

# ANALISI DI VARIANZA DELLE CARATTERISTICHE PRESTAZIONALI DEGLI OPEN GRADE

Maurizio Crispino (1) Federico Fiori (2)  
Emanuela Marforio (3) Luigi Mussio(4)

- (<sup>1</sup>) D.I.I.A.R. – Politecnico di Milano – e-mail: [maurizio.crispino@polimi.it](mailto:maurizio.crispino@polimi.it)  
(<sup>2</sup>) D.I.I.A.R. – Politecnico di Milano – email: [federico.fiori@polimi.it](mailto:federico.fiori@polimi.it)  
(<sup>3</sup>) D.I.I.A.R. – Politecnico di Milano – e-mail: [fam.marforio@alice.it](mailto:fam.marforio@alice.it)  
(<sup>4</sup>) D.I.I.A.R. – Politecnico di Milano – e-mail: [luigi.mussio@polimi.it](mailto:luigi.mussio@polimi.it)

## Riassunto

Lo studio delle prestazioni e della durabilità del sistema composito caratterizzato dalla presenza del conglomerato bituminoso e della malta cementizia come materiale innovativo è nato dall'esigenza di sviluppare nuove tecnologie, mirate alla ricerca di un valido compromesso tra le pavimentazioni rigide e flessibili. Un campo prove per la sperimentazione in vera grandezza ha permesso l'estrazione di materiale: carote e travetti, da estrarre e sottoporre a diversi test, mirati a valutare modulo di rigidezza e resistenza a fatica. Dopo un'ampia descrizione dei test eseguiti, sono passati in rassegna tutti i risultati dell'analisi di varianza, a tre vie senza interazioni tra le celle, e correlati per determinare la composizione del sistema. Un ampio confronto con altri materiali impiegati nella realizzazione delle pavimentazioni è condotto a partire dai conglomerati bituminosi tradizionali fino ad altre tipologie di Open Grade.

## Abstract

The requirement to develop new technologies, in order to find a valid compromise between rigid and flexible pavements, promotes the study of performances and durability of a composite system, characterized by presence of open grade asphalt and cement mortar as innovative material. A test field for experimentation in true largeness allowed the extraction of cylindrical specimens and beams, to extract and to subject to various tests, with the aim to estimate the stiffness modulus and the fatigue resistance. After a wide description of executed tests, a three ways variance analysis without interactions among cells is performed and its results are correlated in order to determine the best composition of the system. A wide comparison with other materials used in the realization of pavements is done starting from traditional bituminous conglomerates and arriving to other types of Open Grade.

## Introduzione all'analisi multivariata

La validazione dati e dei modelli si fonda sulle varie tecniche dell'analisi multivariata, di volta in volta, studiando la variabilità e l'interdipendenza fra gli attributi, entro una classe di oggetti. Essa prende in considerazione insiemi di dati, ciascuno dei quali, relativo ad un oggetto della classe, contiene i valori osservati di certe variabili statistiche. Questi insiemi possono talvolta essere completi; mentre più spesso sono estratti da insiemi più grandi. A loro volta, le variabili osservate, sono in generale campioni estratti da variabili casuali, di tipo continuo o, raramente, discreto. Da una tale complessità di premesse, lo studio dell'analisi multivariata è variamente articolato.

- ❑ Semplificazione strutturale. Gli insiemi di dati devono essere ricondotti, se possibile, in forme più semplici con cambi di variabili, in particolare, con trasformazioni capaci di sciogliere variabili connesse in variabili indipendenti.
- ❑ Classificazione degli oggetti. L'analisi dell'insieme di dati deve porre in evidenza la presenza di gruppi (clusters), ovvero di sottoinsiemi di oggetti, caratterizzati da valori preferenziali degli attributi o di parte di essi, cercando di ricondurre a poco le notevoli variabilità presenti.
- ❑ Raggruppamento degli attributi (clustering). L'analisi degli insiemi di dati deve far ricadere, per quanto possibile, differenti variabili in un unico gruppo.
- ❑ Analisi della connessione. Gli insiemi di dati devono essere studiati rispetto alla dipendenza vaga e generica o meno fra le variabili contenute (ovvero all'essere in connessione di queste ultime).
- ❑ Analisi della dipendenza funzionale. Gli insiemi di dati devono essere studiati rispetto alla dipendenza funzionale o meno fra le variabili contenute (ovvero all'essere in regressione di

queste ultime), con particolare riferimento alla dipendenza lineare o correlazione (ovvero all'essere in regressione lineare o correlate).

- Costruzione e verifica d'ipotesi. Il confronto probabilistico, fra statistiche campionarie e valori teorici di riferimento, permette di formulare un giudizio critico sui risultati ottenuti nelle varie tappe dell'analisi.

In particolare, un insieme di dati, stratificato per regioni (dati a referenza spaziale, o spazio-varianti) e/o per epoche (dati variabili nel tempo, o tempo-varianti), come pure per classi di valore (dati varianti in uno spazio astratto), è raggruppabile in celle tridimensionali. L'informazione dentro le celle è rappresentata da una variabile statistica multidimensionale i cui elementi sono i valori degli attributi dell'informazione stessa. Nella tecnica di analisi di varianza si considera la varianza generale come somma di due componenti, varianza spiegata e varianza residua, ortogonali tra loro. Lo scopo è quello di massimizzare la varianza spiegata (che descrive il fenomeno secondo il modello stratificato) rispetto a quella residua (che è legata, invece, al carattere statistico dei dati e, in conclusione, alla loro accidentalità).

Il giudizio statistico è poi formulato tramite l'inferenza statistica (multivariata), ovvero quella parte dell'analisi multivariata dedicata alla costruzione e verifica d'ipotesi (per problemi di controllo di qualità, oppure controllo e confronto d'ipotesi di altri problemi). Infatti il confronto probabilistico permette di formulare un giudizio critico sui risultati ottenuti, nelle varie tappe dell'analisi multivariata. Così con le usuali strategie dell'inferenza statistica, vari tests multipli consentono di discriminare, tanto stime di parametri da campioni normali, quanto statistiche di modelli non-parametrici (ovvero modelli distribution free).

Gli oggetti del giudizio critico sono, come detto, i risultati ottenuti nelle varie tappe dell'analisi multivariata, in particolare: frequenze relative, contingenze, medie campionarie o altri indicatori del centro (di una popolazione), varianze campionarie o altri indicatori della dispersione, coefficienti di correlazione campionari o altri indicatori della correlazione. I test multipli si differenziano per i diversi oggetti in esame, come pure per la distribuzione di appartenenza della popolazione cui i campioni si riferiscono, se normale, oppure altra (spesso sconosciuta). In generale, comunque, tutti i campioni sono supposti fra loro indipendenti, mentre se quest'ipotesi non è soddisfatta, occorre procedere, come per i modelli non-parametrici, adottando strategie assolutamente generali, ma assai poco potenti, cioè meno capaci di discriminare fra ipotesi alternative vicine, a parità di numerosità dei campioni. Per quanto riguarda i loro fondamenti teorici, questi si richiamano, in generale:

- alla definizione assiomatica di probabilità;
- alla legge dei grandi numeri ed al limite centrale della statistica;
- ai teoremi della normalità: conservazione per trasformazioni lineari, identità fra indipendenza ed incorrelazione;
- alla definizione della variabile casuale  $\chi^2$  (chi quadrato);
- alla definizione delle variabili casuali  $t$  (t di Student) e  $F$  (F di Fisher);
- al calcolo di probabilità estremali (ad es. di Kolmogorov–Smirnov e di Hawkins), ove necessario;
- al teorema di decomposizione ortogonale degli scarti;

e nello specifico, alla trasformazione della distribuzione di probabilità, secondo requisiti da definirsi, caso per caso, così come è costruito un determinato test multiplo.

### **Analisi di varianza a tre vie senza interazioni fra le celle:**

Un insieme di dati, stratificato per regioni (dati a referenza spaziale, o spazio-varianti) e/o per epoche (dati variabili nel tempo, o tempo-varianti), come pure per classi di valore (dati varianti in uno spazio astratto), è raggruppabile in celle bidimensionali o tridimensionali. L'informazione

dentro le celle è rappresentata da una variabile statistica multidimensionale i cui elementi sono i valori degli attributi dell'informazione stessa. Nella tecnica di analisi di varianza si considera la varianza generale come somma di due componenti, varianza spiegata e varianza residua, ortogonali tra loro. Lo scopo è quello di massimizzare la varianza spiegata (che descrive il fenomeno secondo il modello stratificato) rispetto a quella residua (che è legata, invece, al carattere statistico dei dati e, in conclusione, alla loro accidentalità).

Nell'analisi di varianza a tre vie senza interazione tra le celle, il modello funzionale adottato è il seguente, mentre il modello stocastico prevede osservazioni indipendenti e di uguale precisione:

$$\begin{aligned} \bar{a} + \hat{a}_i + \hat{a}_j + \hat{a}_k &= s_{ijkl}^o + \hat{v}_{ijkl} & \forall i = 1, \dots, I \\ & & \forall j = 1, \dots, J \\ & & \forall k = 1, \dots, K \\ & & \forall l = 1, \dots, L = 1 \end{aligned}$$

Nella espressione precedente  $I$  è il numero trattamenti,  $J$  è il numero di blocchi,  $K$  è il numero di strati ed  $L=1$  è la dimensione della variabile statistica. Il numero complessivo di osservazioni (si veda, a riguardo, la tabella sottostante) è  $m = I \cdot J \cdot K \cdot L = 30$ , ed il numero di parametri è  $n = I + J + K + 1 = 11$ .

10711	10444	7438	8999	10451
19273	17944	12957	18358	17912
10813	9992	9384	9320	12473
18864	17278	17916	16791	19018
10396	10395	8070	10057	10245
18630	18458	15318	18612	19724

A tali equazioni è necessario aggiungere alcune condizioni di vincolo, a causa della sovrapparametrizzazione dovuta ad incognite non linearmente indipendenti che determinano singolarità del sistema. Pertanto per ogni gruppo di parametri, si è imposto che la loro somma fosse esattamente nulla, ovvero una matrice di coefficienti tutti unitari a basso peso ( $10^{-2}$ ) è stata addizionata alla matrice normale per ogni gruppo di parametri.

I valori numerici ad esse relativi possono, ancora, essere facilmente ottenuti risolvendo, a minimi quadrati, i suddetti sistemi di equazioni d'osservazione e, tramite essi, propagando, opportunamente, le varianze delle osservazioni.

#### SOLUZIONE, SQM E TEST T DI STUDENT

MEDIA:					1
1	1.3875	0.6263	2.2153	2.0700	
TRATTAMENTI:					2
2	-0.3929	0.4946	-0.7944	2.0700	
3	0.3929	0.4946	0.7944	2.0700	
BLOCCHI:					5
4	0.0906	0.2010	0.4511	2.0700	
5	0.0210	0.2010	0.1047	2.0700	
6	-0.2027	0.2010	-1.0089	2.0700	
7	-0.0185	0.2010	-0.0922	2.0700	
8	0.1096	0.2010	0.5453	2.0700	

STRATI:	3			
9	-0.0426	0.3305	-0.1289	2.0700
10	0.0310	0.3305	0.0939	2.0700
11	0.0116	0.3305	0.0350	2.0700

Dalla tabella si evidenzia ovviamente la somma nulla dei parametri di ciascun gruppo, ma soprattutto la non significatività di alcun parametro, fatto salvo la media generale. Eppure come specificato nel prosieguo, l'analisi di varianza a tre vie è del tutto significativa. Infatti dalle stime dei parametri si ricavano, immediatamente, le varianze spiegate, relative ai trattamenti  $T$ , ai blocchi  $B$  ed agli strati  $S$ , concordemente al modello di riferimento prescelto:

$$\sigma_T^2 = \frac{1}{I-1} \sum_{i=1}^I \hat{a}_i^2 \quad \sigma_B^2 = \frac{1}{J-1} \sum_{j=1}^J \hat{a}_j^2 \quad \sigma_S^2 = \frac{1}{K-1} \sum_{k=1}^K \hat{a}_k^2$$

e la varianza residua:

$$\sigma_R^2 = \frac{1}{\nu} \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K \sum_{l=1}^L \hat{v}_{ijkl}^2$$

dove il denominatore rappresenta i gradi di libertà del sistema:  
 $\nu = I \cdot J \cdot K \cdot L - (I-1) - (J-1) - (K-1) - 1$

Un giudizio sull'analisi di varianza e sulla validità dei modelli adottati discende dalla comparazione fra la varianza spiegata e residua. In questa sede, si utilizza sempre lo schema rigido di comparazione, confrontando direttamente la varianza spiegata e la varianza residua. Di conseguenza, si ha:

ANALISI DI VARIANZA:			
TRATTAMENTI E TEST F DI FISHER	2	15.7972	3.4400
BLOCCHI E TEST F DI FISHER	5	1.2715	2.6600
STRATI E TEST F DI FISHER	3	0.0993	3.0500

dove il test F di Fisher passa, in abbondanza, per quanto riguarda i trattamenti, ovviamente significativi che caratterizzano la prova in esame (e danno senso e significato all'intero trattamento statistico dei dati), mentre lo stesso test non passa per quanto riguarda sia i blocchi che gli strati. Eppure è possibile rilevare come i blocchi diano un po' più di significato al test, rispetto agli strati, in quanto cambiare blocco vuol dire operare in un altro sito, mentre cambiare strato vuol dire semplicemente cambiare campione nello stesso sito. Come evidente, il cambio di sito innesca una certa variabilità, per quanto debole ed indesiderata, mentre il cambio di campione nello stesso sito ne innesca una certamente minore, altrettanto indesiderata e comunque trascurabile.

Queste considerazioni danno ragione della effettiva validità dell'analisi di varianza, rispetto al semplice test di significatività dei parametri, effettuato con la distribuzione t di Student. In effetti, confrontando qualitativamente la varianza a priori con il sigma zero a posteriori si può osservare una significativa riduzione della dispersione (a riguardo, si noti che, data la loro stretta dipendenza, potrebbero essere confrontati solo in termini non - parametrici) ed il test chi quadrato passa con un'abbondanza addirittura sbalorditiva:

MEDIA, SQM E				
N. DATI E PARAMETRI	1.4067	0.4181	30	11
GDL, SIGMA ZERO E				
TEST CHI QUADRATO	22	0.0988	0.1857	33.9000

Controlli numerici danno ragione della sostanziale correttezza del modo di procedere. Infatti tanto il numero di condizione del sistema (calcolato sulla norma dell'estremo superiore delle matrici normale ed inversa), quanto le ridondanze locali delle singole osservazioni (calcolate in base al teorema di decomposizione ortogonale della varianza) sono relativamente prossimi ad uno:

NUMERO DI CONDIZIONE	0.8364
RIDONDANZE LOCALI	0.7333

cosa che ben garantisce di una sufficiente lontananza tanto dalla condizione di singolarità e mal – condizionamento del sistema, quanto dalla condizione di inaffidabilità e non – controllabilità delle singole osservazioni. Da ultimo per quanto riguarda l'eventuale presenza di dati anomali, occorre segnalare come il test tau di Thompson evidenzi due osservazioni, su trenta, appena oltre la soglia di criticità:

DATI ANOMALI, SCARTI, SQM, TEST TAU DI THOMPSON					
14	1.2957	0.2393	0.0846	2.8270	1.9386
16	1.7916	-0.1830	0.0846	-2.1616	1.9386

cosa che assicura della sufficiente bontà dei dati rilevati ed esclude particolari necessità di eliminare dati anomali, prima di effettuare il calcolo di statistiche ottimali. Nell'immediato prosieguo, si riportano le stime delle osservazioni compatibili con il modello funzionale adottato, precedute dai valori delle osservazioni (divisi per  $10^4$ ) e seguite dal loro scarto quadratico medio:

1	1.0711	1.0426	0.0510	16	1.7916	1.6086	0.0510
2	1.9273	1.8284	0.0510	17	0.8070	0.8034	0.0510
3	1.0813	1.1163	0.0510	18	1.5318	1.5892	0.0510
4	1.8864	1.9020	0.0510	19	0.8999	0.9335	0.0510
5	1.0396	1.0968	0.0510	20	1.8358	1.7192	0.0510
6	1.8630	1.8826	0.0510	21	0.9320	1.0071	0.0510
7	1.0444	0.9730	0.0510	22	1.6791	1.7929	0.0510
8	1.7944	1.7588	0.0510	23	1.0057	0.9876	0.0510
9	0.9992	1.0467	0.0510	24	1.8612	1.7734	0.0510
10	1.7278	1.8324	0.0510	25	1.0451	1.0616	0.0510
11	1.0395	1.0272	0.0510	26	1.7912	1.8473	0.0510
12	1.8458	1.8130	0.0510	27	1.2473	1.1352	0.0510
13	0.7438	0.7492	0.0510	28	1.9018	1.9210	0.0510
14	1.2957	1.5350	0.0510	29	1.0245	1.1157	0.0510
15	0.9384	0.8229	0.0510	30	1.9724	1.9015	0.0510

Si noti altresì come, essendo tutti uguali gli scarti quadratici medi delle stime suddette, il teorema di decomposizione ortogonale della varianza faccia sì che anche lo scarto quadratico medio di tutti gli scarti residui sia sempre uguale e pari a 0.0846.

Infine giova ribadire come i test di validazione dei modelli, quali quelli qui presentati e svolti, permettano di sottoporre a verifica, mediante opportuni controlli e confronti d'ipotesi, le stime effettuate come, del resto, tutti i risultati ottenuti nell'ambito della statistica.

Da ultimo, si noti come siano 40 le prove realmente effettuate (su tre differenti valori di temperatura: 5°, 20° e 40°), anche se non regolarmente distribuite come le prime 30, prese in considerazione. In questo caso, l'analisi di varianza ha dato risultati del tutto comparabili e talvolta rafforzativi, ad ulteriore conferma delle attese, pur presentando variazioni negli indici, qui costanti, per l'irregolarità diffusasi nella tabella dei dati e, di conseguenza, nei risultati ottenuti.

## **Conclusione**

L'analisi di varianza effettuata ha fornito i risultati attesi, in maniera assolutamente ottimale. Del resto, è particolarmente difficile non accogliere favorevolmente risposte che dichiarano significative le metodiche e non – significativi tanto i siti, quanto le ripetizioni, dove i secondi sono tuttavia un po' più rilevanti delle ultime. Eppure tutto ciò potrebbe far sorgere il dubbio di un'analisi pilotata e, di conseguenza, alcune considerazioni / precisazioni devono essere messe in chiaro, come esposto nell'immediato prosieguo.

- ❑ Una qualsiasi analisi statistica fallisce, quasi certamente, se chi la conduce non dispone delle capacità operative e dei mezzi tecnici per effettuarla a regola d'arte.
- ❑ Una qualsiasi analisi statistica deve obbligatoriamente muoversi sulla base di alcune aspettative, fondate sulle preconoscenze dei fenomeni e/o dei processi oggetto di studio.
- ❑ Una qualsiasi analisi statistica può confermare liberamente o, altrettanto liberamente, smentire quanto ipotizzato, sulla base delle suddette preconoscenze.
- ❑ In entrambi i casi, i dati raccolti portano a calcolare statistiche campionarie di alcuni modelli, come stime dei valori teorici della popolazione, in generale, incogniti ed inconoscibili.
- ❑ Successivamente le statistiche campionarie sono confrontate con i valori di riferimento, rappresentativi dei sopraccitati valori teorici, dove l'aderenza o meno determina una conferma od una smentita.
- ❑ Nel primo caso infatti, l'aderenza probabilistica fra statistiche campionarie e valori di riferimento porta a confermare le ipotesi effettuate ed i modelli prescelti.
- ❑ Nel secondo caso invece, la non – aderenza probabilistica fra statistiche campionarie e valori di riferimento porta viceversa a smentire le ipotesi effettuate.
- ❑ Pertanto l'accadere del secondo caso richiede di abbandonare i modelli prescelti, modificare le ipotesi effettuate e riconsiderare le preconoscenze, per quanto possibile.
- ❑ Al contrario, l'accadere del primo caso è certamente un'utile validazione sperimentale di modelli, ipotesi e preconoscenze, anche se non costituisce alcuna dimostrazione.
- ❑ Resta da notare come una dimostrazione non sia una verifica sperimentale concreta, ma un'operazione logica astratta per cui basta un solo controesempio, mentre non servono innumerevoli esempi.

Un'ultima importante considerazione / precisazione non attiene all'epistemologia della ricerca, ma all'etica del ricercatore. Questi infatti non deve falsare i risultati ottenuti, in alcun modo, alterandoli verso una qualche direzione cui vuole far giungere la ricerca in corso. Proprio per questo, è auspicabile che sappia resistere a lusinghe e/o minacce di tipo ideologico, politico, economico, accademico, ecc., compreso il prestigio personale, che possono gravemente compromettere la libertà della ricerca e del ricercatore stesso.

E' un compito arduo, perché ancora più delle minacce, le lusinghe possono essere subdole e sottotaciute. Ciononostante la stima di un ricercatore ed il valore effettivo di una ricerca stanno sostanzialmente nella capacità di rifiutare clausole compromissorie invalidanti e nell'eventuale denuncia pubblica del venire meno dei mezzi materiali, per poter impostare liberamente, sviluppare nel modo dovuto e condurre a termine con dignità la ricerca intrapresa.

## **Ringraziamenti**

Gli autori ringraziano la Impresa Bacchi s.r.l. di Carpiano (MI) e, in particolare, l'ing. Loretta Venturini ed il suo staff tecnico per la preziosa collaborazione offerta.

## **Bibliografia**

Fischer F.E. (1973): Fundamental Statistical Concepts. Canfield Press, San Francisco.  
Sachs L. (1984): Applied Statistics. A handbook of Techniques. Springer, New York.