

UNIVERSITÀ DEGLI STUDI DI PADOVA

FACOLTÀ DI SCIENZE STATISTICHE

Corso di laurea specialistica in
SCIENZE STATISTICHE, DEMOGRAFICHE E SOCIALI



TESI DI LAUREA

**La povertà in Perù:
dimensioni del disagio femminile in
diversi contesti geografici**

RELATORE: *Ch.ma Castiglioni Maria*

LAUREANDO: Da Re Riccardo

ANNO ACCADEMICO 2007-2008

INDICE

0. PREMESSA.....	5
1. INTRODUZIONE.....	15
1.1. IL CONTESTO.....	15
1.1.1. Il Perù.....	15
1.1.2. Il Dipartimento di Cusco.....	25
1.2. DEMOGRAPHIC AND HEALTH SURVEYS (DHS).....	34
1.2.1. Il questionario.....	35
1.2.2. Aspetti operativi dell'indagine.....	37
1.2.3. Dati e variabili utilizzati nello studio.....	37
1.3. METODOLOGIE UTILIZZATE.....	38
1.3.1. Analisi delle corrispondenze.....	38
1.3.2. Analisi di segmentazione.....	40
1.3.3. Analisi di regressione logistica.....	42
2. PRIME ANALISI A LIVELLO NAZIONALE.....	45
2.1. PROPOSTE DI INDICATORI E MACROAREE D'INTERVENTO.....	45
2.2. SELEZIONE DELLE VARIABILI.....	52
2.2.1. Ipotesi iniziali e problema dei <i>missing values</i>	52
2.2.2. Età al matrimonio non supplementare.....	59
2.2.3. Evidenziare lo stato coniugale.....	60
2.2.4. Scelta delle variabili che definiscono l'abitazione.....	61
2.2.5. Scelta delle variabili per definire la presa di decisioni.....	68
2.2.6. Conclusioni.....	70
2.3. LE NON RISPOSTE E IL PROBLEMA DELLA <i>PRIVACY</i>	71
2.3.1. Analisi di regressione logistica.....	71

3. CONFRONTI FRA REGIONI.....	85
3.1. COSTA NORD: TUMBES.....	89
3.2. COSTA SUD: ICA.....	98
3.3. ZONA METROPOLITANA: LIMA.....	107
3.4. GIUNGLA BASSA: LORETO.....	112
3.5. GIUNGLA ALTA: AMAZONAS.....	121
3.6. ALTIPIANO ANDINO: APURIMAC.....	127
3.7. CONCLUSIONI.....	135
4. CUSCO.....	141
4.1. IL DISAGIO E LE MANCATE RISPOSTE.....	141
4.2. RISULTATI SENZA <i>MISSING VALUES</i>	147
4.3. CONCLUSIONI.....	155
5. CONCLUSIONI.....	159
6. BIBLIOGRAFIA.....	163
7. ALLEGATI.....	165

0. PREMESSA

L'idea base dalla quale è partito questo studio, è nata a conclusione di uno stage di quattro mesi svolto nella primavera del 2006 presso una ONG privata, MIDE, che opera senza fini di lucro con il microcredito, aiutando principalmente le donne *campesinas* della zona andina del sud-est del Perù. Uno dei problemi sollevati era quello di monitorare le condizioni di vita delle donne; per affrontare questo problema, ci si confronta quotidianamente con dei problemi che molte persone cercano da tempo di risolvere. Alcuni di questi problemi riguardano: la scelta degli strumenti di misura (per esempio, quali indicatori utilizzare), che cosa si intende per sviluppo, e qual è la popolazione obiettivo.

L'esigenza di trovare degli indicatori efficaci e comuni a tutti i Paesi in via di sviluppo, risponde a necessità socio-politiche molto importanti, fra le quali:

- poter confrontare fra di loro regioni dalle caratteristiche culturali molto diverse, dislocate in diverse parti del mondo (ma anche all'interno di uno stesso continente), differenti per motivi geografici, storici o politici;
- valutare l'impatto sociale sulla popolazione di talune politiche o azioni, economiche e non, sia da parte di enti pubblici che privati.

Il tema è ovviamente complesso e dibattuto: se a livello mondiale si cerca di semplificare e uniformare il metodo di valutazione e il confronto descrittivo fra le regioni "in via di sviluppo", dall'altra parte sorge il problema dell'incapacità di questi indicatori di cogliere le varie sfaccettature delle realtà locali. Inoltre viene spesso sollevato il problema della forzata esigenza di voler quantificare a tutti i costi situazioni e tematiche soggettive, che i numeri rischiano sempre di non cogliere adeguatamente o rendere "fredde" ed inutilizzabili per studi successivi. Inoltre spesso si aggiunge il problema di raccolta dati in questi Paesi dove può non essere presente un sistema statistico efficiente; è difficile quindi essere esigenti, e il rischio è quello di ritrovarsi a dover maneggiare dati poveri rispetto alle aspettative. Nel nostro lavoro, si è scelto di non considerare singoli indicatori, ma di focalizzarci sull'estrazione delle variabili che meglio descrivono i contesti regionali.

Quello che noi chiamiamo “sviluppo” (e di conseguenza, come suo contrario, parleremo di “condizioni di disagio”) corrisponde veramente a ciò che le persone definite “povere” hanno progettato e vogliono per il loro futuro? Sono in grado degli indicatori statistici, costruiti per valutare la bontà di un progetto che ha come focus principale la gente, di afferrare veramente quei cambiamenti soggettivi che ogni singola persona desidera?

Dopo l'introduzione nel 1990 del concetto di sviluppo umano, basato principalmente sulle teorie del premio nobel Amartya Sen, viene considerato un nuovo modo di vedere lo sviluppo: esso viene definito come il processo di ampliamento delle scelte della gente e, per la prima volta, l'aspetto economico, sul quale erano basati tutti gli indicatori precedenti, assume un ruolo relativo e viene affiancato da altre misure. Tali scelte possono essere infinite e cambiare nel tempo ma, come affermato nel rapporto UNDP del 1990, a qualsiasi livello di sviluppo le tre caratteristiche essenziali per valutare il miglioramento di un individuo sono: la possibilità di condurre una vita lunga e sana, di acquisire conoscenze, e di accedere alle risorse necessarie ad un tenore di vita dignitoso (i tre “ambiti” sui quali si calcola l'Indice di Sviluppo Umano).

Da quel momento in poi, molteplici sono stati i tentativi di trovare algoritmi ed indicatori sempre più idonei, e per comprendere l'ampiezza della proposta è sufficiente ricercare fra le presentazioni di differenti organizzazioni che operano nell'ambito dello sviluppo. Questo accade sia a livello mondiale, ma soprattutto a livello locale, dove ogni struttura che opera nell'ambito sociale cerca di descrivere la realtà circostante e di valutare se il proprio operato ottiene gli effetti desiderati, basandosi di volta in volta su variabili diverse a seconda del contesto in cui lavora.

La questione principale è che non esiste un “indicatore corretto o scorretto”; esistono alcuni indicatori standardizzati per misurare i risultati finanziari, ma il tema dell'impatto sociale è molto più complesso. Gli indicatori che sono utili per un determinato programma in una zona geografica, non sono necessariamente adeguati anche per altri programmi, ma è necessario adattarli alle esigenze locali.

Per il nostro studio, ci si è basati sui quadri di lavoro di alcune grandi e piccole organizzazioni, preferibilmente operanti in America latina, perché più vicine alla realtà analizzata.

La parola povertà è inoltre legata troppo spesso al problema delle pari opportunità tra uomini e donne. In nessuna società le donne godono gli stessi diritti ed opportunità degli uomini, e questo non dipende dal reddito nazionale o altri indicatori economici. Le differenze di genere esistono in qualsiasi ambito, dall'educazione alla sanità (divario però che si sta riducendo in molti Paesi), dalle retribuzioni economiche nel lavoro alla partecipazione politica. Secondo le Nazioni Unite, la povertà è femminile: il 70% dei poveri sono donne che si trovano in situazioni di disagio ed impotenza, a causa di salari più bassi di quelli degli uomini, dell'inaccessibilità ai ruoli decisionali, della scarsa autostima.... Questo è uno dei motivi che ha spinto ad ideare altri indici, oltre all'ISU (Indice di Sviluppo Umano), come l'ISG (Indice di Sviluppo correlato al Genere) o l'MPG (Misura di attribuzione di Potere correlato al Genere), che tenessero conto di queste problematiche.

Le donne sono però anche più affidabili rispetto agli uomini: più difficilmente dilapidano i loro stipendi in alcool, sono maggiormente legate all'abitazione e al futuro dei figli. E' per questo che molti progetti di sviluppo e finanziamento (in primo piano il microcredito) vengono svolti affidandosi alle donne e non necessariamente al "padrone" di casa, che normalmente è l'uomo. Ed è per questo che molti studi vengono svolti sui dati raccolti da campioni femminili, come avverrà nel nostro caso.

La nostra ricerca non ha alcun obiettivo risolutivo: il focus delle analisi sarà quello di approfondire la problematica della ricerca di variabili ed indicatori "socialmente determinanti" per la condizione della donna, studiando a fondo la situazione di una nazione, il Perù, per evidenziarne sia le caratteristiche simili presenti in tutta la popolazione, sia le particolarità che delineano invece ogni regione geografica presente nel territorio. Sarà interessante estrarre quelle variabili comuni a tutti, utili per determinare eventuali linee base di operato; ma soprattutto capire come le problematiche e le esigenze delle persone cambino

drasticamente a seconda della zona geografica, dell'etnia, del luogo di residenza...

Il Perù si presta particolarmente bene a questo scopo, perché la sua vastità offre differenze enormi (anche all'interno delle singole regioni): gli stili di vita cambiano da nord a sud, dalle città alle campagne, dalla costa agli altipiani alti fino a oltre i 5000 metri, alla giungla, dalla stessa Lima, metropoli che da sola accoglie quasi un terzo della popolazione del Paese, a qualsiasi altro centro abitato; inoltre sono presenti in tutto il territorio diversi gruppi etnici con lingua e tradizioni proprie.

Questa varietà e ricchezza si contrappone alla diffusa povertà che sembra invece accomunare un'alta percentuale della popolazione. Secondo l'INEI, l'istituto nazionale di statistica e informatica, il 51,6% dei peruviani vivono in condizioni di povertà, cioè vivono in abitazioni le cui spese totali sono inferiori ad una soglia base di consumo: tale "linea di povertà" varia tra i 147,39 dollari mensili nella giungla rurale, ai 260,21 a Lima metropolitana. Invece il 19,2% (vale a dire circa 6 milioni di abitanti) della popolazione nazionale vive in una situazione definita di povertà estrema, cioè hanno delle spese inferiori ad una soglia base di alimenti: tale "linea di povertà estrema" fluttua fra i 95,01 dollari mensili nella giungla rurale, ai 138,18 nella giungla urbana (dati del 2004, "Indagine Nazionale sulle Abitazioni e sulle Condizioni di Vita e Povertà").

La differenza fra "povertà" e "povertà estrema", è che la prima si riferisce a quei cittadini che non sono in grado di accedere ad un paniere di beni e servizi minimi ritenuti essenziali (è sufficiente non coprirne uno per essere considerati poveri), mentre la seconda riguarda solo l'alimentazione. L'Organizzazione Mondiale della Sanità stima che ad un essere umano servono 2318 calorie giornaliere per sopravvivere (tale cifra varia col sesso, l'età e la fisionomia dell'individuo); quindi, ad esempio, chi riesce ad alimentarsi in modo adeguato, ma non è in grado di accedere all'educazione, alla salute o a qualsiasi infrastruttura pubblica, è considerato "povero", ma non in una situazione di "povertà estrema".

Le percentuali, precedentemente riportate, ovviamente cambiano drasticamente da regione a regione, e all'interno di esse, anche spostandosi dalle zone rurali a quelle urbane, come si vede dalla tabella seguente

Tabella 0.1. Percentuale della popolazione che vive sotto la “linea di povertà” e la “linea di povertà estrema” per aree geografiche (2004)

	Incidenza Povertà Totale	Incidenza Povertà Estrema
Perù	51,6	19,2
Area		
Urbana	43,3	11,5
Rurale	72,5	40,3
Zona geografica		
Metropoli	36,6	3,4
Costa	40,8	8,1
Altipiano	67,7	36,5
Giungla	59,5	26,4

Fonte: Indagine Nazionale sulle Abitazioni, promossa dall'INEI (Istituto Nazionale di Statistica ed Informatica peruviano), 2004

Queste differenze sono evidenziate anche dall'ISU, che posiziona nel 2005 il Perù all'87esimo posto sulle 177 Nazioni analizzate, con un valore di 0,773; ma ovviamente questo indice raggiunge soglie molto più basse in alcuni dipartimenti, come si vedrà nei prossimi capitoli.

La domanda quindi a cui rispondere per poter operare in un ambito così complesso è: cosa accomuna e cosa divide le regioni del Perù? Salute, abitazione, sviluppo lavorativo e capitale, sicurezza alimentare, educazione, autostima: sono sempre tutti temi egualmente importanti, o i veri obiettivi futuri e le necessità imminenti di queste persone differiscono dai nostri canoni di benessere? In che ambiti si può operare con le medesime metodologie, e dove invece bisogna approfondire i rapporti e capire le particolarità del luogo per poter offrire un aiuto più mirato e non grossolano? Dove alcuni indicatori, normalmente utilizzati, rischiano di perdere significato?

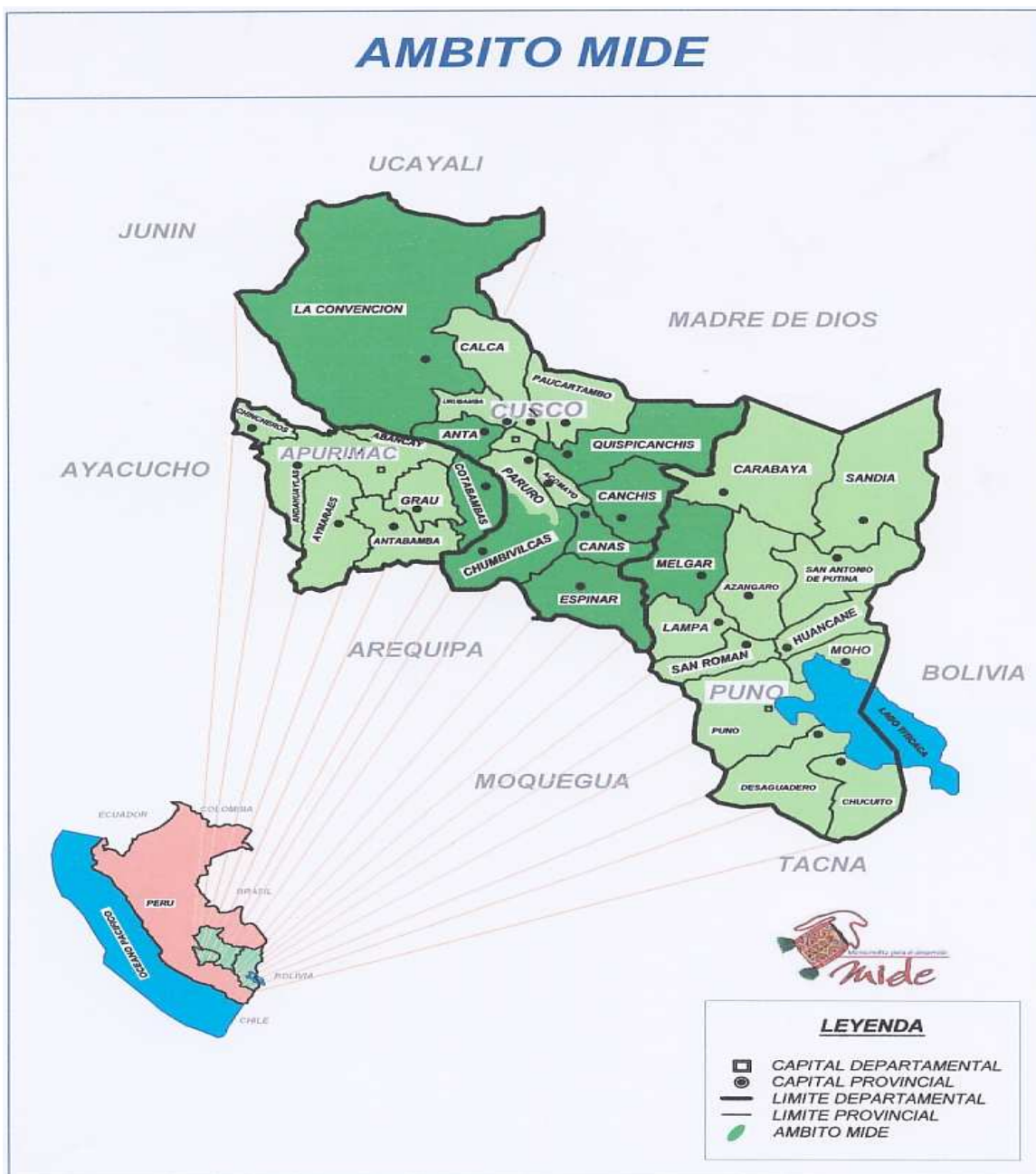
Presentiamo ora le linee caratterizzanti dell'organizzazione presso la quale si svolto lo stage in Perù.

MIDE lavora quasi esclusivamente con donne, e le sue clienti vivono spesso in condizioni di povertà estrema; i prestiti erogati sono molto piccoli, soprattutto all'inizio della collaborazione: essi mediamente fluttuano dai US\$14 ai US\$83. Il numero di clienti attivi al settembre 2004 era di 3673, e la percentuale di mora

(a 30 giorni dal prestito) del 5,8%. MIDE lavora nei dipartimenti di Cusco, Apurimac e Puno.

I prestiti sono impiegati prevalentemente nel settore agricolo, nell'allevamento e nella produzione artigianale. Alle destinatarie dei prestiti si chiede spesso di organizzarsi in gruppi solidali con un numero di unità limitato, che rispondono vicendevolmente del debito contratto, sviluppando doti collaborative e di solidarietà. Le donne *campesinas* con cui MIDE lavora appartengono all'etnia quechua: parlano quasi esclusivamente la lingua quechua, spesso sono del tutto analfabete e all'interno della famiglia vivono in condizioni di totale subalternità al marito, subendone spesso la violenza. Hanno mediamente 4 o 5 figli, e tra i figli le bimbe subiscono un'eguale discriminazione: se infatti ai maschi viene data qualche possibilità di frequentare la scuola, le bambine ne vengono escluse. Dare credito a queste donne significa dunque promuoverne l'emancipazione, stimolando un vero e proprio processo di autoconsapevolezza e di presa di coscienza dei propri diritti e delle proprie potenzialità. L'obiettivo principale di MIDE è lo sviluppo delle comunità agricole della zona, tramite il sistema di microcredito, e grazie a mirati programmi di autostima della donna, presa di coscienza e organizzazione in gruppi: in questo modo si cerca di sostenere le donne sia moralmente che materialmente.

Mappa 0.1. Ambito di intervento di MIDE nei Dipartimenti di Cusco, Apurimac e Puno.



La particolarità e la ricchezza di questo stage è dovuta al fatto che il territorio dove MIDE agisce è molto vario: si passa dalle comunità andine che vivono oltre i 5000 metri (per esempio nella zona di Espinar) alle donne che risiedono nella cosiddetta giungla alta (attorno ai 1000 metri). Le attività che svolgono sono distinte in ogni zona: commercializzazione di prodotti agricoli, allevamento di piccoli (pollame, maiali, pecore, porcellini d'india) e grandi animali (bovini e alpaca), produzione di formaggi, compravendita e fabbricazione di oggetti d'artigianato, vendita di frutta e produzione di cacao e caffè.

Tali diversità hanno a volte portato al fallimento di certi metodi di erogazione del credito, e alcune dinamiche di gruppo che funzionano nella maggior parte del territorio, in altre aree erano assolutamente inconcepibili. Questo perché le condizioni climatiche e la cultura fanno sì che le esigenze siano diverse, e le donne stesse siano più o meno diffidenti verso questi interventi esterni. Se questo accade a livello regionale, cosa può avvenire in un contesto nazionale o addirittura mondiale?

I dati utilizzati nelle analisi sono stati forniti da Measure DHS (Demographic and Health Surveys). Il dataset più recente disponibile è quello del 2000: esso è stato ricavato somministrando un questionario standard a 29422 donne di età compresa fra i 15 e i 49 anni, residenti su tutto il territorio nazionale. Il campione è stratificato per dipartimento. Il tasso di risposta è stato del 94.6%. Il questionario si sviluppa in 12 sezioni, che toccano vari argomenti, dalla storia della donna a quella della sua famiglia, dalla salute e la condizione dell'abitazione a questioni più delicate, quali la violenza familiare o i metodi contraccettivi conosciuti e utilizzati. Per ulteriori informazioni su DHS, sul campionamento, sul questionario e sulla selezione delle variabili da utilizzare si rimanda il lettore ai capitoli successivi.

Le nostre analisi non includeranno tutti i dati. Si è deciso di usare solo le informazioni riguardanti le 18237 donne che hanno avuto almeno una volta nella vita un partner (ci si riferisce quindi alle donne sposate, conviventi, separate, divorziate e vedove); questa suddivisione è legata al tentativo di creare due gruppi omogenei al loro interno, le cui caratteristiche sono troppo differenti per essere studiate assieme: di fatto le donne nubili coprono un ruolo (ad esempio di figlie) all'interno della famiglia ben diverso da quelle già sposate, che invece sono investite di maggiore responsabilità familiare. Inoltre quest'ultime, essendosi relazionate almeno una volta nella vita con un partner, hanno risposto ad una serie di domande del questionario (ad esempio sulla violenza domestica) che permettono un'analisi differente da quella che si farebbe con le "mai coniugate". Il loro inserimento potrebbe quindi distogliere l'attenzione invece che favorire il confronto: sarebbe più utile uno studio distinto, che però non verrà compiuto in questo lavoro.

Chiarite le premesse che hanno indotto ad iniziare questo lavoro, e descritte schematicamente le fonti e il dataset utilizzato, riportiamo ora un riassunto di come lo studio sarà strutturato.

In prima istanza si cercherà di delineare un quadro generale del contesto a cui i nostri dati fanno riferimento, sia del Perù che del Dipartimento di Cusco, dove era stato effettuato lo stage. Inoltre verranno riportate e descritte brevemente le metodologie d'analisi che si utilizzeranno in queste pagine, e verrà compiuto un approfondimento sul campionamento per la raccolta dei dati e sulla struttura del questionario (capitolo 1).

Lo studio vero del dataset inizia con un primo esame delle relazioni di interdipendenza dei dati a livello nazionale tramite l'analisi multipla delle corrispondenze. Il vero obiettivo di questo capitolo è però la scrematura delle variabili, in modo da sceglierne un ristretto gruppo significativo degli argomenti trattati dal questionario, da utilizzare successivamente nel confronto fra regioni. Alle conclusioni si aggiungerà anche un'analisi approfondita dei valori mancanti dovuti al disagio creato dalla mancanza di privacy in alcune interviste (capitolo 2).

Nella fase successiva si confronteranno alcuni dipartimenti, opportunamente scelti per rappresentare le diverse zone geografiche del Paese: Lima metropolitana, costa del Nord e costa del Sud, altipiano, giungla alta e giungla bassa. All'analisi multipla delle corrispondenze si affiancherà l'analisi di segmentazione condotta sulle coordinate assunte dalle modalità nei due assi fattoriali, per poter meglio comprendere quali sono le variabili che hanno maggiormente contribuito alla formazione di essi (capitolo 3).

Un approfondimento verrà effettuato per il dipartimento di Cusco, trasversale a tre delle zone geografiche sopra elencate, in quanto sede principale dello stage (capitolo 4).

Infine nelle conclusioni si cercherà di riassumere quanto emerso, col fine di rispondere ai quesiti sopra citati: quali variabili sono comuni a tutto il territorio e quindi utilizzabili per la costruzione di opportuni indicatori, e quali invece differiscono caratterizzando regioni, aree geografiche e piccole etnie locali? (capitolo 5)

1. INTRODUZIONE

1.1 IL CONTESTO

Sarà presentato sinteticamente il contesto a cui i nostri dati fanno riferimento, ossia il Perù e il Dipartimento di Cusco, dove era stato effettuato lo stage.

1.1.1 Il Perù

Il Perù è una nazione dell'America meridionale. E' una Repubblica presidenziale, e l'attuale capo di governo è Alan Garcia.

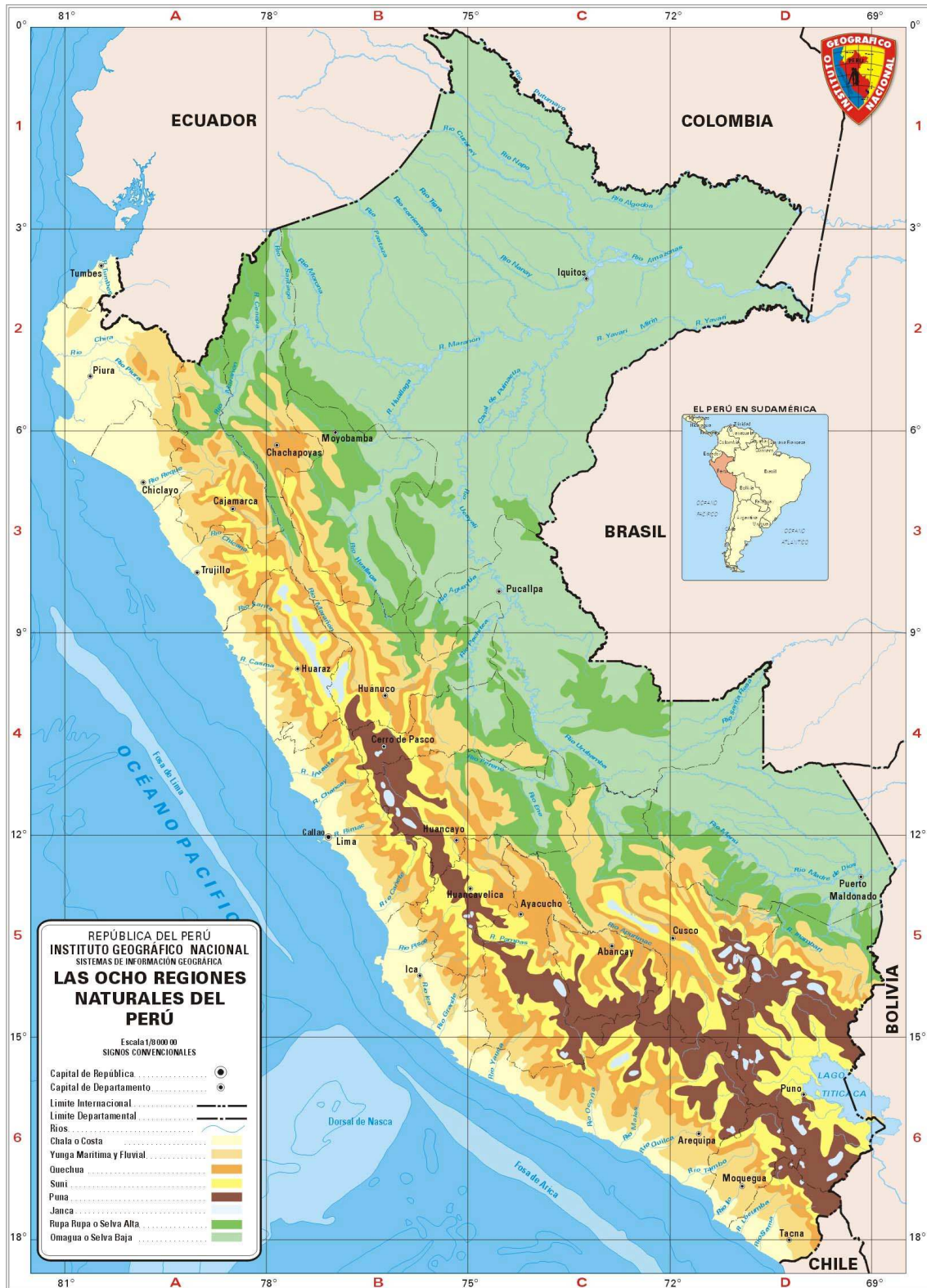
E' un paese geograficamente variegato situato sulla costa dell'Oceano Pacifico. Si estende su un territorio di 1.285.216 kmq (compresi 4.996 kmq della parte peruviana del lago Titicaca al confine con la Bolivia) ed è, per grandezza, il terzo paese dell'America Latina. Con un'estensione che va da pochi chilometri a sud dell'equatore fino ai 18° di latitudine sud, confina a nord con l'Ecuador e la Colombia, ad est con il Brasile e la Bolivia ed a sud con il Cile.

Il sistema montagnoso, la Cordigliera delle Ande, che percorre il Paese da Nord a Sud, dà luogo a tre zone morfologiche che si sviluppano parallelamente in senso longitudinale: la *Costa* (fascia costiera sull'Oceano Pacifico), la *Sierra* (gli altipiani costituiti dalle Ande peruviane) e la *Selva* (la giungla della zona amazzonica orientale). Le caratteristiche di queste zone fanno sì che il Perù sia, dal punto di vista naturale, un Paese molto variegato in cui coesistono una grande ricchezza di paesaggi e di ecosistemi. Procedendo dal litorale pacifico verso l'interno, si incontra:

- ❖ La zona bassa e prevalentemente desertica della *Costa*, che costituisce l'11% del territorio, stretta tra il mare e la montagna ed attraversata perpendicolarmente dai fiumi che scendono dal versante occidentale delle Ande con corso irregolare. L'agricoltura che, rispetto al resto del Paese, è qui praticata con tecniche più moderne, e che si sviluppa soprattutto nelle valli irrigate dai fiumi, si limita a produzioni stagionali, dipendenti dalla disponibilità idrica e dalla frequenza delle piogge. In questa zona, l'unica di facile accesso e servita dalle principali vie di comunicazione stradale, si trovano le principali città.

- ❖ La parte centrale, la *Sierra* (il 32% del territorio nazionale), è costituita dalla Cordigliera delle Ande, che presenta vette altissime e gole profonde. Il clima freddo di queste zone raggiunge temperature molto basse ed i pochi appezzamenti di terreno coltivabili forniscono una produzione appena sufficiente al sostentamento delle popolazioni che abitano questa zona e che praticano l'agricoltura con tecniche tradizionali. In alternativa, si esercita l'allevamento di piccoli e grandi animali, e l'artigianato.
- ❖ La parte più orientale del Paese, la *Selva*, che occupa la maggior estensione della superficie territoriale del Perù (57%) presenta una vegetazione rigogliosissima essendo la zona più ricca di acque che, scendendo dalle Ande, formano il bacino del Rio delle Amazzoni. Costituita dalle pendici orientali dei rilievi andini e dalla pianura amazzonica, con clima tipico delle foreste pluviali, e ricoperta dalla fitta foresta spesso inaccessibile, è la zona più ricca di vegetazione e di risorse naturali, ma la meno abitata del Paese. Si divide in *Selva Alta* e *Selva Baja*.

Mapa 1.1. Le regioni naturali del Perù



Demografia

In Perù risiedono 26.152.265 abitanti, per una densità di 21 ab/kmq (dati censimento 2005, fonte INEI), dei quali oltre 7.700.000 solo nella capitale, Lima

(8.630.000 se si considera l'agglomerato urbano). Le donne sono il 50,1% della popolazione totale.

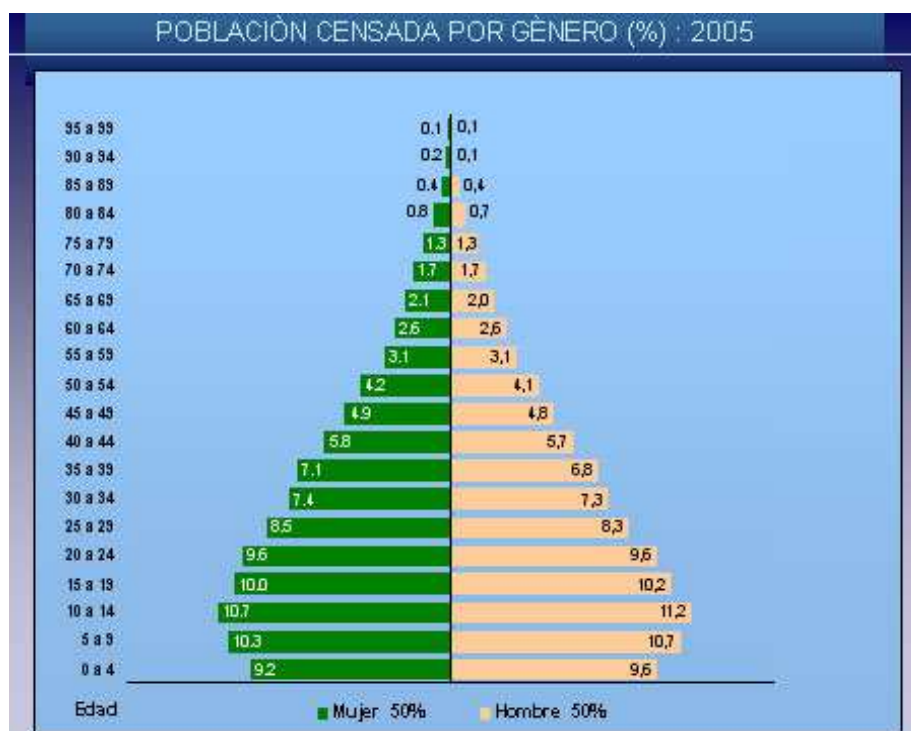
Ecco alcuni indicatori e alcune informazioni sulla situazione demografica del Paese (Fonte: "Calendario Atlante De Agostini 2008", Istituto Geografico De Agostini).

Crescita annua:	1,5% (2000-2005)
Incremento naturale (per 1000 abitanti):	14,6 ‰ (2005)
Fecondità (figli per donna):	2,6 (2005)
Natalità (per 1000 abitanti):	20,9‰ (2005)
Mortalità (per 1000 abitanti):	6,3‰ (2005)
Mortalità infantile (per 1000 nati vivi):	23‰ (2005)
Speranza di vita (anni):	M 68 F 73 (2005)
Popolazione urbana:	72,6% (2005)

Gruppi etnici:	Quechua 47,1%, Meticci 32%, Creoli 12%, Aymarà 5,4%, altri 3,5%
Lingue:	aymarà, quechua e spagnolo quelle ufficiali
Religione:	cattolici 89,2%, protestanti 6,6%, altri 4,2%

Ai fini dell'analisi che si condurrà in questa tesi, è importante conoscere la distribuzione d'età della popolazione, soprattutto di quella femminile. Si vede per esempio che la percentuale delle donne fra i 15 e i 49 anni (come si vedrà più avanti, la fascia d'età del campione di donne utilizzato nel nostro studio) è del 52,7% rispetto al totale della popolazione femminile.

Grafico 1.1. Piramide d'età della popolazione peruviana, per sesso, al censimento 2005 (valori per 100 femmine e per 100 maschi, rispettivamente)



Fonte: INEI

Dipartimenti amministrativi

Il territorio peruviano è suddiviso in 24 Dipartimenti amministrativi, 25 se si considera la Provincia Costituzionale di Callao (all'interno del territorio di Lima). I 24 Dipartimenti sono a loro volta frazionati in 194 aree intermedie, dette Province, e in 1828 divisioni minori o Distretti.

Tabella 1.1. Popolazione e densità per Dipartimento (al censimento 2005)

Dipartimenti	Popolazione	Densità
Amazonas	389.700	10
Ancash	1.039.415	29
Apurimac	418.882	20
Arequipa	1.140.810	18
Ayacucho	619.338	14
Cajamarca	1.359.023	41
Cusco	1.171.503	16
Huancavelica	447.054	20
Huanuco	730.871	20
Ica	665.592	31
Junin	1.091.619	25
La Libertad	1.539.774	60
Lambayeque	1.091.535	77
Lima	7.819.436	225
Loreto	884.144	2
Madre de Dios	92.024	1

Moquegua	159.306	10
Pasco	266.764	11
Piura	1.630.772	45
Puno	1.245.508	17
San Martin	669.973	13
Tacna	274.496	17
Tumbes	191.713	41
Ucayali	402.445	4
Callao (provincia costituzionale)	810.568	5.514

Fonte: "Calendario Atlante De Agostini 2008", Istituto Geografico De Agostini, Novara

Mappa 1.2. I Dipartimenti amministrativi del Perù



Istruzione e Sanità¹

L'istruzione è gratuita ed obbligatoria dai 7 ai 16 anni.

Gli studenti sono 7.677.987 (2005), mentre i laureati sono 55.923 (2004). Il 12,3% della popolazione è analfabeta.

La spesa dello Stato per l'istruzione nel 2005 era pari al 2,4% del PIL.

Alcuni indicatori per la sanità sono:

Posti letto ospedalieri: 1,2 ogni 1000 ab. (2005)

Medici: 1,2 ogni 1000 ab. (2004)

Alcune malattie: AIDS 0,6% degli adulti (2005), 206 casi di tubercolosi ogni 100.000 abitanti (2005) e 293 di malaria (2003)

Spesa dello Stato per la sanità: 1,2% del PIL (2005)

Povert , ISU e altri indicatori

L'indice di sviluppo umano tiene conto dei seguenti fattori:

- il reddito, rappresentato dal prodotto interno lordo (Pil) individuale, dopo una trasformazione che tiene conto sia del potere di acquisto della valuta, sia del fatto che l'aumento del reddito non determina un aumento del benessere in modo lineare (l'aumento di benessere   molto maggiore quando il Pil passa da 1000 a 2000 dollari che quando passa da 15.000 a 16.000);
- il livello di sanit , rappresentato dalla speranza di vita alla nascita.
- il livello d'istruzione, rappresentato dall'indice di alfabetizzazione degli adulti (moltiplicato per due) e dalla percentuale di iscrizioni alle scuole superiori.

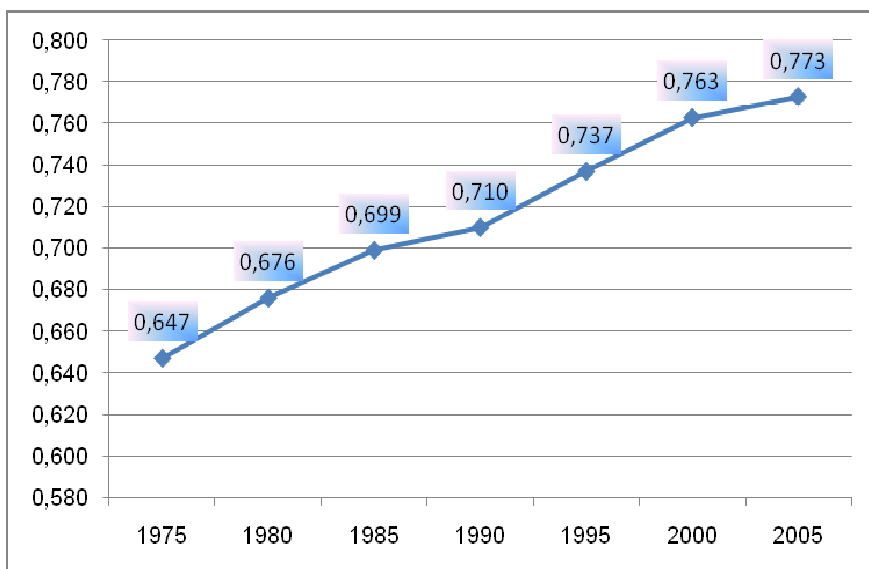
Dal rapporto UNDP 2007-08, che fa riferimento ai dati del 2005, il Per  occupa l'87esima posizione del *ranking* mondiale basato sull'Indice di Sviluppo Umano, (ISU 0,773), preceduto dalla Giordania e seguito dal Libano e dall'Ecuador. Secondo l'UNDP, il Per  si colloca nella fascia media di sviluppo umano, che va da ISU 0,500 a ISU 0,800.

¹ Fonte: "Calendario Atlante De Agostini 2008", Istituto Geografico De Agostini, Novara

Rispetto agli altri paesi Caraibici e del Sud America, il Perù si colloca sotto la media continentale (che supera lo 0,800), avvicinandosi molto di più a quella dei paesi arabi e dell'Est asiatico (di poco inferiore al valore peruviano).

Nel grafico è riportato il trend di tale indice per il Perù dal 1975:

Grafico 1.2. Indice di Sviluppo Umano per il Perù dal 1975 al 2005



Separando gli indicatori sui quali è costruito l'ISU, il Perù scende alla 93esima posizione per quanto riguarda sia la speranza alla nascita che il reddito individuale, mentre sale alla 63esima posizione per la percentuale di alfabetizzazione e persino alla 43esima sull'iscrizione alla scuola secondaria.

La particolare morfologia del territorio, e le diversità culturali presenti all'interno del Paese, fanno sì che tale indice cambi in modo significativo da regione a regione. Nella tabella seguente vengono riportati i 24 Dipartimenti più la Provincia di Callao, ordinati per l'Indice di Sviluppo Umano; sempre nella medesima tabella, viene indicata la distribuzione in percentuale della popolazione tra le 4 fasce geografiche individuate in precedenza, alle quali si aggiunge Lima metropolitana.

Tabella 1.2. ISU 2003 per dipartimento, e suddivisione della popolazione per area geografica

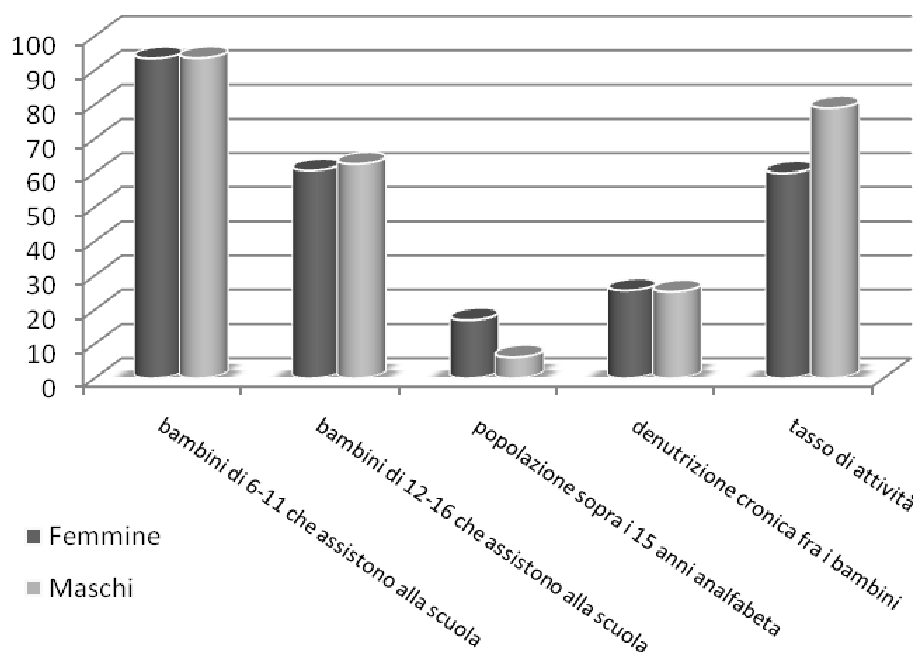
Dipartimento	ISU 2003 dati PNUD	Popolazione (%) per zona morfologica (dati DHS 2000)				
		Metropoli	Costa	Sierra	Selva Alta	Selva Baja
Lima	0,720	84,5	10,9	4,6	0,0	0,0
Callao	0,712	100,0	0,0	0,0	0,0	0,0
Tacna	0,665	0,0	92,3	7,7	0,0	0,0
Ica	0,662	0,0	100,0	0,0	0,0	0,0
Arequipa	0,653	0,0	14,4	85,6	0,0	0,0
Moquegua	0,650	0,0	74,6	25,4	0,0	0,0
Lambayeque	0,616	0,0	98,8	1,2	0,0	0,0
Tumbes	0,609	0,0	100,0	0,0	0,0	0,0
Madre de Dios	0,601	0,0	0,0	0,0	0,0	100,0
Junin	0,595	0,0	0,0	78,0	22,0	0,0
Pasco	0,591	0,0	0,0	76,3	23,7	0,0
La Libertad	0,585	0,0	70,6	29,4	0,0	0,0
Ucayali	0,563	0,0	0,0	0,0	4,2	95,8
Ancash	0,558	0,0	30,8	69,2	0,0	0,0
Piura	0,556	0,0	91,4	8,6	0,0	0,0
Puno	0,549	0,0	0,0	100,0	0,0	0,0
Loreto	0,525	0,0	0,0	0,0	0,0	100,0
San Martin	0,525	0,0	0,0	5,6	26,3	68,0
Cusco	0,511	0,0	0,0	91,6	7,0	1,4
Ayacucho	0,509	0,0	0,0	91,5	3,3	5,3
Amazonas	0,502	0,0	0,0	17,3	60,1	22,6
Cajamarca	0,491	0,0	3,9	79,3	16,8	0,0
Apurimac	0,488	0,0	0,0	100,0	0,0	0,0
Huanuco	0,476	0,0	0,0	67,9	32,1	0,0
Huancavelica	0,464	0,0	0,0	100,0	0,0	0,0

Si nota immediatamente che, ad eccezione di Arequipa, i Dipartimenti situati nelle zone montuose del Paese possiedono gli ISU più bassi.

Differenze fra i sessi

In Perù molte differenze di genere si sono col tempo appiattite, come ad esempio quelle sull'istruzione. Esistono ancora molte differenze però sul campo lavorativo e politico. Nel seguente grafico riportiamo alcuni indici a confronto tra maschi e femmine.

Grafico 1.3. confronto tra condizioni maschili e femminili secondo alcuni indicatori. Valori percentuali (2000, fonte INEI)



Le differenze di genere ovviamente non si possono riassumere in poche cifre, benché a volte molto significative. DHS (Demographic and Health Surveys), di cui parleremo approfonditamente nel prossimo paragrafo, campiona da anni gruppi di donne che vivono nei cosiddetti Paesi in via di sviluppo, per cercare in qualche modo di cogliere quali sono gli aspetti più disagianti della loro vita e quali invece i punti di forza su cui far leva per cercare di uscire dalla povertà. Nella tabella riportiamo alcuni indici sulle donne del Perù da loro calcolati dal 1986 al 2000. Da essi si vede che molti miglioramenti rispetto al passato, soprattutto in ambito sanitario.

Tabella 1.3. Evoluzioni temporali di alcuni indicatori sulla condizione femminile in Perù, 1986-2000

Indicatori	1986	1992	1996	2000
percentuale di donne senza istruzione	10,9	6,1	6,2	5,1
percentuale di donne senza accesso a giornali, televisione e radio	47,6	59,8	52,7	17,5
Numero medio di figli per donna	4,1	3,5	3,5	2,8
percentuale di donne che attualmente usano metodi moderni di pianificazione familiare	23	32,8	41,3	50,4
età mediana al primo matrimonio per le donne di 25-49 anni	20,5	21,1	20,9	21,4
età mediana al primo rapporto sessuale per le donne di 25-49 anni	18,9	19,4	18,9	19
numero ideale di figli	2,6	2,5	2,5	2,4
mortalità infantile (per 1000 nati vivi)	75,1	54,5	42,8	33,3
mortalità dei bambini sotto i 5 anni (per 1000 nati vivi)	110,2	77,5	59,1	46,7
percentuale di nascite che ricevono cure pre-natali da personale specializzato	54,8	66,1	67,3	82,6
percentuale di nascite in un istituto sanitario	-	48,4	49,6	57,9
percentuale di bambini che ricevono vaccini	-	57,7	63	66,3

Fonte: DHS

1.1.2 Il Dipartimento di Cusco

Il Dipartimento di Cusco è suddiviso in 13 Province e 108 Distretti, molto diversi fra loro. La peculiarità di questa regione, infatti, è quella di raggruppare dei territori che vanno dalla Giungla bassa (0 metri sul livello del mare), alle zone Alto Andine (fino a 6372m s.l.m.).

- *Regione alto andina.* Ubicata fra i 3500 e i 6372 metri s.l.m., è una zona dove le condizioni di vita sono molto difficili, e il clima vincola completamente le attività economiche, basate principalmente sull'allevamento di ovini, bovini e camelidi sudamericani, o la produzione agricola di patate.
- *Regione interandina.* Situata fra i 2500 e i 3500 metri s.l.m., rappresenta il 23% del territorio regionale, e vi ci risiede il 63% della popolazione totale; qui si concentrano le maggiori attività agricole ed economiche, grazie anche allo sviluppo dei collegamenti stradali.
- *Regione amazzonica e pre-amazzonica,* comprende nella sua totalità la provincia de La Convencion, una delle più estese e meno densamente popolate del Paese, e alcuni distretti di altre province.

Popolazione

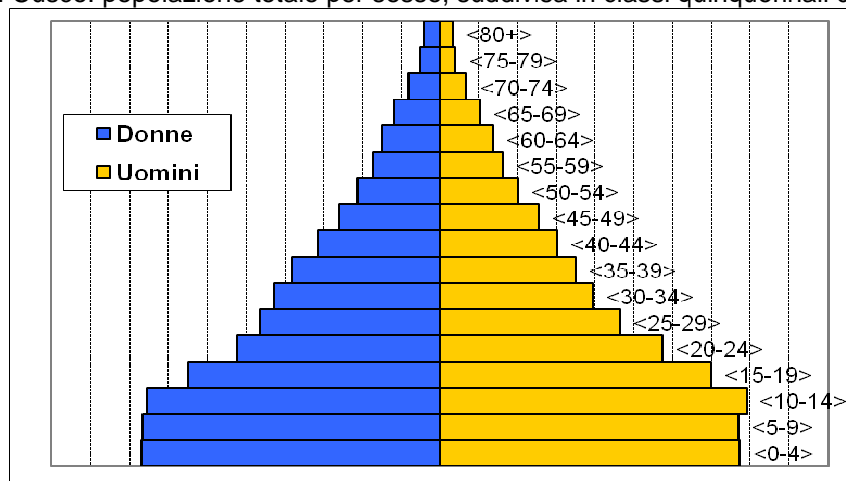
Secondo i dati dell'ultimo censimento generale del 2005, nella Regione di Cusco risiedono 1171503 persone, delle quali il 48,2% in zone rurali.

La densità è 16,3 ab./km², ma tale numero varia moltissimo da zona a zona, raggiungendo limiti minimi nella Provincia de La Convencion (5,5 ab./ km²), e massimi in quella di Urubamba (39,6 ab./ km²),

Le donne sono il 50,3% del totale, però tale percentuale rimane inferiore al 50% fino ai 24 anni d'età.

La piramide d'età mostra che la maggior parte della popolazione è di giovane età: il 35,04% infatti ha meno di 15 anni.

Grafico 1.4. Cusco: popolazione totale per sesso, suddivisa in classi quinquennali d'età (2003)



Caratteristiche economiche

Il tasso di occupazione (per la popolazione di almeno 15 anni compiuti) nel Dipartimento di Cusco è del 52,2%. Sul totale della popolazione economicamente attiva, appena il 31,5% è di sesso femminile (percentuale che raggiunge soglie minime del 16% in alcune province).

Fra le attività economiche più importanti nella regione c'è sicuramente il turismo. In generale esso sta crescendo a un tasso del 4% in tutto il Perù, ed è la seconda fonte di entrate del Paese dopo il settore minerario e prima di quello della pesca; ma questi risultati sono dovuti soprattutto a Cusco (dichiarata patrimonio dell'UNESCO), una delle città con maggiori riconoscimenti internazionali per una serie di organizzazioni del mercato turistico mondiale, e

al consistente numero di siti archeologici sparsi in tutto il territorio circostante, fondamentalmente della cultura Inka, pre Inka e coloniale. Inoltre ad attirare molti turisti sono anche i distinti ecosistemi naturali e l'abbondante biodiversità che rendono questa Regione la principale meta turistica del Paese.

È necessario però sottolineare il fatto che questo settore ancora non porta un effetto benefico alla popolazione autoctona, visto che nelle aree circostanti le mete turistiche si registra chiaramente una carenza economica e una precarietà della qualità di vita delle famiglie, soprattutto nelle zone rurali. Tale contraddizione si spiega col fatto che sono le imprese straniere a beneficiare della maggior percentuale delle entrate economiche portate dai turisti.

Il settore economico che coinvolge la frazione più cospicua della popolazione è quello agricolo (52,8% del totale degli occupati), in modo particolare in alcune zone più isolate, come la Provincia di Paruro (89,2%). In generale l'agricoltura regionale dagli anni '80 soffre di una forte decapitalizzazione, acuita da fattori climatici avversi e dall'instabilità della politica agraria; inoltre essa è caratterizzata dal predominio di piccoli agricoltori, dalla continua parcellizzazione delle terre, dai bassi livelli di redditività, dall'incapacità di organizzazione dei contadini e dallo scarso sviluppo tecnologico. Problemi che si accentuano nelle zone "Altas" (oltre i 4000 metri sul livello del mare) della Regione.

Fra i prodotti principali che vengono esportati ci sono il caffè, il cacao, l'achiote² e il mais bianco gigante.

Da uno studio fatto da CONVEAGRO³ nel 2005, è possibile ottenere il profilo del tipico produttore agricolo di Cusco:

- Grado di istruzione: educazione primaria 26.9%, secondaria 50,2% e superiore 2,1%; il restante 20,8% è analfabeta;
- Il 47% si dedica all'agricoltura, il 20% all'allevamento, il 33% ad entrambi;
- Il 36% frequentano corsi di miglioramento tecnico;
- Il 65,8% non è in grado di pianificare la propria attività in modo lungimirante;
- Bassa capacità di negoziazione;

² Frutto che cresce in territorio peruviano, il cui utilizzo risale agli antichi Incas come pianta medicinale

³ *Convención Nacional del Agro Peruviano*

- Scarsa autostima;
- Si abitua alla sua situazione, e non lotta per i suoi diritti.

In relazione ai quattro tipi di agricoltura in cui è suddivisa la produzione nazionale (agricoltura da esportazione, estensiva, per il mercato interno e di sussistenza), la tipologia che predomina nella regione è quella estensiva tradizionale (49%), seguita da quella dei piccoli produttori per il mercato interno; l'agricoltura di sussistenza invece riguarda il 9% del territorio, e si colloca nelle terre marginali della "sierra" e della "selva".

Rispetto all'allevamento, quello degli animali "maggiori" (ovini, bovini e camelidi) domina nelle regioni "alte" e della "sierra", mentre quello degli animali "minori" (principalmente dei *cuyes*, porcellini d'india, che nelle zone andine è uso comune consumare come cibo) assume grande importanza nelle aree della Giungla e in alcune zone più impervie della "sierra".

Molto arretrata è ancora l'attività di trasformazione dei prodotti e delle risorse agrarie; i livelli sono molto bassi soprattutto rispetto alle tecnologie moderne utilizzate, alla capacità di gestione e ai mercati dove vendere i prodotti. Gli sviluppi più importanti si sono verificati nel campo della panetteria, fresatura, cioccolateria e nella lavorazione dei latticini.

Lo sviluppo di alcune zone rispetto ad altre è favorito dall'accessibilità dei mercati, e quindi dall'espansione delle infrastrutture principali di comunicazione. La qualità di alcune strade principali (totalmente asfaltate e mantenute in un regolare stato di transitabilità) ha permesso la generazione di un'adeguata dinamica di relazioni fra le aree produttive e i centri popolati situati in questi assi strategici.

Però la rete stradale non raggiunge con gli stessi livelli di qualità tutte le province della Regione, e questo sfavorisce lo sviluppo di talune aree e non permette alla popolazione di accedere a mercati diversi da quelli locali. In particolare, a distinguersi per il basso livello di accessibilità sono le zone amazzoniche e quelle alto andine, spesso anche totalmente isolate per alcuni periodi dell'anno.

Situazione sociale

Il dipartimento di Cusco si caratterizza per un alto indice di mortalità. Fra le principali cause figurano in primo luogo le infezioni respiratorie acute e le infermità del sistema digestivo.

Il secondo gruppo di cause è riconducibile ai traumi in seguito ad incidenti esterni. Il terzo gruppo è costituito dalle malattie dell'apparato circolatorio.

Il tasso di denutrizione cronica fra i bambini sotto i cinque anni è del 43,2% (a livello nazionale la percentuale è del 7,1%); questo numero deve essere valutato con attenzione, visto che in realtà nasconde una problematica molto più critica, dal momento che tale indice raggiunge cifre anche superiori al 70% nelle cosiddette "province alte".

La regione fornisce un sistema educativo difettoso che non copre la totalità della popolazione in età scolastica e gli adulti analfabeti. Anche i contenuti curriculari non si adeguano alle realtà locali, da quelle rurali a quelle urbane, da quelle andine a quelle amazzoniche. Le famiglie con bambini o bambine che non frequentano la scuola sono il 2,8%.

Il tasso di analfabetismo della popolazione di almeno 15 anni è del 10%.

Rispetto ai servizi ritenuti di base, nel dipartimento di Cusco il 61% delle abitazioni non fruisce di un servizio di fognature, e il 36% non ha acqua corrente.

Dalla seguente tabella è possibile avere un quadro generale della situazione nella regione, confrontata con quella nazionale:

Tabella 1.4. Indicatori di povertà per la regione di Cusco e per il Perù

<i>PRINCIPALI INDICATORI</i>	<i>CUSCO</i>	<i>PERU'</i>
Tasso di mortalità materna (ogni 100.000 nati vivi)	163,0 *	185,0 ^
Tasso di mortalità infantile (su 1.000 nati vivi)	65,0 *	33,7 ^
Tasso di denutrizione cronica nei bambini sotto i 5 anni (%)	43,2 *	7,1 ^
Speranza di vita alla nascita	70,7 *	69,3 ^
Tasso di natalità 1995-2000 (per 1000)	28,8 *	23,5 ^
Tasso di analfabetismo (% della popolazione di 15 anni e più)	10,0 **	8,4 ^^

Fonti: * PNUD/INEI-IDH DEPARTAMENTAL 2000

** INEI, Censimento 2005

^ Pan American Health Organization - regional office of the World Health Organization (dati per il 2000)

^^ The Economic Commission for Latin America (ECLA) (dati per il 2005)

Un altro modo per capire le condizioni di vita della popolazione è attraverso l'Indice di Sviluppo Umano (ISU), elaborato dal Programma delle Nazioni Unite per lo Sviluppo Umano (PNUD). Rispetto a questo indice, calcolato per l'anno 2003, Cusco occupa la posizione 19 a livello nazionale (su 25, comprendendo quindi anche Callao), con un valore pari a 0.511 (valore peggiorato rispetto al 2000, ed inferiore alla media nazionale di 0.079)⁴.

Osservando i valori delle province di Cusco, si osserva che 6 di esse rientrano nella fascia più bassa di ISU (ultima fra esse, Paucartambo, con ISU 0.404, che occupa la 193esima posizione nel ranking nazionale su 194 province), mentre le restanti, ad eccezione della provincia di Cusco che possiede un ISU alto (0,606), si collocano nella fascia denominata "medio-bassa".

Una nota negativa della regione, è che l'ISU della maggior parte delle sue province è diminuito rispetto al 2000.

Nella tabella sono riportati anche i valori degli indicatori che vengono utilizzati per il calcolo dell'ISU; si osserva che alcune province si collocano in modo molto diverso rispetto ad essi nel ranking nazionale: ad esempio Canchis possiede il reddito pro capite familiare più alto, ma la speranza di vita alla nascita più bassa della regione.

Entrando ancora più nel dettaglio, per descrivere meglio la situazione locale, è più utile osservare come si distribuiscono i distretti di Cusco negli intervalli ISU stabiliti su scala nazionale; prestando attenzione alla mappa e allo schema riassuntivo proposto in seguito, si nota che Cusco possiede il 49,1% dei suoi distretti nel quantile più basso (quello dei più poveri), percentuale che è la più alta a livello nazionale (seconda solo alla regione di Huanuco).

A conferma di questo è il fatto che ben 7 distretti di Cusco rientrano nella classifica dei 20 più poveri del Perù, mentre nessuno fra i 20 più ricchi.

⁴ Questo indice unisce indicatori di speranza di vita, raggiungimento educativo e reddito familiare pro capite. I valori riportati fanno riferimento all'ultimo lavoro del PNUD sul Perù (2004), utilizzando come fonti INEI (Istituto Nazionale di Statistica peruviano), MEF (Ministero di economia e Finanza) e MINEDU (Ministero dell'Educazione peruviano)

Tabella 1.5. Indice di Sviluppo Umano provinciale del 2003 per il Dipartimento di Cusco, suddiviso nei quattro indici principali che lo determinano⁵

Dipartimento	Indice di Sviluppo		Speranza di vita alla nascita		Alfabetismo		Immatricolazione secondaria		Reddito familiare pro capite	
	ISU	ranking ⁶	años	ranking	%	ranking	%	ranking	in soles	ranking
Cusco	0,511	19	66,4	23	83,7	18	68,1	19	221,2	17
Cusco	0,607	32	71,4	36	94,5	18	90,1	37	305,3	45
Canchis	<u>0.518*</u>	100	61.6#	192	82,5	103	<u>87.5*</u>	49	<u>214.5*</u>	109
Espinar	0,507	113	65,5	148	83,7	94	73,6	106	193,5	140
Anta	0,504	117	63,7	173	78,8	126	79,4	86	190,6	143
Urubamba	0,499	126	65,9	145	<u>84.9*</u>	85	67,4	121	205,7	120
La Convención	0,480	141	<u>67.7*</u>	108	84,1	90	56,1	154	184,8	152
Calca	0,479	143	65,3	154	77,9	133	64,6	128	194,0	138
Canas	0,461	161	62,1	190	80,0	121	63,7	130	187,3	150
Chumbivilcas	0,429	182	63,5	178	68,8#	168	52,5	164	173,4#	166
Quispicanchis	0,425	187	63,0	188	75,7	144	47,1	182	189,8	147
Acomayo	0,418	189	64,0	170	76,7	139	40,8	189	193,8	139
Paruro	0,407	191	63,4	181	70,7	164	41,1	188	182,3	156
Paucartambo	0.404#	193	64,5	165	71,0	162	37.1#	192	177,5	161

Tabella 1.6. distribuzione dei distretti di Cusco rispetto ai quantili dell'ISU nazionale nel 2003

	Quantili 2003					Distretti numero	ISU Cusco	
	[0.3227, 0.4579]	[0.4580, 0.5045]	[0.5048, 0.5508]	[0.5509, 0.6009]	[0.6011, 0.7883]		2003	2000
Cusco	49,1	24,1	16,7	7,4	2,8	108	0,511	0,537

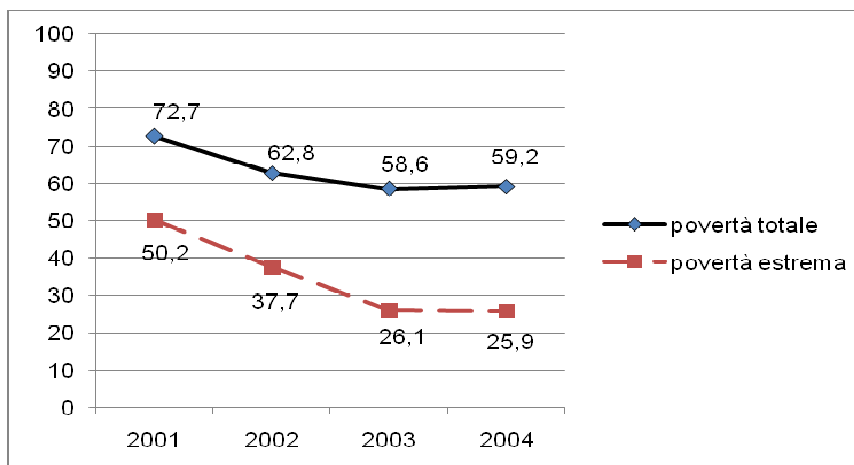
Queste informazioni sulla situazione di disagio sociale, in cui vivono la maggior parte degli abitanti del dipartimento, sono confermate dall'Indagine Nazionale delle Abitazioni sulle Condizioni di Vita e Povertà (ENAHO): Cusco rientra fra i cinque dipartimenti con maggior povertà estrema.

Nel grafico successivo viene riportato il variare della percentuale di popolazione che vive in condizioni di povertà e povertà estrema, dal 2001 al 2004.

⁵ Gli indici sottolineati e marcati con l'asterisco si riferiscono ai valori più alti (escludendo quelli della provincia di Cusco, che sono più alti sempre); quelli col simbolo “#” sono invece i più critici

⁶ Il ranking per il dipartimento fa riferimento a 25 (i 24 dipartimenti più la provincia di Callao); quello per le province viene calcolato rispetto alle 194 province presenti sul territorio nazionale

Grafico 1.5. Incidenza della povertà totale e della povertà estrema dal 2001 al 2004.

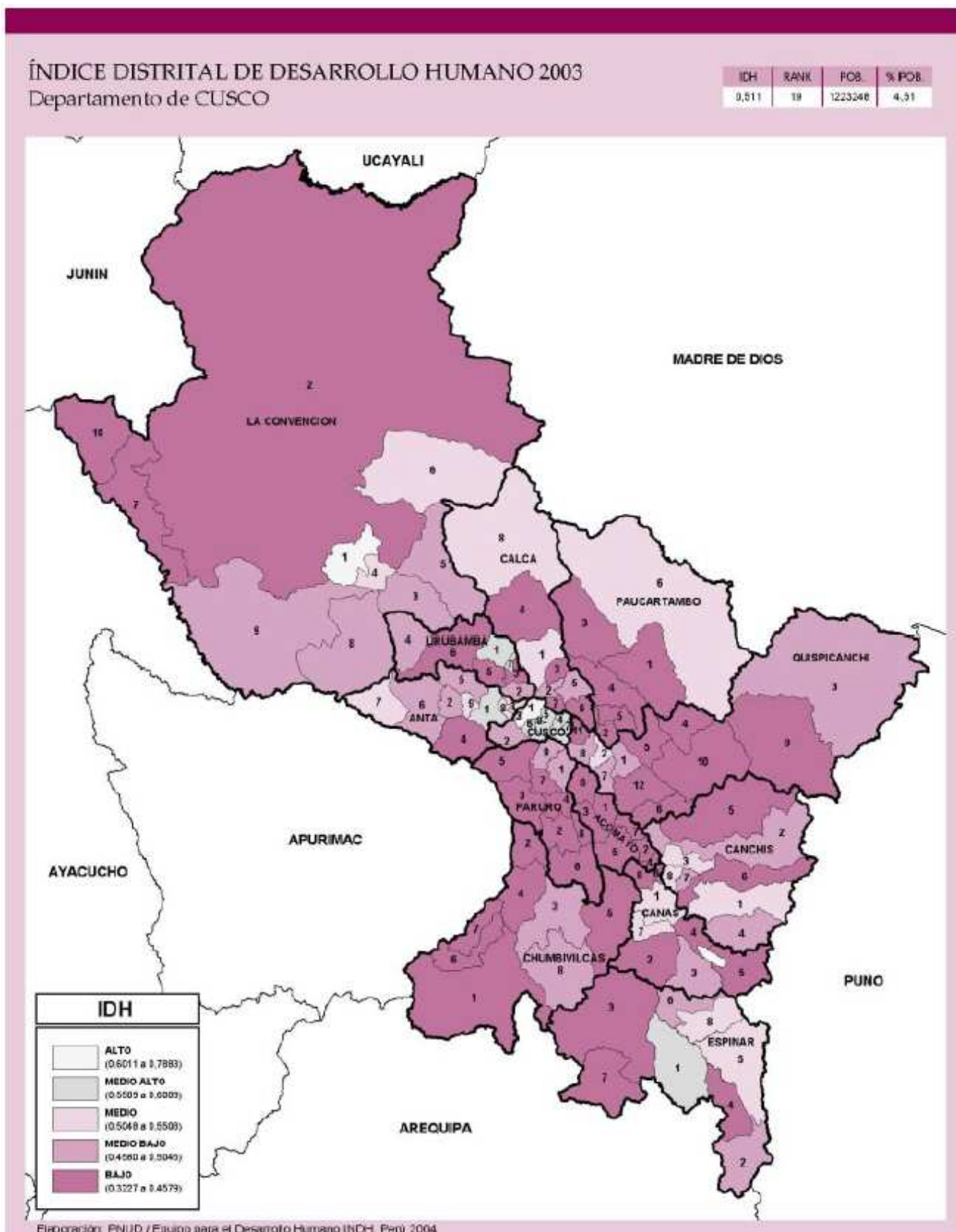


Infine, analizzando l'ultima tabella, si osserva l'enorme differenza del tasso di mortalit  infantile rispetto ad alcune variabili demografiche, quali il sesso, l'et  della madre al parto, la durata dell'intervallo fra un parto e il successivo: quando la madre   molto giovane o quando l'intervallo fra i parti   inferiore ai due anni, il rischio di mortalit  infantile diventa molto alto, superando il 100per mille.

Tabella 1.7. Mortalit  infantile rispetto ad alcune variabili sociali e demografiche (valori riferiti al 2000)

	Mortalit� Infantile per 1000
Totale	84.0
Sesso del neonato	
maschio	62.9
femmina	90.2
Ordine di nascita	
1	94.4
2 o 3	95.5
4-6	85.8
7 +	47.7
Et� della madre al parto	
< 20	137.6
20-29	82.4
30-39	64.0
40-49	71.4
Durata dell'intervallo fra i parti	
< 2 anni	116.4
2 anni	74.9
3 anni	76.9
4+ anni	53.5

Mapa 1.3. ISU 2003 nei distretti del Dipartimento di Cusco



1.2 DEMOGRAPHIC AND HEALTH SURVEYS (DHS)

I dati che si utilizzeranno nelle analisi che seguono sono stati forniti da Measure DHS (Demographic and Health Surveys).

Dal 1984 il progetto Measure DHS ha promosso più di 200 indagini in 75 Paesi, col fine di comprendere il livello di salute e lo sviluppo della popolazione nei Paesi in via di sviluppo. L'obiettivo strategico di Measure DHS è di migliorare e istituzionalizzare la raccolta e l'uso dei dati riguardanti molti Paesi per un monitoraggio programmato e per la valutazione di decisioni politiche sullo sviluppo. DHS fu fondata da USAID (United States Agency for International Development) col contributo di altri donatori.

DHS ha guadagnato una reputazione mondiale per la raccolta di dati, rappresentativi a livello nazionale, e per l'accurata distribuzione di essi; sono dati sulla fecondità, sulla pianificazione familiare, sulla salute di madri e figli, sulla sopravvivenza dei figli, su HIV/AIDS, sulla malaria e sulla nutrizione.

Fra i servizi offerti da DHS c'è: la raccolta e l'analisi di dati, la collaborazione con enti locali e strutture internazionali, la diffusione di dati e risultati, e l'aiuto tecnico ed informatico alle organizzazioni locali durante le rilevazioni.

I tre tipi d'indagini più frequentemente promosse, ad intervalli regolari, sono: le Indagini sulla Demografia e Salute, le indagini specifiche sugli indicatori per il monitoraggio dei programmi sull'AIDS, e le indagini sui servizi statali offerti alla popolazione su salute e pianificazione familiare.

Le indagini sulla Demografia e Salute, dalle quali sono stati ricavati i dati per il nostro studio, sono rappresentative, a livello nazionale, con campioni ampi (solitamente da 5000 a 30000 persone). Esse procurano dati per un'ampia gamma di indicatori di monitoraggio e di valutazione d'impatto su popolazione, salute e nutrizione. Tipicamente le indagini DHS sono condotte ogni 5 anni per permettere un confronto temporale.

L' "*Encuesta Demografica y de Salud Familiar del 2000*" (Endes2000) in Perù è una rilevazione che si realizza all'interno della quarta fase del programma mondiale delle rilevazioni sulla Demografia e Salute, conosciuto come *DHS+*.

Il campione utilizzato per l'indagine è un campione probabilistico, selezionato in tre fasi: selezione dei centri abitati nel primo stadio, dei conglomerati abitativi

nel secondo, e delle abitazioni nel terzo. Il campione prevedeva 1414 conglomerati, proporzionati fra centri urbani e campagna per ogni dipartimento: quindi circa 50 per dipartimento, con l'eccezione di Lima dove se ne selezionarono 226.

Il campione finale permette di stimare indicatori per le abitazioni e per le donne fra i 15-49 anni d'età, grazie al numero consistente di variabili, sia a livello dipartimentale che per le tre regioni geografiche (costa, sierra e selva).

Furono visitate oltre 33mila abitazioni, nelle quali furono intervistate 29.423 donne eleggibili (delle quali 27.843 hanno compilato il questionario in modo completo). Il campione è rappresentativo della popolazione totale del Paese, e nessuna area geografica è stata omessa deliberatamente dal disegno campionario.

Il tasso di risposta è stato del 94,6%.

1.2.1 Il questionario

Nella rilevazione di Endes2000 si applicarono due questionari, uno per l'abitazione e l'altro individuale per le donne nelle abitazioni campionate.

L'obiettivo base del programma DHS è che i dati siano confrontabili tra le varie nazioni. Perché questo accada è stato sviluppato un modello di questionario standard, al quale ogni nazione può aggiungere delle sezioni specifiche a seconda delle esigenze (per esempio, nel caso del Perù, la sezione sulla Violenza domestica).

Nel tempo questi modelli standard sono stati modificati e rivisti, cercando però sempre di mantenere le stesse codifiche; se questo è quindi un aspetto che facilita il confronto fra Paesi e l'imputazione dei dati, da un punto di vista più operativo invece complica sia la comprensione da parte di studiosi esterni, sia la ricerca nel dataset delle variabili presenti nel questionario.

Nel dataset finale inoltre sono state create delle variabili riassuntive, o sono stati standardizzati alcuni dati.

Questionario sull'abitazione

Vengono rilevate alcune caratteristiche generali della famiglia attraverso il sesso del capo famiglia, l'età, l'utilizzo delle strutture sanitarie pubbliche,

l'attività economica, il livello di istruzione e altre informazioni generali. Nello stesso tempo raccolgono informazioni sull'abitazione, come la presenza d'acqua potabile, la vicinanza a strutture sanitarie, il possesso di elettrodomestici e mezzi di trasporto, il combustibile utilizzato per cucinare, il numero di stanze usate per dormire, il materiale dominante del pavimento, dei muri e del soffitto.

Questionario individuale sulla donna

I temi principali di analisi presenti in questo questionario, sono suddivisi in 12 sezioni.

- I. **Caratteristiche generali riguardanti il passato dell'intervistata:** luogo di residenza nell'infanzia, livello d'istruzione e alfabetismo, periodicità nella lettura di quotidiani, lingua parlata;
- II. **Comportamenti riproduttivi:** anno di nascita dei figli ed eventuali anni di morte, gravidanze, conoscenza dei giorni fertili, aborti;
- III. **Contraccezione:** conoscenza e uso di metodi contraccettivi e fonti di somministrazione, chi decide nella coppia la pianificazione familiare e ragioni per non usare i metodi moderni;
- IV. **Salute:** utilizzo di strutture sanitarie per il parto e controlli post parto, peso dei neonati, copertura antitetanica, vaccini e malattie, alimentazione dei figli;
- V. **Nuzialità:** stato coniugale e informazioni sull'attività sessuale;
- VI. **Preferenze di fecondità:** desiderio di avere più figli e confronto con i desideri del partner, pianificazione familiare;
- VII. **Storia del marito e lavoro della donna:** età, istruzione e situazione economica dello sposo, esperienza lavorativa della donna, remunerazione e presenza di aiuti nel badare ai figli da parte di parenti e amici;
- VIII. **AIDS e altre malattie sessualmente trasmissibili (STD):** conoscenze e pregiudizi sulle malattie sessualmente trasmissibili;
- IX. **Mortalità femminile e familiare:** storia demografica (fino all'eventuale momento di decesso) di figli, parenti e fratelli;

- X. **Violenza familiare:** violenza fisica e psicologica sulle donne, violenza subita dai genitori in passato, forme di violenza sui figli come metodo educativo, a chi chiede aiuto in caso di violenza domestica;
- XI. **Peso e altezza:** misure dei figli nati dal 1995, dell'intervistata e del marito;
- XII. **Anemia:** livelli di emoglobina delle intervistate e dei loro figli nati dopo il 1995.

1.2.2 Aspetti operativi dell'indagine

Tre furono le fasi operative principali dell'indagine:

- la selezione e la formazione di 116 intervistatori e 62 supervisori (3 settimane)
- la rilevazione dei dati, iniziata il 17 Luglio simultaneamente nei 24 Dipartimenti, e conclusasi il 22 Novembre del 2000
- inserimento dei dati, verifica delle inconsistenze, calcolo degli errori e analisi dei dati

1.2.3 Dati e variabili utilizzati nello studio

In questo studio non è stato necessario utilizzare tutti i dati forniti da DHS: l'attenzione si è incentrata sul questionario individuale rivolto alle sole donne. Inoltre, per le motivazioni fornite in Premessa, si è ritenuto opportuno concentrare le analisi solo sulle 18764 intervistate che avevano avuto almeno una volta nella vita un partner fisso (quindi le donne sposate, conviventi, separate, divorziate e vedove).

Per quanto riguarda le variabili, nel capitolo successivo si descriverà come alla fine vengano selezionate solo 33 variabili sulle 4223 disponibili nel dataset. Poiché l'obiettivo del lavoro non è centrato su un tema specifico, ma ha la pretesa di vagliare differenti aspetti della vita dell'intervistata, idee e opinioni personali, sono state passate in rassegna trasversalmente quasi tutte le sezioni del questionario. Eccezione sono le sezioni sull'anemia, sul peso ed altezza e mortalità materna. Questa esclusione è dovuta o per disinteresse sul tema trattato, o per difficoltà di interpretazione del questionario per i motivi che si elencavano precedentemente. Per quest'ultima ragione, purtroppo sono state

sfruttate poco alcune sezioni che si ritenevano importanti ai fini dell'analisi, quali, ad esempio, quella sulla salute infantile.

1.3 METODOLOGIE UTILIZZATE

Come anticipato nella Premessa, il fenomeno trattato è molto vasto ed articolato: aspetti sociali si intrecciano continuamente con quelli economici e demografici, e questo è il motivo della complessità dell'ambizioso questionario di DHS, che cerca di cogliere quante più sfaccettature possibili del disagio presentato dalle donne intervistate. Di conseguenza anche i nostri dati sono molti ed illeggibili con semplici tabelle di frequenza; per questo motivo essi sono stati trattati con analisi statistiche multivariate, per cercare di studiare le relazioni presenti in molteplici variabili appartenenti a macroaree differenti.

Descriviamo sinteticamente ora le metodologie utilizzate durante l'analisi, e alle quali si fa riferimento nei prossimi capitoli.

1.3.1 Analisi delle corrispondenze

L'analisi delle corrispondenze è un metodo di studio multidimensionale particolarmente adatto per l'esame e la sintesi di informazioni di tipo qualitativo, specie se si ha la necessità di analizzare un'ampia matrice di dati. Inoltre, data la possibilità di considerare più variabili simultaneamente, permette di evidenziarne le relazioni interne e di ricavare le principali componenti associative del fenomeno oggetto di studio.

Esistono due tipi di Analisi delle Corrispondenze: *Semplice* e *Multiplo*. La prima analizza matrici di dati costituite dall'incrocio di due variabili, la seconda permette invece un'analisi di ampie matrici a più variabili. Le variabili studiate possono essere sia di natura qualitativa che dicotomica, ma anche quantitativa se opportunamente ridotte in classi ordinali; inoltre è possibile decidere di non utilizzare alcune variabili "attivamente" per la ricerca delle componenti principali, ma proiettarle solo successivamente nel grafico come "illustrative".

L'analisi delle corrispondenze è una particolare analisi fattoriale: essa quindi consente di estrarre nuove variabili - i fattori - che hanno la proprietà di riassumere l'informazione rilevante contenuta nelle innumerevoli caselle delle

tabelle di dati. Ciascun fattore è rappresentabile da una dimensione spaziale (o asse), al cui centro (o baricentro) c'è il valore "0" e che si sviluppa in modo bipolare verso le estremità negativa e positiva, in modo tale che gli oggetti collocati su poli opposti sono quelli più diversi tra loro. La nomenclatura del fattore viene di solito assegnata osservando appunto le modalità agli estremi degli assi.

In genere le modalità proiettate sugli assi cartesiani sono rappresentate da dei punti che assumono la forma di nuvole addensate attorno all'origine, che diventano sempre più rade quanto più ci si allontana dal centro degli assi. La forma più comune è quella ellissoide (con l'asse maggiore dell'ellisse nella direzione del fattore più importante), ma ci possono essere altre distribuzioni dei punti. Gli esempi più frequenti di forme sono:

- a triangolo attorno ad un asse fattoriale, che sta ad indicare la presenza di una dimensione principale unipolare, mentre la seconda (che corre lungo la "base" del triangolo isoscele) assume un ruolo decisamente minore;
- a ferro di cavallo, detto effetto Guttman, che indica una sostanziale unidimensionalità delle frequenze osservate;
- a nuvole separate, che di solito dimostrano la presenza di due o più insiemi di dati che non si compenetrano (in questo caso è preferibile svolgere analisi differenti)

Per interpretare i fattori si utilizzano una serie di misure di sintesi, quali:

- il *coseno quadrato*, che indica il contributo che un determinato fattore fornisce alla spiegazione della modalità;
- l'*inerzia*, ovvero il contributo relativo di una modalità alla varianza totale.

Scelte effettuate per lo studio

Nel nostro caso specifico il dataset è composto in prevalenza da variabili qualitative; fatto questo che sottolinea l'adeguatezza dell'analisi delle corrispondenze. Verrà condotta un'analisi multipla delle corrispondenze perché il numero di variabili analizzate è superiore a due.

Per studiare i fattori e la quota di inerzia da loro spiegata, si farà riferimento sia al chi-quadrato, sia agli “aggiustamenti” proposti da Greenacre e da Benzecri; essi ritengono che gli ammontari di inerzia forniti dagli autovalori sono una misura pessimistica della bontà del modello, e propongono delle diverse trasformazioni degli autovalori. In particolare Benzecri indica un limite, 1 fratto il numero di variabili attive, sotto il quale l’autovalore, e quindi il fattore ad esso associato, diviene trascurabile.

Oltre al coseno quadrato e all’inerzia precedentemente spiegati, si farà riferimento ad ulteriori misure di sintesi che saranno di volta in volta illustrate durante le analisi.

Nei grafici che compariranno nei prossimi capitoli, l’asse verticale delle ordinate rappresenterà la prima dimensione, mentre quello orizzontale delle ascisse la seconda dimensione.

L’analisi è stata condotta con il programma SAS.

1.3.2 Analisi di segmentazione

L’analisi di segmentazione è una tecnica multivariata di analisi che permette, a partire da una variabile dipendente, di studiare le relazioni tra variabili mediante la suddivisione progressiva del campione in sottoinsiemi di unità il più possibile omogenei al loro interno. Si tratta perciò di una tecnica di analisi asimmetrica che richiede la specificazione di una variabile dipendente e un insieme di variabili esplicative.

Questa metodologia di analisi è applicabile qualunque sia la scala della variabile dipendente (qualitativa, quantitativa o dicotomica), così come quelle delle variabili esplicative.

La segmentazione si distingue poi in base al numero di sottoinsiemi che possono essere formati nella suddivisione del gruppo di unità. Si dice binaria un’analisi che consente la partizione del campione a due vie; se invece di considerare suddivisioni a tre vie si parla di analisi ternaria; la segmentazione multipla infine consente la partizione del campione in un numero qualsiasi di vie.

Al primo passo il campione viene diviso in due o più sottoinsiemi caratterizzati dal valore di una variabile esplicativa; tali gruppi possono poi essere ulteriormente suddivisi sulla base dei valori che le unità assumono in corrispondenza di un'altra variabile indipendente. Le variabili vengono scelte in base alla loro capacità di formare gruppi il più omogenei possibile al loro interno. Con il termine "impurità" si intende la disomogeneità di un gruppo rispetto alla variabile dipendente. Per valutare la bontà della partizione normalmente viene calcolato un indice di miglioramento dell'impurità che segnala appunto il grado di omogeneità della suddivisione in relazione alla variabile dipendente in esame. Il processo viene interrotto secondo una precisa regola di arresto, che può essere la dimensione minima dei gruppi, la minima capacità esplicativa della migliore suddivisione, la minima devianza del gruppo genitore o un numero massimo di passi del processo.

L'analisi della segmentazione consente di ottenere un grafico, detto albero o dendrogramma, in cui i nodi rappresentano i gruppi di unità nei diversi stadi dell'analisi, i rami sono i criteri utilizzati per la suddivisione del campione e le foglie rappresentano i nodi terminali per i quali non si è ritenuto utile suddividere ulteriormente il gruppo di unità.

La tecnica della segmentazione è una procedura *stepwise* che consente quindi di selezionare le variabili esplicative escludendo quelle che non apportano un contributo significativo nella suddivisione del campione sulla base della variabile dipendente scelta.

Scelte effettuate per lo studio

L'analisi di segmentazione non verrà usata per scopi predittivi, ma la si affiancherà allo studio delle corrispondenze regionali per descrivere al meglio le relazioni fra le variabili. Nello specifico, utilizzando i punteggi fattoriali ottenuti dalle unità rispetto alle due dimensioni individuate durante l'analisi delle corrispondenze, si costruiranno i dendrogrammi ponendo come variabili dipendenti le dimensioni stesse.

Le variabili esplicative inserite sono state quelle che hanno svolto un ruolo attivo nell'analisi delle corrispondenze: tali variabili sono sia di natura

categoriale, sia quantitativa, sia dicotomica. Si è deciso di non inserire le variabili supplementari.

L'analisi è stata condotta con il programma SPSS 15. Il programma da noi utilizzato comprende gli algoritmi CHAID (Chi-squared Automatic Interaction Detector), per la segmentazione multipla di una variabile qualitativa, e CART (o CRT, Classification and Regression Trees) per la segmentazione binaria di variabili su qualsiasi scala.

Vista la natura quantitativa della variabile dipendente si è optato per l'analisi CRT, che prevede la sola partizione binaria: nel caso quindi che la variabile esplicativa contenga un numero maggiore di modalità, queste verranno accorpate in due sottoinsiemi.

L'algoritmo scelto suddivide il campione iniziale in sottoinsiemi massimizzando l'omogeneità interna ai gruppi formati. Ad ogni passo della procedura verrà segnalata la variabile per la quale vengono formati i nuovi sottoinsiemi e l'indice di miglioramento, ovvero di riduzione dell'impurità del nodo rispetto al gruppo genitore.

Abbiamo deciso di arrestare l'algoritmo di analisi quando i nodi comportano un miglioramento dell'omogeneità interna dei gruppi inferiore all'1%, o quando contengono un numero di unità inferiore a 50. Questa decisione non è molto selettiva, ma è stata ritenuta soddisfacente vista la natura descrittiva della nostra analisi. Per motivi analoghi, si è deciso di non forzare l'algoritmo "pagando" una penale (di devianza spiegata) di simmetria dell'albero.

1.3.3 Analisi di regressione logistica

L'analisi di regressione logistica è un modello di regressione applicato quando la variabile y dipendente è dicotomica. L'obiettivo è quello di studiare il possesso o meno di un determinato attributo dicotomico, rispetto ad un insieme di variabili esplicative. Si predilige la logistica all'analisi di regressione lineare, perché in questo modo si ha la possibilità di ottenere una stima di Y (la cui distribuzione è binomiale) compresa fra 0 e 1.

Indicando con p la probabilità che l'evento osservato si verifichi (e quindi che $Y=1$), e con x_i le q variabili esplicative, la funzione di regressione logistica assume la seguente forma:

$$\text{logit}(p) = \beta_0 + \sum_{i=1}^q \beta_i x_i = \mathbf{X} \boldsymbol{\beta} ,$$

dove con il termine “logit” si esprime il logaritmo naturale del rapporto fra probabilità di “successo” e quella di “insuccesso”. La funzione di Y si può quindi scrivere come una funzione logistica:

$$p = \frac{e^{x\beta}}{1 + e^{x\beta}} .$$

La stima dei parametri ignoti β è effettuata con il metodo della massima verosimiglianza, che consiste nella massimizzazione della probabilità di osservare l’insieme di dati osservato, in funzione di β .

La selezione delle variabili esplicative da inserire nel modello finale è una selezione *stepwise*, che consiste nella scelta di una variabile alla volta, prediligendo quelle che meglio spiegano la variabile dipendente.

Per verificare la bontà del modello complessivo considerato, è sufficiente confrontarlo con quello “saturato”, cioè quello che, avendo tanti parametri quante variabili esplicative presenti, si adatta perfettamente ai dati. La quantità utilizzata per tale confronto è detta “devianza”, ed è definita da:

$$D = -2 \log \left\{ \frac{\max[L(\textit{considerato})]}{\max[L(\textit{saturato})]} \right\} .$$

Questo rapporto di verosimiglianza D si distribuisce asintoticamente come un χ_{n-b}^2 , dove n è il numero di osservazioni e b quello dei parametri. Valori piccoli di D significano buon adattamento.

La significatività complessiva del modello può anche essere valutata calcolando la differenza tra la devianza del modello avente la sola intercetta e quella del modello considerato. Contrariamente a prima, tale differenza dovrà essere grande per far vedere che le variabili esplicative introducono una quantità di informazione significativa rispetto alla sola intercetta.

Scelte effettuate per lo studio

Nel nostro studio l’analisi di regressione logistica verrà utilizzata solo nel paragrafo 4 del prossimo capitolo, riguardante i problemi di assenza di privacy.

Prima dell'analisi vera e propria, è stato ritenuto opportuno effettuare una selezione preliminare delle variabili predittive, studiandone le singole dipendenze rispetto alla variabile dipendente Y . Per le variabili nominali e ordinali si è utilizzato la statistica Chi quadrato, per quelle dicotomiche si è stimato il Rischio Relativo, mentre per ogni variabile quantitativa abbiamo costruito un modello logistico ad un predittore.

Per le analisi è stato utilizzato il pacchetto operativo SAS.

2. PRIME ANALISI A LIVELLO NAZIONALE

Come introdotto nella premessa, l'obiettivo della ricerca è comprendere quali sono le variabili che meglio descrivono o determinano le situazioni di povertà nelle diverse regioni geografiche del paese. Per far questo partiremo da un'analisi a livello nazionale, per selezionare un campione di variabili da utilizzare successivamente per il confronto regionale.

Dal dataset fornito da Measure DHS, utilizzeremo solo i dati riguardanti le 18237 intervistate che hanno avuto almeno una volta nella vita un partner (si considerano quindi sposate, conviventi, separate, divorziate e vedove); questo numero è il risultato anche dell'eliminazione delle 527 "visitanti", cioè di quelle intervistate che non risiedono nell'abitazione campionata: esse infatti sono un fattore di disturbo per lo studio, perché non rispondono a molte domande del questionario che indagano sulle caratteristiche della casa e le dinamiche interne ad essa.

Il dataset è molto ricco: offre 4223 variabili. Per selezionare quelle utili all'analisi delle condizioni di povertà rifletteremo innanzitutto su alcune proposte di indicatori quotidianamente utilizzati da grosse organizzazioni che lavorano nell'ambito sociale, successivamente si cercherà di individuare alcune macroaree di intervento, e infine si selezioneranno le variabili presenti nel questionario.

2.1 PROPOSTE DI INDICATORI E MACROAREE D'INTERVENTO

Passiamo ora in rassegna le proposte di indicatori (e quindi di conseguenza le variabili a cui fanno riferimento) che si possono trovare in letteratura; esse ovviamente varieranno a seconda della grandezza e peculiarità dell'area geografica, e della tipologia d'intervento nel territorio. Ovviamente molti indicatori sono costruiti per valutare progetti specifici sul territorio, e quindi inutilizzabili ai nostri fini. Per ovviare a questo problema, e per non distrarsi

dall'obiettivo dello studio, abbiamo confrontato le idee di molte organizzazioni, soffermandoci soprattutto su quelle che lavorano col microcredito, e in contesti simili a quelli peruviani.

Ciò che accomuna la maggior parte delle grosse organizzazioni, è cercare di seguire una linea comune di lavoro per raggiungere gli stessi risultati di sviluppo: non è raro infatti sentire citati gli Obiettivi di Sviluppo del Millennio delle Nazioni Unite per il 2015 (Dichiarazione firmata nel settembre del 2000) nei programmi di lavoro, e molti indicatori si rifanno a quelli proposti dal PNUD (United Nations Development Programme) per monitorare il raggiungimento degli obiettivi. Quindi, prima di approfondire le proposte delle organizzazioni, esaminiamo tali Obiettivi di Sviluppo del Millennio e gli indicatori del PNUD:

- 1) Sradicare la povertà estrema e la fame: quindi garantire la sicurezza alimentare (numero di pasti al giorno, varietà di cibi, durata dei periodi di insicurezza) e ridurre il numero di persone che vivono con meno di un dollaro al giorno (e riescono ad accedere a servizi basici quali acqua, elettricità, ecc.)
- 2) Garantire l'istruzione primaria universale: eliminare quindi l'analfabetismo, sia maschile sia femminile, e garantire la qualità dell'insegnamento.
- 3) Promuovere la parità di sessi e l'autonomia delle donne: sia nell'educazione infantile, sia nel tasso di alfabetizzazione, sia nel lavoro e sia nelle decisioni politiche;
- 4) Ridurre la mortalità infantile: quindi ridurre i tassi di mortalità e alzare le percentuali di bambini vaccinati;
- 5) Migliorare la salute materna: ridurre il tasso di mortalità materno e alzare la percentuale di parti con assistenza di personale medico specializzato;
- 6) Combattere l'HIV/AIDS, la malaria ed altre malattie: cercare di bloccarne l'incidenza e ridurre gli effetti;
- 7) Garantire la sostenibilità ambientale: s'intendono quindi politiche e programmi ambientali nei Paesi, ridurre la percentuale di popolazione senza un accesso all'acqua potabile e ottenere un miglioramento significativo della vita di almeno 100 milioni di abitanti degli *slum* (baraccopoli);
- 8) Sviluppare un partenariato mondiale per lo sviluppo.

In conformità a questo, molte organizzazioni propongono indicatori di valutazione d'impatto sociale, alcuni sulle condizioni generali delle donne, altri legati chi più all'area economica, chi più a quella sanitaria, chi più a quella agricola. Presentiamo di seguito le proposte di un piccolissimo campione di organizzazioni, principalmente operanti in Perù o comunque in realtà simili del Sud America:

- Il Consorzio Imp-Act è un gruppo di organizzazioni che promuovono e supportano i servizi di microfinanza e la valutazione dell'impatto sulla povertà. Esso è attivo in diversi Paesi, e per ogni realtà propone indicatori differenti, sia a livello economico (ingressi, tipologia di lavoro, rete di mercati, etc.), sia di misura "sociale" (di sicurezza alimentare, salute, istruzione della donna e dei figli, sull'abitazione, etc.), sia sull'operato delle organizzazioni stesse (soddisfazione dei clienti, valutazione dell'assistenza al cliente e alla capacità di quest'ultimo di riuscire a risparmiare nelle proprie attività commerciali, etc.).
- RFR (Rete Finanziaria Rurale dell'Ecuador), una Corporazione privata senza fini di lucro, incentra le sue valutazioni più sulla capacità dei clienti di riuscire ad inserirsi attivamente nel contesto economico circostante: misurano quindi l'abilità nel riuscire ad accedere ai mercati, a partecipare a reti economiche, a guadagnare e soprattutto risparmiare per il futuro.
- La Fondazione Ford, che opera in disparati ambiti in tutto il mondo, per quanto riguarda la realtà cusquena propone indicatori sociali per valutare l'impatto del microcredito su: spese ed entrate familiari, proprietà, garanzie dell'impresa, scolarizzazione, condizioni dell'abitazione (proprietà, qualità di vita e materiali di costruzione), accesso a servizi pubblici, accesso a servizi sanitari, partecipazione delle clienti in comunità, presa di decisioni.

Al posto di quelli appena elencati, si è scelto alla fine di seguire lo schema di indicatori di FONDECAP, organizzazione che lavora in Perù nella costruzione di mercati per le zone rurali più povere, con l'obiettivo di aiutare i clienti sviluppandone le capacità di lavoro e relazioni. La scelta è caduta sul loro schema per la sua chiarezza e la sua completezza, in quanto copre molteplici

aree di intervento e sviluppo; inoltre la suddivisione in “stadi di povertà”, con descrizione sintetica per ognuno di essi delle condizioni delle persone, facilita la scelta delle variabili dal questionario, e l’eventuale suddivisione in modalità coerenti con la realtà che si vuole studiare.

Riportiamo di seguito le 14 aree individuate da FONDECAP, e fra parentesi qual è la condizione “soglia” delle donne che si possono definire **non povere**:

1. Terra e bestiame (possedere animali minori e di grandi dimensioni e/o avere una terra propria)
2. Accesso all’istruzione (gli uomini concludono la secondaria, le donne almeno il ciclo primario)
3. Attività economica e mercato (attività con un margine di accumulazione e partecipazione a mercati regionali)
4. Contributo nel lavoro della donna (lavoro della donna riconosciuto in famiglia e nella comunità)
5. Partecipazione della donna nell’organizzazione comunale (partecipazione simile per uomini e donne, e vengono riconosciute a quest’ultima eventuali posizioni di guida)
6. Capacità di negoziazione della donna (le donne negoziano individualmente)
7. Violenza domestica (maltrattamento occasionale, ma riconosciuto come anormale, e quindi spesso non ammesso o denunciato)
8. Autostima (donne e figli hanno cura personale)
9. Pianificazione familiare (donne conoscono e sanno mettere in pratica metodi di pianificazione non tradizionali)
10. Numero di figli (meno di 6 figli sotto i 7 anni)
11. Stato di nutrizione dei figli (no denutrizione cronica)
12. Alcolismo (no alcolismo nella coppia; si beve solo alle feste)
13. Condizioni dell’abitazione (mobilio completo, presenza di elettrodomestici, stanze al secondo piano, tetto e pareti di materiali sicuri, dislocazione in centro abitato, o con un terreno attorno)
14. Distribuzione del credito (prestito per attività in mercati anche non locali, che permettono accumulo)

A questo punto, per ogni punto dello schema si sono cercate le corrispondenti variabili nel questionario e se n'è valutata la qualità controllando eventuali filtri, facili da individuare, e il numero di non risposte.

Purtroppo le domande a disposizione non coprono tutte le aree individuate da FONDECAP, o le soddisfano solo in modo parziale o laterale; altre volte alcune domande interessanti sono state scartate per troppi valori mancanti o modalità di risposta troppo generali o specifiche (per esempio la variabile "lavoro della donna" che prevede un numero di modalità eccessivo, ed un eventuale accorpamento, che comunque non avrebbe ridotto di molto le modalità, avrebbe rischiato di provocare errori di valutazione). Le sezioni che rimarranno scoperte sono la 1, la 3, la 6, la 11 e la 14.

Di seguito riportiamo uno schema delle variabili selezionate per ogni macroarea, con una breve descrizione, e alcune variabili che non si è potuto utilizzare per i motivi precedentemente elencati. In allegato inoltre si trova, per ognuna di esse, il codice e la nomenclatura utilizzata in tutti i prossimi output (sia tabelle che grafici) per le variabili e le loro modalità.

Alla fine dello schema sono state aggiunte le variabili a carattere generale che descrivono il contesto geografico dove vive l'intervistata e la sua storia personale.

Un'ultima nota: non sono presenti variabili specifiche nel questionario per coprire l'area 5, la partecipazione della donna nell'organizzazione comunale (quelle più interessanti presentano troppi valori mancanti); abbiamo però ampliato il senso del tema basandoci sulle altre fonti bibliografiche: non essendoci un'area distinta sulla conoscenza del territorio, e ritenendola noi fondamentale per lo studio, abbiamo incluso in questo settore le domande che potevano in qualche modo valutare la conoscenza dell'intervistata della realtà circostante e la capacità di utilizzare i servizi pubblici in caso di difficoltà.

Inoltre è stata inserita un'ulteriore macroarea, non presente fra quelle di FONDECAP: la conoscenza dell'AIDS e delle malattie sessualmente trasmissibili (STD è la sigla inglese, che verrà utilizzata durante le analisi), e i pregiudizi verso coloro che ne sono affetti.

Tabella 2.1. Variabili inseriti ed escluse dall'analisi, rispetto all'area di studio

AREA DI STUDIO	VARIABILI INSERITE	CARATTERISTICHE E MODALITA'	VARIABILI ESCLUSE	MOTIVAZIONE ESCLUSIONE
1. Terra e bestiame	-	-	Proprietà del terreno	Valori mancanti
2. Accesso all'istruzione	Istruzione donna	4 modalità ordinate: nessuna, primaria, secondaria, superiore	Lettura di giornali e riviste	Modalità non chiare
	Alfabetizzazione ⁷	Dicotomica	-	-
	Differenza uomo-donna	Variabile creata manualmente		
3. Attività economica e mercato	-	-	Lavoro del partner	Modalità troppo specifiche, inaccorpabili
			Lavoro della donna	Modalità troppo specifiche, inaccorpabili
			Capacità di accumulo	Valori mancanti
4. Contributo donna nel lavoro	Donna lavora attualmente	Dicotomica	-	-
5. partecipazione nella organizzazione comunale	Sai dove comprare i preservativi?	Dicotomica	In caso di violenza, chiedi aiuto a istituzioni pubbliche?	Valori mancanti
			In caso di violenza, chiedi aiuto ad amici?	Valori mancanti
			Nell'ultimo anno, usati centri sanitari	Valori mancanti
6. Capacità di negoziazione	-	-	-	-
7. Violenza domestica	Maltrattamenti subiti dal partner	Dicotomica: violenza fisica	Il padre maltrattava la madre	Non significativa
	Usare le maniere forti per l'educazione dei figli	Dicotomica: opinione sulla violenza sui figli	Chi punisce i figli	Valori mancanti
	Il partner la offende verbalmente	Dicotomica: violenza morale	-	-
8. Autostima	Chi ha l'ultima parola sulle decisioni riguardanti la salute	3 modalità: l'intervistata, l'intervistata con altre persone, altre persone	-	-

⁷ Con la modalità "analfabeta" includiamo le donne che non sanno leggere o che sanno leggere solo piccole parti di frasi. Essa viene utilizzata ad integrazione della variabile "titolo di studio" in quanto trasversale alle prime due modalità ("nessuna" e "primaria")

	Chi ha l'ultima parola sulle decisioni riguardanti acquisti importanti	3 modalità: l'intervistata, l'intervistata con altre persone, altre persone			
9. Pianificazione familiare	Discussione della PF col partner	Dicotomica	L'ultimo figlio era pianificato?	Valori mancanti	
	Metodi contraccettivi utilizzati	Resa dicotomica: metodi moderni contro metodi tradizionali o folkloristici	Attuale metodo contraccettivo usato	Non idoneo per tutte le intervistate	
10. Numero di figli	Numero di figli negli ultimi 5 anni	4 modalità (l'ultima è ">3")	Totale figli	Inutile ai nostri fini	
			Numero di fratelli della donna	Difficile il recupero della domanda filtro	
11. Stato di nutrizione dei figli	-	-	Varietà di alimenti nella dieta settimanale	Valori mancanti	
12. Alcolismo	Donna ubriaca	3 modalità: no, a volte, spesso	-	-	
13. Condizioni dell'abitazione	Presenza di elettricità	Dicotomica	Presenza di...	Le altre variabili sugli elettrodomestici sono state scartate durante le analisi	
	Presenza di un frigo	Dicotomica			
	Materiale delle pareti	Dicotomica: grezzo contro finito	Materiale tetto / pavimento		scartate durante le analisi
	Tipo di bagno	3 modalità: assente, latrina, con acqua	Acqua potabile		scartata durante le analisi
14. Distribuzione del credito	-	-	-	-	
0. Variabili generali	Età dell'intervistata	In classi di 5 anni	-	-	
	Età al primo matrimonio /convivenza	Continua ma resa divisa in classi in base alla frequenza	Durata del rapporto	Inutile ai nostri fini	
	Stato coniugale	Sposata, convivente, vedova, separata, divorziata		-	
	Gruppo etnico	Modalità raggruppate in "spagnolo" e "minoranza"			
	Dipartimento	24: Lima è unito a Callao			
	Regione geografica	Metropoli, Costa, Altipiano, Giungla Alta, Giungla Bassa			
	Urbanizzazione	Dicotomica: città contro campagna	Urbanizzazione	Con distinzione fra "città grande" e "città piccola": troppo dispersivo	

	Sesso del capo-famiglia	Maschio, Femmina	
	Privacy	No privacy durante l'intervista => no risposte alle domande sulla violenza	-
15. Malattie sessualmente trasmissibili	Un insegnante può avere l'AIDS	Dicotomica: è una di una serie di domande sul pregiudizio rispetto ai malati di AIDS	-
	Conosci le STD (malattie sessualmente trasmissibili)	Dicotomica	
	Conosci l'AIDS	Dicotomica	

2.2 SELEZIONE DELLE VARIABILI

2.2.1 Ipotesi iniziali e problema dei *missing values*

Come detto in precedenza, sarà utilizzata l'analisi multipla delle corrispondenze per cercare di cogliere la struttura interna della nostra matrice di dati. All'inizio riporteremo, in modo veloce e schematico, i primi risultati conseguiti con le variabili selezionate a priori; tali variabili saranno in seguito confermate o cambiate con altre a seconda del loro apporto più o meno utile allo studio. Alla fine descriveremo dettagliatamente i grafici e le tabelle ottenute, e cercheremo di trarne alcune conclusioni.

L'analisi delle corrispondenze consente di distinguere tra variabili attive e supplementari. Inizialmente vengono impiegate nelle analisi 21 variabili attive e 5 illustrative (tra queste particolare attenzione deve essere rivolta ad età al matrimonio, perché alla fine la impiegheremo come variabile attiva). Le classi d'età al matrimonio sono state decise in base alle frequenze e ai risultati.

Si considerano come supplementari (lo stato coniugale è stato aggiunto in un secondo momento):

- gruppo etnico (spagnolo, minoranza)
- età dell'intervistata (15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49)
- urbanizzazione (urbano, rurale)
- età al matrimonio (<15, 15-17, 18-20, 21-30, >30)

- regione geografica (Lima metropolitana, costa, altipiano, giungla alta, giungla bassa)
- stato coniugale (sposata, convivente, vedova, divorziata, separata).

In un primo momento, sono stati esaminati nell'analisi anche i valori mancanti: con l'opzione "*missing*", SAS infatti considera le osservazioni che contengono dati mancanti, evitando quindi la perdita di unità statistiche. Si evidenzia immediatamente la presenza di due fattori principali: il primo (asse verticale) spiega, secondo la scomposizione del Chi-quadrato, il 12,24% dell'inerzia totale, il secondo (asse orizzontale) ne spiega l'8,54%: complessivamente quindi i due fattori spiegano quasi il 21%, percentuale che aumenta ad oltre il 26% con la correzione di Greenacre (15,58%+10,43%), e al 90,18% con quella di Benzecri (64,61%+25,27%).

Tabella 2.2. Elenco dei primi 6 fattori individuati con la correzione di Greenacre

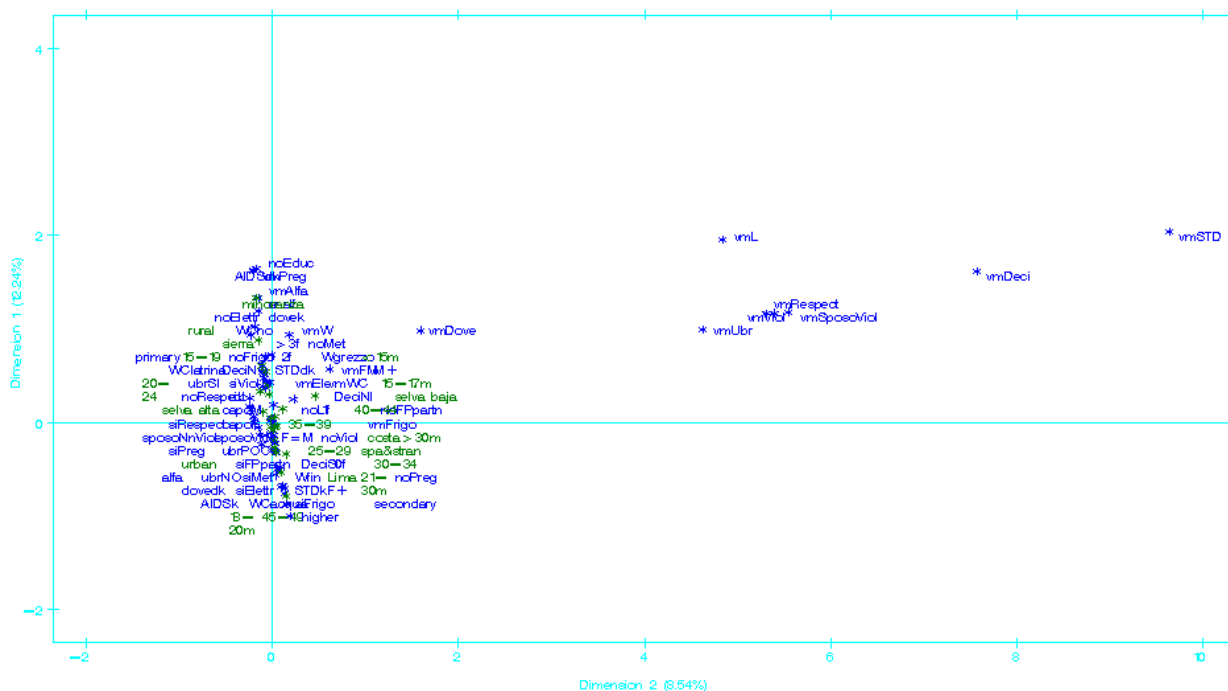
Scomposizione dell'inerzia corretta di Greenacre

Inerzia principale	Inerzia corretta	Percentuale	Percentuale cumulativa	3 6 9 12 15
0,26	0,23	15,58	15,58	*****
0,18	0,16	10,43	26,02	*****
0,10	0,08	5,33	31,35	*****
0,09	0,07	4,63	35,98	*****
0,09	0,07	4,42	40,40	*****
0,06	0,05	3,13	43,53	****
.
.
.
0,00	0,00	0,00	100,00	.
Totale	1,49	100,00		

Dal grafico si desume però che il risultato è singolare: quasi tutti i valori mancanti si collocano lungo il semiasse positivo delle ascisse; in modo particolare, i valori mancanti delle tre variabili sulla violenza, fisica, verbale e sui figli, contribuiscono maggiormente alla spiegazione dell'inerzia del secondo fattore, rispettivamente col 23,67% 23,18% e 22,71%, seguito subito dopo dall'alcolismo (20,76%). Ma anche le modalità *missing* che spiegano percentuali minori (tutte inferiori al 2%) hanno un ruolo importante per capire la peculiarità

della situazione: infatti quasi tutti i valori mancanti delle variabili le cui modalità determinano maggiormente la prima dimensione, sono più importanti per la seconda.

Grafico 2.1. Rappresentazione grafica delle modalità attive ed illustrative, rispetto ai primi due assi fattoriali, trovati con l'analisi delle corrispondenze sul Perù, con l'inserimento dei *missing values*



Nelle analisi successive dovremo eliminare i *missing* in modo definitivo, perché non è statisticamente plausibile che essi “occupino” totalmente un asse, particolarità che potrebbe influenzare le conclusioni e complicare l’interpretazione delle altre modalità nel grafico. La loro importanza però non permette di ignorarli completamente: non sono casuali, e probabilmente sono fortemente legati alla vergogna nel rispondere per una situazione di disagio materiale o psicologico che ora non riusciamo a vedere; sarà quindi dedicato in seguito un paragrafo intero per poter meglio analizzarli.

Nello studio delle corrispondenze senza valori mancanti, esiste un fattore dominante e un secondo molto meno importante. Secondo la scomposizione dell’inerzia corretta di Greenacre, il primo fattore spiega quasi il 18% dell’inerzia

totale, mentre il secondo poco più del 7%. Complessivamente quindi arrivano a spiegare circa il 25% dell'inerzia.

Tabella 2.3. Elenco dei primi 6 fattori individuati con la correzione di Greenacre

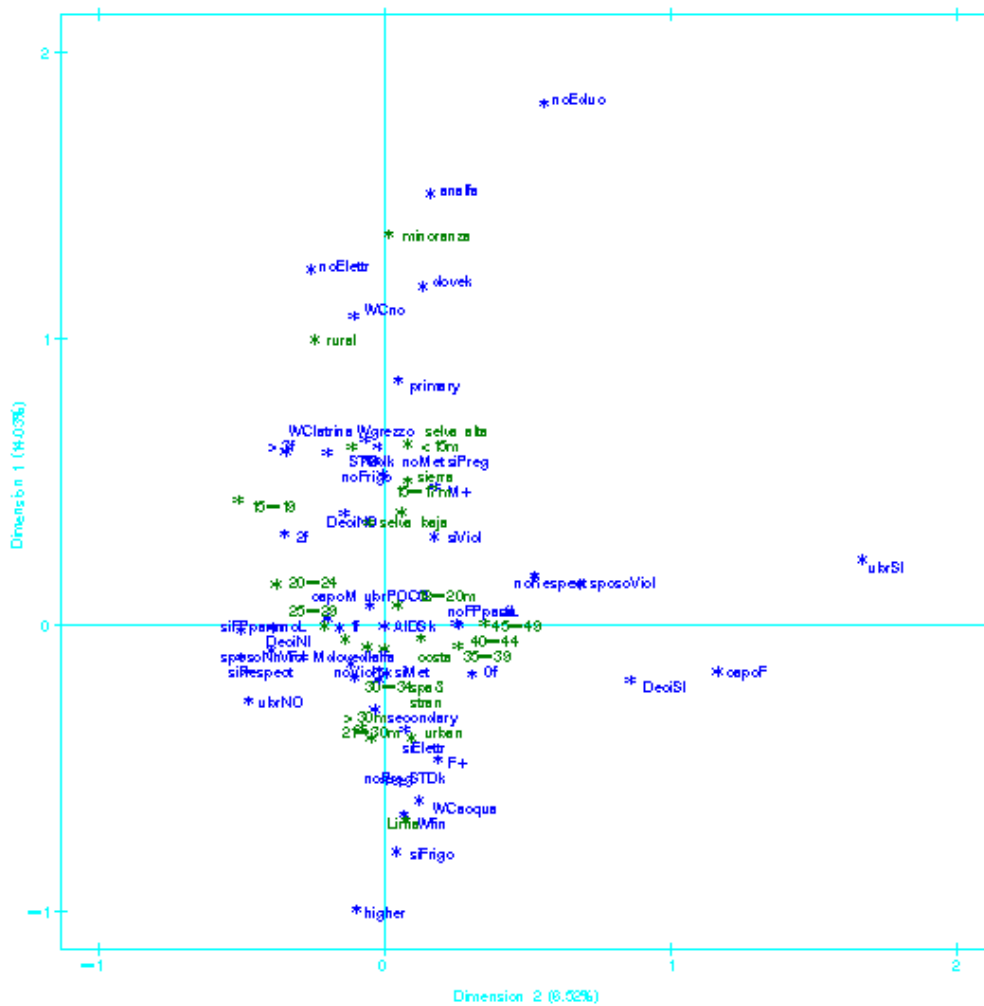
Scomposizione dell'inerzia corretta di Greenacre

Inerzia principale	Inerzia corretta	Percentuale	Percentuale cumulativa	4 8 12 16 20
0,19	0,16	17,74	17,74	*****
0,09	0,07	7,32	25,06	*****
0,08	0,06	6,19	31,24	*****
0,06	0,04	4,83	36,07	*****
0,06	0,04	4,72	40,80	*****
0,05	0,04	3,90	44,70	*****
.
.
.
0,01	0,00	0,54	100,00	*
Totale	0,92	100,00		

Proporzioni simili si ottengono usando la scomposizione del Chi-quadrato, mentre utilizzando l'algoritmo di Benzecri il primo spiega da solo oltre l'87% dell'inerzia totale, mentre il secondo appena il 7% (il restante è suddiviso fra gli altri 7 selezionati).

Si osservi ora il grafico. In colore blu vengono riportate le modalità delle variabili attive, mentre in verde sono segnate quelle delle variabili supplementari. A prima vista si capisce che l'asse verticale è dominato dalle modalità riguardanti l'istruzione delle donne e l'abitazione, mentre quello orizzontale dai comportamenti delle intervistate.

Grafico 2.2. Rappresentazione grafica delle modalità attive ed illustrative rispetto ai primi due assi fattoriali trovati con l'analisi delle corrispondenze per il Perù



In particolare, il semiasse positivo delle ordinate è caratterizzato dalla presenza di:

- no istruzione (cioè nessun titolo di studio)
- analfabetismo
- no elettricità
- conoscenza dove comprare preservativi
- no bagno interno
- istruzione della donna superiore al proprio partner

Dalla parte opposta invece vengono proiettate le seguenti modalità:

- livello di istruzione superiore
- si frigo
- bagno con acqua corrente

- muro con materiali finiti
- conoscenza delle STD
- no pregiudizi sui malati di AIDS

Le prime modalità che si incontrano nel semiasse positivo delle ascisse sono:

- alcolismo
- capo-famiglia donna
- capacità di prendere decisioni
- partner violento
- violenza verbale

Hanno invece coordinata negativa sull'asse orizzontale:

- no alcolismo
- partner non violento
- pianificazione familiare
- no violenza verbale
- presa di decisioni con altre persone
- non lavora attualmente

La proiezione sul grafico delle variabili supplementari indica che:

- le minoranze etniche abitano solitamente nelle zone più rurali e sono caratterizzate dai problemi più materiali, quali abitazioni senza bagno, elettricità e con muri precari;
- viceversa, invece, gli abitanti di Lima, e in genere chi abita nelle zone urbane, vivono in un contesto economico più agiato e hanno un'istruzione superiore (semiasse negativo delle ordinate);
- le abitanti della *sierra* e della giungla, alta e bassa, molto vicini fra loro, si collocano nel secondo quadrante, e sono caratterizzate da un'istruzione non superiore a quella primaria, e comunque inferiore a quella del partner, dall'uso di metodi contraccettivi non moderni, dal pregiudizio verso chi è affetto da AIDS, dalla non conoscenza delle STD, dall'idea che i figli si possano educare con metodi violenti, dall'impossibilità di poter prendere decisioni da sole e da abitazioni costruite con materiali

- grezzi e scadenti, e senza elettrodomestici; se inoltre vivono in zone rurali, le case non hanno neanche bagni ed elettricità;
- le donne giovani si trovano nel primo quadrante, quelle di mezza età (30-34) lungo il semiasse negativo delle ordinate, mentre quelle più anziane nel semiasse positivo delle ascisse: questo ad indicare che le più giovani sono alle prese più con i problemi economici, mentre quelle di età superiore hanno la testa rivolta principalmente al rapporto col partner e ai difficoltà che questa relazione può comportare;
 - le intervistate che hanno lasciato casa in età giovane per vivere col proprio partner si collocano nel primo quadrante e vivono in case povere; viceversa, raggiungono un livello di istruzione superiore e vivono in modo più agiato le donne che si sposano dopo i 20 anni.

La lettura del grafico è supportata dalla tabella seguente, dove vengono evidenziate le prime modalità per importanza nella spiegazione dell'inerzia di ogni dimensione.

Tabella 2.4. Migliori modalità per contributo parziale all'inerzia delle due dimensioni

Dimensione 1		Dimensione 2	
noElettr	0,088	ubrSI	0,141
siFrigo	0,067	capoF	0,108
analfa	0,063	DeciSI	0,108
primary	0,060	sposoViol	0,108
Wfin	0,055	sposoNnViol	0,079
Wgrezzo	0,054	noRespect	0,072
WCacqua	0,054	siRespect	0,066
higher	0,053	siFPpartn	0,045
WCno	0,052	DeciNI	0,042
noFrigo	0,049	noL	0,033
STDdk	0,041		
dovek	0,041		
STDk	0,039		
noPreg	0,037		
siPreg	0,036		
noEduc	0,029		

Le nostre ipotesi vengono confermate: l'asse verticale è dominato chiaramente dal livello di studio, dal possesso di beni materiali e dalla condizioni dell'abitazione, mentre quello orizzontale, benché meno forte, è caratterizzato da quelle modalità più psicologiche-comportamentali-decisionali, come alcool,

violenza (fisica, verbale e verso i figli), pregiudizi, lavoro donna e presa di decisioni. L'idea è quindi quella di denominarli come "disagio materiale" il primo, e "autostima" o "disagio sociale" il secondo.

Dopo queste analisi preliminari, abbiamo introdotto alcuni cambiamenti che riassumiamo ora schematicamente.

2.2.2 Età al matrimonio non supplementare

Il primo accorgimento è stato quello di utilizzare l'età al matrimonio come variabile attiva, perché fra le supplementari è quella meno spiegata dalle due dimensioni trovate.

Essa, una volta resa attiva nella determinazione dei fattori, si posiziona lungo l'asse verticale e, in particolare, le donne sposate prima tendono ad avere problemi più materiali: non è determinante come l'istruzione, ma si colloca allo stesso livello dell'abitazione. In particolare, le donne che si sono sposate più giovani sono caratterizzate da un'istruzione primaria, dalla non conoscenza delle malattie sessualmente trasmissibili, e da abitazione con la presenza di latrine al posto di normali servizi igienici; le donne sposate a 18-20 anni hanno problemi simili, a livello di abitazione (bagni all'aperto e assenza di elettrodomestici), e sono caratterizzate da un livello di istruzione inferiore a quello del partner e dal pregiudizio rispetto a chi è portatore di AIDS; le intervistate coniugate in fascia media d'età si collocano nel grafico proprio all'origine degli assi, mentre le donne sposate dopo i 30 anni si distinguono per un'istruzione superiore a quella del marito (o della persona con cui convivono), per la conoscenza delle STD e per l'assenza di pregiudizi verso chi ne è afflitto; vivono normalmente in città, in un'abitazione con presenza di elettricità e altre comodità di base.

Questa descrizione è comoda per capire le differenze all'interno della variabile stessa ma, osservando i valori di inerzia spiegata e confrontando la proiezione delle modalità come attive e come supplementari (coordinate molto simili), si evince che l'età al matrimonio ai fini dell'analisi delle corrispondenze è abbastanza inutile; quindi la differenza col tenerla come supplementare è praticamente nulla.

- in generale, tutte le donne attualmente senza un partner si trovano lungo il semiasse orizzontale positivo, cioè dove si presume ci siano problemi a livello di autostima; a differenza però delle divorziate, le separate e le vedove sono molto meno influenzate dal fattore economico: le prime, che comunque si trovano sempre nel quarto quadrante, sono caratterizzate, ovviamente, dalla modalità capo-famiglia donna, e dalla capacità di prendere decisioni; le seconde invece si posizionano nel secondo quadrante, e sono legate ai problemi di alcolismo;
- le sposate e le conviventi invece sono sull'asse delle ordinate: nella parte negativa (quindi discretamente benestanti dal punto di vista economico) le prime, in quella positiva le conviventi. Da quanto si evince dal grafico, le sposate si distinguono per un livello d'istruzione elevato (almeno secondario), per l'uso di metodi contraccettivi moderni, per la presenza di elettrodomestici dell'abitazione; esse vivono principalmente in città costiere, e sono di origine ispanica. Diversamente invece le conviventi sono più giovani, sono soggette a violenze e sono incapaci di prendere decisioni; vivono principalmente nella giungla (alta e bassa), in case senza elettrodomestici.

2.2.4. Scelta delle variabili che definiscono l'abitazione

Inizialmente alcune variabili erano state selezionate a priori per logica o conoscenze pregresse; molte di queste riguardavano l'abitazione, come, ad esempio, il materiale delle pareti, la tipologia di bagno, la presenza di elettricità e del frigo (la prima di quest'ultime due era stata scelta per individuare i più disagiati, la seconda, quasi sottoinsieme della prima, per indicare i maggiormente benestanti).

Il passo successivo è ora quello di scegliere definitivamente (o comunque confermare) le variabili più utili per definire i fattori. Tre sono i tratti caratteristici di un'abitazione: il materiale di costruzione, il possesso di alcuni beni (quali gli elettrodomestici basici e mezzi di trasporto più comuni), e la presenza di bagni ed acqua potabile.

- i. Rispetto ai materiali di costruzione con cui la casa è costruita, confrontiamo la variabile "muro", selezionata inizialmente, con le variabili "tetto" e "pavimento" delle abitazioni; la variabile che descrive il soffitto

- con “SqCos” si intendono i coseni al quadrato delle coordinate rispetto agli assi (minore è l'angolo fra l'asse e il vettore che unisce l'origine alla coordinata, maggiore è il coseno, fino ad 1, quando cioè una modalità sta esattamente sopra l'asse): essi quindi descrivono il contributo relativo che il fattore fornisce alla spiegazione della modalità della variabile;
- nelle colonne “Dim1” e “Dim2”, SAS assegna, ad ogni modalità, un punteggio da 0 a 2 che implica l'entità del contributo che quella modalità fornisce all'inerzia di quella dimensione;
- infine, l'ultima colonna “Best” assume valore 1 o 2 a seconda che la modalità contribuisca maggiormente alla prima o alla seconda dimensione.

Tabella 2.5. Misure a confronto per le variabili che definiscono l'abitazione

modalità	coord 1	coord 2	contr 1	contr 2	SqCos1	SqCos2	Dim1	Dim2	Best
Wfin	-0,71	0,11	0,056	0,003	0,511	0,012	1	0	1
Wgrezzo	0,71	-0,11	0,056	0,003	0,511	0,012	1	0	1
Ffin	-0,63	0,07	0,050	0,002	0,549	0,007	1	0	1
Fgrezzo	0,52	-0,06	0,003	0,000	0,015	0,000	0	0	1
Fnat	0,92	-0,10	0,068	0,002	0,493	0,006	1	0	1
Rfin	-0,56	0,14	0,031	0,005	0,260	0,018	1	0	1
Rgrezzo	0,47	-0,12	0,026	0,004	0,260	0,018	0	0	1

Anche dalla tabella, si conferma subito l'appartenenza delle tre variabili alla prima dimensione.

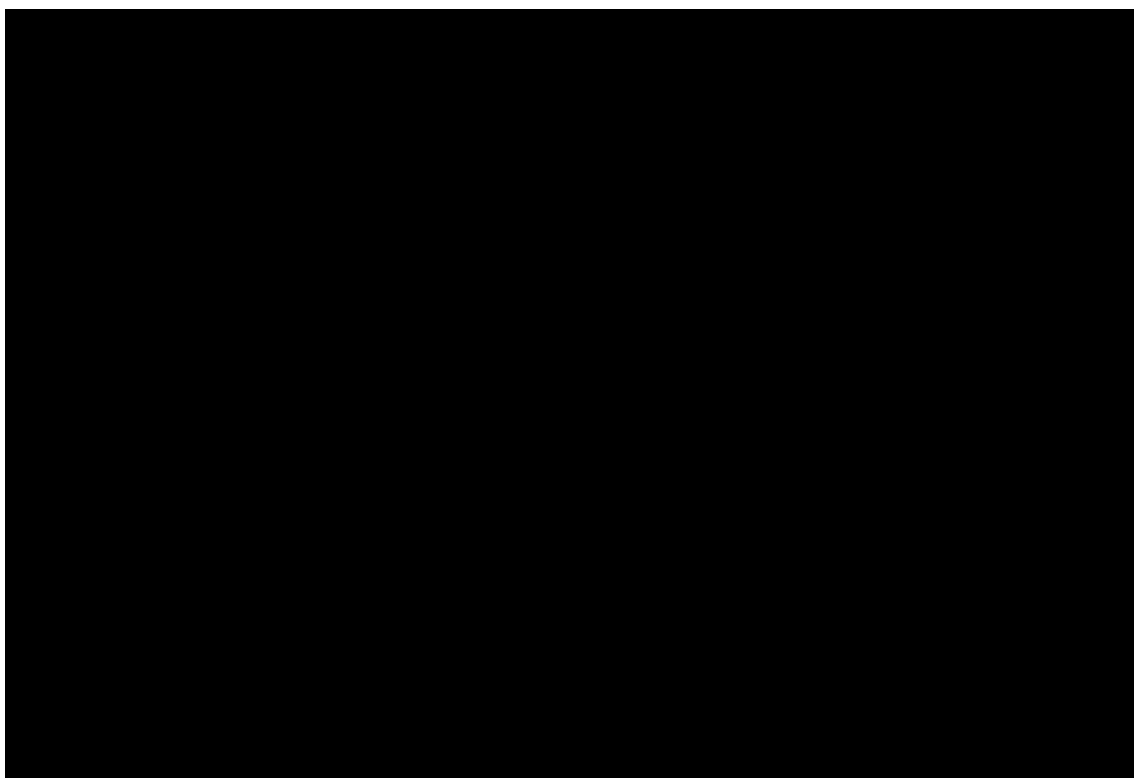
Come primo passo, si esclude il tetto perché meno importante nell'individuazione del fattore: esso infatti è quello con il coseno al quadrato minore, e una delle due modalità non risulta significativa per la spiegazione dell'inerzia della prima dimensione.

Ai fini dello studio, è necessario tenere solo una variabile che descriva il materiale di costruzione; è quindi necessario sceglierne una sola fra muro e pavimento, anche perché le due sono fra di loro molto correlate. La scelta è difficile, perché apparentemente sembrerebbero di uguale importanza: la variabile pavimento però risulta essere molto debole nella modalità mediana “grezzo”, e allo stesso tempo sarebbe una forzatura aggregarla a “naturale”.

In conclusione quindi si preferisce tenere “muro”.

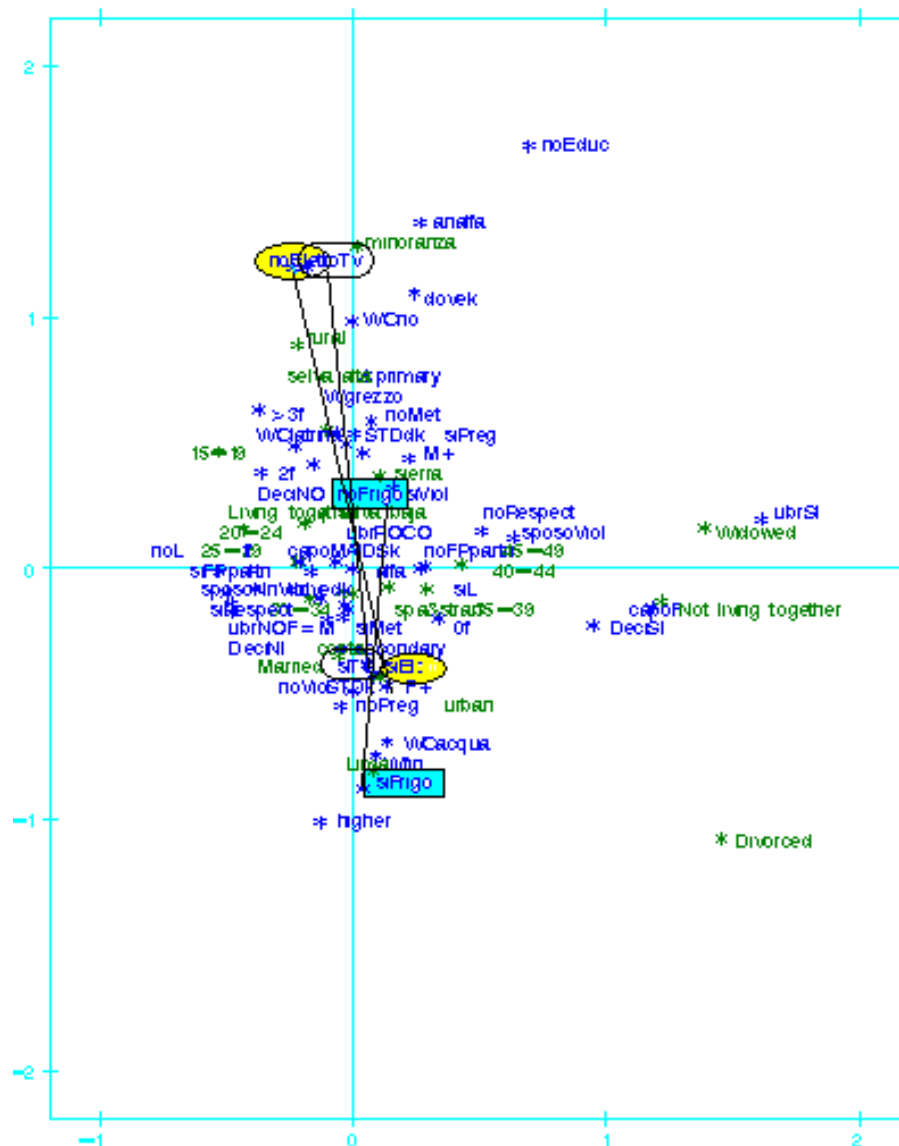
- ii. Fra i beni posseduti dalle famiglie rilevati nel questionario, si era inizialmente inserita la variabile frigo come protettiva della presenza di disagio economico-materiale (almeno a livello potenziale), e elettricità come predittiva; queste variabili sono state individuate in base alle loro frequenze e a delle ipotesi iniziali bibliografiche. Nel dataset sono presenti però altre variabili potenzialmente discriminanti, quali il possesso di radio, televisione, bicicletta, moto e macchina. Si è deciso di non considerare a priori la radio, in quanto la percentuale di chi la possiede è troppo alta, e nemmeno i mezzi di trasporto, perché viceversa sono in pochi coloro che ne usufruiscono o possono permetterseli, in modo particolare al di fuori delle zone costiere. Invece si proverà il confronto con la variabile televisione.

Grafico 2.5. Percentuale di intervistate che posseggono alcuni elettrodomestici e mezzi di trasporto



Nel grafico delle corrispondenze sono evidenziate con un rettangolo le modalità di “frigo”, con un cerchio scuro quelle di “elettricità”, e con uno chiaro quelle di “televisione”.

Grafico 2.6. Confronto delle variabili sui beni materiali nella rappresentazione grafica delle modalità rispetto ai primi due assi fattoriali trovati con l'analisi delle corrispondenze



Si nota subito che “frigo” ed “elettricità” si posizionano sull’asse verticale e contribuiscono alla spiegazione solo del primo fattore, però in maniera differente: chi non possiede l’elettricità in casa spiega circa il 9% dell’inerzia totale del fattore “disagio materiale”, mentre chi la possiede appena il 2,6%; il contrario avviene invece per il frigo, dove la modalità “presenza” spiega una percentuale maggiore dell’assenza dell’elettrodomestico. Questo a conferma del valore protettivo, rispetto alla presenza del disagio, della presenza di un frigo all’interno dell’abitazione, contro il valore predittivo di chi non ha l’elettricità.

L'inserimento della variabile "televisione" non sembra portare grandi miglioramenti: essa infatti nel grafico quasi si sovrappone ad "elettricità". Alla fine si preferisce quest'ultima, perché spiega maggior inerzia.

Tabella 2.6. Misure a confronto per le variabili che definiscono il possesso di beni

modalità	coord1	coord2	contr1	contr2	SqCos1	SqCos2	dim1	dim2	Best
noElettr	1,33	-0,28	0,092	0,009	0,509	0,022	1	0	1
siElettr	-0,38	0,08	0,026	0,002	0,509	0,022	1	0	1
noFrigo	0,58	-0,02	0,044	0,000	0,454	0,001	1	0	1
siFrigo	-0,79	0,03	0,061	0,000	0,454	0,001	1	0	1
noTv	1,34	-0,22	0,083	0,005	0,447	0,013	1	0	1
siTv	-0,33	0,06	0,021	0,001	0,447	0,013	0	0	1

- iii. L'ultima variabile che si desiderava inserire per descrivere l'abitazione doveva essere in grado di misurare la comodità e funzionalità dei servizi presenti all'interno di essa. La prima bozza ha puntato sui servizi igienici, opportunamente divisi in tre modalità: assenza di un bagno strutturale ("Wcno"), la presenza di un bagno comune all'aperto o comunque di una latrina (Wclatrina"), l'esistenza di un bagno con acqua corrente (Wcacqua").

Un altro servizio fondamentale però è anche la possibilità di accedere ad acqua potabile. Questa variabile raccoglie varie possibilità di fonti, che sono accorpabili sempre in tre tipologie: acqua in tubature dentro l'abitazione ("siACQUA"), acqua potabile da sorgente pubblica ("wellWATER"), acqua piovana o di fiume ("noACQUA")

Tabella 2.7. Frequenze assolute della variabile "acqua" e "bagno", e test del Chi-quadrato

		ACQUA			Totale
		no	well	si	
BAGNO	no	2119	478	2098	4695
	latrina	1622	966	3104	5692
	acqua	100	168	7207	7475
	Totale	3841	1612	12409	17862

Valori mancanti: 375

Statistica	Gradi di libertà	Valore	Prob.
Chi quadrato	4	4926.8859	<.0001

Le due variabili sono fra loro correlate, come si vede dalla tabella. È quindi presumibile che, salvo particolari risultati durante l'analisi delle corrispondenze, alla fine se ne selezionerà soltanto una.

Ecco i risultati dell'analisi delle corrispondenze.

Grafico 2.7. Confronto delle variabili sui servizi dell'abitazione nella rappresentazione grafica delle modalità rispetto ai primi due assi fattoriali trovati con l'analisi delle corrispondenze

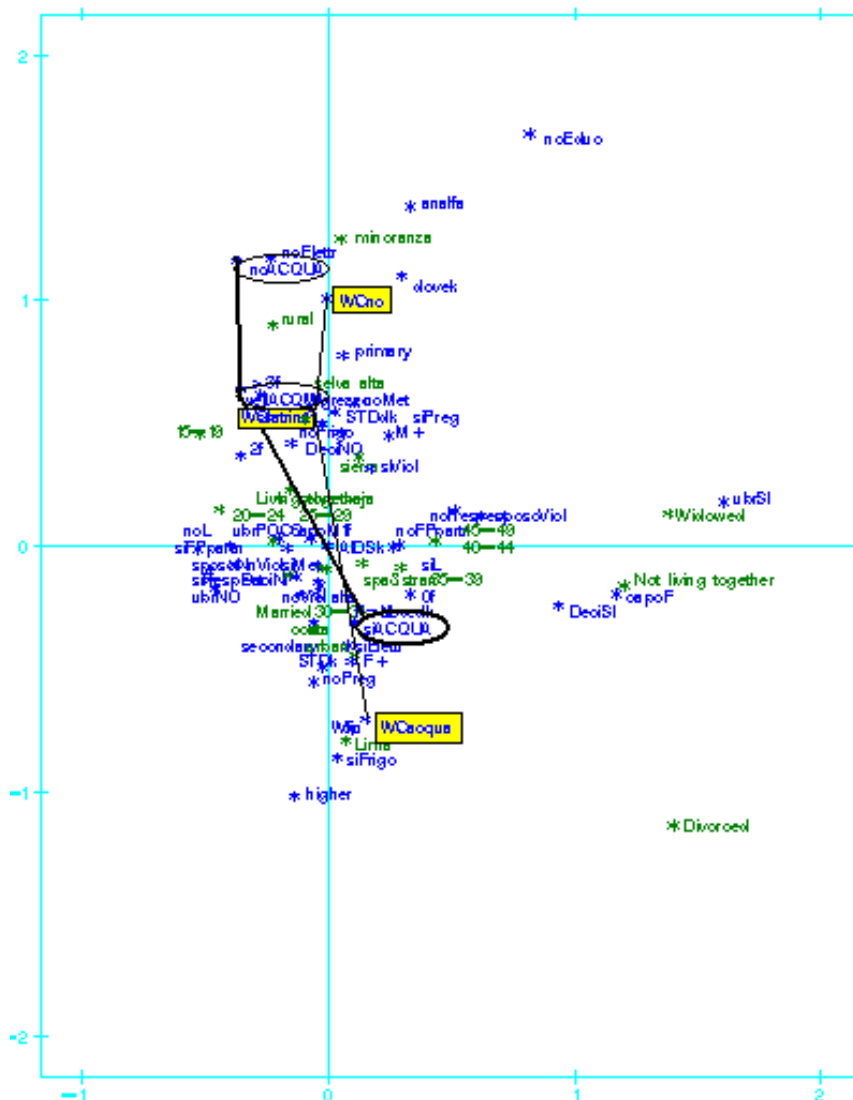


Tabella 2.8. Misure a confronto per le variabili che definiscono il possesso di beni

modalità	coord1	coord2	contr1	contr2	SqCos1	SqCos2	dim1	dim2	Best
WCacqua	-0,63	0,14	0,054	0,006	0,535	0,026	1	0	1
WClatrina	0,65	-0,24	0,026	0,008	0,147	0,020	1	0	1
WCno	1,14	-0,10	0,052	0,001	0,267	0,002	1	0	1
siACQUA	-0,30	0,11	0,016	0,005	0,314	0,040	0	0	1
wellACQUA	0,73	-0,29	0,009	0,003	0,043	0,007	0	0	1
noACQUA	1,21	-0,42	0,051	0,014	0,254	0,030	1	0	1

Dal grafico delle corrispondenze e dalla tabella risulta evidente la conferma della scelta iniziale: le due variabili seguono andamenti simili nel grafico, ma la “fonte d’acqua” risulta significativa, per la spiegazione dell’inerzia del primo fattore, solo con la modalità “no acqua”, e comunque sempre in modo minore della modalità corrispondente di “bagno”.

In conclusione quindi si confermano le variabili selezionate all’inizio dello studio. Tutte e quattro sono spiegate molto bene dal primo fattore.

2.2.5 Scelta delle variabili per definire la presa di decisioni

Un problema simile si presenta per la scelta delle variabili che meglio identificano le donne che riescono a prendere decisioni da sole; inizialmente si era selezionata “chi ha l’ultima parola sull’acquisto di beni costosi” (1=donna 2=donna con altri 3=altri). Si è deciso di confrontarla con “ultima parola sulla salute”, dove la prima modalità ha maggior frequenza della corrispondente nell’altra variabile.

Esistono altre variabili utili per valutare la capacità della donna di imporsi nelle decisioni della vita quotidiana, che però sono state ritenute meno idonee al nostro scopo; in particolare vengono escluse a priori le variabili “chi ha l’ultima parola sulle visite ai parenti” e “chi ha l’ultima parola su cosa cucinare”, in quanto la terza modalità si riduce drasticamente, e la variabile “chi ha l’ultima parola sui piccoli acquisti” perché si comporta in modo simile a “salute”.

La proiezione sul grafico delle modalità delle due variabili selezionate è inaspettato.

Osserviamo separatamente il comportamento delle due variabili selezionate:

- “acquisti grandi” : le donne che prendono decisioni da sole influenzano il secondo fattore in modo significativo (12% della spiegazione dell’inerzia), e si collocano fra le disagiate; esse sono principalmente donne separate o divorziate; quelle invece che prendono decisioni con il supporto di altre persone si collocano dalla parte opposta dell’asse, e questo probabilmente è dovuto ad un sostegno morale da parte di marito o familiari; stranamente invece coloro che non prendono decisioni si collocano nella parte positiva del primo fattore, cioè “disagio materiale”: questo avviene probabilmente perché viene percepito molto di più il malessere di una condizione economica disagiata, più che il diverso ruolo nella presa delle decisioni;
- “salute”: l’andamento della variabile è simile all’altra, ma non sovrapponibile; infatti coloro che prendono decisioni sulla salute da sole influenzano molto meno il secondo fattore, mentre sono più accentuate le altre due modalità, sia per i primo che per il secondo fattore!

Complessivamente quindi la prima variabile spiega più inerzia, ma questo è dovuto solo al fatto che la modalità “si” è particolarmente importante; la seconda variabile invece è più utile per le altre due modalità.

In conclusione si terranno entrambe le variabili, anche per gli ambiti diversi che descrivono.

2.2.6 Conclusioni

In questi paragrafi abbiamo spiegato la motivazione dell’inclusione di alcune variabili al posto di altre, presenti nel questionario, appartenenti alla stessa macroarea. Alla fine le analisi hanno selezionato un ristretto gruppo che non si discosta di molto da quello scelto inizialmente.

L’elenco delle variabili, che verrà utilizzato per tutte le successive analisi, è presente nell’Allegato 1; per tutte le variabili vengono riportate le etichette con cui si è lavorato in SAS e in SPSS (e che verranno spesso citate per comodità durante lo studio in tabelle e grafici) e la descrizione delle modalità.

2.3 LE NON RISPOSTE E IL PROBLEMA DELLA *PRIVACY*

Si è deciso in precedenza di lavorare senza i valori mancanti: il fatto che essi occupino un intero fattore non è statisticamente accettabile. Non si può però trascurare questo evento, perché i *missing values*, come vedremo, evidenziano chiaramente un problema di disagio sociale.

La maggior parte dei valori mancanti delle variabili selezionate sono dovuti ad una variabile filtro “privacy”, che rileva l’eventuale presenza di altre persone durante l’intervista. Questo fatto ha condizionato le risposte alle domande della sezione sulla violenza domestica perché la presenza del partner, di un altro familiare o di un esterno al momento dell’intervista non dava loro la tranquillità di rispondere liberamente.

Per lo stesso motivo, in caso di risposta “mancanza di privacy” troviamo tutti valori mancanti alle domande su: alcolismo, violenza verbale, violenza fisica, opinione sulla violenza come metodo di istruzione dei figli. Queste unità statistiche non verranno considerate nelle analisi sulle singole aree, ma in questa sede ci chiediamo: quali variabili influiscono di più su una situazione di disagio tale da non garantire alle intervistate la libertà di risposta ad un questionario? Quali caratteristiche accomunano queste donne?

Per rispondere a questi interrogativi, è stata eseguita un’analisi di regressione logistica confrontando le 543 intervistate che hanno riscontrato problemi di privacy con il resto del campione. La variabile “privacy” sarà quindi posta come dipendente, e si utilizzeranno le altre 24 variabili come predittive.

2.3.1 Analisi di regressione logistica

Cosa determina quindi il disagio dell’intervistata nel rispondere alle domande sulla violenza familiare?

Durante l’analisi, verrà considerata come dipendente la variabile “privacy”, con modalità “mancanza di privacy” contro “resto della popolazione”.

Selezione preliminare delle variabili predittive

Prima di iniziare l’analisi vera e propria, è necessario svolgere un’analisi preliminare delle potenziali variabili predittive, con lo scopo di scremare il numero iniziale decisamente elevato, e quindi scomodo da interpretare,

eliminando quelle variabili che rapportate singolarmente con la variabile dipendente risultano già statisticamente non significative.

A tale scopo sono state suddivise le variabili predittive in tre sottogruppi: variabili nominali e ordinali, variabili dicotomiche e variabili quantitative; se ne valuterà la significatività attraverso, rispettivamente, indici di associazione (nello specifico è stato utilizzato il Chi-quadrato), rischio relativo e test del rapporto di verosimiglianza su modelli logistici ad una variabile.

i. Variabili nominali/ordinali

Nello specifico, le percentuali di persone per ogni modalità che sono soggette ad un disagio di mancanza di privacy sono riassunte nella tabella successiva.

Tabella 2.10. Percentuali di “no privacy” per le modalità delle variabili nominali

Variabile	Modalità	%no privacy	Variabile	Modalità	%no privacy
Recere geografica	Lima metropoli	2.36	Stato coniugale	Sposata	2.62
	Resto costa	1.61		Convivente	3.73
	Sierra	2.48		Vedova	0.90
	Selva alta	1.85		Divorziata	0.00
	Selva bassa	7.26		Separata	2.27
Istruz. MF	F+	2.27	Servizi igienici	No wc	2.67
	F=M	2.88		Latrina	4.13
	M+	3.31		Wc con acqua	2.25
Decisione acquisti	No	3.45	Decisione salute	No	3.21
	Insieme	2.95		Insieme	3.42
	Da sola	2.34		Da sola	2.64
Età al primo matrimonio	x<15	4.14	Età donna	15-19	3.21
	15<x<17	3.70		20-24	2.36
	18<x<20	2.73		25-29	2.86
	21<x<30	2.34		30-34	2.84
	x>30	2.56		35-39	3.38
				40-44	3.32
Titolo di studio	No istruzione	4.87		45-49	2.93
	Primaria	3.44			
	Secondaria	2.38			
	Superiore	2.03			

Il primo passo dell’analisi è stato quello di studiare la dipendenza delle variabili ordinali e di quelle nominali con la variabile “privacy”, ponendole singolarmente a confronto in una tabella a doppia entrata. L’indice di associazione utilizzato è quello del Chi-quadrato.

Nella tabella che segue vengono riportate le variabili considerate e il valore della statistica χ^2 .

Tabella 2.11. Statistica del χ^2 per variabili nominali/ordinali

VAR	DF	Valore χ^2	Prob.	Bontà
Sregion	4	235.27	<.0001	***
educazioneMF	2	5.54	0.063	*
decisione	2	9.53	0.008	**
decisionA	2	7.76	0.022	**
V106 (istruzione)	3	42.41	<.0001	***
wc	2	42.25	<.0001	***
V501 (stato matr.)	4	29.21	<.0001	***
Eta matr	4	25.11	<.0001	***
V013 (età)	6	6.76	0.343	-

Nota: “-” non significativo; “*” significativo al 90%;
 “**” significativo al 95%; “***” significativo al 99.9%

Si osserva che:

- L'unica variabile nominale non correlata con la privacy è l'età della donna, quindi non sarà considerata in questa parte di studio;
- Diversamente da quanto emerso nelle analisi precedenti, dove solo alcune modalità alternativamente risultavano importanti per spiegare i fattori, l'età al primo matrimonio sembra essere importante per capire il fenomeno;
- il titolo di studio massimo è inversamente proporzionale al crescere dei problemi di privacy; si comporta in modo simile la variabile sulla differenza di istruzione fra uomo e donna;
- le due variabili sulla capacità di prendere decisioni hanno ancora una volta andamenti un po' diversi: fra le tre modalità, si differenziano dalle altre le donne che prendono decisioni da sole sulla salute, mentre per i grandi acquisti quelle che non possono prenderle neanche assieme a qualcun altro;
- sorprendentemente alta è la percentuale delle donne proveniente dalla giungla bassa: tale risultato è inaspettato perché fino ad ora sembrava che le donne provenienti da altopiano, giungla alta e bassa si comportassero in modo abbastanza simile;
- infine, anche la variabile wc ha un andamento percentuale non lineare come ci si aspetterebbe.

ii. Variabili dicotomiche

La seconda categoria include le variabili predittive dicotomiche (1= presenza dell'attributo, 0= assenza di esso). In tabella 2.12 riportiamo le percentuali di "no privacy" per ogni modalità.

Tabella 2.12. Percentuali di "no privacy" per le modalità delle variabili dicotomiche

Modalità	% no privacy	Modalità	% no privacy
capofamiglia Maschio	3.10	dove conosce	2.64
capofamiglia Femmina	2.23	dove non conosce	4.20
urbano	2.62	FP col partner	2.52
rurale	3.45	no FP col partner	3.21
etnia spagnola	2.73	muro finito	2.00
etnia minore	4.30	muro grezzo	3.37
alfabeta	2.59	si elettricità	2.67
analfabeta	4.38	no elettricità	3.53
lavora	2.81	si frigorifero	2.16
no lavora	3.14	no frigorifero	3.28
no pregiudizio	2.19	conosce AIDS	2.64
si pregiudizio	2.99	non conosce AIDS	4.44
metodi contraccettivi moderni	2.71	conosce STD	2.36
metodi contraccettivi tradizionali	3.68	non conosce STD	3.34

Per studiare la dipendenza nella mancanza di privacy rispetto a queste variabili, si è calcolata una stima del rischio relativo: i valori ottenuti sono riportati nella tabella riassuntiva 2.13, insieme ad un intervallo di confidenza al 95% che serve a determinarne la significatività (se il valore 1, cioè l'assenza di dipendenza, è incluso nell'intervallo, il valore non è statisticamente significativo).

Tabella 2.13. Rischio relativo per variabili dicotomiche e intervallo di confidenza al livello 95%

VARIABILE	Modalità (1) Presenza	Valore RR	Limite di confidenza al 95% inferiore	Limite di confidenza al 95% superiore
Pregiudizio	si	0.73*	0.59	0.89
Ethnicity	spagnola	0.62*	0.51	0.76
Urbanizzazione	rurale	1.33*	1.12	1.58
Alfabetizzazione	alfabetizzata	0.58*	0.48	0.70
Lavorando	si	0.89	0.75	1.06
Conosce AIDS	si	0.58*	0.48	0.71
Conosce STD	si	0.70*	0.58	0.84
Non Conosce Dove	si	1.61*	1.34	1.95
Capofamiglia	donna	0.71*	0.54	0.94
FP con il partner	si	0.78*	0.65	0.94
Uso metodi mod. contracc.	metodi moderni	0.73*	0.61	0.87

Elettricità	si	0.75*	0.63	0.89
Frigo	si	0.65*	0.53	0.80
Muro	moderno	0.58*	0.48	0.71

“**” = la stima è significativa al 95%

Si osserva che:

- solo la variabile “lavorando” non determina in modo significativo la mancanza di privacy dell’intervistata;
- la quasi totalità delle modalità 1, cioè presenza della caratteristica (seconda colonna della tabella), ad eccezione di “non conosce dove”, sono “protettivi” rispetto al disagio della mancanza di privacy.

iii. Variabili quantitative

Nell’ultimo sottoinsieme di variabili, sono stati costruiti tanti modelli regressivi logistici con intercetta e un predittore quante sono le variabili d’interesse: per valutare se e quanto un predittore determina la mancanza di privacy si prende in considerazione la significatività dei coefficienti angolari. Nella tabella successiva vengono quindi elencati i parametri dell’intercetta, della variabile indipendente, e le loro significatività.

All’unica variabile quantitativa presente fra quelle selezionate (figli) sono state aggiunte l’età e l’età alla prima unione, non suddivise in classi.

Tabella 2.14. Modelli logistici con significatività per l’intercetta e la variabile indipendente

privacy=...	intercetta	Pr> χ^2	Coefficiente angolare	Pr> χ^2
Figli 5	-3.67	< .0001	0.19	< .0001*
Età	-3.73	< .0001	0.01	0.15
Età matrimonio	-2.68	< .0001	-0.04	0.0001*

Si osserva che:

- l’età della donna, anche non divisa in classi, è non significativa. Non verrà quindi considerata nell’analisi successiva;
- al crescere dell’età al primo matrimonio, diminuisce la probabilità che una donna si trovi in una situazione di disagio nel rispondere al questionario;
- il numero di figli è direttamente proporzionale alla mancanza di privacy.

Informazioni sui dati da analizzare

Dopo una prima scrematura delle variabili esplicative di interesse, si inizia l'analisi per trovare un modello in grado di spiegare la variabile privacy. Verrà utilizzato il metodo "stepwise", che consente di seguire passo per passo la selezione delle variabili da inserire nel modello.

Una volta specificato al programma qual è la probabilità di successo che il modello deve considerare (mancanza di privacy), SAS elimina automaticamente le osservazioni dove sono presenti valori mancanti.

Successivamente sono state riparametrizzate le variabili nominali ed ordinali: per quest'ultime si è usata come confronto la modalità più alta (nel livello di istruzione, indipendenza nelle decisioni, e tipologia di servizi igienici); per quelle nominali (regione geografica e stato coniugale) invece, non essendoci informazioni a priori sufficienti per effettuare una scelta mirata, sono state selezionate quelle che fino ad ora si sono comportate in modo meno anomalo (rispettivamente, quindi, *selva* bassa e non conviventi).

SREGION	1	1	0	0	0
	2	0	1	0	0
	3	0	0	1	0
	4	0	0	0	1
	5	-1	-1	-1	-1
decisione	0	1	0		
	1	0	1		
	2	-1	-1		
decisioneA	0	1	0		
	1	0	1		
	2	-1	-1		
V106	0	1	0	0	
	1	0	1	0	
	2	0	0	1	
	3	-1	-1	-1	
wc	0	1	0		
	1	0	1		
	2	-1	-1		
V501	1	1	0	0	0
	2	0	1	0	0
	3	0	0	1	0
	4	0	0	0	1
	5	-1	-1	-1	-1

Vengono riportati i quattro modelli più significativi, le cui analisi sono così caratterizzate:

- i. immissione di tutte le variabili esplicative selezionate dall'analisi preliminare;

- ii. esclusione della variabile “pregiudizio”, perché portatrice di troppi valori mancanti;
- iii. esclusione della variabile che indica lo stato coniugale della donna, perché l’importanza della modalità “divorziate” influenza troppo l’immissione della variabile nel modello;
- iv. esclusione forzata di “sregion”, che sarà sempre la prima variabile ad entrare nel modello, per studiare il comportamento delle altre variabili.

Prima di iniziare a descrivere i risultati, è importante sottolineare che l’obiettivo di questo sotto-studio non sarà quello di selezionare il modello migliore in assoluto che definisce la variabile “privacy”, ma più semplicemente quello di capire quali variabili, e quindi quali disagi o caratteristiche contestuali, ruotano attorno alle donne che hanno avuto difficoltà nel rispondere alle domande sulla violenza domestica.

Un’ultima nota: tutti i modelli che verranno mostrati sono stati provati sia con la variabile “età al matrimonio” quantitativa, sia con quella suddivisa in 5 classi di frequenza, e quindi ordinale. Poiché i risultati non cambiano in maniera significativa (i predittori selezionati sono i medesimi, e i coefficienti molto simili), vengono presentati solo i modelli con la variabile quantitativa.

Selezione stepwise per il primo modello

Il primo modello presentato è quello che verifica gli effetti di tutte le variabili, selezionate nell’analisi preliminare, hanno sulla variabile risposta.

I passi della selezione stepwise sono riassunti in tre punti:

- i. selezione della prima variabile da inserire nel modello (precedentemente era stato calcolato il modello con la sola intercetta), cioè quella che meglio spiega la variabile dipendente;
- ii. verifica della significatività del nuovo modello complessivo rispetto al modello con la sola intercetta, tramite il test Score, il Wald test e il test del Rapporto di Verosimiglianza; quest’ultimo, nello specifico, è dato dalla differenza fra la log-verosimiglianza del modello con la sola intercetta e quella del modello selezionata nel passaggio in esame;

- iii. infine controllo del Chi-quadrato residuo, che è dato dalla differenza fra la log-verosimiglianza del modello attuale e la log-verosimiglianza del modello con tutte le variabili esplicative: se tale test è significativo allora si procede con la immissione di altre variabili esplicative, altrimenti ci si ferma perché ulteriori variabili non aggiungerebbero alcun contributo significativo al modello.

Nel nostro caso, dopo aver immesso l'**intercetta** che risulta significativa, viene selezionata per prima la variabile **sregion**.

I test dell'ipotesi globale nulla riportati di seguito ci dicono che il modello attuale complessivo è migliore di quello con la sola intercetta, e il test del Chi-quadrato residuo ci permette di proseguire l'analisi:

Test dell'ipotesi globale nulla: BETA=0			
Test	Chi-quadrato	DF	Pr > ChiQuadr
Rapp. verosim.	128.2235	4	<.0001
Score	161.2816	4	<.0001
Wald	141.8160	4	<.0001

Test del Chi-quadrato residuo			
Chi-quadrato	DF	Pr > ChiQuadr	
85.6889	27	<.0001	

Al passo successivo SAS fa entrare la variabile esplicativa **ethnicity** che, come confermato dai test sotto riportati, migliora il modello e ci permette di proseguire nella selezione.

Test dell'ipotesi globale nulla: BETA=0			
Test	Chi-quadrato	DF	Pr > ChiQuadr

Rapp. verosim.	144.3943	5	<.0001
Score	178.7490	5	<.0001
Wald	157.8831	5	<.0001

Test del Chi-quadrato residuo

Chi-quadrato	DF	Pr > ChiQuadr
66.0163	26	<.0001

Omettendo ora i singoli passaggi successivi, SAS seleziona successivamente le seguenti variabili: **v785 (STD), decisioneA, FPpartner, wc, muro, figli5.**

Nell'ultimo passo invece entra nel modello la variabile **v501 (stato coniugale)**, ma il test del Chi-quadrato residuo non è significativo, in quanto non aggiunge alcun contributo significativo al modello, e quindi si rimuove la variabile. La costruzione del modello in questo modo termina perché l'ultima variabile immessa viene rimossa dal criterio della statistica di Wald.

Riassumendo, ecco il riepilogo della selezione per passi:

Passo	Imnesso	Rimosso	Chi-quadrato	Wald	Descrizione
1	sregion		161,28		regione geografica
2	ethnicity		19,26		gruppo etnico
3	v785		9,46		conoscenza STD
4	decisioneA		8,19		decisione salute
5	FPpartner		5,60		programmazione familiare
6	wc		7,73		servizi igienici
7	muro		4,94		muro
8	figli5		4,37		figli negli ultimi 5 anni
9	v501		9,30		stato civile
10		v501		7,55	stato civile

Stima dei parametri e dei rapporti crociati per il primo modello

Nella tabella successiva vengono riportati, per ogni variabile, la stima del parametro, il suo errore standard, il valore del test di Wald ($2 \times \text{stima} / \text{errore standard}$), e la significatività di tale test.

Tabella 2.15. Analisi delle stime di massima verosimiglianza

Parametro	DF	Stima	Errore standard	Chi-quadrato	Pr > ChiQuadr
Intercept	1	-2.3547	0.3793	38.5306	<.0001
SREGION	1	-0.1291	0.1553	0.6912	0.4058
SREGION	2	-0.3401	0.1185	8.2440	0.0041
SREGION	3	-0.4393	0.1202	13.3668	0.0003

SREGION	4	1	-0.3308	0.1905	3.0150	0.0825
decisioneA	0	1	-0.3018	0.1100	7.5247	0.0061
decisioneA	1	1	0.2470	0.0874	7.9906	0.0047
wc	0	1	-0.2815	0.1040	7.3249	0.0068
wc	1	1	0.1704	0.0807	4.4576	0.0347
ethnicity		1	-0.7002	0.1916	13.3490	0.0003
V785		1	-0.3105	0.1155	7.2242	0.0072
FPpartner		1	-0.3117	0.1203	6.7066	0.0096
muro		1	-0.2978	0.1417	4.4141	0.0356
figli5		1	0.1307	0.0625	4.3658	0.0367

Si nota subito che non tutte le stime sono significative al 95%, in particolare la modalità Lima.

Successivamente si possono sottolineare altre particolarità: le variabili dicotomiche sono tutti fattori protettivi, mentre si comportano in maniera anomala quelle nominali.

Questi aspetti vengono accentuati anche dalla stima dell'odds ratio (costruito con l'esponenziale della stima), riportato di seguito:

Tabella 2.16. Stime degli *odds ratio*

Effetto		Stima puntuale	95% Limiti di confidenza di Wald	
SREGION	1 vs 5	0.427	0.286	0.635
SREGION	2 vs 5	0.267	0.196	0.363
SREGION	3 vs 5	0.242	0.179	0.326
SREGION	4 vs 5	0.269	0.167	0.433
decisioneA	0 vs 2	0.700	0.501	0.979
decisioneA	1 vs 2	1.212	0.946	1.552
wc	0 vs 2	0.675	0.474	0.962
wc	1 vs 2	1.061	0.806	1.396
ethnicity		0.497	0.341	0.723
V785		0.733	0.585	0.919
FPpartner		0.732	0.578	0.927
muro		0.742	0.562	0.980
figli5		1.140	1.008	1.288

Risulta quindi:

- che all'aumentare del numero di figli, la probabilità di avere problemi di privacy aumenta;
- sono fattori protettivi il conoscere le STD, l'essere di etnia spagnola, il discutere col partner la Pianificazione Familiare e l'avere i muri dell'abitazione in materiale non grezzo;
- fra le regioni geografiche, determina più facilmente il non riuscire a rispondere alle domande sulla violenza familiare la Selva Bassa, seguita da Lima metropoli;

Bontà della classificazione del primo modello

In conclusione, riportiamo la percentuale di coppie di osservazioni concordanti (cioè quando, prese due osservazioni con valori di Y differenti, il valore stimato di quella per cui Y=0 è inferiore al valore stimato dell'altra), che è pari a 66.8%.

Associazione di probabilità previste e risposte osservate

Percentuale concordante	66.8	Di di Somers	0.379
Percentuale discordante	28.9	Gamma	0.396
Percentuale legato	4.3	Tau-a	0.019
Coppie	5124471	c	0.689

L'ultima tabella di questo paragrafo invece fornisce, a diversi livelli "soglia" di probabilità, la classificazione dei soggetti: il numero di soggetti classificati correttamente e non a seconda del livello della variabile risposta, la percentuale complessiva di correttamente classificati, la sensibilità (la percentuale di casi per cui Y=1 correttamente classificati dal modello), la specificità, i falsi positivi (la percentuale di casi per cui Y=0 erroneamente classificati) e i falsi negativi.

Tabella 2.17. Classificazione dei soggetti

Livello Prob.	Corretta		Non corretta		Corretto	Percentuali		POS falso	NEG falso
	Evento	Non- evento	Evento	Non- evento		Sensi- bilità	Speci- ficità		
0.000	363	0	14117	0	2.5	100.0	0.0	97.5	.
0.020	235	8432	5685	128	59.9	64.7	59.7	96.0	1.5
0.040	152	11846	2271	211	82.9	41.9	83.9	93.7	1.8
0.060	81	13146	971	282	91.3	22.3	93.1	92.3	2.1
0.080	39	13718	399	324	95.0	10.7	97.2	91.1	2.3
0.100	17	13968	149	346	96.6	4.7	98.9	89.8	2.4
0.120	6	14065	52	357	97.2	1.7	99.6	89.7	2.5
0.140	2	14101	16	361	97.4	0.6	99.9	88.9	2.5
0.160	1	14105	12	362	97.4	0.3	99.9	92.3	2.5
0.180	1	14109	8	362	97.4	0.3	99.9	88.9	2.5
0.200	1	14110	7	362	97.5	0.3	100.0	87.5	2.5
0.220	0	14115	2	363	97.5	0.0	100.0	100.0	2.5
0.240	0	14115	2	363	97.5	0.0	100.0	100.0	2.5
0.260	0	14117	0	363	97.5	0.0	100.0	.	2.5

Gli altri modelli e confronti

Non verranno riportati ora tutti i passaggi che hanno caratterizzato la scelta delle variabili negli altri modelli ipotizzati.

Nella tabella seguente sono messi a confronti i predittori della variabile "privacy", ricavati nei quattro studi.

Tabella 2.18. Predittori selezionati nei 4 modelli studiati

I: tutti i predittori		II: il primo senza "pregiudizio"		III: il secondo senza "stato coniugale"		IV: il terzo senza "sRegion"
sRegion		sRegion		sRegion		Wc
Ethnicity		Istruzione		Istruzione		Dove
STD		Ethnicity		Ethnicity		Murø

DecisioneA		STD		STD		Alfabeta
FP partner		Wc		Wc		Capo famiglia
Wc		Stato coniugale		Dove		FP partner
Muro		Dove		Muro		Figli 5
Figli 5		Muro		Elettricità		
s. coniugale		Elettricità		Capo famiglia		
s. coniugale	<i>rimosso</i>	Uso contraccettivi		AIDS		
		Uso contraccettivi	<i>rimosso</i>	AIDS	<i>rimosso</i>	

La prima particolarità che salta agli occhi è che le uniche variabili sempre presenti, tranne ovviamente “sregion” che è stata esclusa dall’ultimo modello forzatamente, sono “muro” e “wc”; questo porta a pensare che l’importanza dell’abitazione e del contesto familiare dove si vive siano molto più di meri indici di povertà fisica, e che invece vadano a toccare sfere di disagio psicologico molto profonde.

Altre caratteristiche evidenziate dal confronto dei quattro modelli, anche rispetto alla tabella 2.19 sui rapporti crociati, sono:

- la presa di decisione sui temi inerenti alla salute è presente solo nel primo modello (quello con i valori mancanti causati dalla variabile “pregiudizio”)
- nell’ultimo modello vengono stravolti gli schemi presenti nei precedenti, e in particolare:
 - la variabile istruzione viene sostituita dall’alfabetizzazione
 - ritornano alcune variabili scomparse nei due modelli precedenti, quali il numero di figli e la discussione della Pianificazione Familiare col partner;
- quando viene esclusa la variabile “stato coniugale”, compare immediatamente il “sesso del capofamiglia”; questo a sottolineare l’ovvia correlazione fra le due variabili;
- la conoscenza delle STD è molto più importante della conoscenza dell’AIDS. Probabilmente questo è dovuto dal fatto che il nostro modello ha il limite di predire meglio la modalità Y=0 (visto lo sbilanciamento dei dati verso questa modalità), e si era visto in precedenza che sapere chi non conosce l’AIDS è molto più importante per capire chi soffre di un certo disagio, piuttosto che per prevedere chi sta bene;
- non conoscere dover comprare i preservativi fa aumentare la probabilità di non rispondere per mancanza di “privacy”;

- le variabili che non vengono mai usate sono (oltre a quelle già escluse nell'analisi preliminare, cioè "età al matrimonio" e "lavorando"): differenza di istruzione fra Uomo e Donna, presa di decisione sui grandi acquisti, frigo, uso di metodi contraccettivi moderni, urbanizzazione e conoscenza dell'AIDS;
- il terzo modello aveva incluso la variabile "stato coniugale" ma, dalla significatività delle stime dei rapporti crociati, si vede che solo la modalità "divorziate" è significativa;
- la variabile "wc" presenta la particolarità di avere la modalità 0, che ci aspetteremmo portatrice di disagio, protettiva rispetto alla variabile dipendente; è però anche vero che essa non risulta significativa nell'ultimo modello.

Tabella 2.19. Confronti fra le stime degli *odds ratio* dei 4 modelli. Per le variabili dicotomiche, la stima è riferita alla modalità "1"

Variabile	Rapporto	Stima I	Sign	Stima II	Sign	Stima III	Sign	StimaIV	Sign
Sregion	1 vs 5	0.43	*	0.43	*	0.42	*	esclusa	
Sregion	2 vs 5	0.27	*	0.25	*	0.24	*	esclusa	
Sregion	3 vs 5	0.24	*	0.27	*	0.26	*	esclusa	
Sregion	4 vs 5	0.27	*	0.25	*	0.25	*	esclusa	
DecisioneA	0 vs 2	0.70	*	-		-		-	
DecisioneA	1 vs 2	1.21	No	-		-		-	
Wc	0 vs 2	0.67	*	0.72	*	0.74	No	0.77	No
Wc	1 vs 2	1.06	No	1.08	No	1.09	No	1.36	*
Ethnicity		0.50	*	0.63	*	0.65	*	-	
STD		0.73	*	0.73	*	0.73	*	-	
FPpartner		0.73	*	-		-		0.81	*
Muro		0.74	*	0.74	*	0.73	*	0.69	*
Figli5		1.14	*	-		-		1.10	*
Istruzione	0 vs 3	-		1.70	*	1.67	*	-	
Istruzione	1 vs 3	-		1.15	No	1.16	No	-	
Istruzione	2 vs 3	-		0.96	No	0.98	No	-	
Stato coniugale	1 vs 5	-		1.34	No	esclusa		esclusa	

Stato coniugale	2 vs 5	-		1.53	*	<i>esclusa</i>		<i>esclusa</i>	
Stato coniugale	3 vs 5	-		0.41	No	<i>esclusa</i>		<i>esclusa</i>	
Stato coniugale	4 vs 5	-		<0.001	No	<i>esclusa</i>		<i>esclusa</i>	
Dove		-		1.33	*	1.32	*	1.34	*
elettricità		-		1.28	*	1.26	*	-	
Capo-famiglia		-		-		0.77	No	0.72	*
Alfabetizzazione		-		-		-		0.87	*

Nota: nella colonna della significatività, il simbolo “*” indica che la stima è significativa, mentre il “No” indica che non è significativa.

3. CONFRONTI FRA REGIONI

Nel capitolo precedente si selezionano 33 variabili sulle 4223 disponibili nel dataset, attraverso uno studio preliminare e l'analisi delle corrispondenze. Da quest'ultima si estraggono due fattori che sembrano descrivere bene la variabilità dei dati: il primo comprende quelle variabili che descrivono la situazione economico-materiale delle intervistate; il secondo invece raccoglie quegli aspetti più introspettivi e psicologici, delineando un quadro più sociale del disagio percepito dalle donne campionate.

In questo capitolo ci si focalizzerà sul vero obiettivo dello studio, cioè quello di entrare in profondità nelle differenze locali per coglierne le differenze e le similitudini. Si ipotizza che poche variabili rimarranno importanti (ai fini dell'identificazione delle donne economicamente e socialmente svantaggiate) indistintamente su tutto il territorio nazionale; mentre, presumibilmente, ogni area geografica avrà delle peculiarità tali da non poter applicare ai suoi abitanti alcun modello generale di previsione o descrizione.

Lo studio di tutti i singoli Dipartimenti sarebbe lungo e ripetitivo: per questa analisi, sono stati quindi selezionati 6 di essi, sui quali si studieranno le relazioni presenti fra le 33 variabili estratte nello studio a livello nazionale.

I 6 Dipartimenti scelti cercano di rappresentare le 5 aree geografiche presenti nel Perù, riassunte nella variabile "sregion": l'area Metropolitana (nella fattispecie identificabile nella capitale Lima), la Costa, l'Altipiano (la "sierra"), la Giungla Alta (cioè quei territori pre-amazzonici) e la Giungla Bassa (la "selva"). Per ogni area è stata selezionata quindi una Regione, ad eccezione della Costa, dove si è preferito distinguere fra costa nord e costa sud, per via delle differenze climatiche, culturali ed economiche presenti fra esse.

La zona morfologica è quindi la variabile principale di stratificazione: i Dipartimenti selezionati sono quelli con la totalità delle intervistate (o con la più alta percentuale di esse) appartenenti alla regione geografica d'interesse; in questo modo è possibile mantenere nelle analisi la stessa numerosità

campionaria iniziale. Ad esempio, gli unici due dipartimenti il cui territorio è totalmente pianeggiante (e quindi il 100% del campione appartiene alla modalità Costa), sono Ica e Tumbes, selezionati a rappresentanza, rispettivamente, della Costa Sud e di quella Nord.

Le percentuali di campione per zona morfologica sono già state riportate in tabella 1.2. Di seguito vengono presentate tali percentuali ordinate e riassunte per zona morfologica di predominanza (il primo Dipartimento elencato all'interno di ogni zona, è quello selezionato per l'analisi)

Tabella 3.1. Dipartimenti del Perù suddivisi per zona morfologica dominante, e percentuale di donne campionate che risiedono in essa

Zona Metropolitana	%	Sierra	%
Lima	84,54	Apurimac	100
		Huancavelica	100
Costa	%	Puno	100
Ica	100	Cusco	91,59
Tumbes	100	Ayacucho	91,47
Lambayeque	98,76	Arequipa	85,63
Tacna	92,33	Cajamarca	79,28
Piura	91,43	Junin	77,96
Moquegua	74,56	Pasco	76,27
La Libertad	70,61	Ancash	69,21
		Huanuco	67,90
Selva Baja	%		
Loreto	100	Selva Alta	%
Madre de Dios	100	Amazonas	60,12
Ucayali	95,79		
San Martin	68,03		

Si nota che la scelta del dipartimento è stata obbligata per la zona metropolitana e quella della *selva alta*; invece Apurimac e Loreto sono state selezionate, fra quelle con la totalità della popolazione appartenente alla zona morfologica principale, in base alla loro posizione strategica rispetto al territorio nazionale.

Nella Premessa e nel primo Capitolo si sono evidenziate le enormi differenze esistenti fra la Costa, la Sierra e la Selva. Si è evidenziato come i Dipartimenti situati in altura presentino gli ISU più bassi, e che il 67,7% della popolazione che vi ci risiede non riesca a superare la soglia di povertà.

Queste differenze sono ovviamente presenti anche nei 6 Dipartimenti campionati per lo studio. Nella tabella seguente, ad esempio, vengono mostrati gli ISU e gli indicatori che lo compongono.

Tabella 3.2. Indice di Sviluppo Umano (2003) per i 6 Dipartimenti selezionati, e la sua scomposizione

	ISU 2003 <i>[0-1]</i>	Speranza di vita alla nascita <i>(anni)</i>	Alfabetismo <i>(%)</i>	Immatricolazione secondaria <i>(%)</i>	Raggiungimento educativo <i>(%)</i>	Reddito familiare pro capite <i>(dollari)</i>
Lima (metropoli)	0,720	73,5	96,5	95,8	96,1	652,2
Ica (costa sud)	0,662	72,0	95,7	97,3	96,8	412,9
Tumbes (costa nord)	0,609	70,3	94,8	84,9	88,2	346,9
Loreto (giungla bassa)	0,525	68,8	92,0	64,8	73,9	212,3
Amazonas (giungla alta)	0,502	69,3	84,0	58,1	66,7	205,6
Apurimac (altipiano)	0,488	65,9	70,1	69,0	69,4	185,1

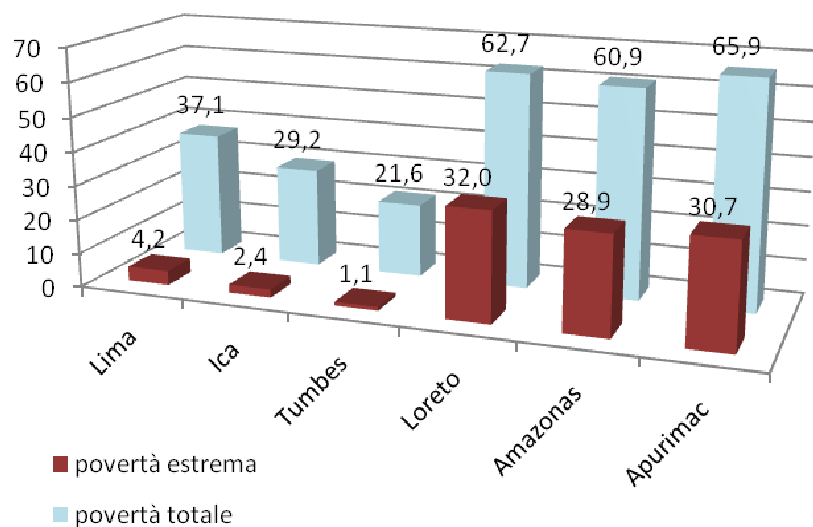
Fonte: INEI, MEF, MINEDU. **Elaborazione dati:** PNUD / Equipo para el Desarrollo Humano, Perú 2004

Questi indici, per quanto generici, sottolineano subito la presenza di due realtà molto distinte sotto tutti i punti di vista: quella della costa e delle grandi città, e quella dalla Cordigliera delle Ande fino alla Giungla. Gli indicatori dei tre dipartimenti sulla costa sono infatti sempre più alti rispetto a quelli relativi alle altre tre Regioni; ma anche all'interno di essi ci sono interessanti differenze:

- Il Dipartimento della costa sud (Ica) è molto più simile a quello di Lima piuttosto che a quello della costa nord;
- A est della cordigliera, la povertà sembra crescere al crescere dell'altitudine;
- Significativa la differenza fra alfabetismo e immatricolazione secondaria: sull'altipiano si studia di più rispetto alla zona amazzonica, ma la percentuale di analfabeti è decisamente più alta;
- Gli abitanti di Apurimac sono i più poveri e quelli con speranza di vita minore rispetto ai Dipartimenti selezionati.

Quanto detto si osserva anche rispetto alla percentuale di persone che vivono sotto la "linea di povertà", come si evince dal prossimo grafico.

Grafico 3.1. Incidenza della Povertà Totale e della Povertà Estrema sulla popolazione residente in 6 Dipartimenti (percentuale)



Questa breve introduzione vuole far riflettere sulle possibili differenze che emergeranno nei paragrafi successivi. Per ogni Dipartimento si procederà ora nel seguente modo:

- le relazioni fra le variabili (suddivise in variabili attive e variabili supplementari) saranno valutate tramite l'uso dell'analisi delle corrispondenze, che ci aiuterà ad individuare i fattori principali;
- sugli assi cartesiani, si riporteranno solo le modalità che apportano un contributo significativo alla spiegazione dell'inerzia;
- con le coordinate assunte dalle singole unità sugli assi (i punteggi fattoriali), si effettuerà l'analisi di segmentazione, inserendo tutte le variabili "attive", anche quelle non proiettate nell'analisi delle corrispondenze perché poco significative.

Non si faranno analisi dei missing values in questo capitolo. La scelta è dovuta al fatto che in 4 aree la proporzione di non risposte è piuttosto contenuta (tabella 3.3). Diversa è la situazione nella giungla bassa, ma soprattutto nella *sierra*, sull'altipiano andino. In questo capitolo ci concentriamo solo sul confronto fra aree geografiche, mentre nel prossimo capitolo, relativo al Dipartimento di Cusco (anch'esso situato nella *sierra*, e soggetto ad un'alta riduzione del campione), valuteremo la selezione operata nel campione a causa dell'alta percentuale di missing values.

Tabella 3.3. Percentuale di riduzione dei dati per le aree geografiche con l'eliminazione dei *missing values*

Lima	area metropolitana	6%
Tumbes	costa nord	4%
Ica	costa sud	8%
Apurimac	altipiano andino	46%
Amazonas	giungla alta	10%
Loreto	giungla bassa	23%

Per evitare di ripetere le stesse informazioni ad ogni analisi, informiamo subito che la variabile “conoscenza dell'aids” non risulterà significativa in nessuna delle regioni selezionate. Essa quindi non sarà più nominata in questo capitolo. Inoltre non verrà proiettata la variabile supplementare “gruppo etnico” nelle due regioni della costa e in Amazonas, perché la totalità del campione appartiene alla modalità “spagnola”. Per motivo analogo, non si considera la variabile “urbanizzazione” in Lima metropolitana.

3.1 COSTA NORD: TUMBES

L'analisi multipla delle corrispondenze evidenzia la presenza di due fattori principali: il primo spiega il 18,5% dell'inerzia totale, mentre il secondo il 12,28%. Complessivamente quindi arrivano a spiegare poco più del 28% dell'inerzia.

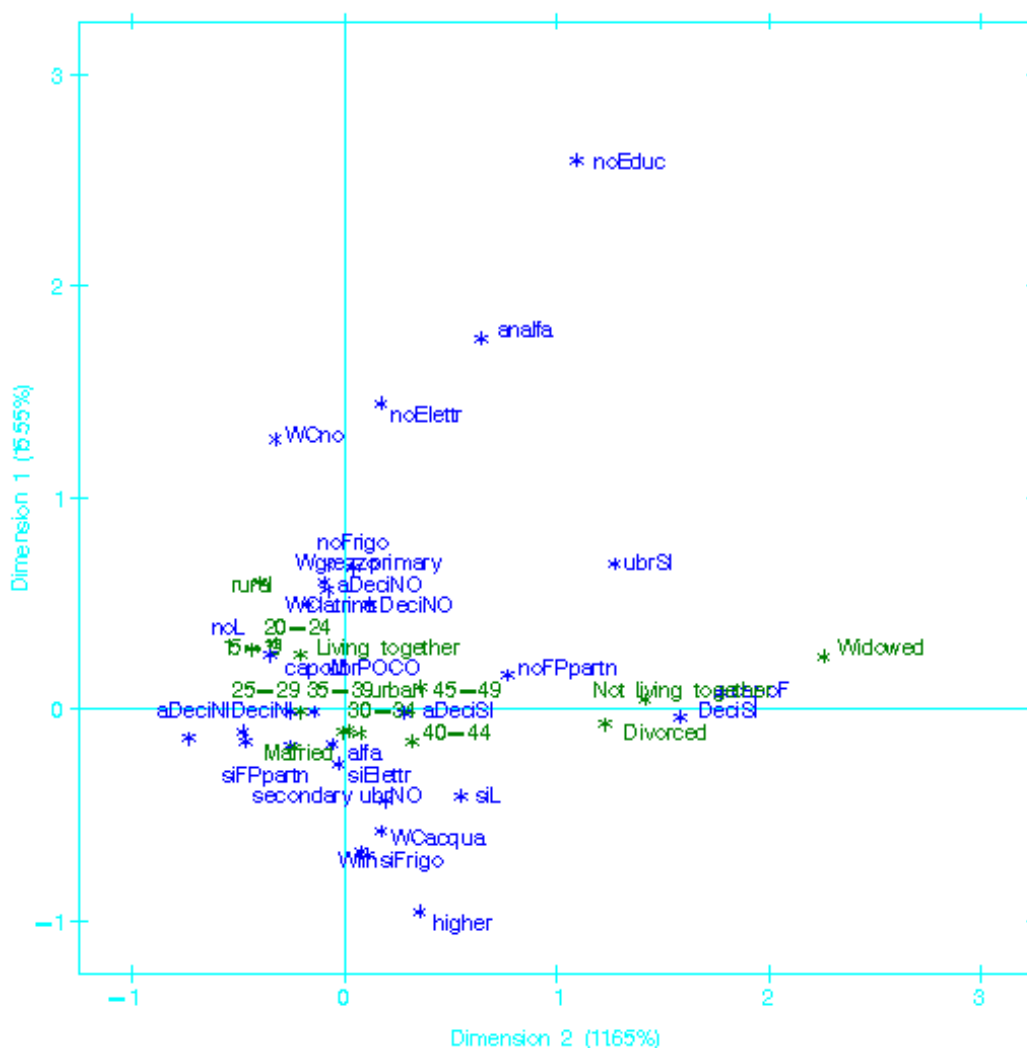
Con la scomposizione corretta di Greenacre tale percentuale cresce a circa il 36%, come si vede dalla tabella che segue.

Tabella 3.4. Scomposizione dell'inerzia con la correzione di Greenacre

Scomposizione dell'inerzia corretta di Greenacre				
Inerzia principale	Inerzia corretta	Percentuale	Percentuale cumulativa	4 8 12 16 20
0,25	0,20	20,85	20,85	*****
0,20	0,15	15,14	35,99	*****
0,14	0,09	9,50	45,50	*****
0,11	0,07	7,33	52,82	*****
0,10	0,06	6,39	59,21	*****
0,10	0,06	5,97	65,18	*****
0,10	0,05	5,69	70,87	*****
0,09	0,05	5,27	76,15	*****
0,08	0,05	4,70	80,84	*****
0,08	0,04	4,05	84,89	*****
0,08	0,04	3,91	88,80	*****
0,06	0,03	2,89	91,69	****
0,06	0,02	2,57	94,26	***
0,05	0,02	2,32	96,58	***
0,05	0,02	1,81	98,39	**
0,04	0,02	1,61	100,00	**
Totale	0,96	100,00		

Si osservi ora il grafico delle corrispondenze. A prima vista si conferma il fatto che l'asse verticale è dominato dalle modalità riguardanti l'istruzione delle donne e l'abitazione, mentre quello orizzontale dai comportamenti delle intervistate.

Grafico 3.2. Rappresentazione grafica delle modalità rispetto ai primi due assi fattoriali trovati con l'analisi delle corrispondenze per il Dipartimento di Tumbes



Nel semiasse positivo delle ordinate la variabile “titolo di studio” spicca per importanza rispetto alle altre variabili che descrivono l'istruzione; le prime quattro modalità che caratterizzano tale semiasse sono:

- nessun titolo di studio
- analfabetismo

- no elettricità
- assenza di un bagno interno

Dalla parte opposta invece vengono proiettate le seguenti modalità:

- livello di istruzione superiore
- si frigo
- muro con materiali finiti
- bagno con acqua corrente

Le prime modalità che si incontrano nel semiasse positivo delle ascisse sono:

- capo-famiglia donna
- capacità di prendere decisioni sui grandi acquisti
- alcolismo (che però influisce anche sul semiasse positivo delle ordinate, perché si colloca nel secondo quadrante)
- nessuna pianificazione familiare

Hanno invece le coordinate più piccole sull'asse orizzontale:

- presa di decisioni sulla salute con altre persone
- presa di decisioni sui grandi acquisti con altre persone
- capacità di pianificazione familiare

La proiezione sul grafico delle tre variabili supplementari (età, urbanizzazione, stato civile) indica che:

- le donne divorziate sono collocate vicino alle donne separate e alle vedove (fatto che non accadeva a livello nazionale), e si trovano agli estremi del semiasse positivo delle ascisse (secondo fattore);
- le modalità della fascia d'età non distano molto dall'origine degli assi, e quindi non parrebbero essere spiegate bene dai due fattori; volendo effettuare comunque una suddivisione, si vede che rispetto al primo fattore le donne più giovani e quelle più anziane (fino ai 29 anni e sopra i 45) sono nella parte superiore del grafico, quindi fra le più povere; invece, rispetto al secondo fattore, la soglia d'età fra le più o le meno disagiate è 35 anni (le donne più anziane sono nel semiasse positivo);

- le donne che vivono in zone rurali si collocano nel semiasse positivo del primo fattore, cioè fra le disagiate economicamente.

Sugli assi cartesiani non abbiamo riportato, per semplificarne la lettura, le modalità delle variabili con contribuiscono molto poco all'individuazione dei due fattori. Queste variabili sono: differenza di istruzione fra donna e uomo, violenza dello sposo sia fisica che verbale, opinione sulla violenza, metodi contraccettivi, conoscenza di dove comprare i preservativi, pregiudizio verso chi ha l'AIDS, età al matrimonio, conoscenza delle malattie sessualmente trasmissibili.

La lettura del grafico è supportata dalla tabella seguente, dove vengono evidenziate le prime modalità per importanza nella spiegazione dell'inerzia di ogni dimensione. Sono state selezionate quelle modalità a cui SAS fornisce un punteggio almeno uguale a 1 (in una scala da 0 a 2) per indicare l'entità del contributo all'inerzia della dimensione.

Tabella 3.5. Migliori modalità per contributo parziale all'inerzia delle due dimensioni

Dimensione 1		Dimensione 2	
noElettr	0,112	DeciSI	0,204
analfa	0,093	capoF	0,188
Wgrezzo	0,083	noFPpartn	0,108
WCno	0,082	siFPpartn	0,068
Wfin	0,081	DeciNI	0,065
siFrigo	0,079	aDeciNI	0,064
noFrigo	0,069	ubrSI	0,060
higher	0,069	siL	0,056
WCacqua	0,066		
noEduc	0,061		
primary	0,051		

L'asse verticale è dominato chiaramente dal livello di studio, dal possesso di beni materiali e dalla condizioni dell'abitazione, mentre quello orizzontale è caratterizzato dalla capacità di prendere decisioni, dal lavoro e dall'alcolismo.

Le stesse modalità sono anche fra quelle con i coseni quadrati più elevati, con l'aggiunta di alcune modalità sul materiale di costruzione della casa per quanto riguarda il primo fattore, e sulla presa di decisioni per il secondo fattore.

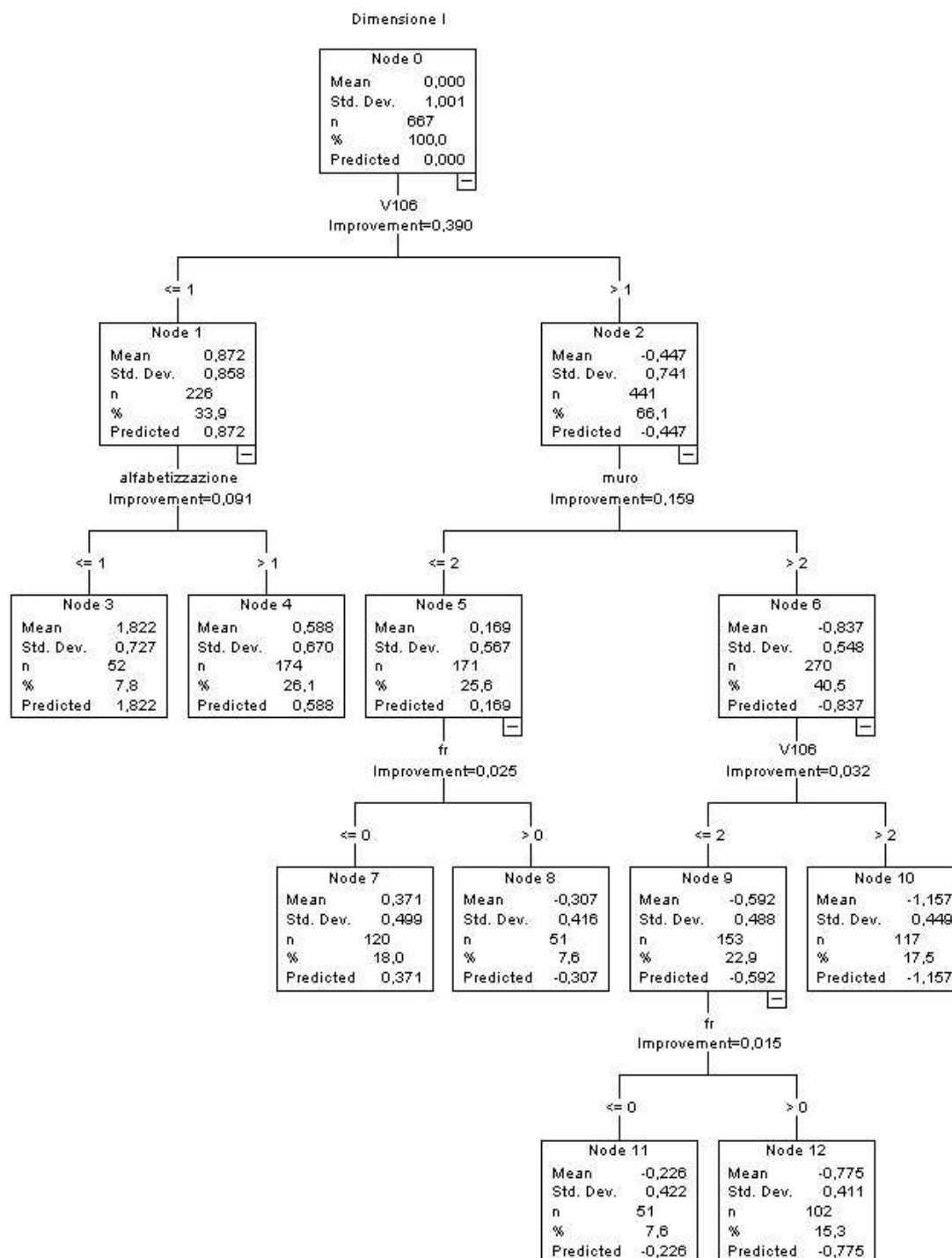
Tabella 3.6. Migliori contributi che le dimensioni forniscono alla spiegazione della variabilità delle modalità (coseni quadrati)

Dimensione 1		Dimensione 2	
Wfin	0,457	DeciSI	0,514
Wgrezzo	0,457	capoM	0,452
WCacqua	0,430	capoF	0,452
siFrigo	0,414	siFPpartn	0,369
noFrigo	0,414	noFPpartn	0,369
noElettr	0,368	DeciNI	0,368
siElettr	0,368	noL	0,191
analfa	0,284	siL	0,191
alfa	0,284	aDeciNI	0,177
WCno	0,268	aDeciSI	0,161
higher	0,244	ubrSI	0,135

L'analisi multipla delle corrispondenze ci ha permesso di indagare le relazioni fra le variabili ritenute maggiormente caratterizzanti. Per avere una visione più dettagliata di quelle che hanno maggiormente contribuito alla formazione degli assi fattoriali e alla successiva rappresentazione grafica delle modalità, si procede con l'analisi di segmentazione sulle coordinate assunte dalle singole modalità.

Ecco il dendrogramma per la prima dimensione. Nell'output, per ogni nodo, viene riportato la coordinata media, la deviazione standard e il numero di unità (e la percentuale rispetto al totale); per ogni suddivisione è segnalata la percentuale di riduzione dell'impurità (*improvement*).

Grafico 3.3. Dendrogramma delle coordinate sul primo asse fattoriale



Le variabili che sembrano meglio discriminare la prima dimensione sulla base del loro punteggio assunto nell'asse delle ordinate sono proprio quelle che registravano un coseno quadrato elevato (contributo che il fattore fornisce alla spiegazione dell'inerzia delle variabili) nell'analisi delle corrispondenze.

La prima suddivisione divide le donne intervistate fra coloro che hanno un titolo di studio almeno secondario e coloro che hanno la licenza elementare o nessun titolo di studio. La riduzione dell'impurità è del 39%. Per le donne con bassa

istruzione, l'unica altra variabile discriminante è l'alfabetizzazione (con un ulteriore guadagno in termini di omogeneità dei gruppi formati pari al 9,1%), mentre per le altre entrano in gioco il materiale delle pareti dell'abitazione, il possesso del frigorifero e il possesso di una laurea o di un diploma superiore a quello secondario.

I nodi terminali sono 7, di cui 3 con coordinata media positiva (che significa presenza di disagio economico-materiale); secondo quanto emerge, si può affermare quindi che le persone più "povere" siano individuabili semplicemente da un titolo di studio basso o da abitazioni caratterizzate sia da muri di materiale non resistente sia dall'assenza del frigorifero.

Fra le variabili evidenziate nell'analisi delle corrispondenze, non compaiono a sorpresa né l'elettricità né i servizi igienici.

Ecco il riassunto dei nodi terminali, ordinati rispetto alla media, e della loro composizione:

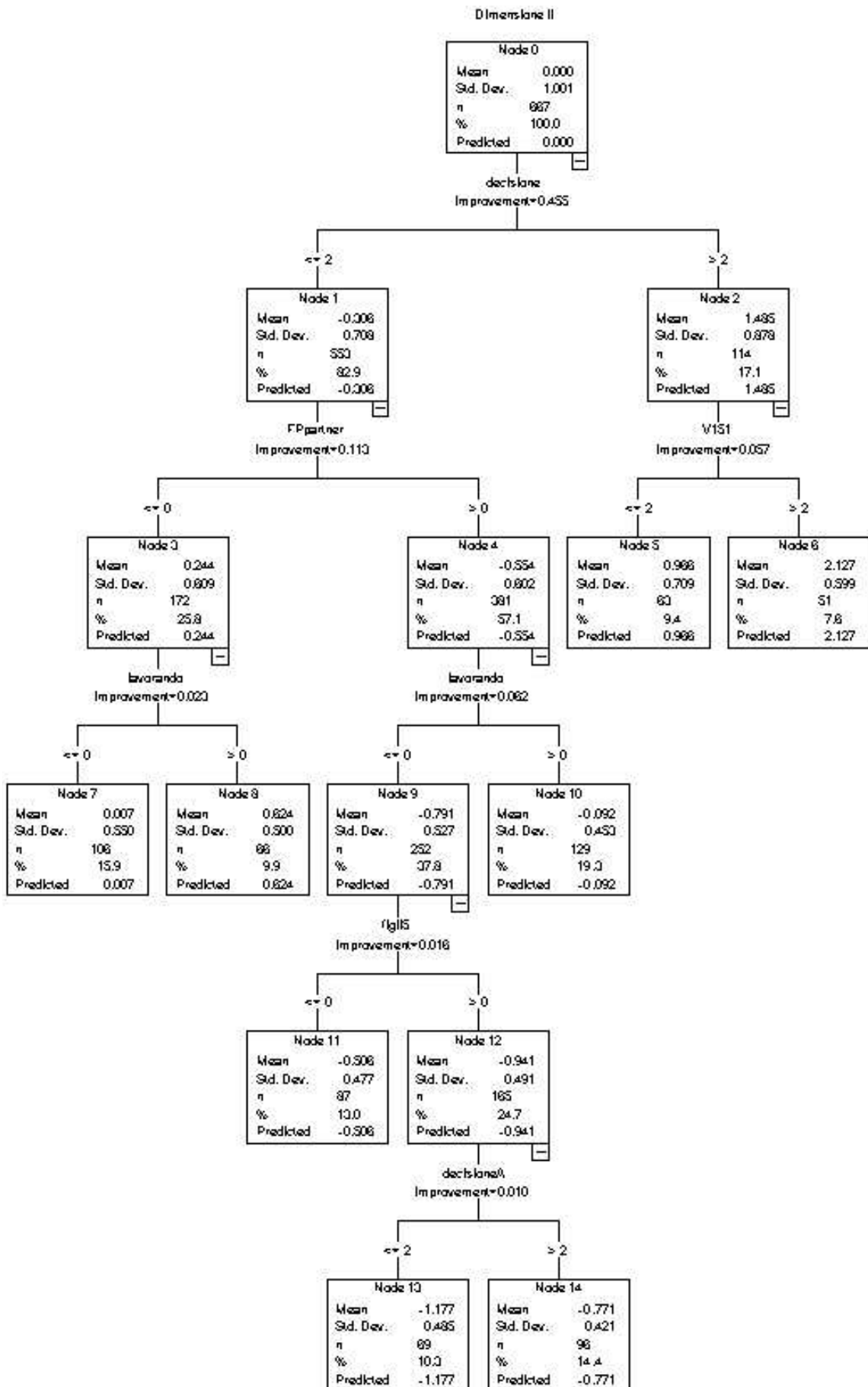
Tabella 3.7. Nodi terminali del dendrogramma per la prima dimensione, con numerosità, media e descrizione del gruppo di donne

Nodo	N°	%	Media	Descrizione
3	52	7,8%	1,82	Donne analfabete
4	174	26,1%	0,59	Donne con titolo di studio elementare o inferiore
7	120	18,0%	0,37	Donne con titolo di studio almeno secondario, con muri grezzi e senza frigorifero
11	51	7,6%	-0,23	Donne con titolo di studio secondario, con muri finiti e senza frigorifero
8	51	7,6%	-0,31	Donne con titolo di studio almeno secondario, con muri grezzi e con frigorifero
12	102	15,3%	-0,77	Donne con titolo di studio secondario, con muri finiti e con frigorifero
10	117	17,5%	-1,16	Donne con titolo di studio superiore, con muri finiti

Complessivamente, l'albero sembra spiegare una percentuale di variabilità che si aggira attorno al 71%.

Analogamente procediamo con la seconda dimensione.

Grafico 3.4. Dendrogramma delle coordinate sul secondo asse fattoriale



L'interpretazione dell'albero questa volta è più complessa, probabilmente anche a causa del fatto che il secondo asse fattoriale ha una capacità esplicativa ridotta rispetto a quella dell'altra dimensione.

Dal dendrogramma sembra emergere che la seconda dimensione spieghi, più che il disagio sociale della donna, il suo grado di indipendenza. In realtà abbiamo visto in precedenza che le due cose sono strettamente legate: in una realtà più "arretrata" e maschilista, una donna che vive da sola (per scelta o perché rimasta vedova) è sì più indipendente, ma anche maggiormente sottoposta a problemi organizzativi e sociali; questo fatto spesso si tramuta in disagio, e gli effetti possono essere, ad esempio, l'alcolismo, come emergeva dal grafico delle corrispondenze.

Ma se precedentemente si era dato più rilievo al disagio, ora l'algoritmo di segmentazione seleziona solo variabili che in qualche modo hanno a che fare con la capacità da parte delle donne di prendere decisioni, col fatto di essere sole (e quindi capofamiglia), di riuscire a lavorare e avere figli, che viene visto come un aiuto prezioso alla sopravvivenza economica. Ad esempio, la prima variabile selezionata è la capacità di prendere decisioni da soli sui grandi acquisti: essa da sola spiega il 45,5% della variabilità.

Questa nuova interpretazione è d'aiuto ai fini del nostro studio, visto che il microcredito, e molte politiche sociali, cercano di lavorare principalmente con le donne, perché più affidabili, ma che spesso sono troppo soggette al volere dei mariti; molte organizzazioni non governative aiutano le donne disagiate, ma il requisito fondamentale (di partenza o da raggiungere collaborando nel tempo) è che siano in grado di prendere decisioni sul proprio futuro e che siano indipendenti rispetto alle scelte economiche.

Riassumiamo ora i nodi terminali che l'analisi ha prodotto in tabella 3.8.

Tabella 3.8. Nodi terminali del dendrogramma per la seconda dimensione, con numerosità, media e descrizione del gruppo di donne

Nodo	N°	%	Media	Descrizione
6	51	7,6%	2,13	Donne capofamiglia in grado di prendere decisioni da sole sui grandi acquisti
5	63	9,4%	0,97	Donne, non capofamiglia, in grado di prendere decisioni da sole sui grandi acquisti
8	66	9,9%	0,62	Donne che non prendono decisioni da sole, che non fanno programmazione familiare col partner, ma che lavorano
7	106	15,9%	0,01	Donne che non prendono decisioni da sole, che non fanno programmazione familiare col partner e che non lavorano
10	129	19,3%	-0,09	Donne che non prendono decisioni da sole, che fanno programmazione familiare col partner e che lavorano
11	87	13,0%	-0,51	Donne che non prendono decisioni da sole, che fanno programmazione familiare col partner, che non lavorano e senza figli
14	96	14,4%	-0,77	Donne che non prendono decisioni da sole, che fanno programmazione familiare col partner, che non lavorano, con figli e che decidono sulla salute
13	69	10,3%	-1,18	Donne che non prendono decisioni da sole, che fanno programmazione familiare col partner, che non lavorano, con figli e che non decidono sulla salute

Complessivamente, l'albero spiega una percentuale di variabilità che si aggira attorno al 74%.

3.2 COSTA SUD: ICA

Fra i 6 Dipartimenti studiati, abbiamo visto che Ica è quello che possiede gli indicatori sulla scolarizzazione più alti; rispetto a Tumbes poi, ha l'ISU più alto e anche tutti gli indici che lo compongono sono maggiori; viceversa, è la regione costiera settentrionale che, secondo l'INEI, ha percentuali più basse di popolazione che vive sotto la soglia di povertà, sia totale che estrema. Andremo ora a vedere se queste differenze, benché minime, si rifletteranno anche sulla determinazione dei fattori.

Dopo alcuni primi tentativi, si decide di non considerare le seguenti variabili, perché nessuna delle loro modalità è importante per la spiegazione dell'inerzia dei due fattori principali: età al matrimonio, opinione sulla violenza sui figli, metodi contraccettivi.

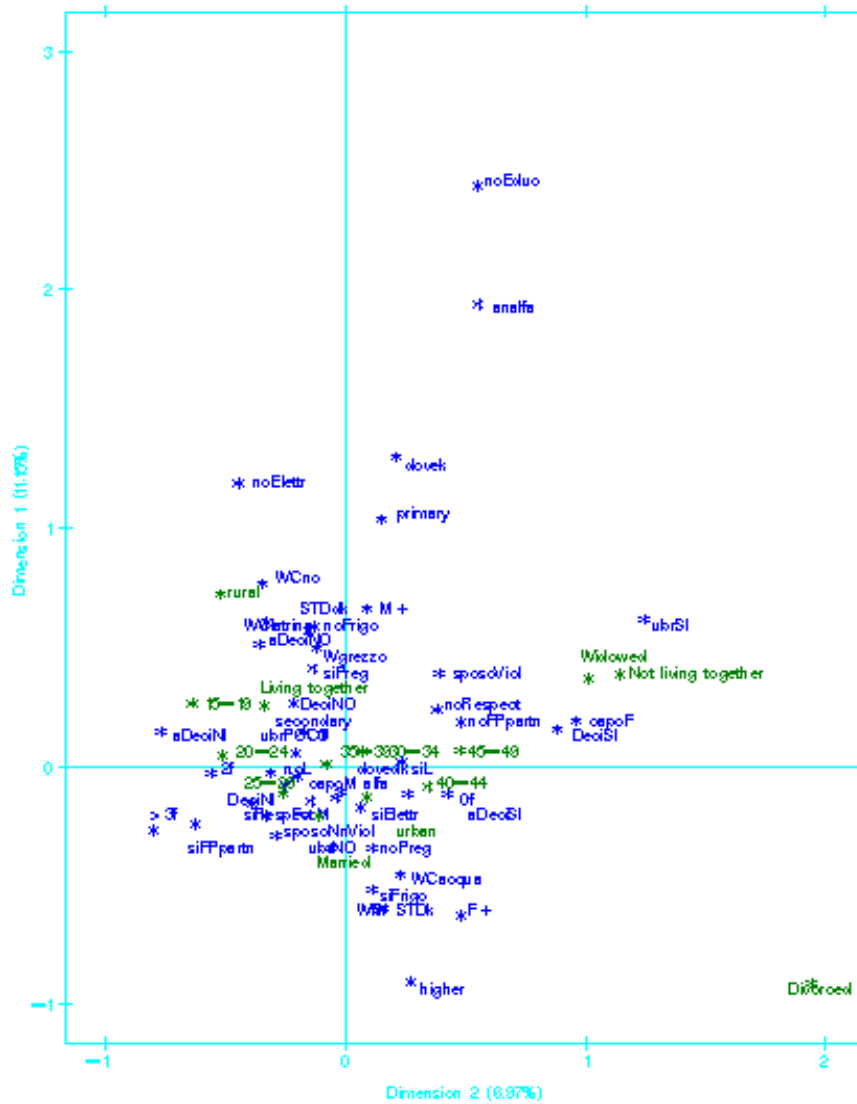
L'analisi multipla delle corrispondenze che si ottiene evidenzia la presenza di un unico fattore principale, perché il secondo spiega appena l'8% dell'inerzia totale (secondo la correzione di Greenacre). Abbiamo però osservato che molte variabili che si collocano sulla dimensione principale, contribuiscono molto poco alla spiegazione dell'inerzia; questo può voler dire che alcune di esse sono una ripetizione di informazione. Dopo ulteriori esperimenti, si è scoperto che è possibile eliminare temporaneamente le variabili alfabetizzazione, differenza titolo di studio uomo-donna, elettricità, dove e pregiudizio. Con questo accorgimento, il secondo fattore torna a spiegare una percentuale di inerzia consistente, come riportato dalle seguenti tabelle che confrontano le due scomposizioni corrette di Greenacre.

Tabella 3.9. Confronto della scomposizione dell'inerzia con la correzione di Greenacre con tutte le variabili (a sinistra) e senza quelle meno significative (a destra)

Scomposizione dell'inerzia corretta di Greenacre				Scomposizione dell'inerzia corretta di Greenacre			
Inerzia principale	Inerzia corretta	Percentuale	Percentuale cumulativa	Inerzia principale	Inerzia corretta	Percentuale	Percentuale cumulativa
0,16	0,14	13,87	13,87	0,18	0,15	14,68	14,68
0,10	0,08	8,01	21,88	0,14	0,10	10,43	25,11
0,09	0,07	6,95	28,83	0,11	0,08	7,61	32,72
0,08	0,06	5,95	34,77	0,09	0,06	6,27	39,00
0,07	0,05	5,09	39,86	0,09	0,06	5,61	44,60
0,07	0,05	4,68	44,54	0,08	0,05	5,46	50,06
0,06	0,04	4,27	48,82	0,08	0,05	5,07	55,13
0,06	0,04	4,06	52,87	0,08	0,05	4,81	59,94
0,06	0,04	3,96	56,84	0,07	0,04	4,46	64,40
0,06	0,04	3,77	60,60	0,07	0,04	4,35	68,75
0,05	0,04	3,60	64,20	0,06	0,04	3,83	72,58
0,05	0,03	3,46	67,66	0,06	0,04	3,74	76,33
0,05	0,03	3,37	71,03	0,06	0,04	3,56	79,88
0,05	0,03	3,03	74,06	0,06	0,03	3,16	83,04
0,04	0,03	2,80	76,87	0,05	0,03	2,86	85,90
0,04	0,03	2,57	79,44	0,05	0,03	2,54	88,44
0,04	0,03	2,51	81,95	0,05	0,02	2,44	90,88
0,04	0,02	2,43	84,38	0,05	0,02	2,31	93,18
0,04	0,02	2,30	86,68	0,04	0,02	2,12	95,30
0,04	0,02	2,10	88,78	0,04	0,02	1,68	96,98
0,03	0,02	1,96	90,74	0,03	0,02	1,54	98,52
0,03	0,02	1,76	92,50	0,03	0,01	1,48	100,00
0,03	0,02	1,57	94,07				
0,03	0,01	1,46	95,54	Totale	1,00	100,00	
0,03	0,01	1,44	96,98				
0,03	0,01	1,34	98,32				
0,02	0,01	1,19	99,51				
0,01	0,00	0,49	100,00				
Totale	1,00	100,00					

Da questo confronto quindi si restituisce importanza alla seconda dimensione. Nel grafico delle corrispondenze però riporteremo i risultati ottenuti col primo tentativo, al fine di non trascurare alcuna informazione.

Grafico 3.5. Rappresentazione grafica delle modalità rispetto ai primi due assi fattoriali trovati con l'analisi delle corrispondenze per il Dipartimento di Ica



Il primo semiasse positivo è ancora una volta caratterizzato dall'istruzione, ma questa volta appaiono tutte e tre le variabili che la contrassegnano (titolo di studio, alfabetizzazione e differenza maschio-femmina); le prime modalità che si incontrano sono:

- nessun titolo di studio
- analfabetismo

- non conoscenza di dove comprare i preservativi
- no elettricità
- licenza elementare

Esse sono seguite da: no bagno, titolo di studio maggiore dell'uomo, non conosce le STD.

Dalla parte opposta invece vengono proiettate le seguenti modalità:

- livello di istruzione superiore
- titolo di studio maggiore della donna
- muro con materiali finiti
- conoscenza delle STD
- si frigo
- bagno con acqua corrente

Le prime modalità che si incontrano nel semiasse positivo delle ascisse sono:

- alcolismo (che però influisce anche sul semiasse positivo delle ordinate, perché si colloca nel secondo quadrante)
- capo-famiglia donna
- capacità di prendere decisioni sui grandi acquisti
- capacità di prendere decisioni sulla salute

Hanno invece le coordinate più piccole sull'asse orizzontale:

- 3 figli o più
- presa di decisioni sui grandi acquisti con altre persone
- capacità di pianificazione familiare
- 2 figli

Differentemente da quanto accade per Tumbes, la presa di decisioni sui temi inerenti la salute, la differenza di istruzione fra la donna e il suo partner, la conoscenza delle STD e il numero di figli, sembrano giocare un ruolo più importante nella determinazione dei fattori principali.

La proiezione sul grafico delle tre variabili supplementari (età, urbanizzazione, stato civile) indica che:

- le donne divorziate si posizionano isolate nel quarto quadrante, come accadeva a livello nazionale;
- l'età della donna si appiattisce lungo l'asse orizzontale, seguendo l'ordine crescente delle sue modalità;
- le modalità supplementari che si collocano agli estremi dell'asse delle ordinate sono "rurale" (semiasse positivo) e "sposate" (semiasse negativo).

La lettura del grafico è supportata dalla tabella seguente, dove vengono evidenziate le prime modalità per importanza nella spiegazione dell'inerzia di ogni dimensione.

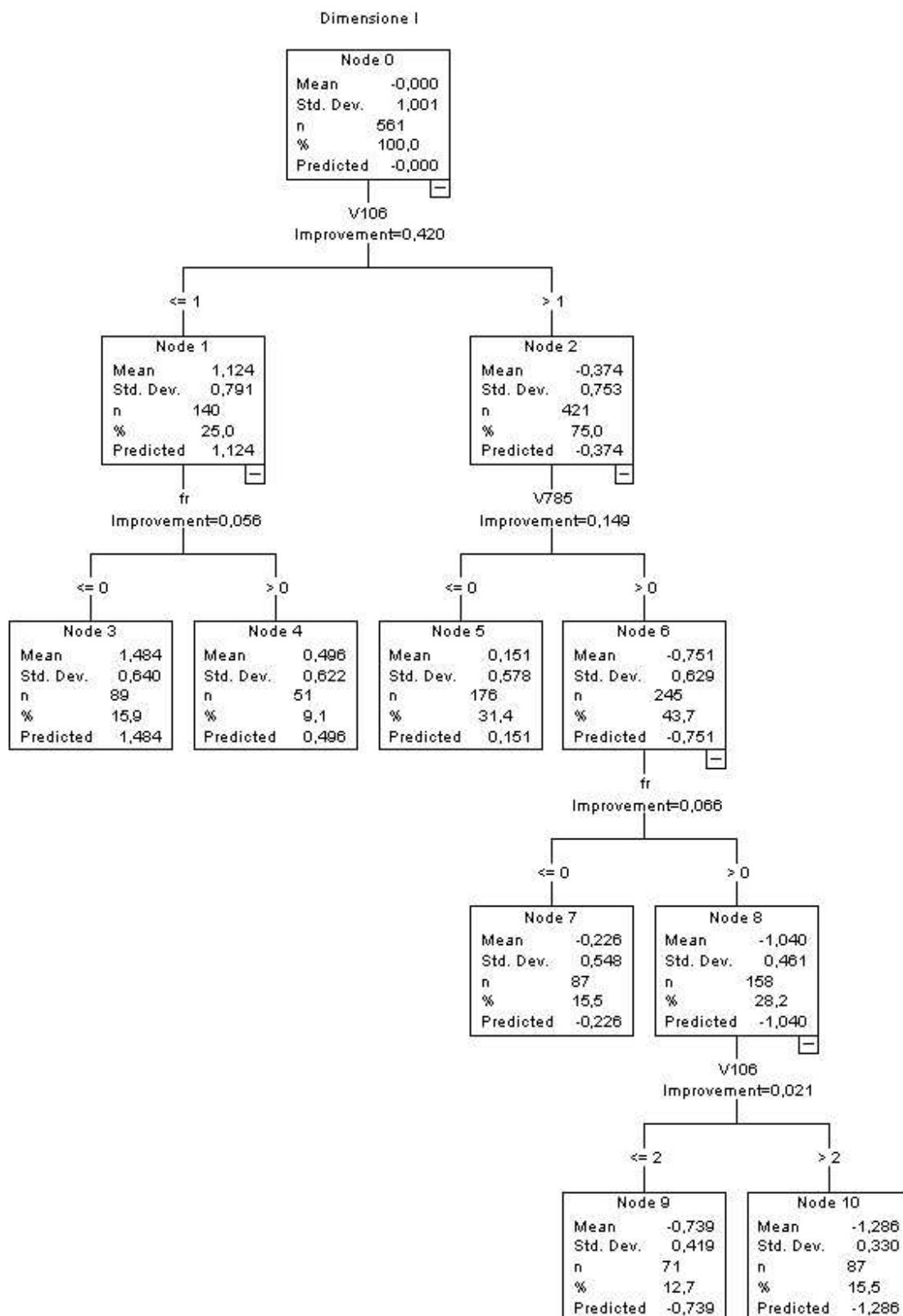
Tabella 3.10. Migliori modalità per contributo parziale all'inerzia delle due dimensioni

Dimensione 1		Dimensione 2	
primary	0,081	DeciSI	0,113
analfa	0,077	ubrSI	0,089
higher	0,071	siFPpartn	0,087
noElettr	0,057	capoF	0,082
STDk	0,055	noFPpartn	0,067
STDdk	0,053	aDeciNI	0,058
Wfin	0,052	DeciNI	0,043
noFrigo	0,052	Of	0,038
siFrigo	0,046	noRespect	0,035
Wgrezzo	0,043	sposoViol	0,033
dovek	0,041	siRespect	0,030
WCacqua	0,039	aDeciSI	0,025
M+	0,036	sposoNnViol	0,025
WClatrina	0,034	2f	0,022
noEduc	0,030	noL	0,021
siPreg	0,025	capoM	0,017
WCno	0,024	WCacqua	0,016

La tabella conferma quanto detto in precedenza sull'asse verticale. Fra le modalità più importanti per la spiegazione della seconda dimensione invece scompaiono le donne con più di 3 figli (probabilmente a causa della ridotta frequenza), mentre sembrano rilevanti le violenze subite, sia fisiche che verbali.

Ora procediamo con l'analisi di segmentazione sui punteggi fattoriali delle singole modalità. Abbiamo utilizzato tutte le variabili, al di là della loro significatività nell'analisi delle corrispondenze. Ecco il dendrogramma per la prima dimensione.

Grafico 3.6. Dendrogramma delle coordinate sul primo asse fattoriale



L'albero termina in 6 foglie finali, ma le variabili che compaiono sono solo 3, ripetute più volte: il titolo di studio, il frigorifero e la conoscenza delle malattie sessualmente trasmissibili. Questo numero ridotto da un lato certamente

sorprende, dall'altro però conferma quanto emerso all'inizio della ricerca dei fattori, cioè che molte variabili presenti sull'asse verticale ripetono le medesime informazioni; questo quindi confermerebbe la loro esclusione, e quindi la presenza di un secondo fattore più significativo (vedi tabella 3.8 con confronto scomposizione di Greenacre).

La prima suddivisione separa, anche per il Dipartimento di Ica, le donne intervistate fra coloro che hanno un titolo di studio almeno secondario e coloro che hanno la licenza elementare o nessun titolo di studio. La riduzione della variabilità è del 42%. Per queste ultime, l'unica altra variabile discriminante è questa volta il possesso del frigorifero (con un ulteriore guadagno in termini di omogeneità dei gruppi formati pari al 5,6%). Le donne invece che posseggono un titolo di studio superiore a quello elementare, vengono suddivise in base alla conoscenza delle STD (ultima ramificazione che riconduce alle donne più "povere", cioè alla coordinata positiva). Coloro che hanno risposto in modo affermativo alla domanda sulla conoscenza, vengono successivamente suddivise in base, di nuovo, alla variabile frigo e titolo di studio.

Nella tabella vengono riportati i nodi terminali. Si vede che quelli con media positiva (quindi quelli composti dalle donne più "povere") sono caratterizzati o da un titolo di studio basso o dall'ignorare cosa siano le malattie sessualmente trasmissibili.

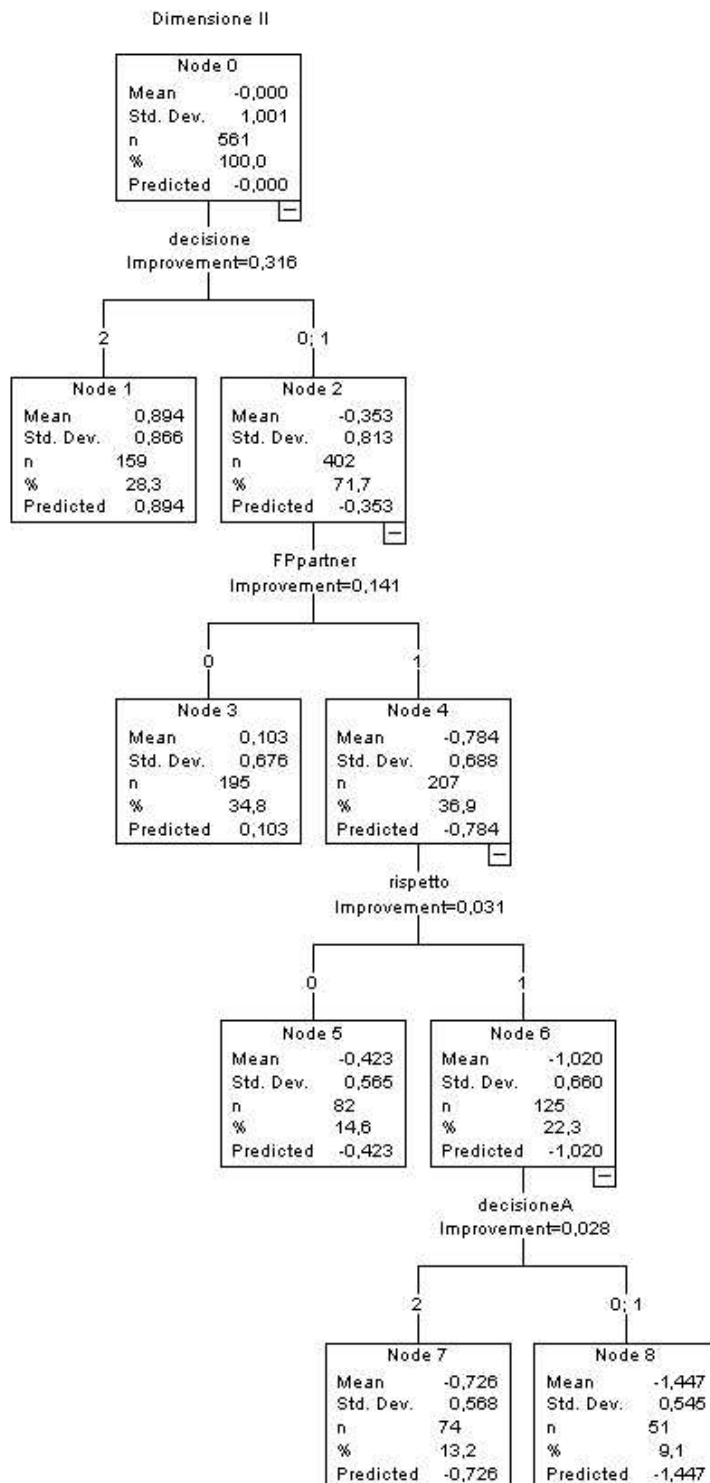
Tabella 3.11. Nodi terminali del dendrogramma per la prima dimensione, con numerosità, media e descrizione del gruppo di donne

Nodo	N°	%	Media	Descrizione
3	89	15,9%	1,48	Donne con titolo di studio elementare o inferiore, senza frigorifero
4	51	9,1%	0,49	Donne con titolo di studio elementare o inferiore, con frigorifero
5	176	31,4%	0,15	Donne con titolo di studio superiore, ignoranti le STD
7	87	15,5%	-0,22	Donne con titolo di studio superiore, conoscenti le STD, senza frigorifero
9	71	12,7%	-0,74	Donne con licenza secondaria, conoscenti le STD, con frigorifero
10	87	15,5%	-1,29	Donne con laurea, conoscenti le STD, con frigorifero

Complessivamente, l'albero sembra spiegare una percentuale di variabilità che si aggira attorno al 71%.

Analogamente procediamo con la seconda dimensione.

Grafico 3.7. Dendrogramma delle coordinate sul secondo asse fattoriale



Il dendrogramma è completamente asimmetrico; per caratterizzare il semi-asse positivo è infatti sufficiente trovare quelle donne che sanno prendere da sole le

decisioni sui grandi acquisti. Questa variabile supera per importanza sia il fatto di essere capofamiglia sia i livelli di alcolismo.

Più complessa invece la suddivisione nell'altro semiasse: le variabili che di volta in volta vengono selezionate per individuare le donne che si collocano agli estremi negativi sono, in ordine, la pianificazione familiare col partner, la violenza verbale subita e le decisioni sui temi inerenti la salute.

A differenza di quanto avviene per il Dipartimento di Tumbes, ora non si può più parlare solo di "indipendenza della donna", perché le variabili questa volta sono più strettamente legate al maltrattamento e all'incapacità di potersi esprimere. Anche il fatto che la presa di decisione sui grandi acquisti sia la variabile discriminante principale fra semiasse positivo e negativo, mentre la presa di decisione sui temi inerenti la salute sia l'ultima variabile che seleziona le donne poste all'estremo negativo, è significativo: si è infatti visto anche in analisi precedenti che la prima determina il grado di indipendenza della donna, mentre la seconda, legata ad una sfera più delicata quale è la salute, è implicitamente connessa al rapporto col partner.

Un'interpretazione potrebbe essere che in un unico fattore (il secondo) è unito il tema dell'indipendenza, al di là del loro disagio percepito, con quello del disagio percepito a causa della presenza opprimente del marito. Questa spiegazione ovviamente implicherebbe un'analisi separata per le donne attualmente conviventi e quelle invece che vivono sole.

Riassumiamo ora i nodi terminali che l'analisi ha prodotto.

Tabella 3.12. Nodi terminali del dendrogramma per la seconda dimensione, con numerosità, media e descrizione del gruppo di donne

Nodo	N°	%	Media	Descrizione
1	159	28,3%	0,89	Donne in grado di prendere decisioni da sole sui grandi acquisti
3	195	34,8%	0,10	Donne non in grado di prendere decisioni da sole sui grandi acquisti, che non fanno pianificazione familiare col partner
5	82	14,6%	-0,42	Donne non in grado di prendere decisioni da sole sui grandi acquisti, che fanno pianificazione familiare col partner e che dal quale subiscono violenza verbale (no rispetto)
7	74	13,2%	-0,73	Donne non in grado di prendere decisioni da sole sui grandi acquisti, che fanno pianificazione familiare col partner, dal quale sono rispettate e che decidono da sole sulla salute
8	51	9,1%	-1,45	Donne non in grado di prendere decisioni da sole sui grandi acquisti, che fanno pianificazione familiare col partner, dal quale sono rispettate ma che non decidono da sole sulla salute

Complessivamente, l'albero spiega una percentuale di variabilità che si aggira attorno al 52%.

3.3 ZONA METROPOLITANA: LIMA

Continuiamo ad analizzare la fascia costiera, ma spostandoci però in zona urbana, cioè nella città metropolitana di Lima, capitale del Perù. Essa desta particolare interesse, perché ovviamente in una grande città si instaurano dinamiche sociali diverse da quelle del resto del Paese, e la tipologia di "poveri" sarà diversa da quella dei *campesinos* che vivono nelle campagne o nelle alture. Rispetto agli altri Dipartimenti, Lima vanta l'ISU più alto, ma la proporzione di abitanti che vivono in condizioni di povertà è superiore a quello di altre regioni costiere.

Essendo interessati alla zona metropolitana più che alla regione Lima, le analisi verteranno unicamente su quell'84,54% degli abitanti "limeñi" che vivono in città, e non quelli nelle campagne subito al di fuori di essa.

Si procede con l'analisi delle corrispondenze. Dopo alcuni tentativi, si decide di non considerare alcune variabili poco significative, al fine di snellire il grafico e semplificare le analisi: numero di figli, opinione sulla violenza, decisioni sulla salute, elettricità, lavoro, sesso del capofamiglia, decisione sugli acquisti, pianificazione familiare.

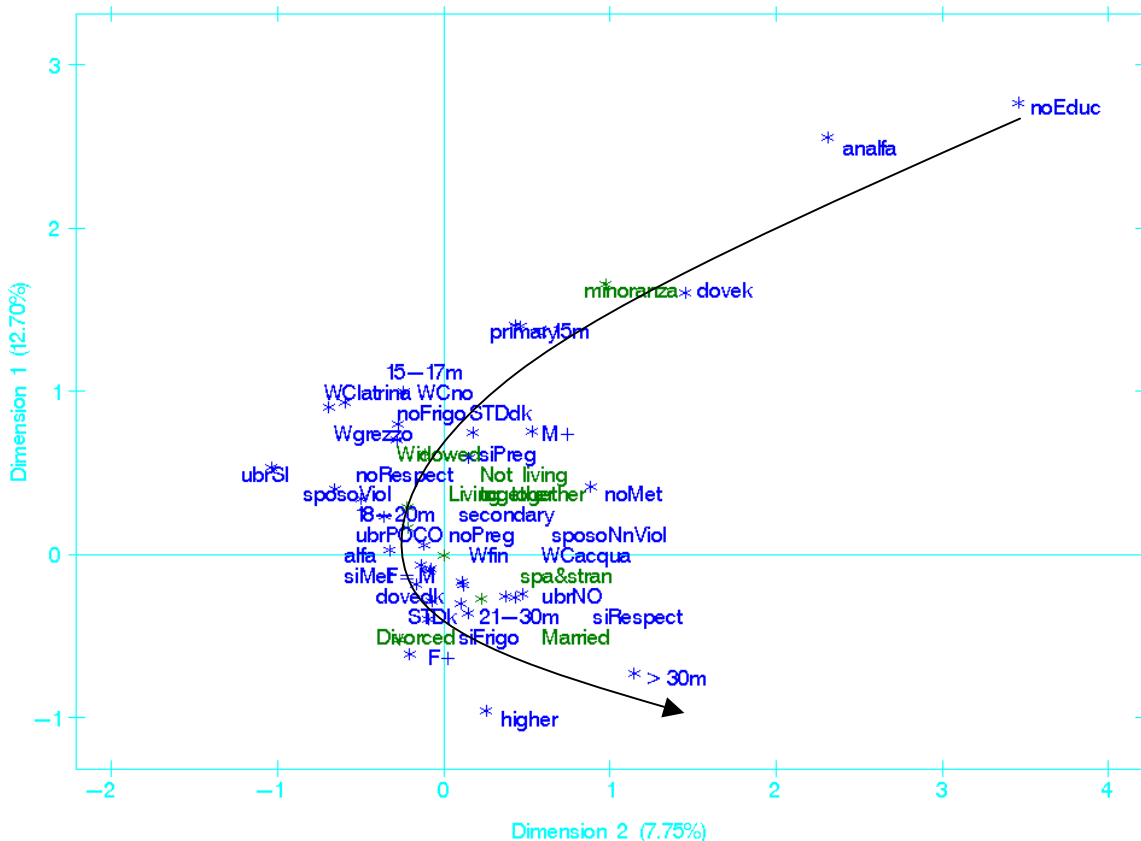
Dall'analisi risulta che esiste un unico fattore principale (12,7% di inerzia spiegata secondo la scomposizione del Chi-quadrato), mentre il secondo spiega meno dell'8% dell'inerzia totale.

Tabella 3.13. scomposizione dell'inerzia con la correzione di Greenacre

Scomposizione dell'inerzia corretta di Greenacre				
Inerzia principale	Inerzia corretta	Percentuale	Percentuale cumulativa	3 6 9 12 15
0,20	0,16	16,29	16,29
0,12	0,09	8,91	25,20
0,10	0,07	6,95	32,15
0,09	0,06	6,23	38,38
0,08	0,06	5,58	43,95
0,08	0,05	5,10	49,05
0,08	0,05	4,85	53,90
0,07	0,05	4,63	58,53
0,07	0,05	4,49	63,02
0,07	0,04	4,09	67,11
0,07	0,04	3,97	71,08
0,06	0,04	3,58	74,65
0,06	0,04	3,50	78,15
0,06	0,03	3,44	81,59
0,06	0,03	3,16	84,75
0,06	0,03	3,11	87,87
0,05	0,03	3,00	90,87
0,05	0,03	2,59	93,46
0,04	0,02	2,18	95,64
0,04	0,02	1,98	97,63
0,04	0,02	1,71	99,34
0,02	0,01	0,66	100,00
Totale	1,00	100,00		

Dal grafico delle corrispondenze si capisce il motivo della debolezza della seconda dimensione.

Grafico 3.8. Rappresentazione grafica delle modalità rispetto ai primi due assi fattoriali trovati con l'analisi delle corrispondenze per il Dipartimento di Lima



Come si vede, i dati presentano la classica forma a ferro di cavallo, o effetto Guttman: questo indica una sostanziale unidimensionalità delle frequenze osservate, e il fattore non corre lungo gli assi coordinati, ma lungo il ferro di cavallo.

La larghezza di tale ferro di cavallo nella parte centrale fa supporre che probabilmente la seconda dimensione ha una qualche importanza nell'interpretazione, ma sostanzialmente il fattore che meglio descrive i dati è unico. Per tale motivo è meglio ordinare le modalità rispetto alla loro posizione lungo la freccia presente nel grafico.

Le modalità che si trovano nella parte superiore del “ferro di cavallo” sono:

- nessun titolo di studio
- analfabetismo
- non conoscenza di dove comprare i preservativi
- sposate prima dei 15 anni d'età
- licenza elementare
- no servizi igienici e bagno rudimentale (latrina)
- no frigo
- muro grezzo
- uomo con istruzione maggiore
- pregiudizi nei confronti dei portatori di AIDS

Alla fine della freccia invece si incontrano:

- donne sposate dopo i trent'anni
- livello di istruzione superiore
- no violenza verbale
- donna con titolo di studio superiore a quello del partner
- si frigo

Agli estremi di questo fattore a “ferro di cavallo” compaiono, per la prima volta, l'età al matrimonio e il pregiudizio verso i portatori di AIDS. Inoltre si nota subito come vengano mischiate le variabili più “materiali” con quelle più “sociali”.

L'effetto Guttman, e il conseguente mescolamento delle variabili fra le due dimensioni iniziali, si vede anche dalla tabella dei contributi parziali alle inerzie dei due fattori.

Tabella 3.14. Migliori modalità per contributo parziale all'inerzia delle due dimensioni

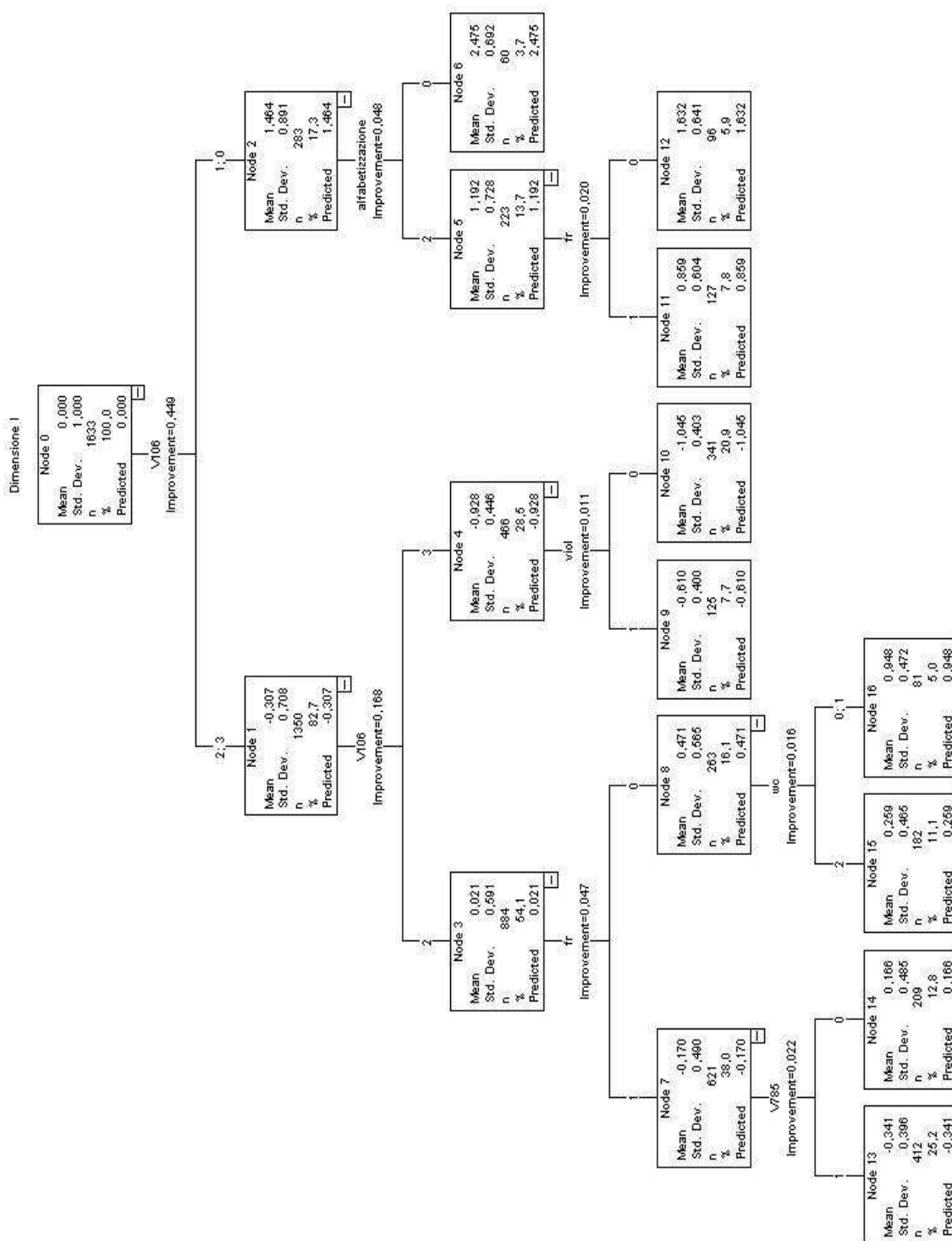
Dimensione 1		Dimensione 2	
primary	0,116	analfa	0,115
higher	0,093	sposoViol	0,100
analfa	0,086	noEduc	0,069
STDdk	0,069	sposoNnViol	0,066
noFrigo	0,063	dovek	0,064
Wgrezzo	0,052	noRespect	0,062
M+	0,050	noMet	0,061
dovek	0,047	ubrSI	0,053
siPreg	0,041	siRespect	0,046
WClatrina	0,036	ubrNO	0,044
STDk	0,036	M+	0,040
noEduc	0,027	WClatrina	0,035
15-17m	0,026	Wgrezzo	0,034
<15m	0,025	secondary	0,034
21-30m	0,024		
siFrigo	0,024		

Per quanto concerne le variabili illustrative, abbiamo deciso di non proiettare la fascia d'età, perché le modalità di collocavano tutte al centro e rendevano impossibile la lettura. Per quanto riguarda invece il gruppo etnico e lo stato civile si nota che:

- le minoranze etniche si pongono all'inizio del "ferro di cavallo", cioè fra le persone più disagiate;
- nella parte inferiore della freccia ci sono invece le donne sposate, seguite dalle donne divorziate e dalle donne di etnia spagnola;
- le altre modalità dello stato civile si collocano assieme vicino all'origine, nel primo quadrante.

Nonostante l'effetto Guttman e quanto detto fin'ora, decidiamo di svolgere comunque l'analisi di segmentazione per vedere se emerge qualche informazione in più sui primi due fattori individuati dal programma. Ecco il dendrogramma per la prima dimensione.

Grafico 3.9. Dendrogramma delle coordinate sul primo asse fattoriale



Al primo impatto, si osserva subito che l'albero si estende molto più in larghezza che in profondità, che esso è dominato nelle prime tre ramificazioni da variabili che riguardano l'istruzione, e che alle solite variabili "materiali" si aggiunge la violenza fisica.

Al terzo livello della ramificazione, cioè dopo la suddivisione in donne analfabete, con titolo di studio elementare, secondario e superiore, compaiono le variabili frigo, violenza fisica, bagno e conoscenza delle STD.

L'albero termine in 9 foglie finali, e spiega quasi l'80% della variabilità.

L'analisi di segmentazione è stata condotta anche sulla seconda dimensione (asse delle ascisse nel piano cartesiano). Come anticipato, il secondo asse fattoriale ha una capacità esplicativa molto ridotta, e quindi non riportiamo qui il suo dendrogramma, ma ne riassumiamo solo le caratteristiche:

- inserendo tutte le variabili esplicative l'algoritmo suddivideva il campione in 8 parti;
- la variabile per la quale vengono distinte le donne è quella relativa alla violenza fisica subita dalle donne, con una guadagno in termini di omogeneità all'interno dei gruppi pari al 30% circa;
- le altre variabili selezionate sono: alcolismo, presa di decisione su salute e sugli acquisti, sesso del capofamiglia e violenza verbale.

Non compare in nessuno dei due dendrogrammi l'età al matrimonio, che invece risultava importante per il fattore individuato nell'analisi delle corrispondenze. E, ancora una volta, le variabili titolo di studio e analfabetismo vengono preferite alla differenza di titolo di studio fra maschio e femmina, nonostante quanto emerso nelle corrispondenze.

3.4 GIUNGLA BASSA: LORETO

Iniziamo ora ad analizzare i 3 Dipartimenti situati a est della Cordigliera delle Ande. Partiremo dal più basso per altitudine, Loreto, fino ad arrivare a quello situato sulla *sierra*, Apurimac.

Fra le tre regioni, Loreto è quella con ISU più alto. Per quanto riguarda invece la percentuale di persone che vivono in condizioni di povertà, avevamo visto che tale percentuale tende a crescere al crescere dell'altitudine se parliamo di povertà generale, mentre è più alta per le regioni più basse (cioè quelle nella foresta Amazzonica) se ci riferiamo alla soglia di povertà estrema. Questo ci fa

ipotizzare quindi che saranno discriminanti le variabili riguardanti i servizi basici piuttosto che quelle riguardanti il titolo di studio e il lavoro.

Iniziamo con l'analisi multipla delle corrispondenze. Anche in questo caso alcune variabili sono così poco significative per la spiegazione dei due assi fattoriali principali, che non vengono considerate per non generare confusione; tali variabili sono: opinione sulla violenza, pianificazione familiare, metodi contraccettivi, numero di figli, lavoro e età al matrimonio.

L'analisi mostra la esistenza di un fattore molto importante, che da solo spiega il 17,52% dell'inerzia (percentuale che sale a quasi il 23% considerando la correzione di Greenacre), mentre la seconda dimensione contribuisce appena con il 7,79% (8,79% con Greenacre).

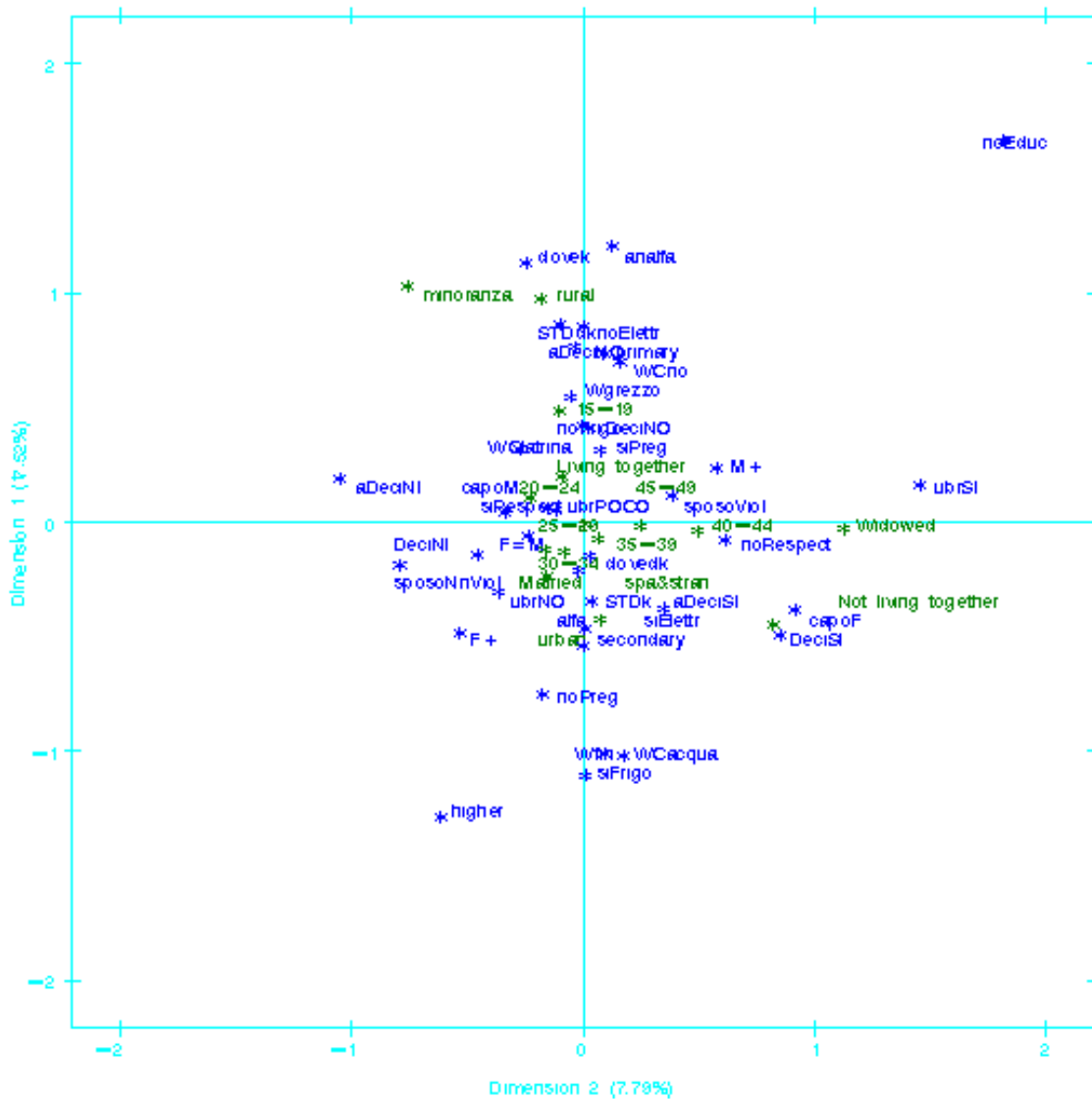
Tabella 3.15. scomposizione dell'inerzia corretta di Greenacre

Scomposizione dell'inerzia corretta di Greenacre				
Inerzia principale	Inerzia corretta	Percentuale	Percentuale cumulativa	5 10 15 20 25
0,25	0,22	22,92	22,92	*****
0,11	0,08	8,79	31,71	*****
0,09	0,07	7,04	38,75	*****
0,09	0,06	6,60	45,35	*****
0,08	0,06	5,82	51,17	*****
0,07	0,05	4,97	56,14	*****
0,07	0,04	4,56	60,69	*****
0,07	0,04	4,45	65,14	****
0,06	0,04	4,23	69,37	****
0,06	0,04	3,81	73,18	****
0,05	0,03	3,50	76,68	****
0,05	0,03	3,41	80,09	***
0,05	0,03	3,13	83,21	***
0,05	0,03	2,77	85,99	***
0,04	0,02	2,53	88,52	***
0,04	0,02	2,03	90,55	**
0,04	0,02	1,97	92,52	**
0,03	0,02	1,74	94,27	**
0,03	0,02	1,60	95,87	**
0,03	0,01	1,45	97,32	*
0,03	0,01	1,25	98,57	*
0,02	0,01	0,82	99,39	*
0,02	0,01	0,61	100,00	*
Totale	0,96	100,00		

Anche in questo caso la presenza di un unico fattore dominante può essere spiegata dal contesto socio-economico della regione: le alte percentuali di persone al di sotto della “linea di povertà estrema” fanno pensare che i problemi economici e quelli più socio-psicologico, nomenclature utilizzate fin’ora per descrivere i fattori, vengano percepiti come un’unica situazione generale di disagio dalle donne che si trovano in situazioni di difficoltà estrema.

Procediamo però per passi, e vediamo se questa tesi trova conferma dalle analisi dei dati.

Grafico 3.10. Rappresentazione grafica delle modalità rispetto ai primi due assi fattoriali trovati con l’analisi delle corrispondenze per il Dipartimento di Loreto



Prima di iniziare l'analisi degli assi cartesiani, si nota che le donne senza titolo di studio e quelle col più alto titolo di studio si scostano dall'asse delle ordinate ponendosi, rispettivamente, nel primo e nel terzo quadrante. Questo significa che gli estremi della variabile "titolo di studio" influenzano entrambi i fattori.

Escludendo queste due modalità, il semiasse positivo delle ordinate è contrassegnato da:

- analfabetismo
- non conoscenza di dove comprare i preservativi
- no elettricità
- disinformazione sulle STD

Nel semiasse negativo invece sono presenti:

- si frigo
- muro con materiali finiti
- servizi igienici con acqua corrente
- no pregiudizio sui malati di AIDS

L'asse delle ascisse invece è caratterizzato da molte meno modalità; in un estremo infatti si contraddistinguono solo le donne soggette ad alcolismo, mentre nel lato opposto ci sono le donne che prendono decisioni non da sole e quelle che non subiscono violenze domestiche.

Le variabili supplementari (età, urbanizzazione, stato civile, gruppo etnico) si comportano nel seguente modo:

- le modalità di urbanizzazione si collocano esattamente agli estremi dell'asse verticale: questo è molto importante, perché significa che, in una regione senza grandi città di rilievo, riuscire a vivere in una zona urbanizzata migliora le condizioni di vita;
- le donne vedove sono situate nel semiasse positivo della seconda dimensione, fra i problemi di alcolismo e la violenza verbale;
- le donne che vivono da sole sono ovviamente capofamiglia e fanno prendere da sole le decisioni sui grandi acquisti; la differenza rispetto alle donne nelle stesse condizioni di altre regioni, è che esse si collocano nel

quarto quadrante, e quindi vivono anche in una condizione economica più stabile;

- non ci sono donne divorziate nel campione;
- le intervistate più giovani sono le più povere, mentre le altre fasce d'età sono disperse attorno all'origine degli assi;
- le donne appartenenti a qualche minoranza etnica sono nel primo quadrante, molto vicino al semiasse del disagio "materiale", mentre le donne di origine ispanica sono nel quadrante opposto.

La lettura del grafico è supportata dalla tabella seguente, dove vengono evidenziate le prime modalità per importanza nella spiegazione dell'inerzia di ogni dimensione.

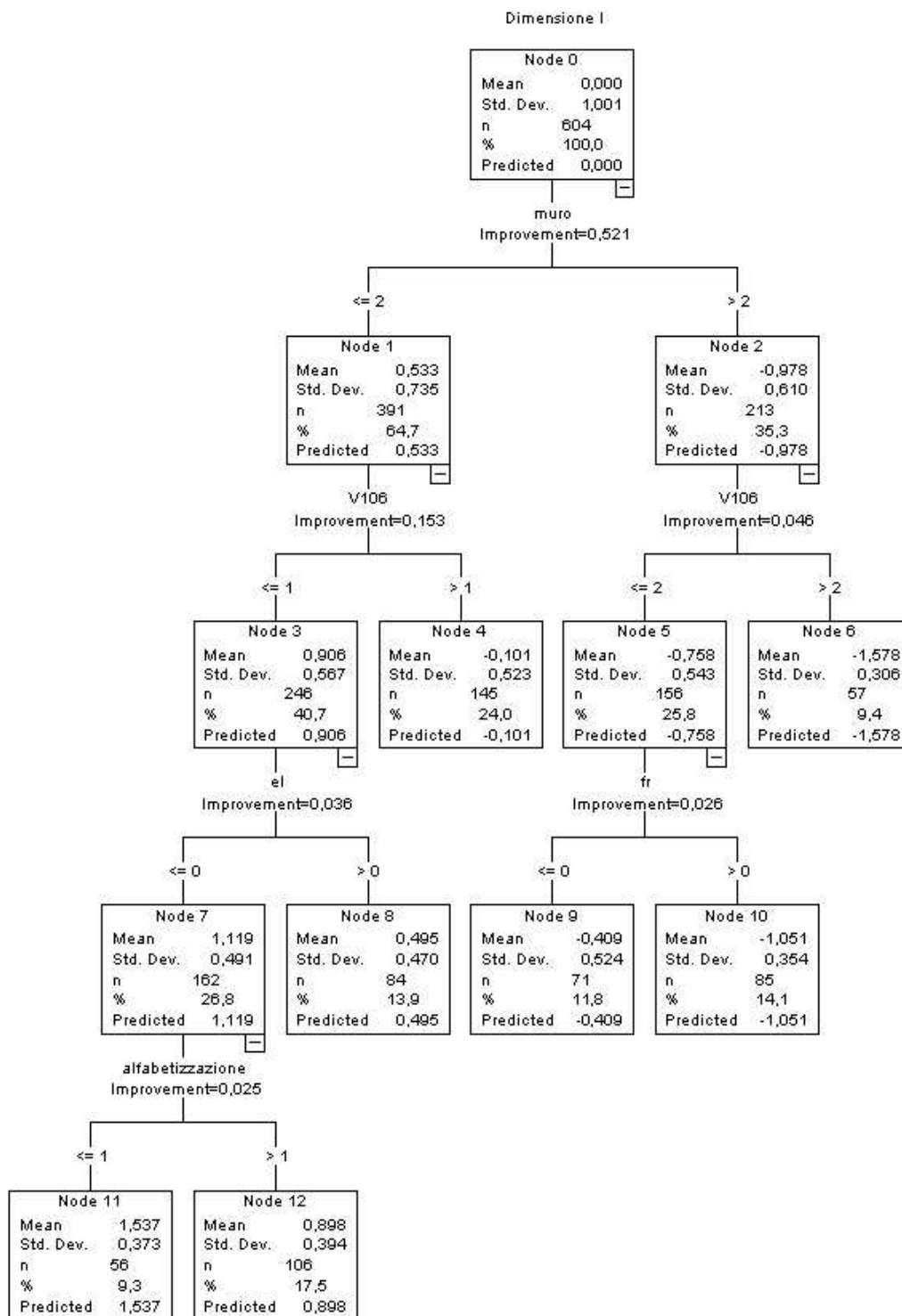
Tabella 3.16. Migliori modalità per contributo parziale all'inerzia delle due dimensioni

Dim 1		Dim 2	
Wfin	0,089	ubrSI	0,120
siFrigo	0,084	aDeciNI	0,113
WCacqua	0,083	DeciSI	0,105
noElettr	0,070	DeciNI	0,103
primary	0,059	noRespect	0,075
analfa	0,054	capoF	0,068
STDdk	0,052	M+	0,061
higher	0,052	sposoNnViol	0,054
Wgrezzo	0,048	sposoViol	0,045
siElettr	0,044	siRespect	0,041
noPreg	0,042	aDeciSI	0,039
dovek	0,036		
WCno	0,036		
aDeciNO	0,035		
noFrigo	0,032		

Le prime quattro modalità per la prima dimensione riguardano tutte variabili relative all'abitazione, questo a conferma delle nostre ipotesi iniziali. Per quanto concerne la seconda dimensione, le modalità che meglio ne spiegano l'inerzia sono le medesime che si collocano agli estremi dell'asse delle ascisse nel grafico delle corrispondenze

Passiamo ora all'analisi di segmentazione sui punteggi fattoriali delle singole modalità, utilizzando tutte le variabili, anche quelle escluse nell'analisi delle corrispondenze.

Grafico 3.11. Dendrogramma delle coordinate sul primo asse fattoriale



Fra i dendrogrammi costruiti, questo è sicuramente quello più simmetrico (almeno nella prima parte); la prima suddivisione per la prima volta utilizza una variabile diversa dal titolo di studio: le donne vengono separate in due gruppi in base al materiale di costruzione delle pareti dell'abitazione. La riduzione dell'impurità è del 52,1%.

L'istruzione viene introdotta solo al secondo livello: le donne che vivono in case con i muri di materiale grezzo vengono distinte fra quelle senza titolo di studio e quelle con licenza elementare, mentre quelle che vivono in abitazioni più solide si distinguono fra titolo di studio secondario e superiore. Successivamente, per le intervistate con istruzione più bassa, entra in gioco la presenza di elettricità nell'abitazione e, dove non è presente, l'alfabetizzazione. Invece, per quelle con licenza superiore, diventa distintiva la variabile frigorifero.

In generale si osserva che per appartenere al gruppo di donne meno disagiate (coordinata negativa) è sufficiente abitare in una casa con muri in mattoni o aver frequentato almeno la scuola di primo livello.

Nella tabella vengono riportati i nodi terminali. Si vede che quelli con media positiva (quindi quelli composti dalle donne più "povere") sono caratterizzati o da un titolo di studio basso o dall'ignorare cosa siano le malattie sessualmente trasmissibili.

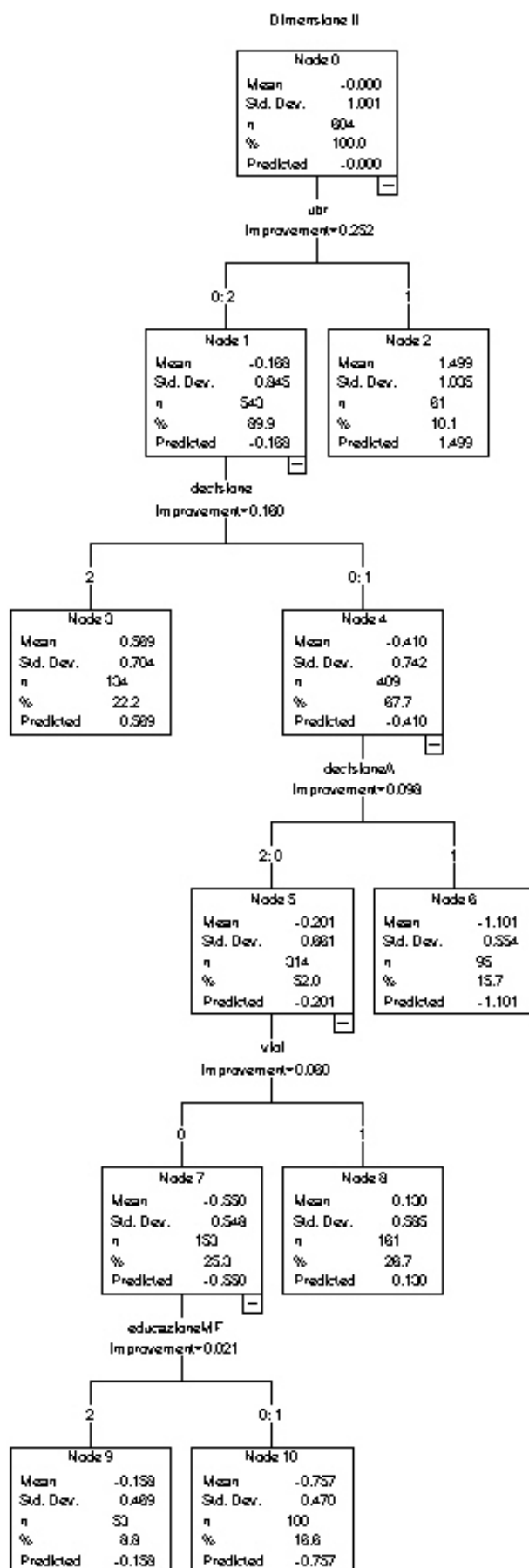
Tabella 3.17. Nodi terminali del dendrogramma per la prima dimensione, con numerosità, media e descrizione del gruppo di donne

Nodo	N°	%	Media	Descrizione
11	56	9,3%	1,54	Donne analfabete, in abitazioni con muri di materiale grezzo e senza elettricità
12	106	17,5%	0,90	Donne senza titolo di studio, in abitazioni con muri di materiale grezzo e senza elettricità
8	84	13,9%	0,49	Donne senza titolo di studio, in abitazioni con muri di materiale grezzo ma con elettricità
4	145	24,0%	-0,10	Donne con licenza elementare, in abitazioni con muri di materiale grezzo
9	71	11,8%	-0,41	Donne con titolo di studio secondario, in abitazioni con muri di materiale finito, ma senza frigorifero
10	85	14,1%	-1,05	Donne con titolo di studio secondario, in abitazioni con muri di materiale finito e con frigorifero
6	57	9,4%	-1,58	Donne laureate o diplomate, in abitazioni con muri di materiale finito

Complessivamente, l'albero sembra spiegare una percentuale di variabilità che si aggira attorno all'81%.

Analogamente procediamo con la seconda dimensione.

Grafico 3.12. Dendrogramma delle coordinate sul secondo asse fattoriale



Il dendrogramma assume una forma molto complessa, e le variabili si alternano in modo confusionario. Ad ogni livello la ramificazione o viene interrotta o viene inserita un'unica variabile alla volta. Le variabili che compaiono sono, in ordine: alcolismo, decisioni sui grandi acquisti, decisioni sulla salute, violenza fisica, differenza di istruzione fra uomo e donna.

Da questa particolare suddivisione, è difficile etichettare il secondo fattore come "disagio sociale" o come "livello di indipendenza" come era accaduto finora in altre regioni. Probabilmente l'unico ordine che viene seguito lungo l'asse delle ascisse è quello relativo alla presenza o no di un uomo opprimente nella casa. Ma sicuramente l'unica informazione interessante è quella relativa alla prima suddivisione: per la prima volta la variabile "alcolismo" viene selezionata con un ruolo importante nell'analisi di segmentazione.

La decisione migliore è forse quella di riportare la descrizione dei singoli nodi terminali senza ulteriori commenti.

Tabella 3.18. Nodi terminali del dendrogramma per la seconda dimensione, con numerosità, media e descrizione del gruppo di donne

Nodo	N°	%	Media	Descrizione
2	61	10,1%	1,50	Donne con problemi di alcolismo
3	134	22,2%	0,57	Donne che bevono poco, e che prendono da sole decisioni sui grandi acquisti
8	161	26,7%	0,13	Donne che bevono poco, che non prendono da sole decisioni sui grandi acquisti e che subiscono violenze fra le mura domestiche
9	53	8,8%	-0,16	Donne che bevono poco, che non prendono da sole decisioni sui grandi acquisti, che non subiscono violenze e con titolo di studio maggiore a quello del partner
10	100	16,6%	-0,76	Donne che bevono poco, che non prendono da sole decisioni sui grandi acquisti, che non subiscono violenze e con titolo di studio uguale o inferiore a quello del partner
6	95	15,7%	-1,10	Donne che bevono poco, che non prendono da sole decisioni sui grandi acquisti e solo in compagnia quelle sulla salute

Complessivamente, l'albero spiega una percentuale di variabilità che si aggira attorno al 59%.

3.5 GIUNGLA ALTA: AMAZONAS

Per il Dipartimento di Amazonas considereremo solo quel 60,12% delle donne campionate che vivono nella zona geografica denominata “selva alta”, il che equivale a dire 484 donne sulle 805 campionate inizialmente: tale numero scende a 434 escludendo le interviste con valori mancanti. È il valore più basso fra quelli che si useranno in questo studio, ed è limitato per condurre un’analisi di segmentazione. Non essendo però interessati ad una classificazione per scopi previsivi, ma solo con l’obiettivo di descrivere le informazioni contenute nei dati, si è deciso di usarla comunque in relazione con l’analisi delle corrispondenze.

Differentemente da quanto accade per Loreto, l’analisi multipla delle corrispondenze evidenzia la presenza di due fattori principali che complessivamente spiegano il 22,64% dell’inerzia totale. In tabella è riportata la scomposizione corretta di Greenacre.

Tabella 3.19. scomposizione dell’inerzia corretta di Greenacre

Scomposizione dell’inerzia corretta di Greenacre				
Inerzia principale	Inerzia corretta	Percentuale	Percentuale cumulativa	4 8 12 16 20
0,19	0,17	17,47	17,47	*****
0,12	0,10	10,33	27,80	*****
0,10	0,07	7,66	35,45	*****
0,09	0,06	6,53	41,98	*****
0,08	0,06	6,27	48,25	*****
0,07	0,05	5,19	53,43	*****
0,07	0,05	4,97	58,40	*****
0,06	0,04	4,49	62,89	*****
0,06	0,04	3,92	66,81	*****
0,06	0,04	3,86	70,67	*****
0,05	0,03	3,45	74,12	****
0,05	0,03	3,30	77,43	****
0,05	0,03	3,24	80,66	****
0,05	0,03	2,87	83,53	****
0,04	0,03	2,77	86,30	***
0,04	0,02	2,45	88,75	***
0,04	0,02	2,16	90,90	***
0,03	0,02	1,76	92,66	**
0,03	0,02	1,72	94,38	**
0,03	0,01	1,50	95,89	**
0,03	0,01	1,41	97,30	**
0,03	0,01	1,19	98,49	*
0,02	0,01	1,01	99,50	*
0,02	0,00	0,50	100,00	*
Totale	0,95	100,00		

significa che gli estremi della variabile “titolo di studio” influenzano entrambi i fattori.

Infine desta interesse la netta unipolarità del secondo fattore, e lo scostamento nel primo quadrante delle modalità nella “punta” del triangolo che normalmente si collocano esattamente sul semiasse positivo.

Il semiasse negativo delle ordinate è contrassegnato da:

- nessun titolo di studio
- analfabetismo
- non conoscenza di dove comprare i preservativi
- no servizi igienici

Nel semiasse positivo invece sono presenti:

- titolo di studio alto
- si frigorifero
- muro con materiali finiti
- servizi igienici con acqua corrente
- donne con titolo di studio maggiore di quello del partner
- diploma di scuola secondaria

Per l'asse delle ascisse invece indicheremo solo quelle con coordinata positiva:

- capofamiglia donna
- presa di decisioni sui grandi acquisti
- alcolismo
- presa di decisioni sulla salute

Fra le modalità delle variabili illustrative, solo “urban” si posiziona significativamente sull'asse verticale (fra le donne con diploma secondario, senza pregiudizi, che conoscono le STD e che vivono in case almeno con l'elettricità), mentre solo le donne che vivono da sole (vedove e separate) in quello delle ordinate (esattamente nella punta del triangolo immaginario: donne capofamiglia in grado di prendere decisioni da sole sugli acquisti).

La lettura del grafico è supportata dalla tabella con le migliori modalità per importanza nella spiegazione dell'inerzia di ogni dimensione.

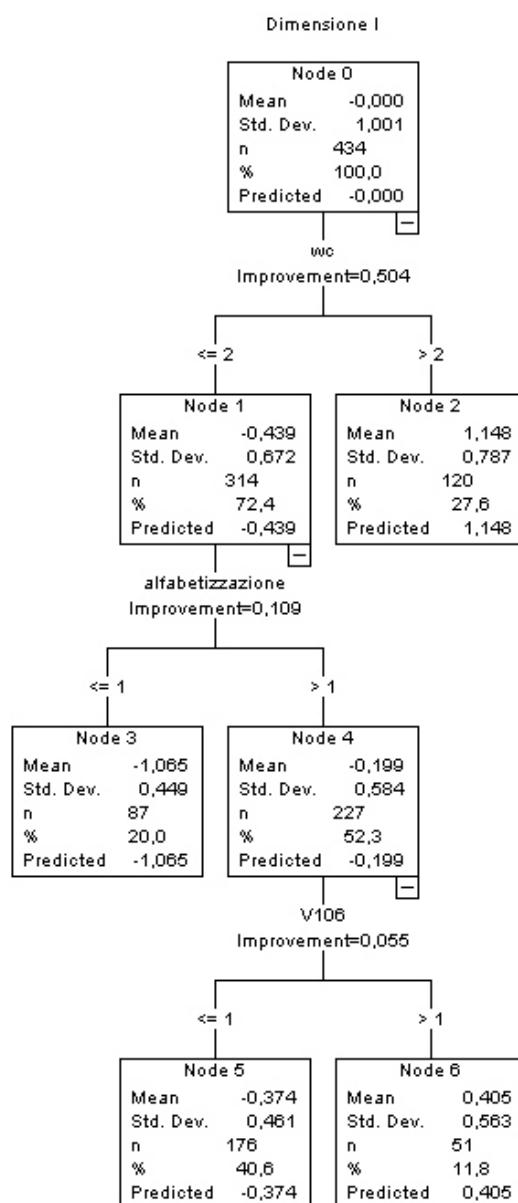
Tabella 3.20. Migliori modalità per contributo parziale all'inerzia delle due dimensioni

Dimensione 1		Dimensione 2	
WCacqua	0,112	capoF	0,162
siFrigo	0,080	DeciSI	0,158
siElettr	0,075	noRespect	0,083
analfa	0,067	aDeciSI	0,076
secondary	0,062	sposoViol	0,067
higher	0,060	siRespect	0,067
dovek	0,048	siFPpartn	0,046
noElettr	0,048	sposoNnViol	0,046
STDk	0,045	ubrSI	0,042
Wfin	0,037	aDeciNO	0,037
noPreg	0,036	noFPpartn	0,029
STDdk	0,034		
F+	0,034		
primary	0,031		
noEduc	0,030		
WCno	0,026		

Le prime modalità per la prima dimensione riguardano tutte variabili relative all'abitazione, e per la prima volta è più caratterizzante possedere l'elettricità piuttosto che non poterne usufruire. Per la seconda dimensione, oltre alle modalità presenti sulla punta del "triangolo", compaiono la violenza, sia verbale che fisica, e la pianificazione familiare.

Il processo di segmentazione crea due alberi poco profondi, inserendo solo 3 variabili esplicative per ogni dimensione.

Grafico 3.14. Dendrogramma delle coordinate sul primo asse fattoriale



La variabile “servizi igienici” viene preferita dal processo di segmentazione all’istruzione. Le donne meno disagiate (sul semiasse positivo) si possono individuare o dalla presenza di un bagno con acqua, o dal titolo di studio superiore a quello primario. Viceversa, le donne in condizioni di povertà sono caratterizzate da un basso livello di istruzione.

Tabella 3.21. Nodi terminali del dendrogramma per la prima dimensione, con numerosità, media e descrizione del gruppo di donne

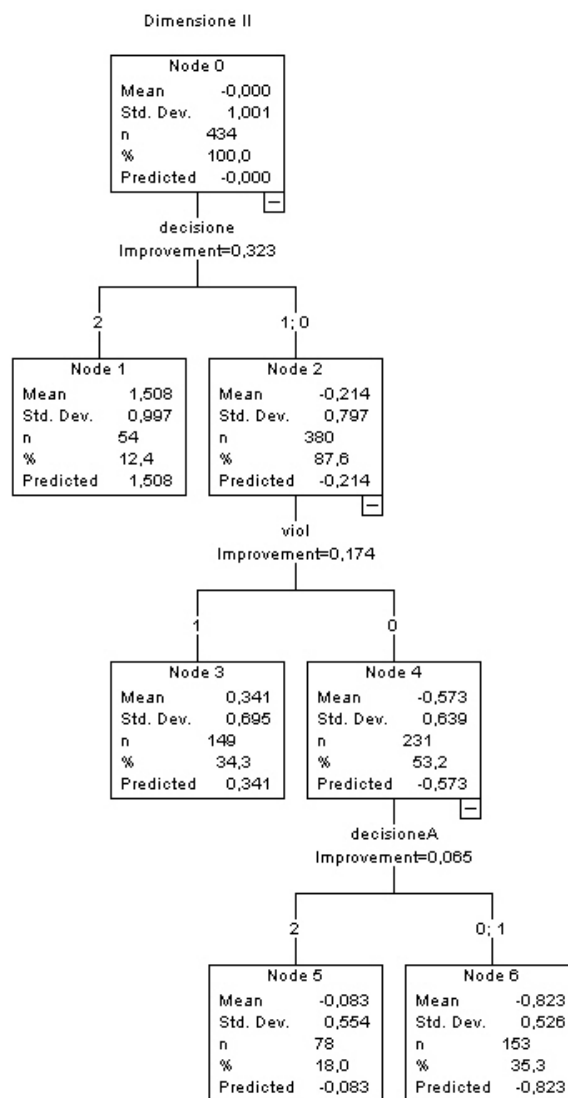
Nodo	N°	%	Media	Descrizione
------	----	---	-------	-------------

2	120	27,6%	1,15	Donne in abitazioni con servizi igienici con acqua
6	51	11,8%	0,40	Donne in abitazioni senza servizi igienici, con titolo di studio almeno secondario
5	176	40,6%	-0,38	Donne in abitazioni senza servizi igienici, senza titolo di studio o con licenza elementare
3	87	20,0%	-1,06	Donne analfabete in abitazioni senza servizi igienici

Complessivamente, l'albero spiega il 67% della devianza.

L'analisi di segmentazione seleziona solo tre variabili anche per la seconda dimensione.

Grafico 3.15. Dendrogramma delle coordinate sul secondo asse fattoriale



La variabile “decisione sui grandi acquisti” domina l’analisi, separando subito le unità con coordinate positive (quelle sulla “punta del triangolo”) dal resto dei

dati. Le altre ramificazioni hanno il difficile compito di dividere in gruppi omogenei tutti i dati concentrati attorno al baricentro degli assi cartesiani. Inaspettatamente, viene selezionata prima la variabile “violenza”, che nell’analisi delle corrispondenze non aveva suscitato particolare interesse. Successivamente, le donne che non subiscono violenze fra le mura domestiche, vengono ripartite fra quelle che prendono decisioni da sole sulla salute e quelle che o non decidono o lo fanno solo se con qualche altra persona.

Come in altre analisi, lascia riflettere la presenza nel semiasse positivo sia delle donne che prendono da sole le decisioni sia di quelle il cui sposo è protagonista di atti di violenza; le valutazioni quindi di natura “sociale” si uniscono ad un livello di indipendenza che il più delle volte è causato dal vivere senza un partner.

Tabella 3.22. Nodi terminali del dendrogramma per la seconda dimensione, con numerosità, media e descrizione del gruppo di donne

Nodo	N°	%	Media	Descrizione
1	54	12,4%	1,51	Donne che prendono da sole le decisioni sui grandi acquisti
3	149	34,3%	0,34	Donne che non prendono da sole le decisioni sui grandi acquisti e che subiscono violenze domestiche
5	78	18,0%	-0,08	Donne che non prendono da sole le decisioni sui grandi acquisti, che non subiscono violenze domestiche e che decidono da sole sulla salute
6	153	35,3%	-0,82	Donne che non prendono da sole le decisioni sui grandi acquisti, che non subiscono violenze domestiche e che non decidono sulla salute

Complessivamente, l’albero spiega una percentuale di variabilità che si aggira attorno al 56%.

3.6 ALTIPIANO ANDINO: APURIMAC

Apurimac, come molti altri Dipartimenti montani, è caratterizzato da un ISU molto basso, dovuto principalmente agli indicatori sulle condizioni di salute (speranza di vita alla nascita), sul reddito e sull’alta percentuale di analfabetismo (circa il 30%); viceversa, il tasso di scolarizzazione è superiore

quantomeno a quello delle regioni amazzoniche. Apurimac inoltre ha la percentuale più alta di povertà, fra i 6 Dipartimenti analizzati.

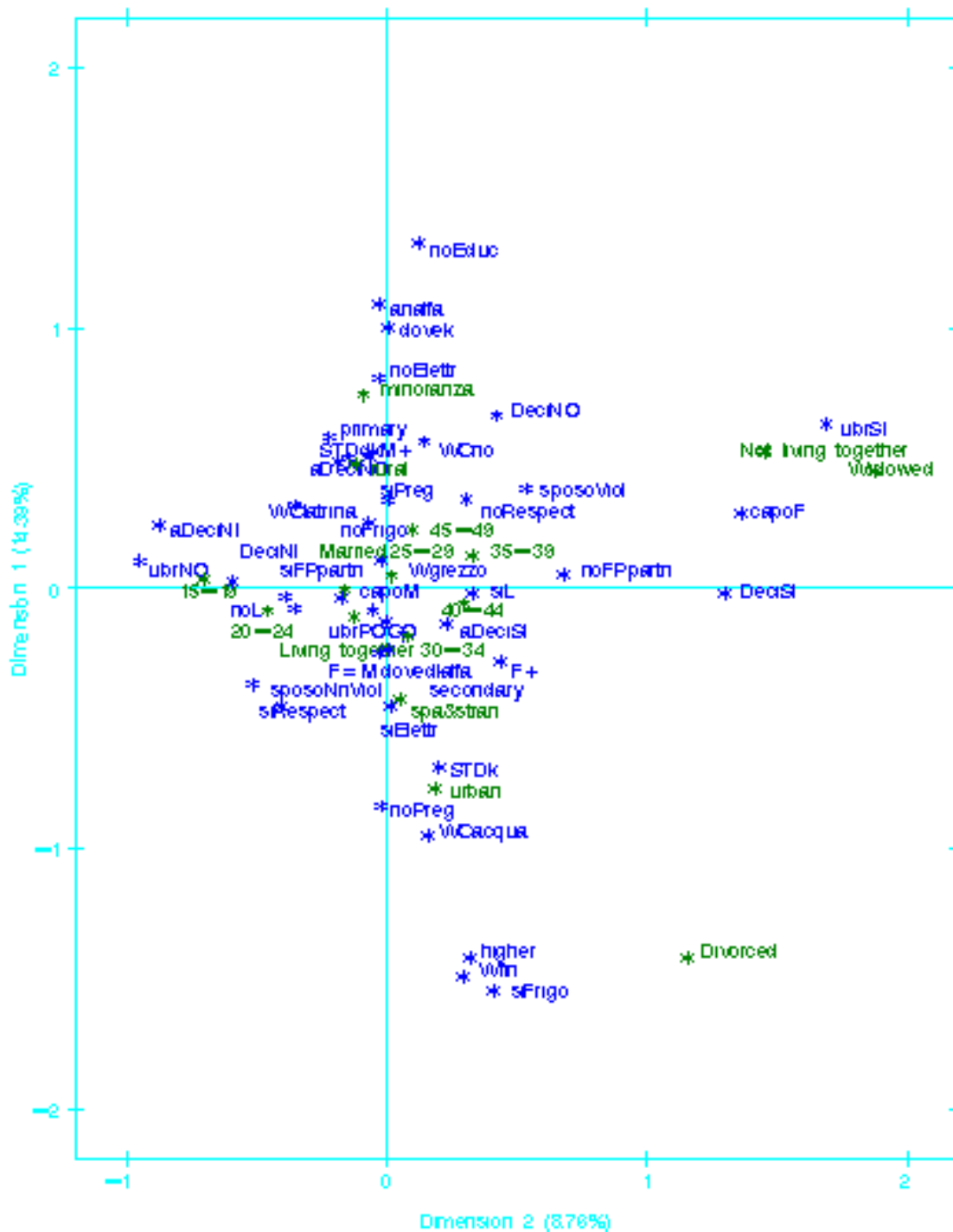
Con l'analisi multipla delle corrispondenze si ricavano due fattori principali, che congiuntamente arrivano a spiegare circa il 23% della devianza, secondo la scomposizione del Chi quadrato. Tale percentuale sale al 28,5% con la correzione di Greenacre, come descritto in tabella.

Tabella 3.23. Scomposizione dell'inerzia corretta di Greenacre

Scomposizione dell'inerzia corretta di Greenacre				
Inerzia principale	Inerzia corretta	Percentuale	Percentuale cumulativa	4 8 12 16 20
0,20	0,17	18,26	18,26	*****
0,12	0,10	10,24	28,50	*****
0,09	0,07	7,47	35,97	*****
0,08	0,06	5,88	41,85	*****
0,07	0,05	5,50	47,36	*****
0,07	0,05	4,98	52,33	*****
0,06	0,04	4,69	57,02	*****
0,06	0,04	4,51	61,53	*****
0,06	0,04	3,95	65,48	****
0,05	0,03	3,58	69,06	****
0,05	0,03	3,47	72,53	****
0,05	0,03	3,24	75,77	****
0,05	0,03	3,01	78,79	****
0,04	0,03	2,91	81,70	****
0,04	0,02	2,51	84,20	***
0,04	0,02	2,29	86,50	***
0,04	0,02	2,21	88,71	***
0,04	0,02	2,09	90,80	***
0,03	0,02	1,92	92,72	**
0,03	0,02	1,74	94,46	**
0,03	0,01	1,48	95,94	**
0,03	0,01	1,45	97,39	**
0,02	0,01	1,15	98,54	*
0,02	0,01	0,99	99,52	*
0,01	0,00	0,48	100,00	*
Totale	0,94	100,00		

Nell'analisi non verranno considerate età al matrimonio, opinione sulla violenza, metodi contraccettivi e numero di figli.

Grafico 3.16. Rappresentazione grafica delle modalità rispetto ai primi due assi fattoriali trovati con l'analisi delle corrispondenze per il Dipartimento di Apurimac



L'asse delle ordinate presenta agli estremi sempre due modalità del titolo di studio, mentre per la prima volta l'asse orizzontale è caratterizzato dalla variabile alcolismo sia all'estremo positivo che a quello negativo.

Nello specifico, il semiasse positivo delle ordinate è contrassegnato da:

- nessun titolo di studio

- analfabetismo
- non conoscenza di dove comprare i preservativi
- no elettricità

Nel semiasse negativo invece sono presenti:

- si frigorifero
- muro con materiali finiti
- titolo di studio alto
- servizi igienici con acqua corrente
- no pregiudizi verso i malati di AIDS
- conoscenza delle STD

Il semiasse delle ascisse è contraddistinto nella sua parte positiva da:

- alcolismo
- capofamiglia donna
- presa di decisioni sui grandi acquisti

mentre nel lato negativo da:

- nessun problema di alcolismo
- decisioni sulla salute con altre persone
- decisioni sugli acquisti con altre persone

Le variabili supplementari (età, stato civile, gruppo etnico ed urbanizzazione) hanno le seguenti caratteristiche:

- le minoranze etniche vivono in condizioni precarie, senza elettricità o servizi igienici, e gli uomini hanno un titolo di studio superiore rispetto a quello delle donne, che mediamente si ferma al massimo alla licenza elementare;
- hanno meno incertezze economiche le donne che dimorano in città;
- le donne divorziate, pur collocandosi, come in altri casi precedentemente analizzati, isolate nel quarto quadrante, sono molto più vicine all'asse verticale;
- le donne che vivono da sole sono strettamente legate ad un eccessivo consumo di alcolici;

- per la prima volta in questo studio, le donne sposate hanno coordinata y positiva, benché molto prossima allo zero, mentre sono le conviventi a collocarsi fra le donne più agiate;
- le ragazze più giovani (19-24 anni) non hanno problemi di alcolismo, ma nello stesso tempo non lavorano ed hanno difficoltà a prendere decisioni da sole.

La lettura del grafico è supportata dalla tabella con le migliori modalità per importanza nella spiegazione dell'inerzia di ogni dimensione.

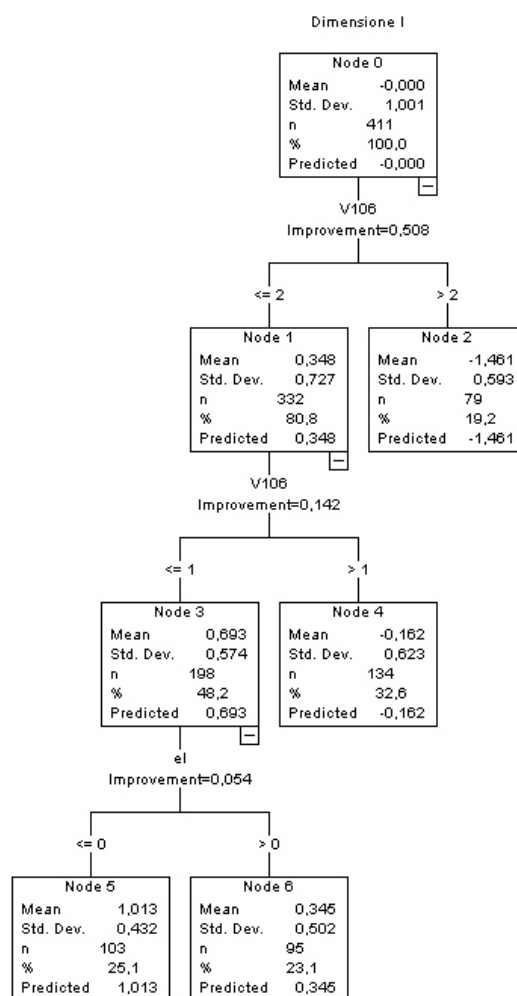
Tabella 3.24. Migliori modalità per contributo parziale all'inerzia delle due dimensioni

Dimensione 1		Dimensione 2	
higher	0,107	DeciSI	0,131
siFrigo	0,094	ubrSI	0,114
WCacqua	0,081	capoF	0,095
noElettr	0,065	noFPpartn	0,077
analfa	0,058	sposoViol	0,064
noPreg	0,056	sposoNnViol	0,061
STDk	0,056	aDeciNI	0,058
Wfin	0,042	noL	0,057
STDdk	0,040	ubrNO	0,046
primary	0,039	siFPpartn	0,044
siElettr	0,036	DeciNI	0,042
WCno	0,033	siRespect	0,032
noEduc	0,032		
dovek	0,031		
siRespect	0,024		
M+	0,024		

Fra le variabili sull'istruzione, l'analfabetismo assume un ruolo secondario rispetto al titolo di studio. Le modalità che meglio spiegano l'inerzia della seconda dimensione sono le stesse collocate all'estremo positivo dell'asse orizzontale.

Il processo di segmentazione crea un primo albero poco profondo, essendo sufficiente l'inserimento di sole due variabili per bloccare l'algoritmo.

Grafico 3.17. Dendrogramma delle coordinate sul primo asse fattoriale



Per individuare le donne sul semiasse negativo, è sufficiente isolare al primo passo quelle col titolo di studio più alto, e successivamente quelle con la licenza secondaria: le restanti, con un livello d'istruzione inferiore a questo, hanno coordinata y positiva, ed è sufficiente solo un'ulteriore suddivisione fra chi ha la possibilità di utilizzare la corrente elettrica e chi invece è costretta farne a meno.

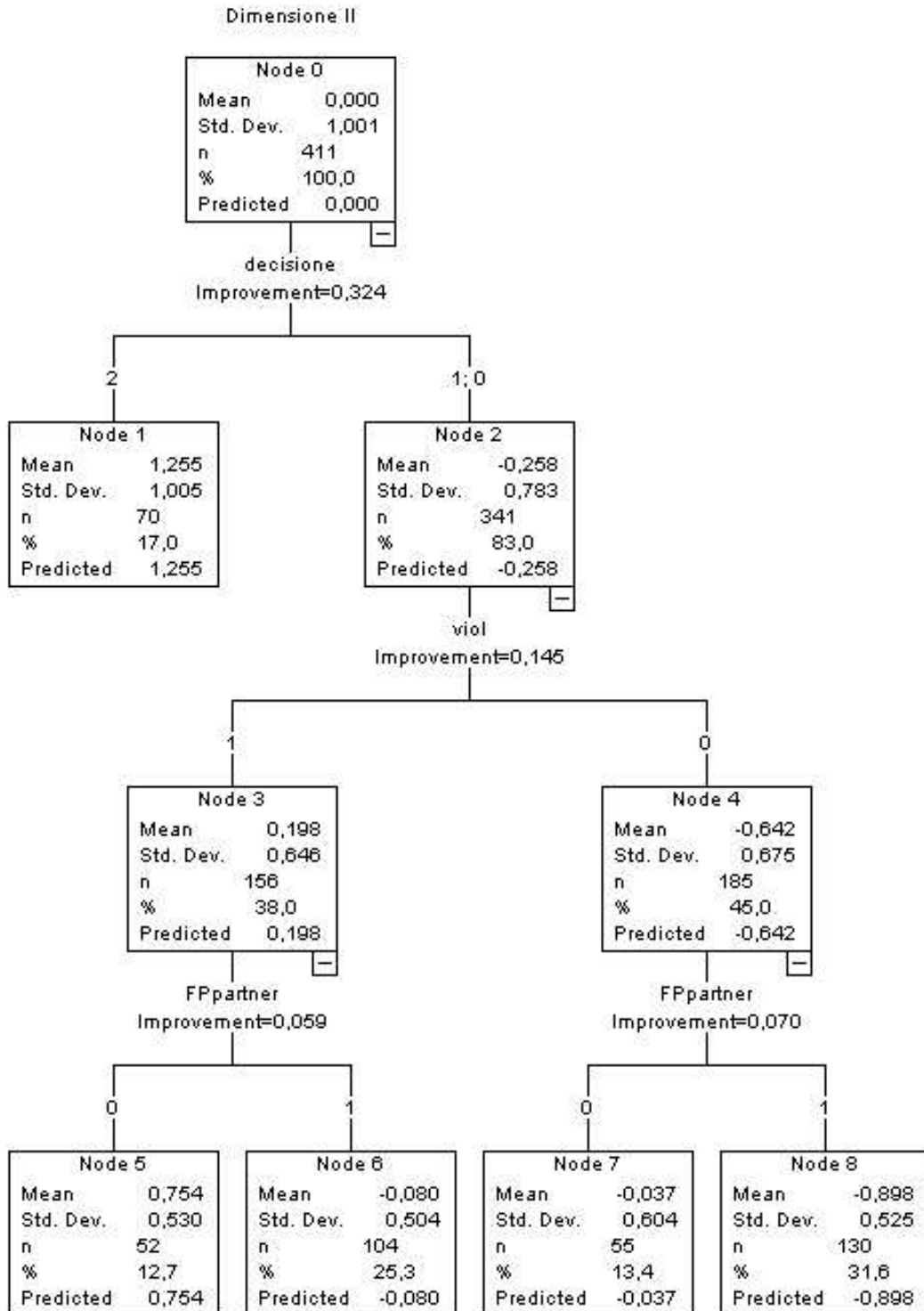
Tabella 3.25. Nodi terminali del dendrogramma per la prima dimensione, con numerosità, media e descrizione del gruppo di donne

Nodo	N°	%	Media	Descrizione
5	103	25,1%	1,01	Donne con basso titolo di studio, senza elettricità
6	95	23,1%	0,34	Donne con basso titolo di studio, con elettricità
3	134	32,6%	-0,16	Donne con licenza secondaria
1	79	19,2%	-1,46	Donne con laurea o diploma

L'albera spiega all'incirca il 70% della devianza complessiva.

Il dendrogramma con la seconda dimensione come variabile dipendente è più complesso, benché selezioni comunque poche variabili predittive.

Grafico 3.18. Dendrogramma delle coordinate sul secondo asse fattoriale



La variabile “decisione sui grandi acquisti” è la prima ad essere selezionata, arrivando a spiegare il 32,4% della devianza: le 70 donne che sono in grado di decidere da sole formano già un nodo terminale, mentre le restanti vengono ulteriormente suddivise fra chi subisce e chi non subisce le violenze da parte del partner, infine, entrambi i due sottogruppi vengono ripartiti in base alla variabile “pianificazione familiare col partner”.

Quello che si nota subito è che i nodi terminali “6” e “7” si scambiano inaspettatamente di ordine, ma che in realtà hanno media molto simile: probabilmente le donne non soggette a violenze fisiche ma che non hanno la possibilità di pianificare la loro vita percepiscono un disagio molto simile a chi ha magari la possibilità di fare pianificazione familiare, ma ha uno sposo violento nei suoi confronti; se quest’interpretazione fosse realistica, allora sarebbe opportuno far convergere i rami in un unico nodo terminale.

Questa dipendenza dallo sposo, può far pensare che il dialogo col partner (o semplicemente la possibilità di poterlo fare) sia la chiave di lettura di questa dimensione. Per un approfondimento sulla nomenclatura di quest’asse fattoriale, rimandiamo però al prossimo paragrafo conclusivo, dove verrà effettuato un confronto fra le 6 regioni.

Tabella 3.26. Nodi terminali del dendrogramma per la seconda dimensione, con numerosità, media e descrizione del gruppo di donne

Nodo	N°	%	Media	Descrizione
1	70	17,0%	1,25	Donne che prendono da sole le decisioni sui grandi acquisti
5	52	12,7%	0,75	Donne che non decidono da sole, che sono soggette a violenza fisica e che non possono prender parte alla pianificazione familiare
7	55	13,4%	-0,03	Donne che non decidono da sole, che sono soggette a violenza fisica e che prendono parte alla pianificazione familiare
6	104	25,3%	-0,08	Donne che non decidono da sole, che non sono soggette a violenza fisica ma che non possono prender parte alla pianificazione familiare
8	130	31,6%	-0,90	Donne che non decidono da sole, ma che non sono soggette a violenza fisica e possono prender parte alla pianificazione familiare

Complessivamente, l’albero spiega una percentuale di variabilità che si aggira attorno al 60%.

3.7 CONCLUSIONI

Il confronto dei 6 Dipartimenti ha messo in luce la presenza di alcune variabili utili su tutto il territorio nazionale per cercare di interpretare le situazioni di maggior povertà; ma soprattutto quest'analisi ci ha permesso di scoprire quanto varino le caratteristiche e le esigenze delle donne a seconda dell'area geografica.

L'analisi multipla delle corrispondenze ha evidenziato la presenza di alcune dinamiche interne ai dati molto diverse da regione a regione. In alcuni dipartimenti la variabilità della matrice di dati può essere ben spiegata da due fattori, uno che delinea il profilo delle intervistate in base a delle evidenze economiche-materiali, l'altro che cerca di coglierne il grado di indipendenza e il disagio più socio-psicologico. In altri dipartimenti invece il livello di povertà può essere così totalizzante per alcune intervistate, che per spiegare la variabilità tra le donne è sufficiente un unico fattore che accomuna assieme molteplici aspetti negativi di diversa natura; la causa di questa situazione di precarietà è ovviamente dovuta al contesto circostante: passiamo da un'economia di sussistenza nella foresta amazzonica, dove solo il poter vivere in un centro urbano viene percepito come una realtà di ricchezza, a una realtà sovrastante quale è una metropoli, dove le aspettative sono tante ma solo il non aver un titolo di studio può chiudere molte porte e creare frustrazioni interiori.

L'analisi di segmentazione, a supporto di quella delle corrispondenze, permette di fare una buona sintesi fra tutte le variabili in gioco, per vedere quali riescono a differenziare meglio i gruppi di donne in base alle due dimensioni; ad ogni passo viene selezionata una variabile esplicativa alla volta che divide il campione in due sottogruppi il più omogenei possibile al loro interno. Questo procedimento permette quindi di confermare o confutare le ipotesi sui fattori formulate durante l'analisi delle corrispondenze, come ad esempio è avvenuto nel nostro studio per più di una regione rispetto alla seconda dimensione.

Tornando alle differenze fra i le regioni geografiche, nella seguente tabella sono riassunte, in ordine di comparsa, le variabili esplicative che l'algoritmo di segmentazione ha selezionato per la **prima dimensione** (le prime in ogni colonna sono le prime a comparire e quindi, con la successiva ramificazione, sono le prime che spiegano le percentuali maggiori di devianza; in grigio le

variabili legate all'istruzione, sottolineate quelle che fanno riferimento all'abitazione).

Tabella 3.27. Variabili esplicative selezionate nell'analisi di segmentazione sulla prima dimensione, per regione geografica

Costa Nord	Costa Sud	Metropoli	Giungla Bassa	Giungla Alta	Altipiano
titolo di studio alfabetizzazione <u>muro</u> <u>frigo</u>	titolo di studio <u>frigo</u> conoscenza STD	titolo di studio alfabetizzazione <u>frigo</u> violenza fisica <u>servizi igienici</u> conoscenza STD	<u>muro</u> titolo di studio <u>elettricità</u> <u>frigo</u> alfabetizzazione	<u>servizi igienici</u> alfabetizzazione titolo di studio	titolo di studio <u>elettricità</u>

Dalle analisi emerge che in tutti i Dipartimenti costieri il titolo di studio è la prima variabile discriminante, seguita spesso dall'alfabetizzazione e dal possesso di elettrodomestici; il materiale di costruzione delle abitazione viene selezionato solo a Tumbes, regione molto più agricola rispetto ad Ica, mentre la conoscenza delle STD nelle due regioni costiere più centrali.

Si è osservato poi che le donne della metropoli sono le uniche a mescolare le due dimensioni, quella economica e quella sociale, in un unico fattore: l'albero di segmentazione è infatti quello più ramificato, e compaiono anche i maltrattamenti fisici subiti dal partner.

L'istruzione invece assume un ruolo secondario (ed è sempre accompagnata dalla variabile alfabetizzazione) nelle regioni amazzoniche, dove il materiale di costruzione delle case o la presenza di servizi igienici identificano meglio le donne benestanti da quelle più disagiate. In particolare poi, nella giungla bassa, compaiono anche tutte le altre variabili legate all'abitazione, quali elettricità e frigorifero, quasi a sottolineare questo rapporto più materiale fra ricchezza e possesso.

Nell'altipiano, dove la scolarizzazione è più alta rispetto alla giungla, formando quindi un più elevato divario con l'alta percentuale di donne analfabete, il titolo di studio torna ad essere la prima variabili previsiva della segmentazione, non accompagnato da "alfabetizzazione". Il livello di istruzione è accompagnato unicamente dall'elettricità, risaltando quindi questa disparità a cui si accennava.

L'analisi della **seconda dimensione** risulta più complicata; questo è dovuto sia all'alto numero di variabili coinvolte, sia al fatto che in alcune regioni essa spiega una percentuale di inerzia così ridotta da essere quasi inglobata nel primo fattore (Lima e Loreto), sia perché le variabili che la determinano variano da una regione all'altra.

Durante lo studio di ogni regione geografica, infatti, il nome del secondo asse fattoriale si è evoluto assumendo varie forme in base alle variabili coinvolte: all'inizio era "disagio socio-psicologico", successivamente si è passati per "livello di indipendenza della donna", fino a "dialogo col partner" nell'ultima regione (Apurimac).

I dendrogrammi sono sicuramente di difficile interpretazione e molto diversi fra loro per le variabili selezionate, ma esiste una struttura di base analoga per ogni regione: è infatti presente quasi ovunque una certa simmetria unilaterale sinistra (cioè dove le coordinate sono negative). Per avere un'idea più chiara di cosa si intende, facciamo riferimento all'ultimo albero di segmentazione di Apurimac, più semplice rispetto agli altri, ma che alla fine presenta una struttura molto simile ad essi. Quest'albero chiarifica bene alcune dinamiche:

- il secondo fattore da una parte isola le donne sole (coordinate positive), ma quest'indipendenza sviluppata da vedove e separate (le divorziate spesso si collocano in un'altra parte del grafico), non è percepita come un punto di forza (come si ipotizzava all'inizio dell'analisi), ma è spesso legata a problemi di alcolismo;
- agli estremi del semiasse negativo invece si collocano sempre le donne che hanno la possibilità di prendere decisioni solo se assieme a qualche altra persona (partner o familiari), ma che comunque sono forti della presenza di uomo al loro fianco, spesso non violento (maltrattamenti fisici o verbali) e che lascia loro la possibilità di pianificare assieme la vita familiare: questo viene percepito come una situazione non disagiata, benché a volte sia anche legata al fatto che la donna non lavori o che l'abitazione non possieda tutti i comfort desiderabili, ma che probabilmente è anche quello che il contesto circostante prevede come "normale";

- al centro del grafico invece ci sono le violenze, che timidamente fanno slittare le coordinate delle unità verso il lato positivo, benché esso sia già occupato dalle donne sole.

Il secondo asse fattoriale quindi rappresenta due realtà: l'indipendenza nelle decisioni e il rapporto col partner; in alcune aree geografiche queste due forme di disagio si fondono in un'unica dimensione, in altre invece i due aspetti non convivono bene assieme. Probabilmente in queste regioni un approfondimento potrebbe andare nella direzione di separare le donne che attualmente vivono senza un partner (vedove, conviventi e divorziate) dalle rimanenti.

La tabella riassuntiva che viene presentata, con le variabili che compaiono negli alberi di segmentazione, ci permette di confermare quanto appena detto, e ci aiuta ad entrare, per quanto possibile, nel confronto fra regioni morfologiche (le prime in ogni colonna sono le prime a comparire e che quindi, con la successiva ramificazione, spiegano le percentuali maggiori di devianza; in grigio le due variabili sulla presa di decisione, sottolineate quelle che fanno riferimento alla violenza subita dalle donne, sia verbale che fisica).

Tabella 3.28. Variabili esplicative selezionate nell'analisi di segmentazione sulla seconda dimensione, per regione geografica

Costa Nord	Costa Sud	Metropoli	Giungla Bassa	Giungla Alta	Altipiano
dec. acquisti	dec. acquisti	<u>violenza fisica</u>	alcolismo	dec. acquisti	dec. acquisti
FP partner	FP partner	alcolismo	dec. acquisti	<u>violenza fisica</u>	<u>violenza fisica</u>
secco cf	<u>violenza verbale</u>	dec. salute	dec. salute	dec. salute	FP partner
lavoro	dec. salute	dec. acquisti	<u>violenza fisica</u>		
figli		secco cf	istruzione MF		
dec. salute		<u>violenza verbale</u>			

A causa della complessità del secondo fattore, come appena detto, sarà inutile analizzare troppo nel dettaglio la tabella. Ma già osservando la variabile che determina la prima suddivisione per ogni regione morfologica si ricavano informazioni interessanti: la possibilità di prendere decisioni sui grandi acquisti in modo indipendente (che abbiamo visto essere relazionata al sesso del capofamiglia) è utilizzata dal programma al primo passo della ramificazione in quasi tutte le regioni, ad eccezione di Lima e di Loreto, i due Dipartimenti dove la seconda dimensione è molto debole e spiega una bassa percentuale di

inerzia. Nella foresta amazzonica, comunque, tale variabile viene sostituita dal consumo di alcolici da parte della donna, che abbiamo sottolineato in precedenza essere la sfaccettatura negativa dell'indipendenza dell'intervistata. Lima invece ha un andamento diverso, e la prima variabile selezionata è la violenza fisica del partner (inoltre è l'unica regione dove anche la violenza verbale assume un'importanza nodale).

Rispetto alla violenza sulle donne, Tumbes è l'unico Dipartimento dove essa non è fondamentale per il raggruppamento dei dati; viceversa, esso è l'unico dove compaiono il lavoro della donna e il numero dei figli, fatto spiegato dall'elevata percentuale di persone residenti in zone rurali.

Infine si nota che nella giungla bassa compare una variabile legata all'istruzione, benché sia quella più "sociale" delle tre utilizzate nello studio.

In generale quindi riassumiamo quanto emerso nel capitolo nei seguenti 7 punti schematici:

- I. la regione morfologica assume un'importanza strategica nell'individuazione delle tipologie di disagi percepiti dalle donne; in particolare, le differenze maggiori si hanno fra i Dipartimenti situati sulla costa e quelli ad Est della Cordigliera Andina
- II. la realtà metropolitana della capitale segue una dinamica non associabile ad alcun'altra area geografica
- III. all'interno delle regioni geografiche, le differenze maggiori sono fra le persone che dimorano in zone urbanizzate e i *campesinos*
- IV. una parte dei dati sono spiegati bene da un primo fattore denominato "disagio economico" o "disagio materiale": esso è identificabile dal titolo di studio nella Costa, dall'alfabetizzazione e dalle condizioni dell'abitazione nella *sierra* e nella *selva*
- V. nella *sierra* il divario di ricchezza della popolazione probabilmente è maggiore, visto che vengono selezionate modalità agli antipodi, quali il titolo di studio elevato da una parte, ed il non possesso di elettricità dall'altro
- VI. in alcune regioni esiste un secondo fattore, per quanto più debole rispetto al primo: esso riguarda sia il livello di indipendenza della donna, sia il rapporto col partner (dialogo e violenza subita);

probabilmente sarebbe migliorabile se si isolassero le donne che vivono da sole dal resto del campione, quali le conviventi, le vedove e le divorziate

- VII. la violenza fisica gioca un ruolo fondamentale nella metropoli, mentre l'alcolismo si accentua maggiormente della giungla
- VIII. le modalità della conoscenza di dove comprare i preservativi vengono citate spesso nelle nomenclature del primo asse fattoriale (in particolare la non conoscenza), ma la variabile non è mai così discriminante da essere selezionata dall'algoritmo di segmentazione
- IX. non vengono mai prese in considerazione durante le analisi, in quanto poco utili per spiegare i fattori latenti nei dati, le variabili sull'uso di metodi contraccettivi moderni e sull'opinione riguardo la violenza sui figli come metodo educativo

4. Cusco

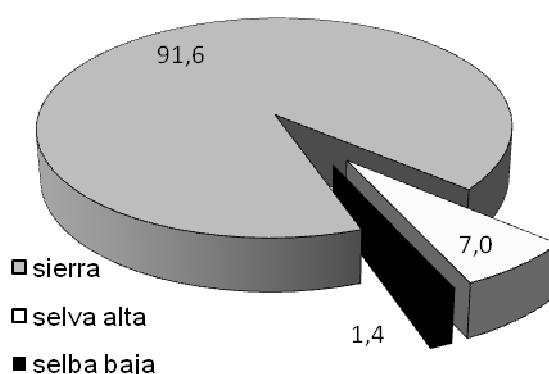
4.1 IL DISAGIO E LE MANCATE RISPOSTE

Si vuole concludere lo studio con un approfondimento sul Dipartimento di Cusco, in quanto sede principale dello stage presso la ong MIDE.

Nel Capitolo 1 erano già stati riportati alcuni indicatori che caratterizzano la regione. Ricordiamo solamente che l'indice di Sviluppo Umano di Cusco, nel 2003 pari a 0,511, è uno dei più bassi del Perù; il Dipartimento inoltre è inserito fra le 5 regioni con più alta percentuale di abitanti che vivono sotto la "linea di povertà estrema".

Cusco è una zona che desta particolare interesse perché il suo territorio attraversa le tre zone geografiche più orientali (benché con percentuali decisamente disomogenee, come riportato dal grafico), e quindi i programmi di microcredito, o di sviluppo in generale, devono essere pensati diversamente a causa delle differenti esigenze delle donne. Gli ISU più bassi si incontrano nelle province "alte" e nei distretti amazzonici della regione La Convención.

Grafico 4.1. Percentuale di abitanti per zona geografica

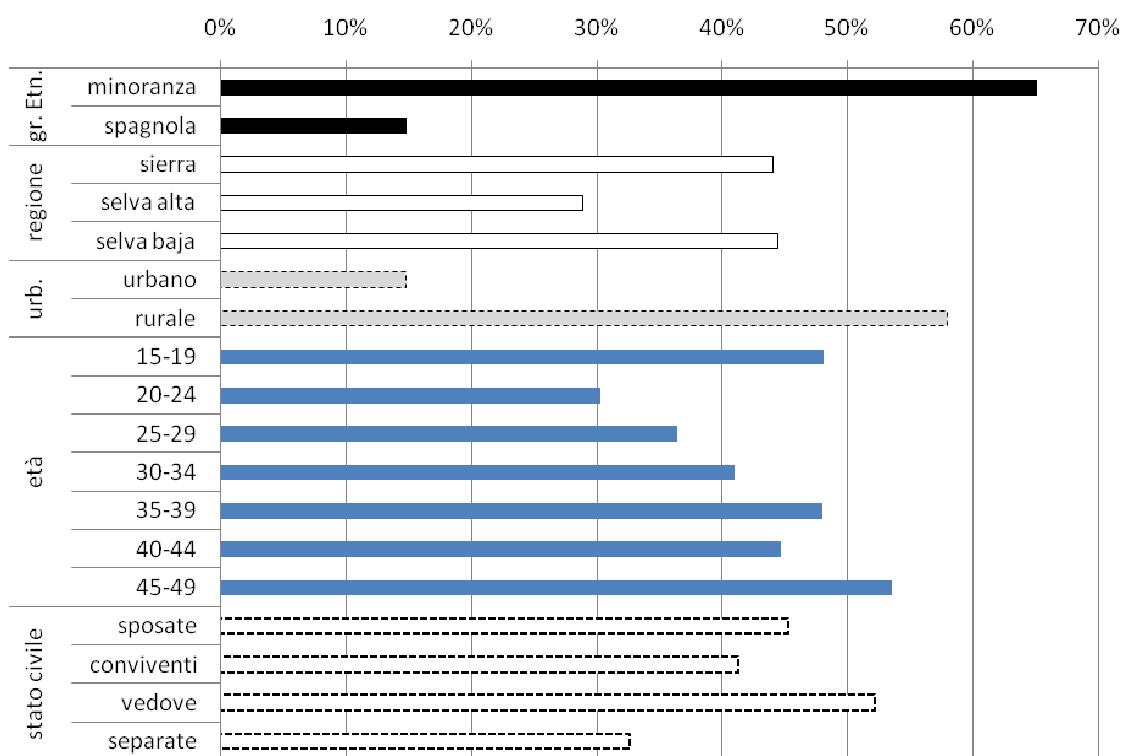


Rispetto al dataset DHS, Cusco vanta un altro primato negativo: insieme ad Apurimac ed a pochi altri dipartimenti posizionati sull'altipiano andino, ha la più alta percentuale di valori mancanti nelle risposte del questionario; le cause sono dovute sia all'imbarazzo provocato da alcune singole domande, sia alla "mancanza di privacy" che non ha permesso di rispondere agli argomenti più delicati, quali ad esempio la violenza domestica. Per questo motivo, le 642

donne campionate (escluse come sempre le “mai coniugate” e le “non residenti”) si riducono drasticamente a 366 quando andiamo ad analizzare i dati senza i *missing values*.

Le non rispondenti ovviamente non sono distribuite in modo casuale: alcune modalità delle variabili selezionate per lo studio subiscono una riduzione di frequenza molto maggiore rispetto ad altre. L’esempio più significativo è quello della variabile collegata alla domanda “Conosci l’AIDS?”, dove la risposta negativa scompare completamente nell’analisi senza i *missing*.

Grafico 4.2. Percentuale di riduzione dei dati per le modalità delle variabili “illustrative” con l’eliminazione dei *missing*



Rispetto alle cinque variabili che verranno utilizzate nell’analisi multipla delle corrispondenze come illustrative, il grafico risalta la disparità di riduzione delle frequenze fra modalità, dopo l’eliminazione dei *missing*.

Il confronto fra le minoranze etniche e gli abitanti di origine spagnola è quello più significativo: il campione iniziale, che prevedeva circa il 56% delle intervistate appartenenti a qualche minoranza etnica (quechua, aymara, etc.), si riduce in modo asimmetrico (riduzione del 65%), tanto che alla fine nelle analisi

siamo costretti ad utilizzare un dataset con una percentuale di appartenenti a gruppi etnici poco superiore al 34%.

Divari tanto profondi si creano anche per la variabile “urbanizzazione”, e per alcune modalità delle altre variabili illustrative. È evidente quindi che le non rispondenti alle domande “sensibili” hanno caratteristiche ben definite, che le riconducono alla fascia più disagiata della popolazione.

In tabella sono presenti le riduzioni in percentuali anche per le variabili che sono state finora utilizzate come “attive” nelle analisi.

Tabella 4.1 Percentuale di riduzione dei dati per le modalità delle variabili “attive” con l’eliminazione dei *missing*

 sesso capo famiglia 		 elettricità 		 FP col partner 	
M	43%	no	69%	no	45%
F	44%	si	27%	si	38%
 età al matrimonio 		 frigorifero 		 decisione acquisti 	
<15	68%	no	47%	no	56%
15-17	43%	si	13%	assieme	40%
18-20	44%	 muro 		da sola	36%
21-30	34%	grezzo	44%	 decisione salute 	
>30	45%	finito	19%	no	58%
 educazione MF 		 servizi igienici 		assieme	47%
F+	16%	no wc	56%	da sola	32%
M=F	38%	latrina	57%	 dove preservativi 	
M+	53%	wc con acqua	13%	si	23%
 titolo di studio 		 violenza domestica 		no	70%
nessuno	83%	no	49%	 consumo alcolici 	
primario	52%	si	36%	no	40%
secondario	10%	 opinione violenza 		si	45%
superiore	4%	no violenza	42%	poco	41%
 alfabetizzazione 		si violenza	41%	 metodi contraccettivi 	
no	74%	 rispetto dal partner 		no	59%
si	21%	no	40%	si	35%
 n° di figli 		si	44%	 pregiudizio 	
0	36%	 conoscenza AIDS 		si	4%
1	44%	no	100%	no	9%
2	47%	si	6%	 lavorando 	
3+	57%	 conoscenza STD 		no	26%
		no	55%	si	46%
		si	10%		

Queste sono le caratteristiche delle intervistate che trovano difficoltà nel rispondere alle domande più delicate:

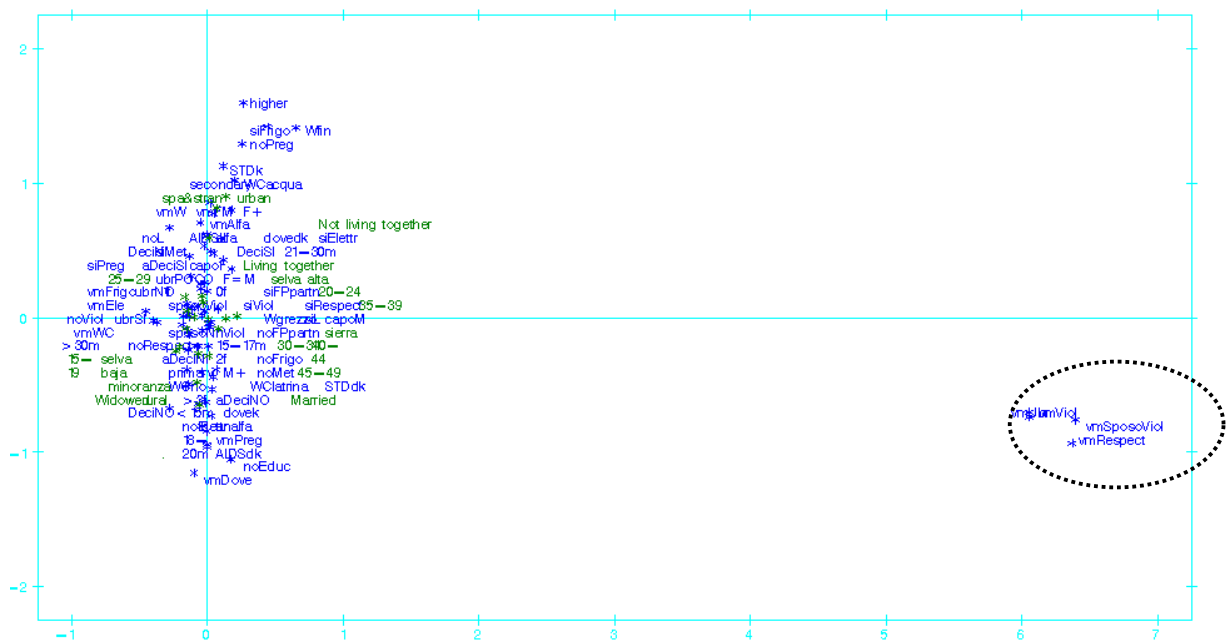
- Età al primo matrimonio inferiore ai 15
- Analfabete o senza titolo di studio, o per lo meno inferiore a quello del partner

- Elevato numero di figli
- Abitazioni prive di alcuni servizi o elettrodomestici primari
- Ignoranza sulle malattie sessualmente trasmissibili
- Partner che non permette loro di prendere decisioni né di programmare la vita familiare
- Non conoscenza di strutture di base, e scarso uso di metodi contraccettivi moderni
- Al momento dell'intervista lavoravano

Le uniche variabili che cambiano di poco le proporzioni fra modalità sono quelle sulla violenza, sul sesso del capofamiglia, sul pregiudizio verso i malati di AIDS e sul consumo di alcolici.

A conferma della ricchezza informativa espressa dai valori mancanti, abbiamo effettuato l'analisi delle corrispondenze forzando l'inserimento dei *missing*. Il grafico che ne emerge è molto simile a quello ottenuto a livello nazionale, con l'unica differenza che solo quattro modalità determinano, in modo totalizzante, il semiasse positivo della seconda dimensione. Queste modalità sono i valori mancanti delle variabili "sposo violento", "rispetto", "alcolismo" e "opinione violenza".

Grafico 4.3. Rappresentazione grafica delle modalità attive ed illustrative, rispetto ai primi due assi fattoriali, trovati con l'analisi delle corrispondenze per il Dipartimento di Cusco, con l'inserimento dei *missing values*



Il primo fattore spiega, secondo la scomposizione del Chi-quadrato, l'11,87% dell'inerzia totale; il secondo (asse orizzontale con i *missing*) ne spiega quasi l'8%: complessivamente quindi i due fattori spiegano quasi il 20%, percentuale che aumenta ad oltre il 24,59% con la correzione di Greenacre (14,96 e 9,63%), e all'86,19% con quella di Benzecri (63,34 e 22,85).

Le modalità sono condensate vicino al centro del piano cartesiano, e sono di difficile lettura. È possibile però individuare le modalità che si posizionano agli estremi dell'asse verticale (primo fattore); il semiasse positivo è caratterizzato da:

- Livello di istruzione superiore
- si frigo
- no pregiudizi verso i malati di AIDS
- conoscenza delle STD

In esso sono presenti le modalità "illustrative" che individuano le persone che abitano in città e quelle di origine spagnola.

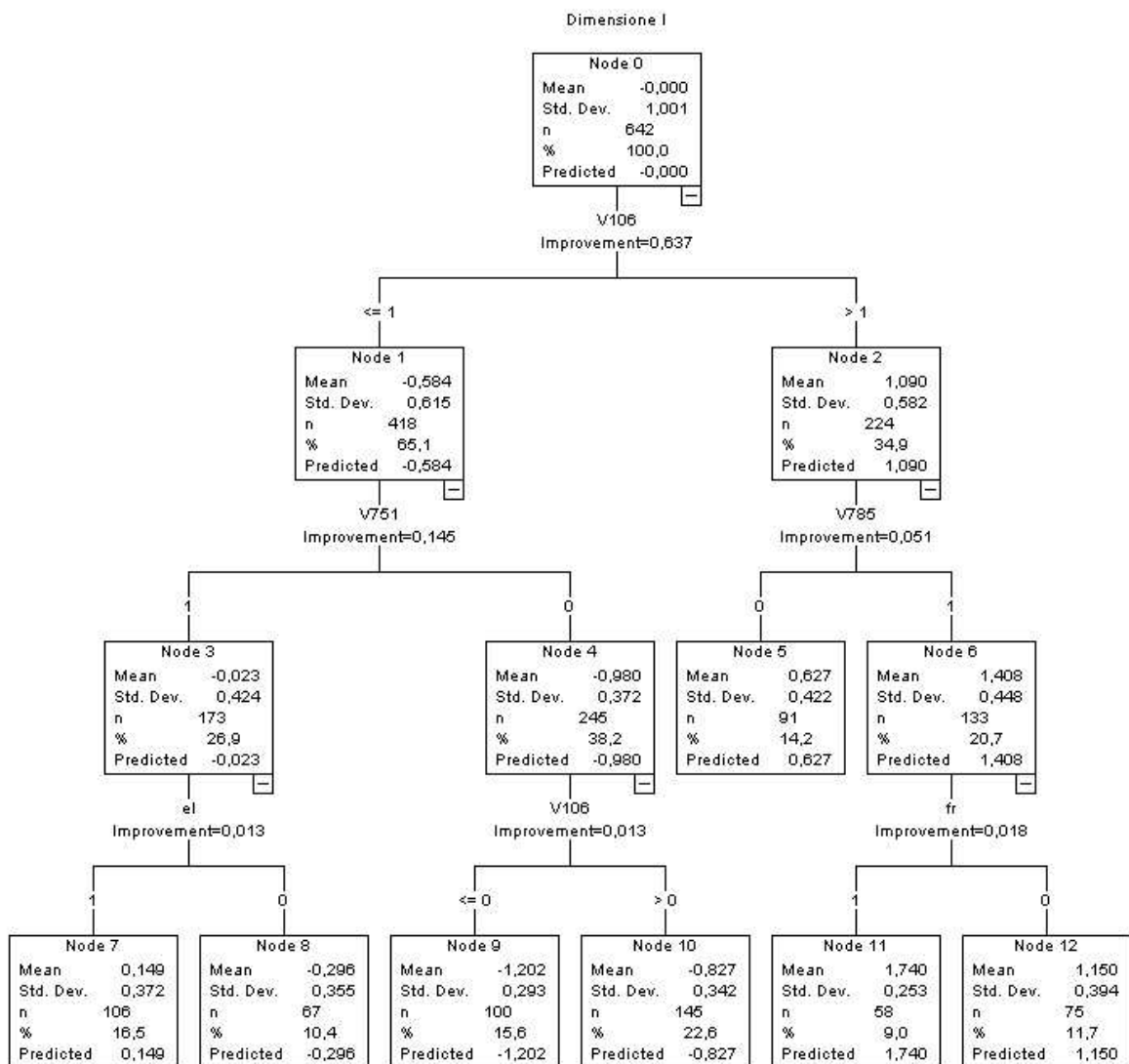
Le prime modalità che si incontrano nel semiasse negativo sono:

- non risposte alla domanda "dove comprare preservativi"
- nessun titolo di studio
- sposate a 18-20 anni
- non conoscenza dell'AIDS
- non risposte sul pregiudizio

La "nuvola" di valori mancanti determina un estremo del secondo fattore, senza però spiegarlo veramente: infatti abbiamo costruito un albero di segmentazione con i punteggi fattoriali assunti dalle unità rispetto alla seconda dimensione, ma le variabili selezionate non sono quelle alle quali appartengono le non risposte evidenziate nel grafico. Inoltre la devianza complessiva spiegata dal modello è così bassa (minore del 7%) che ci ha fatto desistere immediatamente dall'interpretare i risultati.

Non è sensato quindi continuare l'analisi con tutti i dati per i motivi sopra elencati. Concludiamo il discorso sui valori mancanti, riportando il dendrogramma della prima dimensione; se infatti da una parte la “nuvola” di missing fa perdere di significatività il secondo fattore, dall'altra la presenza di tutte le 642 donne rende significativa una variabile che era scomparsa nei capitoli precedenti, cioè la conoscenza dell'AIDS.

Grafico 4.4. Dendrogramma delle coordinate sul primo asse fattoriale



Per la prima ramificazione l'algoritmo di segmentazione seleziona, come per molti altri Dipartimenti, la variabile “titolo di studio”, dividendo le donne fra diploma inferiore o superiore da una parte, e quello elementare dall'altra. Questo passo porta ad una spiegazione del 63,7% della devianza.

Le donne meno istruite vengono poi ulteriormente raggruppate in base alla conoscenza dell'AIDS: per coloro che risposero negativamente alla domanda, torna ad essere importante il titolo di studio, con la definitiva suddivisione fra licenza elementare e nessun titolo di studio; viceversa, alle donne che conoscono l'AIDS si andrà a guardare la presenza o meno dell'elettricità in casa.

Alle donne più istruite si chiede invece se conoscono le STD: chi risponde affermativamente verrà ulteriormente suddiviso in base al possesso del frigorifero.

Ecco la tabella riassuntiva dei nodi terminali.

Tabella 4.2. Nodi terminali del dendrogramma per la prima dimensione (trovata con i *missing values*), con numerosità, media e descrizione del gruppo di donne

Nodo	N°	%	Media	Descrizione
11	58	9,0%	1,74	Donne con titolo di studio almeno secondario, che conoscono le STD e che posseggono un frigorifero
12	75	11,7%	1,15	Donne con titolo di studio almeno secondario, che conoscono le STD e che non posseggono un frigorifero
5	91	14,2%	0,63	Donne con titolo di studio almeno secondario ma che non conoscono le STD
7	106	16,5%	0,15	Donne con titolo di studio inferiore a quello secondario, che conoscono l'AIDS e che usufruiscono dell'elettricità
8	67	10,4%	-0,30	Donne con titolo di studio inferiore a quello secondario, che conoscono l'AIDS e che non usufruiscono dell'elettricità
10	145	22,6%	-0,83	Donne con licenza elementare, che non conoscono l'AIDS
9	100	15,6%	-1,20	Donne senza titolo di studio, che non conoscono l'AIDS

Complessivamente, l'albero spiega una percentuale di variabilità pari all'87,6%.

4.2 RISULTATI SENZA *MISSING VALUES*

Con le 366 osservazioni rimaste dopo l'eliminazione dei valori mancanti, procederemo con l'analisi multipla delle corrispondenze e con quella di segmentazione, parallelamente quindi a quanto compiuto per le altre regioni. Nel paragrafo precedente abbiamo osservato che le donne che non rispondono ad alcune domande "sensibili", vivono in una condizione di disagio a prescindere dalle loro risposte alle altre sezioni del questionario. Ora invece cercheremo di individuare le donne in situazione di povertà nella restante parte del campione.

L'analisi delle corrispondenze individua un primo fattore molto forte che spiega il 16,29% dell'inerzia, secondo la scomposizione corretta di Greenacre. Il secondo fattore invece contribuisce con una percentuale di spiegazione di poco superiore al 7%.

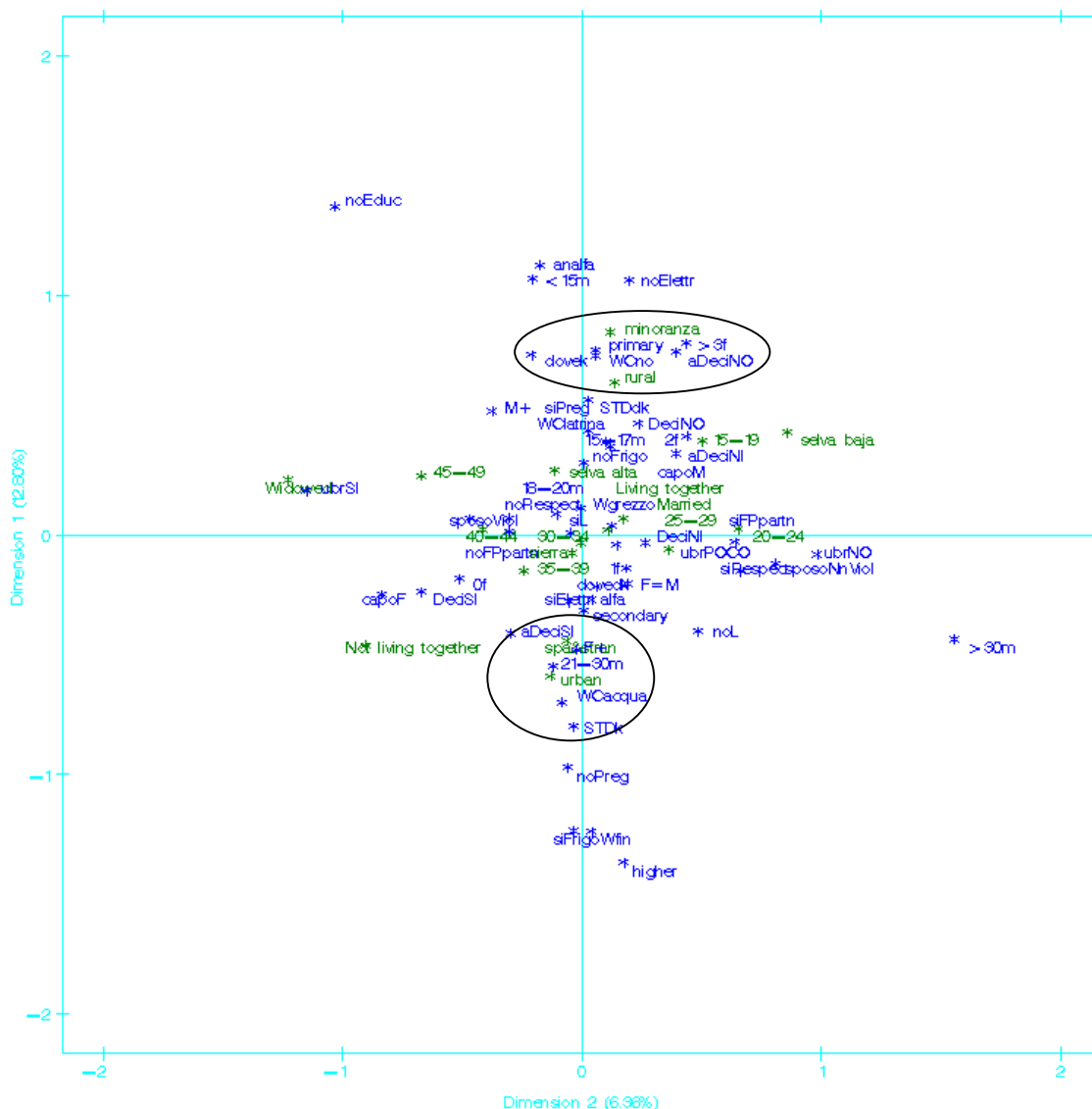
Tabella 4.3. elenco dei primi 7 fattori individuati con la correzione di Greenacre

Scomposizione dell'inerzia corretta di Greenacre

Inerzia principale	Inerzia corretta	Percentuale	Percentuale cumulativa	3 6 9 12 15
0,20	0,18	16,29	16,29	*****
0,10	0,08	7,27	23,56	*****
0,09	0,07	6,00	29,56	*****
0,08	0,06	5,15	34,70	*****
0,07	0,05	4,85	39,56	*****
0,07	0,05	4,26	43,82	*****
0,06	0,04	3,91	47,73	*****
0,01	0,00	0,19	100,00	

Viene riportato il grafico ottenuto dall'analisi. In esso si è deciso di non proiettare la variabile "conoscenza dell'AIDS" perché una modalità ha frequenza zero (come spiegato nel paragrafo precedente), né le modalità di "metodi contraccettivi moderni" e "opinione sulla violenza" perché il loro contributo alla spiegazione dei due fattori è minimo.

Grafico 4.5. Rappresentazione grafica delle modalità attive ed illustrative rispetto ai primi due assi fattoriali trovati con l'analisi delle corrispondenze per il Dipartimento di Cusco



L'asse delle ordinate presenta agli estremi delle modalità riguardanti l'istruzione (titolo di studio "higher" da un lato, e analfabetismo dall'altro). Per la prima volta una modalità della variabile "età al matrimonio" si posiziona esattamente sull'asse verticale in modo evidente. Le donne senza titolo di studio si scostano molto dall'asse verticale, collocandosi nel primo quadrante.

Nello specifico, il semiasse positivo delle ordinate è contrassegnato da:

- analfabetismo
- donne sposate prima di aver compiuto 15 anni
- no elettricità

Un secondo gruppo di punti rappresenta: licenza elementare, 3 figli o più, sa dove comprare i preservativi, no servizi igienici e no decisioni sulla salute.

Nel semiasse negativo invece sono presenti:

- titolo di studio alto
- si frigorifero
- muro con materiali solidi
- no pregiudizi verso i malati di AIDS

Il gruppo di punti immediatamente successivo comprende: conoscenza STD, wc con acqua, età al matrimonio fra i 21 e i 30 anni.

Le variabili che finora abbiamo visto occupare l'asse orizzontale, si scostano leggermente da esso in entrambi gli estremi. Questo ad evidenziare il fatto che poche variabili sono spiegate totalmente dal secondo fattore.

L'asse delle ascisse è contraddistinto, nella sua parte positiva, da:

- donne coniugate la prima volta ad un'età superiore ai 30 anni
- no consumo di alcolici
- sposo non violento
- programmazione familiare col partner
- rispetto della donna

Mentre hanno coordinata negativa:

- problemi di alcolismo
- capofamiglia donna
- decisioni sugli acquisti da sola
- 0 figli
- sposo violento

Le variabili supplementari (età, stato civile, gruppo etnico, urbanizzazione e regione morfologica) hanno le seguenti caratteristiche:

- sul semiasse positivo del primo fattore si collocano distintamente le minoranze etniche e le *campesinas*: esse vivono in abitazioni senza servizi igienici, hanno 3 figli o più, hanno la licenza elementare e non prendono decisioni sulla salute;
- sul semiasse opposto troviamo le donne che dimorano in città di etnia spagnola: esse si sposarono fra i 21 e i 30 anni, vivono in abitazioni con bagno, conoscono le STD e prendono da sole le decisioni sulla salute;

- le donne vedove sono strettamente legate ai problemi derivati dal consumo di alcolici;
- le modalità dell'età sono distribuite lungo l'asse orizzontale, ma comunque molto vicine all'origine: solo le donne più anziane del campione si scostano da esso, ponendosi nel primo quadrante, vicino alle donne vedove;
- il piccolo campione intervistato nella zona amazzonica si trova isolato nel primo quadrante: le altre due modalità della regione geografica invece sono piuttosto vicine fra loro al centro degli assi cartesiani;
- le donne separate, nel terzo quadrante, vivono da sole in condizioni di tranquillità economica.

La lettura del grafico è supportata dalla tabella con le migliori modalità per importanza nella spiegazione dell'inerzia di ogni dimensione.

Tabella 4.4. Migliori modalità per contributo parziale all'inerzia delle due dimensioni

Dimensione 1		Dimensione 2	
higher	0,086	ubrSI	0,136
siFrigo	0,073	sposoNnViol	0,119
noPreg	0,071	ubrNO	0,071
STDk	0,064	sposoViol	0,069
analfa	0,060	siRespect	0,068
noElettr	0,058	siFPpartn	0,065
WCacqua	0,057	DeciSI	0,061
primary	0,056	Of	0,045
WCno	0,050	capoF	0,045
STDdk	0,045	noRespect	0,031
siPreg	0,032	noFPpartn	0,031
dovek	0,031	noEduc	0,030
Wfin	0,030	aDeciSI	0,025
aDeciNO	0,029	M+	0,021
noEduc	0,026		
aDeciSI	0,023		
21-30m	0,023		

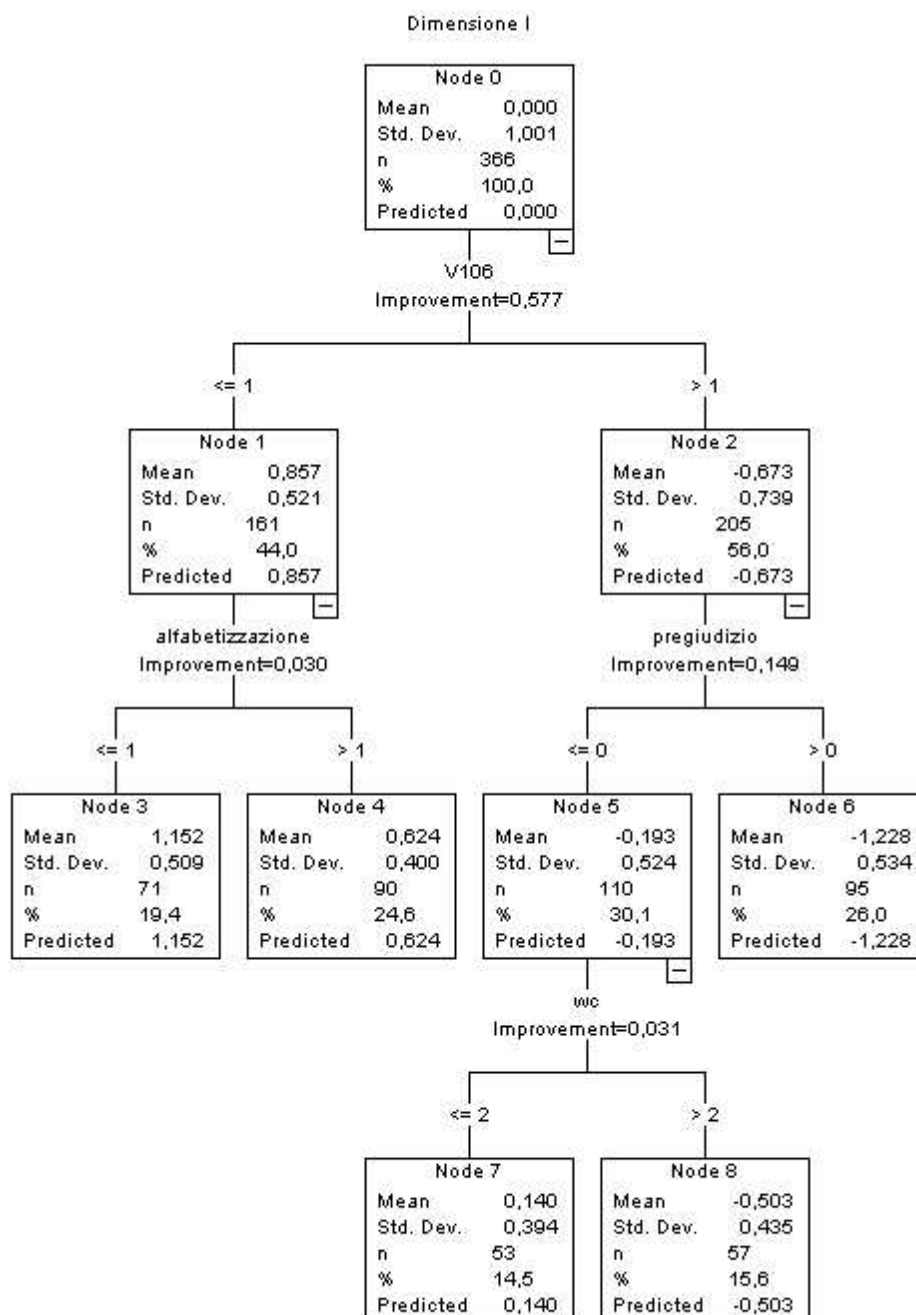
Il possesso di un titolo di studio elevato spiega l'8,6% dell'inerzia del primo fattore. Fra le variabili sull'abitazione, solo la presenza di un elettrodomestico in casa entra fra le prime modalità. Coprono un ruolo importante il pregiudizio e la conoscenza delle STD.

Per il secondo fattore, le due variabili sul consumo di alcolici e la violenza domestica spiegano con le loro modalità quasi il 40% dell'inerzia. La differenza

più evidente con le regioni studiate precedentemente, è lo scarso contributo portato dalle variabili sulla presa di decisione.

Il processo di segmentazione, per la prima dimensione, crea un albero basato sull'istruzione: la prima ramificazione, che spiega il 57,7% della devianza, suddivide le donne in un gruppo con un alto titolo di studio e in uno caratterizzato da un basso (o nullo) titolo di studio.

Grafico 4.6. Dendrogramma delle coordinate sul primo asse fattoriale



Dopo la prima suddivisione, l'algoritmo sceglie altre due variabili legate in qualche modo alla formazione intellettuale dell'intervistata: le donne sul semiasse positivo vengono differenziate in base alla variabile "analfabetismo"; invece alle donne più istruite è sufficiente vedere se hanno pregiudizi nei confronti dei malati di AIDS.

Solo al terzo livello viene selezionata una variabile riguardante l'abitazione (caratteristiche dei servizi igienici), che fa da "spartiacque" fra coordinate positive e negative sul primo asse fattoriale.

Rispetto al dendrogramma della prima dimensione con i *missing values*, si osserva che in realtà, nonostante le diverse variabili selezionate, il primo fattore segue sempre lo stesso schema: al primo livello viene selezionata una variabile riguardante l'istruzione, al secondo compaiono delle variabili che in qualche modo hanno a che fare con il livello culturale dell'intervistata, e solo al terzo livello le condizioni dell'abitazione diventano discriminanti.

Ecco il riassunto dei nodi terminali, in ordine decrescente rispetto alla coordinata media.

Tabella 4.5. Nodi terminali del dendrogramma per la prima dimensione, con numerosità, media e descrizione del gruppo di donne

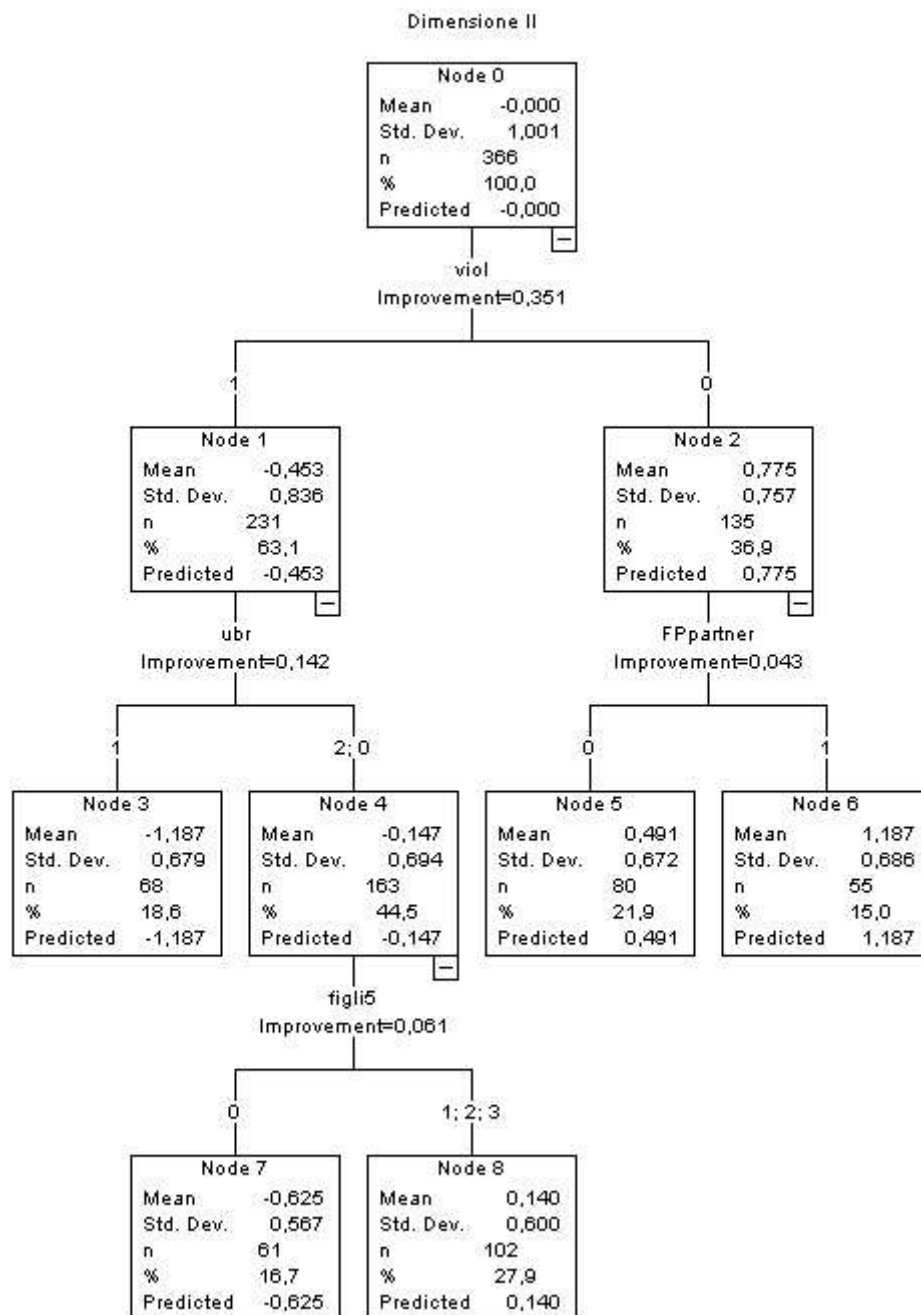
Nodo	N°	%	Media	Descrizione
3	71	19,4%	1,15	Donne analfabete
4	90	24,6%	0,62	Donne con titolo di studio primario
7	53	14,5%	0,14	Donne con alto titolo di studio, con pregiudizi verso i malati di AIDS, senza bagno o con una latrina rudimentale nella propria abitazione
8	57	15,6%	-0,50	Donne con alto titolo di studio, con pregiudizi verso i malati di AIDS, che vivono in abitazioni con servizi igienici moderni
6	95	26,0%	-1,23	Donne con alto titolo di studio, senza pregiudizi verso i malati di AIDS

L'albera spiega quasi il 79% della devianza complessiva.

Il dendrogramma per la seconda dimensione è costruito utilizzando le prime variabili presenti in tabella 4.3, cioè quelle che maggiormente ne spiegavano l'inerzia nell'analisi delle corrispondenze. Compaiono quindi la violenza

domestica, il consumo di alcolici, la pianificazione familiare e, per la seconda volta in questo studio, il numero di figli avuti negli ultimi cinque anni (questa variabile era già stata selezionata per discriminare la seconda dimensione in Tumbes, però si trovava al quarto livello di un albero molto più complesso, e spiegava appena l'1,6% della variabilità).

Grafico 4.7. Dendrogramma delle coordinate sul secondo asse fattoriale



La variabile “violenza domestica” è la prima ad essere selezionata, arrivando a spiegare il 35,1% della devianza: il primo gruppo, più consistente rispetto all’altro, è composto da 231 donne: esse hanno un partner che le maltratta fisicamente dentro le mura domestiche. Esse vengono ulteriormente suddivise in base al consumo di alcolici; per coloro che non hanno problemi di alcolismo (consumo basso o nullo), si guarda il numero di figli avuti nei 5 anni antecedenti l’intervista: la coordinata media di quelle che non ne hanno avuti rimane negativa, mentre le altre donne si posizionano sul semiasse positivo.

Le 135 donne che non subiscono violenza fisica vengono divise in base alla pianificazione familiare col partner: quelle che riescono a dialogare con lo sposo occupano l’estremo positivo dell’asse.

Tabella 4.6. Nodi terminali del dendrogramma per la prima dimensione, con numerosità, media e descrizione del gruppo di donne

Nodo	N°	%	Media	Descrizione
6	55	15,0%	1,19	Donne non soggette a violenza fisica e che prendono parte alla pianificazione familiare
5	80	21,9%	0,49	Donne non soggette a violenza fisica e che non prendono parte alla pianificazione familiare
8	102	27,9%	0,14	Donne soggette a violenza fisica, che consumano una quantità limitata di alcolici, con almeno un figlio avuto nei 5 anni antecedenti l’intervista
7	61	16,7%	-0,62	Donne soggette a violenza fisica, che consumano una quantità limitata di alcolici, che non hanno avuto figli negli ultimi 5 anni
3	68	18,6%	-1,19	Donne soggette a violenza fisica, con problemi di alcolismo

Complessivamente, l’albero spiega una percentuale di variabilità che si aggira attorno al 60%.

4.3 CONCLUSIONI

In questo capitolo abbiamo rimarcato l’importanza delle non risposte, soprattutto in una regione come Cusco dove un’alta percentuale della popolazione vive al di sotto della soglia di povertà estrema. Le donne che si imbarazzano nel rispondere a domande “sensibili” sono anche quelle che vivono in condizioni di disagio ed emarginazione, per via del basso livello di

istruzione, o per l'appartenenza ad una minoranza etnica, o perché vivono in una realtà contadina in abitazioni prive dei servizi più elementari.

Fra le intervistate che hanno risposto a tutte le domande, esiste un fattore preminente, basato principalmente su variabili riguardanti l'istruzione. Le condizioni dell'abitazione giocano un ruolo secondario, come accadeva anche per il dipartimento di Apurimac, selezionato a rappresentanza degli abitanti della *sierra* nel capitolo precedente.

Si è osservato che l'albero di segmentazione non cambia molto fra analisi con e senza i *missing values* (nonostante le diverse variabili selezionate): al primo livello è discriminante l'istruzione, al secondo il livello culturale dell'intervistata, e solo al terzo compaiono variabili riguardanti le condizioni dell'abitazione. Questo conferma che le non risposte influenzano principalmente la seconda dimensione.

Il risultato della seconda analisi senza *missing* è quindi generalizzabile per quanto riguarda la prima dimensione, nonostante l'elevato numero di non risposte.

Il secondo fattore individuato con l'analisi multipla delle corrispondenze è molto meno forte, in quanto spiega una quantità secondaria di varianza. Da esso però si ricavano comunque delle informazioni importanti. In primo luogo si osserva che l'asse fattoriale non può più essere definito come grado di indipendenza della donna (come avviene in altre regioni), perché le due variabili sulla presa di decisione vengono rimpiazzate, per importanza discriminante, dal consumo di alcolici. Inoltre le donne che vivono da sole (vedove e conviventi) si staccano dall'estremo dell'asse in direzioni opposte: di questa condizione di solitudine, esse quindi avvertono molto di più le conseguenze economiche (positive o negative che siano) che le conseguenze sociali.

Questo spiegherebbe anche il motivo per cui la prima suddivisione delle intervistate rispetto alla seconda dimensione venga eseguita sulla base della variabile "violenza fisica": in una realtà contadina molto arretrata, è segnale di sviluppo e presa di coscienza il riuscire a comunicare col proprio partner, piuttosto che il riuscire a prendere decisioni da sole (un'emancipazione ancora troppo lontana dalla percezione comune).

Rispetto all'analisi con i *missing values*, le variabili posizionate lungo l'asse delle ascisse sono sempre le medesime: nella prima analisi si notavano di meno in quanto appiattite verso il centro del piano cartesiano, a causa della polarizzazione all'estremo positivo dell'asse dei valori mancanti di alcune variabili. Entrambe le analisi delle corrispondenze, comunque, rivelano che il secondo fattore spiega una percentuale molto bassa di inerzia.

Infine si nota che, probabilmente a causa della disomogeneità del campione, l'analisi sui dati del dipartimento di Cusco non è utile per descrivere le tre regioni morfologiche che sono presenti in esso, come invece ci si aspettava sulla base delle esperienze maturate durante lo stage. Questo è dovuto al fatto che la realtà amazzonica, pur costituendo una parte considerevole del territorio totale del Dipartimento, è molto isolata, e quindi le analisi riescono a coglierla solo parzialmente.

5. CONCLUSIONI

Guardando alle caratteristiche del territorio peruviano, risulta evidente come esso sia caratterizzato da una forte frammentazione e diversificazione interna; gli abitanti, nelle diverse regioni geografiche, vivono in contesti che, pur essendo accomunati dalla condizione di precarietà, presentano strutture e caratteristiche molto diverse tra loro, di cui è necessario tener conto nel momento in cui si voglia implementare un qualsiasi programma di sviluppo. Si tratta quindi di uno di quei campi in cui è necessario pensare in modo globale (definendo alcuni standard operativi a livello generale), ma agire in maniera locale (adeguando le procedure e i meccanismi ai bisogni della popolazione locale, personalizzando le politiche d'intervento).

In questo lavoro si è cercato di capire quali variabili fossero comuni a tutto il territorio, e quindi utilizzabili per la costruzione di opportuni indicatori (di confronto fra aree geografiche e utilizzabili per la valutazione d'impatto sociale per le politiche locali), e quali invece differissero caratterizzando le 6 aree geografiche principali presenti in Perù: costa Nord, costa Sud, area metropolitana, giungla alta, giungla bassa e altipiano andino.

Per rispondere a questi interrogativi, abbiamo selezionato delle variabili a partire da alcuni schemi teorici proposti da grosse organizzazioni che lavorano in America Latina, in ambito sociale; successivamente le abbiamo sintetizzate attraverso la ricerca di dimensioni latenti: alla fine ne abbiamo individuate due che spiegano bene le correlazioni tra le variabili selezionate. Il primo fattore riguarda le condizioni economiche e materiali dove la donna vive. Il secondo invece si riferisce al disagio sociale percepito e al grado di indipendenza: il malessere provato dalle donne può infatti esser causato sia da un rapporto non paritario con il partner, che spesso sfocia in violenza domestica, sia da un eccesso di indipendenza che, in una società maschilista, sottopone la donna a problemi organizzativi e relazionali.

Rispetto alla prima dimensione, la variabile comune a tutto il territorio nazionale è il titolo di studio. Il livello di istruzione è la caratteristica che meglio distingue le donne campionate fra coloro che vivono in una condizione di povertà e quelle che invece sono uscite dalla instabilità economica. Il grado discriminante è il raggiungimento della licenza secondaria, ma per individuare le donne che si posizionano all'estremo più povero dell'asse fattoriale è sufficiente selezionare le analfabete, cioè coloro che non sanno leggere.

Le variabili invece che meglio caratterizzano le differenti aree geografiche, sono quelle che fanno riferimento all'abitazione. Già nella costa si riscontrano le prime differenze: il possesso di elettrodomestici (frigorifero) è comune a tutta la fascia costiera, ma a Nord, zona più contadina, è più importante nella discriminazione il materiale con cui sono costruite le pareti della casa.

Il muro gioca un ruolo fondamentale (addirittura superiore a quello dell'istruzione) anche nella *selva*, dove, fra l'altro, tutte le variabili riguardanti l'abitazione spiegano un'alta percentuale di inerzia: la presenza del frigorifero discrimina bene le donne che vivono in condizioni più agiate, mentre l'assenza di corrente elettrica individua quelle più povere.

Nella giungla alta invece la prima variabile discriminante è il livello di bontà dei servizi igienici, mentre nella *sierra*, dove le differenze di ricchezza nella popolazione sono più accentuate, è sufficiente studiare la variabile "presenza di elettricità nell'abitazione".

Inoltre il passaggio da zona urbana a zona rurale è molto più evidente nei Dipartimenti situati ad Est della Cordigliera andina: le due modalità della variabile "urbanizzazione" si collocano esattamente sull'asse, molto più vicine agli opposti che al baricentro del grafico, diversamente da quanto accade nella regione costiera.

Rispetto alla seconda dimensione, la variabile comune a tutto il territorio nazionale è la capacità di prendere decisioni sui grandi acquisti, che discrimina le donne più indipendenti da quelle invece soggette al proprio partner o alla famiglia di origine. L'informazione quindi che questa variabile fornisce è molto preziosa ai fini del nostro studio, visto che il microcredito e molte politiche sociali di organizzazioni che operano con le persone disagiate, cercano di lavorare principalmente con le donne, in quanto più affidabili, ma alle quali si

richiede (o alla partenza o dopo una collaborazione duratura) la capacità di prendere decisioni da sole sul proprio futuro, senza venire influenzate dalla volontà del marito. Questo è il requisito minimo per iniziare un percorso verso la coscienza delle proprie capacità e l'autostima personale.

Già nella giungla bassa, però, questa variabile viene sostituita, come prima variabile discriminante, dal consumo di alcolici: nelle regioni amazzoniche l'indipendenza della donna può essere vista con accezione negativa dalle comunità locali, tanto da venir percepita come situazione disagiata. Nel grafico nelle corrispondenze, la donna capofamiglia, che vive da sola e che è in grado di prendere decisioni viene infatti associata alla modalità riguardante l'elevato consumo di alcolici.

Sempre nella regione amazzonica, la dipendenza con l'uomo è così forte che risulta discriminante la variabile creata con le differenze di titolo di studio fra maschio e femmina.

Fra le altre variabili, ci sono il lavoro e il numero di figli che sono discriminanti solo nella costa Nord (le donne con figli e che non lavorano dipendono molto dalle decisioni del partner), e la pianificazione familiare (variabile utile per capire il rapporto col partner), non rilevante solo nella *selva* alta e bassa.

La regione metropolitana è una realtà a sé stante: l'effetto Guttman presente nei dati indica una sostanziale unidimensionalità delle frequenze osservate, il che evidenzia quindi la presenza di un unico fattore che unisce variabili materiali con variabili sociali.

Infine, nelle regioni della *sierra*, quali Apurimac e Cusco, si riscontra la più alta percentuale di donne che non risponde alle domande più "sensibili" del questionario (ad esempio, sulla violenza domestica o la conoscenza delle malattie sessualmente trasmissibili), anche a causa della presenza del partner al momento dell'intervista, che inibisce la compilazione: questo gruppo di donne non può però essere ignorato durante le analisi, perché descrive la realtà più disagiata del Perù. Esse sono per la maggior parte *campesinas*, appartenenti a gruppi etnici minori, senza istruzione e abitanti in case senza alcuna comodità.

Focalizzando l'attenzione sulle macroaree d'intervento individuate nel capitolo 2, possiamo ritenerci soddisfatti della scelta di alcune variabili che sono risultate essere ben discriminanti per i nostri obiettivi, e sono riuscite a descrivere le realtà studiate. Abbiamo visto comparire nelle varie analisi il problema dell'accesso all'istruzione, della violenza domestica, dell'autostima, della pianificazione familiare, dell'alcolismo e delle condizioni dell'abitazione. Sulla base dei risultati ottenuti in questo lavoro, è possibile sia capire quali sono le caratteristiche socio-culturali che discriminano le donne più povere (e quindi le aree dove è più importante intervenire), sia gettare le basi per la costruzione d'indicatori tarati sulle realtà locali.

Allo stesso tempo, però, sarebbe necessario un approfondimento di quelle aree tematiche che le variabili del questionario non sono riuscite a coprire o hanno toccato in maniera solo parziale. In particolare ci riferiamo alla partecipazione della donna nell'organizzazione sociale e politica della realtà in cui vive, alla sua capacità di negoziazione e alla conoscenza ed uso dei servizi pubblici. Inoltre non abbiamo indagato a fondo tutte le sfaccettature delle macroaree riguardanti il lavoro della donna e lo stato di nutrizione dei figli, a causa dell'elevato numero di valori mancanti e dell'elevato numero di domande filtro che non permetteva di indagare a fondo le caratteristiche di tutto il campione.

6. BIBLIOGRAFIA

COORDINADORA RURAL, *Cuadernos Andinos, 14: De campesinos a empresarios y lideres modernos*, Lima 2005

CONVENCION NACIONAL DEL AGRO PERUANO (CONVEAGRO), *Plan Estrategico de desarrollo agrario regional Cusco 2005 al 2012*, Cusco, 2005

DE AGOSTINI, *Calendario Atlante De Agostini 2008*, Istituto Geografico De Agostini, Novara, 2007

FABBRIS L., *Statistica multivariata*, McGraw-Hill Libri Italia srl, Milano, 1997

UNITED NATIONS DEVELOPMENT (UNDP), *Rapporto su lo sviluppo umano. 6: La parte delle donne*, Rosemberg & Sellier, Torino, 1995

SITOGRAFIA

www.measuredhs.com

www.inei.gob.pe

www.imp-act.org

<http://www.microfinancegateway.org/>

<http://www.undp.org/>

<http://hdr.undp.org/en/>

<http://www.un.org/millenniumgoals/>

<http://www.paho.org/>

<http://www.eclac.org/>

http://stats.uis.unesco.org/unesco/TableViewer/document.aspx?ReportId=143&F_Language=eng

<http://www.fordfound.org/>

<http://www.etimos.it/>

7. ALLEGATO 1

VARIABILI “ILLUSTRATIVE” UTILIZZATE DURANTE LE ANALISI

variabile	variabile d'origine	descrizione	modalità
v012		età dell'intervistata	<i>continua</i>
v013		età in classi da 5 anni	1 = "15-19" 2 = "20-24" 3 = "25-29" 4 = "30-34" 5 = "35-39" 6 = "40-44" 7 = "45-49"
v024		dipartimenti (Lima con Callao)	24 regioni
v023		dipartimenti	25 regioni
v025		urbanizzazione	1='urban' 2='rural'
ethnicity	v131	gruppo etnico	0='minoranza' 1='spa&stran'
v501		stato coniugale	0 = 'Never married' 1 = 'Married' 2 = 'Living together' 3 = 'Widowed' 4 = 'Divorced' 5 = 'Not living together'
sregion		regione geografica	1= 'Lima metropolitana' 2= 'Costa' 3= 'Sierra' 4= 'Selva Alta' 5= 'Selva Bassa'

VARIABILI "ATTIVE" UTILIZZATE DURANTE LE ANALISI

variabile	variabile d'origine	descrizione	modalità
v511		età al primo matrimonio/convivenza	<i>continua</i>
eta_matr	v511	età al primo matrimonio/convivenza in classi di frequenze	1='<15m' 2='15-17m' 3='18-20m' 4='21-30m' 5='>30m'
v106		livello di educazione	0 = "No education" 1 = "Primary" 2 = "Secondary" 3 = "Higher"
educazioneMF	v701-v106	differenza educazione maschi-femmine	0='F+' 1='F=M' 2='M+'
alfabetizzazione	v155	alfabetizzazione	0='analfa' 1='alfa'
V151		sexo del capofamiglia	1='capoM' 2='capoF'
lavorando	v714	donna lavora	0='noL' 1='siL'
figli5	v137	n° figli negli ultimi 5 anni	0='0f' 1='1f' 2='2f' 3='>3f'
uso_metodi_mod	v302	utilizzo di metodi contraccettivi	0='noMet' 1='siMet'
pregiudizio	v779	un insegnante può avere l'AIDS?	0='siPreg' 1='noPreg'
V751		conosci l'AIDS?	0='AIDSdk' 1='AIDSk'
V785		conosci le STD (malattie sessualmente trasmissibili)?	0='STDdk' 1='STDk'
dove	v762az	Non sa dove comprare i preservativi	0='dovedk' 1='dovek'
ubr	d114	alcolismo dell'intervistata	0='ubrNO' 1='ubrSI' 2='ubrPOCO'
viol	s1007	violenza fisica	0='sposoNnViol' 1='sposoViol'
opinioneViol	s1042	violenza come metodo educativo per i figli?	0='noViol' 1='siViol'
rispetto	s1003e	violenza verbale	0='noRespect' 1='siRespect'
FPpartner	v630a	discussione Pianificazione Familiare col partner	0='noFPpartn' 1='siFPpartn'
decisione	v743b	presa di decisione su grandi acquisti	0='DeciNO' 1='DeciNI' 2='DeciSI'

decisioneA	v743a	presa di decisione su salute	0='aDeciNO' 1='aDeciNI' 2='aDeciSI'
muro	v128	materiale delle pareti	2='Wgrezzo' 3='Wfin'
pavimento	v127	materiale del pavimento	1='Fnat' 2='Fgrezzo' 3='Ffin'
tetto	v129	materiale del tetto	2='Rgrezzo' 3='Rfin'
wc	v116	servizi igienici dell'abitazione	0='WCno' 1='WClatrina' 2='WCacqua'
h2o	v113	acqua nell'abitazione	0='noACQUA' 1='wellACQUA' 2='siACQUA'
el	v119	presenza di elettricità nell'abitazione	0='noElettr' 1='siElettr'
fr	v122	presenza del frigorifero nell'abitazione	0='noFrigo' 1='siFrigo'
tv	v121	presenza della tv nell'abitazione	0='noTv' 1='siTv'
priv	v044	manca di privacy nell'abitazione	0='si privacy' 1='no privacy'