

Università degli Studi di Padova
Dipartimento di Scienze Statistiche
Corso di Laurea Magistrale in
Scienze Statistiche



**I costi della Sanità Pubblica in Friuli-Venezia Giulia: profilo degli
assistiti e previsione dei costi sulla base di dataset amministrativi**

Relatore: Prof. Omar Paccagnella
Dipartimento di Scienze Statistiche

Laureanda: Stefania Scapin
Matricola: 1056092

Anno Accademico 2014/2015

Indice

Indice delle Tabelle	v
Indice delle Figure	vii
Introduzione	1
Capitolo 1	5
1.1 Struttura generale ASS	5
1.2 Analisi esplorative	8
1.2.1 Il Medico di Medicina Generale	8
1.2.2 Anagrafica assistiti	11
1.2.3 Esenzioni	13
1.2.4 Servizio ADI	16
1.2.5 Ricoveri ospedalieri	17
1.2.6 Farmaci	20
1.2.7 Specialistiche	24
1.3 Variabili utilizzate nei modelli	26
1.3.1 Variabili risposta	26
1.3.2 Variabili esplicative	28
Capitolo 2	31
2 I modelli	31
2.1 Modelli per variabili binarie	31
2.1.1 Il modello probit	31
2.2 Modelli multilivello	35
2.2.1 Modello multilivello a 2 livelli	37
2.2.1.1 Notazione	37
2.2.2 Coefficiente di Correlazione Intraclasse	37
2.2.3 Modello a intercetta casuale con covariate di livello-1 (Random Intercept Model)	38
2.2.4 Modello a intercetta casuale con covariate di livello-1 e livello-2	39

2.2.5	Modello a pendenza casuale (Random Slope Model)	40
2.2.6	Metodi di Stima	41
2.2.7	Misure di bontà di adattamento	42
2.3	Il modello Tobit	43
2.3.1	Introduzione	43
2.3.2	La distribuzione normale troncata	44
2.3.2.1	Momenti della distribuzione normale troncata	46
2.3.2.2	Modello di regressione troncato	47
2.3.3	Il modello Tobit I	47
2.3.3.1	Introduzione al modello	47
2.3.3.2	Valori attesi del modello tobit	48
2.3.3.3	La funzione di verosimiglianza	49
2.4	La statistica multivariata	51
2.4.1	Vantaggi nell'usare i modelli multivariati	53
2.5	Tobit multivariato	54
2.6	Il modello di Heckman	55
2.6.1	Stima	57
Capitolo 3		59
3 Stima dei modelli		59
3.1	Stima del modello probit e probit multilivello	60
3.1.1	Ricoveri ospedalieri	61
3.1.2	Visite specialistiche	65
3.1.3	Farmaci	69
3.2	Stima dei modelli di regressione lineare e regressione multilivello	72
3.2.1	Ricoveri ospedalieri	73
3.2.2	Visite specialistiche	78
3.2.3	Farmaci	83
3.3	Stima del modello Tobit univariato e Tobit multivariato	88
3.3.1	Ricoveri ospedalieri	90
3.3.2	Visite specialistiche	93
3.3.3	Farmaci	96
3.4	Stima del modello di Heckman	99

3.4.1	Ricoveri ospedalieri	100
3.4.2	Visite specialistiche	104
3.4.3	Farmaci	108
Capitolo 4		113
4	Previsione	113
Conclusioni		123
Appendice A		129
Bibliografia		131
Sitografia		133
Ringraziamenti		135

Indice delle Tabele

Tabella 1 – Suddivisione MMG nei distretti sanitari e nelle ASS	8
Tabella 2 – Descrizione patologia esenzioni	14
Tabella 3 – Classificazione codici DRG	19
Tabella 4 – Classificazione I° livello ATC	21
Tabella 5 – Consumo territoriale classe A-SSN 2007 (popolazione pesata) per il I° livello ATC	23
Tabella 6 – Codici prestazioni visite specialistiche	25
Tabella 7 – Contenuto di alcuni database amministrativi	27
Tabella 8 – Variabili esplicative	30
Tabella 9 – Stime dei modelli probit standard e multilivello, con o senza covariate del medico, dei ricoveri ospedalieri	61
Tabella 10 - Stime dei modelli probit standard e multilivello, con o senza covariate del medico, delle visite specialistiche	65
Tabella 11 - Stime dei modelli probit standard e multilivello, con o senza covariate del medico, della spesa per farmaci	69
Tabella 12 – Stime dei modelli di regressione semplice e multilivello, con o senza covariate del medico, dei ricoveri ospedalieri	73

Tabella 13 - Stime dei modelli di regressione semplice e multilivello, con o senza covariate del medico, delle visite specialistiche	78
Tabella 14 - Stime dei modelli di regressione semplice e multilivello, con o senza covariate del medico, della spesa per farmaci	83
Tabella 15 - Stime del modello Tobit univariato e multivariato, della spesa per ricoveri ospedalieri	90
Tabella 16 - Stime del modello Tobit univariato e multivariato, della spesa per visite specialistiche	93
Tabella 17 - Stime del modello Tobit univariato e multivariato, della spesa per farmaci	96
Tabella 18 - Stima del modello di Heckman della spesa per ricoveri ospedalieri	100
Tabella 19 - Stima del modello di Heckman della spesa per visite specialistiche	104
Tabella 20 - Stima del modello di Heckman della spesa per farmaci	108
Tabella 21 - Stima reale, previsione e alcuni esempi di spesa	120

Indice delle Figure

Figura 1 – Suddivisione dei MMG per genere	9
Figura 2 – Suddivisione campione dei MMG per classi di età e genere	9
Figura 3 – Anni dal conseguimento della laurea, suddivisione per genere MMG	10
Figura 4 – Distribuzione del campione per genere	11
Figura 5 – Distribuzione del campione per classi di età	12
Figura 6 – Luogo servizio ADI	16
Figura 7 – Tipologia di ricoveri ospedalieri	18
Figura 8 – Composizione percentuale del costo totale dei farmaci classificazione ATC I° livello	22
Figura 9 – Previsione della spesa per ricoveri ospedalieri e popolazione totale	115
Figura 10 – Previsione della spesa per ricoveri ospedalieri e popolazione over 75	116
Figura 11 – Previsione della spesa per visite specialistiche, popolazione totale e over 75	117
Figura 12 - Previsione della spesa per farmaci, popolazione totale e over 75	118

Introduzione

Il Servizio Sanitario Nazionale (SSN) è un sistema di strutture e servizi che ha lo scopo di garantire a tutti i cittadini, in condizioni di uguaglianza, l'accesso universale all'erogazione equa delle prestazioni sanitarie, in attuazione dell'art.32 della Costituzione, che recita:

“La Repubblica tutela la salute come fondamentale diritto dell'individuo e interesse della collettività, e garantisce cure gratuite agli indigenti. Nessuno può essere obbligato a un determinato trattamento sanitario se non per disposizione di legge. La legge non può in nessun caso violare i limiti imposti dal rispetto della persona umana”.

Il Ministero della Salute, che ha il compito di provvedere alla salute pubblica, rappresenta l'organo principale del Servizio Sanitario Nazionale.

Attraverso il Piano sanitario nazionale approvato dal Parlamento, il Ministero definisce degli obiettivi generali da raggiungere per la salute pubblica, l'importo del fondo sanitario, il metodo di ripartizione di tale fondo alle Regioni, le quali, mediante una rete di Aziende Unità Sanitarie Locali (ASL), garantiscono l'attività dei servizi sanitari. Le ASL, infatti, sono aziende con personalità giuridica pubblica e sono centri di imputazione di autonomia imprenditoriale. Con la legge di riforma (D.Lgs. 299/99), infatti, la precedente Unità Sanitaria Locale (USL) è divenuta azienda dotata di autonomia organizzativa, gestionale, tecnica, amministrativa, patrimoniale e contabile.

In base al “principio di sussidiarietà” costituzionale, il Servizio Sanitario è articolato secondo diversi livelli di responsabilità e di governo:

- Livello centrale: lo Stato ha la responsabilità di assicurare a tutti i cittadini il diritto alla salute mediante un forte sistema di garanzie, attraverso i Livelli Essenziali di Assistenza (LEA). Questi ultimi sono costituiti dall'insieme delle attività, dei servizi e delle prestazioni che il SSN eroga a tutti i cittadini gratuitamente o con il pagamento di un ticket, indipendentemente dal reddito e dal luogo di residenza.
- Livello regionale: le Regioni hanno la responsabilità diretta della realizzazione degli ideali governativi e della spesa, per il raggiungimento degli obiettivi di salute del Paese. Le Regioni hanno competenza esclusiva nella regolamentazione ed organizzazione dei servizi e di attività destinate alla tutela della salute e dei criteri di finanziamento delle

Aziende Sanitarie Locali e delle Aziende Ospedaliere (anche in relazione al controllo di gestione e alla valutazione della qualità delle prestazioni sanitarie nel rispetto dei principi generali fissati dalle leggi dello Stato).

Il Servizio sanitario nazionale è finanziato attraverso la fiscalità generale e, in particolare, con imposte dirette (addizionale IRPEF, imposta sul reddito delle persone fisiche, ed IRAP, imposta regionale sulle attività produttive) ed indirette (compartecipazione all’IVA, accise sulla benzina), il cui gettito rappresenta un’entrata propria delle Regioni, e residuali trasferimenti erariali. Le risorse sono ripartite tra le Regioni ogni anno dal CIPE (Comitato Interministeriale per la Programmazione Economica) su proposta del Ministero della Salute, d’intesa con le stesse Regioni. Il riparto viene determinato sulla base della numerosità della popolazione residente, e ponderato anche in relazione ai consumi specifici sanitari per età. Altre fonti di finanziamento per le ASL sono redditi, lasciti, donazioni, mutui ma per durata non superiore ai 10 anni (vige infatti un divieto da parte delle ASL di indebitarsi). Le Aziende Sanitarie Locali dispongono anche di entrate proprie derivanti, per esempio, dai ticket sanitari e dalle prestazioni rese a pagamento. Il cittadino, infatti, contribuisce alla spesa sanitaria, attraverso la fiscalità generale, in proporzione al proprio reddito e, laddove non abbia diritto ad una esenzione, attraverso il pagamento di uno specifico ticket per alcune prestazioni comprese nei Livelli Essenziali di Assistenza. Il ticket, introdotto nel 1982, oltre a rappresentare un ulteriore finanziamento dell’assistenza sanitaria a livello territoriale, serve anche a limitare l’inappropriatezza, vale a dire il ricorso a prestazioni sanitarie non necessarie.

Attualmente i ticket riguardano:

- Le prestazioni specialistiche (visite, esami strumentali e analisi di laboratorio)
- Le prestazioni di pronto soccorso
- Le cure termali
- Le prestazioni farmaceutiche

Al cittadino può essere riconosciuto il diritto all’esenzione dal ticket sulla base di particolari situazioni di reddito associate all’età o alla condizione sociale, in presenza di determinate patologie (croniche o rare) o del riconoscimento dello stato di invalidità e in altri casi particolari (gravidanza, diagnosi precoce di alcuni tumori, accertamento dell’HIV, etc.).

Il Nomenclatore dell'assistenza specialistica ambulatoriale riporta le prestazioni specialistiche ambulatoriali erogabili dal Servizio sanitario nazionale e definisce il livello essenziale di assistenza. Il Nomenclatore è stato approvato con il Decreto del Ministro della Sanità del 22 luglio 1996 e contiene una lista di prestazioni identificate da uno specifico codice numerico e le relative tariffe di riferimento nazionale.

Per quanto riguarda le Aziende Ospedaliere, il finanziamento da parte della Regione viene effettuato con la remunerazione a tariffa sulla base dei D.R.G. (Diagnosis Related Groups). Il sistema DRG si può definire come un sistema isorisorse, in quanto orientato a descrivere la complessità dell'assistenza prestata al paziente, partendo dal principio che malattie simili, trattate in reparti ospedalieri simili, comportino approssimativamente lo stesso consumo di risorse umane e materiali.

E' molto importante regolamentare e organizzare il finanziamento delle ASL, poiché la spesa sanitaria pubblica è una delle voci più ingenti che pesano sul bilancio del nostro Paese: essa costituisce, infatti, circa il 7% del PIL (Istat, 2012). Anche se la spesa sanitaria italiana (2.344\$ pro capite, anno 2011) è molto inferiore rispetto ad altri Paesi europei, a parità di potere di acquisto, come Paesi Bassi (spesa pro capite 4.054\$, anno 2011) e Lussemburgo (3.794\$ spesa pro capite, anno 2011), un obiettivo che da sempre il Governo italiano si pone è quello di ridurre i costi per la sanità. Per tale motivo è molto importante studiare la spesa pubblica sanitaria per cercare di contenere i costi andando a focalizzarsi sugli eventuali sprechi o comportamenti scorretti delle persone coinvolte nel Sistema Sanitario Nazionale.

Questo elaborato vuole, quindi, analizzare la spesa sanitaria attraverso la stima di modelli che permettano di stilare il profilo dei pazienti che più gravano sul Sistema Sanitario. Inoltre, un altro obiettivo è quello di individuare se una possibile variabilità tra medici di medicina generale può influire negativamente sulla spesa pubblica. Infine, ipotizzando che certe caratteristiche dei pazienti non cambino negli anni, fornire una previsione dei costi.

L'analisi riguarda la spesa per ricoveri ospedalieri, visite specialistiche e farmaci prescritti agli abitanti del Friuli-Venezia Giulia, anno 2007. I dataset a disposizione sono di tipo amministrativo e racchiudono caratteristiche anagrafiche degli assistiti e dei medici di medicina generale a loro associati. La spesa di una Regione è sicuramente influenzata dall'età o dal genere degli assistiti, ma un altro fattore che può contraddistinguere tale spesa è la diffusione di patologie a livello territoriale. Una proxy di tale caratteristica può essere

l'esenzione per patologia che indica quanto una malattia è diffusa e quanto grava sulla sanità regionale.

La prima parte dell'elaborato si concentrerà sull'analisi esplorativa dei vari dataset, dai quali sono stati esclusi i bambini fino a 14 anni (perché seguiti da pediatri) e quei pazienti a cui non era associato nessun medico di medicina generale. Inoltre, verrà descritta la suddivisione Azienda-Distretto che caratterizza il Friuli-Venezia Giulia (essendo una regione a statuto speciale, l'ASL cambia denominazione passando a ASS, Azienda per i Servizi Sanitari). A completare le analisi esplorative saranno una serie di valutazioni riguardanti i tre costi in esame: spese farmaceutiche, ricoveri ospedalieri e visite specialistiche. Queste spese sono molto importanti perché sono usate come variabile risposta nei modelli statistici della mia analisi.

Nel Capitolo 2 vi sarà una descrizione teorica dei modelli utilizzati e nel Capitolo 3 le stime vere e proprie di questi modelli. Inizialmente viene stimato un modello probit e un modello probit multilivello, entrambi senza e con covariate dei Medici di Medicina Generale. In questo modo ho cercato di analizzare la presenza di un'eventuale variabilità di comportamento tra i medici stessi. Successivamente, data la natura continua delle variabili risposta, ho stimato un modello di regressione lineare semplice e multilivello, con e senza covariate del medico. Infine, ho stimato i modelli Tobit univariati, Tobit multivariati e la loro estensione Tobit II (modello di Heckman).

Il Capitolo 4, invece, riguarda un esercizio di previsione dei tre costi sulla base della previsione della popolazione del Friuli-Venezia Giulia dal 2011-2060 predisposta dall'Istat¹ con un modello per componenti (cohort component model).

Il Capitolo finale racchiude le conclusioni, le quali rivedono i principali risultati dei modelli stimati e mettono in evidenza le differenze riscontrate tra i vari modelli e tra i diversi tipi di spesa. Inoltre, viene descritta in breve la previsione dei tre costi in analisi, anche in relazione alla previsione della popolazione friulana. Infine, si conclude l'elaborato con alcune considerazioni circa pro e contro dell'uso di dataset amministrativi.

¹ www.demo.istat.it

Capitolo 1

1.1 Struttura generale ASS

L'obiettivo principale che ogni società civile si pone è la tutela della salute dei cittadini, il suo grado di attuazione misura in gran parte il benessere di cui gode la popolazione. Il Friuli-Venezia Giulia è una di quelle regioni che meglio si contraddistingue in questa direzione e nel 2007 conta, per una popolazione di circa 1.200.000 abitanti, 15 milioni di prestazioni ambulatoriali, 450 mila accessi al pronto soccorso, 200 mila ricoveri ospedalieri, 90 mila interventi chirurgici, 7 milioni di accessi al medico di medicina generale; l'aspettativa di vita è di 76 anni per gli uomini e di 82 per le donne².

La regione Friuli-Venezia Giulia, nel 2007, era suddivisa in sei Aziende per i Servizi Sanitari, che coprivano tutte le zone della regione:

- ASS 1 – Triestina
- ASS 2 – Isontina
- ASS 3 – Alto Friuli
- ASS 4 – Medio Friuli
- ASS 5 – Bassa Friulana
- ASS 6 – Friuli Occidentale

Secondo i principi contenuti nell'ordinamento Nazionale Regionale del Servizio Sanitario, fatti propri nell' "Atto Aziendale", l'Azienda "riconosce come propria fondamentale missione il raggiungimento da parte della popolazione del più elevato livello possibile di salute intesa come benessere fisico, mentale e sociale" e persegue, a sostegno di ciò "gli obiettivi sanitari, di

² Guida ai servizi per la salute in Friuli-Venezia Giulia

salute, di prestazioni, di struttura...previsti dalla programmazione nazionale e regionale, ... assicurando i livelli di assistenza previsti e garantendo gli interventi di prevenzione, diagnostici terapeutici, educativi più indicati, con il minimo rischio possibile per la persona e con il miglior grado di soddisfazione possibile per gli interventi ricevuti, per i contatti umani con gli operatori e per gli esiti.”

Ogni azienda ha le sue direzioni ma le ASS del Friuli hanno fatto proprie alcune linee guida che intendono sviluppare a beneficio dei cittadini, come:

- assicurare equità nella salute, abbattendo il divario attribuibile alle differenti possibilità di accesso alle cure ed ai servizi tra le diverse aree geografiche e tra i diversi gruppi di popolazione presenti nel territorio aziendale, anche rispetto agli standard di salute e di servizi presenti nelle altre aree regionali;
- migliorare la qualità della vita, garantendo che le capacità fisiche e mentali si possano sviluppare e realizzare in modo da ricavare il massimo beneficio possibile attraverso la diffusione di stili di vita salutari;
- arricchire di salute la vita, mediante un' efficace riduzione della malattia e della disabilità, in particolare verso le frazioni di popolazione più anziane;
- aggiungere anni alla vita, riducendo le morti premature.

Le attività dell'Azienda si articolano su due livelli, quello strategico e quello operativo, che si realizza attraverso i distretti sanitari, i dipartimenti territoriali, il coordinamento sociosanitario.

Ogni ASS friulana è suddivisa in più distretti:

- ASS 1 – Triestina
 - Distretto 1
 - Distretto 2
 - Distretto 3
 - Distretto 4
- ASS 2 – Isontina
 - Distretto Alto Isontino
 - Distretto Basso Isontino
- ASS 3 – Alto Friuli
 - Distretto “Gemonese, Canal del Ferro, Val Canale”
 - Distretto “Carnia”

- ASS 4 – Medio Friuli
 - Distretto di Cividale
 - Distretto di Codroipo
 - Distretto di San Daniele
 - Distretto di Tarcento
 - Distretto di Udine
- ASS 5 – Bassa Friulana
 - Distretto Est
 - Distretto Ovest
- ASS 6 – Friuli Occidentale
 - Distretto Urbano
 - Distretto Ovest
 - Distretto Sud
 - Distretto Nord
 - Distretto Est

Il distretto è l'articolazione organizzativa mediante la quale l'Azienda per i Servizi Sanitari assicura una risposta coordinata e continuativa ai bisogni sanitari dei cittadini residenti nei comuni afferenti.

Il distretto è centro di riferimento dei cittadini per tutti i servizi dell'Azienda per i Servizi Sanitari, sede di integrazione dei servizi sanitari con quelli socio-assistenziali del territorio e può operare in modo coordinato con strutture private di volontariato che offrano servizi sanitari e socio-assistenziali.

1.2 Analisi esplorative dei dati

1.2.1 Il Medico di Medicina Generale

Il Medico di Medicina Generale (MMG) rappresenta una figura centrale dell'assistenza sanitaria: a lui viene affidata la responsabilità complessiva della tutela della salute del proprio assistito, nei cui confronti pratica attività di diagnosi, terapia, riabilitazione, prevenzione individuale ed educazione sanitaria.

La Tabella 1 evidenzia la suddivisione territoriale dei Medici di Medicina Generale nelle varie ASS della Regione. Si può notare come le ASS con la concentrazione maggiore di medici siano la ASS 4 del Medio Friuli e la ASS 6 del Friuli Occidentale. La ASS 4 comprende, infatti, la città di Udine e parte della sua provincia, la ASS 6 racchiude tutta la provincia di Pordenone. Le grandi città hanno, sicuramente, una densità di abitanti più elevata rispetto ai paesi di provincia, perciò le ASS necessitano di un numero maggiore di Medici di Medicina Generale.

AZIENDA	DISTRETTO					Totale
	1	2	3	4	5	
ASS 1 - Triestina	49	69	49	51	0	218
ASS 2 - Isontina	61	46	0	0	0	107
ASS 3 - Alto Friuli	32	39	0	0	0	71
ASS 4 - Medio Friuli	41	33	46	43	131	294
ASS 5 - Bassa Friulana	46	41	0	0	0	87
ASS 6 - Friuli Occidentale	43	31	42	41	73	230
Totale	272	259	137	135	204	1,007

Tabella 1 - Suddivisione MMG nei distretti sanitari e nelle ASS

La Figura 1 rappresenta l'istogramma del genere dei MMG.

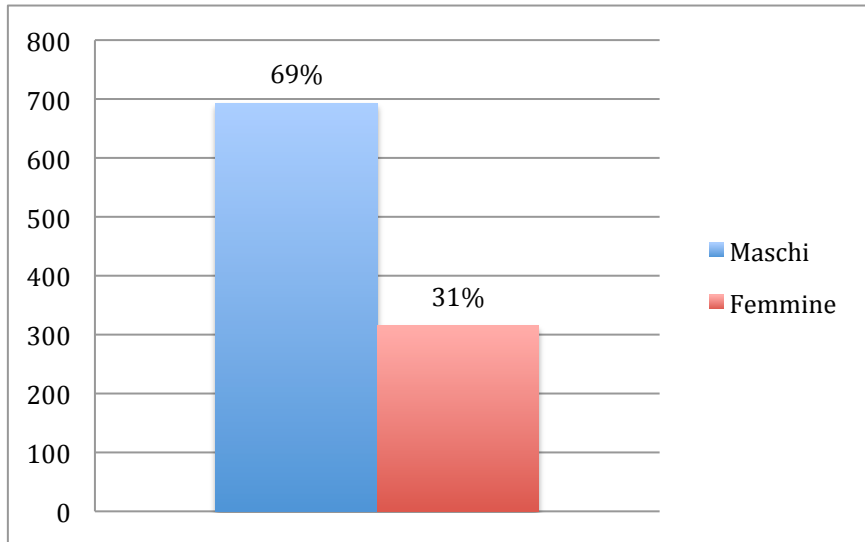


Figura 1 – Suddivisione dei MMG per genere

Si può notare che su 1007 medici di medicina generale, la maggior parte sono di genere maschile (circa il 69%).

Divido l'età dei MMG in classi di dieci anni e in relazione al genere (Figura 2).

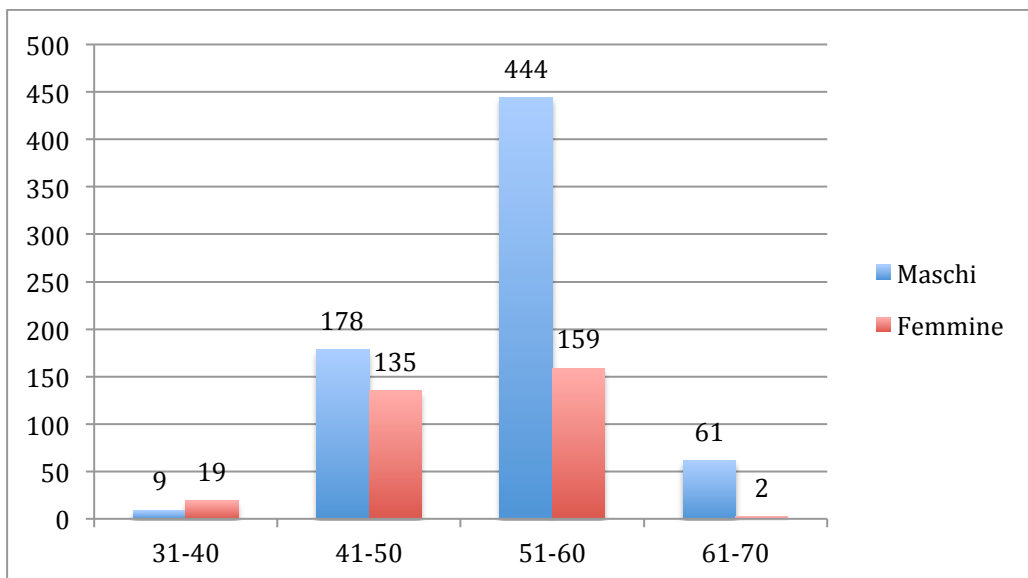


Figura 2 – Suddivisione del campione dei MMG per classi di età e genere

Si noti come la classe più numerosa sia quella dei 51-60 anni. Possiamo definirli medici con più esperienza e la maggior parte di loro sono di genere maschile. La differenza di numerosità

tra medici maschi e medici femmine, infatti, si evidenzia maggiormente nei medici che hanno oltre 50 anni di età. Per quanto riguarda i medici più giovani non si nota una grande differenza tra uomini e donne. Questo denota che in passato le persone che studiavano questa disciplina erano per la maggior parte di sesso maschile, mentre negli ultimi tempi si può osservare un maggior equilibrio di genere.

Si consideri la variabile anni dalla laurea, confrontata con il genere del MMG (Figura 3).

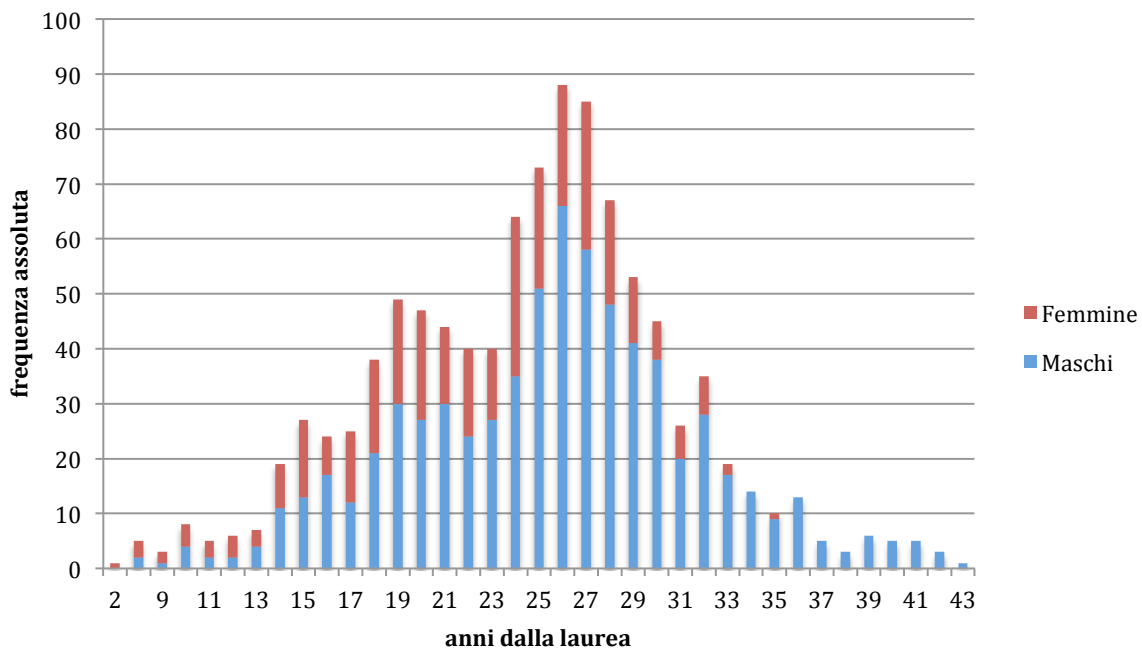


Figura 3 - Anni dal conseguimento della laurea, suddivisione per genere del MMG

L'andamento complessivo sembra essere quello di una gaussiana, gli anni dalla laurea vanno da un minimo di 2 a un massimo di 43. Poche persone hanno conseguito la laurea da più di 32 anni, probabilmente sono persone anziane e che stanno andando in pensione, e sono per la quasi totalità di genere maschile. Altrettante poche persone hanno conseguito il titolo da meno di 15 anni e questo potrebbe derivare da restrizioni nelle politiche di reclutamento. I MMG più frequenti nel territorio del Friuli sono quelli che mediamente hanno conseguito il titolo da 26 anni. Se ci si focalizza sul genere rispetto agli anni dalla laurea, si nota, anche in questo grafico, che molte sono le giovani donne che svolgono questo lavoro rispetto al passato, infatti pochissime dirigono una disciplina da più di 30 anni.

In media, gli assistiti a carico di un medico di medicina generale sono 1090, ma ben il 24% dei MMG supera la soglia di assistiti massimi.

1.2.2 Anagrafica assistiti

Il Friuli-Venezia Giulia conta, nel 2007, 1.198.494 abitanti e 1.129 medici di medicina generale.

L'analisi si focalizzerà solo sugli assistiti seguiti dai Medici di Medicina Generale. Ho deciso, quindi, di togliere dal dataset iniziale i pazienti con età inferiore ai 14 anni, i quali sono solitamente seguiti da pediatri (le cure e le visite legate ai bambini possono essere diverse e più particolari rispetto a quelle degli adulti).

Il campione finale analizzato sarà, quindi, pari a 1.068.732 pazienti.

Come si vede dalla Figura 4, il numero di assistiti di sesso femminile è leggermente maggiore rispetto a quello degli assistiti di sesso maschile.

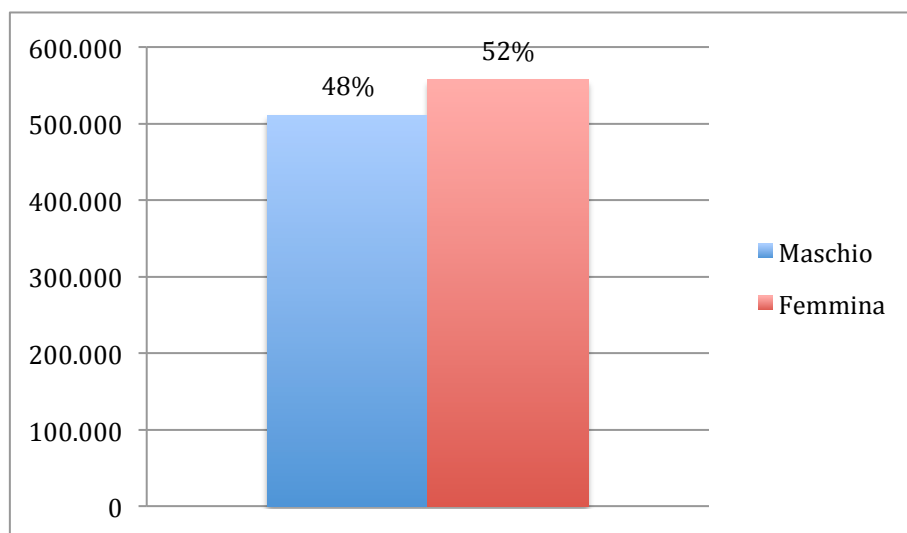


Figura 4 - Distribuzione del campione per genere

La suddivisione in classi di età è evidenziata nella Figura 5.

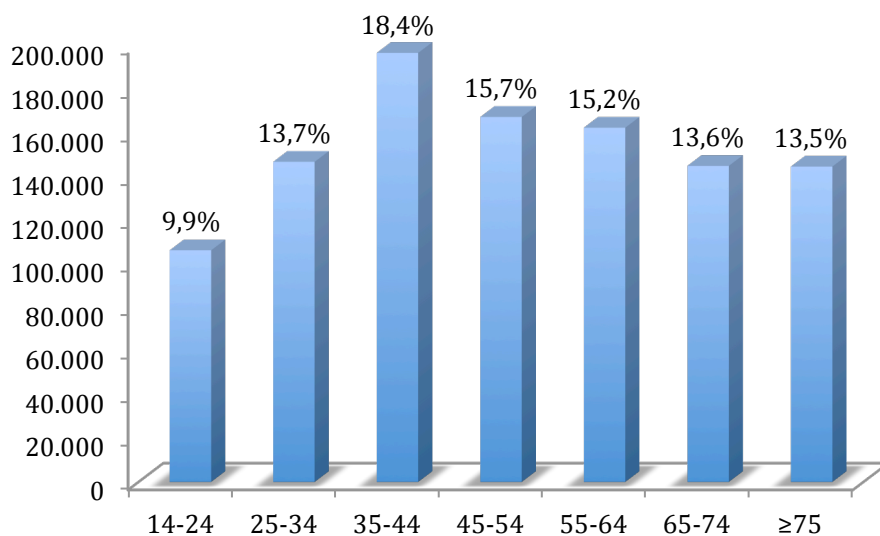


Figura 5 – Suddivisione del campione per classi di età

La classe più numerosa è caratterizzata dagli assistiti con un'età compresa tra i 35 e 44 anni, circa il 18% della popolazione. Ben il 27% circa, però, ha più di 65 anni, fascia che comprende per lo più pensionati.

L'età media e mediana della popolazione è di 45 anni (deviazione standard 23,16). Il range dell'età degli assistiti va da un minimo di 14 anni ad un massimo di 119 anni. L'età media degli abitanti di sesso femminile è di 47 anni (deviazione standard 23,85), mentre gli individui di sesso maschile hanno un'età media di 43 anni (deviazione standard 22,23).

1.2.3 Esenzioni

In Friuli-Venezia Giulia, nel 2007, erano previste le seguenti esenzioni dal pagamento del ticket:

- Esenzione per gli ultra 65enni. L'esenzione per i farmaci è concessa solo ai pensionati con più di 65 anni che hanno come unica fonte di reddito la pensione minima o la pensione/assegno sociale.
- Esenzione per invalidità. Gli invalidi civili, di guerra, per servizio o per lavoro non devono pagare alcuna tariffa per le prestazioni specialistiche, di diagnostica strumentale e di laboratorio, di fisioterapia né la quota fissa per i farmaci.
- Esenzioni per patologia. Gli affetti da patologie specifiche, malattie tumorali o in attesa di trapianto d'organo sono esenti dal pagamento del ticket per le prestazioni specialistiche di diagnostica strumentale e di laboratorio, di fisioterapia inerenti con la patologia. E' dovuta, invece, la quota fissa per farmaci: 1 euro per ogni confezione prescritta fino ad un massimo di tre confezioni per ricetta.
- Esenzione finalizzata alla tutela della gravidanza. Alle donne in stato di gravidanza viene riconosciuta l'esenzione per le prestazioni di diagnostica strumentale e di laboratorio e per le prestazioni specialistiche secondo quanto previsto da uno specifico controllo.

Secondo le disposizioni regionali che vigevano nel 2007, non era più valida l'autocertificazione (tranne per i disoccupati) per quel che riguarda l'esenzione farmaci per età e con limiti di reddito.

I dati amministrativi a disposizione raccolgono le esenzioni per patologia e gravidanza.

Gli assistiti che hanno almeno un'esenzione per patologia e gravidanza sono 309.207 (29% del totale assistiti del Friuli-Venezia Giulia), ma tra questi ce ne sono alcuni che presentano più di un'esenzione. In particolare 107.596 (10%) pazienti hanno avuto 2 esenzioni e 83.422 (7,8%) pazienti hanno avuto un numero di esenzioni che va da un minimo di 3 a un massimo di 31. Le esenzioni totali a carico del SSN sono quindi 500.225.

Nella Tabella 2 sono riportate le più diffuse esenzioni per patologia nel Friuli-Venezia Giulia, per il periodo analizzato. Vengono riportate in ordine crescente di frequenza.

Codice	Descrizione patologia	Numerosità
056	Tiroidite di Hashimoto	8.869
S39	Invalidità di servizio cat. 6-8	11.584
GM0	Gravidanza	11.647
019	Glaucoma	15.363
025	Colesterolo	18.978
S57	Invalidità civile al 100%	19.021
002	Affezioni sistema circolatorio	20.277
S52	Invalidità civile (assegno accomp.)	21.333
002A	Malattie cardiache e del circolo polmonare	21.714
S51	Invalidità civile >2/3	23.224
031A	Ipertensione arteriosa senza danno d'organo	45.479
013	Diabete	51.033
048	Patologie neoplastiche maligne	59.367
031	Ipertensione arteriosa con danno d'organo	76.212

Tabella 2 - Descrizione patologia esenzioni

Un'analisi esplorativa di alcune esenzioni per patologia mostra, per esempio, come l'età media delle donne esenti per gravidanza è di 33 anni (deviazione standard 5,24).

L'esenzione che viene concessa più frequentemente è quella per l'ipertensione arteriosa con danno d'organo. L'età media delle persone che soffre di questa patologia è di 69 anni (deviazione standard 11,64) e il range va dai 14 ai 104 anni perciò copre tutte le varie fasce d'età; la mediana si aggira sui 70 anni.

L'età media degli esenti per patologie neoplastiche maligne è di 65 anni (deviazione standard 13,6) e il range d'età varia dai 14 ai 104 anni. Il 55% delle esenzioni sono concesse alle donne.

L'età media degli assistiti che soffre di diabete è di 67 anni (deviazione standard 13,2) e il 55% è maschio.

Il 62% delle esenzioni per malattie cardiache e del circolo polmonare sono concesse agli assistiti di sesso maschile e l'età media è di 67 anni (deviazione standard 13,2).

L'indennità di accompagnamento è stata istituita nel 1980 e spetta agli invalidi civili totali che, per malattie fisiche o psichiche, si trovino nell'impossibilità di deambulare senza l'aiuto permanente di un accompagnatore o, non essendo in grado di compiere gli atti quotidiani della vita, hanno bisogno di un'assistenza continua. L'indennità di accompagnamento, a

differenza di altri benefici economici concessi agli invalidi, è indipendente dall'età e dalle condizioni reddituali della persona. Nel 2007, nel Friuli-Venezia Giulia gli invalidi civili con assegno di accompagnamento erano per il 70% donne con un'età media di 78 anni (deviazione standard 17,09). La mediana è di 83 anni e questo denota una distribuzione con asimmetria negativa la cui forma è caratterizzata da una coda allungata verso sinistra.

Gli invalidi civili al 100% sono anch'essi per lo più donne (circa il 64%) e l'età media è di 72 anni (deviazione standard 15,6). Chi usufruisce di tale esenzione ha diritto a: fornitura gratuita di ausili e protesi previsti dal nomenclatore nazionale; collocamento obbligatorio se presente capacità lavorativa residua; esenzione dalla partecipazione alla spesa sanitaria (esclusa la quota fissa); tessera di libera circolazione gratuita. Inoltre viene concessa la pensione di inabilità per le persone di età compresa tra 18 e 65 anni, nel rispetto dei limiti reddituali.

Gli assistiti con un'invalidità maggiore del 75% hanno un'età media di 65 anni (deviazione standard 15,3) e sono per il 58% donne. A questo tipo di invalidità viene riconosciuto un assegno mensile concesso solo alle persone di età compresa tra i 18 e i 65 anni, prive di impiego, nel rispetto dei limiti di reddito per usufruirne. E' incompatibile con altri redditi pensionistici. Per chi supera i 65 anni di età è previsto l'assegno sociale dell'INPS.

Gli esenti per invalidità di servizio sono per il 96% uomini con un'età media di 66 anni (deviazione standard 10,5).

L'esenzione per chi soffre di tiroidite di Hashimoto è concessa per il 92% alle donne con un'età media di 48 anni (deviazione standard 13,3).

1.2.4 Servizio ADI

L'assistenza domiciliare è stata definita dall' Organizzazione Mondiale della Sanità³ come “la possibilità di fornire presso il domicilio del paziente quei servizi e quegli strumenti che contribuiscono al mantenimento del massimo livello di benessere, salute e funzione”.

L'assistenza domiciliare integrata (ADI) è un sistema di interventi e servizi sanitari offerti a domicilio, intendendo per domicilio sia l'abitazione del paziente, sia una struttura comunitaria, casa di riposo o struttura residenziale permanente. Le forme di Assistenza Domiciliare Integrata sono due: semplice e complessa. La prima include prestazioni infermieristiche o riabilitative più semplici, come medicazioni, prelievi del sangue o cambi di catetere, ed è rivolta a persone non totalmente autosufficienti, in genere anziane. L'assistenza Domiciliare Integrata complessa comprende un insieme di cure mediche, infermieristiche, riabilitative e assistenziali, che riguardano persone gravemente ammalate non autosufficienti, con necessità complesse. Gli interventi possono essere di diverse intensità: a bassa intensità sanitaria (ADI 1), con interventi di assistenza di base presumibilmente per un lungo periodo di presa in carico; a media o alta intensità (ADI 2 o ADI 3), rivolta a cittadini affetti da gravi patologie.

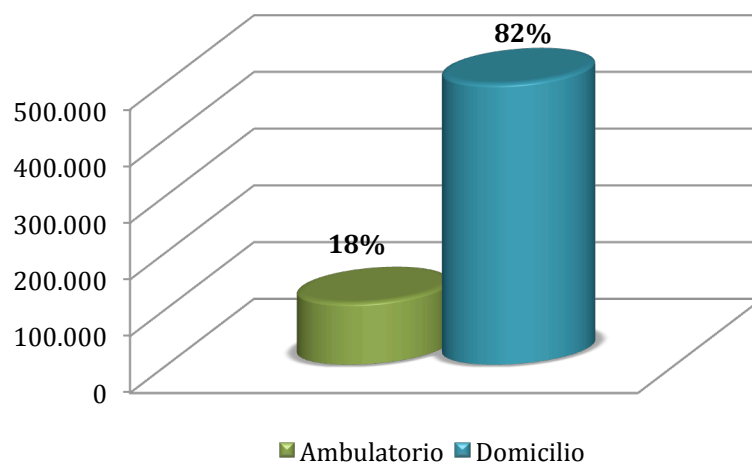


Figura 6 - Luogo servizio ADI

³ <http://www.who.int>

Nel 2007, 48.989 assistiti hanno usufruito del servizio di assistenza domiciliare integrata almeno una volta, ma la media per persona è di 27 volte (deviazione standard 38,2). Il range di visite va da 1 a 524 e la mediana è di 14 volte. Il 25% degli assistiti ha richiesto più di 34 visite a domicilio. In totale le visite effettuate a carico del SSN sono state 605.141 per la regione del Friuli-Venezia Giulia nell'anno 2007. L'età media degli assistiti che hanno usufruito di tale servizio è di 76 anni (deviazione standard 14,3) e la mediana è di 80 anni. La Figura 6 evidenzia che più dell'80% del servizio ADI è stato svolto nel domicilio del paziente.

1.2.5 Ricoveri ospedalieri

Il numero di persone che sono state ricoverate nel 2007 in Friuli-Venezia Giulia sono 134.629. Gli assistiti ricoverati due volte nello stesso anno sono 34.664 e 20.842 persone sono state ricoverate più di due volte. In totale, i ricoveri ospedalieri a carico del SSN sono stati 199.026. Il range dei ricoveri va da 1 a 24 e solo il 5% degli assistiti è stato ricoverato più di 3 volte.

Il ricovero ospedaliero può essere:

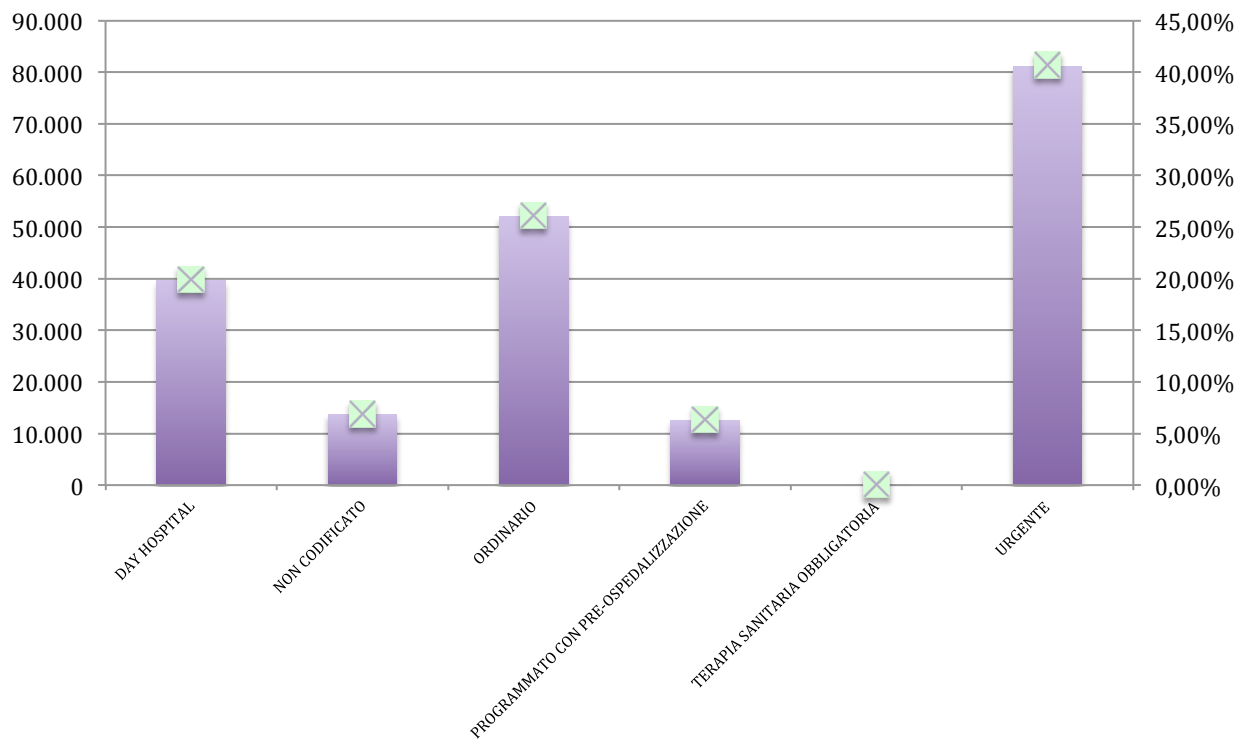
- Programmato o ordinario quando viene disposto dal medico specialista successivamente a visita in cui ha rilevato la necessità di attuare un piano diagnostico terapeutico di regime di degenza;
- Urgente se il paziente giunge attraverso il Pronto Soccorso, in questo caso il ricovero viene deciso dal Medico del Dipartimento di Emergenza e Accettazione.

Talvolta il ricovero programmato è preceduto da una pre-ospedalizzazione che consiste nella fase che precede il ricovero per intervento chirurgico programmato e serve ad effettuare indagini, visite ed esami necessari all'intervento.

Oltre al ricovero ordinario sono possibili ricoveri in regime di:

- Day Hospital, il cosiddetto "ospedale di giorno": consiste in uno o più ricoveri di una giornata ciascuno con rientro serale a domicilio, in relazione al piano diagnostico terapeutico più adatto ai problemi di salute della persona.
- Day Surgery: è un Day Hospital ma in questo caso il piano terapeutico è di tipo chirurgico e si attua nell'arco di una sola giornata (con eventuale pernottamento) tra il lunedì e il venerdì.

La Figura 7 rappresenta i diversi tipi di ricoveri ospedalieri avvenuti in Friuli Venezia Giulia nel 2007.



Nota: Asse y principale: istogramma valori assoluti, Asse y secondaria: indicatore valori percentuali

Figura 7 - Tipologia di ricoveri ospedalieri

Il 41% dei ricoveri ospedalieri è di tipo urgente, seguiti poi dai ricoveri ordinari e dai day hospital. La media di degenza dei ricoveri è di 7 giorni.

Ogni assistito, dopo essere stato dimesso dall'ospedale, viene caratterizzato da un codice DRG. Si tratta di un sistema di classificazione dei pazienti che viene utilizzato come base per il finanziamento delle Aziende Ospedaliere. I Diagnosis-related groups (DRG) sono l'equivalente in italiano dei Raggruppamenti Omogenei di Diagnosi ossia ROD⁴. Il sistema DRG si può definire come un sistema isorisorso, in quanto orientato a descrivere la complessità dell'assistenza prestata al paziente, partendo dal principio che malattie simili, trattate in reparti ospedalieri simili, comportano approssimativamente lo stesso consumo di risorse umane e materiali. Tale sistema è stato quindi creato per poter predire la quantità ed il

⁴ <http://www.salute.gov.it>

tipo di risorse utilizzate per assistere i pazienti. I DRG consentono di mettere in relazione tra loro i seguenti elementi: input, cioè le risorse impiegate (uomini, materiali, attrezzature), output, cioè i servizi diagnostici, terapeutici ed alberghieri prestati ai pazienti, outcome, cioè il risultato conseguito dai pazienti in termini di miglioramento del loro stato di salute e qualità, cioè la validità tecnico-scientifica dei procedimenti utilizzati.

La Tabella 3 contiene i codici DRG, introdotti con Decreto Ministeriale il 21 novembre 2005 (classificazione versione 19)⁵, relativi agli interventi più frequenti in Friuli-Venezia Giulia nel 2007. Le varie tipologie di ricoveri ospedalieri sono riportate in ordine crescente di frequenza. Inoltre, è stato calcolato il costo totale del singolo codice DRG, ottenuto come somma cumulata della spesa per ricoveri di tutti i pazienti che sono caratterizzati dallo stesso codice DRG.

DRG	FREQ.	DESCRIZIONE	COSTO TOTALE
88	2028	Malattia polmonare cronica ostruttiva	7.495.888
39	2060	Interventi sul cristallino con o senza vitrectomia	3.702.278
89	2112	Polmonite semplice e pleurite, età > 17 anni con CC (con patologie complicanti)	10.900.000
231	2207	Escissione locale e rimozione di mezzi di fissaggio intramidollare eccetto anca e femore	4.746.495
229	2285	Interventi su mano o polso eccetto interventi maggiori sulle articolazioni, senza cc	4.387.368
364	2456	Dilatazione e raschiamento, conizzazione eccetto per neoplasie maligne	4.609.464
119	2499	Legatura e stripping di vene	5.533.712
14	2738	Malattie cerebrovascolari specifiche eccetto attacco ischemico transitorio	14.500.000
410	2830	Chemioterapia non associata a diagnosi secondaria di leucemia acuta	13.800.000
162	2973	Interventi per ernia inguinale e femorale, età > 17 anni senza CC	6.945.405
381	3284	Aborto con dilatazione e raschiamento, mediante aspirazione o isterotomia	5.814.522
359	3654	Interventi su utero e annessi non per neoplasie maligne, senza cc	11.400.000
209	4514	Interventi su articolazioni maggiori e reimpianti di arti inferiori	39.200.000
503	5430	Interventi sul ginocchio senza diagnosi principale di infezione	14.000.000
127	6113	Insufficienza cardiaca e shock	24.800.000
391	6811	Neonato normale	4.485.392
373	6991	Parto vaginale senza diagnosi complicanti	13.400.000

Tabella 3 - Classificazione codici DRG

L'intervento più frequente dei ricoveri ospedalieri è il parto vaginale senza diagnosi complicanti (6991 casi). Se, invece, ci focalizziamo sui ricoveri per patologia, possiamo definire il Friuli-Venezia Giulia una regione che ha molti casi di insufficienza cardiaca e shock (6113 interventi nel 2007). Tuttavia, il costo totale più oneroso a carico del SSN è quello che riguarda le articolazioni maggiori e reimpianti di arti inferiori. In media, il costo per un singolo intervento su articolazioni maggiori e reimpianti di arti inferiori è pari a 8.864 €. Un

⁵ Rapporto annuale sull'attività di ricovero ospedaliero - Anno 2006

altro degli interventi più costosi a livello di singola unità è la polmonite semplice (codice DRG 89), che costa 5.161 € per ogni caso.

1.2.6 Farmaci

I medici di medicina generale prescrivono ai propri pazienti i farmaci necessari alla cura; la legge prevede una prima distinzione tra farmaci “contemplati” e “non contemplati”; questi ultimi non rientrano tra quelli erogati a carico del servizio sanitario pubblico e devono essere indicati sul cosiddetto “ricettario bianco”. I farmaci contemplati, invece, si suddividono in “non prescrivibili” (fascia C) e “prescrivibili” (fascia A). I medicinali di fascia C sono quelli utilizzati per patologie di lieve entità e comprendono anche i medicinali di automedicazione, non soggetti a ricetta medica. Questi medicinali sono interamente a carico del paziente. I medicinali di fascia A, invece, sono essenziali per la cura di patologie gravi, croniche e acute, e sono a carico del SSN.

Gli assistiti che nel 2007 hanno acquistato almeno un farmaco di fascia A sono 764.787 ma il range dei farmaci acquistati va da 1 a 261 con una media per persona di 18 farmaci (deviazione standard 18,72).

I farmaci sono identificati dal codice AIC, un codice numerico a nove cifre che indentifica univocamente ogni formato di ogni farmaco, quindi non solo ogni farmaco ha un codice univoco ma formati diversi dello stesso farmaco hanno codici AIC diversi.

Un altro tipo di classificazione dei farmaci è il Sistema di classificazione Anatomico Terapeutico e Chimico (ATC), usato per la classificazione sistematica dei farmaci ed è controllato dall'Organizzazione Mondiale della Sanità. In questo sistema di classificazione i farmaci sono suddivisi in gruppi sulla base degli organi o apparati su cui agiscono e delle loro proprietà chimiche, farmacologiche e terapeutiche. La classificazione è di tipo alfanumerico e suddivide i farmaci in base ad uno schema costituito da 5 livelli gerarchici.

Il primo livello contiene il Gruppo Anatomico principale (contraddistinto da una lettera dell'alfabeto). In totale ve ne sono 14.

Il secondo livello contiene il Gruppo Terapeutico principale (contraddistinto da un numero di due cifre).

Il terzo livello contiene il Sottogruppo Terapeutico Farmacologico (contraddistinto da una lettera dell'alfabeto).

Il quarto livello contiene il Sottogruppo Chimico-Terapeutico Farmacologico (contraddistinto da una lettera dell'alfabeto).

Il quinto livello contiene il Sottogruppo Chimico (contraddistinto da un numero di due cifre) ed è specifico per ogni singola sostanza chimica.

I prodotti farmaceutici sono classificati sulla base dell'uso terapeutico (indicazione) principale, seguendo il principio che tutti i prodotti con formulazione analoga, cioè paragonabili per ingredienti, dose unitaria e via di somministrazione, possono avere un solo codice ATC. Tuttavia, uno stesso principio attivo può ricevere più di un codice ATC nel caso siano presenti prodotti chiaramente diversi per indicazione terapeutica, dosaggio e via di somministrazione.

La Tabella 4 contiene la classificazione ATC di primo livello dei farmaci concessi in Friuli-Venezia Giulia nel 2007 (frequenza, importo totale lordo, costo unitario).

ATC	Totale confezioni vendute	Primo livello - Gruppo anatomico principale	Costo totale	Costo unitario
A	2.820.342	Tratto alimentare e metabolismo	35.100.000	12,45
B	1.492.872	Sangue e organi eritropoietici	9.616.719	6,44
C	7.901.192	Apparato cardiovascolare	98.000.000	12,40
D	84.169	Farmaci dermatologici	1.048.422	12,46
G	748.034	Apparato genito-urinario e ormoni sessuali	9.509.519	12,71
H	512.965	Preparazioni ormonali sistematiche, esclusi ormoni sessuali e insulina	2.978.123	5,81
J	1.157.583	Anti-infettivi per uso sistemico	16.900.000	14,60
L	150.338	Antineoplastici e immunomodulatori	11.400.000	75,83
M	836.113	Apparato muscolo-scheletrico	8.342.517	9,98
N	1.529.033	Sistema nervoso	22.900.000	14,98
P	50.473	Prodotti antiparassitari, insetticidi e repellenti	314.588	6,23
R	699.527	Apparato respiratorio	16.000.000	22,87
S	334.379	Organi sensori	4.231.869	12,66
V	70.900.000	Vari	840.166	0,01
	89.217.020	Totali	237.181.923	

Tabella 4 – Classificazione I° livello ATC

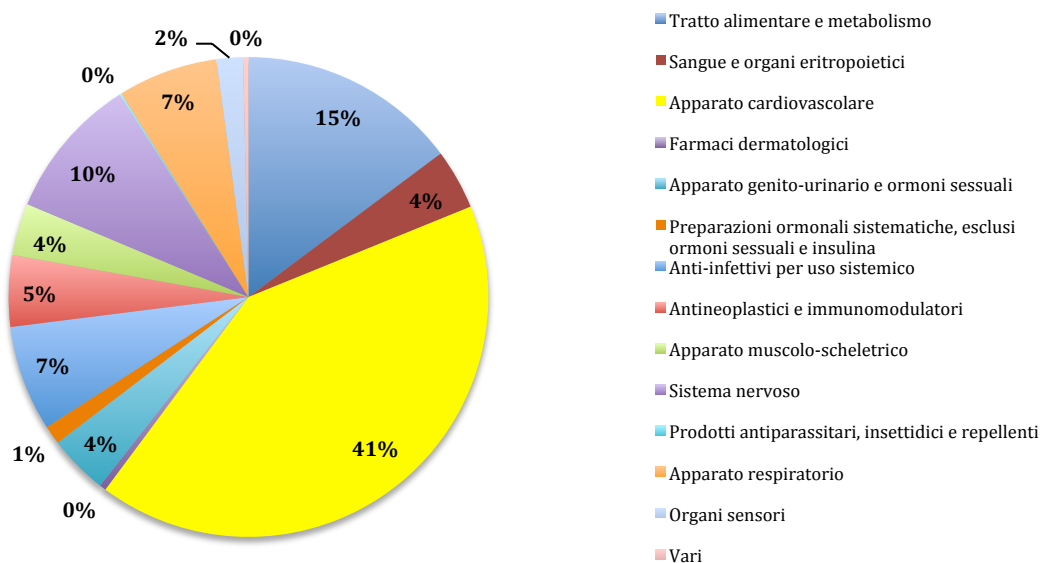


Figura 8 - Composizione percentuale del costo totale dei farmaci classificazione ATC 1° livello

I farmaci che pesano di più sul bilancio del Sistema Sanitario del Friuli-Venezia Giulia nel 2007 sono quelli che riguardano l'apparato cardiovascolare (Figura 8). Circa il 41% del costo totale dei farmaci è attribuibile a questa voce. Il costo unitario del singolo farmaco non è tra i più elevati (12,40 euro) ma le confezioni vendute sono molte, 7.900.000 circa. Il costo unitario più ingente, invece, è attribuibile agli antineoplastici e immunomodulatori (75,83 euro per farmaco), anche se il costo totale pesa solo per il 5% sul bilancio del Sistema Sanitario. Il costo per farmaco per problemi di apparato respiratorio è di 22,87 euro ma nel complesso questa spesa pesa per il 7% sul costo totale dei farmaci. La seconda spesa che grava di più sul costo totale è quella attribuibile ai farmaci del tratto alimentare e metabolismo, che copre circa il 15% del costo totale dei farmaci, con 2.800.000 confezioni vendute e un costo unitario di 12,45 euro.

Il sistema di classificazione ATC è usato negli studi di farmacoutilizzazione in concomitanza a un indice, il cui acronimo è DDD (Defined Daily Dose, dose definita giornaliera). Questo indice nasce perché il principale problema delle analisi in campo farmaceutico è l'adozione di una metodologia che consenta di produrre stime affidabili dell'esposizione della popolazione ai farmaci, di porre a confronto i consumi di farmaci o classi di farmaci diversi e, infine, di confrontare l'uso di farmaci in aree diverse. La metodologia adottata a livello internazionale è denominata, per l'appunto, ATC/DDD.

Il modo apparentemente più semplice per misurare i volumi di prescrizione dei farmaci è di contare il numero di confezioni prescritte e, eventualmente, di raggrupparle per principio attivo o per gruppo terapeutico. Tuttavia, questo metodo ha grossi limiti che ne sconsigliano l'uso. Per uno stesso principio attivo, le varie preparazioni in commercio possono differire per il numero di unità posologiche contenute: confezioni dello stesso farmaco con numero diverso di compresse saranno conteggiate sempre con un pezzo, sebbene il numero di dosi unitarie fornite al paziente sia diverso. Anche nel confronto di molecole diverse all'interno di un gruppo terapeutico, il conteggio dei pezzi può offrire una misura distorta dell'esposizione ai farmaci. Per ovviare a questi inconvenienti si usa misurare i consumi di farmaci proprio attraverso la DDD, che è diventata l'unità di misura standard internazionale della prescrizione farmaceutica. La DDD è definita come la dose media di un farmaco assunta giornalmente da un paziente adulto, con riferimento all'indicazione terapeutica principale del farmaco stesso. La DDD, quindi, rappresenta la dose di mantenimento della terapia e non la dose iniziale, né la dose raccomandata, né prescritta, e va considerata unicamente come strumento tecnico per misurare la prescrizione dei farmaci. I dati vengono espressi in numero di DDD per 1000 abitanti al giorno (DDD/1000abitanti/die): ciò consente di confrontare volumi di prescrizioni relativi a popolazioni diverse ma anche di ottenere una stima approssimativa dell'esposizione ad un determinato farmaco o gruppi di farmaci in una popolazione⁶.

Tipo ATC	Primo livello - Gruppo anatomico principale	Spesa lorda pro capite	%	DDD/1000 ab die	%
A	Tratto alimentare e metabolismo	32,93	14,8%	139,76	7,6%
B	Sangue e organi eritropoietici	9,02	4,1%	135,02	7,3%
C	Apparato cardiovascolare	91,94	41,3%	130,55	7,1%
D	Farmaci dermatologici	0,98	0,4%	112,18	6,1%
G	Apparato genito-urinario e ormoni sessuali	8,92	4,0%	123,12	6,7%
H	Preparazioni ormonali sistematiche, esclusi ormoni sessuali e insulina	2,79	1,3%	148,95	8,1%
J	Anti-infettivi per uso sistemico	15,85	7,1%	133,69	7,3%
L	Antineoplastici e immunomodulatori	10,69	4,8%	123,92	6,7%
M	Apparato muscolo-scheletrico	7,83	3,5%	123,99	6,7%
N	Sistema nervoso	21,48	9,7%	136,02	7,4%
P	Prodotti antiparassitari, insetticidi e repellenti	0,30	0,1%	179,83	9,8%
R	Apparato respiratorio	15,01	6,7%	147,21	8,0%
S	Organi sensori	3,97	1,8%	129,86	7,1%
V	Vari	0,79	0,4%	74,09	4,0%
		222,51	100 %	1838,19	100 %

Tabella 5 - Consumo territoriale classe A-SSN 2007 (popolazione pesata) per il I° livello ATC

⁶ Bollettino di informazione sui farmaci, bimestrale del Ministero della Salute, 2002

Dalla Tabella 5 si nota che la spesa lorda pro capite maggiore è attribuibile ai farmaci per l'apparato cardiovascolare (41,3%). Il rispettivo DDD/1000 abitanti è di 130,55 e costituisce il 7% del totale dei DDD. Questo vuol dire che su 1000 abitanti sono state prescritte 130 DDD di quel farmaco al giorno, cioè il 13% degli assistiti ha ricevuto in media ogni giorno una DDD di quel farmaco. Il DDD più elevato è attribuibile ai prodotti antiparassitari, insetticidi e repellenti (9,8% del totale dei DDD): 179 persone su 1000 hanno ricevuto ogni giorno una DDD di quel farmaco. Altri DDD/1000 abitanti die elevati sono per esempio le preparazioni ormonali sistematiche esclusi ormoni sessuali e insulina (8,1 % del totale). 148 persone su 1000 hanno ricevuto ogni giorno una DDD di quel farmaco. Anche i farmaci per l'apparato respiratorio sono molto venduti tra gli assistiti, circa 147 persone su 1000 hanno ricevuto ogni giorno una DDD del farmaco.

1.2.7 Visite Specialistiche

Nei Centri Unici di Prenotazione (CUP) è possibile prenotare, muniti della relativa impegnativa/richiesta del medico, visite e prestazioni specialistiche erogate dalle Aziende per i Servizi Sanitari.

Per quanto riguarda il Friuli-Venezia Giulia, 740.227 persone hanno effettuato almeno una visita specialistica nel 2007 e 656.626 sono gli assistiti che hanno avuto due visite specialistiche. Il 25% dei pazienti ha effettuato più di 27 visite specialistiche. La media di visite per assistito è di 22, il range va da 1 a 805. In totale, il numero di visite a carico del SSN sono state circa 13.300.000.

La Tabella 6 riporta i codici prestazioni la cui frequenza supera le 100.000 visite specialistiche, assieme al costo totale e il costo unitario.

CODICE PRESTAZIONE	Descrizione	Frequenza	%*	Costo totale	Costo unitario
89.01	Anamnesi e valutazione, definite brevi	461.257	3,46	7.380.112	16,00
89.52	Elettrocardiogramma	137.286	1,03	1.990.647	14,50
89.7	Visita generale	450.467	3,38	11.500.000	25,53
90.04.5	Alanina aminotransferasi (ALT) (GPT) [S/U]	480.155	3,60	864.279	1,80
90.09.2	Aspartato aminotransferasi (AST) (GOT) [S]	453.128	3,40	815.630	1,80
90.10.5	Bilirubina totale e frazionata	127.337	0,96	305.608	2,40
90.14.3	Colesterolo totale	402.326	3,02	683.954	1,70
90.16.3	Creatinina [S/U/dU/La]	503.155	3,78	955.994	1,90
90.22.5	Ferro [S]	157.605	1,18	378.252	2,40
90.23.5	Fosfatasi alcalina	158.972	1,19	333.841	2,10
90.25.5	Gamma glutamil transpeptidasi (gamma GT) [S/U]	349.973	2,63	664.948	1,90
90.27.1	Glucosio [S/P/U/dU/La]	508.291	3,82	609.949	1,20
90.28.1	Hb-Emoglobina glicata	107.638	0,81	1.410.058	13,10
90.37.4	Potassio [S/U/dU/(Sq)ER]	252.482	1,90	479.715	1,90
90.38.4	Proteine (elettroforesi delle) [S]	138.755	1,04	943.534	6,80
90.40.4	Sodio [S/U/dU/(Sg)Er]	220.261	1,65	418.495	1,90
90.42.1	Tireotropina (TSH)	172.797	1,30	2.004.445	11,60
90.42.3	Tiroxina libera (FT4)	139.506	1,05	1.855.430	13,30
90.43.2	Trigliceridi	384.397	2,89	730.354	1,90
90.43.3	Triiodotironina libera (FT3)	110.316	0,83	1.467.203	13,30
90.43.5	Urato [S/U/dU]	228.509	1,72	342.763	1,50
90.44.1	Urea [S/P/U/dU]	173.519	1,30	329.686	1,90
90.44.3	Urine esame chimico fisico e microscopico	455.565	3,42	1.321.139	2,90
90.56.5	Antigene prostatico specifico (PSA)	125.222	0,94	1.953.463	15,60
90.62.2	Emocromo: Hb, GR, GB, HCT, PLT, IND. DERIV., F.L.	671.542	5,04	3.424.864	5,10
90.72.3	Proteina C reattiva (Quantitativa)	123.481	0,93	1.197.766	9,70
90.75.4	Tempo di protrombina (PT)	153.672	1,15	507.117	3,30
90.82.5	Velocità di sedimentazione delle emazie (VES)	213.090	1,60	255.708	1,20
91.49.2	Prelievo di sangue venoso	1.037.725	7,79	3.424.493	3,30
	[...]				
	TOTALE	13.323.367			

Tabella 6 - Codici prestazioni visite specialistiche

*rispetto al totale delle visite specialistiche

Si può notare che la visita specialistica più frequente è il prelievo di sangue venoso (codice 91.49.2), 1.037.725 visite in un anno, che corrisponde al 7,8% del totale delle visite specialistiche, seguita poi dall'emocromo (codice 90.62.2), 671.542 visite ovvero il 5,04% del totale delle visite specialistiche. La terza visita più frequente riguarda il glucosio (codice 90.27.1), 508.291 visite specialistiche cioè il 3,82% del totale. Nonostante queste siano le visite più frequenti, in realtà non sono le più costose. Il costo totale più oneroso è attribuibile

alla visita generale, 11.500.000 € e una frequenza di 450.467 visite, che porta ad avere un costo unitario per visita di 25,53 euro, il più costoso appunto. Un altro costo unitario elevato è attribuibile all'anamnesi e valutazione, definite brevi, circa 16 euro per visita, seguita poi dall'antigene prostatico specifico, 15,60 € e dall'elettrocardiogramma che costa 14,50 € per esame.

1.3 Variabili utilizzate nei modelli

Sulla base delle informazioni contenute nei dataset amministrativi a disposizione, seleziono le variabili a mio parere più interessanti per lo svolgimento di questo elaborato. Di seguito sono descritte le variabili risposta e le variabili esplicative utilizzate.

1.3.1 Variabili risposta

Dall'analisi preliminare dei dati si può notare come le uniche variabili che mi permettono di valutare il costo a carico della Sanità siano quelle relative ai ricoveri ospedalieri, alle visite specialistiche e ai farmaci.

Decido di studiare queste variabili separatamente poiché, se le analizzassi a livello aggregato, non potrei cogliere le influenze che certe variabili esplicative hanno su una variabile risposta piuttosto che un'altra. Certamente si parla sempre di costi, ma in primo luogo la numerosità delle variabili risposta è diversa, dato che, ad esempio, una persona può aver fatto un ricovero ospedaliero e non aver eseguito nessuna visita specialistica nel 2007. A livello aggregato questa particolarità non si noterebbe. In secondo luogo, i tre tipi di spesa pesano in modo diverso sulla Sanità, studiarle a livello aggregato non permetterebbe di capire quale sia quella che influenza di più la spesa pubblica. Inoltre, è ipotizzabile che le variabili esplicative non siano tutte ugualmente significative o non significative per tutti e tre i tipi di costi. Le analisi svolte dimostreranno che, pur parlando sempre di costi, ci sono delle differenze a livello individuale e di singola variabile esplicativa.

Nello specifico, le variabili risposta utilizzate sono le seguenti:

- *tariffa ricovero* : intesa come somma cumulata per persona, delle tariffe attribuibili a tutti i ricoveri avvenuti nel 2007 per quella singola persona;
- *tariffa lorda visite specialistiche*: somma cumulata per persona, di tutte le visite specialistiche effettuate nel 2007 dal singolo. Ogni visita è identificata da un codice prestazione, che fa riferimento al tipo della visita effettuata e all'importo della visita stessa;
- *importo totale lordo farmaci*: somma cumulata dell'importo di tutte le confezioni di farmaci di fascia A acquistati da una persona nell'arco dell'anno 2007.

Servizi	N° assistiti che hanno effettuato almeno una spesa positiva	N° totale prestazioni	Costo totale a carico del SSN	Assistiti che hanno effettuato una spesa positiva		Totale assistiti	
				Spesa media (€)	Dev. St.	Spesa media (€)	Dev. St.
Ricoveri	134.629	199.026	739.000.000	5.491,61	7651,83	616,61	3095,83
Specialistiche	740.227	13.323.367	191.000.000	257,83	598,72	159,24	486,93
Farmaci	764.787	89.217.020	237.000.000	310,25	538,87	197,98	455,55
Popolazione totale	1.198.494						

Tabella 7 - Contenuto di alcuni database amministrativi

La Tabella 7 evidenzia alcune analisi descrittive delle variabili dipendenti in analisi. Si può notare che il numero di persone che sono state ricoverate almeno una volta nel 2007 in Friuli-Venezia Giulia sono 134.629, circa l' 11% del totale della popolazione. Molte di più, invece, sono le persone che hanno effettuato una visita specialistica (740.227 assistiti, 62% della popolazione) o che hanno acquistato almeno un farmaco (764.787 assistiti, circa il 64% della popolazione). Nonostante le persone ricoverate siano state una percentuale piccola del totale della popolazione, la voce ricoveri ospedalieri, in realtà, è la più onerosa per il Sistema Sanitario Nazionale. Il costo totale dei ricoveri ammonta a 739 milioni di euro, nel 2007, per la sola Regione del Friuli-Venezia Giulia. Se ci focalizziamo sulle sole persone ricoverate, la spesa media per ogni assistito è di 5.491 € con una deviazione standard di 7.651. Il costo di degenza è molto elevato e la variabilità tra gli assistiti è altrettanto elevata. Ci possono essere ricoveri molti costosi e ricoveri meno costosi, dipenderà sicuramente dai giorni di degenza, dal tipo di ricovero e dal tipo di cura adottata. Se, invece, ci focalizziamo sul totale della

popolazione (che tiene conto anche delle persone non ricoverate), la spesa per ricoveri scende di molto ed ammonta a 616 € (deviazione standard 3.095).

I farmaci venduti nel 2007 in Friuli-Venezia Giulia sono circa 89 milioni, con un costo a carico del SSN di 237 milioni di euro. La spesa media per farmaci di una persona che vive in questa regione ammonta a 197 € (dev. stand. 455).

Le visite specialistiche compiute nel 2007 in Friuli-Venezia Giulia sono 13.323.367, con un costo a carico del SSN di 191 milioni di euro. In media, ogni persona residente in FVG costa alla Sanità pubblica circa 160 euro di visite specialistiche (dev. stand. 486).

1.3.2 Variabili esplicative

In questo paragrafo vengono presentate le variabili esplicative utilizzate nei modelli che verranno approfonditi in seguito. Le variabili si distinguono principalmente in: caratteristiche degli assistiti e caratteristiche dei Medici di Medicina Generale.

Dapprima ho inserito alcune caratteristiche degli assistiti, come il sesso e l'età. Per quanto riguarda l'età, ho deciso di creare delle dummy in modo tale da non lasciare la variabile continua come in origine. Ho diviso l'età dai 14 ai 75+ anni in dummy di dieci anni ciascuna. Penso sia molto importante segmentare i pazienti in modo da individuare le categorie più o meno onerose per la Sanità pubblica regionale. Si può pensare che al crescere dell'età una persona abbia maggiori problemi di salute rispetto a un ragazzo giovane. C'è anche da notare, però, che le persone anziane che rimangono in vita sono tendenzialmente quelle in buona salute o comunque senza malattie importanti che gravano sul sistema sanitario.

Altre variabili che ho inserito nei modelli sono il decesso o meno di una persona nell'arco del 2007 e nell'arco del 2008 (l'estrazione del dataset amministrativo è avvenuta diversi anni dopo il 2007, per cui sono disponibili informazioni anche sui decessi del 2008). Si può ipotizzare che una persona alla fine del suo ciclo di vita abbia un carico maggiore di spese per la sanità. Ho deciso, inoltre, di inserire il servizio di assistenza domiciliare integrata, creando due dummies, una se il servizio di assistenza è stato effettuato nel domicilio dell'assistito e l'altra se in ambulatorio. Infine ho creato una serie di dummies che si riferiscono alle esenzioni che

sono più diffuse in Friuli Venezia Giulia. Ho voluto inserire le variabili che riguardano le esenzioni per patologia perché possono essere utilizzate come proxy della salute individuale.

Le variabili finora descritte sono usate come base per tutti i modelli. In altri modelli, invece, ho inserito anche le caratteristiche del Medico di Medicina Generale, come il sesso e l'età. Un'altra variabile utile all'analisi è la variabile continua "anni dalla laurea", usata come proxy dell'esperienza del medico. Infine, verrà utilizzata anche una variabile dummy che assume valore 1 se gli assistiti di un medico superano la soglia massima imposta dalla legge (1500 assistiti), 0 altrimenti.

Di seguito sono riportate tutte le variabili esplicative descritte sopra.

VARIABILE		DESCRIZIONE
Variabili di base		
Sesso	Femmina	Dummy che assume valore 1 se l'assistito è femmina, 0 se uomo
Età	25-34	Dummy che vale 1 se l'assistito ha un'età compresa tra i 25 e i 34 anni, 0 altrimenti
	35-44	Dummy che vale 1 se l'assistito ha un'età compresa tra i 25 e i 34 anni, 0 altrimenti
	45-54	Dummy che vale 1 se l'assistito ha un'età compresa tra i 25 e i 34 anni, 0 altrimenti
	55-64	Dummy che vale 1 se l'assistito ha un'età compresa tra i 25 e i 34 anni, 0 altrimenti
	65-74	Dummy che vale 1 se l'assistito ha un'età compresa tra i 25 e i 34 anni, 0 altrimenti
	75+	Dummy che vale 1 se l'assistito ha un'età compresa tra i 25 e i 34 anni, 0 altrimenti
Decesso	2007	Dummy che vale 1 se l'assistito è deceduto nell'arco dell'anno 2007, 0 altrimenti
	2008	Dummy che vale 1 se l'assistito è deceduto nell'arco dell'anno 2008, 0 altrimenti
ADI	Ambulatorio	Dummy che vale 1 se l'assistito ha ricevuto assistenza domiciliare in ambulatorio
	Domicilio	Dummy che vale 1 se l'assistito ha ricevuto assistenza domiciliare nel proprio domicilio
Esenzioni	Affezioni sistema circolatorio	Dummy che vale 1 se l'assistito ha ricevuto almeno un' esenzione per affezioni del sistema circolatorio nel 2007, 0 altrimenti
	Malattie cardiache	Dummy che vale 1 se l'assistito ha ricevuto almeno un' esenzione per malattie cardiache nel 2007, 0 altrimenti
	Diabete	Dummy che vale 1 se l'assistito ha ricevuto almeno un' esenzione per diabete nel 2007, 0 altrimenti
	Glaucoma	Dummy che vale 1 se l'assistito ha ricevuto almeno un' esenzione per glaucoma nel 2007, 0 altrimenti

	Ipertensione arteriosa con danno d'organo	Dummy che vale 1 se l'assistito ha ricevuto almeno un' esenzione per ipertensione arteriosa con danno d'organo nel 2007, 0 altrimenti
	Ipertensione arteriosa senza danno d'organo	Dummy che vale 1 se l'assistito ha ricevuto almeno un' esenzione per ipertensione arteriosa senza danno d'organo nel 2007, 0 altrimenti
	Patologie neoplastiche maligne	Dummy che vale 1 se l'assistito ha ricevuto almeno un' esenzione per patologie neoplastiche maligne nel 2007, 0 altrimenti
	Invalità civile >2/3	Dummy che vale 1 se l'assistito ha ricevuto almeno un' esenzione per invalidità civile >2/3 nel 2007, 0 altrimenti
	Invalità civile con assegno accompagnamento	Dummy che vale 1 se l'assistito ha ricevuto almeno un' esenzione per invalidità civile con assegno accompagnamento nel 2007, 0 altrimenti
	Invalità civile al 100%	Dummy che vale 1 se l'assistito ha ricevuto almeno un' esenzione per invalidità civile al 100% nel 2007, 0 altrimenti
	Gravidanza	Dummy che vale 1 se l'assistito ha ricevuto almeno un' esenzione per gravidanza nel 2007, 0 altrimenti
	Altro	Dummy che vale 1 se l'assistito ha ricevuto almeno un' esenzione per qualsiasi altro tipo di esenzione che non siano quelle descritte sopra nel 2007, 0 altrimenti
Variabili di secondo livello		
Sesso	Medico maschio	Dummy che vale 1 se il medico è maschio, 0 altrimenti
Età MMG		Variabile continua dell'età dei MMG
Anni dalla laurea		Variabile continua calcolata come 2007-anno di laurea del MMG, divisa per 100
Anni dalla laurea²		Variabile continua calcolata come anno dalla laurea al quadrato, divisa per 1000
Supera max assistiti		Variabile dicotomica che vale 1 se il medico ha superato la soglia degli assistiti massimi, 0 altrimenti

Tabella 8 - Variabili esplicative

Capitolo 2

2. I modelli

In questo capitolo saranno presentate le principali caratteristiche dei modelli stimati in questa analisi. Il primo obiettivo che mi sono posta è cercare di stilare il profilo degli assistiti più onerosi per il SSN. Decido di considerare i modelli per variabili binarie, come il modello probit, e il modello di regressione lineare per quanto riguarda le variabili continue. L'obiettivo successivo è quello di individuare se esiste una certa variabilità tra i medici. Se così fosse, si potrebbero individuare i medici così detti *iperprescrittori*. Per catturare un eventuale effetto della gerarchia nei dati decido di stimare modelli multilivello. Successivamente presenterò il modello Tobit univariato e multivariato, per poi concludere con i modelli Heckman univariati.

2.1 Modelli per variabili binarie

2.1.1 Il modello probit

Per individuare le caratteristiche degli assistiti che gravano effettivamente sul Sistema Sanitario Regionale del Friuli-Venezia Giulia, ho utilizzato un modello per variabili dipendenti binarie, in particolare un modello probit. I modelli probit sono utilizzati quando la variabile risposta da modellare è dicotomica, ossia assume solo due modalità. In questi casi, le modalità della variabile dipendente vengono codificate in “1” e “0”. Nel caso in analisi dispongo di tre variabili dipendenti, che vengono codificate nel modo seguente:

-ricoveri: variabile dipendente che vale 1 se l'assistito ha effettuato almeno un ricovero ospedaliero nell'arco del 2007, 0 altrimenti;

-*specialistiche*: variabile dipendente che vale 1 se l'assistito ha effettuato almeno una visita specialistica nell'arco del 2007, 0 altrimenti;

-*farmaci*: variabile dipendente che vale 1 se l'assistito ha acquistato almeno un farmaco di fascia A nell'arco del 2007, 0 altrimenti.

L'obiettivo dei modelli a risposta binaria è descrivere la "scelta" (y_i) fra due alternative discrete, ovvero descrivono $\Pr(y_i = 1|x_i)$, con x_i insieme di variabili esplicative. Si provi a considerare un modello lineare che spieghi y_i in funzione delle caratteristiche x_i :

$$y_i = x_i' \beta + \varepsilon_i$$

Sotto l'ipotesi che $E(\varepsilon_i|x_i) = 0$, in questo modello

$$E(y_i|x_i) = x_i' \beta$$

Poiché y_i è una variabile con distribuzione di Bernoulli:

$$\begin{aligned} E(y_i|x_i) &= 1 \cdot \Pr(y_i = 1|x_i) + 0 \cdot \Pr(y_i = 0|x_i) = \\ &= \Pr(y_i = 1|x_i) \end{aligned}$$

Ne consegue che il modello stimato spiega la probabilità che un evento si realizzi:

$$y_i = \Pr(y_i = 1|x_i) + \varepsilon_i$$

ovvero la probabilità che l'evento si realizzi è una funzione lineare:

$$\Pr(y_i = 1|x_i) = x_i' \beta$$

Questo modello è detto *linear probability model (LPM)*, che può essere stimato mediante OLS.

Tuttavia, lo stimatore OLS risente di alcuni problemi:

- ε_i non ha una distribuzione normale, è distribuito come una Bernoulli perché può assumere solo due valori;
- ε_i è eteroschedastico, la varianza dell'errore dipende dalle x ;
- questi modelli non impongono la restrizione che $0 \leq \Pr(y_i|x_i) \leq 1$.

Perciò l'alternativa è usare una funzione del tipo:

$$\Pr(y_i = 1|x_i) = F(x_i'\beta)$$

dove $F(\cdot)$ è una funzione di ripartizione assunta simmetrica attorno allo 0, in modo tale che $F(u) = 1 - F(-u)$ (Hansen, 2012).

Nel caso in analisi F è una normale, $F(u) = \Phi(u)$, e il modello si chiama modello probit.

La specificazione di questo modello può essere ottenuta anche per mezzo di un approccio a variabile latente:

$$y_i^* = x_i'\beta + \varepsilon_i$$

$$\varepsilon_i \sim F(\cdot)$$

$$y_i = \begin{cases} 1 & \text{se } y_i^* > 0 \\ 0 & \text{altrimenti} \end{cases}$$

Perciò,

$$\begin{aligned} \Pr(y_i = 1|x_i) &= \Pr(y_i^* > 0|x_i) \\ &= \Pr(x_i'\beta + \varepsilon_i > 0|x_i) \\ &= \Pr(\varepsilon_i > -x_i'\beta|x_i) \\ &= 1 - F(-x_i'\beta) \\ &= F(x_i'\beta) \end{aligned}$$

Nella maggior parte delle applicazioni dei modelli a risposta binaria, lo scopo primario è quello di spiegare gli effetti delle diverse variabili esplicative x_i sulla probabilità $\Pr(y_i = 1|x)$.

Per tali modelli, la direzione dell'effetto di x_i su $E(y_i^*|x_i) = x_i'\beta$ e su $E(y_i|x_i) = P(y_i = 1|x_i) = F(x_i'\beta)$ è data dal segno della stima del parametro associato a x_i . Al contrario, la magnitudo delle stime dei singoli parametri non è molto utile in quanto non forniscono l'effetto marginale a differenza dei modelli di regressione lineare (Wooldridge, 2012).

Se x_i è una variabile continua, il suo effetto marginale su $p(x) = P(y_i = 1|x_i)$ è ottenuto da:

$$\frac{\partial p(x)}{\partial x_i} = f(x_i' \beta) \beta_i, \quad \text{dove } f(z) \equiv \frac{dF}{dz}(z).$$

Nel caso dei modelli probit, $F(\cdot)$ è una funzione di densità cumulata, perciò $f(z) > 0$ per ogni z . Pertanto, l'effetto marginale di x_i su $p(x)$ dipende da x attraverso la quantità positiva $f(x_i' \beta)$; questo significa che l'effetto marginale ha sempre lo stesso segno di β_i .

Se x_1 è una variabile esplicativa binaria, l'effetto marginale di x_1 nel passare da 0 a 1, tenendo fisse tutte le altre variabili, è il seguente:

$$F(\beta_0 + \beta_1 + \beta_2 x_2 + \dots + \beta_k x_k) - F(\beta_0 + \beta_2 x_2 + \dots + \beta_k x_k).$$

Nella grande maggioranza dei casi, i parametri presenti nei modelli a scelta binaria vengono stimati per mezzo del metodo della massima verosimiglianza. Per costruire la verosimiglianza, serve la distribuzione condizionale di una singola osservazione. In generale, il contributo alla verosimiglianza dell'osservazione i con $y_i = 1$ è data da $\Pr(y_i = 1|x_i)$, interpretata come funzione del vettore dei parametri ignoti β , e analogamente per $y_i = 0$ (Veerbek, 2010).

La funzione di verosimiglianza per l'intero campione è dunque data da

$$L(\beta) = \prod_{i=1}^N P(y_i = 1|x_i; \beta)^{y_i} P(y_i = 0|x_i; \beta)^{1-y_i}$$

E la funzione di log-verosimiglianza è la seguente:

$$\log L(\beta) = \sum_{i=1}^N y_i \log F(x_i' \beta) + \sum_{i=1}^N (1 - y_i) \log [1 - F(x_i' \beta)]$$

Questa equazione definisce una funzione obiettivo da massimizzare rispetto a β .

2.2 Modelli multilivello

Dopo aver tracciato i profili degli assistiti che gravano sul sistema sanitario del Friuli-Venezia Giulia, introduco nella mia analisi i modelli multilivello in modo da studiare quanta variabilità totale viene spiegata dalla variabilità tra i medici di medicina generale. Questo tipo di modello differisce dal modello di regressione lineare multivariato standard perché contiene più di un termine di errore, più precisamente un termine d'errore per ogni livello. La variabile dipendente sarà sempre di livello-1 perché quello che interessa è spiegare cosa succede al livello più basso della gerarchia.

Il principale aspetto di cui i modelli multilivello tengono conto è che la struttura dei dati della popolazione in analisi è di tipo gerarchico. Questi modelli si sono sviluppati negli ultimi 30 anni per rispondere al problema della ricerca sociale che cerca di capire la relazione tra individui e società. Il concetto generale è che gli individui interagiscono con il contesto sociale di appartenenza, ovvero che le persone vengono influenzate dal gruppo sociale o dal contesto in cui vivono e che tali gruppi vengono influenzati a loro volta dalle persone che compongono quel determinato gruppo. Gli individui e i gruppi sociali sono concepiti come un sistema gerarchico di individui annidati all'interno dei gruppi, con individui e gruppi definiti a livelli diversi di questo sistema gerarchico. Questi sistemi possono essere osservati a diversi livelli gerarchici e le variabili possono essere definite ad ogni livello. Tutto ciò porta allo studio delle relazioni tra variabili che caratterizzano gli individui e le variabili dei gruppi, un tipo di ricerca che viene chiamata appunto analisi multilivello (Hox, 2011).

Il livello più basso è associato al livello 1 che in questa analisi è costituito dagli abitanti del Friuli-Venezia Giulia, con età maggiore di 14 anni, che nel 2007 erano assegnati a un MMG. Il livello 2, ovvero i gruppi, è invece costituito dai medici di medicina generale operanti nel Friuli-Venezia Giulia nel 2007.

Nell'analisi multilivello, le variabili possono essere definite a qualsiasi livello della gerarchia. Alcune di queste variabili possono essere definite direttamente al loro livello naturale, ovvero per quanto riguarda gli assistiti possiamo introdurre la variabile che riguarda l'età e per i medici la variabile che riguarda l'anno di laurea. Inoltre possiamo "spostare" le variabili da un livello ad un altro per aggregazione o disaggregazione. Aggregazione significa che le variabili

in un livello inferiore vengono spostate a un livello superiore. Disaggregazione è lo spostamento di variabili a livello inferiore.

Storicamente, i problemi dei modelli multilivello hanno portato ad analisi il cui approccio era quello di spostare tutte le variabili, aggregandole o disaggregandole, in un unico livello di interesse per poi svolgere un'analisi multivariata standard, o analisi della varianza o altri metodi di analisi. Tuttavia, analizzare variabili definite a livelli diversi in un unico livello comune è inadeguato e porta a vari tipi di problemi.

Il primo problema è di tipo statistico. Se i dati vengono aggregati il risultato è che diversi valori provenienti da molte sub-unità sono raggruppati in un minor numero di valori per un minor numero di unità di livello superiore. Questo porta a una grande perdita di informazione e quindi ad una perdita di precisione. Se, al contrario, i dati vengono disaggregati, il risultato è che i singoli dati provenienti dal ristretto numero di unità superiori sono divisi in molti più valori per un numero maggiore di sub-unità. I test statistici ordinari considerano che i valori disaggregati siano, in genere, informazioni indipendenti provenienti dall'insieme delle unità di livello inferiore. Invece, nelle situazioni in cui i dati sono gerarchicamente organizzati, tale ipotesi viene generalmente a cadere. Il comportamento degli individui è infatti influenzato dal contesto sociale nel quale sono inseriti e le caratteristiche di un gruppo sono influenzate dagli individui che formano il gruppo stesso: gli individui e il contesto sociale nel quale vivono possono essere visti come un sistema gerarchico di individui e gruppi, nel quale gli individui e i gruppi agiscono a livelli diversi. I test statistici tradizionali, basati sull'assunto di indipendenza, producono stime distorte degli errori standard e, di conseguenza, i risultati che si ottengono possono apparire "impropriamente" significativi. Ciò significa che l'analista si troverà ad avere molti risultati significativi quando in realtà è falso (Ruscione, 2013).

Il secondo problema è di tipo concettuale. Se l'analista non è molto attento all'interpretazione dei risultati può commettere l'errore di analizzare i dati ad un livello e formulare conclusioni rispetto ad un altro livello. Questo tipo di errore è definito *ecological fallacy*, ovvero interpretare dati aggregati a livello individuale (Hox, 2011).

2.2.1 Modello multilivello a due livelli

Per individuare i principali effetti prodotti dall'esistenza di una struttura gerarchica delle osservazioni si può far riferimento ad un modello a due soli livelli di raggruppamento. Si supponga che le i unità di primo livello (assistiti), o unità elementari, siano aggregate in J gruppi di unità di secondo livello (medici di medicina generale) e si assuma che il totale delle unità elementari raggruppate entro il j -esimo gruppo siano pari a n_j ($j = 1, \dots, J$). Della variabile oggetto di interesse Y , osservata sulle unità elementari, si desidera indagare in merito al legame con l'insieme delle variabili esplicative X . Si suppone che tale legame sia di tipo lineare e, inoltre, che esso possa variare, da gruppo a gruppo, in relazione all'azione di una variabile esplicativa Z , che interviene al secondo livello.

2.2.1.1 Notazione

Introduco la notazione principale utilizzata nella stima dei modelli multilivello:

- j = indice delle unità di livello-2, $j=1, 2, \dots, J$
- i = indice delle unità di livello-1 all'interno del gruppo j , $i=1, 2, \dots, n_j$
- y_{ij} = valore assunto dalla variabile risposta per l'individuo i del gruppo j
- X_{ij} = variabile esplicativa a livello-1 che esprime una caratteristica dell'unità i -esima del gruppo j -esimo
- Z_j = valore assunto dalla variabile esplicativa di livello-2 per il gruppo j -esimo

2.2.2 Coefficiente di Correlazione Intraclasse

Il coefficiente di correlazione intraclasse (ICC) è dato dal rapporto tra la varianza tra i gruppi e la varianza totale. La varianza tra i gruppi viene anche chiamata variabilità between, mentre la varianza totale è data dalla somma della variabilità between con quella within, ossia entro i gruppi.

$$\rho = \frac{\text{variabilità between}}{\text{variabilità between} + \text{variabilità within}}$$

dove $0 \leq \rho \leq 1$, valori alti di ρ implica una forte variabilità tra le unità di secondo livello.

Ci sono diversi modi per definire questo indicatore, ma Kreft e De Leeuw (1998) lo presentano, ad esempio, come una misura dell'omogeneità dei gruppi.

L'inserimento nel modello di covariate di primo livello dovrebbe contribuire a spiegare la variabilità individuale, riducendo la componente residua di primo livello (variabilità within). L'inserimento di variabili di secondo livello dovrebbe, invece, spiegare la variabilità tra i gruppi e ridurre la varianza between, determinando una diminuzione dell'ICC.

2.2.3 Modello ad intercetta casuale con covariate di livello-1 (Random Intercept Model)

Il modello ad intercetta casuale è un modello per dati gerarchici in cui solo l'intercetta varia casualmente tra i gruppi. Le variabili di primo livello contribuiscono a spiegare la variabilità del fenomeno sia a livello di gruppo che a livello individuale, mentre le variabili di secondo livello (variabili di contesto) spiegano la variabilità del fenomeno a livello di gruppo.

Introduco per semplicità di espressione una sola variabile di primo livello nel modello, l'equazione del modello in questione è quindi la seguente:

$$Y_{ij} = \alpha_j + \beta_{1j}X_{1ij} + \varepsilon_{ij}$$

dove

$\alpha_j = \gamma_{00} + u_{0j}$, è l'intercetta casuale che riflette l'effetto di gruppo

$\beta_{1j} = \beta_1$ è il coefficiente di regressione che si suppone rimanga costante per tutti i gruppi

In questo modo l'effetto della variabile esplicativa X_1 su Y è lo stesso per tutti i gruppi, cambia solamente il livello medio della risposta.

In forma compatta

$$Y_{ij} = \gamma_{00} + \beta_1 X_{1ij} + u_{0j} + \varepsilon_{ij}$$

con:

$\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$ è la componente d'errore di livello-1;

$u_{0j} \sim N(0, \sigma_{u0}^2)$ è la componente d'errore di livello-2;

$$\text{Cov}(\varepsilon_{ij}, u_{0j}) = 0$$

La parte deterministica del modello è rappresentata da $\gamma_{00} + \beta_1 X_{1ij}$, mentre $u_{0j} + \varepsilon_{ij}$ rappresenta la parte casuale o aleatoria.

Questo modello viene talvolta definito modello a componenti di varianza per il fatto che la varianza residua è scomposta in due componenti, una per ogni livello di gerarchia:

$$\text{Var}(Y_{ij} | X_{1ij}) = \sigma_\varepsilon^2 + \sigma_{u0}^2$$

2.2.4 Modello a intercetta casuale con covariate di livello-1 e livello-2

Dopo aver verificato la significatività delle variabili esplicative di primo livello e la bontà di adattamento del modello individuato, possono essere inserite nel modello variabili di contesto, o di secondo livello. Le covariate di secondo livello consentono di definire un modello per ogni parametro di primo livello e di ridurre la varianza di secondo livello.

L'equazione del modello sarà la seguente:

$$Y_{ij} = \alpha_j + \beta_{1j} X_{1ij} + \varepsilon_{ij}$$

e in questo caso avremo:

$$\alpha_j = \gamma_{00} + \gamma_{01} Z_{1j} + u_{0j}$$

Questa equazione consente di prevedere il valore medio del fenomeno del gruppo j-esimo a seguito di variazioni di valori della variabile Z_1 . Quindi se γ_{01} è positivo, la media di Y sarà più elevata in quei gruppi che hanno un valore della variabile Z_1 più elevato; viceversa se γ_{01} è negativo, la media di Y sarà minore in quei gruppi che hanno un valore della variabile Z_1 più elevato.

2.2.5 Modello a pendenza casuale (Random Slope Model)

Il random slope model è un modello per dati gerarchici nel quale i gruppi differiscono casualmente tra loro sia per l'intercetta che per i coefficienti di regressione. Il modello a intercetta casuale descritto nel paragrafo precedente può essere visto come un caso particolare del modello a pendenza casuale ma, in questo caso, sia l'intercetta che la pendenza sono *random*, infatti la varianza dei coefficienti di regressione è diversa da zero.

Nel dettaglio, definiamo il modello a pendenza casuale come:

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + \varepsilon_{ij} \quad (1)$$

con

$$\beta_{0j} = \gamma_{00} + u_{0j} \quad (2)$$

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

Sostituendo la (2) nella (1) si ottiene il modello combinato:

$$\begin{aligned} Y_{ij} &= (\gamma_{00} + u_{0j}) + (\gamma_{10} + u_{1j})X_{1ij} + \varepsilon_{ij} = \\ &= \gamma_{00} + \gamma_{10}X_{1ij} + u_{1j}X_{1ij} + u_{0j} + \varepsilon_{ij} \end{aligned}$$

Il modello può quindi essere suddiviso in:

- parte fissa: $\gamma_{00} + \gamma_{10}X_{1ij}$
- parte casuale: $u_{1j}X_{1ij} + u_{0j} + \varepsilon_{ij}$

I termini d'errore sono così distribuiti:

$$\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$$

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{u_0}^2 & \sigma_{u_0 u_1} \\ \sigma_{u_0 u_1} & \sigma_{u_1}^2 \end{bmatrix}\right)$$

$$\text{e } \varepsilon_{ij} \perp \begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix}$$

Gli errori di secondo livello rappresentano gli scarti tra il valore del parametro per il gruppo j e il valore medio del parametro nella popolazione:

$$u_{0j} = \beta_{0j} - \gamma_{00}$$

$$u_{1j} = \beta_{1j} - \gamma_{10}$$

A differenza del modello a intercetta casuale, la varianza residua non è più costante ma dipende dal valore delle covariate. In particolare, per il modello considerato, la varianza residua è una funzione quadratica di X_{1ij} :

$$\text{Var}(Y_{ij} | X_{1ij}) = \sigma_{\varepsilon}^2 + \sigma_{u_0}^2 + 2\sigma_{u_0} \sigma_{u_1} X_{1ij} + \sigma_{u_1}^2 X_{1ij}^2$$

La covarianza condizionata per due individui i e i' appartenenti allo stesso gruppo j è:

$$\text{Cov}(Y_{ij}, Y_{i'j} | X_{1ij}, X_{1i'j}) = \sigma_{u_0}^2 + \sigma_{u_0} \sigma_{u_1} X_{1ij} + \sigma_{u_0} \sigma_{u_1} X_{1i'j} + \sigma_{u_1}^2 X_{1ij} X_{1i'j} \quad i \neq i'$$

Nel random slope model, l'indice ICC perde significato in quanto la correlazione all'interno dei gruppi non è più costante e dipende dal valore assunto dalle covariate (Snijders e Bosker, 1999).

2.2.6 Metodo di stima

La teoria statistica alla base dei modelli multilivello è molto complicata. Per i dati osservati, siamo interessati a stimare i parametri del modello multilivello, ovvero i coefficienti di regressione e le componenti di varianza. Il principale metodo di stima usato nelle analisi multilivello è il metodo della massima verosimiglianza (Maximum Likelihood - ML). Questo metodo produce degli standard error che possono essere usati per test di significatività dei parametri, come il test di Wald. Gli standard error sono asintotici perciò i p-value ottenuti sono spesso un'approssimazione. Il metodo della massima verosimiglianza permette di ottenere delle stime che godono di alcune proprietà importanti, come la consistenza e l'efficienza. Esistono due variazioni del metodo di massima verosimiglianza che spesso sono applicate nelle analisi statistiche. Una è chiamata Full Maximum Likelihood (FML), in questo metodo sia i coefficienti della regressione sia le componenti di varianza sono incluse nella funzione di verosimiglianza. L'altro metodo è chiamato Restricted Maximum Likelihood (REML), in questo caso solo le componenti di varianza sono incluse nella

funzione di verosimiglianza. La differenza tra i due metodi è che FML tratta le stime dei coefficienti della regressione come quantità note quando sono stimate le componenti di varianza, mentre REML le tratta come stime che hanno un certo ammontare di incertezza (Bryk e Raudenbush, 1992; Goldstein, 1995). Dato che REML è più realistico, dovrebbe portare a stime migliori rispetto a FML, specialmente quando il numero dei gruppi non è ampio. Per ottenere le stime, si utilizzano dei metodi iterativi. Innanzitutto, il software statistico genera dei valori iniziali per i veri parametri. Dopo un'iterazione, si ottengono le stime dei Minimi Quadrati Generalizzati (GLS). Se non ci sono problemi nel processo iterativo si raggiunge la convergenza alle stime ML o REML (Hox, 1995).

2.2.7 Misure di bontà di adattamento

Per quanto riguarda i modelli multilivello, non c'è esattamente un equivalente dell' R^2 che indica quanto il modello si adatti bene ai dati come nel caso dei modelli di regressione lineare. Esistono, però, altre misure di bontà di adattamento.

L'indice di correlazione intraclasse (ICC) evidenzia quanta parte della variabilità totale è ascrivibile a variabilità between. Questo indice può essere usato per capire se un modello multilivello è necessario oppure no per l'analisi di un certo dataset e anche quali variabili di livello-2 sono più utili per spiegare la variabilità between.

Il metodo della massima verosimiglianza produce una statistica chiamata devianza, che misura quanto il modello si adatti bene ai dati. In generale, modelli con una bassa devianza si adattano meglio rispetto ai modelli con una devianza più grande. Il metodo REML invece non produce misure di devianza.

L'analisi svolta con i modelli sopra descritti porterà a risultati significativi per quanto riguarda i modelli probit univariati e la regressione univariata svolta solo sulle osservazioni positive. Come già detto, i modelli probit, però, non tengono conto della massa di probabilità dell'evento $y_i^* > 0$, ma si limitano a spiegare una variabile risposta binaria. La retta di regressione, d'altro canto, è stata stimata solo sui valori positivi delle variabili dipendenti, questo perché i dataset a disposizione presentavano molti "0", ovvero persone che non hanno effettuato ricoveri, oppure non hanno eseguito visite specialistiche o non hanno acquistato farmaci. Questo mi porta ad introdurre il modello tobit, che può essere descritto come la combinazione dei modelli precedenti.

2.3 Il modello Tobit

2.3.1 Introduzione

I modelli Tobit si riferiscono a modelli di regressione con dati censurati o troncati in cui il range dei valori assunti dalle variabili dipendenti può essere limitato. In economia, questo tipo di modello fu per la prima volta introdotto da Tobin (1958) in un suo studio definito pioniere. Tobin analizzò la spesa delle famiglie sui beni durevoli usando un modello di regressione che doveva tenere specificatamente conto che la spesa (variabile dipendente) non poteva assumere valori negativi. Tobin chiamò il suo modello: il modello con variabile dipendente limitata. Quest'ultimo e le sue generalizzazioni sono conosciute oggi come modelli Tobit, una frase coniata da Goldberger (1964) per la sua similarità con i modelli probit. Questi modelli sono conosciuti anche come modelli di regressione censurati o troncati. Sono chiamati troncati se le osservazioni al di fuori di uno specifico range sono completamente perse e censurati se possiamo almeno osservare le variabili esogene. Tra gli anni 1958 e gli anni 1970 questi modelli vennero raramente usati ma dai primi anni Settanta iniziarono ad essere largamente sfruttati. Ciò a causa della crescente disponibilità di micro campioni di dati di varie indagini e del recente sviluppo di software che la tecnologia e i computer hanno permesso di sfruttare per studiare milioni di dati che fino a quel momento non potevano essere analizzati (Amemiya, 1985).

Siano $i=1\dots n$ i pazienti del SSN che vivono in Friuli-Venezia Giulia nel 2007, dai quali sono state raccolte informazioni circa le spese per farmaci, per le visite specialistiche e per i ricoveri ospedalieri, y_i , e su alcune loro caratteristiche, raccolte nel vettore x_i .

Se si vuole studiare, per esempio, la spesa per farmaci (variabile dipendente), ci si trova di fronte ad osservazioni del tipo seguente,

$$\left\{ \begin{array}{ll} (y_i > 0, x_i) & \text{per i pazienti che hanno acquistato almeno un farmaco} \\ (0, x_i) & \text{per i pazienti che non hanno acquistato farmaci} \end{array} \right.$$

Occorre un modello che tenga conto di questa caratteristica dei dati, tenendo presente che il valore "0" non rappresenta un valore "mancante" per la variabile dipendente, bensì rappresenta

il comportamento di un gruppo di pazienti, i quali hanno scelto di non acquistare farmaci (anche perché, magari, non ne avevano bisogno), hanno scelto cioè una soluzione ad angolo.

2.3.2 La distribuzione normale troncata

Uno strumento utile per lo studio di questo problema è la distribuzione normale troncata. Sia y il valore osservato della variabile dipendente. Contrariamente a quello che si osserva nel modello di regressione lineare classico, y è solo una parte del valore osservato che è attribuibile alla variabile dipendente latente y^* . In effetti, con il troncamento dal basso, si osserva solo $y=y^*$ se y^* è più grande del punto in cui si tronca, chiamato τ . Così facendo si perdono le osservazioni di y^* che sono più piccole o uguali a τ . Quando ciò succede si assume che la variabile $y|y > \tau$ segue una distribuzione normale troncata. Il problema nel quale bisogna focalizzarsi è che, in realtà, è stata rimossa una parte della distribuzione originale perciò la distribuzione non ha più un'area uguale a uno (Golder, 2003). Per fare in modo che l'area oltre al punto in cui si è troncato la distribuzione sia uguale a 1 bisogna procedere con una riscalatura.

Sia y una variabile dipendente continua distribuita come una normale di media μ e varianza σ^2 e τ una costante.

La funzione di densità di y è:

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} \quad (3)$$

Questa distribuzione normale può essere riscritta come funzione di una distribuzione normale standard $N(0,1)$, sia $z = \frac{y-\mu}{\sigma}$ la variabile standardizzata con funzione di densità:

$$f(z) = \phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right)$$

La (3) può essere quindi riscritta come:

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} = \frac{1}{\sigma} \phi\left(\frac{y-\mu}{\sigma}\right)$$

dove $\phi(\cdot)$ è la funzione di densità della normale standard.

Sia $\Phi(z)$ la funzione di ripartizione di z , che può essere scritta nel modo seguente:

$$P(Y > y) = 1 - \Phi\left(\frac{y - \mu}{\sigma}\right) \quad (4)$$

Dati questi risultati si ottiene la distribuzione di $y|y > \tau$:

$$f(y|y > \tau) = \frac{f(y)}{P(y > \tau)}$$

Usando i risultati da (4):

$$P(Y > \tau) = 1 - \Phi\left(\frac{\tau - \mu}{\sigma}\right) = 1 - \Phi(\alpha)$$

$$\text{dove } \alpha = \frac{\tau - \mu}{\sigma}.$$

La funzione di densità è quindi pari a:

$$f(y|y > \tau) = \frac{f(y)}{1 - \Phi(\alpha)} = \frac{\frac{1}{\sigma} \phi\left(\frac{y - \mu}{\sigma}\right)}{1 - \Phi\left(\frac{\tau - \mu}{\sigma}\right)} = \frac{\frac{1}{\sigma} \phi\left(\frac{y - \mu}{\sigma}\right)}{1 - \Phi(\alpha)} \quad (5)$$

Dato un campione di N osservazioni indipendenti, la funzione di verosimiglianza della distribuzione normale troncata è la seguente:

$$L = \prod_{i=1}^N \frac{f(y)}{1 - \Phi(\alpha)}$$

da cui la log-verosimiglianza:

$$\ln L = \sum_{i=1}^N (\ln[f(y)] - \ln[1 - \Phi(\alpha)])$$

2.3.2.1 Momenti della distribuzione normale troncata

Sia $y \sim N(\mu, \sigma^2)$ e τ il punto in cui viene troncata la distribuzione, allora

$$E[y | y > \tau] = \mu + \sigma \frac{\phi\left(\frac{\tau - \mu}{\sigma}\right)}{1 - \Phi\left(\frac{\tau - \mu}{\sigma}\right)} = \mu + \sigma \lambda(\alpha)$$

e

$$\text{Var}[y | y > \tau] = \sigma^2 \{1 - \lambda(\alpha)[\lambda(\alpha) - \alpha]\} = \sigma^2 [1 - \delta(\alpha)]$$

$$\text{dove } \alpha = \frac{\tau - \mu}{\sigma},$$

$$\delta(\alpha) = \lambda(\alpha)[\lambda(\alpha) - \alpha]$$

e

$$\lambda(\alpha) = \frac{\phi(\alpha)}{[1 - \Phi(\alpha)]}$$

$\lambda(\alpha)$ è il reciproco del rapporto di Mill (*inverse Mill's ratio, IMR*), possiamo definire $\lambda(\alpha)$ *as measuring the amount of truncation* (Long, 1997).

La distribuzione normale troncata, però, presenta degli aspetti negativi. Per esempio, se la funzione è troncata da sotto, la media della distribuzione troncata è più grande della media della distribuzione non troncata. Se la funzione è troncata dall'alto, la media della variabile troncata è più piccola della media originale. Inoltre la varianza della distribuzione troncata è più piccola rispetto alla varianza della distribuzione non troncata (Golder, 2003).

2.3.2.2 Modello di regressione troncato

Sia $y \sim N(X\beta, \sigma^2)$; siamo interessati alla distribuzione di y dato che y è più grande di un certo punto di troncamento, chiamato τ . Abbiamo quindi,

$$E[y | y > \tau] = X\beta + \sigma \frac{\phi\left[\frac{\tau - X\beta}{\sigma}\right]}{1 - \Phi\left[\frac{\tau - X\beta}{\sigma}\right]} \quad (6)$$

Come si può notare la media condizionale non è una funzione lineare di τ, σ, X e β . Se usassimo le stime OLS sbaglieremmo perché OLS ommette tutto ciò che è dopo $X\beta$, perciò ci sarebbe un problema di variabili omesse. Il termine d'errore sarebbe eteroschedastico e le stime di β inconsistenti.

2.3.3 Il modello Tobit I

2.3.3.1 Introduzione del modello

Sulla base dei risultati del precedente paragrafo è possibile introdurre il modello tobit (o modello di regressione normale censurato). Nel caso in analisi si tratta di un modello con soluzione ad angolo, y ha valore 0 con probabilità positiva ma ha distribuzione continua per tutti i valori di y strettamente positivi. Si tratta di risolvere un problema di massimizzazione.

Nel modello per una soluzione ad angolo si vuole fare inferenza su:

- $\Pr(y_i = 0 | x_i)$ cioè la probabilità di avere zero (possibile scelta)
- $E(y_i | x_i, y_i > 0)$ il valore atteso condizionale di un valore positivo

Perciò il valore atteso condizionale (la funzione di regressione) di y_i dato x_i è dato da

$$E(y_i | x_i) = \Pr(y_i = 0 | x_i) \cdot 0 + [1 - \Pr(y_i = 0 | x_i)] \cdot E(y_i | x_i, y_i > 0)$$

Un modello econometrico coerente deve

1. dare una probabilità positiva di ottenere un valore pari a “zero”
2. escludere una previsione negativa

Possiamo scrivere il tutto in termini di un modello con variabile dipendente latente (equazione strutturale del modello Tobit) (Wooldridge, 2002):

$$y_i^* = x_i' \beta + \varepsilon_i$$

dove $\varepsilon_i \sim N(0, \sigma^2)$. Il valore osservato y è definito in questo modo:

$$y_i = \begin{cases} y^* & \text{se } y^* > 0 \\ 0 & \text{se } y^* \leq 0 \end{cases}$$

2.3.3.2 Valori attesi del modello tobit

1. Probabilità di uno zero

$$\Pr(y_i = 0 | x_i) = \Pr(y^* \leq 0 | x_i) = \Pr(\varepsilon_i \leq -x_i' \beta | x_i) = \Pr\left(\frac{\varepsilon_i}{\sigma} \leq -\frac{x_i' \beta}{\sigma} | x_i\right) = 1 - \Phi(x_i' \beta / \sigma)$$

2. Valore atteso condizionale di $y > 0$, che si ottiene a partire dai risultati da (6):

$$E(y_i | y_i > 0; x_i) = X_i \beta + \sigma \frac{\phi\left[\frac{0 - X_i \beta}{\sigma}\right]}{1 - \Phi\left[\frac{0 - X_i \beta}{\sigma}\right]} = X_i \beta + \sigma \frac{\phi\left[\frac{X_i \beta}{\sigma}\right]}{\Phi\left[\frac{X_i \beta}{\sigma}\right]} = X_i \beta + \sigma \lambda(\alpha)$$

dove $\lambda(\alpha)$ è il reciproco del rapporto di Mill.

3. Valore atteso della variabile latente y^* :

$$E[y^*] = X_i \beta$$

4. Valore atteso di y :

$$\begin{aligned} E[y_i | x_i] &= E(y_i | x_i, y_i = 0) \cdot \Pr(y_i = 0 | x_i) + E(y_i | x_i, y_i > 0) \cdot \Pr(y_i = 1 | x_i) = \\ &= 0 + \Phi(x_i' \beta / \sigma) \left(x_i' \beta + \sigma \frac{\phi(x_i' \beta / \sigma)}{\Phi(x_i' \beta / \sigma)} \right) \end{aligned}$$

Questa è la probabilità di essere non censurato moltiplicato per il valore atteso di y dato y non censurato.

2.3.3.3 La funzione di verosimiglianza

Riprendiamo il modello in questione:

$$y_i^* = x_i' \beta + \varepsilon_i \quad \varepsilon | x \sim N(0, \sigma^2)$$

$$y_i = \begin{cases} y^* & \text{se } y^* > 0 \\ 0 & \text{se } y^* \leq 0 \end{cases} \Rightarrow \begin{cases} \text{se } \varepsilon > -x' \beta \text{ oppure } \frac{\varepsilon}{\sigma} > -\frac{x' \beta}{\sigma} \\ \text{se } \varepsilon < -x' \beta \text{ oppure } \frac{\varepsilon}{\sigma} < -\frac{x' \beta}{\sigma} \end{cases}$$

Quindi:

$$\Pr(y = 0 | x) = \Pr\left(\frac{\varepsilon}{\sigma} < -\frac{x' \beta}{\sigma} | x\right) = \Phi\left(-\frac{x' \beta}{\sigma}\right) = 1 - \Phi\left(\frac{x' \beta}{\sigma}\right)$$

$$\Pr(y > 0 | x) = 1 - \Phi\left(-\frac{x' \beta}{\sigma}\right) = \Phi\left(\frac{x' \beta}{\sigma}\right)$$

Dalla (5) sappiamo che:

$$\begin{aligned} f(y | x, y > 0) &= f(\varepsilon | x, \varepsilon > -x' \beta) = \\ &= \frac{\frac{1}{\sigma} \Phi\left(\frac{\varepsilon}{\sigma}\right)}{1 - \Phi\left(-\frac{x' \beta}{\sigma}\right)} = \frac{\frac{1}{\sigma} \Phi\left(\frac{y - x' \beta}{\sigma}\right)}{1 - \Phi\left(-\frac{x' \beta}{\sigma}\right)} \end{aligned}$$

Allora

$$\begin{aligned}
 f(y|x) &= [\Pr(y=0|x)]^{I_{y=0}} \cdot [\Pr(y>0|x) \cdot f(y|x, y>0)]^{I_{y>0}} = \\
 &= \left[1 - \Phi\left(\frac{x'\beta}{\sigma}\right) \right]^{I_{y=0}} \cdot \left[\left(1 - \Phi\left(-\frac{x'\beta}{\sigma}\right) \right) \cdot \frac{\frac{1}{\sigma} \Phi\left(\frac{y-x'\beta}{\sigma}\right)}{1 - \Phi\left(-\frac{x'\beta}{\sigma}\right)} \right]^{I_{y>0}} = \\
 &= \left[1 - \Phi\left(\frac{x'\beta}{\sigma}\right) \right]^{I_{y=0}} \cdot \left[\frac{1}{\sigma} \frac{1}{2\pi} \exp\left\{-\frac{1}{2}\left(\frac{y-x'\beta}{\sigma}\right)^2\right\} \right]^{I_{y>0}}
 \end{aligned}$$

Si tratta di una mistura di distribuzione, una discreta ed una continua dove:

- $I_{y_i=0}$ è la funzione indicatrice dell'evento $y_i=0$: $I_{y_i=0}=1$ se $y_i=0$
- $I_{y_i>0}$ è la funzione indicatrice dell'evento $y_i>0$: $I_{y_i>0}=1$ se $y_i>0$

La funzione di verosimiglianza per il modello tobit è quindi la seguente:

$$L = \prod_i^N \left[1 - \Phi\left(\frac{X_i\beta}{\sigma}\right) \right]^{I_{y_i=0}} \left[\frac{1}{\sigma} \phi\left(\frac{y_i - X_i\beta}{\sigma}\right) \right]^{I_{y_i>0}}$$

Indicando con I_0 l'insieme delle osservazioni per le quali si osserva $y_i=0$ e con I_1 l'insieme delle osservazioni per le quali si osserva $y_i>0$, la log-verosimiglianza diventa pari a:

$$\ln L = \sum_{i \in I_0} \ln\left(1 - \Phi\left(\frac{X_i\beta}{\sigma}\right)\right) + \sum_{i \in I_1} \left(-\ln \sigma + \ln \phi\left(\frac{y_i - X_i\beta}{\sigma}\right)\right)$$

La log verosimiglianza è composta da due parti. La prima è la probabilità che un'osservazione sia censurata, la seconda corrisponde alla classica funzione di regressione per dati non censurati (Green, 2002).

2.4 La statistica multivariata

La statistica multivariata è una tecnica sempre più popolare usata per analizzare dataset complicati. Essa permette di fornire analisi quando sono coinvolte molte variabili indipendenti e/o molte variabili dipendenti, correlate o meno l'una con l'altra. A causa delle complesse analisi statistiche svolte con l'analisi univariata e della disponibilità di software sempre più specifici degli ultimi tempi, oggi giorno l'analisi multivariata è largamente utilizzata. La popolarità dei modelli multivariati, infatti, non è dovuta al caso bensì alla crescente complessità degli attuali studi di ricerca. I metodi di statistica multivariata sono un'estensione dei modelli statistici univariati e bivariati. In particolare, i modelli multivariati sono il caso completo o generale, considerando che le statistiche univariate o bivariate sono casi speciali del modello multivariato.

Il termine statistica univariata si riferisce alle analisi in cui c'è una singola variabile dipendente. Ci possono essere, comunque, più variabili indipendenti, dette anche predittori perché predicono la variabile dipendente Y , chiamata anche variabile risposta.

Con la statistica multivariata si analizzano simultaneamente più variabili dipendenti e indipendenti.

Un aspetto molto importante è l'ortogonalità tra le variabili. Se due variabili sono ortogonali, la loro correlazione è zero e questo porta a molti vantaggi in termini di analisi multivariata. Se tutte le variabili esplicative di un modello sono tra loro ortogonali, questo significa che ciascuna di queste variabili aiuta a predire singolarmente la variabile dipendente.

L'ortogonalità è facilmente illustrata nei diagrammi di Venn, come mostra la Figura 10. I diagrammi di Venn rappresentano la correlazione (o varianza condivisa) come sovrapposizione delle aree tra due o più cerchi.

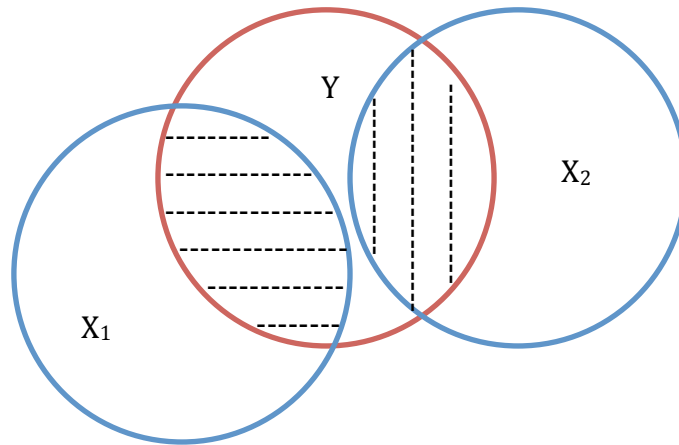


Figura 10 - Diagramma di Venn, non correlazione tra variabili esplicative

In questo caso, la variabile Y è la variabile dipendente e le variabili X_1 e X_2 sono le variabili esplicative. La varianza totale di Y è rappresentata dall'intero cerchio rosso. La sezione con linee orizzontali rappresenta la parte della variabile Y che è spiegata da X_1 e la sezione con linee verticali rappresenta la parte predetta da X_2 . La somma di queste due aree rappresenta la parte di varianza spiegata dalle variabili esplicative. In questo caso le due sezioni non si sovrappongono, appunto perché le due variabili sono ortogonali.

Diversa è la situazione se le variabili esplicative non sono ortogonali.

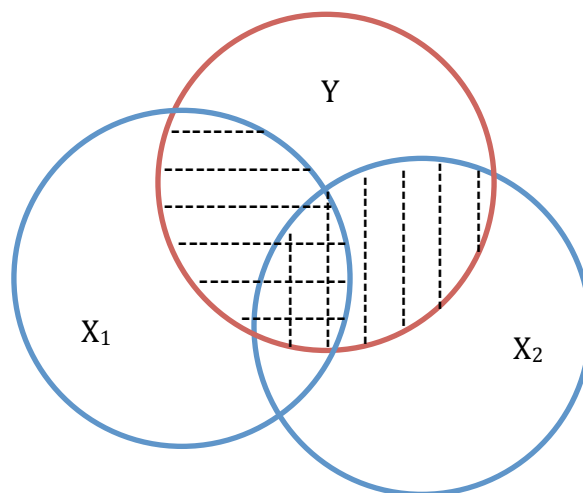


Figura 11 - Diagramma di Venn, correlazione tra variabili esplicative

Nella Figura 11 si nota che le aree di X_1 e X_2 si sovrappongono. Nonostante il contributo indipendente delle due variabili sia lo stesso della Figura 10, la loro predizione congiunta non è la stessa della figura precedente bensì qualcosa di meno dovuto alla sovrapposizione delle aree.

Una decisione importante per l'analista statistico è come gestire la varianza che è predetta da più di una variabile. Nelle analisi standard, la varianza sovrapposta contribuisce alla dimensione delle statistiche riassuntive della relazione generale, ma non è assegnata a nessuna variabile. Nelle analisi sequenziali, invece, l'analista assegnerà una certa priorità alla variabili che entrano nell'equazione che spiega la variabile risposta, e la prima variabile che entra sarà quella che avrà sia la varianza propria che quella sovrapposta con altre variabili. Le altre variabili a bassa priorità verranno inserite solamente con la propria varianza individuale. (Tabachnick and Fidell, 2007).

2.4.1 Vantaggi nell'usare i modelli multivariati

Se ci sono più variabili dipendenti, è probabile che alcune di esse siano correlate tra di loro, perciò usare più regressioni univariate porterebbe ad avere test separati in cui si analizza la stessa varianza.

Inoltre, un altro vantaggio dell'analisi multivariata è che la realizzazione di una serie di test statistici univariati può portare ad accettare erroneamente l'ipotesi nulla, gonfiando così l'errore di I° tipo, che è controllato meglio in un'analisi multivariata.

Un terzo vantaggio è che spesso le analisi multivariate sono più potenti. Su ogni variabile risposta individuale, le differenze possono essere piccole e insignificanti, ma per l'insieme totale delle variabili risposta, l'effetto congiunto può produrre un effetto significativo.

2.5 Tobit multivariato

I modelli trattati fino ad ora analizzano i tre tipi di costi, spesa per ricoveri, visite specialistiche e farmaci, a livello disaggregato. Poiché le spese analizzate si riferiscono allo stesso campo, ovvero la Sanità Pubblica, è probabile che le variabili risposta possano influenzarsi l'una con l'altra. E' ipotizzabile che un ricovero ospedaliero avvenga previa visita specialistica, oppure che la spesa per farmaci di una determinata persona aumenti dopo un controllo specialistico. Decido di introdurre, quindi, il modello Tobit multivariato per cercare di capire se esiste una correlazione tra le variabili dipendenti.

Il modello Tobit multivariato viene stimato in STATA come M-modelli tobit tramite il metodo della massima verosimiglianza simulata. Questo metodo risolve il problema della non trattabilità dell'integrale, sostituendo la funzione di verosimiglianza con una sua approssimazione ottenuta attraverso l'uso di simulazioni: si ottengono così funzioni obiettivo trattabili da un punto di vista computazionale, che possono essere utilizzate per la stima dei parametri. Il comando usa lo stimatore GHK (Geweke-Hajivassiliou-Keane) che calcola le probabilità di una normale multivariata e la relativa funzione usata per delineare numeri casuali per la valutazione degli integrali multidimensionali normali nella funzione di verosimiglianza. Per ogni osservazione, un contributo di probabilità è calcolato per ogni replica e il contributo di verosimiglianza simulata è la media dei valori derivati da tutte le repliche. La funzione di verosimiglianza simulata per il campione nel suo complesso viene quindi massimizzata usando metodi standard, in questo caso la massima verosimiglianza.

2.6 Il modello di Heckman

Per l'ultima parte dell'analisi di questa tesi, ho deciso di adattare ai dati un'estensione del modello tobit, il modello Heckman (o Tobit II). Gli "zeri" presenti nei miei dataset in realtà non sono associati a valori mancanti ma hanno un significato economico vero e proprio. Gli "zeri" sono associati alle persone che non hanno acquistato farmaci, o effettuato una visita specialistica, o un ricovero, probabilmente perché non ne avevano bisogno. Il modello Tobit, però, presenta per costruzione una struttura alquanto restrittiva: le variabili che determinano la probabilità di un'osservazione nulla hanno esattamente la stessa relazione delle variabili che influenzano il livello di un'osservazione positiva (Verbeek, 2010).

Adatto il modello di Heckman ai dati a mia disposizione, per esempio la spesa per farmaci.

L'equazione lineare per la spesa per farmaci è la seguente:

$$\text{equazione di regressione } y_{1j} = x'_{1j}\beta_1 + \varepsilon_{1j} \quad (7)$$

La seguente equazione è chiamata equazione di selezione, riconducibile a un modello probit standard.

$$\text{equazione di selezione } y_{2j}^* = x'_{2j}\beta_2 + \varepsilon_{2j} \quad (8)$$

y_{2j}^* è una variabile latente binaria che indica se l'assistito ha effettuato una spesa per farmaci o meno.

Vale la seguente regola

$$y_{1j} = \begin{cases} x'_{1j}\beta_1 + \varepsilon_{1j} & \text{se } y_{2j} = 1 \text{ ovvero se } y_{2j}^* > 0 \\ \text{non osservata} & \text{se } y_{2j} = 0 \text{ ovvero se } y_{2j}^* \leq 0 \end{cases}$$

Il modello è completato da un'ipotesi sulla distribuzione degli errori $(\varepsilon_{1j}, \varepsilon_{2j})$, di solito normale bivariata con valore atteso nullo, varianza rispettivamente σ_1^2 e σ_2^2 e covarianza σ_{12} .

$$\begin{pmatrix} \varepsilon_{1j} \\ \varepsilon_{2j} \end{pmatrix} \sim NID \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \right]$$

Il modello (8) è un modello probit standard che descrive se un assistito ha effettuato una certa spesa per farmaci oppure no, e per questa ragione è necessario imporre una condizione di normalizzazione, che di solito corrisponde a $\sigma_2^2 = 1$. L'aver effettuato una spesa dipende dalle variabili x_{2j} con coefficienti β_2 . L'equazione (7) descrive la spesa (potenziale) in funzione delle variabili x_{1j} con coefficienti β_1 .

La spesa per farmaci attesa condizionale al fatto che un individuo abbia effettuato tale spesa è:

$$\begin{aligned}
 E(y_{1j} | y_{2j} = 1) &= x'_{1j}\beta_1 + E(\varepsilon_{1j} | y_{2j} = 1) = & (9) \\
 &= x'_{1j}\beta_1 + E(\varepsilon_{1j} | \varepsilon_{2j} > -x'_{2j}\beta_2) = \\
 &= x'_{1j}\beta_1 + \frac{\sigma_{12}}{\sigma_2^2} E(\varepsilon_{2j} | \varepsilon_{2j} > -x'_{2j}\beta_2) = y_{1j} = x'_{1j}\beta_1 + \sigma_{12}\lambda_j + \eta_j \\
 &= x'_{1j}\beta_1 + \sigma_{12} \frac{\phi(x'_{2j}\beta_2)}{\Phi(x'_{2j}\beta_2)}
 \end{aligned}$$

Dove $\lambda_j = \frac{\phi(x'_{2j}\beta_2)}{\Phi(x'_{2j}\beta_2)}$ è noto come inverso del rapporto di Mill, ed è anche chiamato lambda di Heckman. Si noti che $\sigma_{12} = \rho_{12}\sigma_1$, dove ρ_{12} è il coefficiente di correlazione fra i due errori. Da (9) segue che la spesa per farmaci attesa condizionale è pari a $x'_{1j}\beta_1$ solo se $\sigma_{12} = \rho_{12} = 0$. Se i termini d'errore delle due equazioni sono incorrelati, l'equazione della spesa per farmaci può essere stimata in maniera consistente usando i minimi quadrati ordinari. Se, invece, $\sigma_{12} \neq 0$, lo stimatore OLS è affetto da una distorsione dovuta alla selezione campionaria. Proprio il coefficiente di correlazione tra i termini d'errore rende questo modello di selezione campionaria diverso dalla semplice combinazione di un modello di regressione e di un modello probit (Verbeek, 2010).

2.6.1 Stima

La stima dei parametri del modello di Heckman può essere di due tipi: il primo è la massima verosimiglianza, il secondo è una procedura a due passi. In questo elaborato i modelli saranno sempre stimati usando il metodo della massima verosimiglianza, che produce stimatori consistenti e asintoticamente efficienti, con distribuzione asintotica normale.

Il modello può essere interpretato come composto da due parti. La prima descrive il problema della scelta binaria, il contributo dell'osservazione i -esima alla funzione di verosimiglianza è la probabilità di osservare $y_{2j} = 1$ oppure $y_{2j} = 0$. La seconda parte descrive la distribuzione della spesa per farmaci (o visite specialistiche o ricoveri) per i soli individui che hanno avuto una spesa positiva; per questa ragione il contributo alla verosimiglianza è dato da $f(y_{1j} | y_{2j} = 1)$.

La funzione di log-verosimiglianza sarà quindi:

$$\log L(\beta, \sigma_1^2, \sigma_{12}) = \sum_{i \in I_0} \log P(y_{2j} = 0) + \sum_{i \in I_1} [\log f(y_{1j} | y_{2j} = 1) + \log P(y_{2j} = 1)]$$

La componente di scelta binaria è standard, più complicata, però, è la distribuzione condizionale di y_{1j} dato $y_{2j} = 1$. Per questa ragione, è più frequente scomporre in maniera diversa la distribuzione congiunta di y_{1j} e y_{2j} , sfruttando la seguente proprietà:

$$f(y_{1j} | y_{2j} = 1) P(y_{2j} = 1) = P(y_{2j} = 1 | y_{1j}) f(y_{1j})$$

Possiamo, quindi, riformulare la log-verosimiglianza come:

$$\log L(\beta, \sigma_1^2, \sigma_{12}) = \sum_{i \in I_0} \log P(y_{2j} = 0) + \sum_{i \in I_1} [\log f(y_{1j}) + \log P(y_{2j} = 1 | y_{1j})]$$

Massimizzando tale funzione rispetto ai parametri ignoti si ottengono stimatori consistenti e asintoticamente efficienti, con distribuzione asintotica normale (Verbeek, 2010).

Capitolo 3

3. Stima dei modelli

In questo capitolo presenterò i risultati ottenuti dall'applicazione dei vari modelli descritti nel Capitolo 2. I primi modelli che analizzerò sono i modelli probit, considerando come variabile dipendente l'aver oppure no sostenuto una spesa per ricoveri ospedalieri, visite specialistiche e farmaci rispettivamente. Metterò quindi in paragone i modelli probit con i modelli probit multilivello, in modo tale da capire se la struttura gerarchica insita nei dati (individui assistiti da MMG) abbia un effetto significativo nella probabilità di sostenere un costo oppure no. Inoltre, inserirò delle variabili di secondo livello (caratteristiche dei medici di medicina generale) in modo da spiegare con maggior dettaglio, se esiste, la variabilità tra i medici di medicina generale. Lo stesso procedimento verrà quindi proposto per il modello di regressione lineare, con un campione ridotto rispetto ai modelli probit. Infatti, il modello di regressione lineare verrà stimato solo per le osservazioni aventi spesa positiva. Considerando come variabile dipendente il livello sostenuto di spesa (ma solo per gli assistiti che hanno realmente sostenuto una spesa), è possibile studiare il profilo degli assistiti che maggiormente gravano sul SSN del Friuli-Venezia Giulia.

Successivamente, passerò all'analisi delle stime dei modelli tobit univariati e del modello tobit multivariato. Concludo il capitolo con l'analisi delle stime dei modelli Heckman univariati.

Il software da me utilizzato in questa tesi è Stata 12⁷. Per eseguire le stime dei modelli multilivello si è fatto ricorso in particolare al modulo GLLAMM (Generalized Linear Latent and Mixed Models)⁸, che dà stime dei modelli più precise (adattando QGO e QGA per le stime) rispetto ad altri comandi di Stata (per esempio xtmixed), anche se i tempi di stima risultano abbastanza lunghi. La Quadratura Gaussiana Adattiva (QGA) ha una buona accuratezza nella stima dei coefficienti della parte fissa del modello persino in campioni molti ridotti, ma sottostima alcune componenti della varianza anche in campioni di enormi

⁷ <http://www.stata.com>

⁸ <http://www.gllamm.org>

dimensioni. In termini di distorsione, comunque, AGQ produce risultati migliori e le stime più accurate (Paccagnella, 2011).

3.1 Stima dei modelli probit e probit multilivello

Decido di iniziare la mia analisi studiando il comportamento delle variabili esplicative quando la variabile risposta è di tipo binario: aver effettuato una spesa oppure no. Quali sono le caratteristiche degli assistiti che più influenzano la probabilità di aver effettuato un ricovero ospedaliero, oppure una visita specialistica o aver acquistato dei farmaci? Possono le caratteristiche del medico di medicina generale influire su tali comportamenti?

Le Tabelle 9,10,11 riportano le stime ottenute con il modello probit standard e probit multilivello, con e senza covariate del medico, per quanto riguarda i ricoveri ospedalieri, visite specialistiche e farmaci rispettivamente.

In questi modelli è stata inserita la variabile “anni dalla laurea” come proxy dell’esperienza del medico di medicina generale e non la variabile “età”, in quanto fortemente correlata con la prima.

3.1.1 Ricoveri ospedalieri

Ricoveri		Probit				Probit con covariate del medico			
		Standard (1)		Multilivello (2)		Standard (3)		Multilivello (4)	
Variabili esplicative		<i>Coeff</i>	<i>P-value</i>	<i>Coeff</i>	<i>P-value</i>	<i>Coeff</i>	<i>P-value</i>	<i>Coeff</i>	<i>P-value</i>
Sesso	Femmina	0,050	0,000	0,049	0,000	0,050	0,000	0,049	0,000
Età	25-34	0,140	0,000	0,139	0,000	0,140	0,000	0,139	0,000
	35-44	0,147	0,000	0,146	0,000	0,147	0,000	0,146	0,000
	45-54	0,149	0,000	0,150	0,000	0,151	0,000	0,150	0,000
	55-64	0,237	0,000	0,237	0,000	0,238	0,000	0,238	0,000
	65-74	0,365	0,000	0,364	0,000	0,366	0,000	0,365	0,000
	≥75	0,451	0,000	0,450	0,000	0,452	0,000	0,450	0,000
Decesso	2007	1,430	0,000	1,432	0,000	1,430	0,000	1,432	0,000
	2008	0,535	0,000	0,535	0,000	0,535	0,000	0,535	0,000
ADI	Ambulatorio	0,168	0,000	0,196	0,000	0,167	0,000	0,195	0,000
	Domicilio	0,749	0,000	0,757	0,000	0,749	0,000	0,757	0,000
Esenzioni	Affezioni sistema circolatorio	0,260	0,000	0,253	0,000	0,260	0,000	0,253	0,038
	Malattie cardiache	0,429	0,000	0,435	0,000	0,429	0,000	0,435	0,000
	Diabete	0,209	0,000	0,209	0,000	0,209	0,000	0,209	0,000
	Glaucoma	0,046	0,000	0,044	0,000	0,046	0,000	0,044	0,000
	Colesterolo	0,032	0,008	0,025	0,038	0,032	0,008	0,025	0,008
	Ipertensione arteriosa con danno d'organo	0,100	0,000	0,102	0,000	0,100	0,000	0,102	0,000
	Ipertensione arteriosa senza danno d'organo	0,045	0,000	0,053	0,000	0,045	0,000	0,053	0,000
	Malattie neoplastiche maligne	0,528	0,000	0,528	0,000	0,528	0,000	0,528	0,000
	Invalità civile > 2/3	0,243	0,000	0,242	0,000	0,243	0,000	0,242	0,000
	Invalità civile (assegno accomp.)	0,109	0,000	0,099	0,000	0,108	0,000	0,099	0,000
	Invalità civile al 100%	0,234	0,000	0,229	0,000	0,233	0,000	0,229	0,000
	Gravidanza	1,915	0,000	1,932	0,000	1,914	0,000	1,931	0,000
	Altro	0,303	0,000	0,302	0,000	0,303	0,000	0,302	0,000
Costante		-1,714	0,000	-1,715	0,000	-1,623	0,000	-1,622	0,000
Sesso medico	Maschio					-0,0006	0,858	-0,002	0,695
	Supera max assistiti					-0,007	0,047	-0,008	0,176
	Anni dalla laurea					-0,553	0,001	-0,561	0,030
	Anni dalla laurea^2					0,076	0,023	0,077	0,131
σ^2_2				0,0042				0,0040	
ICC				0,42%				0,40%	

Tabella 9 - Stime dei modelli probit standard e multilivello, con o senza covariate del medico, dei ricoveri ospedalieri

Le stime dei parametri del modello probit standard senza covariate del medico sono tutte significative al 5% di livello e questo significa che tutte le caratteristiche degli assistiti inserite influenzano la variabile risposta, aver effettuato un ricovero ospedaliero oppure no. Come spiegato nel capitolo 2, le stime ottenute possono essere utilizzate solo per determinare la direzione dell'effetto di ciascuna variabile, *ceteris paribus*. Infatti, l'effetto di una variazione in x_{ik} dipende dal valore di tutte le x , ma il segno dell'effetto sulla variabile risposta di una variazione in x_{ik} coincide con il segno del coefficiente corrispondente $\hat{\beta}_k$. Dalla Tabella 9 si può notare come gli assistiti donna, tenendo fissi gli altri fattori, abbiano maggior probabilità di aver effettuato un ricovero ospedaliero nel 2007, rispetto agli assistiti maschi. Anche le dummy dell'età hanno tutte segno positivo, perciò gli assistiti dai 25 anni in poi hanno più probabilità di aver effettuato un ricovero ospedaliero rispetto ai pazienti compresi tra i 14 e 24 anni, *ceteris paribus*. In particolare, si nota una relazione crescente con l'età. Per quanto riguarda la variabile esenzione, tutti i coefficienti ad essa associati hanno segno positivo. Si noti che l'esenzione per patologia viene concessa ad un assistito previa una visita specialistica accurata, perciò è ipotizzabile che i pazienti che sono idonei all'esenzione abbiano maggior probabilità di essere ricoverati rispetto a chi non ha nessuna esenzione, proprio perché ci si riferisce a malattie importanti. L'assistenza domiciliare integrata viene concessa a persone non autosufficienti o gravemente malate, perciò, anche in questo caso, il segno positivo di tale stima non sorprende. Le persone decedute nel 2007 e nel 2008 mostrano maggior probabilità di essere state ricoverate rispetto a chi non è deceduto.

Al contrario del modello di regressione lineare, per i modelli di scelta binaria non esiste un'unica misura di adattamento ai dati. Spesso le misure di adattamento ai dati sono basate implicitamente o esplicitamente sul confronto con un modello che contiene come esplicativa solo la costante (Verbeek, 2010). Una misura di adattamento ai dati può essere lo *pseudo R²*, un indicatore che vale 0 se tutte le stime dei coefficienti valgono zero (modello nullo) e vale 1 se il modello genera dei valori previsti che coincidono esattamente con quelli osservati. In questo modello l'indice *pseudo R²* vale 0.14 ed è un valore non molto elevato, ma i modelli Probit si adattano bene ai dati anche con un valore di *pseudo R²* tendenzialmente basso. Inoltre, la numerosità del campione analizzato è molto grande, infatti conta 1.126.348 osservazioni.

Per quanto riguarda il modello probit multilivello corrispondente, ovvero il random intercept model senza covariate del medico (colonna 2), esso mostra che i coefficienti sono tutti significativi e positivi, esattamente come il modello standard. Anche se non parliamo di effetti

marginali, le stime tra i due modelli sono molto simili. L'ICC è pari a 0.42%, un valore abbastanza basso e questo indica la presenza di una scarsa variabilità tra MMG: la variabilità totale è quasi esclusivamente variabilità within.

Passando al modello probit con covariate del medico (colonna 3), si noti come in questo caso le variabili delle caratteristiche degli assistiti siano tutte significative al 5% di livello ad eccezione di quelle relative alle caratteristiche del medico di medicina generale. Le stime delle caratteristiche degli assistiti hanno lo stesso segno di quelle del modello probit senza covariate del medico (colonna 1) e in molti casi addirittura la stima puntuale è esattamente la stessa. Per quanto riguarda le variabili del medico, la maggior parte di esse non ha alcun valore predittivo. Il genere del medico non è significativo al 5%, mentre la variabile che indica l'esperienza lavorativa "anni dalla laurea" è significativa con segno della stima negativo. La variabile "anni dalla laurea al quadrato", invece, è significativa al 5% e non al 1% perciò sono molto cauta ad esprimere un'opinione a riguardo. L'andamento di questa variabile è parabolico, gli assistiti associati a un medico con pochi anni di esperienza hanno meno probabilità di effettuare un ricovero ospedaliero. Questa tendenza, però, si inverte nel tempo, può essere che i medici con molti anni di esperienza siano anche i medici più anziani e questo può portare a sviste dovute all'età tali da incrementare la probabilità che i pazienti abbiano effettuato un ricovero. La dummy che indica se un medico ha superato il numero massimo di assistiti non è significativa all'1% ed è al limite della significatività al 5%, perciò è un dato da prendere con cautela.

L'indice *pseudo R*² del modello probit con covariate del medico vale anche in questo 0.14, esattamente come nel caso senza covariate del medico.

Passando alle stime del modello probit multilivello con covariate del medico (colonna 4), notiamo come la significatività delle variabili abbia un comportamento simile a quella del modello precedente (colonna 3). Le stime delle caratteristiche degli assistiti sono statisticamente tutte significative al 5% di livello ed hanno lo stesso segno dei precedenti modelli. Le variabili delle caratteristiche dei medici di medicina generale sono tutte non significative al 5%, eccetto la variabile anni dalla laurea che è significativa solo al 5% e non all'1%.

L'ICC di questo modello è dello 0.40% e scende leggermente rispetto al modello senza covariate. L'inserimento di variabili di secondo livello fa diminuire solitamente questo indice,

ma già si partiva da una situazione in cui la variabilità tra medici è di fatto assente nello spiegare l'aver effettuato un ricovero o meno.

Per quanto riguarda i ricoveri, i modelli probit (standard e multilivello) mostrano innanzitutto che le variabili di 2° livello, come il sesso del medico, gli anni di esperienza e l'aver superato la soglia di assistiti massimi, non hanno evidenziato nessuna associazione con la variabile dipendente. L'utilizzo di modelli multilivello per cercare di individuare un certo comportamento della popolazione a seconda del medico di appartenenza oppure per capire quali sono i medici con le spese più elevate, ha portato alla conclusione che non c'è alcun effetto significativo nel comportamento degli stessi MMG. Il comportamento degli assistiti non dipende dal medico al quale sono associati, bensì dipende fortemente dalle loro caratteristiche individuali.

Questo risultato non sorprende perché è plausibile pensare che i ricoveri ospedalieri siano dovuti sia a particolari situazioni di salute dei cittadini, sia ad eventi tendenzialmente esogeni ad essi, difficilmente modellabili con un dataset di tipo amministrativo come quello in esame.

3.1.2 Visite specialistiche

Specialistiche		Probit				Probit con covariate del medico			
		Standard (1)		Multilivello (2)		Standard (3)		Multilivello (4)	
Variabili esplicative		Coeff	P-value	Coeff	P-value	Coeff	P-value	Coeff	P-value
Sesso	Femmina	0,319	0,000	0,316	0,000	0,316	0,000	0,315	0,000
Età	25-34	-0,011	0,016	-0,005	0,273	-0,009	0,053	-0,005	0,282
	35-44	0,082	0,000	0,084	0,000	0,085	0,000	0,084	0,000
	45-54	0,197	0,000	0,199	0,000	0,200	0,000	0,200	0,000
	55-64	0,348	0,000	0,352	0,000	0,354	0,000	0,353	0,000
	65-74	0,622	0,000	0,626	0,000	0,628	0,000	0,626	0,000
	≥75	0,530	0,000	0,532	0,000	0,537	0,000	0,533	0,000
Decesso	2007	-0,579	0,000	-0,577	0,000	-0,578	0,000	-0,577	0,000
	2008	-0,122	0,000	-0,118	0,000	-0,120	0,000	-0,118	0,000
ADI	Ambulatorio	1,235	0,000	1,262	0,000	1,232	0,000	1,264	0,000
	Domicilio	0,628	0,000	0,633	0,000	0,627	0,000	0,633	0,000
Esenzioni	Affezioni sistema circolatorio	0,439	0,000	0,443	0,000	0,440	0,000	0,443	0,000
	Malattie cardiache	0,708	0,000	0,712	0,000	0,707	0,000	0,712	0,000
	Diabete	0,663	0,000	0,667	0,000	0,663	0,000	0,667	0,000
	Glaucoma	0,707	0,000	0,703	0,000	0,706	0,000	0,703	0,000
	Colesterolo	0,677	0,000	0,661	0,000	0,675	0,000	0,661	0,000
	Ipertensione arteriosa con danno d'organo	0,392	0,000	0,388	0,000	0,390	0,000	0,389	0,000
	Ipertensione arteriosa senza danno d'organo	0,290	0,000	0,293	0,000	0,289	0,000	0,293	0,000
	Malattie neoplastiche maligne	0,959	0,000	0,957	0,000	0,959	0,000	0,957	0,000
	Invalità civile > 2/3	0,414	0,000	0,413	0,000	0,414	0,000	0,413	0,000
	Invalità civile (assegno accomp.)	-0,027	0,013	-0,028	0,011	-0,027	0,015	-0,028	0,011
	Invalità civile al 100%	0,208	0,000	0,207	0,000	0,207	0,000	0,207	0,000
	Gravidanza	1,747	0,000	1,767	0,000	1,746	0,000	1,767	0,000
	Altro	0,700	0,000	0,694	0,000	0,699	0,000	0,694	0,000
Costante		-0,222	0,000	-0,217	0,000	-0,193	0,000	-0,049	0,174
Sesso medico	Maschio					-0,051	0,000	0,073	0,000
	Supera max assistiti					0,052	0,000	-1,244	0,000
	Anni dalla laurea					0,023	0,859	0,219	0,000
	Anni dalla laurea^2					-0,025	0,333	-0,046	0,000
σ^2				0,012				0,019	
ICC				1,18%				1,86%	

Tabella 10 - Stime dei modelli probit standard e multilivello, con o senza covariate del medico, delle visite specialistiche

In Tabella 10 sono riportate le stime delle variabili del modello probit standard senza covariate del medico, le quali sono tutte significative al 5% ma, a differenza del modello per ricoveri ospedalieri, l'età degli assistiti tra i 25 e 34 anni e l'esenzione per invalidità civile con assegno di accompagnamento non sono significative all'1%. Notiamo che essere donna rispetto ad essere maschio aumenta la probabilità di aver effettuato una visita specialistica, a parità di altre condizioni. Le persone con un'età maggiore di 35 anni hanno maggior probabilità di effettuare una visita specialistica rispetto alla categoria di assistiti con un'età compresa tra i 14 e 24 anni, *ceteris paribus*. La stima della variabile età che comprende le persone tra i 25 e 34 anni è da prendere con cautela. Con un livello di significatività del 5% possiamo definire questa categoria di persone come meno probabile ad aver fatto una visita specialistica rispetto alla categoria di riferimento, ovvero gli assistiti tra i 14 e 24 anni. Entrambe le categorie si riferiscono a ragazzi o comunque persone giovani, perciò credo che la differenza del numero di visite specialistiche sia labile ed è per tale motivo che la significatività di questa variabile sia da prendere con cautela. Le variabili deceduto nel 2007 e deceduto nel 2008 sono entrambe significative e di segno negativo: a parità di altre condizioni, una persona deceduta in uno di questi due anni ha meno probabilità di aver effettuato una visita specialistica nel 2007. Si può ipotizzare che un assistito nei suoi ultimi anni di vita abbia più spese assistenziali, ma per lo più saranno spese per ricoveri, come evidenziato dal modello probit standard in Tabella 9. Le visite specialistiche vengono prescritte dal medico solo in casi di stretto bisogno ed è quindi più probabile che una persona, nei suoi ultimi anni di vita, debba far fronte a ricoveri d'urgenza piuttosto che a visite specialistiche programmate da tempo.

Le variabili che descrivono il servizio di assistenza domiciliare sono entrambe significative e di segno positivo. Gli assistiti che usufruiscono di tale servizio sono per lo più anziani o persone non autosufficienti, perciò è più probabile che abbiano effettuato visite specialistiche rispetto a chi non è idoneo a tale servizio, a parità di altre condizioni. Tutte le stime delle variabili riguardanti le esenzioni per patologia e gravidanza sono significative e di segno positivo: chi ha un'esenzione ha maggiore probabilità di ricorrere a visite specialistiche rispetto a chi non ha nessuna esenzione per patologia, *ceteris paribus*. Questo risultato non sorprende poiché, come sottolineato nel Capitolo 1, le esenzioni per patologia possono essere interpretate come proxy di cattiva salute da parte degli assistiti, che quindi necessitano con maggior frequenza il ricorso a visite specialistiche. Fa eccezione la variabile esenzione per invalidità civile con assegno di accompagnamento che ha segno negativo. L'assegno di accompagnamento viene

concesso solo alle persone non autosufficienti, che necessitano di un aiuto da parte di terze persone perché non riescono a svolgere gli atti quotidiani della vita. Queste persone possono avere un problema fisico o psichico, ma non è detto che soffrano di una malattia che necessiti la prescrizione di visite specialistiche. Il loro stato richiede un aiuto continuo e permanente nel tempo, non è solo una fase della loro vita in cui hanno bisogno di un accompagnatore. Questa indennità, perciò, è concessa solo a seguito di esami dello stato del paziente in cui si riscontra un'invalidità totale e permanente. Gli assistiti ai quali è concessa questa esenzione, hanno minore probabilità di compiere una visita specialistica rispetto a chi non ha nessuna esenzione per patologia, *ceteris paribus*.

L'indice di adattamento ai dati *pseudo R²* vale 0,12. E' un valore molto basso, addirittura inferiore a quello ottenuto nella stima per ricoveri ospedalieri.

Passando all'estensione multilivello del modello probit senza covariate del medico (colonna 2) notiamo subito che le stime sono tutte significative al 5% ed hanno lo stesso segno del modello probit non multilivello, ad eccezione della variabile età 25-34 anni. Quest'ultima risulta essere non significativa (confermando così i dubbi descritti precedentemente): rispetto alla categoria di riferimento, assistiti tra i 14 e 24 anni, non c'è una differenza statisticamente significativa in termini di visite specialistiche con i pazienti tra i 25 e 34 anni. Una differenza maggiore si nota sicuramente con le classi di età dai 35 anni in su, la cui probabilità di aver effettuato una visita specialistica è maggiore rispetto agli assistiti tra i 14 e 24 anni, *ceteris paribus*. Aver stimato un modello multilivello non ha cambiato molto l'interpretazione delle stime. L'ICC è del 1.18%, un valore molto basso che indica che non è presente un effetto legato alla gerarchia nei dati. Anche in questo caso, la variabilità totale nell'aver effettuato oppure no una visita specialistica è ascrivibile quasi completamente a variabilità individuale.

Per quanto riguarda il modello probit standard con covariate del medico (colonna 3), il comportamento delle stime delle variabili di primo livello (caratteristiche degli assistiti) segue quello dei modelli precedenti. Le stime sono tutte significative e il segno è lo stesso dei modelli probit descritti sopra. La dummy dell'età degli assistiti tra i 25 e i 34 anni rimane non significativa al livello del 5%. Le dummy delle esenzioni sono anch'esse tutte significative e di segno positivo, ad eccezione della variabile esenzione per invalidità civile con assegno di accompagnamento.

Passando alle variabili che descrivono le caratteristiche del medico, notiamo che i medici maschi hanno minore probabilità di prescrivere una visita specialistica rispetto ai medici di

sesto femminile, *ceteris paribus*. La variabile continua lineare e quadratica, anni dalla laurea del medico, usata come proxy dell'esperienza, non è significativa perciò non ha nessun valore predittivo. L'esperienza del medico non influisce sul fatto di aumentare o meno la probabilità di prescrivere visite specialistiche ai pazienti.

Il livello di significatività e il segno delle stime delle variabili del modello probit multilivello con covariate del medico (colonna 4) sono le stesse del modello multilivello senza covariate del medico (colonna 2). Inoltre, nel modello con le caratteristiche del medico, notiamo che le variabili associate al medico sono tutte significative, al contrario del modello non multilivello. Infatti, in questo caso anche le variabili riguardanti l'esperienza del medico sono significative. Al crescere dell'esperienza c'è più probabilità che un medico prescriva visite specialistiche per poi diminuire da un certo livello di esperienza in poi.

Sorprende poi il cambiamento di segno delle stime del parametro associato al genere del MMG: a parità di altre condizioni, gli assistiti seguiti da un MMG di sesso femminile tendono ad effettuare minor visite specialistiche rispetto agli assistiti seguiti da un MMG di sesso maschile.

L'ICC del modello multilivello rimane molto basso (1.86%), un valore tuttavia leggermente superiore a quello del modello senza covariate. Questo può significare che una struttura multilivello e l'introduzione di covariate a livello del medico riescono meglio a cogliere quella bassissima variabilità fra MMG insita in questi dati.

3.1.3 Farmaci

Farmaci		Probit				Probit con covariate del medico			
		Standard (1)		Multilivello (2)		Standard (3)		Multilivello (4)	
Variabili esplicative		Coeff	P-value	Coeff	P-value	Coeff	P-value	Coeff	P-value
Sesso	Femmina	0,311	0,000	0,314	0,000	0,312	0,000	0,314	0,000
Età	25-34	0,126	0,000	0,127	0,000	0,127	0,000	0,127	0,000
	35-44	0,233	0,000	0,236	0,000	0,233	0,000	0,236	0,000
	45-54	0,362	0,000	0,371	0,000	0,364	0,000	0,372	0,000
	55-64	0,621	0,000	0,634	0,000	0,625	0,000	0,635	0,000
	65-74	0,948	0,000	0,965	0,000	0,951	0,000	0,965	0,000
	≥75	1,122	0,000	1,143	0,000	1,127	0,000	1,143	0,000
Decesso	2007	-0,560	0,000	-0,560	0,000	-0,559	0,000	-0,560	0,000
	2008	0,047	0,006	0,051	0,003	0,049	0,000	0,050	0,004
ADI	Ambulatorio	0,534	0,000	0,547	0,000	0,527	0,000	0,545	0,000
	Domicilio	0,817	0,000	0,818	0,000	0,816	0,000	0,817	0,000
Esenzioni	Affezioni sistema circolatorio	0,580	0,000	0,584	0,000	0,581	0,000	0,583	0,000
	Malattie cardiache	0,835	0,000	0,840	0,000	0,834	0,000	0,841	0,000
	Diabete	0,905	0,000	0,910	0,000	0,906	0,000	0,910	0,000
	Glaucoma	0,864	0,000	0,867	0,000	0,864	0,000	0,867	0,000
	Colesterolo	0,862	0,000	0,876	0,000	0,865	0,000	0,877	0,000
	Ipertensione arteriosa con danno d'organo	1,023	0,000	1,031	0,000	1,023	0,000	1,031	0,000
	Ipertensione arteriosa senza danno d'organo	1,107	0,000	1,111	0,000	1,106	0,000	1,111	0,000
	Malattie neoplastiche maligne	0,416	0,000	0,420	0,000	0,417	0,000	0,420	0,000
	Invalità civile > 2/3	0,393	0,000	0,392	0,000	0,391	0,000	0,393	0,000
	Invalità civile (assegno accomp.)	-0,045	0,001	-0,036	0,000	-0,044	0,001	-0,036	0,006
	Invalità civile al 100%	0,242	0,000	0,253	0,000	0,241	0,000	0,253	0,000
	Gravidanza	0,359	0,000	0,347	0,000	0,358	0,000	0,347	0,000
	Altro	0,716	0,000	0,724	0,000	0,717	0,000	0,724	0,000
Costante		-0,382	0,000	-0,402	0,000	-0,201	0,000	-0,226	0,000
Sesso medico	Maschio					0,015	0,000	0,012	0,088
	Supera max assistiti					0,068	0,000	0,047	0,000
	Anni dalla laurea					-1,556	0,000	-1,302	0,000
	Anni dalla laurea^2					0,264	0,000	0,184	0,000
σ^2				0,015				0,015	
ICC				1,53%				1,48%	

Tabella 11 - Stime dei modelli probit standard e multilivello, con o senza covariate del medico, della spesa per farmaci

Dalla Tabella 11 (riferita ai risultati del modello sull'acquisto o meno di farmaci), si nota che le stime del modello probit standard sono tutte significative: le caratteristiche degli assistiti hanno un valore predittivo importante. Essere donna aumenta la probabilità di acquistare farmaci rispetto ad essere uomo, *ceteris paribus*. Chi è deceduto nel 2007, invece, presenta meno probabilità di acquistare farmaci rispetto a chi non è deceduto nel 2007. Probabilmente, le persone nell'ultimo anno di vita hanno in generale delle spese assistenziali maggiori ma non attribuibili ai farmaci di fascia A. Come evidenziato nei modelli probit per le spese per ricoveri ospedalieri, è plausibile che le persone malate, che sono decedute nel 2007, siano state ricoverate e che le spese a loro attribuibili siano quelle dei ricoveri ospedalieri e non quelle dei farmaci che si acquistano in farmacia previa ricetta medica. Le persone decedute nel 2008, invece, hanno maggiore probabilità di aver acquistato dei farmaci rispetto a chi non è deceduto nel 2008, *ceteris paribus*. Le persone alle quali è stata attuata l'assistenza domiciliare sono per lo più anziani o persone gravemente malate perciò è comprensibile che le stime di tali variabili aumentino la probabilità di acquistare farmaci, rispetto a chi non usufruisce di tale servizio, a parità di altre condizioni. Anche le variabili riguardanti le esenzioni per patologia e gravidanza sono tutte significative e di segno positivo, indicando, quindi, una maggior probabilità di aver acquistato farmaci. L'unica esenzione con segno negativo è l'esenzione per invalidità civile con assegno di accompagnamento. Le persone alle quali è attribuibile tale esenzione hanno meno probabilità di acquistare farmaci rispetto a chi non ha nessuna esenzione. L'assegno di accompagnamento è indirizzato a persone non autosufficienti e non è detto che esse soffrano di particolari malattie che necessitino l'uso di farmaci.

L'indice di adattamento ai dati *pseudo R²* è 0,17. E' pur sempre un valore basso ma i modelli probit si adattano bene anche a valori bassi di tale misura perciò si può concludere che il modello si adatta discretamente ai dati.

Passando all'estensione multilivello del probit senza covariate del medico (colonna 2) si nota come le stime siano tutte significative al 5% di livello ed abbiano gli stessi segni e grandezze del modello precedente, *ceteris paribus*. L'ICC è dell'1.53%, un indice molto basso che denota la presenza di una scarsa variabilità fra MMG.

Si analizzi ora il modello probit standard con covariate del medico (colonna 3). Anche in questo caso, si può notare che le variabili che riguardano le caratteristiche degli assistiti sono tutte significative ed hanno il medesimo segno del modello probit senza covariate del medico (colonna 1), *ceteris paribus*. Per questo motivo, l'interpretazione della direzione di tali stime è

la stessa. Vorrei porre l'attenzione, però, sulle covariate del medico di medicina generale. Contrariamente al modello probit dei ricoveri ospedalieri, in questo caso tutte le variabili del medico sono significative al 5%. Gli assistiti che hanno come medico di medicina generale un uomo hanno più probabilità di acquistare farmaci di fascia A, rispetto agli assistiti che hanno un medico donna, *ceteris paribus*. A parità di altre condizioni, gli assistiti ai quali è associato un medico il cui carico di pazienti supera il suo massimo, tendono ad aumentare la probabilità di acquistare farmaci. Può essere che un medico con troppi assistiti sia sbrigativo perciò tende a prescrivere più ricette del dovuto, senza magari visitare il paziente e cercare altre soluzioni. E' anche ipotizzabile, però, che un medico che supera il numero massimo di assistiti sia in realtà un bravo medico e molte persone vogliono essere associate a lui, perciò è inevitabile che avendo più pazienti del dovuto prescriva anche più ricette. A parità di altre condizioni, la variabile che indica l'esperienza del medico è significativa e la direzione dell'effetto è negativa ma positiva nella corrispondente variabile quadratica. Si tratta quindi di un andamento parabolico. Probabilmente nei primi anni di professione, i medici tendono a prescrivere meno ricette farmacologiche, magari per inesperienza o per paura di sbagliare terapia. Con il passare degli anni, invece, un medico ha la possibilità di aver a che fare con più pazienti e più malattie. In questo modo è in grado di acquisire quelle competenze tacite che gli permettono di trovare la soluzione al problema prescrivendo il farmaco specifico.

Il modello si adatta discretamente ai dati, l'indice *pseudo R²* è 0,17, esattamente come il modello precedente.

L'estensione multilivello del modello probit con covariate del medico (colonna 4) presenta un ICC dell'1.48%, leggermente più basso rispetto al modello senza covariate. Anche in questo caso, i dati a disposizione non presentano un particolare effetto dovuto alla gerarchia nei dati. Le stime delle variabili delle caratteristiche dei pazienti sono tutte significative al 5% e la direzione dell'effetto è la stessa di tutti gli altri modelli precedenti, *ceteris paribus*. Anche se le variabili delle caratteristiche del medico sono significative al 5%, eccetto il genere, posso concludere che l'ICC mi suggerisce che un modello multilivello non è indicato per questo tipo di dati.

In conclusione, per quanto riguarda i modelli probit della spesa per farmaci, non c'è nessun tipo di variabilità tra i medici che possa giustificare un comportamento particolare degli stessi.

3.2 Stima dei modelli di regressione lineare e regressione multilivello

Dopo aver analizzato il comportamento delle variabili esplicative su una variabile risposta di tipo binario, decido di stimare un modello di regressione lineare quando la variabile dipendente è di tipo continuo. Mi focalizzo quindi solo sul sottogruppo di assistiti che hanno avuto una spesa positiva (per ricoveri, visite specialistiche o farmaci), perciò i modelli che presenterò avranno numerosità diverse. Una persona, per esempio, può essere stata ricoverata nel 2007 ma non aver effettuato nessuna spesa per visite specialistiche in quel determinato anno. Certamente le persone avranno spese diverse, lo scopo è individuare il profilo delle persone che comportano una spesa più onerosa a carico della Sanità pubblica; inoltre verranno stimati dei modelli multilivello per cercare di capire se anche le caratteristiche del medico influenzano tale spesa o meno.

In modo analogo a quanto visto nel Paragrafo 3.1, le Tabelle 12, 13, 14 riportano le stime del modello di regressione lineare e multilivello, con e senza covariate del medico, della spesa per ricoveri, visite specialistiche e farmaci rispettivamente. La variabile dipendente viene qui specificata per mezzo di una trasformazione logaritmica. Per questo motivo, l'interpretazione delle stime dei parametri dovranno avvenire in termini di semi-elasticità.

3.2.1 Ricoveri ospedalieri

Log tariffa ricovero		Regressione				Regressione con covariate del medico			
Variabili esplicative		Semplice (1)		Multilivello (2)		Semplice (3)		Multilivello (4)	
		<i>Coeff</i>	<i>P-value</i>	<i>Coeff</i>	<i>P-value</i>	<i>Coeff</i>	<i>P-value</i>	<i>Coeff</i>	<i>P-value</i>
Sesso	Femmina	-0,058	0,000	-0,059	0,000	-0,058	0,000	-0,058	0,000
Età	25-34	0,116	0,000	0,116	0,000	0,116	0,000	0,116	0,000
	35-44	0,118	0,000	0,119	0,000	0,117	0,000	0,119	0,000
	45-54	0,239	0,000	0,241	0,000	0,239	0,000	0,241	0,000
	55-64	0,356	0,000	0,359	0,000	0,357	0,000	0,359	0,000
	65-74	0,482	0,000	0,484	0,000	0,482	0,000	0,484	0,000
	≥75	0,509	0,000	0,512	0,000	0,509	0,000	0,512	0,000
Decesso	2007	0,341	0,000	0,342	0,000	0,341	0,000	0,342	0,000
	2008	0,314	0,000	0,314	0,000	0,313	0,000	0,314	0,000
ADI	Ambulatorio	0,138	0,000	0,140	0,000	0,137	0,000	0,139	0,000
	Domicilio	0,425	0,000	0,428	0,000	0,426	0,000	0,428	0,000
Esenzioni	Affezioni sistema circolatorio	0,114	0,000	0,112	0,000	0,114	0,000	0,112	0,000
	Malattie cardiache	0,309	0,000	0,310	0,000	0,309	0,000	0,310	0,000
	Diabete	0,107	0,000	0,107	0,000	0,107	0,000	0,107	0,000
	Glaucoma	-0,069	0,000	-0,067	0,000	-0,069	0,000	-0,067	0,000
	Colesterolo	0,014	0,403	0,013	0,423	0,014	0,376	0,014	0,405
	Ipertensione arteriosa con danno d'organo	0,030	0,000	0,032	0,000	0,030	0,000	0,032	0,000
	Ipertensione arteriosa senza danno d'organo	-0,012	0,258	-0,010	0,311	-0,012	0,250	-0,011	0,302
	Malattie neoplastiche maligne	0,272	0,000	0,272	0,000	0,272	0,000	0,272	0,000
	Invalidità civile > 2/3	0,064	0,000	0,063	0,000	0,064	0,000	0,063	0,000
	Invalidità civile (assegno accomp.)	-0,075	0,000	-0,078	0,000	-0,076	0,000	-0,079	0,000
	Invalidità civile al 100%	0,076	0,000	0,075	0,000	0,076	0,000	0,075	0,000
	Gravidanza	0,057	0,000	0,062	0,000	0,057	0,000	0,062	0,000
	Altro	0,097	0,000	0,097	0,000	0,097	0,000	0,097	0,000
Costante		7,718	0,000	7,715	0,000	7,714	0,000	7,712	0,000
Sesso medico	Maschio					0,018	0,001	0,017	0,013
	Supera max assistiti					-0,005	0,290	-0,005	0,422
	Anni dalla laurea					0,103	0,677	0,101	0,737
	Anni dalla laurea^2					-0,050	0,301	-0,049	0,405
σ^2_1				0,763				0,763	
σ^2_2				0,003				0,003	
ICC				0,42%				0,40%	

Tabella 12 - Stime dei modelli regressione semplice e multilivello, con o senza covariate del medico, dei ricoveri ospedalieri

La Tabella 12 mostra le stime dei parametri relative al campione ristretto di 134.629 assistiti, ovvero solo le persone che sono state ricoverate nel 2007.

Sicuramente il segno delle stime del modello di regressione potrà essere diverso da quello del modello probit, appunto perché il modello probit comprende tutto il campione, mentre il modello di regressione solo il sottocampione di persone che hanno avuto una spesa per ricoveri positiva.

Per esempio, il modello probit standard mostra che le donne hanno una maggior probabilità di aver effettuato un ricovero ospedaliero rispetto agli uomini, a parità di altre condizioni. Il modello di regressione, invece, mostra che essere donna porta ad una diminuzione della spesa per ricoveri del 5.8% rispetto agli uomini, *ceteris paribus*. Questo significa che le donne hanno tendenzialmente maggiore probabilità di essere ricoverate, mentre gli uomini gravano di più in termini di costi sul sistema sanitario. Per cercare di capire come mai gli uomini costino di più, faccio riferimento ai DRG (Diagnosis Related Groups), già trattati nelle analisi esplorative. Essi sono un sistema di classificazione dei ricoveri ospedalieri costruiti a partire dalle informazioni contenute nella scheda di accettazione/dimissione ospedaliera. I DRG definiscono delle categorie di ricoveri clinicamente significative ed omogenee al loro interno quanto a consumo di risorse assistenziali e quindi anche rispetto ai costi di produzione dell'assistenza. A differenza delle donne, troviamo che la maggior parte dei ricoveri per interventi sul ginocchio senza diagnosi principale di infezione sono attribuibili agli uomini (circa il 70%, codice drg 503). Questo tipo di ricovero costa, in media, alla Sanità € 2.578 per ogni singolo caso. Inoltre, circa il 94% dei ricoveri ospedalieri relativi agli interventi per ernia inguinale e femorale (codice drg 162) è stato effettuato dagli uomini. Questo ricovero costa mediamente alla Sanità € 2.336 per ogni assistito. Questi due codici drg non fanno parte dei ricoveri ospedalieri più costosi, però si riferiscono agli interventi maggiormente attribuibili agli uomini e non alle donne. In generale, gli altri codici drg si dividono equamente tra i due sessi. Le dummy che riguardano l'età sono tutte significative e notiamo che la magnitudo delle stime cresce con il crescere dell'età. Gli assistiti con un'età tra i 25 e i 34 anni hanno una spesa per ricoveri dell' 11.6% maggiore rispetto agli assistiti di fascia d'età 14-24, *ceteris paribus*. Una grandezza simile viene riscontrata anche per gli assistiti di fascia d'età 35-44, i quali hanno una spesa per ricoveri dell' 11.8% maggiore, rispetto agli assistiti di fascia d'età 14-24. Maggiore è, invece, la differenza di spesa per quanto riguarda le persone tra i 45 e 54 anni. Come detto in precedenza, la grandezza delle stime aumenta con il crescere dell'età, infatti gli

assistiti che fanno parte della fascia d'età 55-64, che sono stati ricoverati nel 2007 in Friuli-Venezia Giulia, hanno una spesa per ricoveri del 35.6% più grande rispetto agli assistiti ricoverati di fascia d'età 14-24. Entrambe le dummy età 65-74 ed età 75+, gravano sul sistema sanitario con una spesa per ricoveri circa il 50% maggiore rispetto agli assistiti dai 14 ai 24 anni, *ceteris paribus*. Queste stime non mi sorprendono, è più probabile che una persona anziana costi di più alla Sanità rispetto a una persona giovane, proprio perché certe malattie, patologie o problematiche subentrano solo con il passare degli anni. In generale le persone giovani sono più in salute rispetto agli anziani. Le stime delle variabili deceduto nel 2007 e deceduto nel 2008 sono entrambe significative e di segno positivo, come nel modello probit standard. Le persone decedute nel 2007 che sono state ricoverate nel medesimo anno, sono caratterizzate da una spesa per ricoveri del 34% maggiore rispetto a chi non è deceduto nel 2007, *ceteris paribus*. Anche le persone decedute nel 2008 hanno una spesa per ricoveri maggiore rispetto a chi non è deceduto. Questa spesa si aggira sul 31% in più rispetto alla spesa di chi è stato ricoverato ma non è deceduto nel 2008. Nemmeno questo risultato mi sorprende, è ipotizzabile che le persone nei loro ultimi anni di vita abbiano bisogno di cure assistenziali maggiori o di essere ricoverati d'urgenza e per periodi lunghi. Questo comporta delle spese più importanti rispetto a chi è in salute o che comunque non presenta malattie gravi tali da portare alla morte. Le dummy relative al servizio di assistenza domiciliare integrata sono entrambe significative e di segno positivo. Un aspetto interessante è la differenza di magnitudo delle stime. Gli assistiti che hanno usufruito del servizio ADI in ambulatorio hanno una spesa per ricoveri del 13,8% maggiore rispetto a chi non ha sfruttato tale servizio, a parità di altre condizioni. Gli assistiti che hanno usufruito del servizio ADI a domicilio, invece, hanno avuto una spesa per ricoveri del 30,9% maggiore rispetto a chi non ha avuto tale servizio, *ceteris paribus*. E' ipotizzabile che le persone che hanno richiesto il servizio ADI a domicilio siano persone che non riuscendo a muoversi (persone con fratture, anziani con gravi problemi di deambulazione, etc.) presentino una gravità maggiore rispetto a chi riesce a usufruire di tale servizio in ambulatorio (per lo più medicazioni, terapie iniettabili). Le dummy che si riferiscono alle esenzioni non sono tutte significative e alcune presentano segni diversi dal modello probit standard. Innanzitutto, le variabili che si riferiscono alle esenzioni per colesterolo e per ipertensione senza danno d'organo non hanno nessun valore predittivo sulla variabile risposta tariffa ricovero, esse non sono significative. Tra tutte le persone che sono state ricoverate nel 2007 e che hanno un'esenzione, quelle che più gravano sul servizio sanitario sono gli esenti per malattie cardiache e gli esenti per malattie

neoplastiche maligne. Essi presentano una spesa per ricoveri ospedalieri del 30% maggiore rispetto a chi non ha alcuna esenzione, *ceteris paribus*. In seguito troviamo gli esenti per diabete, i quali hanno una spesa per ricoveri del 10.7% maggiore rispetto a chi non ha nessuna esenzione. Gli esenti per ipertensione con danno d'organo, invalidità civile 2/3, invalidità civile al 100% e per gravidanza presentano una spesa per ricoveri sicuramente maggiore rispetto a chi non ha nessuna esenzione, però tale percentuale di spesa è minore dell' 8%. Le uniche esenzioni che hanno segno opposto rispetto al modello probit sono: esenzione per glaucoma e per invalidità civile con assegno di accompagnamento. Gli assistiti che hanno una di queste esenzioni hanno una spesa per ricoveri ospedalieri del 7% più piccola rispetto a chi non ha nessuna esenzione, *ceteris paribus*. Si può ipotizzare che le persone con queste esenzioni in realtà non abbiano bisogno di ricoveri ospedalieri d'urgenza o costosi, ma magari di visite specialistiche.

Un indice di adattamento ai dati è l' R^2 , il quale misura la percentuale di varianza della variabile dipendente spiegata dal modello. In questo caso l' R^2 è del 16.2% e ciò significa che il 16.2% della variabilità dei dati è spiegata dalle variabili esplicative. Il modello è caratterizzato dalle sole osservazioni con una spesa per ricoveri positiva perciò il campione è ristretto (anche se la numerosità rimane comunque elevata) e questo può portare a un indice di bontà non molto alto. Il test F porta a rifiutare l'ipotesi nulla di uguaglianza di tutti i coefficienti a zero a parte la costante. Coerentemente con le analisi del modello Probit standard, il ricovero ospedaliero può essere spiegato principalmente solo da caratteristiche individuali degli assistiti di quelle disponibili in questo dataset.

Il comportamento dei coefficienti dell'estensione multilivello del modello di regressione senza covariate del medico, per quanto riguarda i ricoveri ospedalieri, è simile al modello non multilivello. Troviamo la stessa significatività delle variabili e persino la magnitudo delle stime è pressoché uguale. L'ICC si aggira attorno allo 0.42% ed è molto basso. Posso concludere, quindi, che non ci sia un effetto importante della gerarchia dei dati.

Passando al modello di regressione con covariate del medico (colonna 3) si nota che le stime delle caratteristiche degli assistiti presentano la stessa significatività e la stessa grandezza del modello senza covariate del medico. Pongo l'attenzione, quindi, sulle sole caratteristiche del medico di medicina generale. L'unica variabile significativa è il genere del medico, anche se la magnitudo è davvero piccola. Gli assistiti che hanno un medico maschio hanno avuto una spesa per ricoveri maggiore dell'1.8% rispetto a chi ha un medico donna, *ceteris paribus*. Le

variabili dell'esperienza (anni dalla laurea e anni dalla laurea al quadrato) e la dummy "supera massimo assistiti" non sono significative perciò non hanno nessun valore predittivo.

L'indice di bontà di adattamento al modello R^2 è lo stesso del modello senza covariate del medico (16.2%). Il test F porta al rifiuto dell'ipotesi nulla.

L'estensione multilivello del modello di regressione con covariate del medico (colonna 4) si comporta esattamente come il modello precedente. C'è da sottolineare, però, che in questo modello il sesso del medico è significativo al 5% ma non all'1%.

L'ICC di tale modello è leggermente più basso rispetto al modello senza covariate del medico e si aggira attorno allo 0.40%. E' un valore molto basso, che indica una variabilità tra medici di fatto assente (un valore dell'ICC appena maggiore di 0 è nella pratica piuttosto fisiologico).

In conclusione, per quanto riguarda i ricoveri ospedalieri, considerando i risultati sia dei modelli che indicano la presenza di un ricovero, sia dei modelli sulla spesa sostenuta, nelle prossime stime dei modelli si terrà conto solo delle caratteristiche degli assistiti, tralasciando le covariate dei MMG.

3.2.2 Visite specialistiche

Log-tariffa specialistiche		Regressione				Regressione con covariate del medico			
		Semplice (1)		Multilivello (2)		Semplice (3)		Multilivello (4)	
Variabili esplicative		Coeff	P-value	Coeff	P-value	Coeff	P-value	Coeff	P-value
Sesso	Femmina	0,100	0,000	0,095	0,000	0,097	0,000	0,095	0,000
Età	25-34	0,058	0,000	0,066	0,000	0,064	0,000	0,066	0,000
	35-44	0,174	0,000	0,176	0,000	0,178	0,000	0,176	0,000
	45-54	0,284	0,000	0,286	0,000	0,286	0,000	0,286	0,000
	55-64	0,370	0,000	0,376	0,000	0,376	0,000	0,377	0,000
	65-74	0,601	0,000	0,605	0,000	0,608	0,000	0,605	0,000
	≥75	0,505	0,000	0,504	0,000	0,511	0,000	0,504	0,000
Decesso	2007	-0,212	0,000	-0,203	0,000	-0,211	0,000	-0,204	0,000
	2008	0,137	0,000	0,146	0,000	0,138	0,000	0,145	0,000
ADI	Ambulatorio	0,094	0,000	0,151	0,000	0,094	0,000	0,150	0,000
	Domicilio	-0,044	0,000	-0,033	0,000	-0,043	0,000	-0,033	0,000
Esenzioni	Affezioni sistema circolatorio	0,227	0,000	0,219	0,000	0,227	0,000	0,219	0,000
	Malattie cardiache	0,293	0,000	0,305	0,000	0,292	0,000	0,305	0,000
	Diabete	0,363	0,000	0,361	0,000	0,364	0,000	0,361	0,000
	Glaucoma	0,228	0,000	0,220	0,000	0,226	0,000	0,220	0,000
	Colesterolo	0,218	0,000	0,189	0,000	0,218	0,000	0,189	0,000
	Ipertensione arteriosa con danno d'organo	0,167	0,000	0,163	0,000	0,165	0,000	0,162	0,000
	Ipertensione arteriosa senza danno d'organo	0,088	0,000	0,090	0,000	0,086	0,000	0,090	0,000
	Malattie neoplastiche maligne	0,964	0,000	0,958	0,000	0,964	0,000	0,958	0,000
	Invalidità civile > 2/3	0,330	0,000	0,323	0,000	0,330	0,000	0,323	0,000
	Invalidità civile (assegno accomp.)	-0,309	0,000	-0,300	0,000	-0,309	0,000	-0,301	0,000
	Invalidità civile al 100%	0,196	0,000	0,194	0,000	0,196	0,000	0,193	0,000
	Gravidanza	1,289	0,000	1,318	0,000	1,294	0,000	1,317	0,000
Altro	0,501	0,000	0,491	0,000	0,501	0,000	0,491	0,000	
	Costante	4,140	0,000	4,109	0,000	3,471	0,000	3,409	0,000
Sesso medico	Maschio					-0,044	0,000	-0,051	0,000
	Supera max assistiti					0,040	0,000	0,050	0,001
	Anni dalla laurea					5,730	0,000	6,169	0,000
	Anni dalla laurea^2					-1,141	0,000	-1,228	0,000
	σ^2_1			1,384				1,384	
	σ^2_2			0,109				0,043	
	ICC			7,29%				3,01%	

Tabella 13 - Stime dei modelli regressione semplice e multilivello, con o senza covariate del medico, delle visite specialistiche

La Tabella 13 evidenzia le stime dei parametri del modello di regressione semplice e multilivello per quanto riguarda le spese per visite specialistiche effettuate in Friuli-Venezia Giulia nel 2007. Rispetto al modello probit, che racchiudeva tutta la popolazione regionale, i modelli di regressione in questione si baseranno sul sottocampione di 722.128 assistiti, che comprende solo quelle persone che hanno avuto una spesa positiva per visite specialistiche. A tal proposito, si può notare che le stime del modello di regressione sono tutte significative ed alcune hanno segno opposto rispetto al modello probit (Tabella 10).

La prima colonna della Tabella 13 contiene le stime del modello di regressione semplice senza covariate del medico. Essere donna porta ad avere una spesa per visite specialistiche del 10% maggiore rispetto agli uomini, a parità di altre condizioni. Le stime delle dummy relative all'età hanno un andamento che si può definire parabolico. La grandezza di tali stime è direttamente proporzionale al crescere dell'età fino all'età di 75 anni circa, per poi diminuire. Le persone tra i 25 e i 34 anni hanno una spesa per visite specialistiche del 5% maggiore rispetto agli assistiti di età compresa tra i 14 e 24 anni. Le persone con un'età compresa tra i 35 e 44 anni costano (in termini di visite specialistiche) il 17% in più rispetto alla fascia d'età 14-24, *ceteris paribus*. Gli assistiti tra i 45 e 54 anni e tra i 55 e 64 anni, hanno una spesa per visite specialistiche maggiore rispetto alla categoria di riferimento 14-24, rispettivamente del 28% e del 37%. Una grandezza maggiore si riscontra per gli assistiti di età compresa tra i 65 e 74 anni. Questi ultimi hanno una spesa per visite specialistiche del 60% più grande rispetto alle persone tra i 14 e 24 anni, *ceteris paribus*. Gli assistiti compresi nella fascia d'età 75+ costano il 50% in più rispetto alla categoria di riferimento 14-24 ma si può notare che la magnitudo è minore rispetto alla fascia d'età 65-74. E' ipotizzabile che con il crescere dell'età subentrino maggiori problemi di salute, perciò c'è l'esigenza di trovare la causa di queste complicazioni attraverso la prescrizione di visite specialistiche. Una persona anziana di età dai 75 anni in su effettua sicuramente più visite specialistiche di un giovane ragazzo di età compresa tra i 14 e 24 anni, ma, probabilmente, tende a farne progressivamente meno. Presumibilmente, certi sintomi e malesseri compaiono solo fino ad una certa età oppure si concentrano maggiormente nella fascia d'età tra i 65 e 74 anni. Chi è deceduto nel 2007 è caratterizzato da una spesa per visite specialistiche del 21% più piccola rispetto a chi non è deceduto, *ceteris paribus*. E' ipotizzabile, quindi, che negli ultimissimi mesi di vita di una persona si eseguano poche visite specialistiche, magari le visite che vengono fatte sono comprese nel costo per ricoveri ospedalieri (costo che comprende qualsiasi attività svolta nel periodo di degenza dell'assistito). Chi, invece, è deceduto nel 2008, ha una spesa per visite

specialistiche del 13.7% maggiore rispetto a chi non è deceduto nel 2008, a parità di altre condizioni. Passando alle dummy del servizio di assistenza domiciliare integrata si nota che gli assistiti che hanno usufruito di tale servizio in ambulatorio hanno una spesa per visite specialistiche del 9.4% maggiore rispetto a chi non ha usufruito di tale servizio, *ceteris paribus*. Coloro che hanno usufruito del servizio a domicilio, invece, hanno una spesa per visite specialistiche del 4.3% minore rispetto a chi non sfrutta il servizio. Ciò può essere spiegato, in parte, dal fatto che gli assistiti ai quali è indirizzato il servizio a domicilio, in realtà, sono seguiti regolarmente da personale infermieristico, perciò sono consigliati e rassicurati per qualsiasi dubbio circa la loro salute in modo tale da ovviare visite specialistiche magari superflue. Le stime delle dummy che riguardano le esenzioni sono tutte significative e tutte positive, eccetto l'esenzione per invalidità civile con accompagnamento. Gli assistiti con questa esenzione hanno una spesa per visite specialistiche del 30% più piccola rispetto a chi non ha tale esenzione, *ceteris paribus*. Una possibile spiegazione si collega al discorso precedente del servizio ADI a domicilio. Coloro che hanno un assegno di accompagnamento sono per lo più persone non in grado di deambulare da sole e hanno bisogno di assistenza continua. Questo porta ad ipotizzare che è meno probabile che tali persone si spostino per andare a fare qualche visita specialistica (cosa evidenziata anche nel modello Probit, Tabella 10) ma, magari, sfruttino maggiormente il servizio ADI. Passando alle esenzioni che hanno una stima positiva per visite specialistiche, la più onerosa per il Servizio Sanitario è l'esenzione per gravidanza. Gli assistiti con tale esenzione hanno una spesa per visite specialistiche circa del 12,8% maggiore rispetto a chi non ha l'esenzione, *ceteris paribus*. Questo risultato non è sorprendente, una donna in stato di maternità è soggetta a molte visite di controllo per un lungo arco di tempo, cioè 9 mesi. Un'altra esenzione che porta a spese ingenti a carico della Sanità pubblica è l'esenzione per malattie neoplastiche maligne. Gli assistiti che hanno questa esenzione hanno una spesa per visite specialistiche del 96% maggiore rispetto a chi non ha l'esenzione, *ceteris paribus*. Sono circa 60.000 gli assistiti con questa esenzione e, purtroppo, alcune patologie necessitano di cure o visite molto costose che incrementano molto i costi alla Sanità. Altre due esenzioni tra le più ingenti in termini di costi sono l'esenzione per diabete e l'esenzione per invalidità civile >2/3. Queste ultime portano ad avere una spesa per visite specialistiche rispettivamente del 36% e del 33% maggiore rispetto a chi non ha tali esenzioni, *ceteris paribus*. Le esenzioni per diabete (51.000 casi) sono circa il doppio delle esenzioni per invalidità civile >2/3 (23.000 casi) ma il costo a carico della Sanità non è molto differente. Si può ipotizzare che gli assistiti invalidi al 75% sviluppino una serie di complicanze nel tempo,

perciò devono svolgere varie visite e, magari, chi ha l'esenzione per diabete svolge per lo più visite legate alla patologia stessa. Gli assistiti caratterizzati dall'esenzione per malattie cardiache hanno una spesa per visite specialistiche del 29% maggiore rispetto a chi non ha l'esenzione, *ceteris paribus*. I pazienti con esenzione da affezioni del sistema circolatorio hanno un costo per visite specialistiche che grava sul Sistema Sanitario il 23% in più rispetto a chi non ha l'esenzione, *ceteris paribus*. In ordine decrescente di costo troviamo poi l'esenzione per glaucoma, l'esenzione per colesterolo e l'esenzione per invalidità civile al 100%. Queste ultime hanno una spesa per ricoveri rispettivamente del 23%, 22% e 20% maggiore rispetto a chi non ha tali esenzioni, *ceteris paribus*. Una spesa per visite specialistiche minore si riscontra nelle esenzioni per ipertensione arteriosa con danno d'organo e senza danno d'organo, le quali costano alla Sanità rispettivamente il 16% e l'8% in più rispetto a chi non ha tali esenzioni, *ceteris paribus*.

L'indice di bontà di adattamento al modello R^2 è del 13.7%. Nonostante la numerosità del campione sia elevata, l'indice non è molto alto. Il test F porta al rifiuto dell'ipotesi nulla.

La seconda colonna della Tabella 13 presenta le stime del modello di regressione multilivello. Si può notare che la grandezza delle stime della prima e della seconda colonna della Tabella 13 non sono molto diverse, sono tutte significative ed hanno lo stesso segno. L'indice ICC è del 7.3%, molto più alto rispetto al modello di regressione multilivello dei ricoveri ospedalieri (Tabella 12). Questo può significare che è presente una certa variabilità tra medici per quanto riguarda la spesa per visite specialistiche, il comportamento del singolo assistito è in parte influenzato dal comportamento del gruppo di appartenenza (MMG).

Introduco al modello di regressione le caratteristiche del medico di medicina generale in modo da catturare l'effetto che queste variabili possono avere sulla spesa per visite specialistiche (Tabella 13, terza colonna). Le stime delle caratteristiche degli assistiti sono tutte significative e la loro grandezza è molto simile al modello di regressione senza covariate del medico. Mi focalizzo, quindi, sulle sole variabili delle caratteristiche del MMG. Gli assistiti a carico di un medico maschio hanno una spesa per visite specialistiche del 4.3% minore rispetto agli assistiti associati a un medico donna, *ceteris paribus*. Sulla base di modelli non multilivello, si evince che i medici donna tendono sia a prescrivere più visite specialistiche, che a far sostenere spese maggiori. I medici che hanno superato la soglia massima di assistiti causano una spesa per visite specialistiche del 3.9% maggiore rispetto a chi non ha superato la soglia massima, *ceteris paribus*. Le variabili che indicano l'esperienza del medico, "anni dalla

laurea”, sono entrambe significative e di segno opposto. Nello specifico, con il passare degli anni i medici tendono a prescrivere visite specialistiche più onerose ma, raggiunta una certa esperienza, questa tendenza si inverte. A parità di altre condizioni, la stima dell’esperienza del medico segue un andamento parabolico: i medici più anziani hanno delle conoscenze maggiori e tendono a prescrivere visite specialistiche meno costose rispetto ai medici alle prime armi, oppure i medici più giovani tendono ad adottare comportamenti diversi per conoscere meglio gli assistiti che stanno seguendo.

Aggiungendo le covariate del Medico di Medicina Generale, l’indice di bontà di adattamento ai dati R^2 è leggermente aumentato rispetto al modello senza covariate e si aggira sul 14%. Le variabili del MMG hanno un valore predittivo significativo, aiutano a spiegare meglio il modello di regressione della spesa per visite specialistiche. Il test F porta al rifiuto dell’ipotesi nulla.

Passando all’estensione multilivello del modello di regressione con covariate del medico (Tabella 13, colonna 4) si può notare come le stime di tutte le variabili siano significative e molto simili ai modelli precedenti. Degno di nota è il Coefficiente di Correlazione Intraclasse (ICC), il quale è molto più basso rispetto al modello di regressione multilivello senza covariate del medico e si aggira attorno al 3%. Introducendo le covariate di secondo livello è diminuita la variabilità tra i gruppi (variabilità between): tali variabili sono in grado di spiegare la variabilità tra i MMG insita nei dati. Il comportamento degli assistiti è influenzato dal gruppo di appartenenza, in questo caso le caratteristiche dei medici influenzano la spesa per visite specialistiche dei loro pazienti.

3.2.3 Farmaci

Log-importo farmaci		Regressione				Regressione con covariate del medico			
		Semplice (1)		Multilivello (2)		Semplice (3)		Multilivello (4)	
Variabili esplicative		Coeff	P-value	Coeff	P-value	Coeff	P-value	Coeff	P-value
Sesso	Femmina	-0,048	0,000	-0,048	0,000	-0,049	0,000	-0,049	0,000
Età	25-34	0,147	0,000	0,150	0,000	0,150	0,000	0,150	0,000
	35-44	0,331	0,000	0,330	0,000	0,333	0,000	0,330	0,000
	45-54	0,682	0,000	0,678	0,000	0,681	0,000	0,678	0,000
	55-64	1,146	0,000	1,145	0,000	1,148	0,000	1,146	0,000
	65-74	1,569	0,000	1,567	0,000	1,571	0,000	1,568	0,000
	≥75	1,764	0,000	1,761	0,000	1,766	0,000	1,761	0,000
Decesso	2007	-0,529	0,000	-0,524	0,000	-0,529	0,000	-0,524	0,000
	2008	0,001	0,916	0,004	0,709	0,002	0,879	0,004	0,709
ADI	Ambulatorio	0,147	0,000	0,214	0,000	0,146	0,000	0,214	0,000
	Domicilio	0,290	0,000	0,294	0,000	0,292	0,000	0,294	0,000
Esenzioni	Affezioni sistema circolatorio	0,413	0,000	0,400	0,000	0,413	0,000	0,400	0,000
	Malattie cardiache	0,561	0,000	0,573	0,000	0,560	0,000	0,573	0,000
	Diabete	0,759	0,000	0,758	0,000	0,760	0,000	0,758	0,000
	Glaucoma	0,569	0,000	0,567	0,000	0,568	0,000	0,566	0,000
	Colesterolo	0,779	0,000	0,780	0,000	0,780	0,000	0,780	0,000
	Ipertensione arteriosa con danno d'organo	0,844	0,000	0,850	0,000	0,843	0,000	0,850	0,000
	Ipertensione arteriosa senza danno d'organo	0,813	0,000	0,820	0,000	0,812	0,000	0,819	0,000
	Malattie neoplastiche maligne	0,354	0,000	0,352	0,000	0,354	0,000	0,352	0,000
	Invalidità civile > 2/3	0,482	0,000	0,479	0,000	0,481	0,000	0,479	0,000
	Invalidità civile (assegno accomp.)	0,072	0,000	0,081	0,000	0,071	0,000	0,080	0,000
	Invalidità civile al 100%	0,364	0,000	0,366	0,000	0,363	0,000	0,365	0,000
	Gravidanza	-0,387	0,000	-0,380	0,000	-0,383	0,000	-0,380	0,000
	Altro	0,536	0,000	0,535	0,000	0,536	0,000	0,535	0,000
	Costante		3,375	0,000	3,307	0,000	2,886	0,000	2,837
Sesso medico	Maschio					-0,002	0,655	-0,007	0,433
	Supera max assistiti					0,014	0,000	0,070	0,000
	Anni dalla laurea					4,108	0,000	4,436	0,000
	Anni dalla laurea ²					-0,826	0,000	-0,852	0,000
σ^2_1			1,720				1,720		
σ^2_2			0,121				0,099		
ICC			6,58%				5,46%		

Tabella 14 - Stime dei modelli regressione semplice e multilivello, con o senza covariate del medico, della spesa per farmaci

La Tabella 14 evidenzia le stime del modello di regressione semplice e multilivello della spesa per farmaci in Friuli-Venezia Giulia, anno 2007. Il campione di riferimento è di 738.935 persone e comprende solo gli assistiti che hanno avuto una spesa per farmaci positiva.

La prima colonna della tabella racchiude le stime delle caratteristiche degli assistiti ottenute tramite il modello di regressione lineare. La stima della variabile “Femmina” è significativa e di segno negativo. Nonostante il modello Probit (prima colonna, Tabella 11) evidenziasse che le donne hanno maggiore probabilità di acquistare farmaci, la regressione lineare, invece, evidenzia che, tenendo conto delle sole persone che hanno acquistato farmaci, le donne spendono il 4,7% in meno rispetto agli uomini, *ceteris paribus*. Le dummy relative all’età indicano che la spesa per farmaci aumenta con il crescere dell’età. Nel dettaglio, gli assistiti di età compresa tra i 25 e 34 anni hanno una spesa per farmaci del 14.6% maggiore rispetto agli assistiti compresi nella fascia d’età 14-24, a parità di altre condizioni. Coloro che hanno un’età compresa tra i 25 e 34 anni spendono in farmaci il 33% in più rispetto alle persone che acquistano farmaci e hanno un’età compresa tra i 14 e 24 anni. Gli assistiti tra i 35 e 44 anni hanno una spesa per farmaci del 68% maggiore rispetto alla categoria di riferimento 14-24, *ceteris paribus*. Le successive fasce d’età hanno una spesa per farmaci molto onerosa per la Sanità. Infatti, gli assistiti con un’età compresa tra i 55 e 64 anni, tra i 65 e 74 anni e 75+ anni hanno una spesa per farmaci rispettivamente del 114%, 156% e 176% circa maggiore rispetto alla fascia d’età 14-24, *ceteris paribus*. Anche i modelli precedenti, siano spese per ricoveri o spese per visite specialistiche, evidenziavano che al crescere dell’età aumentavano i costi a carico della Sanità. E’ plausibile che più passino gli anni, più una persona abbia dei problemi di salute (le persone giovani costano molto meno alla Sanità pubblica rispetto agli anziani). Gli assistiti deceduti nel 2007 hanno una spesa per farmaci del 53% più piccola rispetto a chi non è deceduto nel 2007, *ceteris paribus*. Le persone possono essere decedute nei primi mesi dell’anno come negli ultimi, perciò, nel complesso, c’è meno probabilità che abbiano acquistato farmaci (cosa evidenziata anche dal modello Probit, prima colonna Tabella 11). La variabile deceduto nel 2008 non è significativa al 5% e ciò sta a indicare che non ha nessun valore predittivo nello spiegare la variabile risposta. Le dummy riguardanti il servizio di assistenza domiciliare integrata sono significative e di segno positivo. Coloro che hanno usufruito del servizio in ambulatorio hanno una spesa per farmaci 14% maggiore rispetto a chi non ha sfruttato tale servizio, *ceteris paribus*. Una percentuale di spesa maggiore è attribuibile al servizio ADI a domicilio, coloro che hanno usufruito di tale servizio hanno una spesa per farmaci quasi del 30% maggiore rispetto a chi non sfrutta il servizio, *ceteris paribus*. In

generale, le persone che vengono assistite a domicilio sono in condizioni peggiori rispetto a chi riesce a muoversi e andare in ambulatorio, ed è per questo che è probabile che abbiano bisogno di più cure e farmaci rispetto agli altri. Le dummy che si riferiscono alle esenzioni sono tutte significative e di segno positivo, eccetto l'esenzione per gravidanza che ha segno negativo. Le persone con questa esenzione spendono in farmaci circa il 38% in meno rispetto a chi non ha l'esenzione, *ceteris paribus*. Le donne in stato di maternità, in realtà, devono fare molta attenzione circa l'assunzione di farmaci poiché possono provocare complicazioni al feto. Per esempio, alcuni tipi di farmaci, come i FANS (farmaci antinfiammatori non steroidei), sono da evitare categoricamente nel terzo trimestre di gravidanza, poiché potrebbero causare la chiusura del dotto di Botallo, un importante vaso sanguigno indispensabile per la corretta circolazione fetale⁹. Per quanto riguarda le esenzioni per patologia, le più onerose in termini di spesa per farmaci sono: l'esenzione per ipertensione arteriosa con danno d'organo, senza danno d'organo, per colesterolo e per diabete. Esse hanno una spesa per farmaci rispettivamente dell' 84%, 81%, 77% e 75% circa maggiore rispetto a chi non ha tali esenzioni, *ceteris paribus*. Queste quattro esenzioni sono quelle maggiormente diffuse nella Regione, eccetto l'esenzione per colesterolo che è stata concessa a quasi 19.000 persone contro le 76.212 dell'esenzione per ipertensione arteriosa con danno d'organo, le 45.479 dell'esenzione per ipertensione arteriosa senza danno d'organo e le 51.033 persone relative all'esenzione per diabete. Da questi dati si può ipotizzare che, in realtà, gli assistiti con problemi di colesterolo devono assumere molti farmaci oppure farmaci molto costosi rispetto agli altri tipi di patologie. In ordine decrescente di grandezza si trova l'esenzione per: glaucoma, malattie cardiache, affezioni sistema circolatorio e malattie neoplastiche maligne. Queste esenzioni hanno una spesa per farmaci rispettivamente del 57%, 56%, 41% e 35% maggiore rispetto a chi non possiede tali esenzioni, a parità di altre condizioni. Passando alle esenzioni per invalidità civile, troviamo che la più ingente a carico del SSN è l'esenzione per invalidità civile >2/3, che costa in termini di spesa per farmaci il 48% in più rispetto a chi non ha tale esenzione. Gli invalidi civili al 100%, invece, hanno una spesa per farmaci del 36% maggiore rispetto a chi non ha questa esenzione, *ceteris paribus*. Gli invalidi civili con assegno di accompagnamento, invece, costano alla Sanità, in termini di farmaci, il 7,2% in più rispetto a chi non ha l'assegno di accompagnamento, *ceteris paribus*. Questa percentuale di spesa per farmaci è molto diversa rispetto alle altre due invalidità civili, probabilmente perché gli invalidi con assegno di accompagnamento, in realtà, non sono persone affette da minoranza

⁹ <http://www.agenziafarmaco.gov.it>

fisica, psichica o sensoriale. L'assegno di accompagnamento è concesso alle persone con più di 65 anni che non sono in grado di deambulare da sole, che presentano difficoltà persistenti a svolgere i compiti e le funzioni proprie della loro età. Per questo motivo, queste persone non hanno bisogno, necessariamente, di cure farmacologiche specifiche poiché non hanno delle patologie particolari, o almeno questo tipo di esenzione non lo prevede.

L'indice di bontà di adattamento R^2 è del 35.7% e, tenendo conto della numerosità campionaria, si può definire un buon indice. Il modello si adatta bene a spiegare i dati. Il test F di uguaglianza di tutti i coefficienti a 0 eccetto l'intercetta porta al rifiuto dell'ipotesi nulla.

Passando alla regressione multilivello (seconda colonna) si noti come le stime del modello siano pressoché identiche rispetto al modello non multilivello. Anche la significatività delle variabili è la stessa. L'unica variabile che ha una grandezza diversa è il servizio ADI ambulatorio: le persone che hanno sfruttato tale servizio hanno una spesa per farmaci del 21% maggiore rispetto a chi non ha usufruito del servizio, *ceteris paribus*. Nel modello non multilivello la stima si aggirava sul 14%. Considerare l'aspetto gerarchico dei dati ha fatto emergere tale differenza di comportamento degli assistiti. Appartenere a un certo gruppo (il MMG) influenza il paziente, in questo caso la spesa per farmaci di coloro che sono stati assistiti in ambulatorio è aumentata. Il coefficiente di correlazione intraclassa è del 6.58% e questo conferma che è presente una certa variabilità tra MMG che può avere un effetto nella spesa sostenuta dai rispettivi assistiti. La spesa per farmaci è influenzata dalle caratteristiche del singolo ma anche dal gruppo di appartenenza, che influisce sul comportamento delle persone al suo interno.

La terza colonna della Tabella 14 racchiude le stime delle caratteristiche degli assistiti e delle caratteristiche dei Medici di Medicina Generale. Le stime delle variabili relative agli assistiti non sono molto diverse dalle stime del modello di regressione lineare senza covariate del medico (colonna 1). La significatività delle stime, invece, è la stessa. Degne di nota sono, quindi, le sole variabili delle caratteristiche del MMG. Il genere del medico non è significativo, essere un medico uomo o un medico donna non influenza la spesa per farmaci, *ceteris paribus*. Invece, i pazienti associati a medici che superano la soglia di assistiti massimi hanno una spesa per farmaci dell'1,4% maggiore rispetto ai medici che non superano la soglia, *ceteris paribus*. Come nel modello di regressione per visite specialistiche (Tabella 13, terza colonna), anche nel modello di regressione della spesa per farmaci l'andamento delle stime della variabile esperienza del medico è di tipo parabolico. Con il passare degli anni dalla laurea i medici

tendono a prescrivere farmaci più costosi per poi invertire tale tendenza dopo un certo livello di esperienza.

L'indice di bontà di adattamento R^2 è del 35.8%, non è molto differente dal modello senza covariate del medico, rimane comunque un discreto indice. Il test F porta a rifiutare l'ipotesi nulla.

Il modello multilivello con covariate del medico (quarta colonna) presenta stime molto simili al modello multilivello senza covariate del medico, per quanto riguarda le caratteristiche degli assistiti. Focalizzandosi, invece, sulle caratteristiche dei MMG, le stime si avvicinano molto al modello non multilivello. Il genere del medico rimane una variabile non significativa, mentre la dummy che indica se un medico ha superato il numero massimo di assistiti è invece significativa. Le persone associate a questi medici hanno una spesa per farmaci del 7% maggiore rispetto ai medici che non superano la soglia massima, *ceteris paribus*. Le stime delle variabili "anni dalla laurea" sono entrambe significative e di segno opposto. L'esperienza del medico, in questo caso, segue un andamento parabolico per quanto riguarda la spesa per farmaci.

L'indice ICC è del 5.46%: introducendo variabili di secondo livello l'indice si è un po' abbassato. Questo porta alla conclusione che è diminuita la variabilità tra i gruppi (variabilità *between*): le covariate del medico aiutano a spiegare in modo migliore tale variabilità e evidenziano che il comportamento degli assistiti dipende dal gruppo (MMG) al quale appartengono.

3.3 Stima del modello Tobit univariato e Tobit multivariato

I modelli precedenti hanno permesso di studiare il comportamento delle variabili esplicative quando la variabile risposta era di tipo binario, perciò quando l'assistito aveva effettuato una spesa oppure no. Questo permetteva di catturare l'associazione tra le variabili esplicative e la variabile risposta solo attraverso una certa probabilità, ma non la magnitudo. Ho deciso, quindi, di stimare dei modelli di regressione in modo da catturare la grandezza delle stime stesse, ma utilizzando dei sottocampioni, ovvero le sole persone che hanno effettuato una spesa positiva. Come detto nel Capitolo 2 (si veda Paragrafo 2.3), limitare l'attenzione alle sole osservazioni positive sarebbe inappropriato poiché il valore atteso condizionale di y_i dato che $y_i > 0$, non è più dato da $x_i\beta$, ma dipende da x_i in modo non lineare attraverso $\phi(\cdot)/\Phi(\cdot)$. Entrambi i modelli, probit e di regressione, presentano, quindi, dei limiti. Inoltre, i valori assunti dalle variabili dipendenti analizzate sono nulli per una parte rilevante della popolazione ma positivi per la parte restante. I modelli finora utilizzati non permettevano di stimare una probabilità positiva di osservare il valore "zero", che ha un significato economico e non è un problema di osservazioni mancanti, ed anche una distribuzione continua sui valori strettamente positivi. Decido, quindi, di introdurre il modello Tobit. In effetti, l'informazione contenuta nel modello tobit combina quella contenuta nel modello di regressione troncata con quella del modello probit che descrive la spesa/non spesa. I parametri β del modello tobit hanno una doppia interpretazione: la prima è l'impatto di una variazione di x_i sulla probabilità di una spesa non nulla, e la seconda è l'impatto di una variazione di x_i sul livello di questa spesa.

Nelle Tabelle 15, 16, 17 sono presentate le stime del modello Tobit univariato e multivariato della spesa per ricoveri, visite specialistiche e farmaci rispettivamente. Decido di introdurre i modelli multivariati in modo tale da catturare, se esiste, una certa correlazione tra le variabili dipendenti, visto si tratta di tre tipi di spesa che si riferiscono allo stesso campo, ovvero la Sanità Pubblica.

I modelli precedenti hanno evidenziato una variabilità non trascurabile del MMG per quanto riguarda l'analisi della spesa per visite specialistiche e farmaci in particolare, variabilità che tuttavia può essere spiegata da certe caratteristiche del MMG stesso. Per questo motivo, nei prossimi modelli vengono inserite, oltre alle caratteristiche degli assistiti, anche le covariate del Medico di Medicina Generale. Per questi settori di spesa è stato considerato solo il

modello Tobit standard e non il modello Tobit multilivello poiché le stime delle variabili dei modelli standard e multilivello erano molto simili e, come detto, certe caratteristiche del MMG sono in grado di spiegare fortemente la variabilità dei medici che emerge dalle analisi multilivello. Per quanto riguarda la spesa per ricoveri, i modelli saranno caratterizzati dalle sole covariate degli assistiti.

Come nei modelli di regressione, anche in questi le variabili dipendenti vengono introdotte per mezzo di una trasformazione logaritmica.

3.3.1 Ricoveri ospedalieri

Log-tariffa ricovero		Tobit univariato		Tobit multivariato		$\beta_{\text{tobit}}/\sigma_{\text{tobit}}$	Probit*
Variabili esplicative		(1)		(2)		(3)	(4)
		<i>Coeff</i>	<i>P-value</i>	<i>Coeff</i>	<i>P-value</i>	<i>Coeff</i>	<i>Coeff</i>
Sesso	Femmina	0,514	0,000	1,095	0,000	0,042	0,050
Età	25-34	1,821	0,000	1,578	0,000	0,147	0,141
	35-44	2,010	0,000	1,734	0,000	0,163	0,148
	45-54	2,166	0,000	1,888	0,000	0,175	0,150
	55-64	3,318	0,000	3,156	0,000	0,268	0,237
	65-74	4,959	0,000	4,981	0,000	0,401	0,365
	≥75	6,020	0,000	6,252	0,000	0,487	0,452
Decesso	2007	12,655	0,000	12,116	0,000	1,023	1,431
	2008	5,826	0,000	5,615	0,000	0,471	0,535
ADI	Ambulatorio	1,978	0,000	2,480	0,000	0,160	0,169
	Domicilio	7,658	0,000	7,564	0,000	0,619	0,749
Esenzioni	Affezioni sistema circolatorio	2,900	0,000	3,148	0,000	0,234	0,260
	Malattie cardiache	4,917	0,000	5,213	0,000	0,397	0,429
	Diabete	2,358	0,000	2,939	0,000	0,191	0,210
	Glaucoma	0,520	0,000	1,126	0,000	0,042	0,046
	Colesterolo	0,456	0,002	1,060	0,000	0,037	0,032
	Ipertensione arteriosa con danno d'organo	1,199	0,000	1,635	0,000	0,097	0,101
	Ipertensione arteriosa senza danno d'organo	0,574	0,000	1,061	0,000	0,046	0,045
	Malattie neoplastiche maligne	6,031	0,000	6,283	0,000	0,488	0,528
	Invalità civile > 2/3	2,885	0,000	3,151	0,000	0,233	0,244
	Invalità civile (assegno accomp.)	1,398	0,000	1,722	0,000	0,113	0,109
	Invalità civile al 100%	2,694	0,000	2,840	0,000	0,218	0,235
	Gravidanza	20,219	0,000	20,537	0,000	1,634	1,916
	Altro	3,540	0,000	4,226	0,000	0,286	0,304
Costante		-20,870	0,000	-21,348	0,000	-1,687	-1,714
σ		12,372	0,000	12,522	0,000		
ρ_{12} (ricoveri-specialistiche)				0,366	0,000		
ρ_{13} (ricoveri-farmaci)				0,231	0,000		
ρ_{23} (specialistiche - farmaci)				0,375	0,000		

*Stime calcolate in precedenza: Tabella 9, prima colonna

Tabella 15 - Stima del modello Tobit univariato e multivariato, della spesa per ricoveri ospedalieri

La Tabella 15 riporta, nella colonna 1 e 2, le stime del modello Tobit univariato e multivariato della spesa per ricoveri ospedalieri. Nella colonna 4 ho inserito le stime del modello probit standard per valutare la correttezza del modello tobit. I modelli tobit e probit, infatti, hanno la stessa struttura ma diversa è l'interpretazione della variabile latente y^* . Nel modello tobit sappiamo il valore di y^* quando $y^* > 0$, mentre nel modello probit sappiamo solo se $y^* > 0$. Da questo si evince che c'è più informazione nel modello tobit, perciò le sue stime dovrebbero essere più efficienti. Se la specificazione del modello tobit è corretta, allora le stime del modello probit sono consistenti per $\beta_{\text{tobit}}/\sigma_{\text{tobit}}$. Come dimostra la Tabella 15, le stime della terza e della quarta colonna sono molto simili perciò il modello tobit è corretto.

Passando all'interpretazione delle stime del modello tobit univariato, prima colonna, si noti come tutte le variabili siano significativamente e positivamente associate con la variabile dipendente. Le donne tendono a spendere di più in ricoveri ospedalieri rispetto agli uomini, *ceteris paribus*. L'età ha un importante valore predittivo. Il coefficiente relativo all'età cresce costantemente nel modello, i pazienti più anziani tendono ad utilizzare più risorse rispetto agli assistiti di età inferiore. Un valore molto elevato è associato alla variabile che si riferisce a chi è deceduto nel 2007, queste persone hanno una spesa per ricoveri sostanziosa, sicuramente maggiore rispetto a chi è deceduto nel 2008. E' probabile che le persone che vengono a mancare siano stati sottoposte a dei ricoveri d'urgenza, caratterizzati magari da molti giorni di degenza e per questo motivo le spese aumentano molto rispetto a chi non è deceduto in quel determinato anno. La stima della variabile del servizio ADI a domicilio è maggiore rispetto a quella del servizio ADI ambulatorio. Come già detto nei modelli precedenti, coloro che sfruttano tale servizio a domicilio sono persone che non sono in grado di deambulare da sole e che hanno continuamente bisogno di assistenza. Si ipotizza, quindi, che questa tipologia di assistiti siano più a rischio di essere ricoverati rispetto a chi non sfrutta il servizio ADI poiché, appunto, la loro condizione di salute è più precaria. Le variabili relative alle esenzioni danno uno sguardo rapido alle patologie più costose a livello di ricoveri ospedalieri. Al primo posto troviamo l'esenzione per gravidanza. Ovviamente la stima di tale variabile è elevata poiché è scontato che una persona con questa esenzione venga ricoverata almeno per qualche giorno. Una patologia che consuma molte risorse è il cancro, seguita poi dalle malattie cardiache. Le esenzioni concesse per malattie neoplastiche maligne sono molte, quasi 60.000. Un'analisi descrittiva dei dataset amministrativi evidenzia che il range di giorni di ricovero va da 1 a 287 con una media di 8 giorni, la mediana si aggira sui 5 giorni. Circa il 90% degli assistiti con questa esenzione sono stati ricoverati per un massimo di 20 giorni. Questi numeri giustificano

la spesa elevata per questo tipo di ricovero, più giorni di degenza caratterizzano un ricovero più i costi aumentano. Le esenzioni concesse per malattie cardiache sono, invece, quasi 22.000, molte meno rispetto alle esenzioni per malattie neoplastiche maligne. Ciò nonostante, il range dei ricoveri va da un minimo di 1 a un massimo di 390 giorni, con circa 8 giorni in media per ogni ricovero e una mediana di 5, numeri molto simili all'esenzione per malattie neoplastiche maligne. In questo caso, è ipotizzabile che chi viene ricoverato per malattie cardiache abbiano bisogno di cure, nel complesso, un po' più costose. Il ricovero ospedaliero di coloro che hanno l'esenzione per malattie cardiache costa, in media, circa 6.000 euro contro i circa 5.000 di coloro che vengono ricoverati ed hanno l'esenzione per malattie neoplastiche maligne. Altre esenzioni onerose a livello di ricoveri sono le esenzioni per affezioni del sistema circolatorio, diabete, invalidità civile >2/3, invalidità civile al 100%. Questi ultimi vengono ricoverati, in media, per 10 giorni, seguiti poi dagli esenti per diabete, 9 giorni in media.

L'indice di bontà del modello, lo pseudo R^2 , è del 7%. Il test di Wald dell'ipotesi che tutti i coefficienti siano nulli, eccetto l'intercetta, porta a rifiutare l'ipotesi nulla. Questo test si distribuisce come una chi-quadrato con 24 gradi di libertà.

La seconda colonna della Tabella 15 riporta le stime del modello tobit multivariato. Le stime dei coefficienti delle variabili esplicative sono molto simili a quelli del tobit univariato, non ci sono particolari differenze fra i due modelli. Per quanto riguarda i coefficienti di correlazione degli errori ρ , si può notare che sono molto bassi. I ricoveri ospedalieri hanno una correlazione del 23.1% con la spesa per farmaci e del 36.6% con la spesa per visite specialistiche. La correlazione è leggermente più elevata tra ricoveri e visite specialistiche poiché, è probabile, che dopo un ricovero ospedaliero vengano prescritte visite specialistiche accurate per prevenire o per tenere sotto controllo le patologie riscontrate nel ricovero. Oppure, determinate visite specialistiche possono essere il preludio di un successivo ricovero ospedaliero. Il test di verosimiglianza che pone i coefficienti di correlazione tutti uguali a 0 porta a rifiutare l'ipotesi nulla.

3.3.2 Visite specialistiche

Log-tariffa specialistiche		Tobit univariato (1)		Tobit multivariato (2)		$\beta_{\text{Tobit}}/\sigma_{\text{Tobit}}$ (3)	Probit* (4)
Variabili esplicative		<i>Coeff</i>	<i>P-value</i>	<i>Coeff</i>	<i>P-value</i>	<i>Coeff</i>	<i>Coeff</i>
Sesso	Femmina	0,872	0,000	0,888	0,000	0,268	0,317
Età	25-34	0,015	0,293	0,003	0,834	0,004	-0,010
	35-44	0,390	0,000	0,376	0,000	0,120	0,085
	45-54	0,845	0,000	0,825	0,000	0,260	0,201
	55-64	1,360	0,000	1,340	0,000	0,418	0,354
	65-74	2,089	0,000	2,072	0,000	0,642	0,629
	≥75	1,862	0,000	1,855	0,000	0,572	0,537
Decesso	2007	-1,075	0,000	-1,050	0,000	-0,330	-0,578
	2008	-0,070	0,014	-0,065	0,024	-0,022	-0,120
ADI	Ambulatorio	1,224	0,000	1,229	0,000	0,376	1,232
	Domicilio	0,712	0,000	0,710	0,000	0,219	0,627
Esenzioni	Affezioni sistema circolatorio	0,809	0,000	0,815	0,000	0,249	0,440
	Malattie cardiache	1,121	0,000	1,127	0,000	0,345	0,707
	Diabete	1,255	0,000	1,265	0,000	0,386	0,663
	Glaucoma	1,045	0,000	1,052	0,000	0,321	0,707
	Colesterolo	1,080	0,000	1,091	0,000	0,332	0,676
	Ipertensione arteriosa con danno d'organo	0,728	0,000	0,757	0,000	0,224	0,390
	Ipertensione arteriosa senza danno d'organo	0,607	0,000	0,645	0,000	0,187	0,290
	Malattie neoplastiche maligne	2,227	0,000	2,234	0,000	0,684	0,960
	Invalidità civile > 2/3	0,886	0,000	0,890	0,000	0,272	0,414
	Invalidità civile (assegno accomp.)	-0,247	0,000	-0,234	0,000	-0,076	-0,027
	Invalidità civile al 100%	0,450	0,000	0,458	0,000	0,138	0,208
	Gravidanza	3,953	0,000	3,963	0,000	1,214	1,747
	Altro	1,714	0,000	1,731	0,000	0,526	0,700
		Costante	-0,069	0,116	-0,136	0,001	-0,021
Sesso medico	Maschio	-0,156	0,000	-0,152	0,000	-0,048	-0,052
	Supera max assistiti	0,155	0,000	0,161	0,000	0,047	0,052
	Anni dalla laurea	4,199	0,000	4,501	0,000	1,290	0,024
	Anni dalla laurea ²	-0,881	0,000	-0,926	0,000	-0,271	-0,025
σ		3,255	0,000	3,263	0,000		
ρ_{12} (ricoveri-specialistiche)				0,366	0,000		
ρ_{13} (ricoveri-farmaci)				0,231	0,000		
ρ_{23} (specialistiche-farmaci)				0,375	0,000		

Tabella 16 - Stima del modello Tobit univariato e multivariato, della spesa per visite specialistiche

*Stime calcolate in precedenza: Tabella 10, terza colonna

La prima colonna della Tabella 16 evidenzia le stime del modello tobit univariato della spesa per visite specialistiche. Rispetto al modello tobit della spesa per ricoveri, si noti l'aggiunta delle covariate del medico di medicina generale, risultate significative nei modelli multilivello. Il genere degli assistiti ha un valore predittivo importante, le donne spendono di più in visite specialistiche rispetto ai maschi, cosa riscontrata anche per i ricoveri ospedalieri. Anche l'età influenza positivamente la spesa per visite specialistiche. Si noti come la dummy dell'età degli assistiti tra i 25 e 34 sia, però, non significativa. La fascia d'età più ingente a livello di spesa per visite specialistiche è quella compresa tra i 65 e 74 anni. In generale, la spesa aumenta costantemente con l'età. Le dummy relative al decesso degli assistiti nel 2007 e nel 2008 sono significativamente e negativamente associate con la variabile dipendente. Soprattutto chi è deceduto nel 2007 ha una spesa per visite specialistiche minore rispetto a chi non è deceduto in quel determinato anno, *ceteris paribus*. Anche il modello probit in colonna 4 sottolinea questa tendenza, chi è deceduto nel 2007 ha meno probabilità di aver effettuato visite specialistiche rispetto a chi non è deceduto, a parità di altre condizioni. E' ipotizzabile che, una persona nell'ultimo periodo della sua vita faccia meno visite specialistiche oppure vengono effettuate durante il ricovero ospedaliero e perciò rientrano nella spesa per ricoveri, che infatti è positiva e ingente per coloro che sono deceduti nel 2007, come evidenziato in Tabella 15. Coloro che hanno usufruito del servizio ADI hanno una spesa per visite specialistiche maggiore rispetto a chi non ha tale servizio. In particolare, le persone che hanno usufruito dell'assistenza domiciliare integrata in ambulatorio hanno una spesa per visite specialistiche maggiore rispetto a chi ha sfruttato il servizio a domicilio. Le variabili relative alle esenzioni per patologia e gravidanza sono tutte significative. Come per la spesa per ricoveri, l'esenzione più onerosa a livello di visite specialistiche è l'esenzione per gravidanza. E' prassi che una donna incinta si sottoponga a varie visite di controllo per monitorare la situazione del bambino nell'arco di tutti i nove mesi. Perciò si tratta di una spesa per visite specialistiche che si protrae nel tempo e questo influisce sull'ammontare della spesa. Un'altra esenzione che porta ad una spesa per visite specialistiche ingente è quella per malattie neoplastiche maligne. Un assistito con questo tipo di esenzione svolge, in media, 47 visite specialistiche all'anno. Bisogna sottolineare, però, che le visite specialistiche svolte possono non essere associate con il tipo di esenzione data, ma può essere che una determinata patologia faccia scaturire altre problematiche che necessitano di controlli approfonditi. Perciò non è detto che le visite specialistiche di un individuo siano tutte inerenti al tipo di esenzione data. Per tale motivo la presenza di esenzioni può essere interpretata, in modo più generale, come presenza di una

salute non buona. Altre esenzioni rilevanti per la spesa per visite specialistiche sono l'esenzione per malattie cardiache, diabete, glaucoma e colesterolo. Coloro che soffrono di malattie cardiache svolgono in media 43 visite all'anno, chi soffre di diabete 42 visite, chi di glaucoma 32 e 36 sono le visite specialistiche che svolgono, in media, gli assistiti esenti per colesterolo. Tendenza opposta è attribuibile agli esenti per invalidità civile con assegno di accompagnamento. Quest'ultimi svolgono in media 30 visite all'anno, ma l'ammontare della loro spesa è più piccola rispetto a chi non ha tale esenzione.

Passando alle caratteristiche del MMG si noti come chi è associato a un medico maschio abbia una spesa per visite specialistiche minore rispetto a chi ha un medico donna. Nei primi anni di esperienza del medico di base la spesa per visite specialistiche è molto alta, probabilmente i medici tendono a prescriberne di più. Con il passare degli anni e acquisendo esperienza, il medico prescrive meno visite, questo andamento parabolico è evidenziato dalle stime delle variabili "anni dalla laurea".

L'indice di bontà del modello, lo pseudo R^2 è del 4,8%. Il test di Wald dell'ipotesi che tutti i coefficienti siano nulli, eccetto l'intercetta, porta a rifiutare l'ipotesi nulla. Questo test si distribuisce come una chi-quadrato con 28 gradi di libertà.

La seconda colonna riporta le stime del modello tobit multivariato, le quali non sono molto diverse da quelle del modello tobit univariato. Inoltre, la correlazione tra la spesa per visite specialistiche e per ricoveri è del 36.6% e tra la spesa per visite specialistiche e farmaci è del 37.5%. Non si tratta di correlazioni forti, anche se è ipotizzabile che dopo una visita specialistica il medico prescriva l'assunzione di alcuni farmaci.

Le stime del modello probit della quarta colonna sono molto simili alle stime del modello tobit riscaldato della terza colonna e questo conferma quanto spiegato precedentemente, le stime del modello tobit sembrano essere robuste.

3.3.3 Farmaci

Log-importo farmaci		Tobit univariato		Tobit multivariato		$\beta_{\text{Tobit}}/\sigma_{\text{Tobit}}$	Probit*
Variabili esplicative		(1)		(2)		(3)	(4)
		<i>Coeff</i>	<i>P-value</i>	<i>Coeff</i>	<i>P-value</i>	<i>Coeff</i>	<i>Coeff</i>
Sesso	Femmina	0,613	0,000	0,629	0,000	0,210	0,312
Età	25-34	0,434	0,000	0,420	0,000	0,149	0,127
	35-44	0,859	0,000	0,845	0,000	0,294	0,234
	45-54	1,495	0,000	1,481	0,000	0,512	0,365
	55-64	2,501	0,000	2,489	0,000	0,857	0,625
	65-74	3,393	0,000	3,383	0,000	1,163	0,952
	≥75	3,753	0,000	3,744	0,000	1,286	1,127
Decesso	2007	-1,003	0,000	-1,001	0,000	-0,344	-0,559
	2008	0,034	0,181	0,031	0,221	0,012	0,049
ADI	Ambulatorio	0,614	0,000	0,630	0,000	0,210	0,527
	Domicilio	0,777	0,000	0,779	0,000	0,266	0,817
Esenzioni	Affezioni sistema circolatorio	0,838	0,000	0,844	0,000	0,287	0,582
	Malattie cardiache	1,142	0,000	1,150	0,000	0,391	0,835
	Diabete	1,512	0,000	1,520	0,000	0,518	0,906
	Glaucoma	1,152	0,000	1,159	0,000	0,395	0,865
	Colesterolo	1,462	0,000	1,472	0,000	0,501	0,865
	Iperensione arteriosa con danno d'organo	1,687	0,000	1,696	0,000	0,578	1,024
	Iperensione arteriosa senza danno d'organo	1,850	0,000	1,861	0,000	0,634	1,107
	Malattie neoplastiche maligne	0,882	0,000	0,901	0,000	0,302	0,417
	Invalità civile > 2/3	0,871	0,000	0,878	0,000	0,299	0,392
	Invalità civile (assegno accomp.)	0,075	0,001	0,087	0,000	0,026	-0,045
	Invalità civile al 100%	0,553	0,000	0,560	0,000	0,190	0,242
	Gravidanza	0,626	0,000	0,680	0,000	0,215	0,358
	Altro	1,543	0,000	1,560	0,000	0,529	0,717
Costante		-0,272	0,000	-0,322	0,000	-0,093	-0,201
Sesso medico	Maschio	0,025	0,000	0,027	0,000	0,009	0,015
	Supera max assistiti	0,147	0,000	0,151	0,000	0,051	0,068
	Anni dalla laurea	-0,044	0,882	0,151	0,613	-0,015	-1,557
	Anni dalla laurea ²	-0,087	0,136	-0,116	0,045	-0,030	0,265
σ		2,918	0,000	2,931	0,000		
ρ_{12} (ricoveri-specialistiche)				0,366	0,000		
ρ_{13} (ricoveri-farmaci)				0,231	0,000		
ρ_{23} (specialistiche-farmaci)				0,375	0,000		

*Stime calcolate in precedenza: Tabella 11, terza colonna

Tabella 17 - Stima del modello Tobit univariato e multivariato, della spesa farmaci

La prima colonna della Tabella 17 racchiude le stime del modello tobit univariato della spesa per farmaci; oltre alle caratteristiche degli assistiti sono state inserite anche le covariate del medico di medicina generale, risultate significative nei modelli precedenti. Come nella spesa per ricoveri e visite specialistiche, anche nella spesa per farmaci le donne spendono di più rispetto agli uomini. Perciò, in generale, le donne gravano maggiormente sulla Sanità pubblica. Le dummy dell'età hanno un importante valore predittivo, sono associate positivamente e significativamente con la variabile risposta. La spesa per farmaci aumenta costantemente con l'età. Le persone decedute nel 2007, invece, tendono a spendere meno in farmaci rispetto a chi non è deceduto. Probabilmente, esclusi decessi improvvisi, una persona in gravi condizioni tende ad essere ricoverata e i farmaci che assume rientrano nella spesa per ricoveri ospedalieri che, come dimostra la prima colonna della Tabella 15, è molto ingente per i deceduti nel 2007. La variabile deceduto nel 2008, invece, non influenza significativamente il livello della spesa per farmaci, a parità di altre condizioni. Gli assistiti che hanno usufruito del servizio ADI, a domicilio o in ambulatorio, influenzano positivamente la spesa per farmaci. Passando all'interpretazione delle stime delle esenzioni si noti come, in questo caso, l'esenzione per gravidanza non sia una tra le spese più ingenti. Le donne incinta, in realtà, non assumono molti farmaci, invece è più probabile che facciano visite specialistiche e ricoveri ospedalieri. Le stime più elevate si riscontrano sulle esenzioni per ipertensione arteriosa con e senza danno d'organo. Le persone con tali esenzioni spendono all'anno rispettivamente, in media, 768 euro e 600 euro in farmaci. Altre esenzioni che portano ad avere una spesa ingente per farmaci a carico del Sistema Sanitario sono quelle concesse agli assistiti con problemi di malattie cardiache, diabete, glaucoma e colesterolo. Per esempio, i pazienti affetti da malattie cardiache spendono, in media, 827 euro all'anno per l'acquisto di farmaci. Il 90% di queste persone, però, spende quasi 1700 euro all'anno in prodotti farmaceutici. Gli assistiti con problemi di diabete spendono all'anno, mediamente, 837 euro e il 90% di queste persone spende più di 1700 euro all'anno per farmaci. Spendono mediamente meno gli assistiti con l'esenzione per glaucoma e colesterolo, rispettivamente 757 e 802 euro.

Le stime delle caratteristiche del medico di medicina generale non sono tutte significative. Il genere del medico e l'aver superato il numero massimo di assistiti sono associati positivamente e significativamente con la variabile risposta. L'esperienza del medico, invece, non è significativa. Gli assistiti che hanno un medico maschio tendono a spendere di più in farmaci rispetto agli assistiti associati a un medico donna, *ceteris paribus*.

L'indice di bontà del modello, lo pseudo R^2 è del 8.8%. Il test di Wald che testa l'ipotesi che tutti i coefficienti siano nulli, eccetto l'intercetta, porta a rifiutare l'ipotesi nulla. Questo test si distribuisce come una chi-quadrato con 28 gradi di libertà.

In colonna 2 troviamo le stime del modello tobit multivariato e si noti come siano molto simili alle stime del tobit univariato. Le uniche stime diverse sono quelle associate alle variabili non significative. La correlazione tra spesa per farmaci e per ricoveri è del 23.1% e tra la spesa per farmaci e visite specialistiche è del 37.5%.

Le stime del modello tobit sembrano essere robuste poiché le stime della colonna 3 e della colonna 4 sono molto simili.

3.4 Stima del modello di Heckman

I modelli Tobit stimati in precedenza sono corretti e sembrano adattarsi abbastanza bene ai dati, però presentano alcuni limiti. Come spiegato nel Capitolo 2, il modello tobit standard impone una struttura che è spesso troppo restrittiva: le variabili che determinano la probabilità di un'osservazione nulla hanno esattamente la stessa relazione delle variabili che influenzano il livello di un'osservazione positiva (Verbeek, 2010). Introduco, perciò, un'estensione del modello Tobit ovvero il modello Tobit II, detto appunto modello di Heckman.

Stata stima il modello di Heckman tramite il metodo della massima verosimiglianza. Le variabili dipendenti utilizzate non hanno subito trasformazioni logaritmiche come nei modelli precedenti, ma sono state riscalate per 1/1000.

Come evidenziato nei modelli precedenti, le stime puntuali dei modelli multivariati sono molto simili a quelle dei modelli univariati. Inoltre, le correlazioni stimate tra le variabili dipendenti sono nel complesso piccole. Per questo motivo sembra del tutto appropriato analizzare separatamente le tre voci di spesa, anziché congiuntamente.

Le Tabelle 18, 19, 20 presentano le stime del modello di Heckman della spesa per ricoveri ospedalieri, visite specialistiche e farmaci. Le variabili di base per i tre modelli sono le caratteristiche degli assistiti e per quanto riguarda la spesa per visite specialistiche e farmaci si introducono anche le covariate del medico di medicina generale risultate significative nei modelli multilivello.

Visti i risultati delle analisi fino a qui condotte, è ragionevole però supporre che le assunzioni poste in essere fino a questo momento (tralasciare sia una struttura multilivello sia un approccio multivariato per le tre voci di spesa), abbiano un impatto nel complesso limitato nell'analisi di questo fenomeno.

3.4.1 Ricoveri ospedalieri

Tariffa ricovero

Variabili esplicative		Equazione di regressione (1)		Equazione di selezione (2)	
		Coeff	P-value	Coeff	P-value
Sesso	Femmina	-0,797	0,000	0,050	0,000
Età	25-34	0,081	0,464	0,141	0,000
	35-44	-0,025	0,816	0,148	0,000
	45-54	0,665	0,000	0,150	0,000
	55-64	0,967	0,000	0,236	0,000
	65-74	1,320	0,000	0,364	0,000
	≥75	0,513	0,000	0,451	0,000
Decesso	2007	1,777	0,000	1,450	0,000
	2008	1,924	0,000	0,538	0,000
ADI	Ambulatorio	0,818	0,000	0,169	0,000
	Domicilio	2,501	0,000	0,757	0,000
Esenzioni	Affezioni sistema circolatorio	0,849	0,000	0,261	0,000
	Malattie cardiache	2,273	0,000	0,433	0,000
	Diabete	0,558	0,000	0,210	0,000
	Glaucoma	-0,634	0,000	0,046	0,000
	Colesterolo	0,112	0,422	0,032	0,008
	Ipertensione arteriosa con danno d'organo	0,112	0,099	0,101	0,000
	Ipertensione arteriosa senza danno d'organo	-0,188	0,034	0,046	0,000
	Malattie neoplastiche maligne	1,791	0,000	0,532	0,000
	Invalità civile > 2/3	0,161	0,109	0,243	0,000
	Invalità civile (assegno accomp.)	-0,817	0,000	0,109	0,000
	Invalità civile al 100%	0,327	0,001	0,236	0,000
	Gravidanza	-2,564	0,000	1,910	0,000
	Altro	0,586	0,000	0,305	0,000
Costante		7,033	0,000	-1,715	0,000
σ		7,345			
ρ		-0,219			

Tabella 18 - Stima del modello di Heckman della spesa per ricoveri ospedalieri

La Tabella 18 riporta le stime del modello di Heckman per quanto riguarda la spesa per ricoveri ospedalieri. La seconda colonna, l'equazione di selezione, descrive come la probabilità di una variabile esplicativa influenzi il fatto di aver effettuato una spesa oppure no. Questo si riconduce ad un modello probit standard (ed infatti le stime sono pressoché uguali a quelle in Tabella 9, prima colonna). Risultati interessanti, invece, emergono dalle stime dell'equazione di regressione. Alcune variabili risultano non significative al 5% come nel modello di regressione svolto sulle sole osservazioni positive (Tabella 12, colonna 1), ma anche altre variabili risultano ora non significative. Nell'equazione di regressione rientrano le sole osservazioni che hanno avuto una spesa per ricoveri maggiore di zero, ovvero quelle che nell'equazione di selezione avevano una certa probabilità di osservare il valore 1, l'aver effettuato una spesa positiva. Il fatto che molte variabili nell'equazione di regressione abbiano significatività diverse rispetto a quelle della regressione semplice effettuata sulle sole osservazioni positive fa capire come il modello Tobit II possa essere maggiormente informativo del modello Tobit.

Nel dettaglio si noti come il sesso abbia un valore predittivo importante sulla variabile risposta. Le donne hanno più probabilità di fare un ricovero ospedaliero, a parità di altre condizioni, ma in media spendono quasi 8.000 euro in meno all'anno rispetto agli uomini. Questa tendenza è emersa anche dalle stime del modello probit in Tabella 9 e del modello di regressione standard sul campione ridotto in Tabella 12. Perciò il modello di Heckman ha solo confermato quello che già i modelli precedenti avevano segnalato. L'equazione di selezione mostra come le dummy dell'età sono tutte significative e di segno positivo. Con il passare degli anni c'è una maggior probabilità di essere ricoverati rispetto ai pazienti compresi nella fascia d'età 14-24, *ceteris paribus*. L'equazione di regressione, però, introduce per la prima volta la non significatività delle dummy dell'età 25-34 e 35-44. Gli assistiti compresi in queste fasce d'età non influiscono significativamente sul livello della spesa per ricoveri rispetto agli assistiti d'età 14-24, *ceteris paribus*. I pazienti deceduti nel 2007 hanno maggior probabilità di essere ricoverati rispetto a chi non è deceduto nel 2007 ma, in media, spende di più in ricoveri ospedalieri chi è deceduto nel 2008. È ipotizzabile che, chi è deceduto l'anno successivo rispetto a quello in analisi, abbia avuto più possibilità di essere ricoverato durante i 365 giorni dell'anno. L'ammontare di spesa attribuito a coloro che sono deceduti nel 2007, invece, può essere relativo a soli pochi giorni o mesi dell'anno, per questo motivo la stima è più piccola. Coloro che hanno usufruito del servizio ADI hanno maggior probabilità di essere ricoverati rispetto a chi non ha usato il servizio, nel dettaglio chi ha sfruttato il servizio a

domicilio spende 2.500 euro in più in ricoveri ospedalieri rispetto a chi non ha usufruito del servizio a domicilio.

Passando alle esenzioni per patologia, si noti come l'equazione di selezione evidenzia che tutti gli assistiti che hanno almeno un'esenzione descritta in tabella, abbiano maggior probabilità di effettuare un ricovero ospedaliero rispetto a coloro che non hanno quel tipo di esenzione. Ma se ci focalizziamo sull'impatto che le esenzioni hanno sul livello di spesa, si noti come alcune di esse non siano significative. Le esenzioni per colesterolo, per ipertensione arteriosa con danno d'organo e invalidità civile >2/3 risultano non significative al 5%, l'esenzione per ipertensione arteriosa senza danno d'organo è non significativa al 10%. Questo significa che i pazienti con tali esenzioni hanno sì maggior probabilità di effettuare un ricovero ospedaliero ma la loro spesa non influisce significativamente sulla spesa complessiva, *ceteris paribus*. Alcune esenzioni influiscono negativamente sul livello della spesa complessiva. Per esempio, gli esenti per gravidanza spendono, in media, 2.564 euro in meno all'anno rispetto agli assistiti che non hanno tale esenzione. In Friuli-Venezia Giulia un ricovero ospedaliero per gravidanza costa alla Sanità mediamente 2.150 euro, e i giorni di degenza sono, in media, 3-4. Gli assistiti con l'esenzione per glaucoma e gli assistiti con l'esenzione per invalidità civile con assegno di accompagnamento spendono rispettivamente 634 € e 817 € in meno rispetto ai pazienti che non hanno tali esenzioni, *ceteris paribus*. In media, un ricovero ospedaliero con l'esenzione per glaucoma costa alla Sanità 4546 euro, con una media di 8 giorni di degenza. Gli assistiti con l'esenzione per invalidità civile con accompagnamento che vengono ricoverati, costano mediamente alla Sanità 4.853 euro, con circa 11 giorni di degenza in media. L'esenzione che costa di più alla Sanità in termini di spesa per ricoveri ospedalieri è l'esenzione per malattie cardiache, seguita poi dall'esenzione per malattie neoplastiche maligne. Gli assistiti con l'esenzione per malattie cardiache che vengono ricoverati, costano mediamente 2.273 euro in più rispetto a coloro che non hanno tale esenzione, *ceteris paribus*. Gli esenti per malattie neoplastiche maligne che sono ricoverati costano, invece, 1791 euro in più rispetto a chi non ha tale esenzione. Analizzando i dataset amministrativi si nota che, in media, un ricovero ospedaliero per gli assistiti con l'esenzione per malattie cardiache costa 5911 euro, con una media di giorni di degenza pari a 8. Un ricovero ospedaliero per le persone con l'esenzione per malattie neoplastiche maligne costa alla Sanità, mediamente, 5140 euro, con una media di 8 giorni di degenza. Altre esenzioni che gravano positivamente sul livello della spesa complessiva per ricoveri sono l'esenzione per affezioni del sistema circolatorio, diabete, e invalidità civile al 100%.

Il test di Wald che ipotizza che tutti i coefficienti dell'equazione di regressione siano pari a 0 eccetto l'intercetta porta a rifiutare l'ipotesi nulla. Questo test si distribuisce come una chi-quadro con 24 gradi di libertà. Il test LR che testa l'incorrelazione tra le equazioni porta a rifiutare l'ipotesi nulla. La correlazione, infatti, è significativa e negativa. Questo segnala, ad esempio, l'esistenza di caratteristiche non osservabili che influiscono positivamente sul fatto di aver effettuato o meno un ricovero ospedaliero, ma negativamente sul livello della quota di spesa relativa ai ricoveri.

Questo è un risultato importante perché significa che studiare separatamente la probabilità di fare un ricovero e la grandezza della spesa del ricovero stesso porta a delle conclusioni interessanti. Il modello Tobit I non avrebbe permesso di captare il segno differente di alcune variabili, che influiscono positivamente sull'effettuare un ricovero ma negativamente sul livello della spesa. Nel modello Tobit standard questi due aspetti rientrano nella stessa stima e per giunta con lo stesso segno.

3.4.2 Visite specialistiche

Tariffa visite specialistiche

Variabili esplicative		Equazione di regressione (1)		Equazione di selezione (2)	
		Coeff	P-value	Coeff	P-value
Sesso	Femmina	-0,003	0,016	0,317	0,000
Età	25-34	0,022	0,000	-0,010	0,051
	35-44	0,035	0,000	0,085	0,000
	45-54	0,040	0,000	0,200	0,000
	55-64	0,050	0,000	0,352	0,000
	65-74	0,112	0,000	0,628	0,000
	≥75	0,095	0,000	0,541	0,000
Decesso	2007	0,130	0,000	-0,584	0,000
	2008	0,327	0,000	-0,118	0,000
ADI	Ambulatorio	0,015	0,002	1,235	0,000
	Domicilio	-0,003	0,381	0,630	0,000
Esenzioni	Affezioni sistema circolatorio	0,032	0,000	0,441	0,000
	Malattie cardiache	0,050	0,000	0,715	0,000
	Diabete	0,079	0,000	0,665	0,000
	Glaucoma	0,044	0,000	0,709	0,000
	Colesterolo	0,020	0,000	0,674	0,000
	Ipertensione arteriosa con danno d'organo	0,049	0,000	0,394	0,000
	Ipertensione arteriosa senza danno d'organo	0,009	0,003	0,291	0,000
	Malattie neoplastiche maligne	0,445	0,000	0,989	0,000
	Invalità civile > 2/3	0,106	0,000	0,417	0,000
	Invalità civile (assegno accomp.)	-0,099	0,000	-0,027	0,015
	Invalità civile al 100%	0,217	0,000	0,224	0,000
	Gravidanza	0,273	0,000	1,741	0,000
	Altro	0,135	0,000	0,703	0,000
		Costante	0,102	0,000	-0,194
Sesso medico	Maschio	-0,004	0,005	-0,052	0,000
	Supera max assistiti	0,005	0,001	0,052	0,000
	Anni dalla laurea	0,526	0,000	0,013	0,923
	Anni dalla laurea^2	-0,108	0,000	-0,023	0,372
σ		0,580			
ρ		-0,120			

Tabella 19 - Stima del modello di Heckman della spesa per visite specialistiche

La Tabella 19 racchiude le stime del modello di Heckman della spesa per visite specialistiche. Come evidenziava il modello Probit in Tabella 10, colonna 3, le donne hanno più probabilità di far visite specialistiche rispetto agli uomini. La stima della variabile nell'equazione di regressione, però, è significativa al 5% e non all' 1%. La differenza sul livello di spesa per visite specialistiche tra uomini e donne è irrisoria. Le variabili dummy dell'età sono tutte significative in entrambe le equazioni ma nell'equazione di selezione l'età dai 25-34 anni è non significativa al 5%. Anche il modello Tobit I evidenziava la non significatività di tale variabile. In generale, però, con il crescere dell'età c'è maggiore probabilità di far visite specialistiche e perciò la spesa aumenta costantemente con l'età, a parità di altre condizioni. Per quanto riguarda le variabili deceduto nel 2007 e deceduto nel 2008, l'equazione di selezione mostra come entrambe queste variabili portino ad avere una probabilità minore di fare visite specialistiche rispetto a chi non è deceduto. Stessa tendenza si può riscontrare nel modello Tobit standard (Tabella 16, colonna 1) e nel modello Probit (Tabella 10, colonna 3). L'equazione di regressione, invece, inverte questa tendenza: le persone decedute nel 2007 o nel 2008 hanno meno probabilità di far visite specialistiche, ma, in media, la loro spesa influisce maggiormente sulla spesa complessiva rispetto a chi non è deceduto, a parità di altre condizioni. Probabilmente, negli ultimi mesi di vita di una persona vengono effettuate meno visite specialistiche ma magari molto costose. Le stime dell'equazione di selezione circa le variabili ADI ambulatorio e ADI domicilio sono significative e positive, come quelle del modello Probit e Tobit di riferimento. Le persone che usufruiscono di tale servizio hanno più probabilità di far visite specialistiche, ma il loro livello di spesa non influisce significativamente sulla spesa complessiva per visite specialistiche, rispetto a chi non sfrutta il servizio ADI. Le stime delle variabili relative alle esenzioni sono tutte significative al 5% sia nell'equazione di selezione sia in quella di regressione. L'unica esenzione non significativa all'1% è quella dell'invalidità civile con assegno di accompagnamento. Gli assistiti con tale esenzione hanno meno probabilità di far visite specialistiche ma, mediamente, spendono 99 euro in meno rispetto a chi non ha tale esenzione, *ceteris paribus*. I pazienti che ricevono un assegno di accompagnamento sono persone non in grado di deambulare da sole ma che, in realtà, non presentano patologie particolari, probabilmente per questo le stime sono negative per entrambe le equazioni. In generale, le persone che fanno una visita specialistica ed hanno un'esenzione per malattie neoplastiche maligne spendono, in media, 445 euro in più rispetto a chi non ha tale esenzione. Questi assistiti sono i più costosi in termini di visite specialistiche, probabilmente necessitano dell'uso di macchinari costosi oppure di terapie molto onerose. Le

visite specialistiche effettuate alle donne con esenzione per gravidanza costano 273 euro in più rispetto a chi non ha tale esenzione, *ceteris paribus*. La maggior parte delle esenzioni per gravidanza concesse nel 2007 trovano un riscontro positivo nella spesa per visite specialistiche del medesimo anno, questo può portare a influire molto sulla spesa complessiva. Gli esenti per invalidità civile al 100% e gli esenti per invalidità civile >2/3 costano rispettivamente 273 e 106 euro in più rispetto a chi non ha l'una o l'altra esenzione. Nel dettaglio, analizzando i dataset amministrativi a disposizione, si evidenzia che gli esenti per gravidanza svolgono all'anno mediamente 42 visite specialistiche (la mediana è di 39 visite, il range va da 1 a 349, l'1% degli esenti svolge più di 120 visite all'anno), con una spesa media per visita di 510 euro. Gli esenti per malattie neoplastiche maligne effettuano in un anno circa 47 visite (la mediana è di 34 visite, il range delle visite va da 1 a 824, l'1% degli esenti svolge più di 251 visite all'anno), che costano mediamente 804 euro l'una. Inoltre, gli esenti per malattie neoplastiche maligne che hanno svolto una visita specialistica nel 2007 sono molti, poco più di 56.000 contro gli 11.000 esenti per gravidanza. Gli esenti per invalidità civile >2/3 e invalidità civile al 100% effettuano rispettivamente 41 e 47 visite specialistiche all'anno, con un costo medio unitario di 504 e 666 euro. Tutte le altre esenzioni presenti in Tabella 19 hanno una spesa positiva per visite specialistiche ma la grandezza della stima non è molto elevata come nelle esenzioni descritte in precedenza.

Passando alle covariate del medico di medicina generale si noti come gli anni di esperienza del medico non influiscano sulla probabilità degli assistiti di fare o meno una visita specialistica. Sono significative, però, le variabili anni dalla laurea nell'equazione di regressione. Il costo per visite specialistiche aumenta quando l'esperienza del medico è poca, per poi diminuire quando il medico acquista maggior esperienza. Le persone associate a un medico maschio hanno meno probabilità di effettuare visite specialistiche e il livello di spesa corrispondente è poco più basso rispetto a quello degli assistiti associati a medici donna, *ceteris paribus*. I medici che hanno più pazienti rispetto al massimo di legge tendono a prescrivere più visite specialistiche rispetto a chi non supera il massimo di assistiti, *ceteris paribus*. Probabilmente, un medico con troppi pazienti è sbrigativo e tende a delegare il problema prescrivendo visite specialistiche.

Il test di Wald che ipotizza che tutti i coefficienti dell'equazione di regressione siano pari a 0 eccetto l'intercetta porta a rifiutare l'ipotesi nulla. Questo test si distribuisce come una chi-quadro con 28 gradi di libertà. Il test LR che testa l'incorrelazione tra le equazioni porta a

rifiutare l'ipotesi nulla. La correlazione, infatti, è significativa e negativa, come per i ricoveri ospedalieri anche se di grandezza inferiore.

Il modello Tobit II è, perciò, valido e mette in risalto alcuni risultati interessanti che i modelli precedenti non avevano catturato. Studiando separatamente l'equazione di selezione e l'equazione di regressione si colgono i diversi comportamenti degli assistiti a seconda si parli di fare o non fare una spesa o del livello della spesa stessa.

3.4.3 Farmaci

Importo spesa per farmaci		Equazione di regressione (1)		Equazione di selezione (2)	
Variabili esplicative		Coeff	P-value	Coeff	P-value
Sesso	Femmina	-0,055	0,000	0,313	0,000
Età	25-34	0,007	0,018	0,127	0,000
	35-44	0,015	0,000	0,232	0,000
	45-54	0,043	0,000	0,362	0,000
	55-64	0,098	0,000	0,620	0,000
	65-74	0,194	0,000	0,947	0,000
	≥75	0,231	0,000	1,130	0,000
Decesso	2007	-0,178	0,000	-0,551	0,000
	2008	0,072	0,000	0,088	0,000
ADI	Ambulatorio	0,038	0,000	0,539	0,000
	Domicilio	0,152	0,000	0,887	0,000
Esenzioni	Affezioni sistema circolatorio	0,176	0,000	0,597	0,000
	Malattie cardiache	0,216	0,000	0,861	0,000
	Diabete	0,291	0,000	0,922	0,000
	Glaucoma	0,182	0,000	0,862	0,000
	Colesterolo	0,243	0,000	0,865	0,000
	Ipertensione arteriosa con danno d'organo	0,216	0,000	1,033	0,000
	Ipertensione arteriosa senza danno d'organo	0,139	0,000	1,112	0,000
	Malattie neoplastiche maligne	0,177	0,000	0,455	0,000
	Invalidità civile > 2/3	0,221	0,000	0,412	0,000
	Invalidità civile (assegno accomp.)	0,069	0,000	-0,031	0,021
	Invalidità civile al 100%	0,226	0,000	0,282	0,000
	Gravidanza	-0,042	0,000	0,356	0,000
	Altro	0,161	0,000	0,731	0,000
	Costante	0,114	0,000	-0,200	0,000
Sesso medico	Maschio	0,002	0,180	0,016	0,000
	Supera max assistiti	-0,002	0,094	0,066	0,000
	Anni dalla laurea	0,604	0,000	-1,602	0,000
	Anni dalla laurea^2	-0,129	0,000	0,274	0,000
σ		0,480			
ρ		-0,223			

Tabella 20 - Stima del modello di Heckman della spesa per farmaci

La Tabella 20 mostra, infine, le stime del modello di Heckman della spesa per farmaci nel 2007 in Friuli-Venezia Giulia. La significatività e il segno delle stime è molto simile al modello Probit (Tabella 11, colonna 3) per quanto riguarda l'equazione di selezione e al modello di regressione per le sole osservazioni con spesa positiva (Tabella 14, colonna 3) per quanto riguarda l'equazione di regressione. Alcune stime, però, sono completamente diverse, alcune variabili non significative nei modelli precedenti risultano significative nel modello di Heckman, oppure accade il contrario.

Nel dettaglio, le donne hanno più probabilità di acquistare farmaci a parità di altre condizioni, ma, mediamente, spendono 55 euro in meno all'anno rispetto agli uomini. Le dummy dell'età sono tutte significative e positive, la probabilità di acquistare farmaci tende a crescere in modo costante con l'età. Anche il livello della spesa aumenta con l'età, per esempio, gli assistiti compresi nella fascia d'età 65-74 spendono in farmaci 194 euro in più rispetto agli assistiti d'età 14-24 anni, *ceteris paribus*. Gli assistiti che hanno più di 75 anni spendono per farmaci, in media, 231 euro in più rispetto agli assistiti di fascia d'età 14-24. Le dummy che si riferiscono al decesso nel 2007 e nel 2008 sono significative in entrambe le equazioni. Chi è deceduto nel 2007 ha meno probabilità di acquistare farmaci, *ceteris paribus*, e anche il livello di spesa è minore rispetto a chi non è deceduto. Come detto in precedenza, la spesa per farmaci di una persona deceduta nel 2007 può essere calcolata per qualche mese come per tutto l'anno, perciò è inevitabile che chi non è deceduto nel medesimo anno abbia più occasioni di acquistare farmaci. Chi è deceduto nel 2008, infatti, ha più probabilità di acquistare farmaci e il livello della spesa è di 72 euro in più rispetto a chi non è deceduto nel 2008, *ceteris paribus*. Le dummy che si riferiscono al servizio ADI domicilio e ambulatorio sono significativamente e positivamente associate con la variabile risposta in entrambe le equazioni. La spesa per farmaci, però, è più ingente per le persone che sfruttano il servizio ADI a domicilio, essi spendono mediamente 152 euro in più in farmaci rispetto a chi non ha il servizio di assistenza domiciliare integrata. Le stime delle esenzioni sono tutte significative al 5% in entrambe le equazioni. L'esenzione per invalidità civile con assegno di accompagnamento, però, non è significativa al 1%. Le persone con tale esenzione hanno meno probabilità di acquistare farmaci ma il livello della loro spesa è di 69 euro maggiore rispetto a chi non ha tale esenzione. Gli assistiti con esenzione per gravidanza hanno più probabilità di acquistare farmaci rispetto a chi non ha questa esenzione ma, in realtà, il loro livello di spesa è minore, a parità di altre condizioni. Le donne in stato di maternità non possono assumere molti farmaci o, per lo meno, possono assumere solo dei farmaci specifici.

Tutte le altre esenzioni descritte in Tabella 20 sono significativamente e positivamente associate con la variabile risposta. Le persone che hanno l'esenzione per diabete spendono mediamente in farmaci 291 euro in più rispetto a chi non ha l'esenzione, *ceteris paribus*. Nell'arco di un anno, gli esenti per diabete spendono in farmaci circa 837 euro in media. Bisogna ricordare, però, che gli esenti per diabete sono molti, circa 46.000, e questo influisce sicuramente sulla spesa complessiva per farmaci. Gli esenti per colesterolo spendono, in media, circa 243 euro in più rispetto a chi non ha questa esenzione. Gli esenti per colesterolo sono quasi 17.000, molto meno rispetto agli esenti per diabete, e la loro spesa media annua in farmaci è di 802 euro. Un'altra spesa ingente per farmaci è attribuibile agli esenti per invalidità civile al 100%, essi costano 226 in più rispetto a chi non ha l'esenzione. Mediamente, queste persone spendono 819 euro in farmaci all'anno. Altre esenzioni che gravano molto sulla spesa per farmaci sono l'invalidità civile $>2/3$, le malattie cardiache e l'ipertensione con danno d'organo. Gli esenti per malattie neoplastiche maligne, invece, spendono in media all'anno 647 euro per l'acquisto di farmaci, circa 177 euro in più rispetto a chi non ha questa esenzione. Rispetto ai costi per ricoveri ospedalieri e visite specialistiche, si noti come nella spesa per farmaci gli assistiti con tale esenzione non siano classificati tra i più costosi.

Passando alle covariate del medico si noti come, al contrario del modello Tobit standard e Probit, le variabili siano tutte significative in entrambe le equazioni. Una variabile non significativa è il sesso del medico nell'equazione di regressione, cosa evidenziata anche nel modello di regressione con campione ridotto (Tabella 14, terza colonna). Gli assistiti associati a un medico maschio hanno maggior probabilità di acquistare farmaci ma ciò non influisce sul livello della spesa complessiva, a parità di altre condizioni. Un'altra variabile non significativa nell'equazione di regressione ma significativa nell'equazione di selezione è la dummy del medico se supera il massimo numero di assistiti. Le persone associate a un medico con un numero di pazienti che supera il massimo consentito dalla legge tendono ad avere maggior probabilità di acquistare farmaci. L'andamento parabolico delle variabili "anni dalla laurea" è interessante. L'equazione di selezione evidenzia che, nei primi anni di esperienza del medico i pazienti hanno meno probabilità di acquistare farmaci, questa probabilità aumenta, però, da una certa età del medico in poi quando si è acquisita una certa esperienza lavorativa. Se analizziamo l'equazione di regressione, però, nonostante i medici con pochi anni di esperienza tendano a prescrivere meno farmaci, la spesa per farmaci dei loro pazienti aumenta di 604 euro all'aumentare di un anno di esperienza del medico. Questa tendenza si inverte quando il medico, dopo molti anni di servizio, acquisisce una buona esperienza e la spesa per farmaci dei

loro pazienti diminuisce con il passare degli anni. Questo risultato non mi sorprende, probabilmente i medici giovani sono restii a prescrivere farmaci, magari per timore di sbagliare cura. I medici più adulti sono più esperti perciò hanno la chiarezza e la lucidità di prescrivere i farmaci consoni. Nel momento in cui, però, i medici con poca esperienza prescrivono farmaci, probabilmente il costo della singola medicina è molto elevato, oppure impongono di non usare farmaci generici e di comprare quelli originali, sicuramente più costosi.

Il test di Wald che ipotizza che tutti i coefficienti dell'equazione di regressione siano pari a 0 eccetto l'intercetta porta a rifiutare l'ipotesi nulla. Questo test si distribuisce come una chi-quadro con 28 gradi di libertà. Il test LR che testa l'incorrelazione tra le equazioni porta a rifiutare l'ipotesi nulla. La correlazione, infatti, è ancora una volta significativa e negativa.

Anche in questo caso, i risultati del modello di Heckman è valido ci indicano che studiare separatamente il comportamento degli assistiti circa il comprare-non comprare un farmaco e circa il livello della spesa a seconda delle caratteristiche degli assistiti, porta a risultati importanti e differenti dai modelli svolti in precedenza. Anche le covariate del medico hanno fatto emergere risultati interessanti, studiare le caratteristiche dei medici ha aiutato a capire il comportamento degli assistiti per quanto riguarda la spesa per farmaci.

Capitolo 4

Previsione

L'analisi svolta porta alla conclusione che le stime ottenute a partire dal modello di Heckman sono quelle che meglio si adattano ai dati a disposizione. Perciò, partendo da questo presupposto, in questo capitolo si vuole proporre un esercizio di previsione dei costi (spesa per ricoveri, visite specialistiche, farmaci) dal 2007 al 2060, utilizzando, appunto, le stime del modello di Heckman e i profili degli assistiti.

I profili utilizzati si riferiscono a donne e uomini suddivisi per le diverse fasce d'età specificate nei modelli. Per ogni profilo viene calcolata la proporzione di persone che hanno determinate caratteristiche, ovvero quante hanno un particolare tipo di esenzione, quante hanno usufruito del servizio ADI, etc. Queste proporzioni verranno moltiplicate per le stime corrispondenti in modo tale da ottenere la previsione per ogni profilo. La previsione ottenuta sarà, quindi, moltiplicata per il totale di categoria e successivamente riscalata per le sole persone che hanno avuto una spesa positiva poiché, nell'equazione di regressione, rientrano solo le osservazioni che nell'equazione di selezione avevano una certa probabilità di osservare tale spesa. La somma della spesa ottenuta per tutte le varie categorie sarà la previsione della spesa totale (per ricoveri, o per visite specialistiche o per farmaci) dell'anno 2007 in Friuli-Venezia Giulia.

La previsione effettuata in questo elaborato, però, non si sofferma al solo anno di riferimento, ovvero il 2007, ma si protrae fino al 2060. Per fare ciò, si è usufruito della previsione della popolazione svolta dall'Istat fino al 2060, che è stata suddivisa per Regione e fasce d'età. Le previsioni demografiche dell'Istat sono predisposte in ragione di standard metodologici riconosciuti in campo internazionale. In particolare, si ricorre al modello per componenti (cohort component model), secondo il quale la popolazione, tenuto conto del naturale processo di avanzamento dell'età, si modifica da un anno al suo successivo sulla base del saldo naturale (differenza tra nascite e decessi) e del saldo migratorio (differenza tra movimenti migratori in entrata e in uscita). Le previsioni sono articolate secondo tre distinti scenari: alto,

centrale, basso. Lo scenario utilizzato in questa analisi è quello definito centrale. Esso comprende un set di stime puntuali ritenute “verosimili” che, costruite in base alle recenti tendenze demografiche, rappresentano quelle di maggiore interesse per gli utilizzatori poiché calcolate sullo scenario considerato più “probabile”¹⁰.

Per realizzare questo esercizio di previsione sono state fatte alcune assunzioni importanti. La prima assunzione è che la *proporzione* di persone con determinate caratteristiche (presenza di esenzioni, ricorso ad assistenza domiciliare, etc.) rimanga sempre la stessa all’interno di ogni classe d’età, per maschi e femmine e nel corso degli anni. Sicuramente, con il passare del tempo, il *totale* di categoria dei profili cambierà, come cambierà anche il numero di persone che hanno accesso a un ricovero, effettuato una visita o acquistato un farmaco. Questo è dovuto, infatti, a un cambiamento demografico che le previsioni Istat tengono in considerazione. Le proporzioni invece non vengono influenzate.

La seconda assunzione è che non cambi nemmeno la *proporzione* di persone di un determinato profilo che hanno una spesa positiva (anche in questo caso però il *totale* delle persone che sostiene una spesa positiva può variare in funzione del cambiamento demografico).

Infine, le previsioni vengono ottenute a partire dalle relazioni emerse con la stima del modello di Heckman, relazioni che riguardano il 2007 e che si suppone quindi continuino a valere anche per gli anni successivi.

Siano Y_j i vari profili (ovvero gli uomini e le donne suddivisi per le diverse fasce d’età) con $j=1,\dots,14$ e x_i le caratteristiche (esenzioni, ADI, etc.) delle varie osservazioni con $i=1,\dots,N$. Sulla base delle evidenze del 2007, definiamo:

- n_{ji} = numero di persone del profilo J con caratteristica x_i
- N_j = numero totale di persone del profilo J
- $\tau_{ji} = \frac{n_{ji}}{N_j}$ = proporzione di persone del profilo J con caratteristica x_i
- k_j = numero di persone del profilo J che hanno avuto una spesa positiva
- $\delta_j = \frac{k_j}{N_j}$ = proporzione di persone del profilo J che hanno avuto una spesa positiva

Il primo scenario ipotizzato è che, appunto, nel corso degli anni non cambi τ_{ji} e neppure δ_j , mentre N_j può cambiare nel tempo (si veda Appendice A).

¹⁰ www.demo.istat.it

Sulla base di queste assunzioni, valori e delle stime ottenute con il modello di Heckman, vengono previste le spese totali della regione Friuli-Venezia Giulia per ricoveri ospedalieri, visite specialistiche e farmaci.

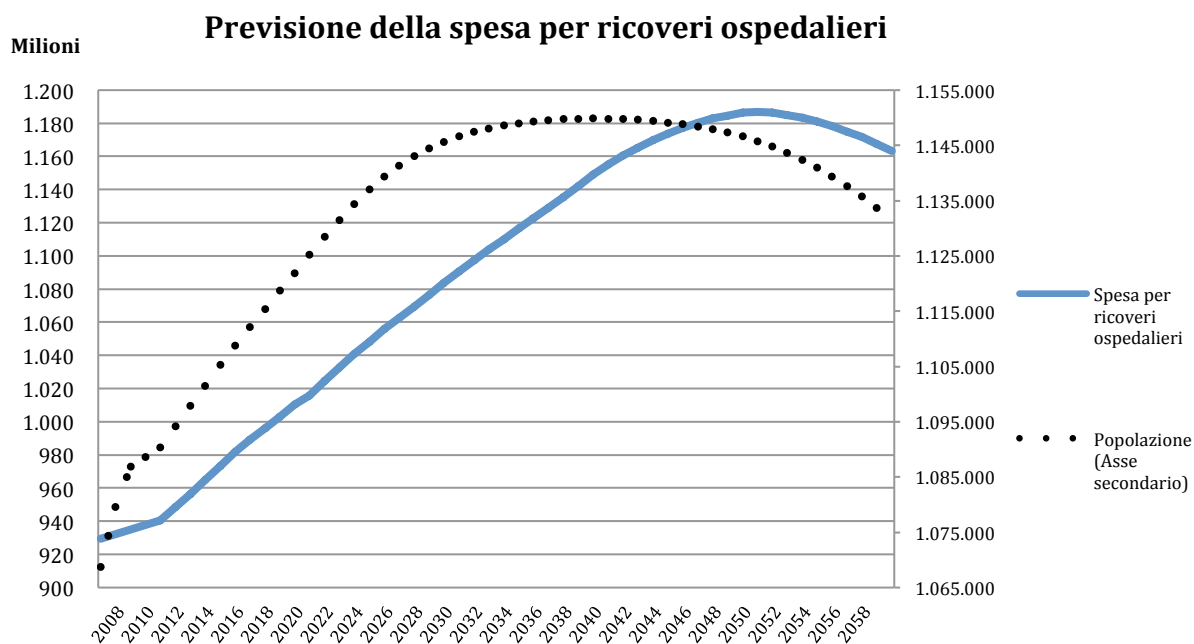


Figura 9 - Previsione della spesa per ricoveri ospedalieri e popolazione totale

In Figura 9 viene riportato l'andamento della spesa per ricoveri ospedalieri nell'arco degli anni 2007-2060, che dimostra essere di tipo parabolico: la spesa aumenta fino ad arrivare a un picco di 1,190 miliardi di euro nel 2050, per poi diminuire. Dal 2007 al 2050 la spesa aumenta di quasi 261 milioni di euro. Notiamo, quindi, un effetto spesa e un effetto demografico. La spesa aumenta perché la popolazione cresce. La popolazione friulana dal 2034 al 2046 sarà sempre maggiore di 1.145.000, anche se il picco si raggiungerà nel 2040, quando la popolazione prevista sarà di 1.149.000 persone. Questo punto non coincide, però, con l'anno del punto massimo della spesa per ricoveri ospedalieri, sembra che la curva dei costi sia traslata di 10 anni. L'andamento delle due curve è lo stesso e ci si aspetterebbe, quindi, una diminuzione dei costi per ricoveri negli stessi anni in cui c'è una diminuzione della popolazione.

Per spiegare questa discrepanza, nella Figura 10 è introdotto l'andamento della popolazione di età maggiore di 75 anni.

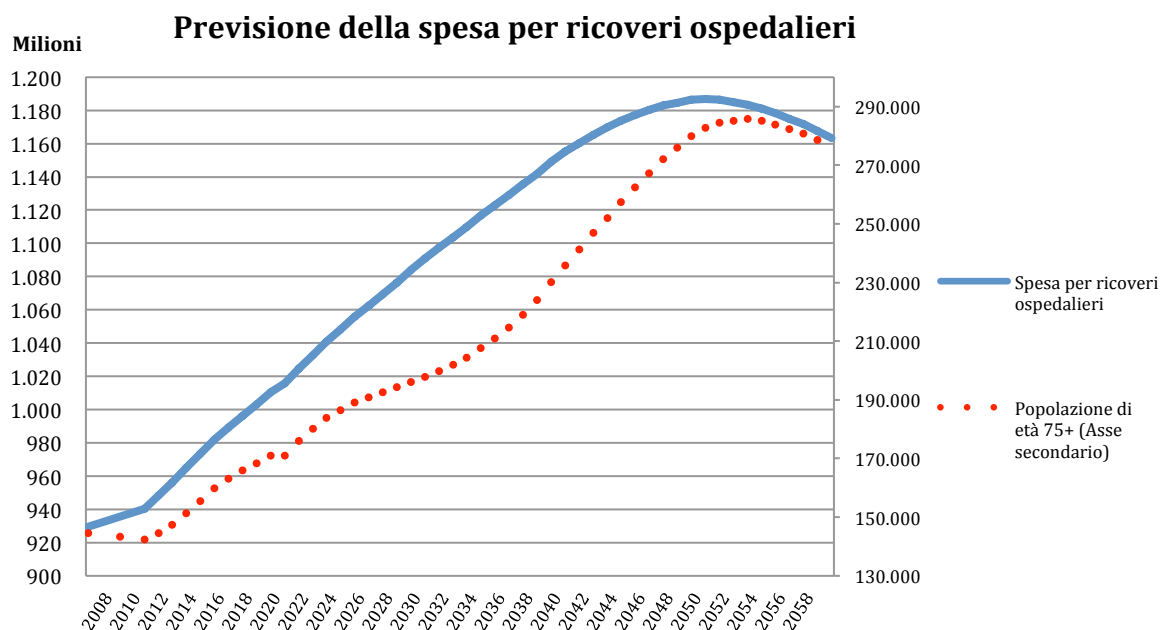


Figura 10 - Previsione della spesa per ricoveri ospedalieri e popolazione over 75

E' bene, infatti, introdurre un aspetto demografico importante, tale da spiegare l'andamento della popolazione e dei costi. Dopo molti anni in cui il tasso di natalità in Italia si aggirava sul 17%, negli anni '60 c'è stato un boom di nascite che hanno fatto raggiungere il tasso di natalità quasi al 20%. L'Italia in quegli anni era caratterizzata da un miracolo economico che ha portato benessere e vita. Le persone nate nel periodo definito baby-boom, ora 40-50 anni, arriveranno nel 2040 ad essere categorizzate come anziane. L'aspettativa di vita aumenta con il passare degli anni e perciò tra 50 anni ci troveremo ad avere molte persone di una certa età che necessitano di cure. L'Italia impiega un secolo, dal 1861, per raddoppiare gli abitanti; raggiungendo i 50 milioni nel 1961; poi, in oltre 40 anni, cresce solo di 8,6 milioni, di cui quasi la metà immigrati. L'aumento degli italiani dal 1965 è sostenuto dalla crescita degli over 65 anni che compensa la carenza di nascite. Un segnale di invecchiamento del Paese è quello dell'indice di vecchiaia (persone over 65 anni ogni 100 bambini 0-14 anni). Questo indice sale da 31 nel 1951 a 142 nel 2006. Ciò significa che, nel 2006, in Italia ci sono 142 over 65 anni contro 100 bambini under 15 anni (Meneghelli, 2008).

La crescita della popolazione è perciò dovuta al fatto che cresce il numero di anziani. Il picco degli over 75 in Friuli-Venezia Giulia dovrebbe arrivare nell'anno 2053, 286.000 persone, per poi diminuire, probabilmente perché verranno a mancare quelle persone nate nel periodo del baby-boom. Questo aumento della popolazione porta, inevitabilmente, ad un aumento ingente dei costi per la Sanità, soprattutto la spesa per ricoveri ospedalieri. Infatti, la curva

della spesa cresce pari passo con la crescita della popolazione per poi diminuire all'incirca nello stesso anno in cui diminuisce la curva della popolazione over 75. La spesa per ricoveri ospedalieri è difatti largamente influenzata dalle persone anziane.

Un andamento dei costi simile si riscontra anche nella previsione di spesa per visite specialistiche e per farmaci (Figura 11 e Figura 12 rispettivamente).

Previsione della spesa per visite specialistiche

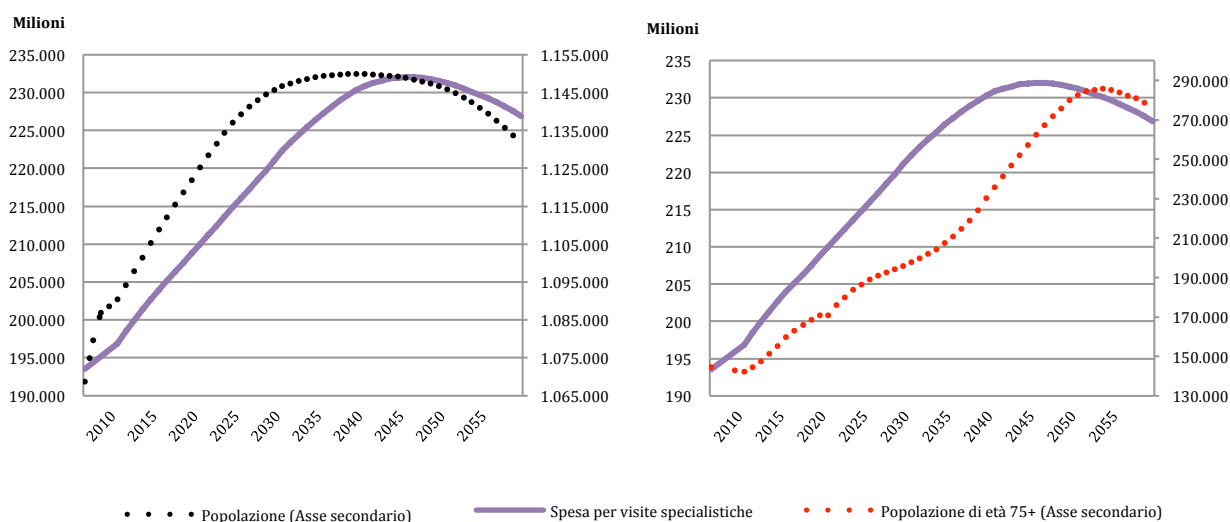


Figura 11 - Previsione della spesa per visite specialistiche, popolazione totale e over 75

L'andamento della previsione della spesa per visite specialistiche è crescente fino a raggiungere un massimo intorno all'anno 2045 di circa 232 milioni di euro. La spesa dal 2007 al 2045 aumenta di circa 40 milioni di euro. Si noti come la curva della popolazione totale raggiunga il suo massimo nell'anno 2040 e la curva della popolazione over 75 nell'anno 2055. Una compensazione tra queste due curve fa in modo che il picco della spesa per visite specialistiche sia nel 2045. In tutti i modelli svolti nell'analisi, si evidenziava come il costo della spesa per visite specialistiche crescesse con l'età. Più il nostro Paese invecchia più ci saranno costi a carico della Sanità Pubblica.

Previsione della spesa per farmaci

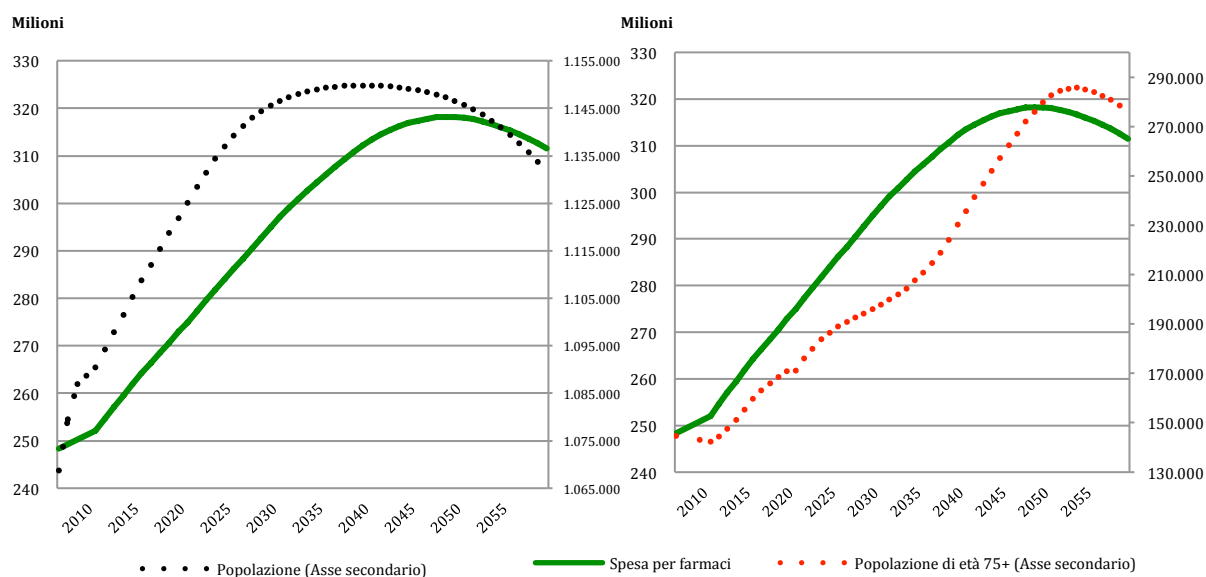


Figura 12 - Previsione della spesa per farmaci, popolazione totale e over 75

Anche la spesa per farmaci ha un andamento crescente, raggiungendo il massimo intorno all'anno 2050 con una spesa pari a quasi 320 milioni di euro. Dal 2007 al 2050 la spesa per farmaci aumenta di circa 70 milioni di euro. Come per la spesa per visite specialistiche, il massimo della spesa per farmaci avviene 10 anni dopo il picco della popolazione totale e qualche anno prima del picco della popolazione over 75. La spesa per visite specialistiche e farmaci è chiaramente influenzata dall'età dei pazienti over 75, ma non così pesantemente come la spesa per ricoveri ospedalieri. Ci saranno altre fasce d'età che compensano il picco della spesa per farmaci, a prescindere dagli anziani.

I grafici della previsione finora analizzati rappresentano il valore assoluto dei costi e il loro andamento nel tempo. Questo, però, non permette di confrontare quanto un costo è cresciuto rispetto ad un altro. Per questo motivo, nella Figura 13 viene calcolata la variazione percentuale della spesa per ricoveri ospedalieri, visite specialistiche e farmaci rispetto all'anno base 2007.

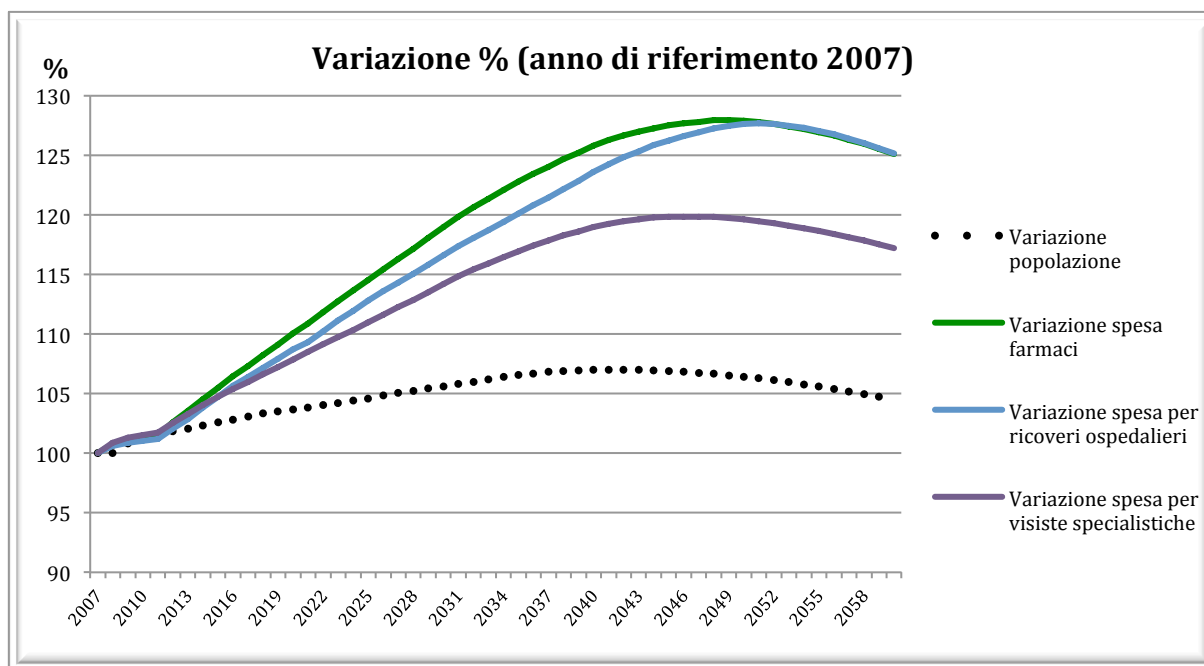


Figura 13 - Variazione % della previsione dei costi per ricoveri ospedalieri, visite specialistiche, farmaci

Dalla Figura 13 è evidente come i costi crescano molto di più a livello percentuale rispetto alla popolazione totale. La popolazione dal 2007 al 2060 cresce al massimo del 7% per poi diminuire dagli anni 2050 in poi. La spesa per visite specialistiche raggiunge la crescita massima del 20% rispetto al 2007 intorno agli anni 2040, per poi diminuire. Si noti come, più passano gli anni più la spesa per visite specialistiche cresce, anche se, in realtà, questa è la spesa che cresce di meno in termini percentuali rispetto agli altri due tipi di costi. Fino al 2014 le tre curve sembrano crescere in modo uguale ma, negli anni successivi, la spesa per farmaci aumenta molto di più rispetto alle altre: infatti è la curva più alta e rimane tale fino al raggiungimento del suo massimo di spesa (+27%) nel 2046. Probabilmente, molte persone tenderanno ad acquistare sempre più farmaci, come per esempio le persone giovani che magari un tempo si avvicinavano con distacco all'uso di prodotti farmaceutici, ma nel futuro il farmaco sarà, magari, indispensabile per curare problemi più o meno gravi. La curva relativa ai ricoveri ospedalieri riflette un andamento crescente dei costi. Rispetto alla curva della previsione dei costi per farmaci, la previsione dei costi per ricoveri cresce più lentamente, però interseca la curva dei farmaci nel 2046 in cui ambedue le curve presentano un aumento della spesa del 27% in più rispetto al 2007. Non è un caso che le spese per ricoveri e per farmaci siano, in realtà, quelle che crescono di più, poiché sono quelle che risentono maggiormente dell'invecchiamento della popolazione, come evidenziato nelle stime finali del modello di

Heckman. Nel 2050 le persone over 75 saranno 285 mila e rappresenteranno circa il 28% della popolazione friulana. Le persone anziane aumenteranno, quindi, del 97% rispetto al 2007.

Da questi grafici è evidente come la spesa Sanitaria Pubblica aumenterà molto nei prossimi cinquant'anni, cosa dovuta all'aumento della popolazione ma soprattutto all'invecchiamento della stessa.

Lo scenario finora analizzato ipotizza che nel corso degli anni non cambi τ_{ji} (proporzione di persone del profilo J con caratteristica x_i) e neppure δ_j (proporzione di persone del profilo J che hanno avuto una spesa positiva). Che cosa accadrebbe se, nel corso degli anni, diminuisse/aumentasse la probabilità δ_j di avere una spesa positiva (ad esempio, se venissero posti in essere provvedimenti legislativi con lo scopo di limitare la spesa pubblica)? Che effetto avrebbe sulla spesa futura? La Tabella 21 mostra alcuni esempi.

	Spesa reale 2007 (milioni di euro)	Previsione 2007 (milioni di euro)	δ_j	
			-1,0%	+1,0%
Ricoveri	716	929	-8,55%	8,55%
Visite specialistiche	186	193	-1,48%	1,48%
Farmaci	235	248	-1,36%	1,36%

Tabella 21 - Stima reale, previsione e variazione di δ_j

La Tabella 21 evidenzia, innanzitutto, come la previsione ottenuta a partire dalle stime del modello di Heckman per il 2007 sovrastimi la spesa per ricoveri, visite specialistiche e farmaci. La spesa reale del 2007 dei farmaci per la popolazione analizzata in questa tesi è di circa 235 milioni di euro, la previsione è di 248 milioni di euro. La previsione sovrastima la spesa di circa 13 milioni di euro, che corrisponde al 5,5% in più della spesa reale. Le variabili utilizzate per spiegare il modello hanno portato a una buona stima della previsione. Minore è la sovrastima della spesa per visite specialistiche: la spesa reale è di 186 milioni di euro e la previsione è di 193 milioni di euro. La stima della previsione conta 7 milioni di euro in più rispetto alla spesa reale, che corrisponde a circa il 3,7%. In generale, per quanto riguarda spesa per farmaci e visite specialistiche, il modello si adatta discretamente ai dati, le variabili utilizzate spiegano bene la variabile risposta e ciò porta a delle previsioni di spesa abbastanza

accurate. Il tutto considerando un aspetto molto importante: tutte queste analisi sono basate su dataset di tipo amministrativo, all'interno dei quali sono completamente assenti informazioni di carattere economico quali il titolo di studio, l'occupazione o i redditi degli assistiti.

Una differenza maggiore si riscontra, invece, nella previsione della spesa per ricoveri ospedalieri. La spesa reale nel 2007 è di 716 milioni di euro e la previsione è di 929 milioni di euro. La differenza tra le due è di 213 milioni di euro, la previsione sovrastima la spesa reale di circa il 29,7%. Si noti, però, che, a differenza degli altri due tipi di spesa, la numerosità del campione della spesa per ricoveri ospedalieri è molto ridotta, solo l'11% della popolazione ha eseguito un ricovero ospedaliero nel 2007 in Friuli-Venezia Giulia. Inoltre, un ricovero ospedaliero può dipendere sia da altre caratteristiche degli assistiti (le già citate titolo di studio, reddito familiare, professione, etc.), sia da fenomeni esterni legati all'insorgere di certe malattie o problemi che necessitano un ricovero ospedaliero.

L'analisi della previsione si conclude ipotizzando diversi scenari: se la probabilità dei vari profili di avere una spesa positiva cambiasse, come varierebbe la previsione della spesa futura?

Per spiegare ciò, si ipotizzi che lo Stato attui delle politiche restrittive circa l'accesso ai servizi ospedalieri oppure limiti le esenzioni per certe patologie o, magari, al contrario, conceda esenzioni per altri tipi di patologie non prese in considerazione fino ad oggi. Inoltre, una politica restrittiva può essere legata ai MMG, i quali possono essere disincentivati a prescrivere visite specialistiche cercando, invece, di essere più meticolosi nel rapporto medico-paziente e nelle visite in ambulatorio (tale comportamento potrebbe evitare controlli specialistici magari inappropriati).

Nel dettaglio, se la proporzione di persone che hanno una spesa positiva per farmaci diminuisce dell'1%, la spesa futura diminuirebbe dell'1,3%. Se la stessa costante venisse, al contrario, aggiunta alla proporzione δ_j , la spesa aumenterebbe della stessa quantità precedente, ovvero dell'1,3%. L'aumento o la diminuzione della spesa per farmaci in termini di spesa non è molto elevato rispetto all'aumento/diminuzione della proporzione di persone che hanno acquistato farmaci. Passando alle visite specialistiche, un aumento dell'1% della proporzione di persone che hanno accesso alle visite specialistiche porta ad un aumento della spesa di quasi 1,5%. Anche in questo caso, come nella spesa per farmaci, un aumento/diminuzione della

proporzione delle persone che hanno accesso alle visite specialistiche porta ad una maggiore/minore spesa rispetto al tasso applicato alle proporzioni ma comunque contenuta.

Molto diverso è il comportamento della spesa per ricoveri al variare di δ_j . Un aumento della probabilità di avere una spesa positiva dell'1% porta ad un aumento della spesa per ricoveri dell'8,5%. Questo dato non è sorprendente perché sono solo 134.629 le persone che hanno avuto un ricovero ospedaliero nel 2007 e rappresentano l'11% della popolazione friulana. Una variazione, anche minima, della proporzione di persone che hanno avuto una spesa positiva porta a un'ingente differenza in termini di spesa. Questo dimostra come, anche se poche persone hanno effettuato ricoveri ospedalieri, in realtà il ricovero è il costo più ingente per la Sanità Pubblica. Una piccola variazione delle persone che hanno accesso ai ricoveri porta ad un grande cambiamento in termini di spesa.

Un aspetto molto importante che nell'esercizio di previsione non è stato incluso è la stima dello standard error, che avrebbe permesso di calcolare degli intervalli di previsione, di notevole interesse economico. L'esercizio di previsione da me svolto, però, presenta diversi scenari, e la stima dello standard error risulta complicata da calcolare. Le uniche stime reali dei costi a disposizione sono quelle dell'anno 2007, la previsione si protrae fino al 2060. Inoltre, per svolgere l'esercizio di previsione mi sono basata sulle stime ottenute dal modello di Heckman, che presenta due equazioni, la prima stima la probabilità di successo/insuccesso, la seconda il livello della spesa. Oltre a ciò, non ho a disposizione la previsione di ogni singola osservazione, bensì solo per le categorie a cui ho fatto riferimento (uomini e donne suddivisi per classi d'età). Per tutti questi motivi, il calcolo di un possibile intervallo di previsione risulta complicato, ma sicuramente di notevole interesse scientifico ed economico, potrebbe quindi essere uno spunto per analisi future.

Conclusioni

Uno dei principali obiettivi che ogni società civile si pone è la tutela della salute dei cittadini e il suo grado di attuazione misura in gran parte il benessere di cui gode la popolazione. In Italia vige il Servizio Sanitario Nazionale (SSN), un sistema di strutture e servizi che hanno lo scopo di garantire a tutti i cittadini, in condizioni di uguaglianza, l'accesso universale all'erogazione equa delle prestazioni sanitarie. Il SSN è finanziato attraverso la fiscalità generale tramite imposte dirette ed indirette, ed è il CIPE (Comitato Interministeriale per la Programmazione Economica) che ripartisce le risorse tra le Regioni, su proposta del Ministero della Salute. Il riparto viene determinato sulla base della numerosità della popolazione residente e ponderato anche in relazione ai consumi specifici sanitari per età. Per questo motivo è molto importante studiare la spesa pubblica sanitaria, anche perché è una delle voci più ingenti che pesano sul bilancio del Paese. Ogni Regione dovrebbe cercare di stilare il profilo degli assistiti più "onerosi" e caratterizzare il comportamento dei Medici di Medicina Generale, in modo tale da catturare sprechi o comportamenti scorretti. Studiare la provenienza dei costi e il loro andamento nel tempo può portare a realizzare delle efficaci politiche economiche di contenimento della spesa pubblica.

In questo elaborato sono stati studiati tre tipi di costi sanitari (ricoveri ospedalieri, visite specialistiche e farmaci), inerenti la regione Friuli-Venezia Giulia per l'anno 2007, tramite l'analisi di dataset di tipo amministrativo.

La popolazione di questa regione conta quasi 1.200.000 di persone, ma l'analisi svolta si è concentrata solo sui cittadini di età maggiore di 14 anni, in quanto le cure e le visite legate ai bambini possono essere diverse e più particolari rispetto a quelle degli adulti. La percentuale di spesa attribuibile ai bambini è veramente bassa, ovvero il 3.1% della spesa per ricoveri ospedalieri, il 2.3% della spesa per visite specialistiche e l'1.1% della spesa per farmaci.

Dalle analisi esplorative sono emersi degli aspetti importanti. Solamente l'11% del totale della popolazione ha effettuato un ricovero ospedaliero nel 2007, a fronte di un 62% circa dei cittadini che ha fatto una visita specialistica e ad un 64% circa di chi ha acquistato almeno un farmaco. In questa tesi sono stati definiti e stimati diversi modelli statistici, ed è interessante

evidenziare come, partendo dallo stesso dataset, modelli differenti (e via via più complessi) abbiano riportato stime differenti dei parametri associati ad alcune variabili. Le variabili esplicative di base inserite nei vari modelli sono le caratteristiche degli assistiti e, se significative, sono state introdotte anche le covariate del medico di medicina generale.

Il primo modello analizzato è un modello per variabili risposta di tipo binario (aver effettuato una spesa oppure no): il modello probit. Sono stati stimati i modelli probit univariati e probit multilivello, con e senza covariate del medico, per cercare di catturare e valutare, se esiste, l'effetto della gerarchia nei dati. Nel modello probit per ricoveri ospedalieri le variabili dei MMG non risultano significative, al contrario di quello che accade nel modello della spesa per farmaci e visite specialistiche. In effetti, è il medico che prescrive le ricette per acquistare farmaci o per effettuare visite specialistiche, mentre per quanto riguarda i ricoveri ospedalieri, il medico di base può avere un ruolo più marginale. Circa il 40% dei ricoveri è di tipo urgente, il restante 60% si suddivide tra day hospital, ricoveri ordinari o programmati con pre-ospedalizzazione. Questi tipi di ricoveri ospedalieri possono essere prescritti dal medico di base, ma anche dal medico specialista previa visita ambulatoriale. Entrando nel dettaglio dei modelli, a parità di altre condizioni, le donne hanno maggiore probabilità di eseguire un ricovero, una visita specialistica o acquistare farmaci, rispetto agli uomini. Questa probabilità aumenta con il passare degli anni e, in generale, gli assistiti che hanno almeno un'esenzione hanno un impatto significativo e positivo sulla variabile risposta, per tutti e tre i tipi di spesa. Le esenzioni possono essere viste come proxy della salute delle persone e quanto una patologia sia diffusa nel territorio. Le persone decedute nel 2007 o nel 2008 hanno più probabilità di aver subito un ricovero ospedaliero; la tendenza inversa caratterizza, invece, le visite specialistiche e i farmaci. Si può ipotizzare che, una persona negli ultimi istanti di vita sia stata ricoverata d'urgenza e perciò le spese per le cure come farmaci o visite rientrano nella spesa per ricoveri ospedalieri.

Il secondo tipo di modello preso in considerazione è il modello di regressione lineare e multilivello, con e senza covariate del medico, considerando solamente le osservazioni che hanno avuto una spesa positiva. Infatti, in questo caso la variabile dipendente è di tipo continuo e si riferisce ai tre tipi di spesa (ricoveri ospedalieri, visite specialistiche e farmaci). In questo modo si è in grado di valutare le caratteristiche degli assistiti che più gravano sulla Sanità Pubblica. Analogamente alle conclusioni delle stime del modello probit, anche nel modello di regressione lineare della spesa per ricoveri ospedalieri le covariate del medico non

risultano statisticamente significative. Analizzando l'effetto delle singole variabili emergono delle interessanti differenze rispetto alle stime del modello probit, *ceteris paribus*. Per esempio, le donne hanno maggiore probabilità di aver effettuato un ricovero ospedaliero, ma, se teniamo conto delle sole persone che hanno fatto un ricovero, gli uomini costano di più alla Sanità Pubblica. Per quanto riguarda le esenzioni, le persone che ricevono l'assegno di accompagnamento hanno più probabilità di effettuare un ricovero o una visita specialistica rispetto a chi non ha l'esenzione, ma in termini di magnitudo spendono meno. Le donne in stato di gravidanza hanno una probabilità maggiore di fare visite specialistiche o ricoveri ospedalieri, ma spendono meno in termini di prodotti farmaceutici. Per quanto riguarda le variabili delle caratteristiche del MMG è emerso che gli assistiti associati a un medico con pochi anni di esperienza costano di più in termini di visite specialistiche e farmaci e con il passare degli anni, questa tendenza mostra una inversione.

Il modello introdotto successivamente è il modello Tobit standard. I valori assunti dalle variabili dipendenti analizzate sono nulli per una parte rilevante della popolazione, ma positivi per la parte restante. In questo caso, il valore "0" non indica un valore "mancante" per le variabili dipendenti, bensì rappresenta un preciso comportamento degli assistiti e quindi ha un particolare significato economico. L'informazione contenuta nel modello tobit combina quella contenuta nel modello di regressione troncata, che descrive l'ammontare della spesa, con quella del modello probit che descrive la spesa/non spesa degli assistiti. Nei modelli stimati emerge come il segno dei parametri sia lo stesso del modello probit, e come le stime del modello tobit riscalate per $1/\sigma_{\text{tobit}}$ si avvicinano molto a quelle del modello probit. Il modello Tobit standard è stato confrontato con il modello Tobit multivariato, in modo tale da analizzare una possibile correlazione tra le variabili risposta. Il risultato emerso è una correlazione bassa e le stime tra i due modelli molto simili.

Il modello Tobit presenta, però, alcuni limiti: per costruzione, le variabili che determinano la probabilità di un'osservazione nulla hanno esattamente la stessa relazione delle variabili che influenzano il livello di un'osservazione positiva. È stato introdotto, quindi, un'estensione del modello Tobit, il modello di Heckman, o Tobit II, che permette di specificare un'equazione per la scelta di una spesa ed una per il livello della spesa, se positiva. A parità di altre condizioni, il modello di Heckman evidenzia come le donne abbiano maggiore probabilità di aver effettuato un ricovero, una visita specialistica o aver acquistato un farmaco. In termini di grandezza della spesa, però, gli uomini spendono di più rispetto alle donne, per tutti e tre i tipi

di costo. L'età è una variabile molto importante poiché mostra come la spesa tenda a crescere con l'età. Le persone decedute nel 2007 o nel 2008 hanno più probabilità di aver effettuato un ricovero ma meno probabilità di aver fatto visite specialistiche o acquistato farmaci. Le persone che usufruiscono del servizio di assistenza domiciliare costano di più rispetto a chi non ha questo servizio; infatti sono persone con problemi di salute o che non sono in grado di deambulare senza l'aiuto di terzi. Per quanto riguarda le esenzioni, l'esenzione per malattie cardiache e per malattie neoplastiche maligne sono quelle che costano di più in termini di spesa per ricoveri ospedalieri. Le donne con l'esenzione per gravidanza hanno sì più probabilità di essere ricoverate ma, in media, spendono meno in ricoveri ospedalieri rispetto a chi non ha questa esenzione. Se si analizza la spesa per visite specialistiche, le donne in stato di gravidanza hanno più probabilità di fare visite specialistiche e il livello della spesa è elevato. Ciò nonostante, la spesa per visite specialistiche più costosa è attribuibile alle esenzioni per malattie neoplastiche maligne. Per quanto riguarda la spesa per farmaci, gli assistiti con l'esenzione per diabete sono coloro che spendono di più in termini di prodotti farmaceutici, seguiti poi dagli esenti per colesterolo.

Il modello di Heckman sembra adattarsi meglio ai dati e per tale motivo i suoi risultati sono stati usati per realizzare una previsione fino al 2060. La previsione dei tre costi sembra sovrastimare la spesa reale, soprattutto nel modello per i ricoveri ospedalieri. Nonostante si preveda che la popolazione friulana cresca del 7% nel giro di 50 anni, le stime delle spese per ricoveri e farmaci dovrebbero crescere del 27%, mentre quelle per visite specialistiche del 20%. Questo è dovuto a un invecchiamento del Paese e l'età è un fattore importante che influenza molto i costi per la Sanità. Il Friuli-Venezia Giulia nel 2050 conterà quasi il doppio delle persone over 75 rispetto al 2007. Sono state fatte alcune ipotesi circa una variazione della proporzione di persone che hanno accesso ai ricoveri ospedalieri, visite specialistiche o farmaci. Una variazione del $\pm 1\%$ non cambierebbe di molto la spesa per visite specialistiche o per farmaci, ma i ricoveri ospedalieri ne risentirebbero certamente di più: se la proporzione di persone che hanno accesso ai ricoveri ospedalieri diminuisse dell'1%, la spesa futura diminuirebbe dell'8,5% rispetto alla previsione.

Concludo questo elaborato evidenziando alcuni aspetti interessanti. Nonostante la mancanza di variabili importanti (come titolo di studio, reddito, status occupazionale) le previsioni ottenute a partire da dataset amministrativi sono molto vicine a quelle reali per quanto riguarda visite specialistiche e farmaci. Per tale motivo, sarebbe molto importante basare le

analisi future su dataset amministrativi che hanno il vantaggio di essere facilmente reperibili e a costo zero, anche se la tempestività non è delle migliori. L'analisi svolta in questa tesi si basa su dati del 2007 che sono stati resi disponibili solo 3 anni dopo il periodo di riferimento.

Inoltre, il dataset fa riferimento ad un periodo non propriamente recente, all'interno del quale certi provvedimenti legislativi non erano ancora stati posti in essere o avevano appena cominciato ad esserlo.

Infine, i risultati ottenuti in questo elaborato sono fortemente influenzati da alcune peculiarità del Friuli-Venezia Giulia, la quale, rispetto ad altre regioni, è sempre stata particolarmente attenta al contenimento delle spese. L'approccio adottato, però, potrebbe essere applicato ad altre regioni italiane e questo permetterebbe di catturare le diversità di un territorio rispetto ad un altro.

Appendice A

Proporzione di persone del profilo J con caratteristiche x_i (τ_{ij}) e proporzione di persone del profilo J che hanno avuto una spesa positiva (δ_j)

T_{ij}	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10	Y11	Y12	Y13	Y14
	F 14-24	F 25-34	F 35-44	F 45-54	F 55-64	F 65-74	F +75	M 14-24	M 25-34	M 35-44	M 45-54	M 55-64	M 65-74	M +75
Decesso 2007	0,000212	0,000379	0,000703	0,001530	0,004394	0,010340	0,060764	0,000313	0,000728	0,001160	0,002907	0,007781	0,019806	0,076074
Decesso 2008	0,000077	0,000379	0,000796	0,001855	0,004430	0,010250	0,062630	0,000553	0,000768	0,001080	0,002990	0,009503	0,021356	0,076472
ADI ambulatorio	0,002676	0,004198	0,005400	0,007541	0,013979	0,037487	0,039802	0,001457	0,001814	0,003080	0,006086	0,013086	0,036672	0,051160
ADI domicilio	0,001136	0,001741	0,002534	0,004771	0,011480	0,032900	0,180669	0,001438	0,002079	0,002580	0,004840	0,011691	0,032079	0,138616
Esen. Sistema circolatorio	0,000847	0,001362	0,002514	0,005048	0,011963	0,025696	0,040672	0,000940	0,001271	0,002890	0,008661	0,028033	0,054115	0,067975
Esen. Malattie cardiache	0,001078	0,001502	0,002979	0,005915	0,016441	0,023583	0,031861	0,001696	0,001563	0,004410	0,014759	0,042074	0,055858	0,058105
Esen. Diabete	0,002984	0,005869	0,008161	0,016191	0,046909	0,079121	0,088490	0,005716	0,004105	0,009709	0,032934	0,089918	0,131476	0,122219
Esen. Glaucoma	0,000135	0,000534	0,001200	0,004036	0,013435	0,029000	0,045687	0,000442	0,000622	0,001540	0,004473	0,012885	0,028224	0,052752
Esen. Colesterolo	0,000347	0,001222	0,002545	0,009348	0,030492	0,043007	0,021417	0,000369	0,001576	0,006370	0,017950	0,032935	0,035092	0,019302
Esen. Ipertensione con	0,000193	0,001081	0,005410	0,024816	0,077775	0,139336	0,162093	0,000350	0,002053	0,008779	0,039815	0,112533	0,173170	0,191487
Esen. Ipertensione senza	0,000154	0,001208	0,005937	0,024611	0,063254	0,088191	0,095307	0,000203	0,002053	0,008189	0,029862	0,068133	0,085943	0,084451
Esen. Malattie neopl. maligne	0,002869	0,008046	0,020378	0,048765	0,091283	0,113848	0,086168	0,003134	0,006449	0,010669	0,021984	0,063960	0,123471	0,150098
Esen. Invalidità civile >2/3	0,002503	0,004100	0,007541	0,014589	0,029708	0,044459	0,051826	0,002987	0,004410	0,008079	0,014142	0,026323	0,039774	0,038465
Esen. Invalidità civile accompagn.	0,002830	0,002850	0,002989	0,004542	0,007050	0,014979	0,124719	0,003190	0,003628	0,003840	0,004556	0,007065	0,014976	0,070681
Esen. Invalidità civile 100	0,001617	0,001811	0,003507	0,006927	0,014558	0,026201	0,078778	0,001917	0,002357	0,003430	0,006691	0,014356	0,025137	0,052314
Esen. Gravidanza	0,017925	0,090314	0,039049	0,000494	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000
Esen. Altro	0,042185	0,061965	0,079473	0,096422	0,117164	0,099167	0,079541	0,036764	0,035953	0,051376	0,074221	0,117938	0,140619	0,145819
Medico maschio	0,638774	0,656431	0,674065	0,689133	0,703879	0,707225	0,706605	0,699039	0,715450	0,717257	0,726527	0,735103	0,734913	0,734654
Supera max assistiti	0,342442	0,322766	0,337071	0,330620	0,321741	0,319788	0,303148	0,331295	0,317053	0,320043	0,323889	0,312205	0,310569	0,303450
Anni dalla laurea	24,8	24,8	24,9	25,5	25,7	25,5	25,4	25,0	25,1	25,0	25,4	25,9	25,7	25,5

δ_j ricoveri	0,0565	0,1384	0,1100	0,0812	0,1007	0,1333	0,2222	0,0459	0,0425	0,0501	0,0708	0,1124	0,1778	0,2741
δ_j specialistiche	0,5485	0,6226	0,6478	0,6847	0,7384	0,8241	0,8030	0,4393	0,3918	0,4421	0,5258	0,6722	0,8143	0,8488
δ_j farmaci	0,4710	0,5790	0,6154	0,6728	0,7831	0,8740	0,9104	0,3776	0,3873	0,4533	0,5512	0,7189	0,8546	0,9182

Bibliografia

- Achen, Christopher H. (1986). *The Statistical Analysis of Quasi- Experiments*. Berkeley, CA: University of California Press;
- Amemiya T., (1995). *Advanced Econometrics*, Harvard University Press, Cambridge, Massachusetts;
- Cappellari L., Jenkins S., (2003). *The Stata Journal. Multivariate probit regression using simulated maximum likelihood*;
- Cappuccio N., Orsi R., (2005). *Econometria*, Il Mulino;
- Cappuccio N., Orsi R., (2011). *Introduzione all'econometria*, Torino, G. Giappichelli Editore;
- Bryk, A.S., & Raudenbush, S.W. (1992). *Hierarchical Linear Models in Social and Behavioral Research: Applications and Data Analysis Methods* (First Edition). Newbury Park, CA: Sage Publications
- De Leeuw J., Meijer E., (2008). *Handbook of Multilevel Analysis*. Springer;
- Ferrari W., (2008). *Guida ai Servizi per la salute in Friuli-Venezia Giulia*;
- Golder M., *Advanced quantitative analysis. Tobit models*. Department of Political Science, Pennsylvania State University;
- Golder M., (2003). *Explaining Variation in the Electoral Success of Extreme Right Parties in Western Europe*. Comparative Political Studies;
- Goldstein H., (2011). *Multilevel Statistical Models*. 4th Edition;
- Greene W., (2002). *Econometric Analysis*, New York University, Fifth Edition; NJ: Prentice Hall;
- Hansen B., (2012). *Econometrics*, University of Wisconsin;
- Heckman, J. J. (1979). *Sample Selection Bias as a Specification Error*. *Econometrica* 47(1): 153-161;
- Hox J. J., (2010). *Multilevel Analysis: Techniques and Applications*, Quantitative methodology series, Second Edition;
- Hox J.J., (1995). *Applied Multilevel Analysis*, TT-Publicaties;

- Kreft I., De Leeuw, (1999). *Introducing Multilevel Modeling*, Sage Publication Ltd;
- Lawrence E., (1997). *Simulated Maximum Likelihood via the GHK Simulator: An Application to the 1988 Democratic Super Tuesday Primary*, The Brookings Institution;
- Long J. Scott, (1997). *Regression Models for Categorical and Limited Dependent Variables. Advanced Quantitative Techniques in the Social Sciences Series*. London: Sage Publication;
- Martini N., Raschetti R., (2008). *L'uso dei farmaci in Italia. Rapporto nazionale anno 2007*;
- Meneghelli G., (2008). *La rivoluzione demografica italiana: dal baby boom alle culle vuote*;
- Paccagnella O., (2011). *Sample size and Accuracy of Estimates in Multilevel Models. New Simulation Results*. Methodology – European Journal of Research Methods for Behavioral and Social Sciences 7, 11-120;
- Piergentili P., Simon G., Paccagnella O., Grassetto L., Rizzi L., Samani F., (2008). *Costi tariffari riportabili al MMG: modelli di risk adjustment basati su dataset amministrativi*;
- Roodman D., (2010). *Estimating fully observed recursive mixed-process models with cmp*;
- Ruscione N.M., (2013). *Utilizzo dell'ICC come indicatore dell'esistenza di una struttura gerarchica*;
- Shenyang G., Mark W. (2015). *Propensity Score Analysis, Statistical Methods and Applications*. Second Edition, Sage Publication, Chapter 4: Sample Selection and Related Models: 85-125;
- Sigelman, Lee & Langche Zeng, (1999). *Analyzing Censored and Sample-Selected Data with Tobit and Heckit Models*;
- Snijders T.A.B., Bosker R.J., (1999). *Multilevel analysis: an introduction to basic and advanced multilevel modelling*. Sage Publications, Londra;
- Stephen W., Anthony S. (2002). *Hierarchical Linear Models, Applications and Data Analysis Methods*. Second Edition;
- Tabachnick G.B., Fidell S.L., (2013). *Using Multivariate Statistics*, Sixth Edition, Pearson;
- Tobin J., (1958). *Estimation of relationship for limited dependent variables*;
- Verbeek M., (2006). *Econometria*, Zanichelli. Titolo originale: *A Guide to Modern Econometrics*, 2004, Second Edition;
- Wooldridge J., (2002). *Introductory Econometrics. A modern approach*, Second Edition, Cambridge: MIT Press;

Sitografia

www.agenziafarmaco.gov.it

www.ass1.sanita.fvg.it

www.ass2.sanita.fvg.it

www.ass3.sanita.fvg.it

www.ass4.sanita.fvg.it

www.ass5.sanita.fvg.it

www.ass6.sanita.fvg.it

www.demo.istat.it

www.drg.it

www.movimentoconsumatori.it

www.ospfe.it

www.salute.gov.it

www.who.int

Ringraziamenti

I dati sono stati forniti dal Ceformed (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.

A termine di questo intenso percorso di studi è doveroso ringraziare tutte le persone che mi sono state vicino e che mi hanno sostenuta in tutti questi anni.

Ringrazio la mia famiglia per avermi sempre supportata, perché è anche grazie ai loro sacrifici che oggi sono riuscita a raggiungere l'obiettivo che mi ero prefissata. Ringrazio mia mamma per aver condiviso con me gioie e dolori di questo percorso, per la sua pazienza e comprensione, ringrazio mio papà che mi ha sempre spronata ad andare avanti anche nei momenti di difficoltà.

Ringrazio Filippo che mi è sempre stato accanto in questi anni e che ha sempre creduto in me. Inoltre, lo ringrazio per la pazienza infinita e per la sua ventata quotidiana di buon umore e ottimismo che mi ha sempre aiutata molto.

Ringrazio la mia migliore amica Silvia che da 20 anni mi è sempre accanto e spalleggia le mie decisioni, mi comprende, mi ascolta e mi supporta.

Ringrazio tutti i miei compagni di Università, per tutte le ore passate insieme a studiare e per tutti i bei momenti trascorsi. Ringrazio Stefania, Alice e Carlo che nei primi anni sono stati fondamentali per me, ringrazio Annalisa, Giada, Camilla, Tiziana, Silvia e Alice che durante gli anni della magistrale hanno rallegrato il mio percorso, per tutte le ore passate a studiare in gruppo e per la collaborazione e sostegno che abbiamo condiviso.

Ringrazio tutti i miei amici che, come me, capiscono le difficoltà dello studente universitario e che sono in prima linea a festeggiare la mia laurea.

Per ultimo, ma non per importanza, ringrazio il mio relatore, il Professor Omar Paccagnella per l'infinita disponibilità, pazienza e comprensione avuta in questi mesi di stesura della tesi.

