

DIPARTIMENTO DI ECONOMIA E FINANZA

# **METODI E ANALISI STATISTICHE**

2020



**UNIVERSITÀ  
DEGLI STUDI DI BARI  
ALDO MORO**



DIPARTIMENTO DI ECONOMIA E FINANZA

# **METODI E ANALISI STATISTICHE**

2020



UNIVERSITÀ  
DEGLI STUDI DI BARI  
ALDO MORO

Tutti i diritti di traduzione, riproduzione e adattamento, totale o parziale, con qualsiasi mezzo (comprese le copie fotostatiche e i microfilm) sono riservati

Toma E., d'Ovidio F. (a cura di) (2020). *Metodi e Analisi Statistiche*, Dipartimento di Economia e Finanza, Università degli studi di Bari *Aldo Moro*.

© Copyright 2020 by Università degli Studi di Bari Aldo Moro  
www.uniba.it

Prima edizione: dicembre 2020

ISBN 978-88-6629-023-0

Gli articoli qui presentati sono stati oggetto, oltre che di valutazione interna, anche di revisione anonima (in “doppio cieco”).

*Editing finale:* F.D. d'Ovidio, E. Toma

# Sommario

Ernesto Toma <b>Presentazione</b> .....	pag. 5
Massimo Bilancia, Giovanni Sansaro <b>Inferenza per modelli a topic latenti: una introduzione</b> .....	« 7
Natalia Leone, Valeria Ancona, Davide Fragnito, Domenico Vitale, Massimo Bilancia <b>Geostatistical analysis of soil reflectance spectra for field-scale digital soil mapping. A case study.</b> .....	« 95
Claudia Marin, Fabio Manca <b>Modelli generativi per la sentiment analysis</b> .....	« 119
Davide Fragnito, Natalia Leone, Valeria Ancona, Domenico Vitale, Antonio Lucadamo <b>Comparison of different multivariate calibrations and ensemble methods for estimating selected soil properties with vis-NIR reflectance spectroscopy.</b> .....	« 135
Maria De Caro, Ilaria Pepe, Paolo Taurisano, Ernesto Toma <b>Studio longitudinale sulle variazioni del (di)stress tra il pre e post-intervento chirurgico per carcinoma mammario in funzione delle strategie di coping adoperate</b> .....	« 163
Laura Antonucci, Corrado Crocetta, Yana Kostiuk <b>Valutazione dei servizi web offerti dall'Università di Foggia ai tempi del COVID-19</b> .....	« 189
Laura Antonucci, Corrado Crocetta, Massimo Russo <b>Analisi della soddisfazione rispetto ai servizi erogati da Sanitaservice ASL Foggia s.r.l.</b> .....	« 205
Nunziata Ribecco, Isabella Stasi, Clelia Punzo, Annalisa Rizzi, Giovanni Tomasicchio <b>Il carcinoma mammario in donne anziane e in donne giovani.</b> .....	« 231

Salvatore Cariello, Monica Carbonara <i>La povertà nei comuni del Mezzogiorno</i> .....	« 255
Ezio Ritrovato <i>Crisi economica, prezzi alimentari e delinquenza nella Puglia di fine Ottocento: correlazione e causalità</i> .....	« 267
Angela Maria D’Uggento, Alessandra Milillo, Barbara Cafarelli <i>È conveniente investire nella finanza sostenibile?</i> .....	« 281
Agata Maria Madia Carucci, Giovanni Vannella <i>Analisi delle strutture evolutive imprenditoriali a livello regionale e comunale: una applicazione su Puglia e Basilicata</i> .....	« 293
Crescenzo Gallo, Alessandro Rinaldi <i>Data Envelopment Analysis: application of the efficient frontier on the financial field in the European and the American scenarios</i> .....	« 323
Mauro Gianfranco Bisceglia <i>La stabilità nelle equazioni differenziali lineari</i> .....	« 343

# Presentazione

Il volume *Metodi e analisi statistiche*, che nel 2020 giunge alla quinta edizione, si conferma un importante momento di confronto per molti ricercatori impegnati sia nell'ambito esclusivamente statistico, che in altri ambiti scientifici che utilizzano il "metodo quantitativo" per analisi matematiche, economiche, storiche, psicologiche, ecc..

Tale metodo diviene in tal senso l'emblema dell'unione di contributi non sempre facilmente classificabili secondo espliciti ed inequivocabili riferimenti a settori scientifico disciplinari. Ci si riferisce, cioè, alle talora troppo rigide distinzioni tra lavori puramente metodologici, ed elaborati prodotti da ambiti diversificati quali la statistica economica, la demografia, la sociologia, la psicologia. Al fine di rendere il volume il più possibile armonico, si è comunque tentato di aggregare le tematiche comuni cercando di rispondere il più possibile a queste esigenze.

Il primo contributo, di Bilancia e Sansaro, rappresenta una interessante e vasta rassegna dei principali risultati esistenti per il modello Latent Dirichlet Allocation (LDA), modello a topic latenti per l'analisi non supervisionata di un corpus di documenti testuali, con i relativi metodi computazionali necessari per la non trattabilità della distribuzione di quello che è a tutti gli effetti un modello Bayesiano gerarchico complesso.

Nel secondo lavoro, di Leone, Bilancia *et al.*, viene presentato un approccio metodologico di tipo multivariato per la mappatura digitale del suolo in un'area di studio del Sud Italia. Gli autori giungono alla conclusione secondo la quale tale approccio sembra essere efficiente e affidabile per mappare la variabilità spaziale del suolo.

Marin e Manca forniscono, invece, una interessante disamina sistematica dei modelli generativi della sentiment analysis, famiglia di metodi statistici ampiamente utilizzati a dati testuali grazie alla generalità e robustezza della metodologia.

Nel quarto contributo, di Vitale *et al.*, si torna a parlare di gestione del suolo. Gli autori, dopo aver sottolineato che tale gestione richiede una corretta valutazione delle proprietà chimiche e fisiche del suolo, spesso ottenuta tramite analisi di laboratorio convenzionali costose e i cui risultati non sono sempre ottenibili in tempi rapidi, propongono un approccio alternativo, più veloce e meno costoso, basato sull'uso di metodologia spettroscopiche e multivariate. Nell'articolo è presente uno studio applicato per stimare alcune proprietà del suolo di base, come sabbia, argilla e carbonio organico, con dati raccolti in un'area agricola del Sud Italia.

Pepe *et al.*, nel quinto lavoro del presente volume, affrontano il ruolo dei fattori stressanti e delle strategie di coping in due fasi cruciali del percorso oncologico, il periodo che precede e segue l'intervento chirurgico da carcinoma mammario della malattia oncologica. La ricerca, realizzata presso il Centro di Senologia multidisciplinare del Policlinico di Bari, con un campione di 135 donne con diagnosi di neoplasia mammaria maligna, ha dimostrato come le variazioni del distress tra il pre e post intervento siano funzione delle strategie di coping adoperate. In particolare, attraverso analisi multivariate, gli autori sono giunti ad evidenziare come le modalità di coping adattive siano predittori

statisticamente significativi della variazione del distress, determinandone l'abbassamento.

Il quinto e sesto contributo hanno come comune denominatore il territorio di riferimento, quello di Foggia. Antonucci, Crocetta e Kostiuk, illustrano i risultati di una indagine sulla efficacia dei servizi web offerti dall'Università di Foggia durante la pandemia da COVID-19. Attraverso l'uso di modelli PLS-PM, gli autori individuano le principali criticità cui dedicare una maggiore attenzione per meglio rispondere alle esigenze degli studenti dell'ateneo dauno. Antonucci, Crocetta e Russo, invece, proseguono gli studi già proposti nella edizione 2019 del presente volume, cercando di individuare i punti di forza e di debolezza per ogni servizio erogato da Sanitaservice ASL FG s.r.l., evidenziando come tale società di servizi abbia sensibilmente migliorato le proprie capacità di problem solving, mantenendo elevati standard qualitativi.

Nell'ottavo contributo si torna ad affrontare la problematica del carcinoma mammario nelle donne. Ribecco et al. confrontano due campioni di donne classificate secondo l'età in "giovani" ed "anziane" al fine di verificare se l'aggressività di sottotipi di carcinoma mammario possa essere influenzata da alcune variabili quali, ad esempio, la dimensione del tumore, il grado istologico dello stesso, il tipo di intervento.

Cariello e Carbonara, dell'ISTAT Puglia, studiano la povertà nei comuni del Mezzogiorno stimando le dimensioni del fenomeno attraverso i dati del Progetto Archivio Integrato Microdati Economici e Demografici (Arch.I.M.E.De). Questi provengono dall'integrazione di basi di dati amministrative e consentono di ottenere microdati relativi alle caratteristiche socio-demografiche e al reddito lordo delle famiglie residenti nei comuni italiani.

Nel decimo contributo, Ritrovato utilizza il "metodo statistico", cui ad inizio di questa presentazione si faceva riferimento, per analizzare le relazioni tra la crisi economica, i prezzi alimentari e la delinquenza nella Puglia di fine Ottocento, fornendo l'ennesima dimostrazione della "contaminazione" statistica negli studi di altri ambiti scientifici.

D'Uggento *et al.* affrontano, invece, l'argomento della "finanza sostenibile" effettuando un confronto tra le banche a prevalente carattere etico e sostenibile (quelle che finanziano in particolare progetti sociali, ambientali e culturali) e l'insieme delle banche europee. L'obiettivo è quello di verificare se le prime mostrino una certa solidità sotto il profilo finanziario-patrimoniale e siano anche dotate di redditività sotto il profilo economico avendo, pertanto, performance simili al complesso delle banche.

Si prosegue con Carucci e Vannella che effettuano una analisi dettagliata delle evoluzioni competitive imprenditoriali, con dettaglio dei dati a livello comunale, utilizzando una applicazione dell'analisi Shift-share su dati derivanti dall'archivio ASIA Unità Locali.

Infine, dopo il contributo in cui Gallo e Rinaldi utilizzano la metodologia DEA (Data Envelopment Analysis) per identificare la frontiera ottimale ed efficiente dei titoli relativi alle aree europee ed americane, Bisceglia fa il punto sul problema della stabilità nelle equazioni differenziali ordinarie di ordine  $k$ , realizzando anche una interessante applicazione di un modello economico.

Certi che il volume potrà riscontrare il favore dei lettori, si ringraziano tutti coloro che vi hanno contribuito, sia Autori che referee.

# Il carcinoma mammario in donne anziane e in donne giovani

Nunziata Ribecco<sup>1\*</sup>, Isabella Stasi<sup>1</sup>, Clelia Punzo<sup>2</sup>,  
Annalisa Rizzi<sup>2</sup>, Giovanni Tomasicchio<sup>2</sup>

<sup>1</sup>Dipartimento di Economia e Finanza, Università degli Studi di Bari Aldo Moro,

<sup>2</sup>Dipartimento dell’Emergenza e dei Trapianti di organi (DETO) – Bari

**Riassunto:** Il tema di ricerca presentato in questo articolo è stato sviluppato nell’ambito di una collaborazione con la sezione della Breast Unit del DETO ed ha riguardato l’analisi di dati del carcinoma mammario in donne anziane e in donne giovani. In particolare, si era interessati a individuare quali fattori prognostici fossero determinanti nella decisione terapeutica in entrambe le fasce d’età. In modo particolare, ci si è concentrati sulla variabile LUMINAL, che identifica da un punto di vista istochimico l’aggressività di sottotipi di carcinoma mammario. È stata studiata l’eventuale relazione tra questo fattore e le variabili dimensione del tumore (T), grado istologico del tumore (G), interessamento linfonodale (N), linfonodo sentinella (LS), TIPO INTERVENTO, PERMEAZIONE ENDOVASCOLARE PERINEURALE e METASTASI A DISTANZA. Pertanto, si sono analizzati due database di donne affette da carcinoma mammario, uno relativo a donne anziane (65-92 anni) e uno relativo a donne giovani (29-45 anni). Su entrambi i campioni, dopo un’iniziale analisi esplorativa per capire l’andamento di ciascun fattore, si è verificato se ci fosse una relazione tra alcune variabili ritenute più rilevanti con riferimento al fenomeno considerato. In conclusione, è stata condotta un’analisi multivariata, al fine di individuare quali variabili possono essere determinanti nelle scelte terapeutiche adottate.

## 1. Introduzione

Il carcinoma mammario rappresenta una crescita incontrollata di cellule atipiche all’interno della ghiandola mammaria. Il fenomeno del carcinoma mammario è un tema ancor oggi molto dibattuto, le cui cause non sono del tutto ben note. In generale,

---

\* Autore corrispondente: [nunziata.ribecco@uniba.it](mailto:nunziata.ribecco@uniba.it)

Il lavoro qui descritto è il risultato dell’impegno congiunto degli autori. In particolare, l’analisi statistica è stata condotta da N. Ribecco e I. Stasi mentre, C. Punzo, A. Rizzi e G. Tomasicchio oltre che mettere i dati a disposizione, hanno contribuito all’interpretazione medica dei risultati.

sono stati associati alla malattia diversi fattori di rischio come: età (la maggior parte dei casi è diagnosticata in donne di età superiore a 50 anni), prima gravidanza dopo i 30 anni, menarca prima dei 12 anni, menopausa dopo i 50 anni, nulliparità e predisposizione genetica (Fondazione Umberto Veronesi – Per il Progresso delle Scienze).

In Italia, secondo l'ultima pubblicazione AIRTUM (Associazione Italiana Registri Tumori, 2019) sui numeri del cancro, il carcinoma mammario rappresenta il tumore più frequentemente diagnosticato (considerando l'intera popolazione, donne e uomini insieme). La sopravvivenza dopo la diagnosi di tumore è uno dei principali indicatori che permette di valutare l'efficacia del sistema sanitario nei confronti della patologia tumorale. La sopravvivenza, infatti, è fortemente influenzata dalla prevenzione secondaria e dalla terapia.

Pertanto, al fine di approfondire gli aspetti del fenomeno e di determinare se la variabile LUMINAL (un indicatore dei diversi sottotipi di carcinoma derivati dall'espressione congiunta di recettori ormonali e fattori di crescita) fosse relazionata alle variabili T (dimensione del tumore), G (grado istologico del tumore), N (interessamento linfonodale), LS (linfonodo sentinella), TIPO INTERVENTO, PERMEAZIONE ENDOVASCOLARE PERINEURALE e METASTASI A DISTANZA, sono stati presi in esame i dati relativi a due campioni, uno di pazienti anziane e l'altro di pazienti giovani.

## **2. Materiali e metodi**

Il primo campione è composto da 150 pazienti, la cui età varia tra i 65 ed i 92 anni, affette da carcinoma mammario e sottoposte a trattamento chirurgico, dal 2006 al 2020, mentre il secondo campione è composto da 38 pazienti la cui età varia tra i 29 ed i 45 anni, sempre affette da carcinoma mammario, sottoposte a trattamento chirurgico dal 2006 al 2019. I dati sono stati messi a disposizione dalla prof.ssa Clelia Punzo in quanto rientrano nella sua casistica personale e riguardano donne sottoposte a trattamento chirurgico presso il reparto U.O.C. Chirurgia Bonomo (2006-2018) – U.O.S.D. Chirurgia Videolaparoscopica (2019 per le donne giovani e 2019-2020 per le donne anziane), Università degli Studi di Bari. Al fine di analizzare i dati, entrambi i database sono stati sottoposti ad alcune modifiche che si sono rese necessarie in quanto vi erano pazienti per le quali si presentavano molti dati mancanti. La soluzione adottata è stata molto accurata affinché non fossero perse informazioni importanti e non fosse compromessa la bontà dei risultati derivati dall'analisi. Per ragioni ovvie, si sono anche omessi i nomi delle pazienti.

Inizialmente sono state individuate le variabili di interesse e, ai fini dell'analisi scelta, sono state trasformate da variabili politomiche in variabili dicotomiche (sì, no; presente, assente; ecc. ...). Pertanto, la variabile PERM. ENDOVASCOLARE PERINEURALE è stata sintetizzata in due modalità SI (presenza) e NO (assenza), anche la variabile TIPO INTERVENTO è stata dicotomizzata nelle seguenti modalità: CC (chirurgia conservativa) e CD (chirurgia demolitiva). Inoltre, le modalità delle variabili LS e METASTASI A DISTANZA sono state indicate con delle etichette come di seguito specificato:

- LS con modalità CI (cellule isolate), MM (micro-metastasi), NEG (negativo), POS (positivo) e NO (non eseguito o perché forme in situ, < 2 cm, o perché forme di dimensioni maggiori e con interessamento clinico dei linfonodi);
- METASTASI A DISTANZA con modalità NO (assenti), P (polmonari), L (linfonodi), LO (linfonodi ossei), LOE (linfonodi ossei epatici), LSE (linfonodali e sospette all'encefalo), ECP (epatiche carcinosi peritoneale) e NDES (non descritte).

Le etichette delle modalità delle restanti variabili sono di seguito indicate:

- G con modalità G1 (grado istologico basso), G2 (grado istologico intermedio) e G3 (grado istologico elevato);
- T con modalità: T1 (fino a 2 cm), T1a ( $\leq 0.5$  cm), T1b (tra 0.5 cm e 1cm), T1c (tra 1 cm e 2 cm), T2 (tra 2 cm e 5 cm), T3 ( $> 5$  cm) e con dimensioni Tis (carcinoma in situ) e T4 che corrisponde a dimensioni molto elevate così definite: estensione diretta alla parete toracica e/o alla cute;
- N con modalità N0 (linfonodi liberi da metastasi), N1 (metastasi nei linfonodi ascellari omolaterali mobili), N1a (metastasi in 1-3 linfonodi ascellari con almeno un deposito  $> 0.2$  cm), N1mic (micro metastasi), N2 (metastasi nei linfonodi ascellari omolaterali fissi), N2a (metastasi nei linfonodi ascellari omolaterali fissi tra di loro o ad altre strutture), N3 (metastasi in uno o più linfonodi omolaterali), N3a (metastasi nei linfonodi sottoclaveari omolaterali), NX (linfonodi non valutabili).

Considerato che lo scopo dello studio è stato quello di evidenziare quali variabili fossero determinanti nelle scelte terapeutiche, dopo un'analisi esplorativa finalizzata a conoscere l'andamento di ciascuna variabile, si è passati a verificare se esiste indipendenza fra le variabili più significative ai fini dell'analisi. Pertanto, per il campione di donne anziane, essendo la numerosità campionaria discretamente elevata, è stata effettuata una verifica d'ipotesi d'indipendenza tramite il test Chi-Quadro, mentre per il campione di donne giovani, poiché la numerosità campionaria non è molto elevata e le frequenze teoriche sono risultate spesso minori di 5, si è fatto ricorso, prevalentemente, al test esatto di Fisher (Wayne W. D., 2012).

Allorché dalla verifica d'ipotesi è risultata una significativa relazione fra le variabili sottoposte ad analisi (rifiuto dell'ipotesi nulla di indipendenza) attraverso il "corrplot", ovvero il grafico della matrice di relazione fra le variabili, è stato possibile evidenziare, per le variabili maggiormente legate tra loro, quali sono le modalità che contribuiscono a determinare, in maniera significativa, questa relazione. Le relazioni positive sono visualizzate in cerchi di colore blu ed indicano attrazione (relazione positiva), mentre le relazioni negative sono visualizzate in cerchi di colore rosso ed indicano repulsione (relazione negativa). Inoltre, l'intensità del colore e la dimensione del cerchio sono proporzionali ai contributi, ossia un colore più intenso ed un cerchio più grande evidenziano un valore elevato che è tanto più elevato quanto più aumenta l'intensità del colore e la dimensione del cerchio. Mentre, un colore meno intenso ed una circonferenza più piccola indicano un valore meno elevato che diventa sempre più piccolo al ridursi dell'intensità e della dimensione della circonferenza (Kassambara, A., 2017).

Infine, si è fatto ricorso all'Analisi delle Corrispondenze Multiple (ACM), tecnica multivariata che, vista la natura dei dati, è la più idonea a raggiungere l'obiettivo preposto. Tale tecnica, in termini generali, si propone di descrivere la struttura delle relazioni sottese alla matrice dei dati oggetto di studio attraverso la collocazione e l'analisi dei punti-modalità delle variabili in uno spazio geometrico-statistico di dimensione ridotta.

L'obiettivo è identificare sia un gruppo di individui con un profilo simile nelle loro risposte alle domande, sia le associazioni tra categorie di variabili. Nell'analisi in oggetto abbiamo scelto di focalizzare lo studio solo sulle variabili e sulle proprie modalità, poiché il nostro scopo è analizzare l'evoluzione del carcinoma mammario sulla base dei fattori che maggiormente spiegano il fenomeno, tralasciando gli individui.

Per le analisi, è stato utilizzato il software statistico R.

### **3. Risultati dell'analisi**

#### *3.1 Analisi su donne anziane*

##### *3.1.1 Studio delle relazioni tra le variabili*

Dopo aver studiato il comportamento delle singole variabili mediante l'analisi esplorativa, siamo passati a valutare se tra le variabili ritenute più importanti ai fini dello studio del fenomeno ci fosse una qualche relazione. Pertanto, abbiamo considerato

la verifica d'ipotesi di indipendenza su tabelle di contingenza a partire dal seguente sistema d'ipotesi:

$$\begin{cases} H_0: \text{le variabili sono indipendenti} \\ H_1: \text{le variabili non sono indipendenti} \end{cases}$$

Di seguito, riportiamo i risultati delle verifiche d'ipotesi di indipendenza con riferimento alle variabili prese in esame (Tab. 1):

**Tabella 1.** Variabili analizzate, valore del test chi-quadro, p-value e decisione presa a seguito del risultato ottenuto.

<i>Variabili analizzate</i>	<i>Valore test <math>\chi^2</math></i>	<i>p-value</i>	<i>Decisione</i>
LUMINAL vs T	0.86	0.35	Accettiamo $H_0$ : le variabili sono indipendenti.
LUMINAL vs G	3.79	<b>0.05</b>	<b>Rifiutiamo <math>H_0</math>: le variabili non sono indipendenti</b>
LUMINAL vs N	0.01	0.91	Accettiamo $H_0$ : le variabili sono indipendenti
LUMINAL vs LS	3.79	<b>0.05</b>	<b>Rifiutiamo <math>H_0</math>: le variabili non sono indipendenti</b>
LUMINAL vs TIPO INTERVENTO	0.29	0.59	Accettiamo $H_0$ : le variabili sono indipendenti
LUMINAL vs PERM. ENDOVASCOLARE PERINEURALE	7.05	<b>0.008</b>	<b>Rifiutiamo <math>H_0</math>: le variabili non sono indipendenti</b>
LUMINAL vs METASTASI A DISTANZA	0.06	0.80	Accettiamo $H_0$ : le variabili sono indipendenti

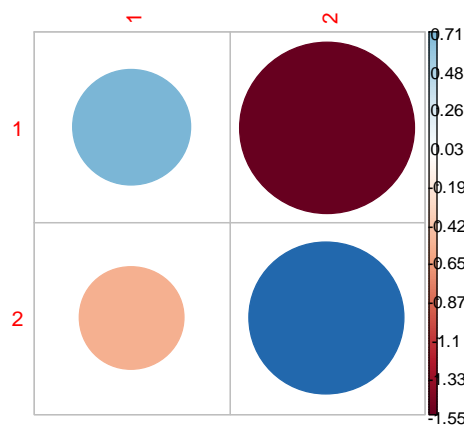
Osservando la Tab. 1, possiamo notare che le coppie di variabili risultate non significative ai fini della verifica di ipotesi, sono le seguenti: LUMINAL e T, LUMINAL ed N, LUMINAL e TIPO INTERVENTO, LUMINAL e METASTASI A DISTANZA. Al contrario, le coppie di variabili risultate significative sono: LUMINAL e G, LUMINAL e LS e LUMINAL e PERM. ENDOVASCOLARE PERINEURALE.

Pertanto, con riferimento alle coppie di variabili per cui abbiamo rifiutato l'ipotesi nulla possiamo ritenere che esiste una relazione la cui intensità potrà essere evidenziata graficamente tramite i “*corrplot*” di seguito riportati.

Come già specificato nel primo paragrafo, il “*corrplot*” è uno strumento grafico che permette di valutare in maniera immediata le relazioni fra le variabili tramite sia la colorazione delle circonferenze che la dimensione.

Analizziamo il “*corrplot*” relativo alle prime coppie di variabili risultate significativamente non indipendenti: LUMINAL e G (Fig. 1).

**Figura 1.** “*Corrplot*” delle variabili LUMINAL e G.

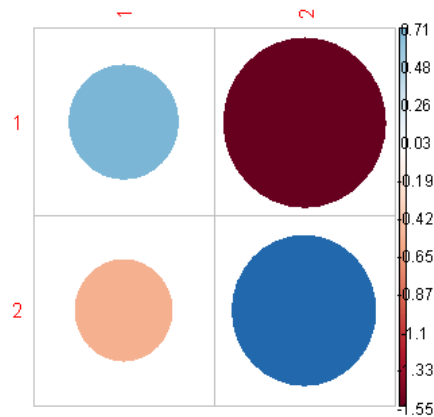


Osservando il “*corrplot*” appare immediatamente evidente che esiste un'associazione fortemente positiva tra GRADO ISTOLOGICO ELEVATO (2 in riga) e TPA (2 in colonna), seguita da una associazione sempre positiva ma meno forte tra GRADO ISTOLOGICO MEDIO-BASSO (1 in riga) e TMA (1 in colonna). Pertanto, sembrerebbe che all'aumentare dei casi in cui il grado istologico è **elevato** corrisponde un incremento di casi in cui il tumore assume forme più aggressive (TPA) mentre, all'aumentare dei casi in cui il grado istologico è **medio-basso** si osserva un aumento di casi in cui il tumore assume forme meno aggressive (TMA).

È evidente che esiste una relazione fortemente negativa tra GRADO ISTOLOGICO MEDIO-BASSO (1 in riga) e TPA (2 in colonna), seguita da una relazione sempre negativa ma meno forte tra GRADO ISTOLOGICO ELEVATO (2 in riga) e TMA (1 in colonna). Questo porta ad ipotizzare che all'aumento dei casi in cui il grado istologico è **medio-basso** corrisponde una riduzione di casi in cui il tumore assume forme più aggressive (TPA) mentre, all'aumentare dei casi in cui il grado istologico è **elevato** si presenta una riduzione dei casi in cui il tumore assume forme meno aggressive (TMA).

Con riferimento alle variabili LUMINAL e LS il “*corrplot*” seguente:

**Figura 2.** “*Corrplot*” delle variabili LUMINAL e LS.



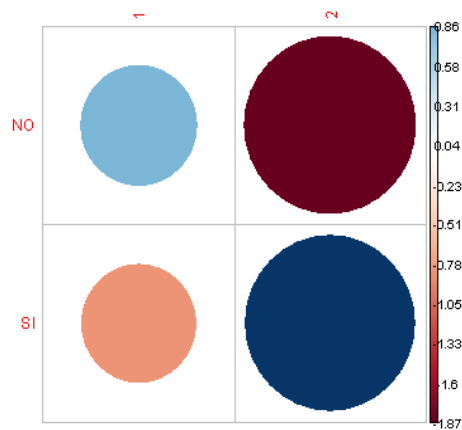
evidenzia la presenza di una relazione fortemente positiva tra LINFONODO SENTINELLA NON ESEGUITO (2 in riga) e TPA (2 in colonna), seguita da una relazione sempre positiva ma meno forte tra LINFONODO SENTINELLA ESEGUITO (1 in riga) e TMA (1 in colonna). Pertanto, sembrerebbe che all'aumentare dei casi in cui il linfonodo sentinella **non** è **eseguito**, vi sia un aumento di casi in cui il tumore assume forme più aggressive (TPA). Inoltre, all'aumento dei casi in cui il linfonodo sentinella è **eseguito**, corrisponde un aumento di casi in cui il tumore assume forme meno aggressive (TMA). Si può ritenere, quindi, che esiste una relazione fortemente negativa tra LINFONODO SENTINELLA ESEGUITO (1 in riga) e TPA (2 in colonna), seguita da una relazione sempre negativa ma meno forte tra LINFONODO SENTINELLA NON ESEGUITO (2 in riga) e TMA (1 in colonna). Possiamo concludere, quindi, che all'aumentare dei casi in cui il linfonodo sentinella è **eseguito**, corrisponde una riduzione di casi in cui il tumore assume forme più aggressive (TPA), mentre all'aumentare dei casi in cui il linfonodo sentinella **non** è **eseguito**, appare una riduzione di casi in cui il tumore assume forme meno aggressive (TMA).

Con riferimento alle variabili LUMINAL e PERM. ENDOVASCOLARE PERINEURALE appare evidente dal grafico (Fig. 3) che esiste una relazione positiva alquanto forte tra presenza di PERM. ENDOVASCOLARE PERINEURALE (SI) e TPA (2 in colonna), seguita da una relazione sempre positiva ma meno forte tra assenza DI PERM. ENDOVASCOLARE PERINEURALE (NO) e TMA (1 in colonna). Ossia, sembrerebbe che una **maggiore presenza** di PERM. ENDOVASCOLARE PERINEURALE (SI), comporti un aumento di casi in cui il tumore assume

forme più aggressive (TPA) mentre ad una **ridotta presenza** di PERM. ENDOVASCOLARE PERINEURALE (NO) corrisponde un aumento di casi in cui il tumore assume forme meno aggressive (TMA).

È evidente, inoltre, che esiste una relazione fortemente negativa tra assenza di PERM. ENDOVASCOLARE PERINEURALE (NO) e TPA (2 in colonna), seguita da una relazione sempre negativa ma meno forte tra presenza di PERM. ENDOVASCOLARE PERINEURALE (SI) e TMA (1 in colonna). Quindi si può ritenere che la presenza o meno di PERM. ENDOVASCOLARE PERINEURALE (SI/NO), comporta una riduzione di casi in cui il tumore assume forme più aggressive (TPA) ed una riduzione di casi in cui il tumore assume forme meno aggressive (TMA).

**Figura 3.** “Corrplot” delle variabili LUMINAL e PERM. ENDOVASCOLARE PERINEURALE



### 3.1.3 Analisi multivariata delle corrispondenze multiple

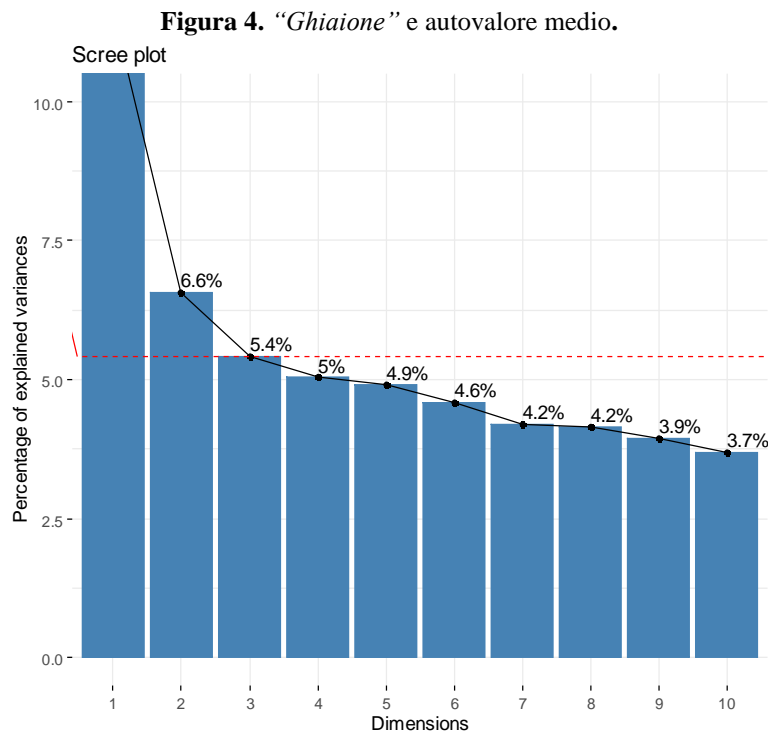
Al fine di individuare quali variabili contribuiscono maggiormente a spiegare la variabilità del fenomeno abbiamo fatto ricorso all'Analisi delle Corrispondenze Multiple. In questo paragrafo, dopo aver descritto le modalità seguite per individuare le variabili, passeremo ad analizzare i risultati ottenuti con riferimento al fenomeno oggetto di studio.

#### Fase 1: scelta del numero di assi da considerare

Per scegliere il numero di dimensioni da conservare per l'interpretazione dei dati non esiste una "regola empirica". Infatti, la scelta viene effettuata tramite il diagramma denominato “*ghiaione*” attraverso il quale è possibile individuare quali dimensioni contribuiscono a spiegare meglio il fenomeno. Il grafico di seguito riportato (Fig. 4)

raffigura sull'asse delle ascisse le diverse dimensioni e su quello delle ordinate la variabilità spiegata. Pertanto, le barre evidenziano la variabilità spiegata da ciascuna dimensione mentre, il punto in cui la spezzata interpolante assume la forma di una curva (il cosiddetto "*gomito*") può essere considerato come indice di una dimensionalità ottimale. Viene, inoltre, calcolato un autovalore medio che rappresenta il punto al disotto del quale i contributi per spiegare la variabilità sono considerati poco rilevanti, mentre gli assi la cui variabilità spiegata si posiziona al disopra di questo punto sono considerati importanti ai fini dell'analisi e vengono inclusi nella soluzione per l'interpretazione dei dati. (Kassambara A., 2017).

Nel grafico ("*ghiaione*") (Fig. 4) la linea rossa tratteggiata è in corrispondenza dell'autovalore medio.

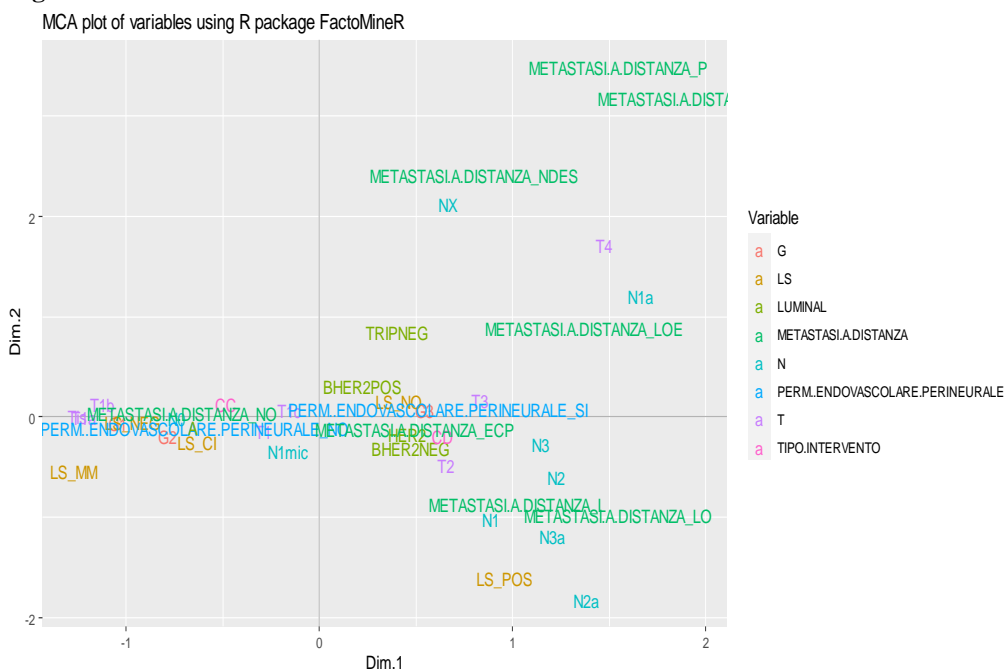


Dall'osservazione del grafico si evince che vanno prese in considerazione solo le dimensioni 1 e 2 in quanto spiegano, rispettivamente, circa l'11.6% e il 6.6% dell'inerzia totale, per un totale cumulativo del 18.2%. La dimensione 3 spiega solo il 5.4% dell'inerzia totale ed è inferiore al valore medio di equivalenza (5.41%) pertanto non viene presa in considerazione per le analisi successive.

### Fase 2: coordinate delle modalità delle variabili

In questa fase si proiettano le modalità delle variabili sugli assi e si osserva la loro collocazione rispetto a questi. L'origine degli assi corrisponde alla media ponderata dei punteggi riferiti alle modalità delle variabili nella rappresentazione fattoriale che vengono definiti profili medi. I punti del grafico che si avvicinano all'origine sono quelli i cui profili sono maggiormente somiglianti con quelli medi. Analogamente i punti che presentano profili che si discostano maggiormente da quelli medi si situano in posizione periferica (Kassambara A., 2017).

**Figura 5.** *Proiezione sull'asse 1 e 2 delle variabili analizzate.*



La Figura 5 è il risultato della proiezione sui due assi delle variabili esaminate. In particolare possiamo evincere che, fatta eccezione per le modalità delle variabili METASTASI A DISTANZA P, METASTASI A DISTANZA LSE, METASTASI A DISTANZA NDES, NX, T4, N1a, TRIPNEG, METASTASI A DISTANZA LOE, N2a, LS POS, N3a, N1, METASTASI A DISTANZA, L e METASTASI A DISTANZA LO che sembrano isolarsi e non essere somiglianti alle altre, tutte le restanti risultano somiglianti e presentano un comportamento alquanto analogo come si può dedurre osservando il loro concentrarsi intorno all'origine degli assi.

In particolare, osserviamo che le modalità che nel grafico sono in posizioni vicine tra loro sono corrispondenti a quelle variabili che sembrano essere in relazione tra

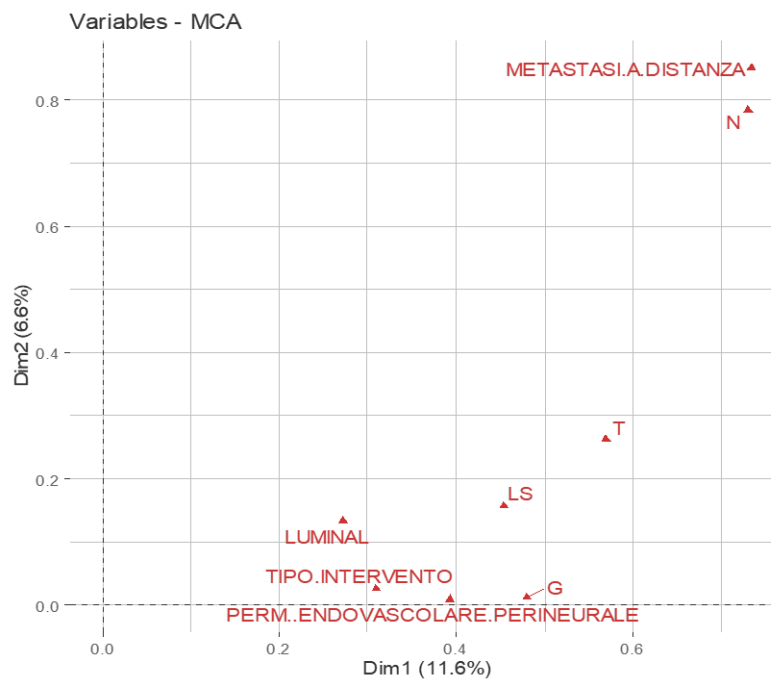
loro. Infatti, a destra dell'origine degli assi troviamo le modalità LUMINAL HER2, LUMINAL BHER2NEG, CD, METASTASI A DISTANZA ECP, T3, G3, PERM. ENDOVASCOLARE PERINEURALE SI e LS NO. Mentre, a sinistra dell'origine troviamo le modalità G2, PERM. ENDOVASCOLARE PERINEURALE NO, LS NEG, METASTASI A DISTANZA NO, N0, T1 e CC.

Le zone maggiormente concentrate inglobano le modalità posizionate nel centro del grafico. Tale rappresentazione grafica non ha l'obiettivo di determinare quali variabili contribuiscono a spiegare maggiormente la variabilità del fenomeno pertanto è opportuno proseguire con l'analisi al fine di individuare le variabili più significative ai fini dell'analisi.

### Fase 3: contributo delle variabili nella definizione delle dimensioni

Una volta stabilito il numero di assi da considerare e dopo aver visualizzato le coordinate delle categorie, il riferimento prioritario va ai contributi delle variabili, cioè alla percentuale di inerzia spiegata dalla modalità della variabile, in corrispondenza di ciascun asse. L'obiettivo è determinare quali variabili contribuiscono maggiormente alla definizione delle diverse dimensioni mantenute nel modello (Kassambara A., 2017).

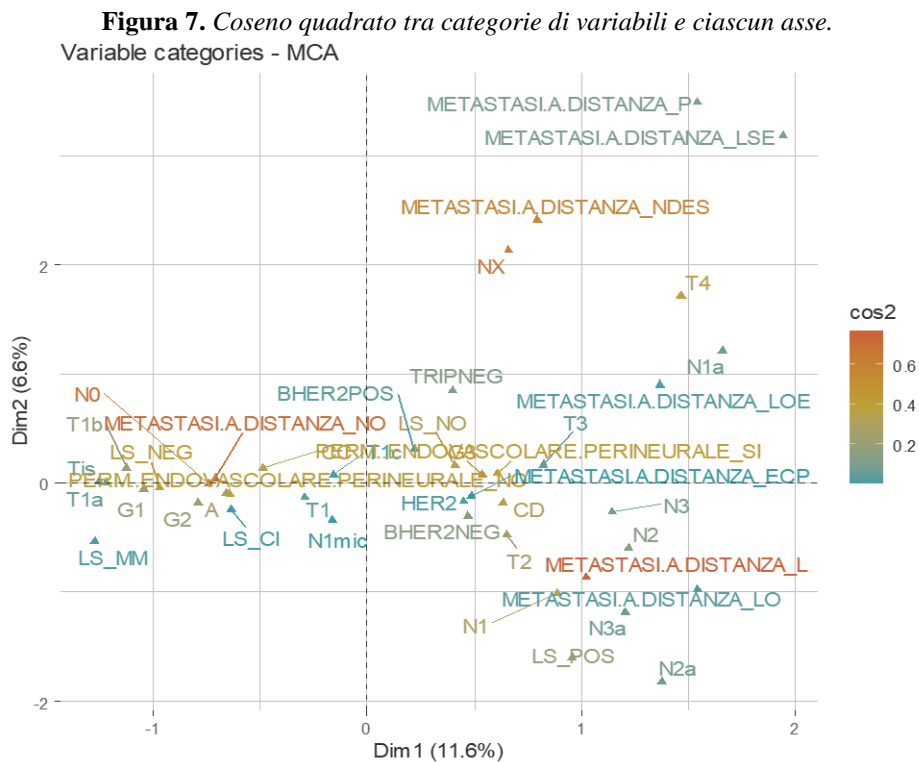
**Figura 6.** Variabili che contribuiscono maggiormente a definire le dimensioni del modello.



Dalla Fig. 6 si osserva che le variabili PERM. ENDOVASCOLARE PERINEURALE, G e TIPO INTERVENTO sono le più correlate alla dimensione 1. Al contrario, il grafico non sembra evidenziare delle variabili correlate alla dimensione 2. Inoltre, ciò che appare è che le variabili METASTASI A DISTANZA e N si discostano in maniera evidente da tutte le altre, creando una nuvola autonoma all'interno del grafico, ciò ci fa pensare che abbiano un andamento indipendente e autonomo dalle altre variabili.

#### Fase 4: qualità della rappresentazione di categorie variabili

La qualità della rappresentazione è chiamata “*coseno quadrato*” (**cos2**) e misura l'intensità della relazione tra categorie di variabili e un asse particolare, consentendo di individuare quale fattore descrive meglio la variabilità del fenomeno. È possibile colorare le categorie di modalità di variabili in base alla loro qualità. (Kassambara, A., 2017).



Dalla Fig. 7 osserviamo che presentano una relazione più forte con la Dimensione 1 quelle modalità che su questa direzione assumono una tonalità di colore prossimo

all'arancio ovvero N0, METASTASI A DISTANZA NO, LS NEG, PERM. ENDOVASCOLARE PERINEURALE NO, PERM. ENDOVASCOLARE PERINEURALE SI e G3. Come evidenziato in precedenza, non ci sono variabili in relazione con la Dimensione 2.

Sostanzialmente, possiamo ritenere che le modalità che meglio descrivono la variabilità del fenomeno sono:

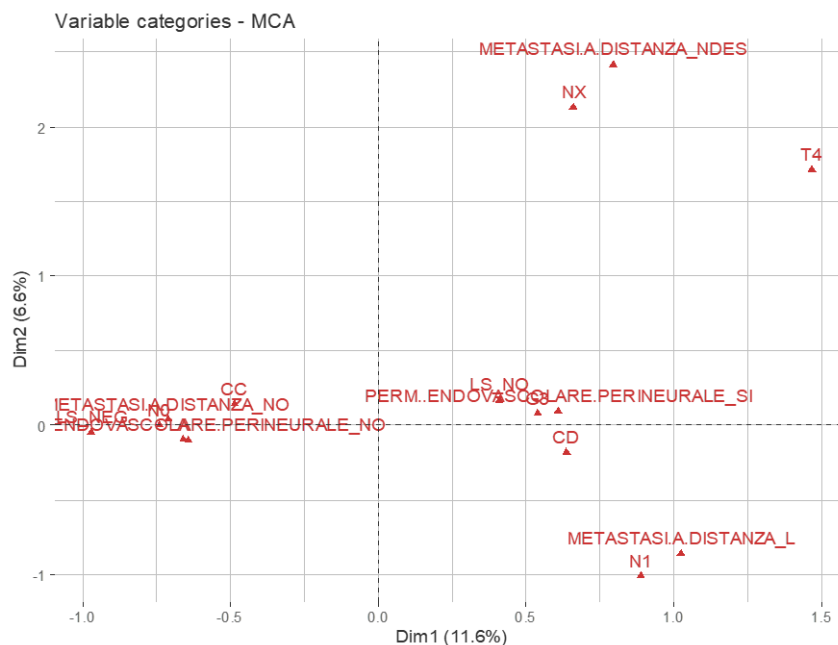
- linfonodi liberi da metastasi;
- assenza di metastasi a distanza;
- linfonodo sentinella negativo;
- assenza di permeazione endovascolare perineurale;
- presenza di permeazione endovascolare perineurale;
- grado istologico elevato.

Fase 5: scelta finale delle modalità più rappresentative

L'analisi congiunta dei contributi e dei coseni quadrati guida la scelta delle modalità da ritenere più significative ai fini della sintesi. Infatti, avendo molte categorie di variabili, è possibile visualizzarne solo alcune, le più importanti sulla base del loro indice di qualità (Kassambara, A., 2017).

Visualizziamo le prime 15 modalità di variabili con un valore più alto di  $\cos^2$ .

**Figura 8. Modalità più rappresentative.**



Nella Fig.8 ritroviamo le stesse modalità delle variabili che fino a questo momento abbiamo ritenuto fornissero un contributo importante per la spiegazione della variabilità dei dati.

In conclusione, escludendo le due modalità che non esplicitano alcun contributo (NX e METASTASI A DISTANZA NDES) in quanto modalità non valutabili, possiamo ritenere che le prime 13 modalità che contribuiscono a spiegare la variabilità del fenomeno sono:

- METASTASI A DISTANZA L (metastasi a distanza linfonodali);
- METASTASI A DISTANZA NO (assenza di metastasi a distanza);
- N1 (metastasi nei linfonodi ascellari omolaterali mobili);
- NO (linfonodi liberi da metastasi);
- LS NEG (linfonodo sentinella negativo);
- LS NO (linfonodo sentinella non eseguito);
- CC (chirurgia conservativa);
- CD (chirurgia demolitiva);
- PERM. ENDOVASCOLARE PERINEURALE SI (presenza di permeazione endovascolare perineurale);
- PERM. ENDOVASCOLARE PERINEURALE NO (assenza di permeazione endovascolare perineurale);
- G3 (grado istologico elevato);
- T4 (estensione diretta alla parete toracica e/o alla cute);
- LUMINAL A.

In modo particolare, possiamo ritenere che in presenza di permeazione endovascolare perineurale e di linfonodo sentinella non eseguito, si rileva un carcinoma mammario più aggressivo, nel quale le cellule tumorali hanno un aspetto anomalo, crescono più rapidamente e si diffondono a distanza. In questo caso, infatti, le pazienti si caratterizzano per un grado istologico molto elevato (G3) e per la chirurgia demolitiva (CD). In contrapposizione, l'assenza di metastasi a distanza, i linfonodi liberi da metastasi, il linfonodo sentinella negativo, l'assenza di permeazione endovascolare perineurale, sono indicatori di una situazione meno grave, in cui le cellule tumorali crescono lentamente e difficilmente si diffondono a distanza. In questo caso, infatti, le pazienti si caratterizzano per il luminal di tipo A e per la chirurgia conservativa (CC).

Notiamo, inoltre, che le metastasi a distanza linfonodali si posizionano vicino alla modalità N1 evidenziando una relazione.

### 3.2 Analisi su donne giovani

#### 3.2.1 Studio delle relazioni tra le variabili

Anche in questo caso, dopo aver studiato il comportamento delle singole variabili mediante l'analisi esplorativa, siamo passati a valutare se tra le variabili ritenute più importanti ai fini dello studio del fenomeno ci fosse una qualche relazione. Pertanto, abbiamo considerato la verifica d'ipotesi di indipendenza su tabelle di contingenza a partire dal seguente sistema d'ipotesi:

$$\begin{cases} H_0: \text{le variabili sono indipendenti} \\ H_1: \text{le variabili non sono indipendenti} \end{cases}$$

Essendo la numerosità del campione non molto elevata ed avendo osservato in diversi casi che le frequenze teoriche erano più piccole di 5 si è fatto ricorso, per la verifica d'ipotesi di indipendenza, al test esatto di Fisher. Questa soluzione ha richiesto, inoltre, la dicotomizzazione di alcune variabili che originariamente erano poliotomiche. Tale scelta, comunque, è stata effettuata valutando attentamente che le esigenze di analisi fossero supportate da giustificazioni a carattere clinico.

Di seguito, riportiamo i risultati delle verifiche d'ipotesi di indipendenza con riferimento alle variabili prese in esame (Tab. 2):

**Tabella 2.** Variabili analizzate, valore del p-value riferito al test esatto di Fisher e decisione presa a seguito del risultato ottenuto.

Variabili analizzate	p-value	Decisione
LUMINAL vs T	0.20	Accettiamo $H_0$ : le variabili sono indipendenti
LUMINAL I vs G	<b>0.02</b>	<b>Rifiutiamo <math>H_0</math>: le variabili non sono indipendenti</b>
LUMINAL vs N	0.35	Accettiamo $H_0$ : le variabili sono indipendenti
LUMINAL vs LS	<b>0.02</b>	<b>Rifiutiamo <math>H_0</math>: le variabili non sono indipendenti</b>
LUMINAL vs TIPO INTERVENTO	0.17	Accettiamo $H_0$ : le variabili sono indipendenti
LUMINAL vs PERM. ENDOVASCOLARE PERINEURALE	0.66	Accettiamo $H_0$ : le variabili sono indipendenti

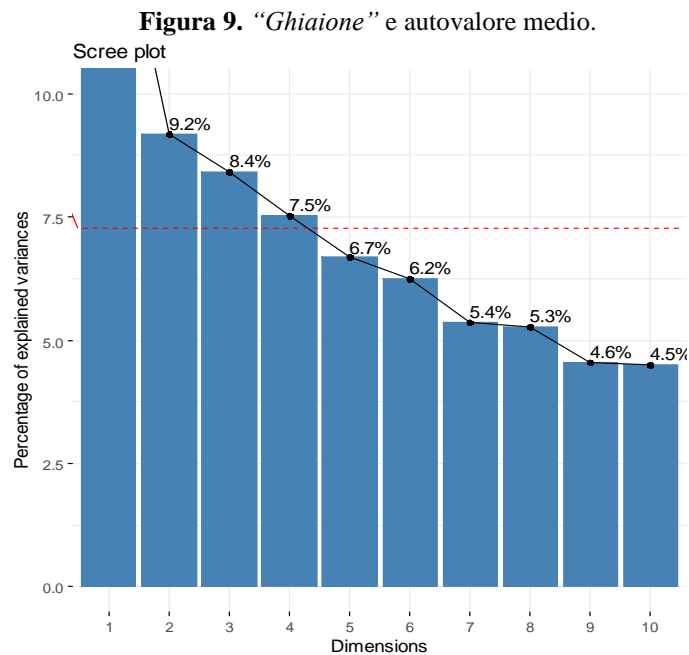
Dalla Tab. 2 si evince che le coppie di variabili risultate non significative sono le seguenti: LUMINAL e T, LUMINAL e N, LUMINAL e TIPO INTERVENTO e LUMINAL e PERM. ENDOVASCOLARE PERINEURALE. Mentre, possiamo osservare che, a differenza di quello che si è riscontrato nelle donne anziane, le uniche coppie di variabili in relazione fra loro sono le seguenti: LUMINAL e G e LUMINAL e LS.

Procedendo in maniera analoga a quanto fatto per le donne anziane, consideriamo, nel paragrafo seguente, i risultati dell'analisi multivariata.

### 3.2.2 Analisi multivariata delle corrispondenze multiple

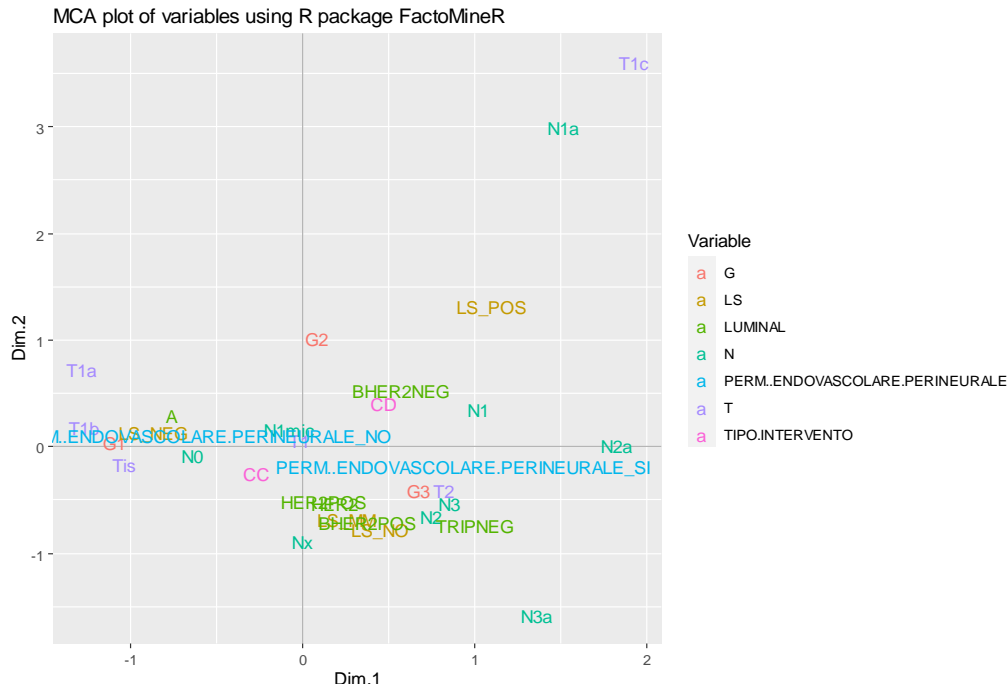
#### *Fase 1: scelta del numero di assi da considerare*

Innanzitutto, esaminiamo il diagramma "ghiaione", in cui l'autovalore medio è in corrispondenza della linea tratteggiata rossa.



Dal grafico (Fig. 9) si evince che solo le prime quattro dimensioni spiegano, rispettivamente, circa il 14.9%, il 9.2%, l'8.4% e il 7.5% dell'inerzia totale, per un totale cumulativo del 40%. La quinta dimensione spiega il 6.7% dell'inerzia totale, inferiore al valore medio di equivalenza (7.27%), pertanto non viene considerata nelle ulteriori analisi.

**Figura 10.** *Proiezione sull'asse 1 e 2 delle variabili analizzate.*



*Fase 2: coordinate delle modalità delle variabili*

Dalla Fig. 10 possiamo evincere che, eccetto per le modalità di variabili N3a, T1c, N1a, T1a, G2 e LS POS che sembrano isolarsi e non essere somiglianti alle altre, tutte le altre tra loro hanno un comportamento molto simile, infatti si concentrano intorno all'origine degli assi.

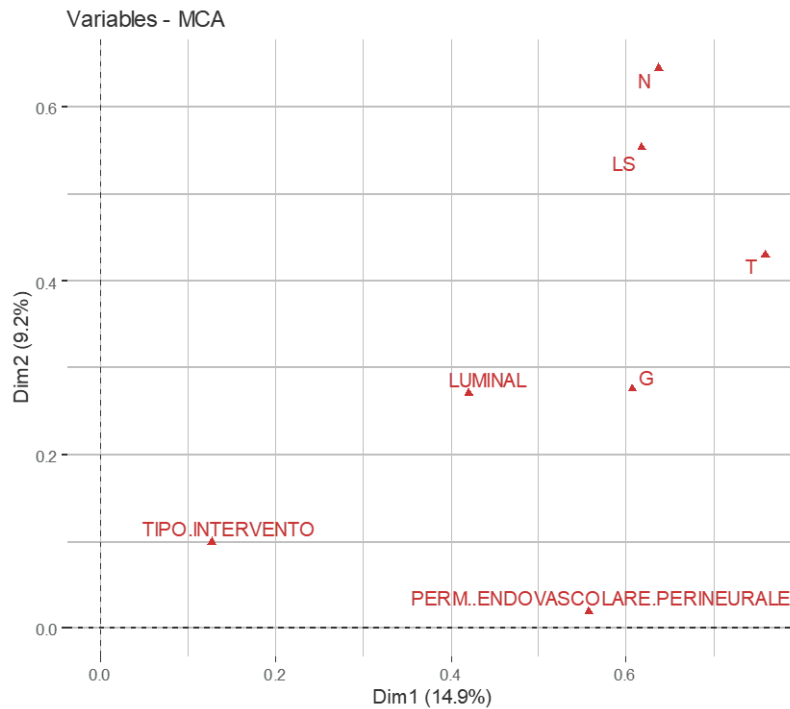
In particolare, le modalità di variabili che più sembrano avere una relazione tra loro poiché risultano vicine nel grafico sono: PERM. ENDOVASCOLARE PERINEURALE SI, G3, T2, N3, N2, LUMINAL TRIPNEG, LUMINAL HER2, LUMINAL HER2POS, LUMINAL BHER2POS e LS NO posizionate a destra rispetto all'origine. Mentre, a sinistra troviamo le modalità G1, PERM. ENDOVASCOLARE PERINEURALE NO, LS NEG, N0, N1mic, T1b, Tis, LUMINAL A e CC. Inoltre, risultano somiglianti tra loro le modalità CD e LUMINAL BHER2NEG.

Notiamo che le zone più concentrate inglobano le modalità posizionate nel centro del grafico ma non possiamo stabilire quali modalità di variabili contribuiscono maggiormente a determinare la variabilità del fenomeno in quanto questo grafico non ha questo scopo.

### Fase 3: contributo delle variabili nella definizione delle dimensioni

Osserviamo i contributi delle variabili, cioè la percentuale di inerzia spiegata dalla modalità della variabile, in corrispondenza di ciascuno degli assi.

**Figura 11.** Variabili che definiscono maggiormente le dimensioni del modello.



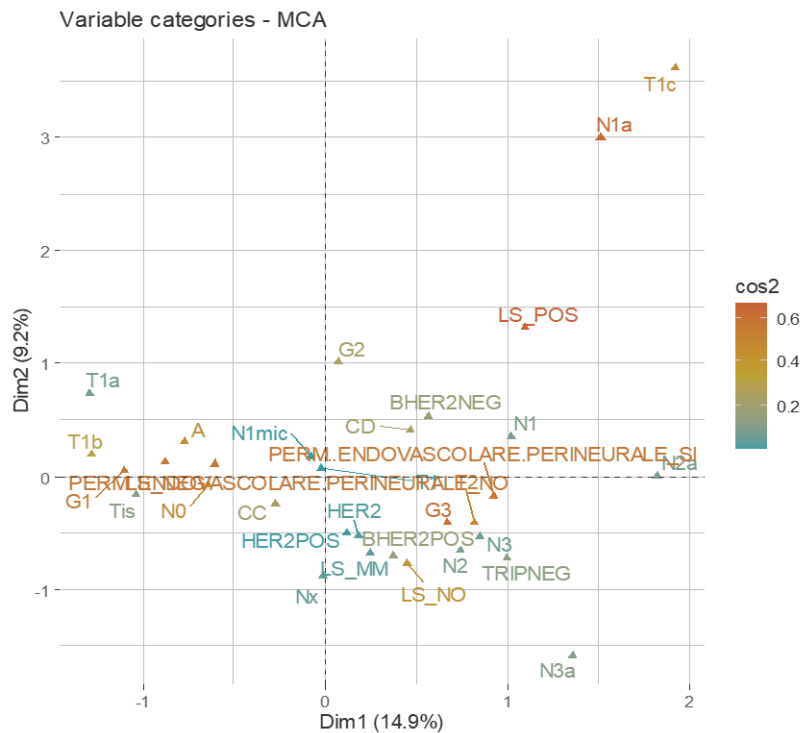
Dalla Fig. 11 si osserva che la variabile PERM. ENDOVASCOLARE PERINEURALE è la più correlata alla dimensione 1. Al contrario, il grafico non sembra evidenziare delle variabili correlate alla dimensione 2. Inoltre, ciò che appare è che le variabili N, LS e T si discostano in maniera evidente da tutte le altre, creando una nuvola autonoma all'interno del grafico e ciò ci fa pensare che abbiano un andamento indipendente e autonomo dalle altre.

### Fase 4: qualità della rappresentazione di categorie variabili

La Fig. 12 evidenzia che le modalità di variabili più associate alla dimensione 1 sono quelle che su questa direzione si presentano con una tonalità di colore sull'arancio ovvero N0, LS NEG, PERM. ENDOVASCOLARE PERINEURALE NO, PERM. ENDOVASCOLARE PERINEURALE SI, G1, G3, T1b, T2 e LUMINAL A. Come

evidenziato in precedenza, non ci sono modalità di variabili associate alla dimensione 2.

**Figura 12.** Coseno quadrato tra categorie di variabili e ciascun asse.



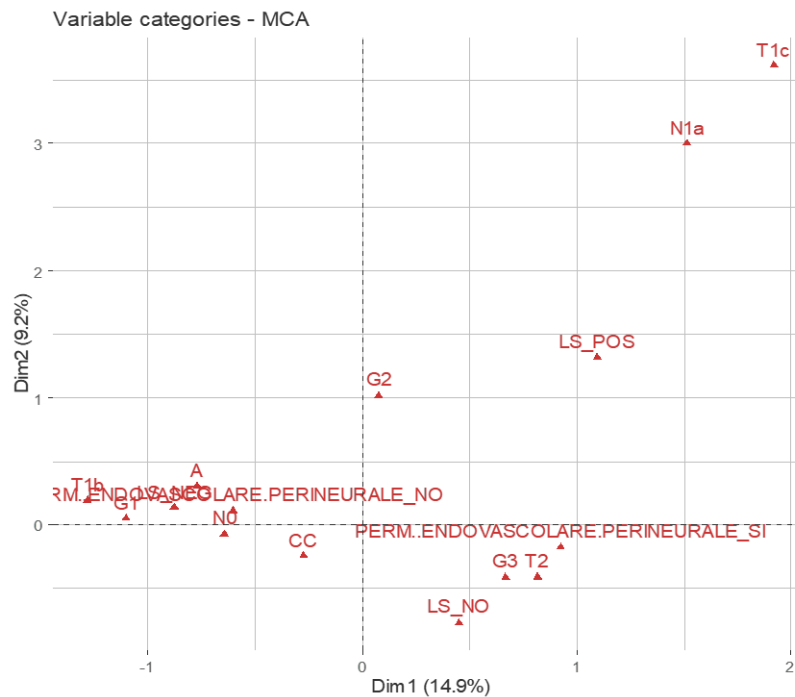
Sostanzialmente, possiamo ritenere che le modalità di variabili che meglio descrivono la variabilità del fenomeno secondo la loro qualità sono:

- linfonodi liberi da metastasi;
- linfonodo sentinella negativo;
- assenza di permeazione endovascolare perineurale;
- presenza di permeazione endovascolare perineurale;
- grado istologico basso;
- grado istologico elevato;
- dimensione del tumore compresa tra 0.5 cm e 1cm;
- dimensione del tumore compresa tra 2 cm e 5 cm nella dimensione massima;
- luminal A.

Fase 5: scelta finale delle modalità più rappresentative

Decidiamo di visualizzare le prime 15 modalità di variabili con un valore più alto di  $\cos^2$ .

**Figura 13.** Modalità più rappresentative.



Nella Fig. 13 ritroviamo le stesse modalità delle variabili che fino a questo momento abbiamo ritenuto fornissero un contributo importante alla spiegazione della variabilità dei dati.

In conclusione, possiamo ritenere che le prime 15 modalità che contribuiscono a spiegare la variabilità del fenomeno sono:

- T1c (dimensione tra 1 cm e 2 cm);
- T1b (dimensione tra 0.5 cm e 1cm);
- T2 (dimensione tra 2 cm e 5 cm);
- N1a (metastasi in 1-3 linfonodi ascellari con almeno un deposito > 0.2 cm);
- N0 (linfonodi liberi da metastasi);
- G1(grado istologico basso);
- G2 (grado istologico intermedio);
- G3(grado istologico elevato);

- LS POS (linfonodo sentinella positivo);
- LS NEG (linfonodo sentinella negativo);
- LS NO (linfonodo sentinella non eseguito);
- CC (chirurgia conservativa);
- LUMINAL A;
- PERM. ENDOVASCOLARE PERINEURALE SI (presenza di permeazione endovascolare perineurale);
- PERM. ENDOVASCOLARE PERINEURALE NO (assenza di permeazione endovascolare perineurale).

In modo particolare, possiamo ritenere che in presenza di permeazione endovascolare perineurale e di linfonodo sentinella non eseguito, si rileva un carcinoma mammario più aggressivo, nel quale le cellule tumorali hanno un aspetto anomalo, crescono più rapidamente e si diffondono a distanza. In questo caso, infatti, le pazienti si caratterizzano per un grado istologico molto elevato (G3) e un tumore dalle dimensioni comprese tra 2 cm e 5 cm nella dimensione massima (T2).

In contrapposizione, l'assenza di permeazione endovascolare perineurale, i linfonodi liberi da metastasi, il linfonodo sentinella negativo e la dimensione del tumore compresa tra 0.5 cm e 1 cm sono indicatori di una situazione meno grave, in cui le cellule tumorali crescono lentamente e difficilmente si diffondono a distanza. In questo caso, infatti, le pazienti si caratterizzano per il luminal di tipo A, per la chirurgia conservativa (CC) e per un grado istologico basso (G1).

#### **4. Conclusioni**

L'analisi sin qui condotta ci porta ad alcune conclusioni riguardanti entrambi i campioni analizzati.

In particolare, con riferimento al campione di donne anziane, riteniamo che la variabile LUMINAL sia indipendente dalle variabili T, N, TIPO INTERVENTO e METASTASI A DISTANZA, mentre risulta essere in relazione con le variabili G, LS e PERM. ENDOVASCOLARE PERINEURALE. Pertanto, possiamo ritenere che in presenza di permeazione endovascolare perineurale e di linfonodo sentinella non eseguito, si rileva un carcinoma mammario più aggressivo, nel quale le cellule tumorali hanno un aspetto anomalo, crescono più rapidamente e si diffondono a distanza. In questo caso, infatti, le pazienti si caratterizzano per un grado istologico molto elevato (G3).

Nel campione da noi esaminato, una maggior percentuale era rappresentata dalle pazienti LUMINAL B. Infatti, in accordo con i dati della letteratura, nelle donne anziane, che non si sottopongono a programmi di screening, si riscontrano forme più aggressive biologicamente che vengono comunque trattate con trattamenti conservativi in rapporto alle comorbidità.

In contrapposizione, i linfonodi liberi da metastasi, il linfonodo sentinella negativo e l'assenza di permeazione endovascolare perineurale, sono indicatori di una situazione prognostica meno grave, in cui le cellule tumorali crescono lentamente e difficilmente si diffondono a distanza. In questo caso, infatti, le pazienti si caratterizzano per il LUMINAL di tipo A e rappresentano una minore percentuale.

Per quello che concerne, invece, il campione di donne giovani, riteniamo che la variabile LUMINAL sia indipendente dalle variabili T, N, TIPO INTERVENTO e PERM. ENDOVASCOLARE PERINEURALE, mentre risulta avere una relazione con le variabili G e LS. In modo particolare, possiamo ritenere che in presenza di linfonodo sentinella non eseguito, si rileva un carcinoma mammario più aggressivo, nel quale le cellule tumorali hanno un aspetto anomalo, crescono più rapidamente e si diffondono a distanza. In questo caso, infatti, le pazienti si caratterizzano per un grado istologico molto elevato (G3). In contrapposizione, i linfonodi liberi da metastasi e il linfonodo sentinella negativo sono indicatori di una situazione prognostica meno grave, in cui le cellule tumorali crescono lentamente e difficilmente si diffondono a distanza. In questo caso, infatti, le pazienti si caratterizzano per il luminal di tipo A e per un grado istologico basso (G1). È dimostrato che, a volte, pur essendo di piccole dimensioni, riscontrate tramite lo screening mammografico, sono in realtà tumori molto eterogenei, in cui prevalgono forme triple negative e ad alto grading (G3) come si evince anche dall'analisi da noi effettuata. Queste forme spesso richiedono il trattamento chemioterapico neoadiuvante prima del trattamento chirurgico.

L'analisi dei due gruppi dimostra come il carcinoma mammario non sia un'unica entità clinica e patologica e necessita di trattamenti personalizzati in base all'età. Infatti i fattori biologici ed immunoistochimici, da noi analizzati, confermano l'estrema eterogeneità del tumore mammario, che va attentamente valutata e riconosciuta.

### **Riferimenti bibliografici**

Anastasiadi Z, Lianos GD, Ignatiadou E, Harissis H V., Mitsis M., Breast cancer in young women: an overview [Internet]. Vol. 69, *Updates in Surgery*, Springer-Verlag Italia s.r.l.; 2017 [cited 2020 Nov 1]. p. 313–7. Available from: <https://pubmed.ncbi.nlm.nih.gov/28260181/>

- Balducci L. Treatment of breast cancer in women older than 80 years is a complex task. Vol. 12, *Journal of Oncology Practice*. American Society of Clinical Oncology; 2016. p. 133–4.
- Bland M., 2009, *Statistica medica*, APOGEO, Milano.
- Ferrigni E, Bergom C, Yin Z, Szabo A, Kong AL. Breast Cancer in Women Aged 80 Years or Older: An Analysis of Treatment Patterns and Disease Outcomes [Internet]. Vol. 19, *Clinical Breast Cancer*, Elsevier Inc.; 2019 [cited 2020 Nov 1]. p. 157–64. Available from: <http://www.clinical-breast-cancer.com/article/S1526820918307055/fulltext>
- Glaser R, Marinopoulos S, Dimitrakakis C. Breast cancer treatment in women over the age of 80: A tailored approach. Vol. 110, *Maturitas*. Elsevier Ireland Ltd; 2018. p. 29–32.
- Gómez-Flores-Ramos L, Álvarez-Gómez RM, Villarreal-Garza C, Wegman-Ostrosky T, Mohar A. Breast cancer genetics in young women: What do we know? Vol. 774, *Mutation Research - Reviews in Mutation Research*. Elsevier B.V.; 2017. p. 33–45.
- Johnson HM, Irish W, Muzaffar M, Vohra NA, Wong JH. Quantifying the relationship between age at diagnosis and breast cancer-specific mortality. *Breast Cancer Res Treat*. 2019 Oct 1;177(3): 713–22.
- Kassambara A., 2017, Practical guide to principal component methods in R: PCA, M (CA), FAMD, MFA, HCPC, factoextra (Vol. 2). STHDA.
- Landis J. R., Kock G. C., 1977, The measurement of observer agreement for categorical data, *Biometrics*, 33, 159 – 174.
- Mills M, Liveringhouse C, Lee F, Nanda RH, Ahmed KA, Washington IR, et al. The prevalence of luminal B subtype is higher in older postmenopausal women with ER+/HER2- breast cancer and is associated with inferior outcomes. *J Geriatr Oncol*. 2020 Aug 26;
- Murphy BL, Day CN, Hoskin TL, Habermann EB, Boughey JC. Adolescents and Young Adults with Breast Cancer have More Aggressive Disease and Treatment Than Patients in Their Forties. *Ann Surg Oncol* [Internet]. 2019 Nov 1 [cited 2020 Nov 1];26(12):3920–30. Available from: <https://pubmed.ncbi.nlm.nih.gov/31376035/>
- Norman. G., Streiner S., 2015, *Biostatistica*, Casa Editrice Ambrosiana, Milano.
- Paluch-Shimon S, Pagani O, Partridge AH, Abulkhair O, Cardoso MJ, Dent RA, et al. ESO-ESMO 3rd international consensus guidelines for breast cancer in young women (BCY3). *Breast* [Internet]. 2017 Oct 1 [cited 2020 Nov 1];35:203–17. Available from: <https://pubmed.ncbi.nlm.nih.gov/28822332/>
- Rosner B., 2000, *Fundamentals of Biostatistics*, 5th ed. Duxbury.
- Wayne W., D., Chad L., C., 2012, *Biostatistics: A Foundation for Analysis in the Health Sciences*, 10th Edition, Casa editrice John Wiley & Sons, Inc.

**Sitografia**

<http://www.sthda.com/english/articles/22-principal-component-methods-videos/71-mca-in-r-using-factminer-quick-scripts-and-videos/>

<http://www.sthda.com/english/articles/31-principal-component-methods-in-r-practical-guide/114-mca-multiple-correspondence-analysis-in-r-essentials/#:~:text=MCA%20is%20generally%20used%20to,The%20associations%20between%20variable%20categories>

<http://www.sthda.com/english/articles/31-principal-component-methods-in-r-practical-guide/113-ca-correspondence-analysis-in-r-essentials/>

<https://www.fondazioneveronesi.it/magazine/tools-della-salute/glossario-delle-malattie/tumore-al-seno>

<https://www.registri-tumori.it/cms/pubblicazioni/i-numeri-del-cancro-italia-2019>

<https://www.aiom.it/linee-guida-aiom-neoplasie-della-mammella-2019/>