx 2\varphi_{0.95} \Rightarrow \sqrt{n} \approx \frac{2S\varphi_{0.95}}{\mu_Z}.$$

Per  $S = 5.01$ ,  $\mu_Z = 2$ ,  $\varphi_{0.95} = 1.64$  si ha:  $n \approx 68$ . Quindi devo avere 68 osservazioni per garantire una potenza del 95% ad un test di livello 0.05.



## Confronto nel caso di campioni indipendenti

Siano  $(X_1, \dots, X_n)$  e  $(Y_1, \dots, Y_m)$  due campioni indipendenti di dimensioni  $n$  ed  $m$ , osservazioni delle v.a.  $X : N(\mu_X, \sigma_X^2)$  e  $Y : N(\mu_Y, \sigma_Y^2)$  rispettivamente, e vogliamo verificare se hanno la stessa media.

A tale scopo, calcoliamo separatamente le medie dei due campioni:

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

$$\bar{Y} = \frac{1}{m} \sum_{j=1}^m Y_j.$$

Distinguiamo due casi.

## **Confronto nel caso di campioni indipendenti (varianze note)**

Nel caso in cui le varianze  $\sigma_X^2$  e  $\sigma_Y^2$  sono note, allora:

$$\bar{X} : N \left( \mu_X, \frac{\sigma_X^2}{n} \right), \quad \bar{Y} : N \left( \mu_Y, \frac{\sigma_Y^2}{m} \right).$$

Quindi:

$$\bar{X} - \bar{Y} : N \left( \mu_X - \mu_Y, \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m} \right).$$

Segue che:

$$U = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} : N(0, 1).$$

## **Confronto nel caso di campioni indipendenti (varianze note)**

Sotto  $H_0$  vera ( $\mu_X = \mu_Y$ ), allora la v.a  $U$  diventa:

$$U_0 = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}}.$$

Quindi, per un test bilaterale di livello  $\alpha$  la regione critica é:

$$D = \{|U_0| > \varphi_{1-\frac{\alpha}{2}}\},$$

e per un test unilaterale la regione critica é:

$$D = \{U_0 > \varphi_{1-\alpha}\}.$$

## **Confronto nel caso di campioni indipendenti (varianze non note)**

Nel caso in cui le varianze  $\sigma_X^2$  e  $\sigma_Y^2$  non sono note, si introduce la varianza combinata:

$$V^2 = \frac{1}{n + m - 2} \left[ \sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{j=1}^m (Y_j - \bar{Y})^2 \right] = \frac{(n - 1)S_X^2 + (m - 1)S_Y^2}{n + m - 2}$$

Nota che  $V^2$  é la media pesata delle stime delle varianze  $S_X^2$  e  $S_Y^2$  con pesi  $n - 1$  e  $m - 1$  rispettivamente.

## **Confronto nel caso di campioni indipendenti (varianze non note)**

Si può dimostrare che se  $\sigma_X^2 = \sigma_Y^2$ , allora  $V^2$  é uno stimatore non distorto della varianza comune e la variabile:

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{V \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

é una Student  $t(n + m - 2)$ .

## **Confronto nel caso di campioni indipendenti (varianze non note)**

Sotto  $H_0$  vera ( $\mu_X = \mu_Y$ ), allora la v.a  $T$  diventa:

$$T_0 = \frac{\bar{X} - \bar{Y}}{V \sqrt{\frac{1}{n} + \frac{1}{m}}}.$$

Quindi, per un test bilaterale di livello  $\alpha$  la regione critica é:

$$D = \{|T_0| > t_{1-\frac{\alpha}{2}}(n + m - 2)\},$$

e per un test unilaterale la regione critica é:

$$D = \{T_0 > t_{1-\alpha}(n + m - 2)\}.$$

## Esempio

I dati in tabella si riferiscono al livello di acido urico sierico misurato in  $n = 10$  soggetti sani ( $X$ ) ed in  $m = 15$  soggetti con sindrome di Down ( $Y$ ).

$X$	9.2	8.3	10.3	11.0	12.0	8.6	9.3	10.3	9.7	9.0
$Y$	10.9	11.3	10.9	10.2	9.3	10.4	10.5	11.3	10.8	12.0
	10.6	10.4	12.2	10.9	10.4					

Si vuole stabilire se esiste una differenza di questa grandezza nelle due classi di soggetti analizzati.

Questo é un caso di campioni indipendenti. Supponiamo di sapere che  $X : N(\mu_X, \sigma_X^2)$  e  $Y : N(\mu_Y, \sigma_Y^2)$ , e vogliamo stabilire se a livello  $\alpha = 0.02$ , possiamo affermare che  $\mu_X \neq \mu_Y$ .

## Esempio (cont.)

Il test da effettuare é di tipo bilaterale in cui vogliamo confrontare l'ipotesi  $H_1 : \mu_X \neq \mu_Y$  contro l'ipotesi  $H_0 : \mu_X = \mu_Y$ . Dalla tabella si ha:  $\bar{X} = 9.8$ ,  $\bar{Y} = 10.8$ . Nell'ipotesi in cui le standard deviation siano note con valori  $\sigma_X = \sigma_Y = 1.1$ , allora dobbiamo confrontare il valore della statistica  $U_0 = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} = -2.23$  con il quantile  $\varphi_{0.99} = 2.33$ .

Poiché  $|U_0| \not\geq 2.33$  allora non possiamo rigettare l'ipotesi nulla e quindi dobbiamo concludere che non c'è differenza tra i due campioni. In questo caso la significatività del test (p-value) é  $\alpha_s = 2[1 - \Phi(|U_0|)] = 0.03 > \alpha$ .



## **Esempio (cont.)**

Se invece non supponiamo note le standard deviation di  $X$  e  $Y$  allora dobbiamo ricorrere alla statistica  $T$ . A tale scopo dalla tabella otteniamo:

$$S_X^2 = 1.30, \quad S_Y^2 = 0.51, \quad V^2 = 0.82.$$

Allora dobbiamo confrontare il valore  $T_0 = -2.71$  con il quantile  $t_{0.99}(23) = 2.49987$ . Poiché  $|T_0| > 2.49987$  allora possiamo rigettare l'ipotesi  $H_0$  in favore dell'ipotesi  $H_1$ . In questo caso la significatività del test (p-value) é  $\alpha_s = 2[1 - F_{23}(|T_0|)] = 0.01 < \alpha$ .

## Multiple hypothesis testing

Quando un gran numero di ipotesi vengono testate simultaneamente gli effetti del caso possono essere critici.

Per illustrare il problema consideriamo un esperimento caso-controllo di microarray. Consideriamo il campione costituito da  $\ell = 47$  profili di espressione genica di colon, ottenuti con U133A Affymetrix chip, che riporta il livello di espressione genica di  $n = 22283$  probe;  $d = 25$  sono tumorali e  $h = 22$  sani.

## Esempio

Organizziamo i dati in una matrice  $X$  costituita da  $n$  righe e da  $\ell$  colonne in cui l'elemento  $x_{ij}$  rappresenta il livello di espressione del gene  $i$  nel soggetto  $j$ . Assumendo i livelli normalmente distribuiti, utilizziamo un t-test per determinare i geni differenzialmente espressi (DE) nelle due condizioni sperimentali. Fissato il livello  $\alpha = 0.01$ , allora per ogni  $i = 1, 2, \dots, n$  dobbiamo decidere tra:

$$H_0 : \mu_D^i = \mu_H^i, \qquad H_1 : \mu_D^i \neq \mu_H^i$$

## Esempio (Cont.)

Eseguiamo i test e troviamo 277 probes DE, i primi 10 sono:

<i>Probe</i>	<i>p - value</i>
201829_at	0.000004
210170_at	0.000029
221626_at	0.000032
203217_s_at	0.000064
215321_at	0.000076
206557_at	0.000098
206501_x_at	0.000113
209792_s_at	0.000149
220264_s_at	0.000177
219482_at	0.000210

Domanda: qual é il numero medio di ipotesi nulle che rigettiamo a livello  $\alpha = 0.01$  nell'ipotesi in cui  $H_0$  é vera per ogni  $i = 1, 2, \dots, n$ ?

## Esempio (Cont.)

Osserva che per definizione di p-value, sotto  $H_0$  vera, si ha:

$$\mathbb{P}\{P \leq \alpha\} = \alpha \quad \text{per ogni } \alpha$$

in quanto noi rigettiamo  $H_0$  se  $P \leq \alpha$  ed  $\alpha$  é la probabilità di rigettare  $H_0$  quando questa é vera. Poiché per definizione  $F(x) = \mathbb{P}\{P \leq x\} = x$  allora  $P$  é uniformemente distribuita in  $[0, 1]$ .

Risposta: il numero medio di ipotesi nulle che rigettiamo a livello  $\alpha$  quando tutte le  $H_0$  sono vere é  $n\alpha$ . Nel nostro esempio é 222.

## Esempio (Cont.)

L'istogramma conferma che i dati non sono associati con il fenotipo di interesse. In questo caso i dati erano casuali!!!

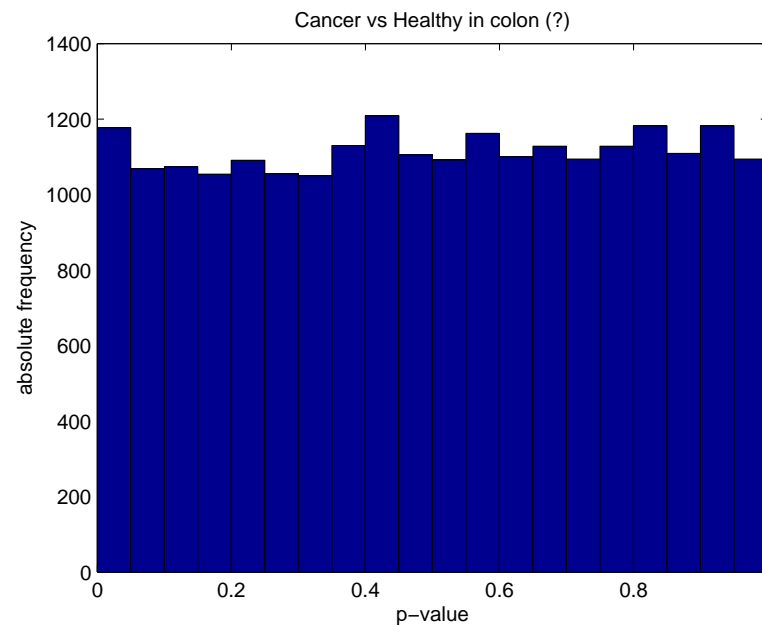


Figure 1: Istogramma dei p-value relativi al nostro esempio.

## Esempio (Cont.)

Nel caso reale abbiamo 1207 geni che superano il test a livello  $\alpha = 0.01$ .

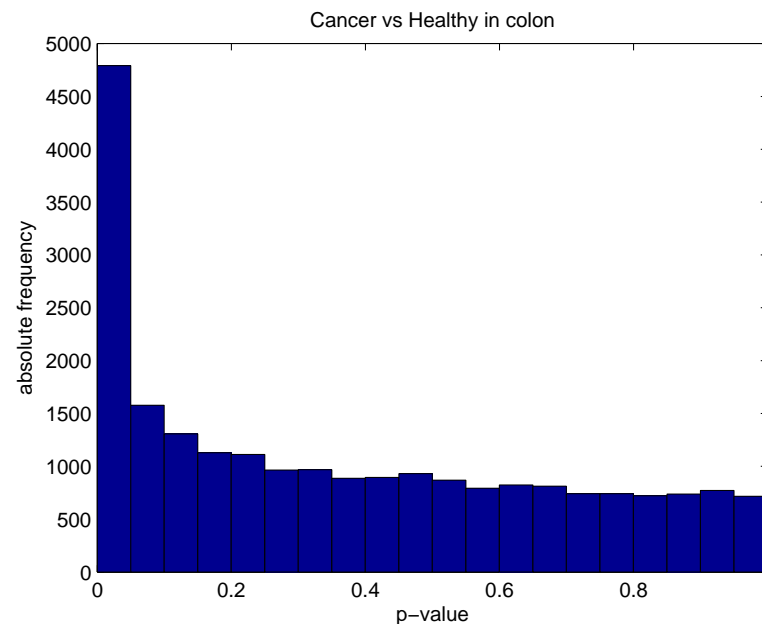


Figure 2: Istogramma dei p-value relativi al caso reale.

## Risultati possibili

Quando applichiamo una soglia di significatività nella verifica di ipotesi multiple, si verificano i seguenti risultati:

	Chiamate significative	Chiamate non significative	Totale
Nulle vere	$F$	$n_0 - F$	$n_0$
Alternative vere	$T$	$n_1 - T$	$n_1$
Totale	$S$	$n - S$	$n$

$S$  é il numero di feature chiamate significative. Di queste,  $F$  sono erroneamente chiamate significative (false positive), mentre  $T$  sono correttamente chiamate significative (true positive). Inoltre,  $n_0$  é il numero di features veramente nulle, mentre  $n_1 = n - n_0$  quelle veramente alternative.



## Procedure di multiple testing

Queste quantità possono essere usate per formare una misura di errore complessiva per ogni livello  $\alpha$ , che fornisce una stima della significatività statistica dei risultati ottenuti.

Essendo  $F$ ,  $T$  ed  $S$  variabili aleatorie, le misure di errore complessivo si scrivono in termini di valore atteso.

## False positive rate

Una misura di errore é il tasso di falsi positivi. In questo caso la procedura consiste nel fissare una soglia  $\alpha$  e di considerare significative tutte le features  $i$  aventi  $P_i \leq \alpha$ . Questa procedura garantisce che il numero medio di falsi positivi é  $\mathbb{E}(F) = n\alpha$ .

Nel caso di analisi di dati di microarray il false positive rate é una misura troppo grande che non fornisce utili indicazioni.

## Family-wise error rate

Un'altra misura di errore é il family-wise error rate (FWER). La procedura consiste nel determinare un livello  $\alpha$  per ogni singolo test di ipotesi tale che:

$$\mathbb{P}\{F \geq 1\} = \alpha_{FWER}.$$

In questo modo vogliamo controllare la probabilità di avere anche un solo falso positivo negli  $n$  confronti. L'obiettivo é di determinare un livello  $\alpha$  per ogni singolo test tale che la probabilità di avere uno o piú falsi positivi sia  $\alpha_{FWER}$ .

## Family-wise error rate (cont.)

Se gli  $n$  confronti sono indipendenti allora:

$$\alpha_{FWER} = \mathbb{P}\{F \geq 1\} = 1 - \mathbb{P}\{F = 0\} = 1 - \binom{n}{0} \alpha^0 (1 - \alpha)^n.$$

Sapendo per Taylor che  $(1 - \alpha)^n \geq 1 - n\alpha$  allora:

$$\alpha_{FWER} = 1 - (1 - \alpha)^n \leq 1 - 1 + n\alpha.$$

Quindi possiamo concludere che:

$$\alpha_{FWER} \leq n\alpha.$$

## Correzione di Bonferroni

Quindi per controllare il FWER, ossia per limitare ad  $\alpha_{FWER}$  la probabilità di avere anche un solo falso positivo, il valore di soglia da utilizzare in ogni test é:

$$\alpha = \frac{\alpha_{FWER}}{n}.$$

Controllare  $\mathbb{P}\{F \geq 1\}$  attraverso la correzione di Bonferroni é utile quando ci aspettiamo che poche feature sono veramente alternative. Il FWER é una procedura troppo conservativa che aumenta sensibilmente il numero di falsi negativi.

Nel nostro esempio ( $n = 22283$ ) dovremmo utilizzare una soglia di  $\alpha = 4 \times 10^{-7}$  per avere  $\alpha_{FWER} = 0.01$ .

## False discovery rate

Consideriamo il rapporto:

$$\frac{\text{number of false positive feature}}{\text{number of significant feature}} = \frac{F}{F + T} = \frac{F}{S} = Q,$$

dove  $Q = 0$  se  $S = 0$ .  $Q$  indica la proporzione di feature false positive tra tutte quelle dichiarate significative.

Si definisce False Discovery Rate il valore atteso di questo rapporto:

$$FDR = \mathbb{E} \left( \frac{F}{S} \right).$$

## False discovery rate (cont.)

Sia  $\alpha$  il livello che abbiamo utilizzato in ogni singolo test, e quindi  $n\alpha$  é il numero atteso di falsi positivi. Detto  $s$  il numero di feature dichiarate significative a livello  $\alpha$ , allora una stima di FDR é:

$$q = \frac{n\alpha}{s}.$$

Quindi, nell'ipotesi di indipendenza delle  $n$  feature, se troviamo  $s$  feature significative a livello  $\alpha$  allora la percentuale di falsi positivi tra le  $s$  feature é  $100q\%$ .

Qual é il livello da usare in ogni singolo test per ottenere un dato valore di FDR?

## La procedura di Benjamini-Hochberg

Eseguiamo gli  $n$  test e siano  $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(n)}$  i corrispondenti p-value ordinati in ordine crescente. Detto  $\gamma$  il valore di FDR desiderato, per  $i = 1, 2, \dots, n$  confrontiamo ogni  $p_{(i)}$  con il valore  $q_i = \frac{i}{n}\gamma$  e sia  $k$  il massimo  $i$  tale che  $p_{(i)} \leq q_i$ . La procedura é la seguente:

a) se non esiste nessun  $i$  tale che  $p_{(i)} \leq q_i$  allora accettiamo tutte le  $n$  ipotesi nulle e concludiamo che nessuna feature é significativa; b) se  $k \geq 1$  allora rigettiamo tutte le ipotesi nulle  $H_{(1)}, \dots, H_{(k)}$  e accettiamo tutte le altre, ossia concludiamo che le prime  $k$  feature sono significative con FDR pari a  $\gamma$ .

La procedura garantisce che  $\mathbb{E}(Q) \leq \gamma$ .



## **Assunzione di indipendenza**

Il principale problema di questo metodo risiede nell'assunzione di indipendenza dei p-value. Questa assunzione non é ragionevole nel caso di dati di espressione genica.

## Esempio

Sia  $\gamma = 0.1$  e consideriamo i seguenti p-value ottenuti in  $n = 20$  test:

$i$	1	2	3	4	5	6	7	8	9	10
$p_{(i)}$	0.001	0.002	0.005	0.011	0.018	0.023	0.032	0.042	0.047	0.052
$q_i$	0.005	0.010	0.015	0.020	0.025	0.030	0.035	0.040	0.045	0.050

$i$	11	12	13	14	15	16	17	18	19	20
$p_{(i)}$	0.060	0.065	0.068	0.073	0.079	0.083	0.089	0.093	0.098	0.123
$q_i$	0.055	0.060	0.065	0.070	0.075	0.080	0.085	0.090	0.095	0.100

Poiché  $k = 7$ , dove 7 é il massimo  $i$  tale che  $p_{(i)} \leq q_i$  allora dichiariamo significative le prime 7 features con un FDR pari al 10%. Osserva che  $(0.035 \cdot 20) / 7 = 0.1$ , ossia il 10% delle feature dichiarate significative sono falsi positivi.