

UNIVERSITÁ DEGLI STUDI DI PADOVA

FACOLTÁ DI SCIENZE STATISTICHE

**CORSO DI LAUREA IN STATISTICA E GESTIONE DELLE
IMPRESE**



**La spesa in farmaci nella regione Friuli Venezia Giulia:
alcune analisi.**

Relatore: Prof. Omar Paccagnella

Laureando: Adriano Biasiolo

Matricola: 573040-GEI

ANNO ACCADEMICO 2010-2011

INDICE

1. INTRODUZIONE	p.3
1.1 L'AGENZIA REGIONALE DELLA SANITÀ	p.3
1.2 IL DATASET	p.4
1.3 SCOPO DELL'ANALISI	p.4
2. ANALISI PRELIMINARI	p.5
2.1 LE VARIABILI SCELTE	p.5
2.2 ALCUNI NUMERI	p.6
2.3 LO STUDIO DELLE VARIABILI	p.7
2.3.1 Età dei pazienti	p.7
2.3.2 Sesso dei pazienti	p.12
2.3.3 Aziende e distretti	p.15
2.3.4 Esperienza e sesso dei medici	p.22
2.3.5 La variabile esenzioni	p.25
3. ANALISI MULTIVARIATE	p.27
3.1 IL MODELLO DI REGRESSIONE LINEARE SEMPLICE	p.27
3.1.1 SAS output	p.29
3.1.2 Bontà del modello	p.30
3.1.3 Studio dei coefficienti	p.30
3.1.4 Conclusioni sul modello	p.33
3.2 IL MODELLO LOGIT	p.34
3.2.1 SAS output	p.35
3.2.2 Stima degli odds ratio	p.36
3.2.3 Bontà del modello	p.37
3.2.4 Studio degli odds ratio	p.37
3.2.5 Conclusioni sul modello	p.39
4. CONCLUSIONI	p.40
Bibliografia	p.41
Alcuni ringraziamenti	p.42

1.INTRODUZIONE

L'oggetto di analisi di questo lavoro è un dataset proveniente dall'Agenzia Regionale della Sanità della regione Friuli Venezia Giulia; i dati sono relativi al 2007 quando l'Agenzia era ancora operativa. Difatti l'Agenzia Regionale della Sanità è in liquidazione dal 2010 a seguito di una legge regionale.

1.1 L'AGENZIA REGIONALE DELLA SANITÀ

L'Agenzia Regionale della Sanità (ARS) della regione Friuli Venezia Giulia, è azienda della Regione, dotata di personalità giuridica pubblica, sottoposta alla vigilanza della Giunta regionale. Fino alla messa in liquidazione aveva finalità di supporto e coordinamento delle Aziende sanitarie regionali, e, nell'osservanza delle rispettive prerogative istituzionali, degli altri organismi che concorrevano al funzionamento del Servizio Sanitario Regionale, svolgendo anche il ruolo di supporto tecnico alla Regione per i compiti alla stessa attribuiti in materia sanitaria. L'ARS organizzava l'attività di supporto e coordinamento verso le aziende sanitarie regionali in particolare verificando l'attuazione degli obiettivi prioritari del Servizio Sanitario Regionale (SSR) e definendo le indicazioni generali progettuali per lo sviluppo e la conduzione del Sistema Informativo Socio Sanitario Regionale (SISSR). In materia di economia sanitaria inoltre l'ARS si occupava di verificare i bilanci delle aziende sanitarie regionali, di supportare le aziende sanitarie regionali con riferimento ad aspetti tecnico contabili e di supporto alla direzione generale. Inoltre l'ARS si occupava di redigere registri regionali quali ad esempio: il registro delle cause di mortalità, il registro dei tumori, il registro delle malattie cardiovascolari e il registro degli incidenti e degli infortuni, tramite i quali si occupava di valutare lo stato di salute della popolazione e gli esiti degli interventi sanitari. Altro compito svolto dall'ARS era l'elaborazione e la diffusione di modelli di valutazione di qualità e di misurazione di performance dell'ambiente sanitario regionale. L'ARS quindi oltre alle finalità di supporto e di coordinamento delle aziende sanitarie locali svolgeva anche la funzione di organismo di controllo delle stesse.

1.2 IL DATASET

Il dataset in esame fa parte di uno più grande che conteneva informazioni molto specifiche su ricoveri ospedalieri, visite specialistiche, accessi al pronto soccorso eccetera. I dati provengono dal sistema informativo sociosanitario regionale. Il dataset in esame è composto da 1.249.321 osservazioni relative a 27 variabili, le osservazioni si riferiscono all'intera popolazione del Friuli Venezia Giulia. Non tutte le variabili sono risultate di interesse per questo studio; alcune di queste, infatti, sono state eliminate o raggruppate ai fini dell'analisi. Per lo studio di questo dataset ho scelto di utilizzare il software SAS tramite il quale sono state effettuate tutte le operazioni. I dati fanno riferimento in particolare alla spesa in medicinali sostenuta dalle aziende sanitarie in Friuli Venezia Giulia nell'anno 2007. La variabile "chiave", quella che identifica ogni singola osservazione è data dall'unione di due variabili, cioè dalla coppia "medico-paziente": ogni volta che un medico ha un nuovo paziente o che un paziente cambia medico di riferimento viene creata una nuova osservazione.

1.3 SCOPO DELL'ANALISI

Lo scopo di questa analisi è di studiare come le variabili oggetto dell'analisi influenzino la spesa in medicinali sostenuta dalle aziende sanitarie regionali. Questo verrà realizzato sia attraverso la presentazione di un ampio set di statistiche descrittive e analisi univariate, sia per mezzo dello studio di appropriati modelli multivariati.

2. ANALISI PRELIMINARI

2.1 LE VARIABILI SCELTE

Il dataset originario presentava un totale di 27 variabili descritte qui a seguito

Nome variabile	Tipologia	Descrizione
Anno	Testuale	Anno di osservazione (inserita solo per uso futuro)
Assistiti	Testuale	Codice identificativo assistiti
Medici	Testuale	Codice identificativo medici
EtaAss	Numerico	Età in anni dell'assistito (anno osservazione – anno di nascita)
DATA_DECESSO	Testuale/Data	Data di decesso
Sesso assistito	Testuale/Dicotomica	F – femmina; M – maschio
CAUSA_PRINCIPALE	Testuale	Identificativo della causa principale di morte
CAUSA_VIOLENTA	Testuale	Identificativo della causa di morte violenta
TIPO	Testuale/Dicotomica	Tipologia di medico (G – medico di medicina generale; P - pediatra)
CAP	Numerico	Codice avviamento postale del medico
EtaMed	Numerico	Età in anni del medico (anno osservazione – anno di nascita)
ExpMed	Numerico	Esperienza in anni del medico (anno osservazione – anno di iscrizione all'anagrafe)
LauMed	Numerico	Distanza dalla laurea del medico in anni (anno di osservazione – anno di laurea)
AZIENDA	Testuale	Azienda sanitaria di riferimento del medico
DISTRETTO	Testuale	Distretto sanitario di riferimento del medico
ASSISTITI IN CARICO	Numerico	Numero di assistiti in carico effettivamente osservati nel periodo
ASSMAX	Numerico	Numero massimo di assistiti
SESSO	Testuale/Dicotomica	Genere del medico (F – femmina; M – maschio)
GG	Numerico	Giorni effettivi di carico dell'assistito col medico di medicina generale
CarPerc	Numerico	Percentuale di carico del medico sui giorni totali di osservazione dell'assistito
Imp2A	Numerico	Importo totale osservato per l'assistito nella categoria di esenzione A (pesando i campi per il numero di ripetizioni)
Imp2N	Numerico	Importo totale osservato per l'assistito nella categoria di esenzione N (pesando i campi per il numero di ripetizioni)
Imp2X	Numerico	Importo totale osservato per l'assistito nella categoria di esenzione X (pesando i campi per il numero di ripetizioni)
Nconf2A	Numerico	Numero totale di confezioni prescritte per l'assistito nella categoria di esenzione A (pesando i campi per il numero di ripetizioni)
Nconf2N	Numerico	Numero totale di confezioni prescritte per l'assistito nella categoria di esenzione N (pesando i campi per il numero di ripetizioni)
Nconf2X	Numerico	Numero totale di confezioni prescritte per l'assistito nella categoria di esenzione X (pesando i campi per il numero di ripetizioni)

Non tutte però sono state rilevanti per l'analisi, in particolare le ultime tre variabili sono state raggruppate in una sola chiamata "spesatot", che descrive l'importo totale in euro che l'azienda ha registrato nel 2007 come spesa in medicinali per la coppia "medico-paziente".

Le variabili che sono state oggetto di studio sono risultate quelle che si sono rivelate influenti rispetto alla variabile "spesatot" oggetto di questa analisi, e sono elencate di seguito:

Nome variabile	Tipologia	Descrizione
Anno	Testuale	Anno di osservazione (inserita solo per uso futuro)
Assistiti	Testuale	Codice identificativo assistiti
Medici	Testuale	Codice identificativo medici
EtaAss	Numerico	Età in anni dell'assistito (anno osservazione – anno di nascita)
Sesso assistito	Testuale/Dicotomica	F – femmina; M – maschio
EtaMed	Numerico	Età in anni del medico (anno osservazione – anno di nascita)
ExpMed	Numerico	Esperienza in anni del medico (anno osservazione – anno di iscrizione all'anagrafe)
AZIENDA	Testuale	Azienda sanitaria di riferimento del medico
DISTRETTO	Testuale	Distretto sanitario di riferimento del medico
SESSO	Testuale/Dicotomica	Genere del medico (F – femmina; M – maschio)
Imp2A	Numerico	Importo totale osservato per l'assistito nella categoria di esenzione A (pesando i campi per il numero di ripetizioni)
Imp2N	Numerico	Importo totale osservato per l'assistito nella categoria di esenzione N (pesando i campi per il numero di ripetizioni)
Imp2X	Numerico	Importo totale osservato per l'assistito nella categoria di esenzione X (pesando i campi per il numero di ripetizioni)

2.2 ALCUNI NUMERI

Per l'anno 2007 la spesa totale in medicinali per le aziende sanitarie della regione Friuli Venezia Giulia è stata di 235.656.835 euro, mentre la spesa media, per coppia medico paziente, è stata di 188,63 euro, con deviazione standard di 444,02. Se si considera solo la parte di popolazione che ha effettivamente registrato una spesa in medicinali, la spesa media è stata di 304,22 euro con deviazione standard di 531,81.

2.3 LO STUDIO DELLE VARIABILI

Per individuare le variabili di interesse sono state effettuate delle analisi descrittive allo scopo di evidenziare andamenti sistematici della spesa in medicinali al variare di fattori come l'età, il sesso, l'azienda o l'esperienza dei medici.

2.3.1 Età dei pazienti

Per lo studio dell'andamento della spesa in medicinali in funzione dell'età dei pazienti è stata creata una nuova variabile qualitativa che si riferisce alle fasce di età; sono state quindi create dieci fasce d'età e, relativamente a queste, si sono registrati i valori di spesa media, di spesa totale e il numero di osservazioni che ne determinano la popolazione. Sono state inoltre individuati due gruppi, uno comprendente tutta la popolazione, uno invece relativo solo a quella parte di soggetti (identificati dalla coppia medico-paziente) che hanno registrato effettivamente delle spese in medicinali per le aziende sanitarie.

Il Grafico 1 mostra il numero di coppie medico paziente che hanno effettivamente registrato spese in medicinali confrontato con il totale della popolazione della regione.

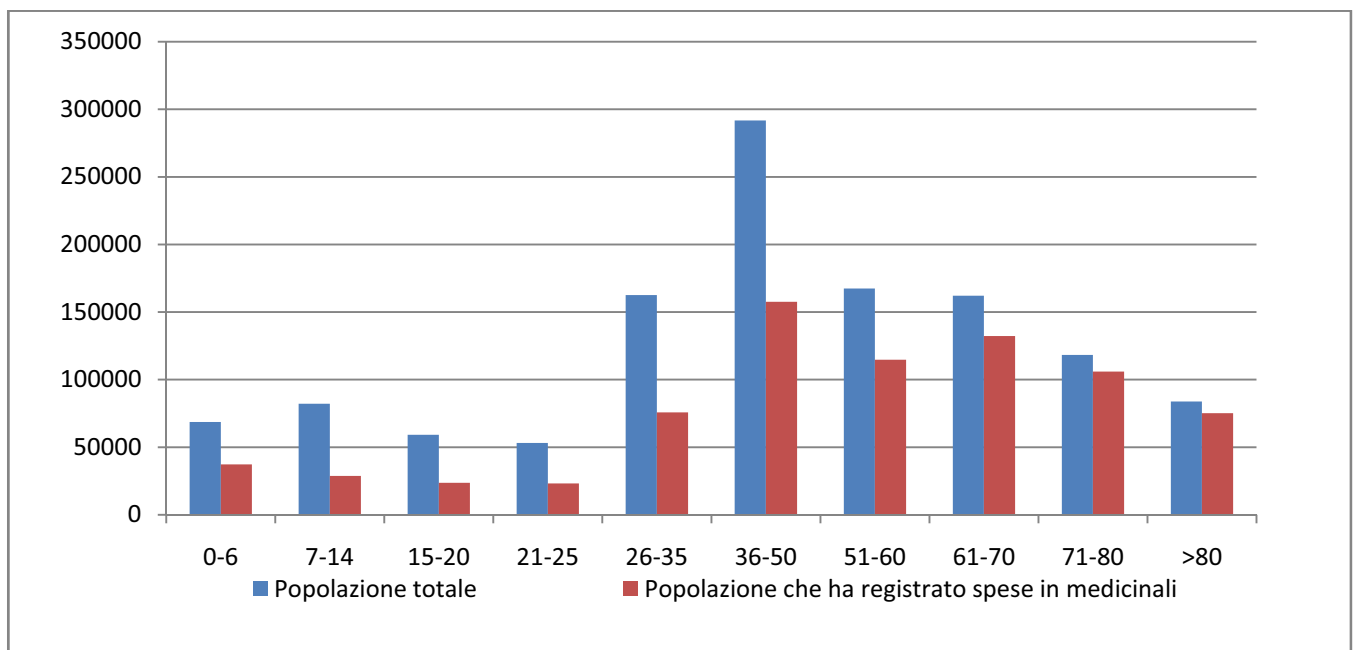


Grafico 1. Confronto tra numero di coppie medico paziente totale e coppie per cui si è effettivamente registrata una spesa in medicinali per fasce d'età.

La Tabella 1 mostra invece le percentuali relative alla frazione della popolazione che ha effettivamente registrato spese in medicinali, per ciascuna classe d'età.

Fasce di età	Percentuale della popolazione che ha registrato spese in medicinali
0-6	54,37
7-14	35,03
15-20	39,94
21-25	43,65
26-35	46,61
36-50	54,02
51-60	68,52
61-70	81,61
71-80	89,63
>80	89,69

Tabella 1. Percentuali della popolazione che ha effettivamente registrato spese in medicinali per ciascuna fascia d'età.

Questa prima analisi segnala forte relazione tra le fasce di età dei pazienti e la spesa in medicinali registrata per ogni coppia medico-paziente. È evidente infatti che la percentuale di soggetti che hanno rilevato spese in medicinali dipende dall'età di questi, in particolare superati i 6 anni questa percentuale aumenta con l'aumentare dell'età.

Successivamente si è analizzato l'andamento della spesa media per paziente e della spesa totale riferita a queste fasce di età. Anche qui si può notare una forte associazione tra età e spesa in medicinali: si è infatti in presenza di un andamento crescente della spesa all'aumentare dell'età dei pazienti fino alla soglia degli ottant'anni, superata la quale la spesa sia media che totale diminuisce; sembra quindi che i cittadini che superano questa soglia godano di salute migliore rispetto ai cittadini della fascia 71-80. Tale andamento è evidente nei grafici 2 e 3 che si riferiscono rispettivamente alla spesa totale ed alla spesa media in medicinali, in particolare il Grafico 3 è particolarmente interessante in quanto non tiene conto delle diverse numerosità delle classi.

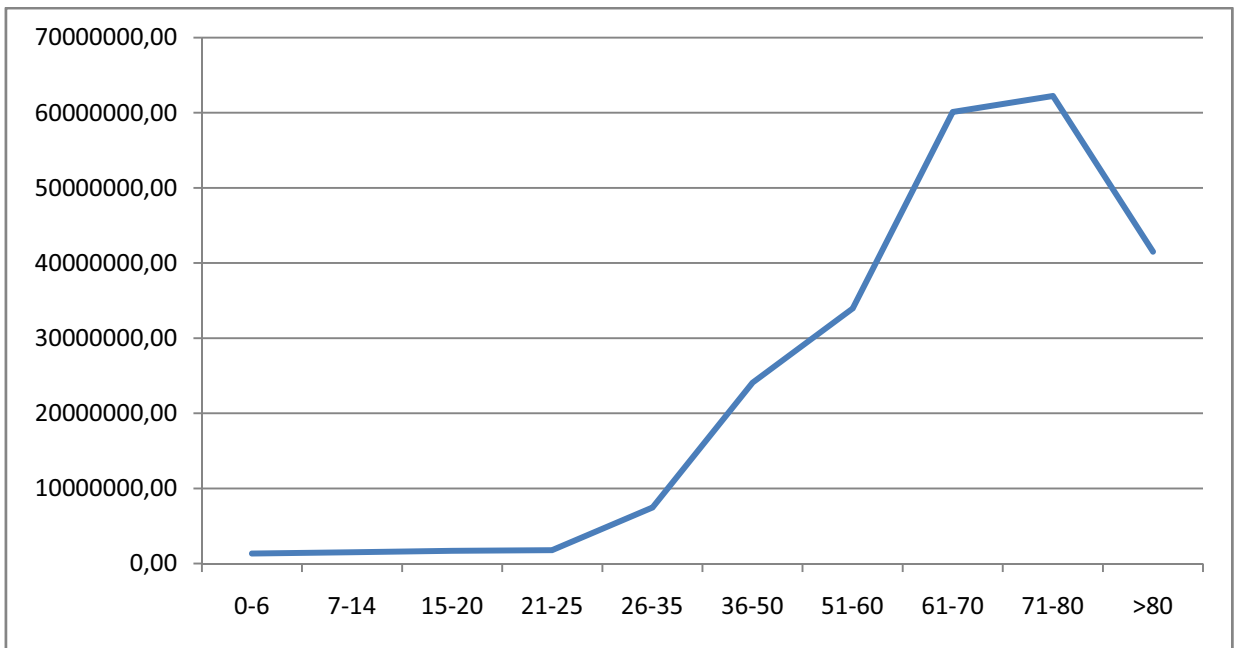


Grafico2. Spesa totale in € in medicinali per fasce di età.

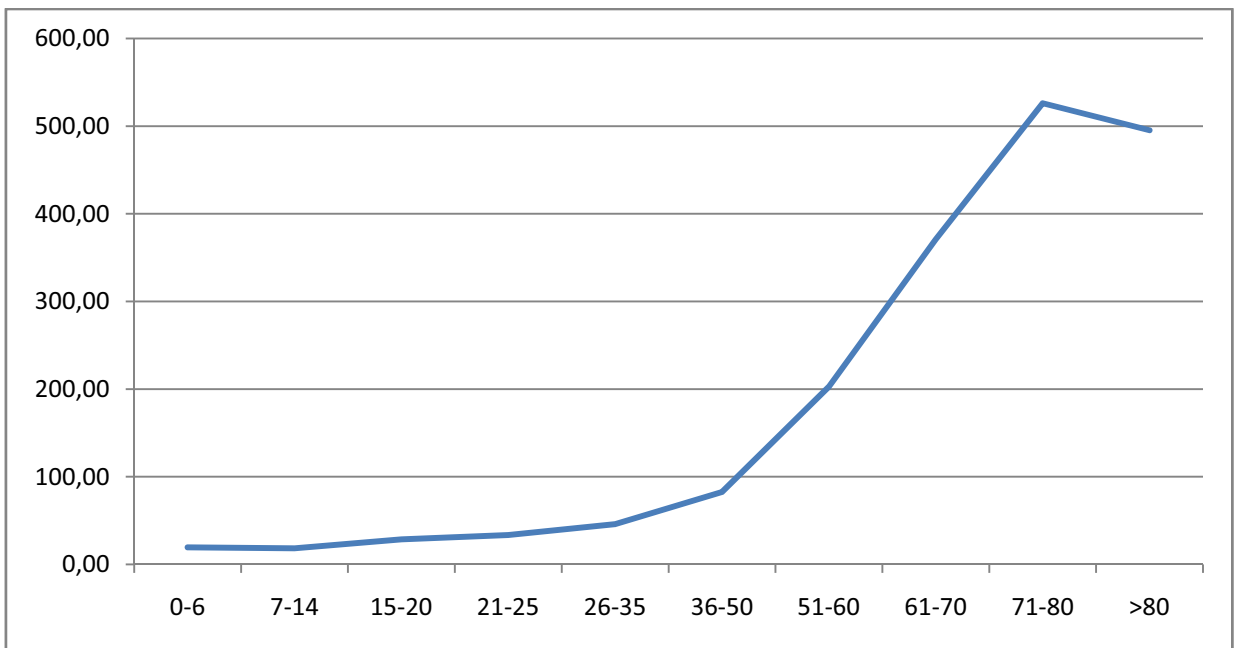


Grafico 3. Spesa media in € in medicinali per fasce di età.

Risulta inoltre di interesse osservare l'andamento del Grafico 4 che mette a confronto la spesa media del totale delle coppie medico paziente e della spesa media delle sole coppie medico paziente che hanno effettivamente registrato una spesa in medicinali; nella prime cinque fasce d'età infatti la mediana risulta essere 0. Questo significa che per queste fasce d'età almeno il 50% delle osservazioni ha spesa in medicinali nulla.

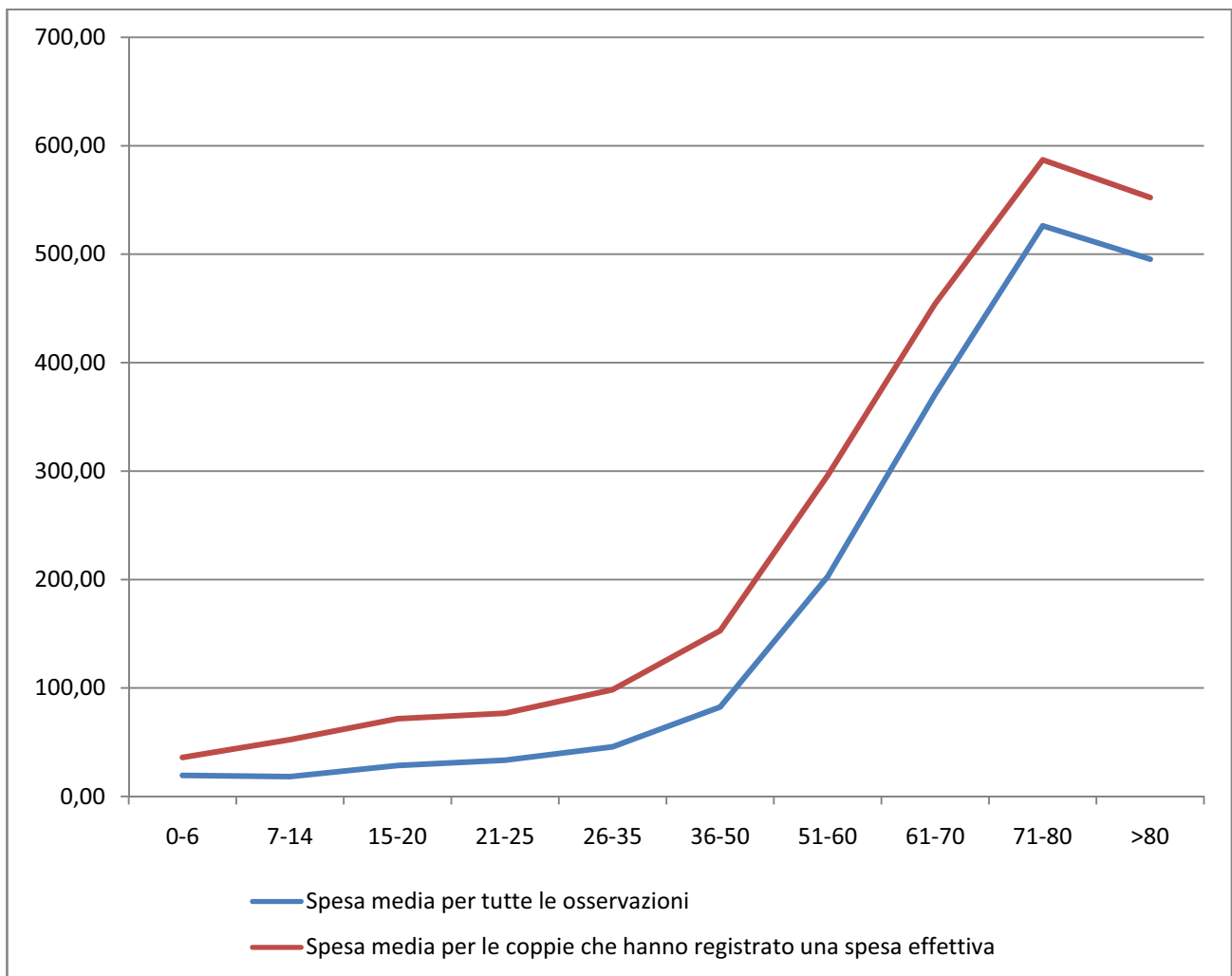


Grafico 4. Confronto tra spesa media in € sul totale delle osservazioni e spesa media in € sulle sole coppie che hanno registrato una spesa in medicinali effettiva.

La linea rossa rappresenta l'andamento effettivo della spesa per medicinali, ovvero l'andamento della parte di popolazione che ha effettivamente generato una spesa, avendo eliminato tutte le osservazioni con spesa nulla; avendo solo osservazioni positive si registra uno spostamento della curva su valori superiori.

Nel Grafico 4 la spesa media massima si registra per la fascia 71-80 con un valore di 526,16 euro nel totale delle osservazioni e di 587,04 euro per le osservazioni con spesa positiva; dalla fascia 0-6 alla fascia 36-50 l'incremento della spesa all'aumentare dell'età sembra essere piuttosto debole, ma giunti a tale fascia si nota un'impennata sulla spesa media. Infatti tra la fascia 36-50 e la fascia 51-60 si passa da una spesa media di 82,51 euro ad una spesa media di 202,76 euro per il totale e da 152,74 euro a 295,93 euro per le osservazioni con spesa positiva. Tale incremento continua fino al valore massimo della fascia 71-80 per poi decrescere fino all'ultima fascia degli over 80. Questo è in linea con il consumo di medicinali descritto dalla Tabella 1. È quindi evidente come superati i 36 anni il consumo in medicinali aumenti notevolmente e che i soggetti appartenenti alle fasce comprese tra la fascia 36-50 e la fascia over 80 sono quelle che contribuiscono maggiormente alla spesa totale in farmaci della regione. Le Tabelle 2 e 3 vogliono mettere a confronto la percentuale di popolazione di ogni fascia d'età sul totale ed il contributo che ognuna di esse ha sulla spesa totale.

Fascia d'età	Popolazione	% della popolazione sul totale
0-6	68708	5,45
7-14	82221	6,58
15-20	59183	4,74
21-25	53219	4,26
26-35	162543	13,01
36-50	291780	23,36
51-60	167461	13,40
61-70	162104	12,98
71-80	118281	9,47
>80	83821	6,71

Tabella 2. Percentuale di popolazione di ogni fascia d'età rispetto al totale.

Fascia d'età	Spesa totale in €	Contributo in % della fascia sulla spesa totale
0-6	1339618,80	0,57
7-14	1505667,95	0,64
15-20	1692168,15	0,72
21-25	1782044,92	0,76
26-35	7446881,57	3,16
36-50	24073436,30	10,23
51-60	33954446,10	14,42
61-70	60110047,50	25,52
71-80	62234330,90	26,42
>80	41518192,90	17,63

Tabella 3. Contributo delle fasce d'età sulla spesa totale.

Questi risultati servono a sottolineare il fatto che l'andamento della spesa totale rilevato dal Grafico 2 è da ricondursi principalmente all'effetto che l'età ha sulla spesa piuttosto che a fattori derivanti dalla numerosità dei gruppi.

2.3.2 Sesso dei pazienti

Il dataset è composto da 647.604 osservazioni relative a pazienti di sesso femminile ed a 601.717 osservazioni relative a pazienti di sesso maschile. Andando ad analizzare questa variabile singolarmente non si ottengono grandi informazioni. Se si confrontano spesa media e spesa totale relativamente a pazienti maschili e femminili si notano solo lievi differenze (Tabella 4).

Genere	Numero osservazioni	Spesa media in €	Spesa totale in €	Deviazione Standard	Standard Error
Donna	647604	192,03	124361601	415,08	0,52
Uomo	601717	184,96	111295234	473,18	0,61

Tabella 4. Statistiche descrittive per la variabile sesso_assistito.

Per verificare se queste differenze sono però statisticamente significative occorre procedere con un opportuno t-test. SAS fornisce due tipi di t-test, il primo considera che i due gruppi abbiano varianza uguale, mentre il secondo considera varianze diverse. Per prima cosa quindi dobbiamo capire in quale dei due casi stiamo lavorando. Per fare ciò SAS calcola il “Folded F test” che assume come ipotesi nulla l’omoschedasticità delle varianze.

Calcolato con la seguente formula:

$$F' = \frac{\max(s_1^2, s_2^2)}{\min(s_1^2, s_2^2)}$$

Dove S_1 e S_2 indicano le deviazioni standard dei due gruppi. Dalla Tabella 5 il test risulta significativo, viene quindi rigettata l’ipotesi nulla di omoschedasticità delle varianze.

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	601716	647603	1.30	<.0001

Tabella 5. Test di omoschedasticità delle varianze per la variabile sesso_assistito.

Possiamo ora calcolare il t-test:

Sesso_assistito	N	Mean	Std Dev	Std Err	Minimum	Maximum
F	647604	192.0	415.1	0.5158	0	30048.5
M	601717	185.0	473.2	0.6100	0	43949.5
Diff (1-2)		7.0706	444.0	0.7950		

Method	Variances	DF	t Value	Pr > t
Pooled	Equal	1.25E6	8.89	<.0001
Satterthwaite	Unequal	1.2E6	8.85	<.0001

Tabella 6. T-test SAS output per la variabile sesso_assistito.

Dalla Tabella 6 possiamo quindi concludere che le differenze tra i due gruppi sono statisticamente significative.

Risulta invece molto utile studiare questa variabile congiuntamente alle altre variabili oggetto dello studio. In particolare risulta interessante studiare il ruolo del sesso rispetto alle fasce d'età;

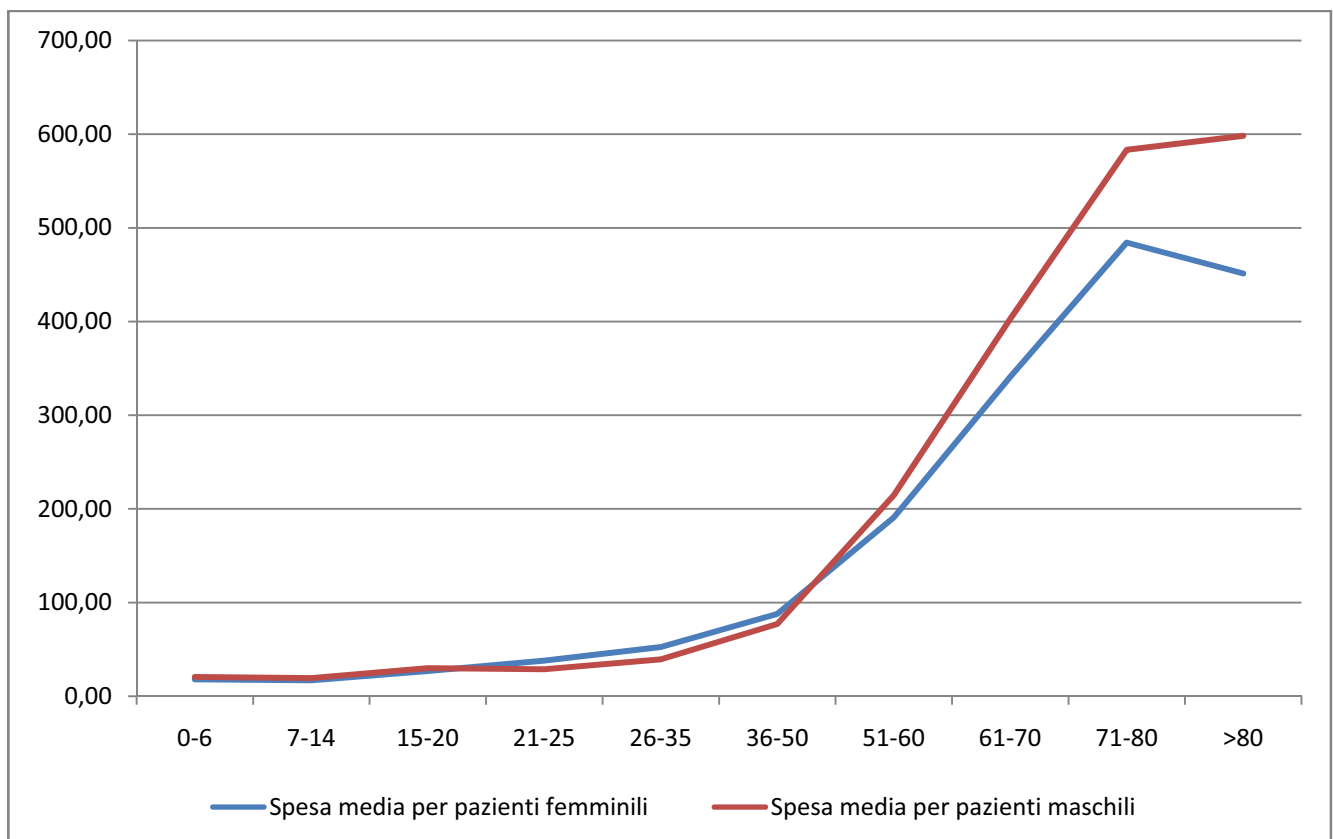


Grafico 5. Spesa media in € maschile e femminile per fasce d'età.

Il Grafico 5 riporta l'andamento della spesa in medicinali delle osservazioni relative a pazienti maschili ed a pazienti femminili all'aumentare dell'età. L'andamento del grafico è simile a quello dei grafici già visti. È però molto interessante notare come superata la fascia 51-60 la spesa media in medicinali dei pazienti maschili risulti essere superiore a quella dei pazienti femminili, con una forbice che si allarga sempre più al crescere dell'età. È molto interessante notare anche come oltre gli 80 anni siano solamente le donne ad evidenziare una diminuzione della spesa media, mentre per gli uomini la spesa media continua a salire. Sono presenti inoltre degli incroci fra le due curve: la spesa media relativa ai pazienti femminili è superiore a quella dei pazienti maschili nelle parte di grafico compresa tra la fascia 15-20 e la fascia 36-50. Questo aumento della spesa in medicinali è probabilmente da imputare al fatto che proprio in quelle fasce d'età le donne sono nel loro periodo fertile. La differenza maggiore si rileva per l'ultima fascia d'età, la spesa media per i pazienti femminili raggiunge il suo massimo nella fascia 71-80, con un valore di 484,34 euro mentre per i pazienti maschili si raggiunge il massimo nella fascia finale over 80, con un valore di 598,30 euro che contrasta di molto con il rispettivo valore della spesa media per i pazienti femminili di 451,33 euro. Per quanto riguarda le osservazioni comprese tra le fasce 15-20 e la fascia 36-50 i valori non sono così contrastanti, la differenza maggiore si nota per la fascia 26-35 che assume valore 52,52 euro per i pazienti femminili e 39,43 euro per quelli maschili, nonostante questo tali differenze sono comunque molto interessanti ai fini dell'analisi.

2.3.3 Aziende e distretti

I dati si riferiscono ad aziende sanitarie regionali. Nell'ordinamento italiano l'azienda sanitaria locale (ASL) è un ente pubblico locale. In passato era un ente strumentale della regione, al quale competeva l'organizzazione finanziaria e gestionale delle prestazioni sanitarie. Dal 1993, secondo prevalente giurisprudenza (Tar Toscana sentenza 17 settembre 2003 n. 5101), ha perso il carattere di organo della Regione, acquisendo una propria soggettività giuridica con un'autonomia che ha poi assunto, stante il disposto dell'art. 3, c. 1 bis del D.Lgs. 502/92 (comma introdotto dal D.Lgs. 19.6.99 n. 229), anche carattere imprenditoriale (“in funzione del perseguimento dei loro fini istituzionali, le unità sanitarie locali si costituiscono in Aziende con personalità giuridica pubblica e autonomia imprenditoriale”), disposizione quest'ultima che ha indotto a ritenere che le aziende sanitarie abbiano assunto la natura di enti pubblici economici. Le aziende sanitarie suddividono il territorio regionale in zone di competenza per ciascuna azienda. Si tratta, quindi, di una suddivisione in aree geografiche ognuna delle quali è di competenza di una singola azienda. Le aziende operanti in Friuli Venezia Giulia nel 2007 erano 6. Ogni azienda è poi suddivisa in distretti, che sono la struttura operativa mediante cui l'Azienda è presente nel territorio.

Il Grafico 6 mostra la distribuzione delle osservazioni per Azienda e la parte di popolazione che ha registrato spese in medicinali.

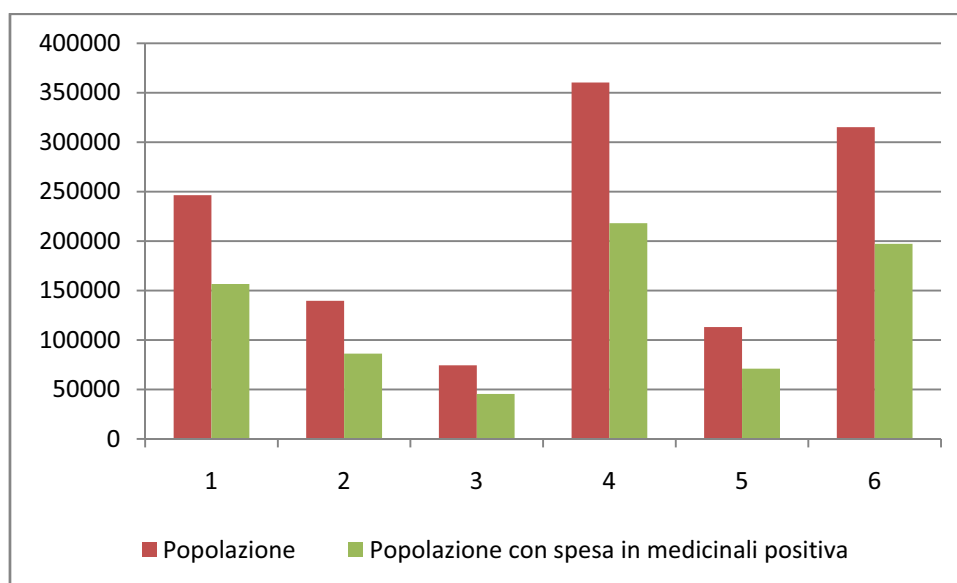


Grafico 6. Confronto tra numero di coppie medico paziente totale e coppie per cui si è effettivamente registrata una spesa in medicinali per azienda.

La Tabella 7 seguente invece mostra la percentuale di popolazione che ha registrato spese in medicinali per ogni azienda.

Azienda	Percentuale della popolazione che ha registrato spese in medicinali
1	63,55
2	61,69
3	61,09
4	60,53
5	62,73
6	62,55

Tabella 7. Percentuali della distribuzione del consumo di medicinali per azienda.

Come si può notare dalle percentuali la distribuzione del consumo di medicinali tra le popolazioni di ogni azienda sembra essere piuttosto omogenea; questo, quindi, potrebbe significare che l'appartenenza di un paziente ad un'azienda rispetto ad un'altra non è legata alla spesa in medicinali registrata.

Andando invece a studiare gli andamenti della spesa media si notano delle interessanti differenze tra aziende. Il Grafico 7 ed il Grafico 8 mostrano rispettivamente la spesa media per azienda relativamente a tutte le osservazioni e la spesa media per azienda relativamente alle sole osservazioni per cui è stata registrata una spesa in medicinali.

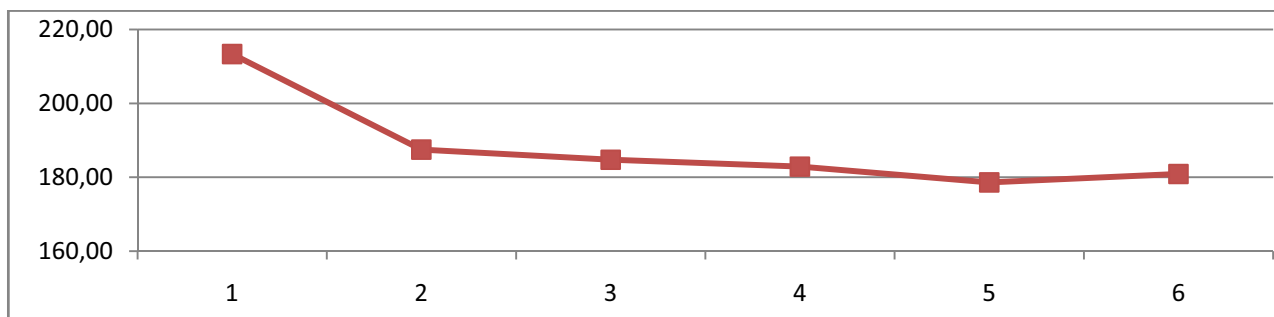


Grafico 7. Spesa media per azienda in €.

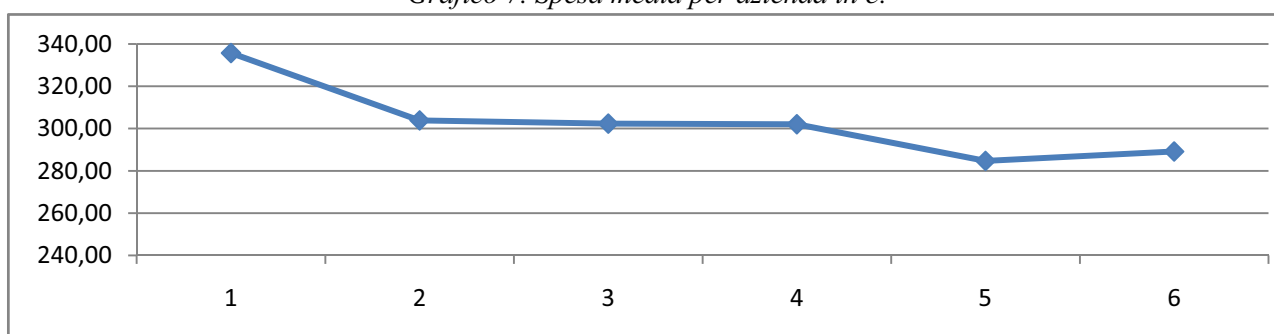


Grafico 8. Spesa media per azienda in € delle osservazioni per cui si è registrata una spesa in medicinali positiva.

Dal Grafico 8 l'azienda 1 mostra la spesa media più alta, 336 euro, mentre l'azienda 5 la più bassa, 235 euro. Dalle analisi precedenti si potrebbe pensare che queste differenze potrebbero essere dovute ad un collegamento tra azienda ed età media, in quanto l'età influisce direttamente (come riscontrato fino ad ora) sulla spesa media e totale. I Grafici 9 e 10 riportano un confronto tra spesa media per azienda ed età media per azienda rispettivamente per le osservazioni totali e per le osservazioni per cui si è registrata una spesa in medicinali positiva. L'asse verticale di sinistra indica la spesa media, mentre l'asse verticale di destra indica l'età media.

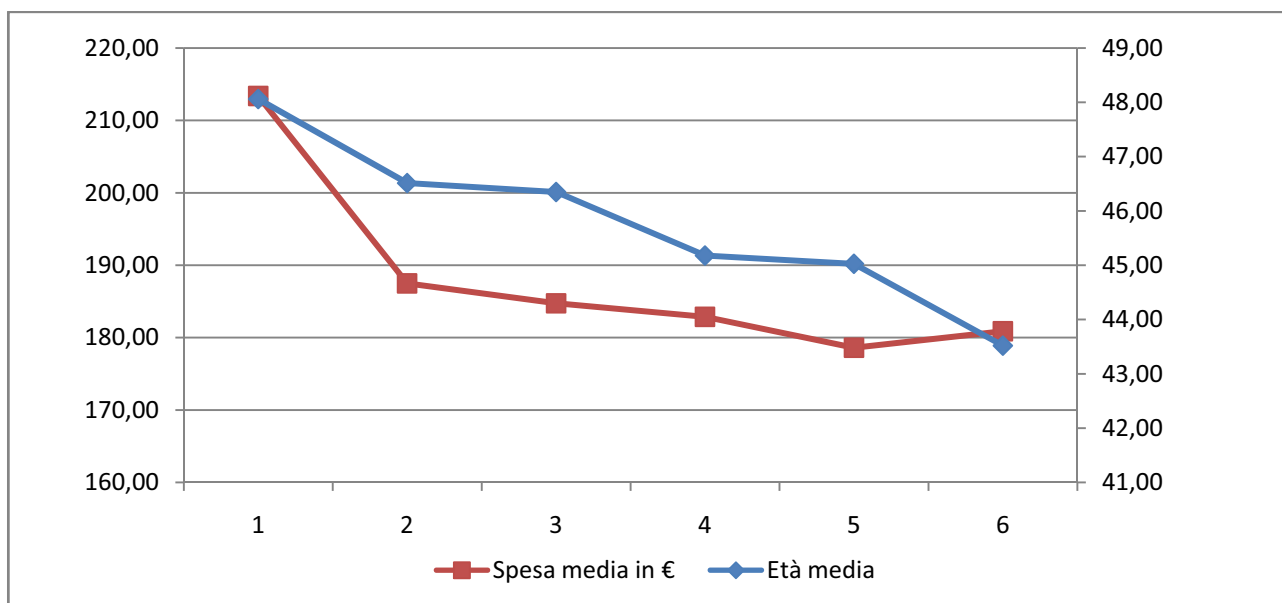


Grafico 9. Spesa media in € per azienda a confronto con età media per azienda.

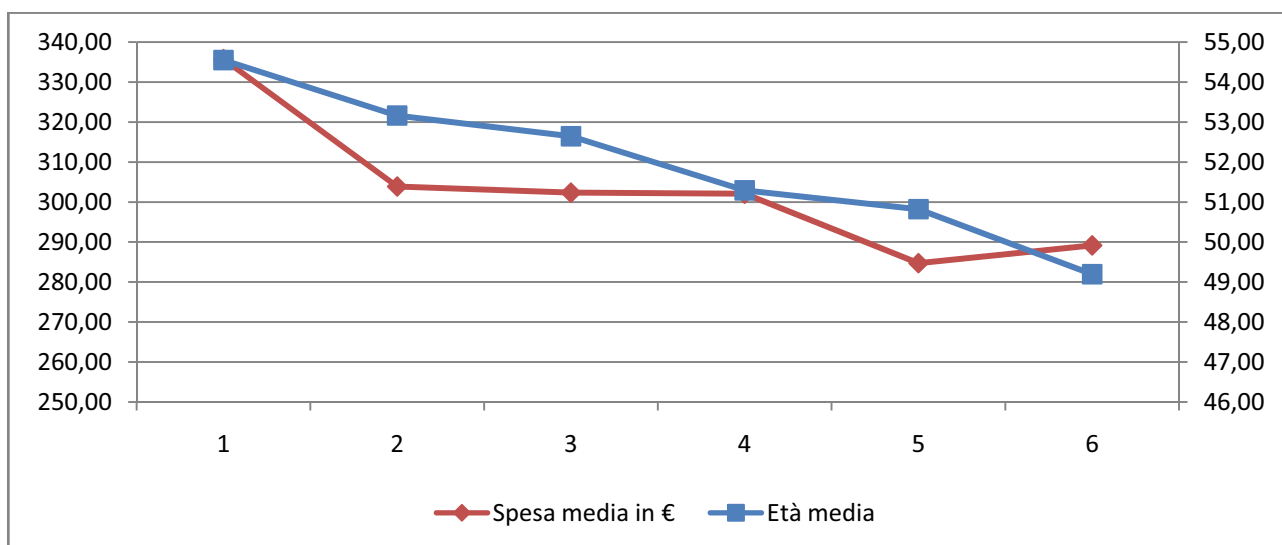


Grafico 10. Spesa media per azienda in € a confronto con età media per azienda relativamente alle solo osservazioni per cui si è registrata una spesa in medicinali positiva.

Questi due grafici sembrano confermare l'ipotesi che la spesa media per azienda sia influenzata dall'età media dell'azienda, fatta eccezione per l'azienda 6, che pur essendo quella con età media inferiore ha spesa media superiore rispetto all'azienda 5. Sembrano esserci in sostanza degli effetti sulla spesa media dovuti alla diversa composizione per età di ogni azienda; il Grafico 11 rappresenta la composizione per fasce d'età di ogni azienda:

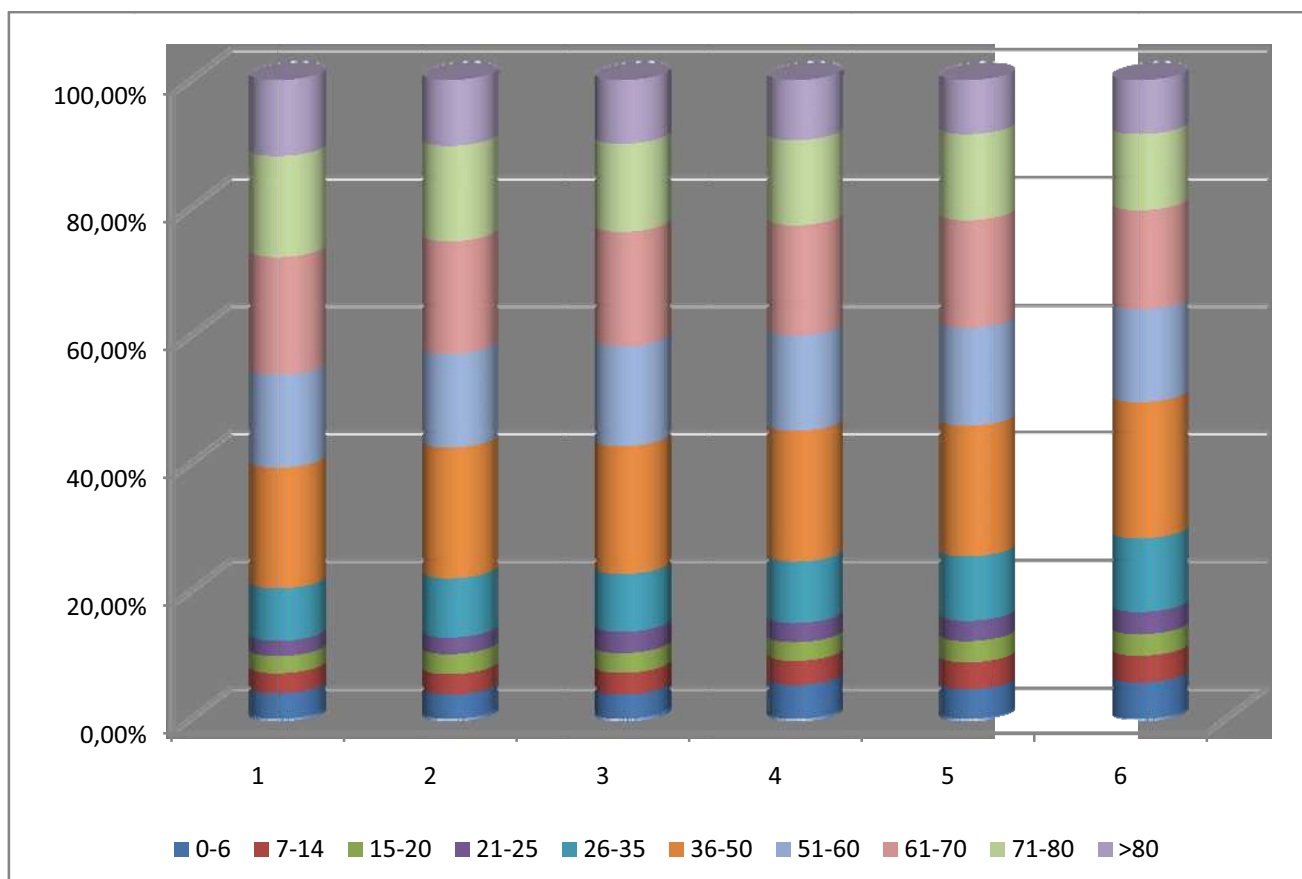


Grafico 11. Composizione delle aziende per fasce d'età.

Dal Grafico 11 si può facilmente notare che l'azienda 1 ha una percentuale di osservazioni relative a pazienti di età superiore a 50 anni maggiore rispetto alle altre, (più del 60% delle osservazioni), mentre per esempio nelle aziende 5 e 6 la percentuale di osservazioni relative a pazienti che superano i 50 anni d'età non raggiunge il 60%. Queste differenze sono conformi con i risultati ottenuti fino ad ora, infatti sembra che effettivamente ci sia un'associazione tra aziende e composizione della popolazione in fasce d'età.

Incrociando le informazioni sulla spesa media per fasce d'età con la spesa media per azienda si ottengono delle variazioni sull'andamento della spesa media rispetto all'età e ogni azienda presenta un andamento leggermente diverso rispetto alle altre, confermando quindi che l'azienda ha un ruolo significativo sulla determinazione della spesa in medicinali per ogni osservazione. Queste differenze sembrano risultare più definite all'aumentare dell'età dei pazienti.

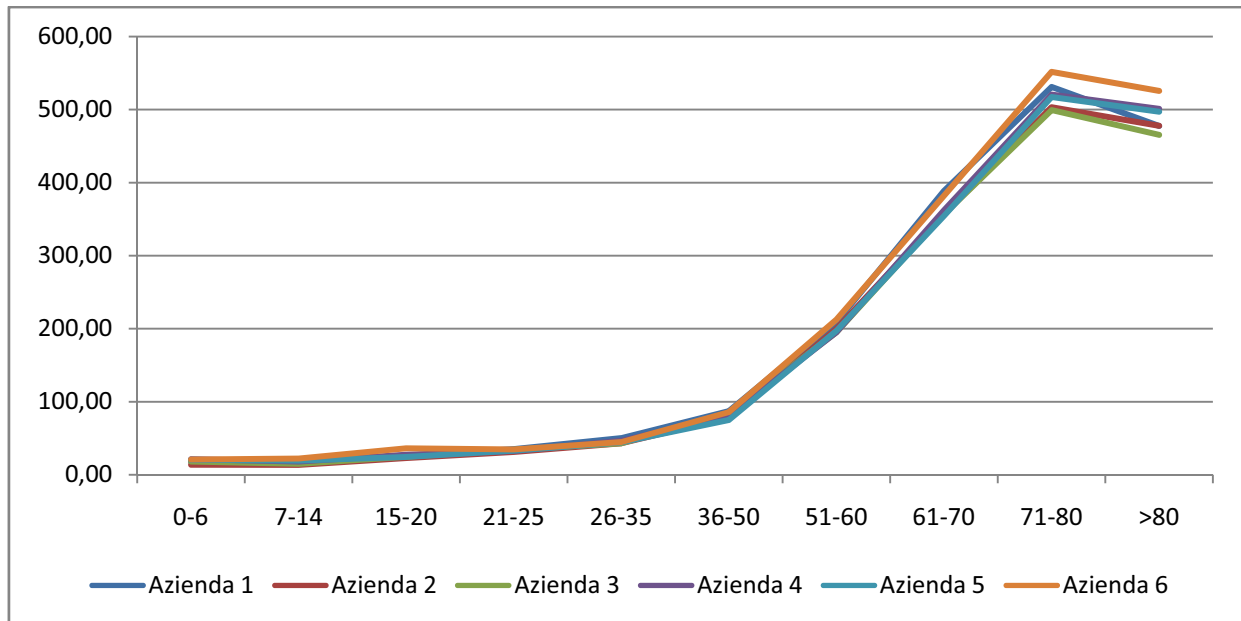


Grafico 12. Andamento della spesa media in € per azienda relativamente alle fasce d'età.

Anche il Grafico 12 ci porta a pensare che ci sia una sorta di associazione tra fasce d'età ed azienda di appartenenza. Per verificare se effettivamente tale associazione esiste occorre a questo punto creare una tabella di contingenza e calcolare gli opportuni test.

La Tabella 8 riporta per ogni cella, nell'ordine: la frequenza assoluta, la frequenza relativa, la frequenza percentuale di riga e infine la frequenza percentuale di colonna. Per la costruzione di questa tabella ho scelto di riferirmi solo alle osservazioni per cui si è rilevata una spesa in medicinali. Si possono notare delle differenze in percentuale tra le varie aziende relativamente alle fasce d'età. Queste possono sembrare lievi ma va considerato che il dataset di riferimento è composto da 774.614 osservazioni, perciò queste differenze potrebbero essere significative.

Table of azienda by fasciaeta											
azienda	fasciaeta										Total
	fascia 0-6	fascia 07-14	fascia 15-20	fascia 21-25	fascia 26-35	fascia 36-50	fascia 51-60	fascia 61-70	fascia 71-80	fascia over 80	
01	6233	4962	4201	3754	12842	29620	22766	28815	24768	18614	156575
	0.80	0.64	0.54	0.48	1.66	3.82	2.94	3.72	3.20	2.40	20.21
	3.98	3.17	2.68	2.40	8.20	18.92	14.54	18.40	15.82	11.89	
	16.69	17.23	17.77	16.16	16.95	18.79	19.84	21.78	23.36	24.76	
02	3276	2763	2702	2239	7980	17658	12648	15185	12867	8876	86194
	0.42	0.36	0.35	0.29	1.03	2.28	1.63	1.96	1.66	1.15	11.13
	3.80	3.21	3.13	2.60	9.26	20.49	14.67	17.62	14.93	10.30	
	8.77	9.59	11.43	9.64	10.53	11.20	11.02	11.48	12.14	11.81	
03	1729	1548	1411	1520	4115	9110	7131	8145	6271	4532	45512
	0.22	0.20	0.18	0.20	0.53	1.18	0.92	1.05	0.81	0.59	5.88
	3.80	3.40	3.10	3.34	9.04	20.02	15.67	17.90	13.78	9.96	
	4.63	5.37	5.97	6.54	5.43	5.78	6.22	6.16	5.92	6.03	
04	11560	8273	6338	6557	20959	44640	32595	37614	29097	20496	218129
	1.49	1.07	0.82	0.85	2.71	5.76	4.21	4.86	3.76	2.65	28.16
	5.30	3.79	2.91	3.01	9.61	20.46	14.94	17.24	13.34	9.40	
	30.95	28.72	26.81	28.23	27.67	28.32	28.41	28.43	27.45	27.26	
05	3283	2998	2357	2268	7158	14517	10896	11954	9551	6015	70997
	0.42	0.39	0.30	0.29	0.92	1.87	1.41	1.54	1.23	0.78	9.17
	4.62	4.22	3.32	3.19	10.08	20.45	15.35	16.84	13.45	8.47	
	8.79	10.41	9.97	9.76	9.45	9.21	9.50	9.04	9.01	8.00	
06	11273	8261	6628	6892	22705	42062	28702	30578	23459	16647	197207
	1.46	1.07	0.86	0.89	2.93	5.43	3.71	3.95	3.03	2.15	25.46
	5.72	4.19	3.36	3.49	11.51	21.33	14.55	15.51	11.90	8.44	
	30.18	28.68	28.04	29.67	29.97	26.69	25.02	23.11	22.13	22.14	
Total	37354	28805	23637	23230	75759	157607	114738	132291	106013	75180	774614
	4.82	3.72	3.05	3.00	9.78	20.35	14.81	17.08	13.69	9.71	100.00

Tabella 8. Tabella di contingenza per aziende e fasce d'età.

Per calcolare il test Chi-Quadro in una tabella a doppia entrata SAS utilizza la seguente formula:

$$Q_P = \sum_i \sum_j \frac{(n_{ij} - e_{ij})^2}{e_{ij}}$$

Dove:

n_{ij} è la frequenza osservata relativamente alla cella (i,j);

e_{ij} è la frequenza attesa nel caso in cui l'ipotesi nulla di indipendenza sia verificata:

Il test Chi-Quadro però è notoriamente influenzato dal numero di celle che compongono la tabella di contingenza: un alto numero di celle può portare a valori di Chi-Quadro decisamente alti. Per tenere conto di questo limite è importante affiancarlo ad altre statistiche come la V di Cramer.

Per calcolare la V di Cramer SAS utilizza la seguente formula:

$$V = \sqrt{\frac{Q_p/n}{\min(R-1, C-1)}}$$

Dove Q_p è il valore assunto dalla statistica Chi-Quadro;

n è il numero delle osservazioni;

R è il numero di righe;

C è il numero di colonne.

La statistica V di Cramer è definita come la radice quadrata del rapporto tra l'indice Chi-Quadro e il suo massimo valore teorico. La V di Cramer assume valore 0 nei casi di "indipendenza perfetta" e 1 nei casi di "dipendenza perfetta".

In pratica quindi la V di Cramer ci dice quanto forte è l'associazione tra le variabili.

Dai test si sono ottenuti i seguenti risultati:

Statistic	DF	Value	Prob
Chi-Square	45	6136.1287	<.0001
Likelihood Ratio Chi-Square	45	6122.9488	<.0001
Mantel-Haenszel Chi-Square	1	4792.1688	<.0001
Phi Coefficient		0.0890	
Contingency Coefficient		0.0887	
Cramer's V		0.0398	

Sample Size = 774614

Il test Chi-Quadro risulta essere fortemente significativo. Viene quindi rifiutata l'ipotesi nulla di "indipendenza tra le variabili". Questo risultato però, come detto precedentemente, potrebbe essere influenzato dall'alto numero di celle della tabella di contingenza, infatti andando a guardare la statistica V di Cramer si registra un valore decisamente basso. Questo potrebbe significare che anche se c'è un'associazione tra azienda ed età dei pazienti questa risulta essere comunque molto debole.

2.3.4 Esperienza e sesso dei medici

Queste variabili si riferiscono al sesso ed agli anni di esperienza dei medici. Se prima le variabili cercavano una corrispondenza tra spesa in medicinali e pazienti, queste ci daranno informazioni sulla correlazione tra spesa in medicinali e medico che ha in cura il paziente. Si è deciso di non considerare la variabile “tipo” in quanto alcuni medici sono sia pediatri che generici. La variabile ExpMed si riferisce agli anni di esperienza lavorativa del medico (anno di osservazione meno anno di iscrizione all’anagrafe). Ai fini dell’analisi è stata creata una nuova variabile che suddivide la variabile ExpMed in 6 fasce, che indicano l’esperienza lavorativa del medico. Questa è stata studiata congiuntamente alla variabile sesso che indica il genere del medico. Per prima cosa abbiamo analizzato come è composta la popolazione, ovvero quale percentuale delle osservazioni occupano i medici uomini e i medici donna sul totale (vedi Grafico 13 e tabella 9).

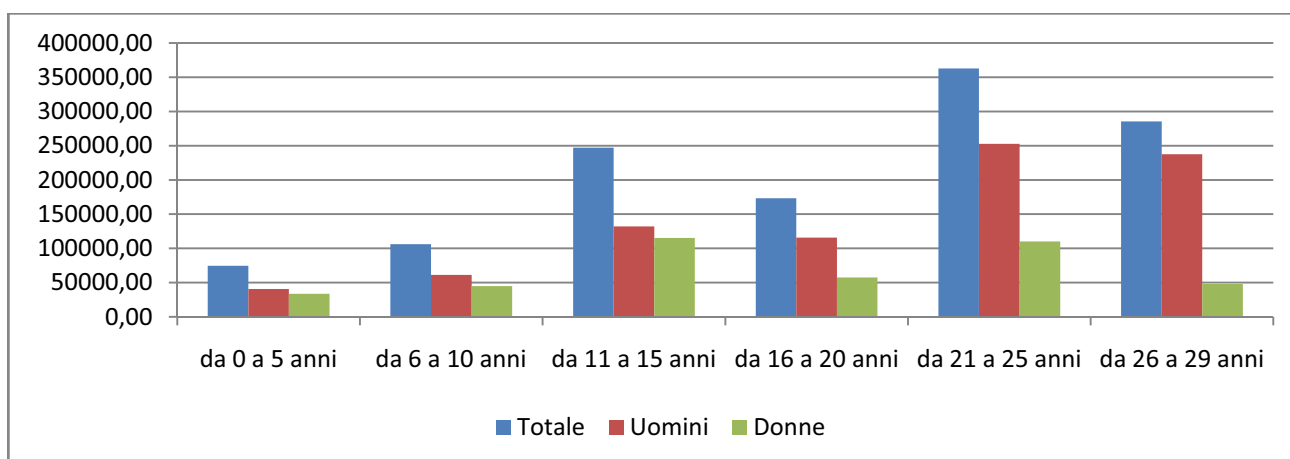


Grafico 13. Popolazione divisa per genere e totale per le fasce d'esperienza dei medici.

Esperienza	Totale	Uomini	Donne	% Uomini	% Donne
da 0 a 5 anni	74500	40779	33721	54,74	45,26
da 6 a 10 anni	106081	61290	44791	57,78	42,22
da 11 a 15 anni	247189	132041	115148	53,42	46,58
da 16 a 20 anni	173316	115736	57580	66,78	33,22
da 21 a 25 anni	362858	252756	110102	69,66	30,34
da 26 a 29 anni	285377	237521	47856	83,23	16,77

Tabella 9. Composizione in % delle fasce di esperienza dei medici per uomini e donne.

Dalla Tabella 9 e dal Grafico 13, si può facilmente notare che la percentuale di medici uomini aumenta all'aumentare degli anni di esperienza, mentre al contrario la percentuale di medici donna diminuisce all'aumentare degli anni di esperienza. Ad esempio nella fascia “da 0 a 5 anni” i medici donna sono il 45,26% mentre quelli uomo sono il 54,74% mentre nell'ultima fascia “da 26 a 29 anni” i medici donna occupano solo il 16,77% mentre gli uomini l'83,23%. Sembra essere presente quindi un'associazione tra anni di esperienza e sesso dei medici; creo quindi una tabella di contingenza (Tabella 10) per verificare l'ipotesi di indipendenza tra le due variabili con un test Chi-Quadro.

Table of sesso by Esperienza							
sesso	Esperienza						Total
	da 0 a 5 anni	da 06 a 10 anni	da 11 anni a 15 anni	da 16 a 20 anni	da 21 a 25 anni	da 25 a 29 anni	
F	33721	44791	115148	57580	110102	47856	409198
	2.70	3.59	9.22	4.61	8.81	3.83	32.75
	8.24	10.95	28.14	14.07	26.91	11.70	
	45.26	42.22	46.58	33.22	30.34	16.77	
M	40779	61290	132041	115736	252756	237521	840123
	3.26	4.91	10.57	9.26	20.23	19.01	67.25
	4.85	7.30	15.72	13.78	30.09	28.27	
	54.74	57.78	53.42	66.78	69.66	83.23	
Total	74500	106081	247189	173316	362858	285377	1249321
	5.96	8.49	19.79	13.87	29.04	22.84	100.00

Tabella 10. Tabella di contingenza per le variabili esperienza e sesso dei medici.

Dalle percentuali di riga le differenze negli anni di esperienza tra medici femmine e medici maschi risultano abbastanza evidenti, la fascia “da 26 a 29 anni” per esempio occupa circa il 12% per i medici femmine e più del 28% per i medici maschi.

Dal test Chi-Quadro si ottengono i seguenti risultati:

Statistic	DF	Value	Prob
Chi-Square	5	65153.9497	<.0001
Likelihood Ratio Chi-Square	5	67705.0455	<.0001
Mantel-Haenszel Chi-Square	1	53911.6303	<.0001
Phi Coefficient		0.2284	
Contingency Coefficient		0.2226	
Cramer's V		0.2284	

Sample Size = 1249321

Il test risulta essere significativo. Viene quindi rifiutata l'ipotesi nulla di indipendenza a favore dell'ipotesi H1 “dipendenza tra le variabili”. La statistica V di Cramer inoltre sembra confermare tale risultato e cioè che ci sia effettivamente una relazione tra il sesso e gli anni di esperienza dei medici.

Andando a studiare l'andamento della spesa media (Grafico 14) si nota una lieve tendenza della stessa all'aumentare degli anni di esperienza fino alla fascia "da 11 a 15 anni" per poi stabilizzarsi. Inoltre, la spesa media dei medici donna risulta essere inferiore alla spesa media dei medici uomo. Questo risultato è interessante perché usualmente quello che ci si aspetta da questo tipo di variabili è un comportamento diverso. Questo può dipendere dal fatto che questo tipo di analisi preliminari non riescono a cogliere le variabili nella visione d'insieme. Sarà interessante nelle successive analisi multivariate osservare se le variabili relative al sesso dei medici ed all'esperienza di essi manterranno questo tipo di andamenti rispetto alla spesa media o se avranno un andamento diverso.

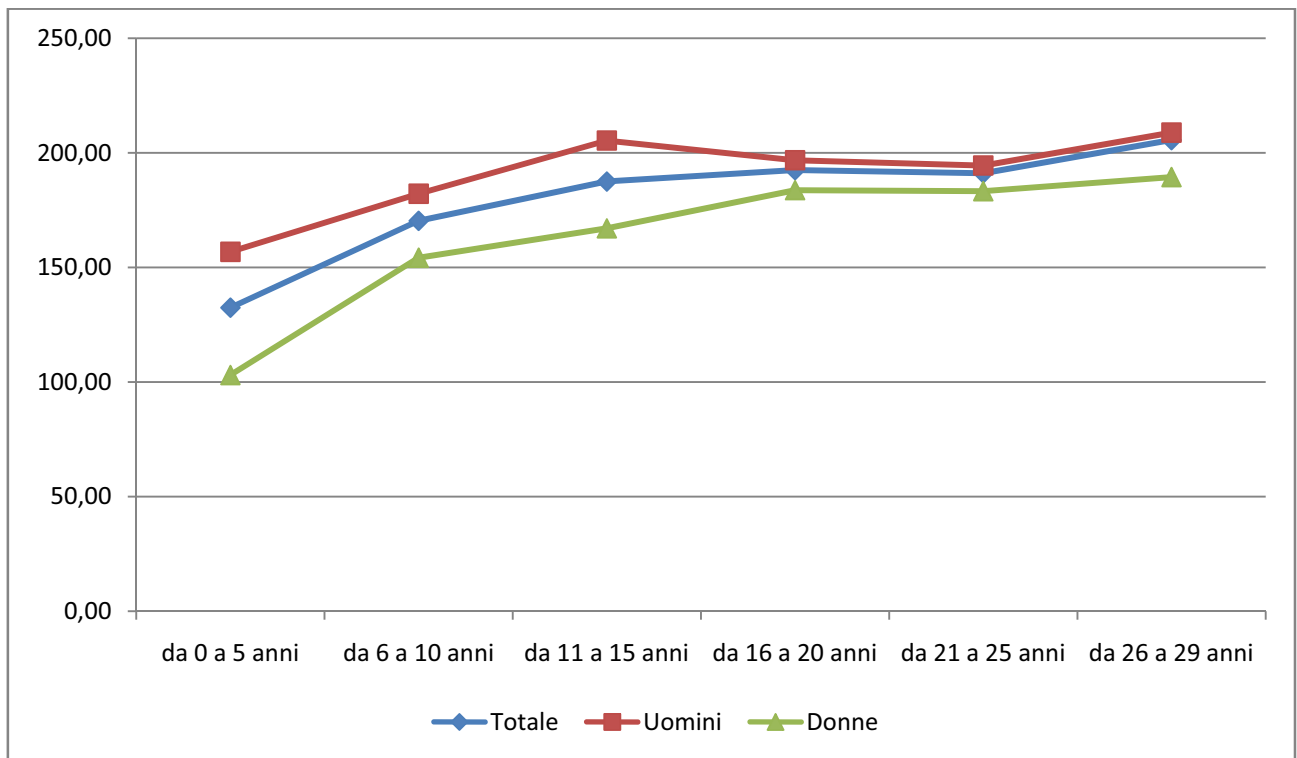


Grafico 14. Spesa media per fasce di esperienza dei medici.

2.3.5 La variabile esenzioni

Oltre al dataset originario soggetto delle analisi preliminari era stato fornito un ulteriore dataset comprendente le esenzioni che si erano registrate prima e durante l'anno 2007 per alcune delle osservazioni contenute nel dataset principale. Il dataset faceva riferimento al solo codice identificativo del paziente e veniva creata una nuova osservazione ogni volta che un paziente veniva reso soggetto di una qualsiasi esenzione; in questo dataset erano presenti 505.932 osservazioni. Questo ha reso difficile l'unione tra questo dataset ed il dataset principale in quanto venivano create osservazioni multiple, non identificabili dalla variabile chiave codificata come la "coppia medico-paziente", ma che differivano tra di loro solo per il tipo di esenzione. Le esenzioni erano definite da una variabile che ne descriveva il tipo e da un codice identificativo. Nel dataset erano descritti e codificati 299 tipi di esenzioni. Il grande numero di esenzioni presenti e l'assenza di una chiave univoca rendevano molto complicata un'analisi approfondita relativamente al tipo di esenzioni che ogni paziente registrava. Per poter analizzare il contributo che i dati contenuti in questo dataset davano alla spesa in medicinali per le aziende sanitarie, è stato deciso di creare una variabile ottenuta dall'unione dei due dataset che desse solo un'indicazione della presenza o meno di esenzioni registrate, sul paziente, prima dell'anno 2007. Per fare ciò, quindi, si è proceduto all'eliminazione di tutte le osservazioni relative all'anno 2007 e si è creata una variabile che assumesse valore 1 se il paziente aveva registrato esenzioni prima del 2007 e 0 altrimenti. Il significato di questa variabile è quindi quello di fare da "Proxy" della salute dei pazienti. Perciò da pazienti che siano già stati precedentemente oggetto di esenzioni ci aspettiamo una spesa in medicinali maggiore rispetto a pazienti che non siano mai stati oggetto di esenzioni, o per lo meno ci aspettiamo che la spesa si manifesti anche per l'anno in corso. La tabella 11 riporta i valori di spesa media e spesa totale in euro suddivisi per presenza o meno di esenzioni prima dell'anno 2007.

Presenza di esenzione prima dell'anno 2007	Numero di osservazioni	Spesa media in €	Spesa totale in €	Deviazione Standard	Standard error
No	953355	84,82	80861614,9	235,91	0,2416
Si	295966	523,02	154795220	711,65	1,3081

Tabella 11. Statistiche descrittive rispetto alla variabile esenzione.

Per verificare se le differenze tra le medie dei due gruppi sono significative si deve procedere con un opportuno t-test. Per prima cosa verifico se è presente eteroschedasticità o omoschedasticità tra le varianze dei due gruppi, il test in Tabella 12 è significativo, rifiuto quindi l'ipotesi nulla di omoschedasticità delle varianze.

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	295965	953354	9.10	<.0001

Tabella 12. Test di omoschedasticità delle varianze per la variabile esenzioni.

Possiamo ora procedere con il test-t, dalla Tabella 13 posso concludere che il test è significativo, le differenze tra le medie dei due gruppi sono quindi statisticamente significative.

esenzione	N	Mean	Std Dev	Std Err	Minimum	Maximum
0	953355	84.8179	235.9	0.2416	0	26227.0
1	295966	523.0	711.6	1.3081	0	43949.5
Diff (1-2)		-438.2	403.0	0.8481		

Method	Variances	DF	t Value	Pr > t
Pooled	Equal	1.25E6	-516.69	<.0001
Satterthwaite	Unequal	316390	-329.41	<.0001

Tabella 13. Test-t SAS output per la variabile esenzione.

I risultati sembrano quindi confermare che da pazienti con esenzioni precedenti ci si possa aspettare spese in medicinali più alte rispetto al resto della popolazione.

3. ANALISI MULTIVARIATE

Nelle pagine precedenti le variabili sono state analizzate singolarmente o a gruppi di due, tramite l'utilizzo di grafici e tabelle di contingenza. Nelle pagine seguenti, invece, studieremo le variabili tramite l'utilizzo delle cosiddette analisi multivariate. Scopo delle analisi multivariate è quello di analizzare simultaneamente le influenze e le relazioni che ogni variabile ha sulle restanti, tramite l'utilizzo di modelli che permettano di non perdere di vista "la visione d'insieme". Sarà quindi interessante notare se questo tipo di analisi darà risultati concordanti o discordanti con quelli ottenuti fino ad ora. In alcuni casi, infatti, studiare le variabili singolarmente non è sufficiente a descriverle completamente e correttamente. Esistono vari tipi di analisi multivariate. Per lo studio del caso in esame ci si è limitati all'utilizzo di due tipologie di modelli: il primo è un modello di regressione lineare semplice ed ha lo scopo di studiare le variabili che influenzano la spesa di medicinali in euro per l'*i*-esima osservazione; il secondo è un modello Logit, che ha lo scopo di studiare la probabilità che si manifesti o non si manifesti la spesa in medicinali per l'*i*-esima osservazione.

3.1 IL MODELLO DI REGRESSIONE LINEARE SEMPLICE

Il modello è stato calcolato sulla base delle sole osservazioni con spesa in medicinali superiore a 0 (numero osservazioni 774.614). Per migliorare l'adattamento del modello ai dati è stata sostituita la variabile risposta "spesatot" con la variabile risposta "logspesa" definita come il logaritmo della variabile "spesatot".

Il modello ottenuto per l'analisi è il seguente:

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \beta_8 X_8 + \beta_9 X_9 + \beta_{10} X_{10} + \varepsilon_i$$

Dove:

Y_i viene definita dalla variabile “logspesa”, cioè il logaritmo della variabile “spesatot”;

β_j indica il coefficiente o parametro relativo alla variabile j ;

β_0 è il coefficiente dell’intercetta del modello;

X_1 si riferisce all’ i -esima osservazione della variabile “EtaAss”;

X_2 si riferisce al sesso dell’assistito vale: 1 se sesso_assistito è uguale a “M”, 0 altrimenti;

X_3 si riferisce all’azienda: vale 1 se la variabile “azienda” è uguale a 1, 0 altrimenti;

X_4 si riferisce all’azienda: vale 1 se la variabile “azienda” è uguale a 2, 0 altrimenti;

X_5 si riferisce all’azienda: vale 1 se la variabile “azienda” è uguale a 3, 0 altrimenti;

X_6 si riferisce all’azienda: vale 1 se la variabile “azienda” è uguale a 4, 0 altrimenti;

X_7 si riferisce all’azienda: vale 1 se la variabile “azienda” è uguale a 5, 0 altrimenti;

L’azienda di riferimento per l’analisi è l’azienda 6; se X_3 , X_4 , X_5 , X_6 , X_7 sono tutte uguali a 0 i risultati sono da riferirsi alle osservazioni con la variabile “azienda” uguale a 6.

X_8 si riferisce all’ i -esima osservazione della variabile “ExpMed”;

X_9 si riferisce alla variabile “sessomed”: vale 1 se la variabile “sessomed” uguale a “M”, 0 altrimenti;

X_{10} si riferisce alla variabile “esenzione” che può assumere i valori 1 o 0 che si riferiscono alla presenza o meno di un’esenzione precedente;

ε_i si riferisce al termine d’errore, assunto a media nulla e varianza costante;

$i = 1, \dots, n$ e n è il numero di osservazioni;

$n = 774.614$.

3.1.1 SAS output

Di seguito vengono elencati i principali output che SAS fornisce per lo studio del modello:

1 - Analisi della varianza:

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	10	810047	81005	48269.0	<.0001
Error	774603	1299933	1.67819		
Corrected Total	774613	2109980			

Tabella 14. Analisi della varianza.

2 - Adattamento del modello ai dati:

Root MSE	1.29545	R-Square	0.3839
Dependent Mean	4.58909	Adj R-Sq	0.3839
Coeff Var	28.22895		

Tabella 15. Indici di adattamento del modello ai dati.

3 - Stima dei parametri:

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	2.62442	0.00584	449.49	<.0001
EtaAss	1	0.03116	0.00007176	434.17	<.0001
sessoass	1	0.10988	0.00298	36.89	<.0001
azienda1	1	0.03108	0.00443	7.02	<.0001
azienda2	1	-0.05915	0.00531	-11.13	<.0001
azienda3	1	-0.03258	0.00676	-4.82	<.0001
azienda4	1	0.00488	0.00403	1.21	0.2258
azienda5	1	-0.04583	0.00568	-8.07	<.0001
ExpMed	1	-0.00083020	0.00020414	-4.07	<.0001
sessomed	1	-0.04067	0.00328	-12.42	<.0001
esenzione	1	1.03923	0.00341	304.40	<.0001

Tabella 16. Stima dei parametri.

3.1.2 Bontà del modello

Dalla Tabella 15 risulta un valore di R-Quadro di 0,3839 che considerata l'alta numerosità è indice di un buon adattamento del modello ai dati. Anche l'analisi della varianza (Tabella 14) sembra confermare l'adeguatezza del modello, infatti il test F è fortemente significativo. Viene quindi rifiutata l'ipotesi nulla " $\beta_1=\beta_2=\beta_3=\beta_4=\beta_5=\beta_6=\beta_7=\beta_8=\beta_9=\beta_{10}=0$ " che mette a confronto il modello completo con il modello ridotto composto dalla sola intercetta. Inoltre quasi tutti i parametri, escluso quello relativo all'azienda 4, sono significativi (vedi Tabella 16). Da questi dati, perciò, possiamo considerare il modello adeguato all'analisi.

3.1.3 Studio dei coefficienti

I coefficienti delle variabili risultano quasi tutti significativi; questo non sorprende in quanto l'alto numero di osservazioni genera Standard Errors molto bassi. Qui di seguito vengono analizzati i coefficienti di ogni variabile.

EtaAss

Variabile	DF	Stima del parametro	Standard Error	T value	Pr > t
EtaAss	1	0,03116	0,00007176	434,17	<,0001

Tabella 17. Stima del parametro relativo alla variabile EtaAss.

Il coefficiente relativo alla variabile EtaAss definita dall'età dei pazienti, vale 0,03116 e risulta fortemente significativo (viene rifiutata perciò l'ipotesi nulla " $\beta_1=0$ "). Secondo il modello, quindi, per ogni anno di età del paziente il logaritmo della spesa in farmaci aumenta di 0,03116 a parità di altre condizioni. Questo è in linea con i risultati ottenuti dalle analisi preliminari che indicavano una forte influenza dell'età sulla spesa, soprattutto per le fasce d'età più elevate.

Sessoass

Variabile	DF	Stima del parametro	Standard Error	T value	Pr > t
Sessoass	1	0,10988	0,00298	36,89	<,0001

Tabella 18. Stima del parametro relativo alla variabile Sessoass.

Il coefficiente relativo alla variabile Sessoass che vale 1 se si tratta di pazienti maschi, 0 altrimenti, è fortemente significativo e vale 0,10988. Quindi, in caso di paziente maschio il logaritmo della spesa in farmaci subirà un incremento di 0,10988 a parità di altre condizioni. Anche questo risultato è in linea con le analisi preliminari. Basti guardare il Grafico 5 dove è chiaro che i pazienti di sesso maschile tendono ad avere spesa media più elevata rispetto ai pazienti di sesso femminile.

Azienda

La variabile azienda in questo modello è definita da 5 variabili dummy che indicano l'azienda di appartenenza dell'osservazione tramite la codifica riportata in Tabella 19.

	Azienda 1	Azienda 2	Azienda 3	Azienda 4	Azienda 5
Se Azienda 1	1	0	0	0	0
Se Azienda 2	0	1	0	0	0
Se Azienda 3	0	0	1	0	0
Se Azienda 4	0	0	0	1	0
Se Azienda 5	0	0	0	0	1
Se Azienda 6	0	0	0	0	0

Tabella 19. Codifica delle variabili dummy riferite all'azienda di appartenenza.

Viene quindi considerata come azienda di riferimento del modello l'azienda 6.

Variabile	DF	Stima del parametro	Standard Error	T value	Pr > t
Azienda 1	1	0,03108	0,00443	7,02	<,0001
Azienda 2	1	-0,05915	0,00531	-11,13	<,0001
Azienda 3	1	-0,03258	0,00676	-4,82	<,0001
Azienda 4	1	0,00488	0,00403	1,21	0,2258
Azienda 5	1	-0,04583	0,00568	-8,07	<,0001

Tabella 20. Stima dei parametri relativi alle variabili Azienda.

I coefficienti relativi ad ogni azienda sono tutti significativi eccetto quello relativo all'Azienda 4. Per questa viene accettata l'ipotesi nulla di non significatività del parametro: l'ipotesi, cioè, che non ci siano differenze significative tra l'Azienda 4 e l'Azienda 6 (azienda di riferimento) a parità di altre condizioni.

Il primo coefficiente della Tabella 20 è positivo. Significa che per le coppie medico-paziente appartenenti alla prima azienda ci sarà un incremento del logaritmo della spesa media rispetto ai pazienti che appartengono all'azienda 6 di 0,03108 a parità di altre condizioni. Gli altri coefficienti sono tutti negativi. Significa che per i pazienti delle aziende 2, 3 e 5 ci sarà una spesa inferiore rispetto a quelli appartenenti all'azienda 6 a parità di altre condizioni. Questi risultati sono in parte in linea con quelli ottenuti dalle analisi preliminari: è vero che l'azienda 1 mostra una spesa media superiore rispetto a tutte le altre aziende, ma è vero anche che dal Grafico 8 l'unica azienda che segnala una spesa media inferiore all'azienda 6 è l'azienda 5; inoltre anche l'azienda 4, il cui coefficiente risulta essere non significativo, segnala una spesa diversa dall'azienda 6, mentre il modello segnala che potrebbero non esserci differenze in termini di spesa per le coppie appartenenti a queste due aziende. Tali risultati differiscono perché le prime erano analisi preliminari che non coglievano la visione d'insieme, mentre queste sono analisi multivariate e studiano le variabili del modello congiuntamente.

ExpMed

Variabile	DF	Stima del parametro	Standard Error	T value	Pr > t
ExpMed	1	-0,0008302	0,00020414	-4,07	<,0001

Tabella 21. Stima del parametro relativo alla variabile ExpMed.

Il coefficiente relativo alla variabile ExpMed che si riferisce agli anni di esperienza dei medici (anno di osservazione meno anno di iscrizione all'anagrafe) è fortemente significativo. Viene, quindi, rifiutata l'ipotesi nulla di non significatività del parametro, e vale -0,0008302. Significa, quindi, che secondo il modello la spesa in medicinali della coppia medico-paziente diminuirà all'aumentare degli anni di esperienza dei medici a parità di altre condizioni. Questo risultato è totalmente discordante con le informazioni ottenute dalle analisi preliminari, dove la spesa media per medicinali aumentava all'aumentare degli anni di esperienza (vedi Grafico 14). Questo è un risultato interessante in quanto, come per i coefficienti delle variabili azienda, dimostra la limitatezza delle analisi preliminari.

Sessomed

Variabile	DF	Stima del parametro	Standard Error	T value	Pr > t
Sessomed	1	-0,04067	0,00328	-12,42	<,0001

Tabella 22. Stima del parametro relativo alla variabile sessomed.

Il coefficiente relativo alla variabile sessomed, che vale 1 se si tratta di medici di sesso maschile e 0 altrimenti, è fortemente significativo e vale -0,04067. Nel caso di medici di sesso maschile, quindi, secondo il modello la spesa avrà una diminuzione a parità di altre condizioni. Anche in questo caso i risultati sono interessanti perché discordi da quelli ottenuti precedentemente (vedi Grafico 14).

Esenzione

Variabile	DF	Stima del parametro	Standard Error	T value	Pr > t
esenzione	1	1,03923	0,00341	304,4	<,0001

Tabella 23. Stima del parametro relativo alla variabile esenzione.

Il coefficiente relativo alla variabile esenzione, che vale 1 se sono presenti esenzioni prima dell'anno di osservazione e 0 altrimenti, è fortemente significativo ed il valore del coefficiente è di 1,03923 che non solo è positivo ma è anche un valore piuttosto alto se confrontato con le stime degli altri coefficienti delle variabili dummy. Ciò, oltre che essere confermato dalle analisi preliminari (vedi Tabella 11) rende bene l'idea che è alla base della variabile esenzione, ovvero che un paziente con esenzioni precedenti a quelle registrate nell'anno di osservazione avrà molto probabilmente una spesa in medicinali superiore rispetto a chi non ha mai registrato esenzioni, a parità di altre condizioni.

3.1.4 Conclusioni sul modello

Il modello costruito sembra approssimare abbastanza bene la spesa in medicinali e dà delle chiare indicazioni su come le variabili studio dell'analisi vadano a influenzare più o meno indirettamente la spesa. Le variabili EtaAss ed esenzione emergono come le principali determinanti della spesa per medicinali. Sono presenti, inoltre, degli scostamenti tra la spesa stimata e la spesa effettiva delle osservazioni. Questo potrebbe dipendere dal fatto che ci siano altre variabili esterne al dataset che vadano ad influenzare la spesa in medicinali. Per migliorare ulteriormente il modello, allo scopo di ottenere una spesa stimata più vicina alla spesa reale, sarebbe, quindi, necessaria una ricerca su nuove variabili poste a spiegare queste differenze. Va ricordato, però, che lo scopo dell'analisi non è quello di creare un modello per predire la spesa in medicinali che si potrebbe verificare in determinate condizioni, ma di capire come le variabili studio dell'analisi vadano ad influenzare la spesa in medicinali sostenuta dalle aziende. Per questo motivo questo modello si può considerare adatto alle analisi.

3.2 IL MODELLO LOGIT

Il modello logit è una soluzione statistica particolarmente adatta nel caso di una variabile risposta dicotomica o binaria, ovvero che può assumere solo due valori. Usualmente questi sono 0 e 1. SAS utilizza per questo tipo di analisi i valori 1 e 2, dove per comodità è stata scelta la codifica 1 “evento”, 2 “non evento”. L’evento considerato è la manifestazione di una spesa in medicinali. Quindi: 1 se per l’osservazione è stata effettivamente registrata una spesa in medicinali, 2 altrimenti. Per far ciò è stata creata una variabile dummy relativa alla presenza o meno di una spesa, denominata *spesatotd*, che assume valore 1 se non è stata registrata alcuna spesa, 0 altrimenti. La codifica di tale variabile secondo SAS è spiegata in Tabella 24. Il dataset di riferimento del modello è l’intero dataset con un totale di 1.249.321 osservazioni.

Response Profile		
Ordered Value	<i>spesatotd</i>	Total Frequency
1	1	774614
2	0	474707

Tabella 24. Codifica e frequenza della variabile *spesatotd*.

Il modello logit ipotizza che:

$$\pi = \Pr(Y = 1 | \mathbf{x})$$

$$\text{logit}(\pi) \equiv \log\left(\frac{\pi}{1 - \pi}\right) = \alpha + \boldsymbol{\beta}'\mathbf{x}$$

Dove \mathbf{X} è la matrice dei regressori del modello;

α è l’intercetta del modello;

$\boldsymbol{\beta} = (\beta_1, \dots, \beta_i)'$ è il vettore dei coefficienti;

Y è la variabile risposta (nel nostro caso la variabile dummy *spesatotd*).

Il modello logit ci fornisce quindi la probabilità (π) che l’evento accada, e tale probabilità è calcolabile come:

$$\Pr(Y=1|\mathbf{x}) = \frac{e^{\alpha + \boldsymbol{\beta}'\mathbf{x}}}{1 + e^{\alpha + \boldsymbol{\beta}'\mathbf{x}}}$$

3.2.1 SAS output

Qui di seguito vengono elencati i principali output che SAS fornisce per lo studio del modello:

Adattamento del modello ai dati:

R-Square	0.1645	Max-rescaled R-Square	0.2238
-----------------	--------	------------------------------	--------

Tabella 25. Coefficienti di R-Quadro del modello logit.

Test per l'ipotesi di nullità dei parametri:

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	224575.338	10	<.0001
Score	197852.454	10	<.0001
Wald	162332.214	10	<.0001

Tabella 26. Test per l'ipotesi di nullità dei parametri del modello logit.

Stima degli odds ratio:

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
EtaAss	1.025	1.025	1.025
sessoass 0 vs 1	1.512	1.500	1.524
azienda 01 vs 06	0.876	0.865	0.886
azienda 02 vs 06	0.834	0.822	0.846
azienda 03 vs 06	0.815	0.801	0.830
azienda 04 vs 06	0.832	0.823	0.841
azienda 05 vs 06	0.925	0.911	0.939
ExpMed	0.993	0.993	0.994
sessomed 0 vs 1	1.073	1.064	1.083
esenzione 0 vs 1	0.220	0.217	0.223

Tabella 27. Stima degli odds ratio.

3.2.2 Stima degli odds ratio

Considerando X una variabile esplicativa dicotomica di una variabile risposta, anch'essa dicotomica, si ha che il coefficiente di regressione di una variabile misura la variazione nel logit di Y corrispondente al possesso dell'attributo X :

$$\log(\psi) \equiv \log(\psi(X=1, X=0)) = \text{logit}(X=1) - \text{logit}(X=0) = (\alpha + 1 \times \beta) - (\alpha + 0 \times \beta) = \beta$$

Dove ψ è detto odds ratio o rapporto crociato.

Quindi:

$$\psi = \exp(\beta)$$

Nel caso di variabili quantitative si ha invece che:

$$\psi = \exp(\beta c)$$

Dove c è l'incremento della variabile quantitativa.

Per l'interpretazione degli odds ratio si considera:

$\psi > 1$ il coefficiente della variabile influisce positivamente sulla probabilità che l'evento si verifichi;

$\psi < 1$ il coefficiente della variabile influisce negativamente sulla probabilità che l'evento si verifichi;

$\psi = 1$ il coefficiente della variabile non influisce sulla probabilità che l'evento si verifichi.

Per il calcolo degli odds ratio nel caso di variabili nominali a più modalità la stima dell'odds ratio dipende dalla modalità di riferimento dei parametri, per esempio nel caso della variabile azienda per cui sono state costruite 5 variabili dummy si fa riferimento all'azienda 6. Ovvero i parametri relativi a queste variabili saranno il logaritmo dell'odds ratio dell'azienda presa in considerazione vs l'azienda 6. Quindi per l'azienda 1 ad esempio si avrà:

$$\log(\psi(\text{azienda } 01, \text{azienda } 06)) = \text{logit}(\text{azienda } 01) - \text{logit}(\text{azienda } 06)$$

Il cui risultato dipende dalla codifica scelta per la variabile. L'interpretazione dell'odds ratio sarà quindi da riferirsi al confronto con il modello che considera la modalità di riferimento.

3.2.3 Bontà del modello

Il valore di R-Quadro descritto dalla tabella 25 è di 0,1645. Considerando che tutti i modelli di tipo logit presentano valori di R-Quadro abbastanza bassi e vista l'alta numerosità delle osservazioni possiamo considerare il valore di R-Quadro sintomo di un buon adattamento del modello ai dati. Inoltre tutti gli odds ratio, risultano significativi (vedi Tabella 27) e tutti i test relativi alla nullità dei parametri, ovvero che mettono a confronto il modello completo contro il modello composto dalla sola intercetta, sono significativi (Tabella 26). Da questi risultati il modello sembra adatto all'analisi.

3.2.4 Studio degli odds ratio

Studiando gli odds ratio si notano alcune differenze sull'influenza che le variabili esplicative hanno sulla variabile risposta rispetto al modello di regressione lineare.

EtaAss

Variabile	Stima dell'odds ratio	95% Wald Confidence Limits	
EtaAss	1,025	1,025	1,025

Tabella 28. Stima dell'odds ratio relativo alla variabile EtaAss.

L'odds ratio relativo alla variabile EtaAss, definita dall'età dei pazienti, vale 1,025 ed è significativo. Indica che all'aumentare dell'età la probabilità che si manifesti una spesa in medicinali per la coppia medico-paziente aumenta a parità di alte condizioni. Questo è in linea sia con i risultati ottenuti dalle analisi preliminari, sia da quelli ottenuti tramite la stima di un modello di regressione lineare. Entrambi, infatti, indicavano un aumento della spesa all'aumentare degli anni di età.

Sessoass

Variabile	Stima dell'odds ratio	95% Wald Confidence Limits	
sessoass 0 vs 1	1,512	1,5	1,524

Tabella 29. Stima dell'odds ratio relativo alla variabile sessoass.

L'odds ratio espresso in tabella 29 vale 1,512 ed è significativo. Indica che per i pazienti di sesso femminile la probabilità che la spesa in medicinali si manifesti aumenta. Questo risultato è molto interessante perché va in contrasto sia con le analisi preliminari sia con i risultati ottenuti dal modello di regressione. Rispetto ai pazienti di sesso maschile i pazienti di sesso femminile hanno una maggiore probabilità di realizzare una spesa, ceteris paribus. Tuttavia, realizzata una spesa i pazienti di sesso femminile tendono a spendere un valore inferiore in medicinali rispetto a pazienti maschi.

Azienda

L'azienda di riferimento per il modello è l'azienda 6.

Variabile	Stima dell'odds ratio	95% Wald Confidence Limits	
Azienda 01 vs 06	0,876	0,865	0,886
Azienda 02 vs 06	0,834	0,822	0,846
Azienda 03 vs 06	0,815	0,801	0,830
Azienda 04 vs 06	0,832	0,823	0,841
Azienda 05 vs 06	0,925	0,911	0,939

Tabella 30. Stima degli odds ratio relativi alle variabili azienda per il modello logit.

Gli odds ratio espressi in tabella 30 sono tutti inferiori a 1 e sono significativi. Indicano quindi che, a parità di altre condizioni, per le coppie medico-paziente appartenenti all'azienda 6 la probabilità che una spesa si manifesti è superiore rispetto alle altre aziende. Questi risultati sono interessanti perché contrastano sia con le analisi preliminari sia con i risultati ottenuti dal modello di regressione. Nonostante i pazienti appartenenti all'azienda 1 tendano a spendere un valore superiore, nell'azienda 6 si tende a prescrivere di più rispetto a tutte le altre aziende.

ExpMed

Variabile	Stima dell'odds ratio	95% Wald Confidence Limits	
ExpMed	0,993	0,993	0,994

Tabella 31. Stima dell'odds ratio relativo alla variabile ExpMed per il modello logit.

L'odds ratio relativo alla variabile ExpMed assume valore 0,993 ed è significativo. Ciò significa che all'aumentare dell'esperienza dei medici la probabilità che si manifesti una spesa in medicinali diminuisce a parità di alte condizioni. Questo risultato è in linea con i risultati ottenuti dal modello di regressione lineare che segnalava una diminuzione della spesa in medicinali all'aumentare degli anni di esperienza dei medici.

Sessomed

Variabile	Stima dell'odds ratio	95% Wald Confidence Limits	
Sessomed 0 vs 1	1,073	1,064	1,083

Tabella 32. Stima dell'odds ratio relativo alla variabile sessomed per il modello logit.

L'odds ratio relativo alla variabile sessomed è significativo, e vale 1,073. Quindi per i medici di sesso femminile la probabilità di manifestazione di una spesa in medicinali aumenta, a parità di altre condizioni. Questi risultati rispecchiano quelli ottenuti dal modello di regressione lineare dove la spesa diminuiva in presenza di medici di sesso maschile.

Esenzione

Variabile	Stima dell'odds ratio	95% Wald Confidence Limits	
Esenzione 0 vs 1	0,220	0,217	0,223

Tabella 33. Stima dell'odds ratio relativo alla variabile esenzione per il modello logit

L'odds ratio della variabile esenzione è significativo e vale 0,220. In particolare, per i pazienti che hanno già registrato esenzioni prima dell'anno di osservazione, la probabilità che una spesa si manifesti aumenta a parità di altre condizioni. Come nel modello di regressione lineare rispetto all'entità della spesa, in questo modello la variabile esenzione dà un'indicazione importante sulla probabilità che la spesa di medicinali si manifesti o meno.

3.2.5 Conclusioni sul modello

Questo modello dà delle chiare indicazioni su come le variabili studiate vadano ad influenzare la probabilità che una spesa in medicinali si manifesti o meno. Nonostante il modello si possa considerare adatto all'analisi sarebbe interessante poter disporre di altre variabili poste a descrivere in modo più chiaro la probabilità di manifestazione di una spesa in medicinali per l'individuo. Anche in questo caso, per migliorare il modello sarebbe quindi necessaria un'ulteriore ricerca rispetto a variabili poste a descrivere la probabilità di avere prescrizioni di medicinali. È interessante però notare come la probabilità di manifestazione di una spesa non sia del tutto legata all'entità di essa, infatti sia la variabile sesso che la variabile azienda danno risultati contrastanti nei due modelli, risultati che probabilmente stanno ad indicare come seppur una variabile possa incidere sull'entità della spesa non è detto che incida allo stesso modo sulla probabilità che essa si manifesti.

4. CONCLUSIONI

In questo lavoro di tesi sono stati osservati gli andamenti sulla spesa in medicinali sostenuta dalle aziende sanitarie del Friuli Venezia Giulia nel 2007 relativamente ad un certo insieme di variabili. Sono state realizzate delle analisi preliminari e delle analisi multivariate. Queste ultime in particolare hanno evidenziato dei risultati e delle conclusioni non sempre coerenti con le analisi preliminari. Questo può essere un indicatore di come possa essere necessario affrontare le variabili di interesse sotto aspetti diversi.

Alcuni di questi aspetti hanno evidenziato la limitatezza delle analisi preliminari. Per le variabili sesso ed esperienza dei medici infatti i risultati delle analisi preliminari si sono rivelati contraddittori, indicazione di come questo tipo di analisi possano essere limitate rispetto ad analisi di tipo multivariato.

Nel considerare i risultati ottenuti va tenuto conto anche della grande numerosità del dataset. Infatti, lavorare con 1.249.321 osservazioni non si è rivelato del tutto semplice, in quanto alcuni di questi risultati sono stati direttamente influenzati da essa.

Se le osservazioni erano in numero decisamente elevato, non si può dire lo stesso delle variabili di interesse. Per quanto si sia rivelato interessante studiare queste variabili spesso si è notata la mancanza di indicatori che descrivessero lo stato di salute del paziente. Abbiamo affrontato i dati tramite variabili che descrivevano aspetti come sesso ed età del paziente, sesso del medico, azienda di appartenenza eccetera, ma non abbiamo potuto studiare dati collegati ad aspetti che potessero considerarsi più determinanti relativamente alla spesa in medicinali. L'unica variabile indicatore dello stato di salute dei pazienti utilizzata è stata la variabile esenzione. Essa infatti è risultata la principale determinante nel modello di regressione per la spesa in medicinali. Questo si può considerare il più grande limite di queste analisi, le variabili studiate infatti si sarebbero rivelate molto più interessanti se affiancate ad informazioni sullo stato effettivo di salute del paziente.

Bibliografia

- *<http://www.ars.sanita.fvg.it>;*
- *SAS Help and Documentation;*
- *Luigi Fabbris: Statistica multivariata, analisi esplorativa dei dati;*

I dati sono stati forniti dal Cefomed (Centro Regionale di Formazione per l'Area delle Cure Primarie) di Monfalcone (GO), che ha autorizzato l'uso per soli fini scientifici. Si ringrazia a tale riguardo il dott. Paolo Piergentili, il dott. Fabio Samani ed il dott. Giorgio Simon.

Alcuni ringraziamenti

Ringrazio il Professor. Paccagnella per essere sempre stato disponibile durante questo lavoro di tesi.

Ringrazio la mia famiglia per avermi permesso di frequentare questi primi anni di università.

I miei amici per la compagnia che non si è mai fatta mancare, in particolare: Steve (che continua a dare spettacolo), Tommaso (anche quando è molesto), Samuele (che nonostante tutto è sempre disponibile a uscire), Stefano (che tira pacchi su pacchi), Alessandro (per tutte le serate al cinema), Juan (per tutte le serate davanti alla play).

Ringrazio tutte le persone che ho conosciuto durante i mesi di erasmus per averla trasformata in un'esperienza indimenticabile.

Un ringraziamento va alle amicizie strette in questi anni di triennale: Indy, Paola, Ricky, Holly, Lu, Helena, Checco; in particolare ringrazio Veronica per le ore in aula studio e per gli incessanti "gossip" che ho dovuto a mio malgrado ascoltare e Giorgia per le sue fobie e per tutta la compagnia che mi ha portato negli ultimi due anni.