

# Appunti su: Regressione Multivariata (MANOVA)

Livio Finos

November 11, 2019

## Contents

<b>1</b>	<b>Regressione Multivariata (MANOVA)</b>	<b>1</b>
1.1	Definizione del modello e Stima dei parametri . . . . .	1
1.2	Stimatori di massima verosimiglianza . . . . .	2
1.3	Verifica di Ipotesi su $\mathbf{B}$ . . . . .	4
1.3.1	Test su $\mu$ con $\Sigma$ nota . . . . .	4
1.3.2	Test su $\mu$ con $\Sigma$ ignota . . . . .	5
1.3.3	Test su $\mathbf{B}$ in presenza di covariate . . . . .	5
1.3.4	$\Lambda$ di Wilks e altre misure di sintesi . . . . .	7

## 1 Regressione Multivariata (MANOVA)

### 1.1 Definizione del modello e Stima dei parametri

Il modello lineare

$$\mathbf{Y}_{n \times p} = \mathbf{X}_{n \times q} \mathbf{B}_{q \times p} + \mathbf{E}_{n \times p}$$

con  $\mathbf{E}$  matrice normale di dati con righe indipendenti e  $\epsilon_i \sim N(0, \Sigma) \forall i = 1, \dots, n$ . Si noti che questo equivale a dire che  $\mathbf{Y}$  è una m.n.d. da  $y_i \sim N(\mathbf{B}'x_i, \Sigma)$ .

La stima della matrice dei parametri  $\mathbf{B}$  e di  $\Sigma$  è realizzata in stretta analogia con il modello univariato. Sia  $\mathbf{H} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{I} - \mathbf{P}$  (vedi anche esercizio ??), la soluzione ai minimi quadrati è:

$$\hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

$$\hat{\Sigma} = S_u = (n - q)^{-1}\mathbf{Y}'\mathbf{H}\mathbf{Y}$$

e, di conseguenza,

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\mathbf{B}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \mathbf{P}\mathbf{Y}$$

e

$$\hat{\mathbf{E}} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{H}\mathbf{Y} = (\mathbf{I} - \mathbf{P})\mathbf{Y}.$$

**Esercizio** Dimostrare che la definizione  $\mathbf{H} = \mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}'$  già usata nel capitolo ?? è un caso particolare di  $\mathbf{H} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$  (e quindi definire  $\mathbf{X}$  che verifica l'equivalenza).

## 1.2 Stimatori di massima verosimiglianza

Per derivare più facilmente le proprietà inferenziali di questi stimatori ai minimi quadrati è utile metterli in relazione con i corrispettivi stimatori di massima verosimiglianza (SMV). Fortunatamente i due metodi portano a risultati analoghi. Lo stimatore  $\hat{\mathbf{B}}$  è lo stesso a cui si perviene tramite approccio dei minimi quadrati mentre  $\hat{\Sigma} = S = (n)^{-1}\mathbf{Y}'\mathbf{H}\mathbf{Y}$ . Per derivare questi risultati, è prima utile dimostrare il seguente:

**Lemma 1.1.** *Siano  $y'_i$ ,  $i = 1, \dots, n$  i vettori riga di una matrice  $\mathbf{Y}$ ,  $x'_i$ ,  $i = 1, \dots, n$  i vettori riga di una matrice  $\mathbf{X}$  e  $\hat{y}_i = \hat{\mathbf{B}}'x_i$ . Fissate una qualsiasi matrice  $\mathbf{B}_*$  e una matrice  $\Sigma$  (definita positiva), vale*

$$\begin{aligned} & \sum_{i=1}^n (y_i - \mathbf{B}_*'x_i)' \Sigma^{-1} (y_i - \mathbf{B}_*'x_i) = \\ & \text{tr}((\mathbf{Y} - \mathbf{X}\mathbf{B}_*)\Sigma^{-1}(\mathbf{Y} - \mathbf{X}\mathbf{B}_*)') = \\ & \text{tr}((\mathbf{Y} - \hat{\mathbf{Y}})\Sigma^{-1}(\mathbf{Y} - \hat{\mathbf{Y}})') + \text{tr}((\hat{\mathbf{Y}} - \mathbf{X}\mathbf{B}_*)\Sigma^{-1}(\hat{\mathbf{Y}} - \mathbf{X}\mathbf{B}_*)') \end{aligned}$$

**Proof 1.2.** *La prima equazione è vera per definizione. Per la seconda uguaglianza si ricordi la proprietà  $\text{tr}(A + B) = \text{tr}(A) + \text{tr}(B)$  per  $A$  e  $B$  matrici quadrate. Ora vale*

$$\begin{aligned} & \text{tr}((\mathbf{Y} - \mathbf{X}\mathbf{B}_*)\Sigma^{-1}(\mathbf{Y} - \mathbf{X}\mathbf{B}_*)') = \\ & \text{tr}((\mathbf{Y} - \hat{\mathbf{Y}} + \hat{\mathbf{Y}} - \mathbf{X}\mathbf{B}_*)\Sigma^{-1}(\mathbf{Y} - \hat{\mathbf{Y}} + \hat{\mathbf{Y}} - \mathbf{X}\mathbf{B}_*)') = \\ & \text{tr}((\mathbf{Y} - \hat{\mathbf{Y}})\Sigma^{-1}(\mathbf{Y} - \hat{\mathbf{Y}})') + \text{tr}((\mathbf{Y} - \hat{\mathbf{Y}})\Sigma^{-1}(\hat{\mathbf{Y}} - \mathbf{X}\mathbf{B}_*)') + \\ & \text{tr}((\hat{\mathbf{Y}} - \mathbf{X}\mathbf{B}_*)\Sigma^{-1}(\mathbf{Y} - \hat{\mathbf{Y}})') + \text{tr}((\hat{\mathbf{Y}} - \mathbf{X}\mathbf{B}_*)\Sigma^{-1}(\hat{\mathbf{Y}} - \mathbf{X}\mathbf{B}_*)') \end{aligned}$$

*Il risultato si ottiene se si osserva che il secondo e terzo termine sono nulli. Infatti, ricordando che vale  $\text{tr}(AB) = \text{tr}(BA)$  per  $A$  e  $B$  matrici qualsiasi, per il secondo termine:*

$$\begin{aligned} & \text{tr}((\mathbf{Y} - \hat{\mathbf{Y}})\Sigma^{-1}(\hat{\mathbf{Y}} - \mathbf{X}\mathbf{B}_*)') = \text{tr}(\mathbf{H}\mathbf{Y}\Sigma^{-1}(\mathbf{P}\mathbf{Y} - \mathbf{X}\mathbf{B}_*)') = \\ & \text{tr}(\mathbf{Y}\Sigma^{-1}(\mathbf{Y}'\mathbf{P}' - \mathbf{B}_*' \mathbf{X}')\mathbf{H}) = \text{tr}(\mathbf{Y}\Sigma^{-1}(\mathbf{Y}'\mathbf{P}'\mathbf{H} - \mathbf{B}_*' \mathbf{X}'\mathbf{H})) = \mathbf{0} \end{aligned}$$

*siccome  $\mathbf{Y}'\mathbf{P}'\mathbf{H} = \mathbf{Y}\mathbf{0} = \mathbf{0}$  e  $\mathbf{B}_*' \mathbf{X}'\mathbf{H} = \mathbf{B}_*'(\mathbf{X}' - \mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') = \mathbf{B}_*\mathbf{0}$ . La stessa conclusione vale per il terzo termine.*

Ora possiamo mostrare il risultato più rilevante:

**Theorem 1.3** (Stima di massima verosimiglianza).  $\hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$  e  $\hat{\Sigma} = S = \mathbf{Y}'\mathbf{H}\mathbf{Y}/n$  sono gli stimatori di massima verosimiglianza di  $\mathbf{B}$  e  $\Sigma$ , rispettivamente.

**Proof 1.4.** Omettendo il termine  $2\pi^{p/2}$  che non ha alcun ruolo nella massimizzazione della log-verosimiglianza, questa può essere esplicitata nel seguente modo:

$$\begin{aligned}
l_l(\mathbf{B}, \Sigma) &\propto -\frac{n}{2} \log |\Sigma| - \frac{1}{2} \sum_{i=1}^n (y_i - \mathbf{B}'x_i)' \Sigma^{-1} (y_i - \mathbf{B}'x_i) \\
&= -\frac{n}{2} \log |\Sigma| - \frac{1}{2} \text{tr} ((\mathbf{Y} - \mathbf{XB}) \Sigma^{-1} (\mathbf{Y} - \mathbf{XB})') \\
&= -\frac{n}{2} \log |\Sigma| - \frac{1}{2} \text{tr} ((\mathbf{Y} - \hat{\mathbf{Y}}) \Sigma^{-1} (\mathbf{Y} - \hat{\mathbf{Y}})') \\
&\quad - \frac{1}{2} \text{tr} ((\hat{\mathbf{Y}} - \mathbf{XB}) \Sigma^{-1} (\hat{\mathbf{Y}} - \mathbf{XB})') \tag{1}
\end{aligned}$$

Per la stima di  $\mathbf{B}$  si noti che la matrice di parametri  $\mathbf{B}$  è stata isolata nel terzo termine; questa quantità è sempre positiva e massimizza la verosimiglianza quando si riduce a 0, cioè quando  $\mathbf{B} = \hat{\mathbf{B}}$  (infatti  $\hat{\mathbf{Y}} = \mathbf{XB}$ ).

Per derivare lo stimatore di  $\Sigma$ , si osservi che per il secondo termine vale:

$$\begin{aligned}
-\frac{1}{2} \text{tr} ((\mathbf{Y} - \hat{\mathbf{Y}}) \Sigma^{-1} (\mathbf{Y} - \hat{\mathbf{Y}})') &= -\frac{1}{2} \text{tr} (\Sigma^{-1} (\mathbf{Y} - \hat{\mathbf{Y}})' (\mathbf{Y} - \hat{\mathbf{Y}})) = \\
&= -\frac{n}{2} \text{tr} (\Sigma^{-1} S) \tag{2}
\end{aligned}$$

(si ricordi che  $\text{tr}(AB) = \text{tr}(BA)$ ).

Quindi

$$l_l(\mathbf{B}, \Sigma) \propto -\log |\Sigma| - \text{tr}(\Sigma^{-1} S).$$

Mostriamo ora che il massimo è raggiunto in  $\Sigma = S$ . Si definisca  $A = \Sigma^{-1} S$  e si usi la proprietà  $|AB| = |A||B|$  (ponendo  $B = \Sigma$  si ha  $|\Sigma| = |S|/|\Sigma^{-1} S|$ ) per riscrivere quest'ultima come

$$l_l(\mathbf{B}, \Sigma) \propto -\log |\Sigma| - \text{tr}(\Sigma^{-1} S) = -\log |S| + \log |A| - \text{tr}(A).$$

$\log |S|$  è una costante e può essere esclusa dalle considerazioni che seguono.

Definiti ora  $\lambda_1, \dots, \lambda_p$  gli autovalori di  $A$  e ricordando che  $|A| = \prod_i \lambda_i$  e  $\text{tr}(A) = \sum_i \lambda_i$ , il problema si riduce alla minimizzazione di  $\sum_i^p (\lambda_i - \log \lambda_i)$ . La funzione  $f(\lambda) = \lambda - \log \lambda$  è minimizzata in  $\lambda = 1$  (ponendo la derivata pari a 0). Se tutti gli autovalori sono unitari  $A = \mathbf{I}$  e quindi  $\Sigma = S$ .

Questa dimostrazione ci permette di dire che gli stimatori  $\hat{\mathbf{B}}$  e  $S$  sono asintoticamente efficienti (raggiungono il limite inferiore di Rao-Cramer), però non garantiscono la stessa proprietà nel finito, specie per campioni piccoli. Anche la non distorsione è garantita asintoticamente; per  $\hat{\mathbf{B}}$  questa risulta valida anche al finito, mentre per ricavare uno stimatore non distorto di  $\Sigma$  è necessario fare ricorso allo stimatore ai minimi quadrati  $S_u = Sn/(n - q)$ .

Ciononostante, l'approccio di massima verosimiglianza resta fondamentale per ricavare una forma generale del test basato sul rapporto di verosimiglianza che vedremo tra poco.

## Proprietà

Anche le proprietà del modello estendono le proprietà del modello lineare univariato.

**Theorem 1.5.** *Dato il modello di regressione (lineare) e gli stimatori appena definiti:*

- $\hat{\mathbf{B}}$  è uno stimatore non distorto di  $\mathbf{B}$
- $E(\hat{\mathbf{E}}) = \mathbf{0}$
- $\hat{\mathbf{B}}$  (in vettore) ha distribuzione normale multivariata
- $\hat{\mathbf{B}} \perp \hat{\mathbf{E}}$
- $nS \sim W_p(\Sigma, n - q)$  (e quindi uno stimatore non distorto è dato da  $S_u = \frac{\sum \hat{e}_u^2}{n - q}$ )

## 1.3 Verifica di Ipotesi su $\mathbf{B}$

Affronteremo ora il problema della verifica d'ipotesi su (sottomatrici di)  $\mathbf{B}$ . In molti casi faremo uso dell'approccio generale del test di rapporto di log-verosimiglianza per ricavare alcuni tra i test già visti nei capitoli precedenti.

È quindi utile definirne la forma generale. Si supponga di voler testare  $H_0 : \mathbf{B} = \mathbf{B}_*$  contro  $H_1 : \mathbf{B} \neq \mathbf{B}_*$  (solitamente  $\mathbf{B}_* = \mathbf{0}$ ). Il rapporto di log-verosimiglianza è definito come  $\mathcal{W} = 2(l_l(\hat{\mathbf{B}}_{|H_1}) - l_l(\hat{\mathbf{B}}_{|H_0})) = 2(l_1 - l_0)$ .

Si noti anche che sotto  $H_0$  e sotto condizioni di regolarità asintoticamente vale  $\mathcal{W} \sim \chi_{pq}^2$ .

### 1.3.1 Test su $\mu$ con $\Sigma$ nota

Sia  $\mathbf{Y}$  una m.n.d da  $y_i \sim N(\mu, \Sigma)$ . Possiamo scrivere questo modello fissando la matrice  $\mathbf{X} = \mathbf{1}_n$ , cioè la media delle righe di  $\mathbf{Y}$  viene modellizzata esclusivamente tramite una costante e quindi  $E(\mathbf{Y}) = \mathbf{1}_n \mu' = \mathbf{1}_n \mathbf{B}$  (cioè  $\mu = \mathbf{B}'$ )

Si voglia verificare l'ipotesi  $H_0 : \mu = \mu_0 = \mathbf{B}'_0$  contro  $H_1 : \mu \neq \mu_0 = \mathbf{B}'_0$  con  $\mu_0$  fissato e  $\Sigma$  nota. In questo caso, la log-verosimiglianza ha forma:

$$l_l(\mathbf{B}, \Sigma) \propto -\frac{n}{2} \log |\Sigma| - \frac{1}{2} \text{tr}((\mathbf{Y} - \mathbf{B}\mathbf{X})\Sigma^{-1}(\mathbf{Y} - \mathbf{B}\mathbf{X})').$$

Sotto  $H_0$ , possiamo sostituire  $\mu = \mu_0$ ; sotto  $H_1$  lo SMV per  $\mathbf{1}_n \hat{\mu}' = \mathbf{1}_n (\mathbf{1}_n' \mathbf{1}_n)^{-1} \mathbf{1}_n' \mathbf{Y} = \mathbf{1}_n \bar{y}$  e usando la decomposizione (1) e la equivalenza (2), si ottiene

$$\begin{aligned} \mathcal{W} = l_1 - l_0 \propto & -\frac{n}{2} \log |\Sigma| - \frac{n}{2} \text{tr}(\Sigma^{-1} S) \\ & + \frac{n}{2} \log |\Sigma| + \frac{n}{2} \text{tr}(\Sigma^{-1} S) + \frac{1}{2} \text{tr}((\hat{\mathbf{Y}} - \mathbf{1}_n \mu'_0) \Sigma^{-1} (\hat{\mathbf{Y}} - \mathbf{1}_n \mu'_0)'). \end{aligned}$$

Si noti anche che  $\mathcal{W} \propto \text{tr}((\hat{\mathbf{Y}} - \mathbf{1}_n \mu'_0) \Sigma^{-1} (\hat{\mathbf{Y}} - \mathbf{1}_n \mu'_0)') = n(\hat{\mu} - \mu_0)' \Sigma^{-1} (\hat{\mu} - \mu_0)$   
Tutto si gioca quindi sulla distanza (di Mahalanobis) tra  $\hat{\mu}$  e il vero (sotto  $H_0$ )

$\mu_0$ . Quindi possiamo fare inferenza usando tale distanza; sappiamo già dal paragrafo ?? che questo è un modo valido per fare inferenza, ora ne ricaviamo anche che è un buon modo.

### 1.3.2 Test su $\mu$ con $\Sigma$ ignota

Il modello e le ipotesi sono le stesse del paragrafo precedente, però non assumiamo  $\Sigma$  nota. Riprendiamo la forma compatta della (log)verosimiglianza come definita in (1). Ricordando che  $\hat{\mathbf{B}}_{|H_0} = \mathbf{B}'_0 = \mu_0$  e  $\hat{\mathbf{B}}_{|H_1} = \hat{\mathbf{B}}' = \bar{y}$  se ne deduce che

$$\hat{\Sigma}_{|H_1} = (\mathbf{Y} - 1\bar{y})'(\mathbf{Y} - 1\bar{y})/n = S$$

e

$$\begin{aligned}\hat{\Sigma}_{|H_0} &= (\mathbf{Y} - 1\mu_0)'(\mathbf{Y} - 1\mu_0)/n \\ &= (\mathbf{Y} - 1\bar{y})'(\mathbf{Y} - 1\bar{y})/n + (1\bar{y}' - 1\mu'_0)'(1\bar{y}' - 1\mu'_0)/n \\ &= S + (\bar{y} - \mu_0)(\bar{y} - \mu_0)'\end{aligned}$$

Sempre con riferimento a (1), si nota che il secondo termine si riduce a  $-np/2$  sia sotto  $H_0$ , sia sotto  $H_1$  (infatti sotto  $H_1$ :  $\text{tr}((\mathbf{Y} - 1\mu')'\hat{\Sigma}^{-1}(\mathbf{Y} - 1\mu')) = \text{tr}(\hat{\Sigma}^{-1}(\mathbf{Y} - 1\mu')'(\mathbf{Y} - 1\mu')) = n\text{tr}(I_p)$  e analogamente sotto  $H_0$ ). Quindi il rapporto delle due massime log-verosimiglianze risulta funzione esclusiva degli stimatori delle varianze sotto le due ipotesi:

$$\mathcal{W} \propto n \log \frac{|S_{H_0}|}{|S_{H_1}|} = n \log \frac{|S + (\bar{y} - \mu_0)(\bar{y} - \mu_0)'|}{|S|}.$$

Ora sfrutteremo due proprietà dei determinati :

- i  $|AB|/|B| = |A|$  (ottenuta da  $|AB| = |A||B|$ ) e
- ii  $|I_p + BC| = |I_n + CB|$  (con  $B_{p \times n}$  e  $C_{n \times p}$ , usata con  $p = 1$ ).

Possiamo notare che  $|S + (\bar{y} - \mu_0)(\bar{y} - \mu_0)'| = |S(I + S^{-1}(\bar{y} - \mu_0)(\bar{y} - \mu_0)')|$  e che  $|(I + S^{-1}(\bar{y} - \mu_0)(\bar{y} - \mu_0)'| = |(1 + (\bar{y} - \mu_0)'S^{-1}(\bar{y} - \mu_0))|$ . Ne consegue che  $\mathcal{W} = n \log(|AB|/|B|) = n \log(|A|) = n \log(|1 + (\bar{y} - \mu_0)'S^{-1}(\bar{y} - \mu_0)|)$  dove, essendo l'argomento uno scalare, il determinante può essere omissso.  $(\bar{y} - \mu_0)'S^{-1}(\bar{y} - \mu_0)$  è esattamente l'espressione della  $T^2$  a meno del fattore  $n - 1$  e quindi ne è funzione monotona.

Questo dimostra che il test  $T^2$  già ricavato tramite il lemma ??, è anche un test del rapporto di verosimiglianza.

### 1.3.3 Test su $\mathbf{B}$ in presenza di covariate

È molto comune che la verifica di ipotesi riguardi solo una parte dei parametri e cioè delle righe di  $\mathbf{B}$  (si noti che se solo alcune colonne sono di interesse per il test, sarà sufficiente escludere le altre variabili/colonne dal modello). Il caso

più tipico è quello in cui si vuole vedere la relazione tra le variabili  $\mathbf{X}$  e  $\mathbf{Y}$  ammettendo però che l'intercetta sia non nulla.

In generale decomporremo la matrice  $\mathbf{X}$  in due sottomatrici  $\mathbf{X}_1$  e  $\mathbf{X}_0$  (di rango rispettivamente  $q_1$  e  $q_0$ ), facenti riferimento rispettivamente ai parametri sotto test  $\mathbf{B}_1$  e a quelli che possono assumere valori non nulli anche sotto l'ipotesi nulla  $\mathbf{B}_0$ .

$$\mathbf{Y} = \mathbf{X}_0\mathbf{B}_0 + \mathbf{X}_1\mathbf{B}_1 + \mathbf{E} \quad (3)$$

L'ipotesi nulla da testare risulta quindi

$$H_0 : \mathbf{B}_1 = \mathbf{0} \quad \forall \mathbf{B}_0$$

Tale approccio si rivela estremamente flessibile. Nel caso di  $c \geq 2$  due o più campioni indipendenti è sufficiente definire  $\mathbf{X}_0$  come la matrice con una sola colonna di uni (l'intercetta) e  $\mathbf{X}_1$  come matrice di  $c - 1$  variabili dummy (indicatrici di gruppo, esclusa la categoria di riferimento). Tale caso particolare è ha volte detto modello MANOVA.

Il caso di 2 (o  $c > 2$ ) campioni appaiati può essere incluso in questo schema definendo in  $\mathbf{X}_0$  le dummy ricavate dagli indicatori dei soggetti e in  $\mathbf{X}_1$  l'indicatore dei due (o  $c$ ) trattamenti. Si noti, inoltre, che è possibile includere il caso di un singolo campione definendo  $\mathbf{X}_0 = \emptyset$  e  $\mathbf{X}_1 = \mathbf{1}_n$ . Più in generale si possono includere disegni sperimentali a blocchi ed anche più complessi in questo schema generale.

**Esercizio** Definire la matrice  $\mathbf{X}$  di un modello di regressione multivariato che formalizzi il confronto di due campioni di numerosità  $n_1$  ed  $n_2$  (le prime  $n_1$  osservazioni di  $\mathbf{Y}$  provengono dal primo campione).

**Esercizio** Formalizzare  $\mathbf{X}_0$  e  $\mathbf{X}_1$  per la verifica dell'ipotesi nulla di uguaglianza dei due campioni.

A partire da questa impostazione, è possibile decomporre la devianza totale in modo del tutto analogo a quello che accade per i modelli univariati. Sia  $\mathbf{H}_0 = \mathbf{I} - \mathbf{X}_0(\mathbf{X}_0'\mathbf{X}_0)^{-1}\mathbf{X}_0'$ , si definiscano le seguenti quantità:

- Devianza Totale (sotto  $H_0$ ):  $T = \mathbf{Y}'\mathbf{H}_0\mathbf{Y}$
- Devianza Residua (sotto  $H_1$ ):  $W = \mathbf{Y}'\mathbf{H}\mathbf{Y}$
- Devianza Spiegata (sotto  $H_1$ ) da  $\mathbf{X}_1$ :  $B = \mathbf{Y}'\mathbf{H}_0\mathbf{Y} - \mathbf{Y}'\mathbf{H}\mathbf{Y} = \mathbf{Y}'(\mathbf{H}_0 - \mathbf{H})\mathbf{Y}$

Si noti che grazie al teorema ?? di Craig - Lancaster possiamo ricavare che le tre matrici casuali sono indipendenti tra loro, seguono una distribuzione di Wishart e la loro somma è una wishart con g.d.l. pari al rango di  $\mathbf{H}_0$ . In ambito univariato, la statistica test F è basata sul rapporto  $F = \frac{B/\text{rango}(\mathbf{H}_0 - \mathbf{H})}{W/\text{rango}(\mathbf{H})}$  la cui distribuzione è nota. Vale anche la pena notare che lo stesso indice di proporzione di varianza spiegata  $R^2 = \frac{B}{T} = \frac{B}{B+W}$ .

In ambito multivariato le statistiche di sintesi non sono altrettanto immediate. Nei casi non banali, la statistica test non può basarsi sul rapporto delle due devianze giacché numeratore e denominatore sono delle matrici. È necessario quindi “sintetizzare” queste due matrici in un’unica quantità.

Si noti che un caso particolare si verifica quando  $\mathbf{X}_1$  ha una sola colonna (e quindi  $\text{rango}(\mathbf{H}_0 - \mathbf{H}) = 1$ ). Nel ricavare la distribuzione della statistica test del rapporto di verosimiglianze (paragrafo 1.3.2) abbiamo sfruttato la proprietà ii. per scoprire che questa segue una  $T^2$ . In generale questo risulta vero non solo per il caso di due campioni, ma più in generale nel caso di un solo predittore sotto test.

Continuando sulla strada segnata del test del rapporto di verosimiglianza, possiamo ricavare il test per il generico problema di regressione. In modo del tutto analogo a quanto mostrato in 1.3.2 (e facendo ancora riferimento ai risultati del lemma 1.1), possiamo notare che il secondo termine in 1 si riduce a  $-np/2$  sia sotto  $H_0$  che  $H_1$ : infatti sotto  $H_1$ :  $\text{tr} \left( (\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}) \hat{\Sigma}^{-1} (\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})' \right) = \text{tr} \left( \hat{\Sigma}^{-1} (\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})' (\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}) \right) = n \text{tr}(I_p)$  e analogamente sotto  $H_0$  usando  $\mathbf{X}_0$  e  $\mathbf{B}_0$ . Il test del rapporto di verosimiglianza quindi si riduce ad una funzione monotona dei determinanti delle varianze (funzione delle devianze) residue sotto  $H_0$  ed  $H_1$ :

$$\mathcal{W} \propto n \log \frac{|S_{H_0}|}{|S_{H_1}|} = n \log \frac{|T|/n}{|W|/n} = n \log \frac{|W + B|}{|W|}.$$

Come anticipato però, la proprietà ii. usata nel paragrafo 1.3.2 non ci permette di giungere a nessuna forma distributiva nota. Siamo quindi costretti a sfruttare i risultati asintotici ( $W$  ha distribuzione asintotica  $\chi^2_{pq_1} - q_1$  e  $p$  dimensioni di  $\mathbf{B}$ ) oppure ricavare la forma della nuova variabile aleatoria.

### 1.3.4 $\Lambda$ di Wilks e altre misure di sintesi

#### $\Lambda$ di Wilks

il rapporto delle verosimiglianze

$$\lambda^{2/n} \propto |W|/|T| = |W| |W + B|^{-1} = |\mathbf{I} + W^{-1}B|^{-1} \sim \Lambda(p, n - q_0 - q_1, q_1)$$

(parametri:  $p$  = colonne  $\mathbf{Y}$ ,  $n - q_0 - q_1$  = rango di  $W$ ,  $q_1$  = rango di  $B$ )

Ora, sotto  $H_0$ ,  $X$  è una matrice di dati da  $N_p(\mu, \Sigma)$ , così dal teorema di Cochran generalizzato si ha, ponendo  $C_1 = \mathbf{H}$  e  $C_2 = \mathbf{H}_0 - \mathbf{H}$

$$W = X' C_1 X \sim W_p(\Sigma, n - q_0 - q_1)$$

$$B = X' C_2 X \sim W_p(\Sigma, q_1)$$

Dal momento che  $B$  e  $W$  sono indipendenti, a condizione che  $n - q_0 - q_1 > 0$ , sotto  $H_0$  la statistica  $\lambda^{2/n}$  si distribuisce come una  $\Lambda$  di Wilks

$$|I + W^{-1}B|^{-1} \sim \Lambda(p, n - q_0 - q_1, q_1)$$

Purtroppo solo in casi particolari - cioè per particolari valori dei parametri- si può utilmente sfruttare la distribuzione al finito di  $\Lambda$ .

**Theorem 1.6.** *Si ha*

$$\Lambda(p, m, n) \sim \prod_{i=1}^n u_i$$

dove  $u_1, \dots, u_n$  sono  $n$  variabili indipendenti e  $u_i \sim \beta(\frac{1}{2}(m+i-p), \frac{1}{2}p), i = 1, \dots, n$ .

*Senza dimostrazione.*

Per valori grandi di  $m$  si può anche utilizzarsi il risultato asintotico.

**Theorem 1.7.**

$$-[m - \frac{1}{2}(p - n + 1)] \log \Lambda(p, m, n) \sim \chi_{np}^2$$

per  $m \rightarrow \infty$ .

*Senza dimostrazione.*

### Altre misure di sintesi della varianza spiegata

Come detto, la statistica di sintesi  $\lambda = |W|/|T|$  non è l'unica possibile. Va anche sottolineato che in ambito multivariato, non è facile stabilire criteri di ottimalità; ad esempio non esiste un test uniformemente più potente. La scelta di  $\lambda$  quindi è giustificabile per l'importanza che ricopre l'approccio di verosimiglianza in tutto l'ambito inferenziale e per il fatto che sia nota la sua distribuzione asintotica, ma non gode di migliori proprietà rispetto alle altre statistiche di sintesi che indicheremo qui di seguito.

Partendo dalla quantità  $A = W^{-1}B$  è possibile stabilire diverse misure di sintesi. Siano  $\lambda_1, \dots, \lambda_p$  gli autovalori non nulli di  $A = W^{-1}B$ . Tra le molte, le più note sono:

- Wilks:  $\prod_i (1/(1 + \lambda_i)) \sim \Lambda$  come già visto
- (Traccia di) Pillai:  $\sum_i (1/(1 + \lambda_i))$
- (Traccia di) Lawley-Hotelling:  $\sum_i \lambda_i$
- (Massima radice di) Roy:  $\max_i \lambda_i$

Nessuna di queste statistiche ha una distribuzione nota (o facile da ricavare) sotto l'ipotesi nulla. I progressi tecnologici degli ultimi decenni, però ci permettono di ottenere queste distribuzioni tramite simulazione.