

**UNIVERSITÀ DEGLI STUDI DI PADOVA**

**FACOLTÀ DI SCIENZE STATISTICHE**

**CORSO DI LAUREA**

**IN STATISTICA E TECNOLOGIE INFORMATICHE**

**TESI DI LAUREA**

**ANALISI DELLE OPZIONI  
DEL CONTRATTO ASSICURATIVO  
RC AUTO**

**RELATORE: CH.MA PROF. LAURA VENTURA**

**LAUREANDO: MARCO ZIERO**

**ANNO ACCADEMICO 2004-05**



# Indice

<b>1</b>	<b>Il contratto assicurativo RC Auto</b>	<b>3</b>
1.1	Lo <i>stage</i> . . . . .	3
1.2	RC Auto - Responsabilità Civile Auto . . . . .	3
1.2.1	Obbligatorietà . . . . .	4
1.2.2	Copertura . . . . .	4
1.2.3	Il Bonus-Malus . . . . .	5
1.3	Fasi importanti . . . . .	5
1.3.1	Tariffe e liberalizzazione tariffaria . . . . .	6
1.3.2	La procedura CID . . . . .	7
1.4	Enti di vigilanza e tutela . . . . .	8
1.4.1	ISVAP . . . . .	8
1.4.2	Fondo di Garanzia Vittime della Strada . . . . .	9
1.5	Situazione attuale . . . . .	9
1.6	Attualità . . . . .	12
1.6.1	Le proposte di Adiconsum . . . . .	12
1.6.2	Una scatola nera contro il caro-polizza . . . . .	13
1.7	Obiettivo della tesi . . . . .	14
<b>2</b>	<b>Analisi esplorativa dei contratti RC Auto</b>	<b>15</b>
2.1	I dati . . . . .	15
2.2	Le variabili . . . . .	16
2.2.1	Età . . . . .	16
2.2.2	Sesso . . . . .	19

2.2.3	Professione . . . . .	20
2.2.4	Cavalli fiscali (hp) . . . . .	22
2.2.5	Sinistri . . . . .	23
2.2.6	Classe di merito . . . . .	25
2.3	Analisi temporali . . . . .	26
2.4	Alcune relazioni tra le variabili . . . . .	29
2.5	Routines - R . . . . .	36
<b>3</b>	<b>Analisi Multivariata</b>	<b>47</b>
3.1	Ricodifica delle variabili del database . . . . .	47
3.2	Analisi delle Corrispondenze . . . . .	48
3.3	Applicazione ai dati RC Auto . . . . .	50
3.4	Routines - R . . . . .	54
<b>4</b>	<b>Conclusioni</b>	<b>55</b>
	<b>Bibliografia</b>	<b>57</b>
4.1	Bibliografia . . . . .	57
4.2	Siti Internet . . . . .	57

# Premessa

Questa tesi nasce dall'esperienza di uno *stage* della durata di 6 mesi svoltosi presso l'agenzia AXA di Cittadella. Essa ha lo scopo di delineare un profilo generale degli assicurati che hanno sottoscritto un contratto assicurativo presso tale agenzia.

Durante l'attività di *stage* ho avuto la possibilità di interagire con il database centrale della compagnia assicurativa, per gestire i contratti assicurativi. Ciò mi ha permesso di utilizzare i dati, in particolare della polizza assicurativa RC Auto, per costruire il dataset sul quale compiere le analisi.

L'analisi si compone di una prima parte, un'analisi descrittiva, e di una seconda parte, sempre esplorativa, che coinvolge l'utilizzo di una tecnica statistica più raffinata. In particolare, nella prima parte si studiano singolarmente le variabili che compongono il dataset, le relazioni nei tre anni considerati ed altri tipi di relazioni bivariate, utilizzando prevalentemente strumenti grafici.

Nella seconda parte, invece, utilizzando come strumento statistico l'Analisi delle Corrispondenze Multiple, si studiano le relazioni fra tutte le variabili, per verificare eventuali legami e riuscire a definire così uno o più ritratti dei clienti della compagnia assicurativa.



# Capitolo 1

## Il contratto assicurativo RC Auto

### 1.1 *Lo stage*

Lo stage si è svolto presso l'Agenzia AXA di Cittadella, sotto la supervisione del Rag. Streliotto Lino e del suo staff, per un periodo di circa 6 mesi. L'attività principale è stata quella di interagire con il database centrale della Compagnia Assicurativa AXA per sottoscrivere nuovi contratti assicurativi e modificarne di già esistenti (rinnovi, modifiche, cancellazioni). Di tutti i contratti che sono stati gestiti, la maggior parte consisteva nel contratto sulla responsabilità civile (RC), il che ha permesso di iniziare da presto ad analizzare il contratto RC Auto e tutte le sue opzioni.

### 1.2 **RC Auto - Responsabilità Civile Auto**

In Italia, per condurre un'automobile è necessario che questa sia assicurata sulla responsabilità civile (RC). Lo stabilisce la legge 990 del 24 dicembre 1969. L'assicurazione per i danni derivanti dalla circolazione dei veicoli - così è definita l'assicurazione RC Auto - impegna la Compagnia a pagare la somma che il titolare della polizza dovrebbe pagare per i danni causati ad un terzo. In altre parole, al posto del responsabile dell'incidente subentra l'impresa che risarcisce il danneggiato nei limiti previsti dal massimale che, attualmente, prevede una

copertura non inferiore a circa 775.000 euro per sinistro, per persona e per danni a cose o animali.

L'assicurazione sulla responsabilità civile ha diversi aspetti che la caratterizzano come l'obbligatorietà, la copertura e la tariffa bonus/malus.

### **1.2.1 Obbligatorietà**

L'assicurazione della responsabilità civile del proprietario e del conducente è obbligatoria. Tutte le altre garanzie, come ad esempio quella relativa al furto e all'incendio, possono essere scelte liberamente. L'obbligo dell'assicurazione di responsabilità civile non riguarda necessariamente il proprietario del mezzo, ma chiunque abbia disponibilità dello stesso. E' importante ricordare che, come è obbligatoria l'assicurazione RC per i veicoli che circolano a norma dell'art. 11 della medesima legge, è obbligatorio per le Compagnie accettare le proposte di assicurazione che ricevono.

### **1.2.2 Copertura**

L'assicurazione RC Auto vale sempre, a meno che non si sia ommesso di pagare il premio. Infatti, trascorsi 15 giorni dalla scadenza della rata, la copertura è sospesa e, quindi, la Compagnia non risarcisce più eventuali danni provocati dal veicolo assicurato. E' bene ricordare che in alcuni casi la Compagnia paga il danno al terzo danneggiato, ma poi si rivale sul proprio assicurato richiedendogli la restituzione dei soldi che ha dovuto versare. Ciò può accadere, ad esempio, se si verifica una delle seguenti circostanze: guida senza patente o con la patente scaduta, revocata o sospesa; guida in stato di ebbrezza o sotto l'effetto di droghe; guida da parte di chi ha superato gli esami, ma non ha ancora la patente oppure che ha una patente non adatta al veicolo che sta guidando ed altri casi ancora, che possono variare secondo la giurisprudenza corrente. Senza contare, poi, talune limitazioni contenute in non poche polizze alle quali occorre prestare particolare attenzione per evitare brutte sorprese. In caso di controlli, per dimostrare che si è assicurati, è necessario esibire, oltre al contrassegno, che

deve essere sempre esposto sul parabrezza, il certificato che contiene i dati essenziali della polizza assicurativa cui è riferito, senza i quali il veicolo può essere sequestrato con conseguenti pesantissime multe.

### 1.2.3 Il Bonus-Malus

Tale formula contrattuale, che regola la stragrande maggioranza delle polizze RC auto, prevede a ogni scadenza progressivi rincari o riduzioni percentuali del premio, rispettivamente se l'assicurato ha causato o no sinistri nel cosiddetto 'periodo di osservazione' (che dura un anno e termina tre mesi prima della scadenza della polizza). Sono stabilite 18 classi di merito, con la 14esima, quella d'ingresso, che corrisponde alla tariffa base. Se, nel periodo di osservazione, l'automobilista non causa incidenti, sale in 13esima (va in bonus) ottenendo uno sconto sul successivo premio; se, invece, causa incidenti, retrocede in 16esima (in malus) e subisce un rincaro. Le variazioni percentuali dei premi cambiano a seconda dell'assicurazione, ma, come regola generale, alla prima classe corrisponde una riduzione del 50% della tariffa base, mentre alla classe 18esima una maggiorazione del 100%. Inoltre, alcune compagnie prevedono ulteriori classi sia in bonus sia in malus, cui corrispondono ulteriori riduzioni o rincari. Se si raffronta il premio stabilito per la classe d'ingresso (la 14) con quello previsto per la classe migliore (la prima) si noterà una differenza pari generalmente alla metà. Se non si causano sinistri si "scende" di una classe, altrimenti con un solo sinistro si "sale" di due classi.

## 1.3 Fasi importanti

Due momenti importanti per la "storia" dell'assicurazione sulla responsabilità civile sono segnati dalla liberalizzazione delle tariffe (1994) e dalla nascita della procedura Convenzione di Indennizzo Diretto, procedura che consente al cliente di essere risarcito direttamente dalla propria Compagnia di Assicurazione.

### 1.3.1 Tariffe e liberalizzazione tariffaria

Sino alla metà del 1994 lo Stato determinava le tariffe. Ora invece, in linea con gli altri paesi comunitari, c'è la piena liberalizzazione tariffaria: per stabilire il costo della copertura assicurativa non si prendono più in considerazione soltanto la potenza fiscale del veicolo, la provincia ove risiede il contraente e la classe bonus-malus di appartenenza, ma anche altri dati ugualmente importanti, come, ad esempio, l'età, il sesso e la professione del contraente, da quanti anni è in possesso della patente, l'eventuale presenza nel veicolo assicurato dell'airbag e/o dell'Abs, eccetera. Ciascuna compagnia, in base alla legge 273/02, è obbligata a rendere noti al pubblico le tariffe e le condizioni generali e speciali del contratto praticate su tutto il territorio nazionale. Le tariffe e le condizioni contrattuali devono essere pubblicizzate da ogni Compagnia non solo presso le agenzie e i punti di vendita, ma anche attraverso i siti Internet che consentono agli assicurati di calcolare e confrontare le diverse offerte e di conoscere anticipatamente le norme che regolano le polizze RC Auto. Le Compagnie devono comunicare l'aumento della tariffa secondo le modalità indicate nel contratto. Se è previsto il tacito rinnovo, cioè il rinnovo automatico della polizza a meno che una delle parti non l'abbia disdetta in tempo utile, l'assicuratore ha due possibilità:

- inviare una comunicazione scritta in cui sia definita con precisione l'entità dell'aumento praticato al contraente entro il termine previsto nel contratto. In caso contrario, la polizza deve essere rinnovata alla vecchia tariffa;
- affiggere nelle proprie Agenzie appositi avvisi per comunicare agli assicurati l'aumento.

Nel primo caso, se non intende accettare l'aumento, il contraente, almeno 30 giorni prima della scadenza della polizza, deve inviare la disdetta mediante raccomandata A.R. o fax, altrimenti il contratto sarà automaticamente rinnovato. In caso di disdetta, che può essere naturalmente formalizzata anche per qualsiasi altro motivo, non vale più il periodo di tolleranza di 15 giorni successivi

alla scadenza della polizza, di cui si è detto. Nel secondo caso, poiché la semplice comunicazione in Agenzia non è una misura particolarmente favorevole al consumatore, il contraente può disdire il contratto anche dopo la scadenza, purché non siano trascorsi 15 giorni dalla stessa. Se non c'è stata nemmeno l'affissione in Agenzia il contratto si risolverà alle 24 del quindicesimo giorno successivo a quello della scadenza. Se l'aumento tariffario - eccezione fatta per quello derivante dal malus - è superiore al tasso programmato d'inflazione, la disdetta potrà essere presentata fino al giorno di scadenza del contratto. Se la Compagnia disdice una polizza RC Auto, il contraente può chiedere in ogni caso di essere nuovamente assicurato dalla medesima Compagnia, che è obbligata ad accettare tale richiesta.

### 1.3.2 La procedura CID

La "Convenzione di indennizzo diretto" (CID) è un sistema che funziona grazie ad un accordo tra le compagnie di assicurazione italiane. In base allo stesso chi subisce un danno in un incidente può chiedere il rimborso direttamente alla propria compagnia, che provvederà successivamente a farsi rimborsare dalla compagnia del danneggiante. Si applica quando sono soddisfatte tutte le seguenti condizioni:

- il modulo blu è firmato da entrambi i conducenti;
- si tratta di urto fra due soli veicoli (esclusi ciclomotori e macchine agricole) assicurati da imprese aderenti alla CID;
- non vi sono danni alle persone;
- il danno al veicolo non supera i 10 milioni di lire (iva compresa).

In tal caso, l'assicuratore:

- raccoglie la denuncia di sinistro e ne rilascia ricevuta all'assicurato;

- indica all'assicurato il numero del sinistro, la persona o l'ufficio incaricato di trattare il danno con il relativo recapito, numero telefonico e reperibilità;
- provvede alla perizia del danno entro 10 giorni da quello in cui è stato messo a sua disposizione il veicolo;
- entro i 15 giorni successivi alla perizia:
  1. in caso di accordo con l'assicurato, paga l'importo concordato e, contestualmente, può chiederne ricevuta;
  2. in caso di disaccordo (ad esempio sull'entità del danno o sulle percentuali di responsabilità del sinistro), paga la somma che ritiene dovuta.

In quest'ultimo caso, l'assicurato incassa la somma e può rivolgersi all'assicuratore dell'altro veicolo per richiedere l'ulteriore parte di danno che ritiene gli debba essere risarcita. Qualora, nel corso della procedura CID, l'assicuratore rilevi fatti che impediscano l'applicazione della procedura stessa, ne informa l'assicurato, invitandolo a rivolgersi all'assicuratore dell'altro veicolo, al quale, contestualmente, trasferisce la pratica.

## 1.4 Enti di vigilanza e tutela

Per tutelare i danneggiati e per vigilare sull'operato delle Compagnie, sono stati istituiti due enti, l'ISVAP e il Fondo di Garanzia Vittime della Strada

### 1.4.1 ISVAP

L'ISVAP - Istituto per la vigilanza sulle assicurazioni private e di interesse collettivo - è un ente dotato di personalità giuridica di diritto pubblico ed è stato istituito con legge 12 agosto 1982, n. 576, per l'esercizio di funzioni di vigilanza

nei confronti delle imprese di assicurazione e riassicurazione nonché di tutti gli altri soggetti sottoposti alla disciplina sulle assicurazioni private, compresi gli agenti e i mediatori di assicurazione. L'ISVAP svolge le sue funzioni sulla base delle linee di politica assicurativa determinate dal Governo.

### **1.4.2 Fondo di Garanzia Vittime della Strada**

Istituito dal legislatore a tutela dei danneggiati, il Fondo di Garanzia interviene, nei limiti dei massimali obbligatori, nei seguenti casi:

1. sinistri causati da veicoli non identificati, solo per danni alle persone;
2. sinistri causati da veicoli non assicurati, per i danni alle persone ed anche, con una franchigia di 500 euro, per i danni alle cose;
3. sinistri causati da veicoli assicurati con imprese poste in "liquidazione coatta amministrativa".

## **1.5 Situazione attuale**

Per dare un quadro generale della situazione nel nostro Paese e per evidenziare il rapporto tra imprese e utenti si riportano estratti del "Protocollo d'intesa sull'assicurazione R.C.Auto tra Governo, Ania e le Associazioni dei consumatori (5 maggio 2003)": **PREMESSA PROGRAMMATICA E CONDIVISIONE DEGLI OBIETTIVI.**

L'assicurazione RC Auto ha una valenza sociale di fondamentale importanza:

- tutela l'interesse dei danneggiati ad essere pienamente risarciti;
- tutela l'interesse degli automobilisti a non dover rispondere con il proprio patrimonio alle richieste dei danneggiati.

E' quindi interesse generale che il "sistema RCA", in tutte le sue articolazioni normative, regolamentari, economiche, aziendali, fiscali, organizzative etc.

possa funzionare in maniera efficace e trasparente, trovando un equilibrio economicamente sostenibile e vantaggioso per imprese e utenti. L'intervenuta liberalizzazione del mercato RC Auto avviata nel 1994 non ha prodotto gli effetti auspicati e ancora oggi il sistema è lontano dalle condizioni di funzionamento ottimali e non ha ancora trovato un punto di equilibrio. Ciò è ampiamente dimostrato dalla recente, tormentata, storia del settore. Qualsiasi realistica ipotesi di intervento sullo scenario attuale non può prescindere da alcuni elementi oggettivi. Tra essi in particolare si segnalano:

- l'alto numero di veicoli circolanti nel nostro paese: 45 milioni;
- la più alta densità di traffico in Europa: 789 veicoli ogni 1000 abitanti e 137 ogni chilometro di strada. In Francia vi sono invece 602 veicoli ogni 1000 abitanti e 40 per km; in Germania 614 ogni 1000 abitanti e 110 ogni km; in Gran Bretagna 478 veicoli ogni 1000 abitanti e 76 per km;
- un basso tasso di manutenzione e sviluppo della rete viaria, aggravato anche da una eccessiva concentrazione (85,6%) del trasporto merci su strada.

Tale situazione "ambientale" influisce pesantemente sugli indici di rischiosità della circolazione stradale:

- circa 4,5 milioni di incidenti risarciti ogni anno;
- oltre 900.000 gli italiani che, nello stesso periodo, riportano danni alla persona;
- 8.000 circa i morti ogni anno per incidenti.

Particolarmente gravi, per le loro implicazioni sociali, sono:

- il continuo forte aumento della incidenza del numero dei sinistri con danno alla persona sul totale (da 9,9% nel 1990 al 22% nel 2000);
- l'elevata conflittualità (1.300 milioni di euro di costi legali annui);

- alcuni fenomeni patologici quali truffe, compagnie fantasma e assicurazioni false, perpetrate a danno sia delle imprese sia dei consumatori;
- forti aumenti delle tariffe RC Auto, particolarmente accentuate in alcune aree territoriali e nei confronti di specifiche categorie di assicurati.

Tutto ciò ha determinato una progressiva incrinatura nel rapporto tra imprese e utenti, che le parti intendono ricostruire positivamente anche attraverso le misure previste dal presente Protocollo.

Tutto ciò premesso si concorda sulla necessità di intervenire con decisione e coerenza sui fattori di crisi dell'assicurazione auto per migliorarne tangibilmente e in via strutturale le condizioni di funzionamento ed in particolare per contenere le cause di costo che hanno influito sui rilevanti aumenti degli ultimi anni; sull'opportunità di individuare alcune aree fondamentali sulle quali intervenire per incidere in maniera organica e strutturale sul costo dell'assicurazione auto, coniugando sinergicamente interventi sulle tariffe - fermo restando il fabbisogno di risorse stimato da ogni impresa per l'equilibrio della gestione tecnica - e sulla prevenzione dei rischi. In particolare si vuole intervenire su:

- *giovani*  
studiando misure di formazione e informazione per un accesso consapevole dei giovani alla guida, per rendere possibili agevolazioni tariffarie ad hoc;
- *ciclomotori e motocicli*  
sviluppando azioni di prevenzione mirate alla riduzione della rischiosità che connota tali veicoli e consentire sconti o riduzioni tariffarie;
- *riduzione del contenzioso*  
per estendere a livello nazionale l'utilizzo di procedure conciliative per la soluzione rapida e positiva delle controversie nei sinistri RC Auto;
- *trasparenza*  
per agevolare la confrontabilità di prodotti e tariffe, le scelte consapevoli degli assicurati e quindi la concorrenza;

- *garanzie*  
per migliorare l'operatività della copertura, anche con interventi legislativi e sulla normativa contrattuale, in particolare per i casi di furto, rottamazione e vendita del veicolo;
- *efficienza del servizio*  
per una gestione dei sinistri più rapida e trasparente, estendendo la procedura CID ai danni alla persona di lieve entità.

## 1.6 Attualità

### 1.6.1 Le proposte di Adiconsum

Negli ultimi anni, il prezzo per le polizze assicurative è salito drasticamente, con aumenti considerevoli per tutte le classi di merito.

C'è chi si è messo in moto per tentare di migliorare la situazione.

Indennizzo diretto e conciliazione potrebbero tagliare del 10% i costi che gravano oggi sulla tariffa RC Auto. Sono queste le proposte di Adiconsum, illustrate in occasione del resoconto annuale sul settore assicurazioni. Secondo i dati dell'organizzazione, a oltre 10 anni dalla liberalizzazione, le differenze di polizza tra una compagnia e l'altra sono rilevanti: per alcuni profili si arriva alla cifra record del 300% in più. La ricerca della polizza più conveniente può ridurre, quindi, in modo significativo il costo dell'assicurazione. “Ma ciò è quasi impossibile - ha dichiarato Paolo Landi, segretario generale di Adiconsum - la comparabilità tra costi resta difficile anche per un consumatore esperto che ricorra ad Internet. In questi ultimi due anni, grazie ai provvedimenti previsti dall'accordo RC Auto sottoscritto nel 2003, gli aumenti delle tariffe sono stati più contenuti, ma riduzioni significative restano però un miraggio”. In particolare, nulla è stato fatto contro le polizze usurarie per i neo-patentati. Per garantire costi assicurativi a portata di consumatore, Adiconsum ha presentato alcune proposte al Ministro delle Attività Produttive e alle compagnie assicurative:

- “polizza virtuosa”, cioè riduzione delle tariffe collegate a comportamenti

virtuosi dell'assicurato, quali ad esempio la disponibilità a ricorrere alle procedure conciliative in caso di contenzioso;

- indennizzo diretto e procedure conciliative;
- comparabilità delle tariffe;
- definizione di alcuni vincoli per impedire polizze “usuarie” ai neopatentati.

Infine, per il prossimo anno sarà realizzato un osservatorio congiunto Aci-Adiconsum-Censis sulla mobilità.

### 1.6.2 Una scatola nera contro il caro-polizza

Il suo nome tecnico è “clear box”: installata sull'automobile permetterà di monitorare, in ogni momento, i comportamenti del conducente. In cambio i guidatori potranno contare su uno sconto della polizza RC-auto non inferiore al 10%. Il Ministero delle Attività Produttive e l'Isvap hanno firmato in questi giorni un accordo per avviare la sperimentazione delle “clear box”, sorta di scatole nere per automobili in grado di registrare i comportamenti dei guidatori e ricostruire con precisione il posizionamento e la dinamica degli incidenti stradali. Al pari delle scatole nere degli aerei o di treni, infatti, le “clear box” aiuteranno a ricostruire i sinistri nei dettagli, smascherando gli imbrogli che simulano gli incidenti stradali allo scopo di truffare le assicurazioni.

La sperimentazione coinvolgerà per ora 5 città: Torino, Milano, Roma, Napoli e Palermo. Gli automobilisti saranno selezionati in un campione definito dal Ministero e dall'Isvap. Chi aderirà alla sperimentazione potrà usufruire, per due anni, di uno sconto sulla polizza Rc-auto non inferiore al 10%. L'installazione delle “clear box” sugli autoveicoli sarà gratuita e la scatola verrà dotata un sistema di rilevazione satellitare con tecnologia Gps e Gprs. Il Movimento Difesa del Cittadino (MDC) ha accolto positivamente l'avvio della sperimentazione della “scatola nera” per le auto. “Approviamo pienamente l'uso della “clear box” - ha commentato Antonio Longo Presidente dell'associazione - Ci auguriamo che il nuovo sistema porti finalmente ad una significativa riduzione

delle tariffe RC Auto, considerando che alcune grandi compagnie già offrivano questa possibilità con riduzioni pari al 20-30%. Rimane solo da capire ora quali saranno i criteri con cui verranno scelti i veicoli su cui installare la scatola nera. Attendiamo inoltre che il Ministero ci comunichi i risultati al termine del periodo di sperimentazione”. Più cauto il giudizio di Adoc, Adusbef, Codacons e Federconsumatori: “Se la scatola nera permetterà di combattere le frodi e di prevenire gli incidenti e se il Ministero ha deciso di investire in questa sperimentazione, c’è da augurarci che il riscontro in termini di diminuzione delle tariffe sarà sostanzioso, almeno superiore al 40%. Una diminuzione minore vorrebbe dire che i soldi spesi non sono stati ben impiegati. Bisognerà anche vigilare sull’aspetto della privacy. I dati in possesso delle compagnie infatti, non dovranno essere utilizzati per nessun altro fine se non quello assicurativo. Non vorremmo trovarci di nuovo di fronte ad un’innovazione tecnologica invasiva e senza vantaggi reali per il consumatore”.

## 1.7 Obiettivo della tesi

Considerando le opzioni del contratto RC Auto, come variabili del database a disposizione presso l’agenzia AXA di Cittadella, ci si è prefitti di delineare i vari profili dei clienti dell’Agenzia AXA di Cittadella, utilizzando un’analisi esplorativa. Essa si compone prima di un’analisi preliminare, prevalentemente grafica, e poi di un’analisi delle corrispondenze ed un’analisi cluster, tramite un processo più approfondito e dettagliato.

L’analisi esplorativa, prima fase dell’analisi dei dati, è in grado non solo di supportare le ipotesi di lavoro di partenza, ma anche di orientare lo studio verso la formulazione di nuove ipotesi, di verificare qualitativamente l’esistenza di opportune assunzioni sulle variabili in gioco e di suggerire eventuali modelli statistici parametrici, da utilizzare a fini inferenziali.

## Capitolo 2

# Analisi esplorativa dei contratti RC Auto

### 2.1 I dati

Il database a disposizione è composto da 300 dati, ciascuno dei quali è caratterizzato da 7 variabili. I dati corrispondono a 300 contratti RC Auto catalogati nell'archivio dell'Agenzia AXA di Cittadella. Tali contratti sono relativi solamente ad autovetture intestate a privati cittadini, escludendo quindi eventuali aziende ed eventuali altre vetture di diverso tipo (motocicli, autocarri, macchine agricole, rimorchi).

Le variabili di ciascun contratto sono: anno, sesso, età, professione, cavalli fiscali (hp), numero di sinistri negli ultimi 5 anni, classe di merito.

La variabile anno può assumere come valori 2004, 2003 e 2002, cioè gli anni in cui sono stati catalogati i contratti oggetti di studio. Le variabili sesso (maschio=0 e femmina=1), età e professione si riferiscono alla persona che ha sottoscritto il contratto con la Compagnia. La variabile professione in particolare, può assumere i valori 1= dipendente, 2=casalinga, 3=imprenditore, 4=commerciante, 5=libero professionista, 6=pensionato/a, 7=artigiano/a, 8=medico, 9=agente rappresentante, 10=insegnante, 11=altro. La variabile cavalli fiscali, indicata con hp, si riferisce ai cavalli fiscali dell'autovettura oggetto del contratto. La

variabile sinistri indica il numero di sinistri registrati negli ultimi 5 anni dalla persona soggetto del contratto (è stato appurato che questa variabile assume valori 0 e 1, e solo in un caso il valore 2. Infine, la variabile classe di merito indica lo stato bonus/malus della persona che sottoscrive il contratto.

E' necessario sottolineare che, per quanto riguarda la variabile classe di merito, essa non rappresenta la "bravura" del conducente, se analizzata in relazione con l'età del soggetto, in quanto non tutti i contraenti hanno sottoscritto un'auto-vettura alla stessa età, e poiché non vi è un limite massimo per sottoscrivere un contratto RC Auto.

Anche per quanto riguarda la variabile sinistri va specificato che se un soggetto cambia Compagnia Assicurativa, con la quale aveva registrato un certo numero di incidenti, nel nuovo contratto, sotto la voce "Numero di sinistri negli ultimi 5 anni", verrà indicato 0 (zero) ed è come se la "storia del soggetto venisse cancellata".

## **2.2 Le variabili**

Si inizia l'analisi studiando separatamente le singole variabili. A questo scopo, di seguito, si riportano le analisi grafiche preliminari delle singole variabili che compongono il database.

### **2.2.1 Età**

Utilizziamo l'istogramma, il boxplot ed il grafico quantile-quantile per studiare la distribuzione della variabile età.

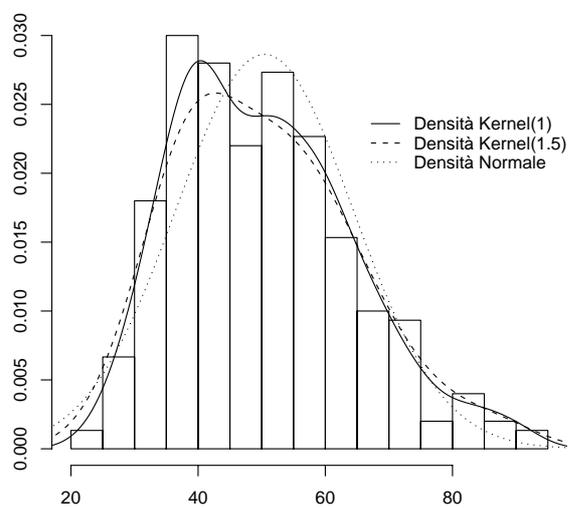


Figura 2.1: Istogramma della variabile età

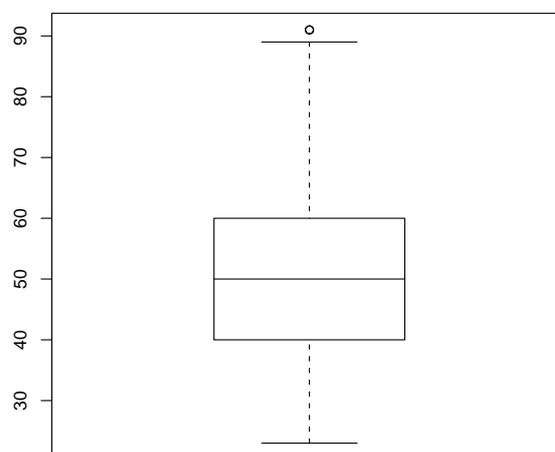


Figura 2.2: Boxplot della variabile età

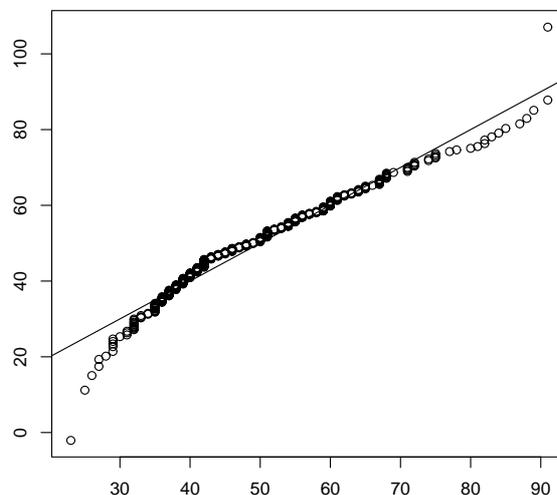


Figura 2.3: Qqplot della variabile età

<b>Min.</b>	<b>1st Qu.</b>	<b>Median</b>	<b>Mean</b>	<b>3rd Qu.</b>	<b>Max.</b>
23.00	40.00	50.00	50.44	60.00	91.00

Tabella 2.1: summary(eta)

Dall'analisi grafica si può notare che la variabile presenta una distribuzione simile alla normale con tuttavia una leggera asimmetria che emerge sia dall'istogramma che dal qqnorm. Ciò è dovuto al fatto che prima dei 18 anni, non è possibile sottoscrivere un contratto RC Auto per le autovetture.

E' interessante notare che prima dei 23 anni non vengono sottoscritti contratti; ciò si può spiegare dal fatto che, puramente da un punto di vista economico, conviene intestare l'autovettura ad uno dei genitori che avranno, presumibilmente, una classe di merito inferiore e quindi un prezzo più basso per l'assicurazione.

### 2.2.2 Sesso

Nel grafico che segue è visualizzata la distribuzione delle frequenze relative della variabile sesso.

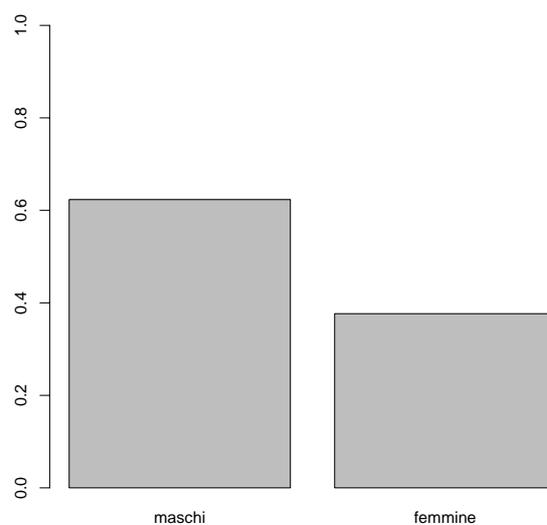


Figura 2.4: Frequenze relative della variabile sesso

Il valore 0 (maschio) presenta una frequenza maggiore rispetto al valore 1 (femmina).

### 2.2.3 Professione

Nel seguente grafico è rappresentata la distribuzione delle frequenze relative della variabile professione.

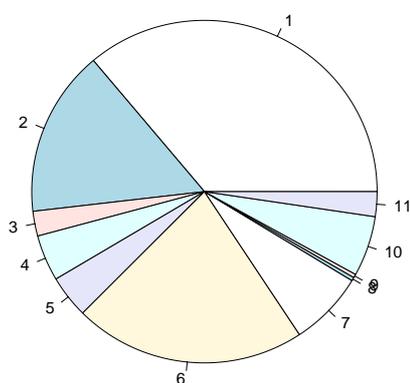


Figura 2.5: Pieplot della variabile professione

<b>id</b>	<b>Professione</b>	<b>Freq. Relativa</b>
1	dipendente	0.360000000
2	casalinga	0.156666667
3	imprenditore	0.023333333
4	commerciante	0.043333333
5	lib. professionista	0.040000000
6	pensionato	0.216666667
7	artigiano	0.070000000
8	medico	0.003333333
9	ag. rappresentante	0.003333333
10	insegnante	0.056666667
11	altro	0.023333333

Tabella 2.2: professione

I valori con frequenza maggiore per la variabile professione sono 1, 2 e 6, rispettivamente dipendente, casalinga e pensionato/a, come si può vedere anche dalla tabella sottostante (Tabella 2.2), la quale riporta la categoria professionale, il relativo numero di riferimento nella Figura 2.5 e le frequenze relative.

### 2.2.4 Cavalli fiscali (hp)

L'analisi grafica della variabile hp è riportata nei grafici che seguono.

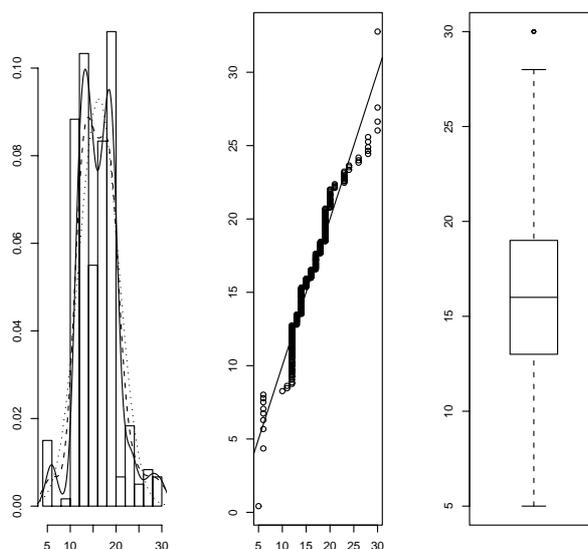


Figura 2.6: Istogramma, Q-QPlot e Boxplot della variabile hp

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
5.00	13.00	16.00	16.21	19.00	30.00

Tabella 2.3: summary(hp)

La variabile hp presenta una distribuzione a campanile compatibile con la distribuzione gaussiana, come si può vedere dal boxplot, ma si nota una cosa interessante, messa in evidenza soprattutto dall'istogramma, ovvero la presenza di un picco nelle frequenze in corrispondenza dei valori più bassi dei cavalli

fiscali delle autovetture. Ciò è dovuto al fatto che tali autovetture (con 5 e 6 cavalli fiscali) sono utilizzate molto dai pensionati, i quali, a loro volta, sono una categoria professionale con una discreta frequenza relativa. Più precisamente, della variabile professione, sono il valore con la seconda maggiore frequenza relativa.

### 2.2.5 Sinistri

Una variabile “importante” in questo dataset è la variabile sinistri (che indica il numero di sinistri registrati negli ultimi cinque anni) in quanto assume solo i valori 0, 1 e 2.

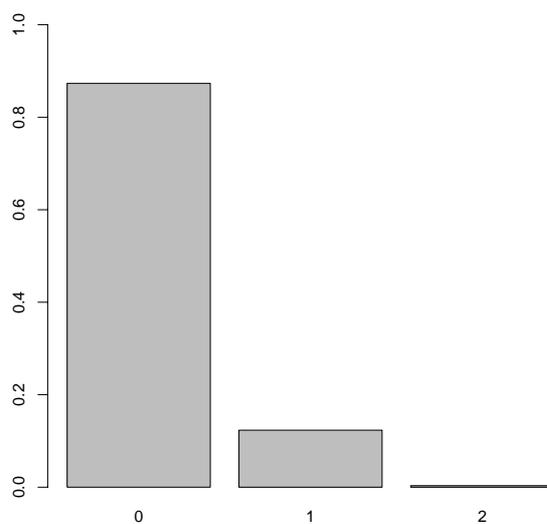


Figura 2.7: Frequenze relative della variabile sinistri

<b>Min.</b>	<b>1st Qu.</b>	<b>Median</b>	<b>Mean</b>	<b>3rd Qu.</b>	<b>Max.</b>
0.00	0.00	0.00	0.13	0.00	2.00

Tabella 2.4: summary(sin)

Decisamente il valore che si presenta con frequenza maggiore, per la variabile  $\text{sin}$  è 0. Mentre il valore 2 si presenta solo in un singolo caso.

Questo indica sia la “virtuosità” degli assicurati, ma anche il comportamento della compagnia assicurativa che tende a “liberarsi” dei maggiori sinistrati, attuando quella che è definita, con terminologia tecnica, una “pulizia di portafoglio”.

### 2.2.6 Classe di merito

L’analisi grafica della variabile  $\text{cl}$  è riportata nei grafici che seguono.

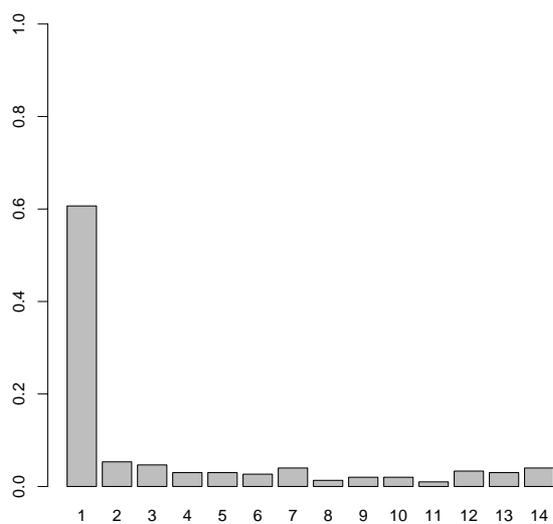


Figura 2.8: Frequenze relative della variabile classe di merito

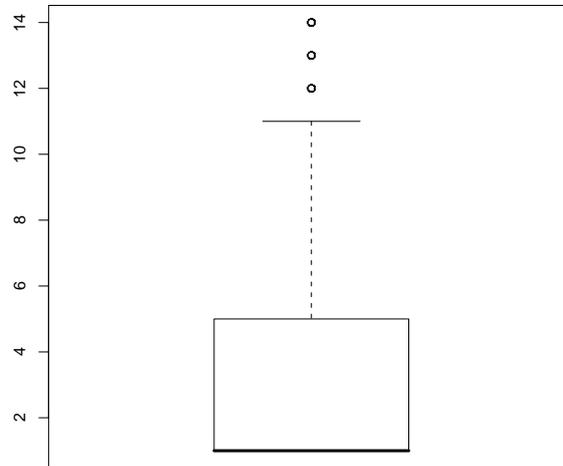


Figura 2.9: Boxplot della variabile classe di merito

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.00	1.00	1.00	3.51	5.00	14.00

Tabella 2.5: summary(cl)

Il valore che si presenta con più frequenza è 1, anche se si nota una certa omogeneità nella frequenza degli altri valori che può assumere la variabile cl.

## 2.3 Analisi temporali

Si procede ora ad un'analisi di tipo temporale, nella quale si mettono in relazione le variabili condizionatamente all'anno.

In Figura 2.10 si analizza il comportamento delle variabili età, hp e sinistri nei tre anni presi in considerazione.

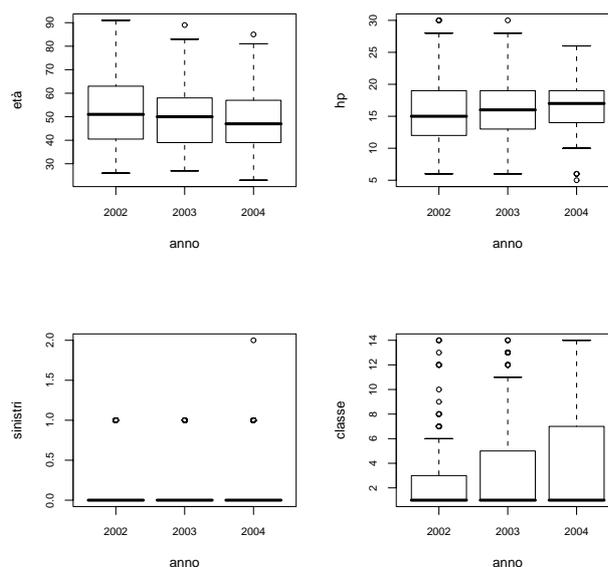


Figura 2.10: Boxplot della variabile anno relazionata con le variabili età, hp, sinistri e classe di merito

Le variabili età ed hp sono caratterizzate da una certa omogeneità nei tre anni anche se, rispettivamente, si delinea un andamento decrescente per età (probabilmente sempre più giovani iniziano ad intestare a se stessi l'autovettura) e crescente per hp (significa che vengono catalogate macchine di una cilindrata sempre maggiore).

Per quanto riguarda le variabili sin e cl, si vede chiaramente come entrambe siano concentrate sul valore 1, che indica rispettivamente il numero di sinistri negli ultimi cinque anni e la classe di merito, anche se quest'ultima, dal 2002 al 2004, tende ad includere classi di merito più "distanti" dalla prima.

Si potrebbe concludere ipotizzando il fatto che, negli ultimi anni, persone sempre più giovani (e quindi con classe di merito maggiore) hanno intestato autovetture

(di cilindrata maggiore, nella media) a proprio nome.

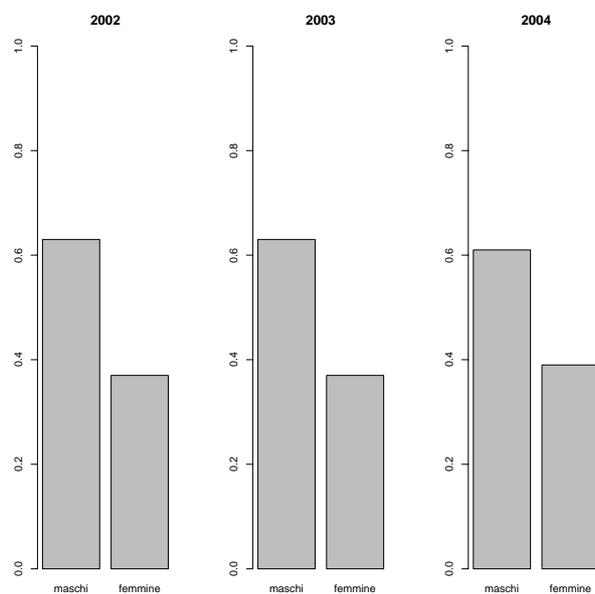


Figura 2.11: Barplot della variabile sesso nei tre anni

Dai barplot in Figura 2.11 si nota che la frequenza dei valori della variabile sesso è simile nei tre anni.

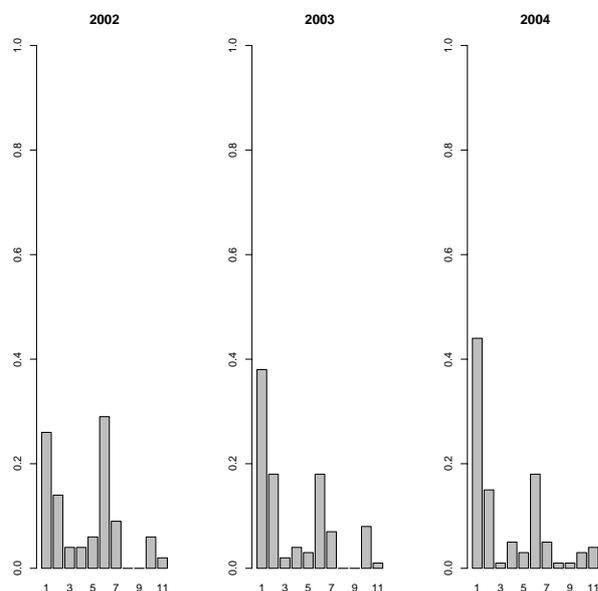


Figura 2.12: Barplot della variabile professione nei tre anni

Dai barplot in Figura 2.12 si evidenzia un andamento crescente della professione 'dipendente' nel corso degli anni ed una diminuzione della frequenza per la categoria 'pensionato'. Le altre professioni mantengono una frequenza relativamente omogenea negli anni considerati.

## 2.4 Alcune relazioni tra le variabili

E' interessante analizzare anche altre relazioni tra le variabili, per denotare eventuali legami. Con i boxplot si possono visualizzare le distribuzioni dell'età, condizionatamente alle altre variabili.

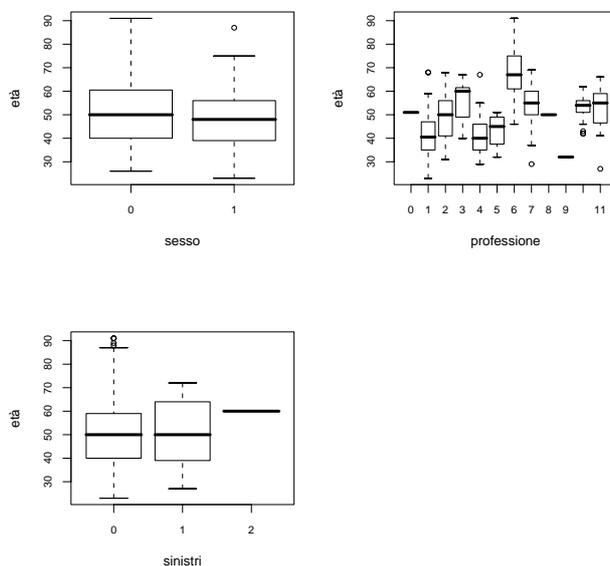


Figura 2.13: Boxplot della variabile età relazionata con le variabili sesso, professione e sinistri

Dalla Figura 2.13 si può notare che i maschi che sottoscrivono un contratto RC Auto sono un po' più "vecchi" rispetto alle femmine, anche se l'età media è pressochè identica. Invece sembra che l'età non influisca sul numero di incidenti, poichè il valore medio dell'età è lo stesso per i valori 0 e 1 della variabile sin (il boxplot per il valore 2 è degenere essendo quel valore rappresentato da un singolo dato).

E' interessante sottolineare il fatto che i valori delle professioni stanno tutti, più o meno, attorno al valore 50 della variabile età, ad esclusione della categoria 'pensionati'.

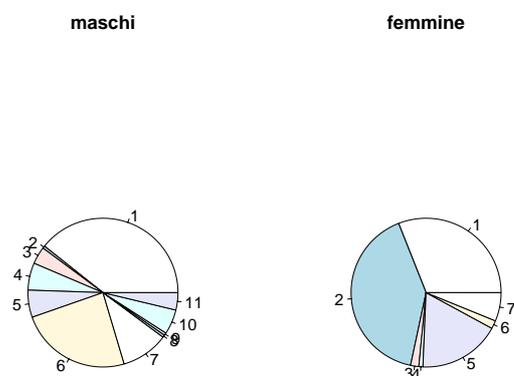


Figura 2.14: Pieplot della variabile professione divisa per sesso

Nella Figura 2.14 sono illustrate le categorie professionali in relazione al sesso. Si nota come il valore 'dipendente' sia molto presente in entrambi i grafici. Il secondo valore con frequenza maggiore nei maschi è il 'pensionato' mentre nelle femmine è la 'casalinga'.

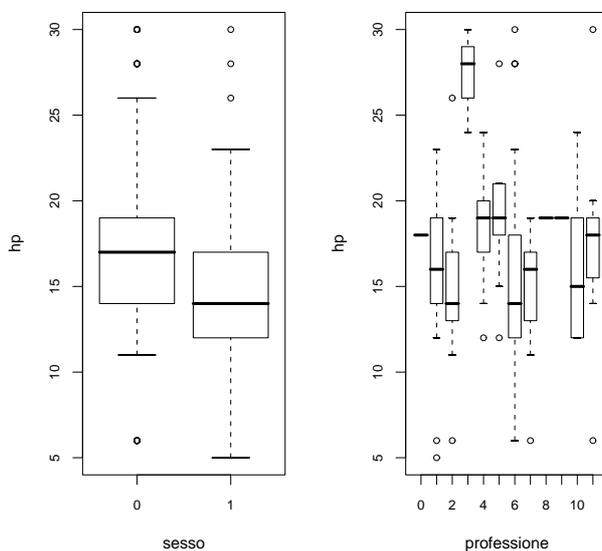


Figura 2.15: Boxplot della variabile hp relazionata con le variabili sesso e professione

Dalla Figura 2.15 si evidenzia il fatto che i maschi tendono ad acquistare autovetture con cavalli fiscali maggiori.

Potrebbe essere interessante vedere il numero di sinistri registrati negli ultimi cinque anni in relazione alle variabili età, hp e classe utilizzando il box-plot.

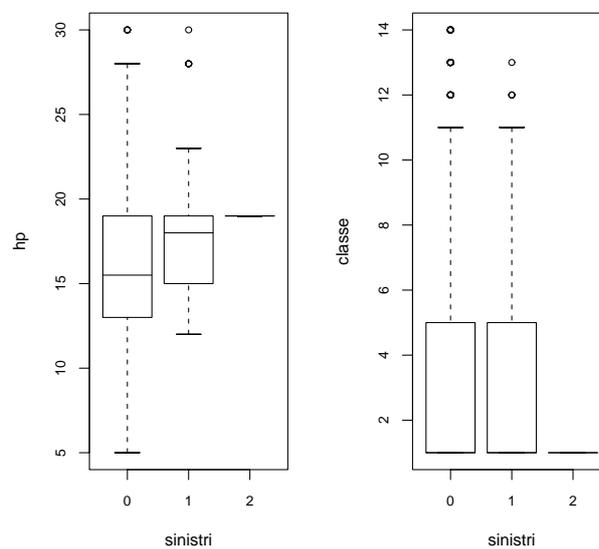


Figura 2.16: Relazione della variabile sinistri con le variabili hp e cl

Dalla Figura 2.16 è interessante notare il fatto che i cavalli fiscali sono, in media, maggiori quando la variabile sin assume valore 1.

Uno strumento utile per rappresentare le variabili quantitative è fornito dalla matrice dei diagrammi di dispersione.

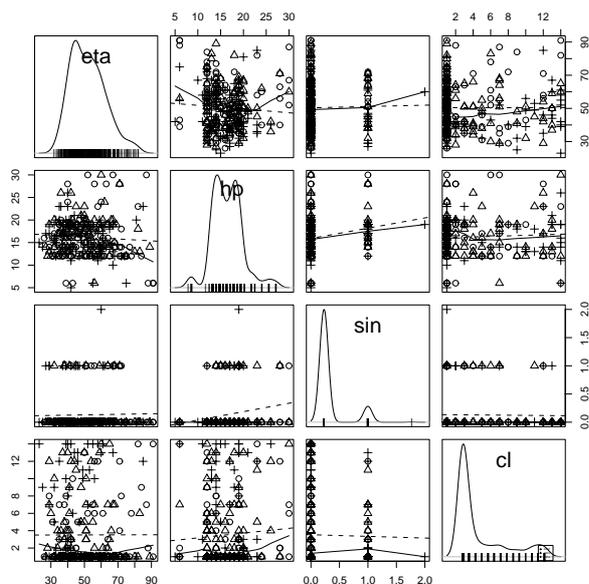


Figura 2.17: Scatterplot

Questo screenshot (Figura 2.17) mostra i diagrammi a dispersione di tutte le variabili numeriche messe a confronto con le altre. La distribuzione marginale stimata con metodo kernel è rappresentata nei grafici sulla diagonale della matrice. Nei singoli diagrammi a dispersione è riportata sia la retta di regressione sia la curva ottenuta con una regressione non parametrica.

I valori degli indici di correlazione sono riportati nella tabella che segue.

	età	hp	sin	cl
età	1.000000000	-0.06862527	0.01988694	0.001664302
hp	-0.068625268	1.000000000	0.18460868	0.057907527
sin	0.019886941	0.18460868	1.000000000	-0.016355824
cl	0.001664302	0.05790753	-0.01635582	1.000000000

Tabella 2.6: cor(dati)

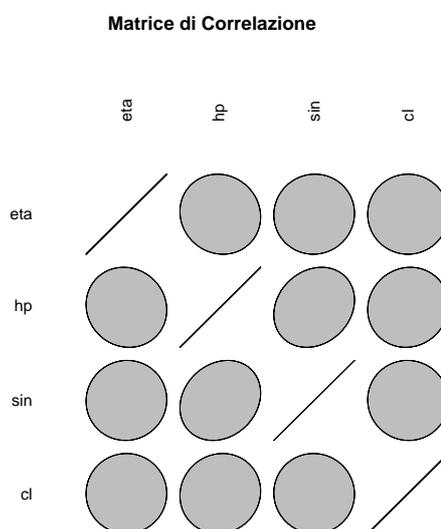


Figura 2.18: Matrice di Correlazione

Dalla Tabella 2.6 e dalla Figura 2.18 si vede che i valori degli indici di correlazione sono piuttosto bassi.

Nel capitolo successivo cerchiamo di studiare le relazioni tra tutte le variabili tramite l'ACOMU (Analisi delle CORRispondenze MULTiple).

## 2.5 Routines - R

```

> dati<-read.table("...",header=TRUE)
> attach(dati)
> colnames(dati)
[1] "anno" "sex"  "eta"  "prof" "hp"   "sin"  "cl"
> dim(dati)
[1] 300   7

#analisi univariata
#analisi variabile "età"
> hist(eta,xlab="",ylab="",main="",prob=TRUE)
> lines(x=1:300,y=dnorm(1:300,mean=mean(eta),sd=sd(eta)),lty=3)
> lines(density(eta,adjust = 1))
> lines(density(eta,adjust = 1.5),lty=2)
> legend(65,0.025,c("Densità Kernel(1)","Densità Kernel(1.5)",
"Densità Normale"), lty=1:3,bty="n")
> boxplot(eta,xlab="",ylab="")
> qqplot(eta,rnorm(10000,mean=mean(eta),sd=sd(eta)),ylab="",xlab="")
> abline(0,1)
> summary(eta)

#analisi variabile "sesso"
> sesso<-c("maschi","femmine")
> sm<-sum(sex==0)
> sf<-sum(sex==1)
> tot<-300
> smr<-sm/tot
> sfr<-sf/tot
> freq<-c(smr,sfr)
> barplot(freq,names.arg=sesso,main="",ylim=c(0,1))

```

```
#analisi variabile "professione"
> tot<-300
> dipendente<-sum(prof==1)
> fdip<-dipendente/tot
> casalinga<-sum(prof==2)
> fcas<-casalinga/tot
> imprenditore<-sum(prof==3)
> fimp<-imprenditore/tot
> commerciante<-sum(prof==4)
> fcom<-commerciante/tot
> lib.professionista<-sum(prof==5)
> flib<-lib.professionista/tot
> pensionato<-sum(prof==6)
> fpen<-pensionato/tot
> artigiano<-sum(prof==7)
> fart<-artigiano/tot
> medico<-sum(prof==8)
> fmed<-medico/tot
> agente.rappr<-sum(prof==9)
> fage<-agente.rappr/tot
> insegnante<-sum(prof==10)
> fins<-insegnante/tot
> altro<-sum(prof==11)
> falt<-altro/tot
> professione<-cbind(dipendente,casalinga,imprenditore,commerciante,
lib.professionista,pensionato,artigiano,medico,agente.rappr,
insegnante,altro)
> fprof<-c(fdip,fcas,fimp,fcom,flib,fpen,
fart,fmed,fage,fins,falt)
> pie(fprof)
> fprof
```

```
#analisi variabile "hp" (cavalli fiscali)
> par(mfcol=c(1,3))
> hist(hp,xlab="",ylab="",main="",prob=TRUE)
> lines(x=1:300,y=dnorm(1:300,mean=mean(hp),sd=sd(hp)),lty=3)
> lines(density(hp,adjust = 1))
> lines(density(hp,adjust = 1.5),lty=2)
> qqplot(hp,rnorm(10000,mean=mean(hp),sd=sd(hp)),ylab="",xlab="")
> abline(0,1)
> boxplot(hp,xlab="",ylab="")
>summary(hp)
```

```
#analisi variabile "sin"
> sinistri<-c("0","1","2")
> tot<-300
> sin0<-sum(sin==0)
> sin1<-sum(sin==1)
> sin2<-sum(sin==2)
> fsin0<-sin0/tot
> fsin1<-sin1/tot
> fsin2<-sin2/tot
> freq<-c(fsin0,fsin1,fsin2)
> barplot(freq,names.arg=sinistri,main="",ylim=c(0,1))
> summary(sin)
```

```
#analisi variabile "cl"
> classe<-c("1","2","3","4","5","6","7",
"8","9","10","11","12","13","14")
> tot<-300
> cl1<-sum(cl==1)
> f1<-cl1/tot
```

```
> c12<-sum(c1==2)
> f2<-c12/tot
> c13<-sum(c1==3)
> f3<-c13/tot
> c14<-sum(c1==4)
> f4<-c14/tot
> c15<-sum(c1==5)
> f5<-c15/tot
> c16<-sum(c1==6)
> f6<-c16/tot
> c17<-sum(c1==7)
> f7<-c17/tot
> c18<-sum(c1==8)
> f8<-c18/tot
> c19<-sum(c1==9)
> f9<-c19/tot
> c110<-sum(c1==10)
> f10<-c110/tot
> c111<-sum(c1==11)
> f11<-c111/tot
> c112<-sum(c1==12)
> f12<-c112/tot
> c113<-sum(c1==13)
> f13<-c113/tot
> c114<-sum(c1==14)
> f14<-c114/tot
> freq<-c(f1,f2,f3,f4,f5,f6,f7,f8,f9,f10,f11,f12,f13,f14)
> barplot(freq,names.arg=classe,main="",ylim=c(0,1))
> boxplot(c1)
> summary(c1)
```

## 40 CAPITOLO 2. ANALISI ESPLORATIVA DEI CONTRATTI RC AUTO

```
#analisi temporali
> par(mfrow=c(2,2))
> boxplot(eta~anno,main="",xlab="anno",ylab="età",cex.axis=0.8)
> boxplot(hp~anno,main="",xlab="anno",ylab="hp",cex.axis=0.8)
> boxplot(sinù~anno,main="",xlab="anno",ylab="sinistri",cex.axis=0.8)
> boxplot(cl~anno,main="",xlab="anno",ylab="classe",cex.axis=0.8)
> m04<-sum(sex[1:100]==0)
> f04<-sum(sex[1:100]==1)
> m03<-sum(sex[101:200]==0)
> f03<-sum(sex[101:200]==1)
> m02<-sum(sex[201:300]==0)
> f02<-sum(sex[201:300]==1)
> mf04<-m04/100
> ff04<-f04/100
> mf03<-m03/100
> ff03<-f03/100
> mf02<-m02/100
> ff02<-f02/100
> freq<-c(mf04,ff04)
> freq2<-c(mf03,ff03)
> freq3<-c(mf02,ff02)
> par(mfcol=c(1,3))
> barplot(freq3,names.arg=c("maschi","femmine"),ylim=c(0,1),main="2002")
> barplot(freq2,names.arg=c("maschi","femmine"),ylim=c(0,1),main="2003")
> barplot(freq,names.arg=c("maschi","femmine"),ylim=c(0,1),main="2004")
> p104<-sum(prof[1:100]==1)
> p204<-sum(prof[1:100]==2)
> p304<-sum(prof[1:100]==3)
> p404<-sum(prof[1:100]==4)
> p504<-sum(prof[1:100]==5)
> p604<-sum(prof[1:100]==6)
```

```
> p704<-sum(prof [1:100]==7)
> p804<-sum(prof [1:100]==8)
> p904<-sum(prof [1:100]==9)
> p1004<-sum(prof [1:100]==10)
> p1104<-sum(prof [1:100]==11)
> p103<-sum(prof [101:200]==1)
> p203<-sum(prof [101:200]==2)
> p303<-sum(prof [101:200]==3)
> p403<-sum(prof [101:200]==4)
> p503<-sum(prof [101:200]==5)
> p603<-sum(prof [101:200]==6)
> p703<-sum(prof [101:200]==7)
> p803<-sum(prof [101:200]==8)
> p903<-sum(prof [101:200]==9)
> p1003<-sum(prof [101:200]==10)
> p1103<-sum(prof [101:200]==11)
> p102<-sum(prof [201:300]==1)
> p202<-sum(prof [201:300]==2)
> p302<-sum(prof [201:300]==3)
> p402<-sum(prof [201:300]==4)
> p502<-sum(prof [201:300]==5)
> p602<-sum(prof [201:300]==6)
> p702<-sum(prof [201:300]==7)
> p802<-sum(prof [201:300]==8)
> p902<-sum(prof [201:300]==9)
> p1002<-sum(prof [201:300]==10)
> p1102<-sum(prof [201:300]==11)
> fp104<-sum(prof [1:100]==1)/100
> fp204<-sum(prof [1:100]==2)/100
> fp304<-sum(prof [1:100]==3)/100
> fp404<-sum(prof [1:100]==4)/100
```

42 *CAPITOLO 2. ANALISI ESPLORATIVA DEI CONTRATTI RC AUTO*

```
> fp504<-sum(prof [1:100]==5)/100
> fp604<-sum(prof [1:100]==6)/100
> fp704<-sum(prof [1:100]==7)/100
> fp804<-sum(prof [1:100]==8)/100
> fp904<-sum(prof [1:100]==9)/100
> fp1004<-sum(prof [1:100]==10)/100
> fp1104<-sum(prof [1:100]==11)/100
> fp103<-sum(prof [101:200]==1)/100
> fp203<-sum(prof [101:200]==2)/100
> fp303<-sum(prof [101:200]==3)/100
> fp403<-sum(prof [101:200]==4)/100
> fp503<-sum(prof [101:200]==5)/100
> fp603<-sum(prof [101:200]==6)/100
> fp703<-sum(prof [101:200]==7)/100
> fp803<-sum(prof [101:200]==8)/100
> fp903<-sum(prof [101:200]==9)/100
> fp1003<-sum(prof [101:200]==10)/100
> fp1103<-sum(prof [101:200]==11)/100
> fp102<-sum(prof [201:300]==1)/100
> fp202<-sum(prof [201:300]==2)/100
> fp302<-sum(prof [201:300]==3)/100
> fp402<-sum(prof [201:300]==4)/100
> fp502<-sum(prof [201:300]==5)/100
> fp602<-sum(prof [201:300]==6)/100
> fp702<-sum(prof [201:300]==7)/100
> fp802<-sum(prof [201:300]==8)/100
> fp902<-sum(prof [201:300]==9)/100
> fp1002<-sum(prof [201:300]==10)/100
> fp1102<-sum(prof [201:300]==11)/100
> freq1<-c(fp104,fp204,fp304,fp404,fp504,
fp604,fp704,fp804,fp904,fp1004,fp1104)
```

```

> freq2<-c(fp103,fp203,fp303,fp403,fp503,
fp603,fp703,fp803,fp903,fp1003,fp1103)
> freq3<-c(fp102,fp202,fp302,fp402,fp502,
fp602,fp702,fp802,fp902,fp1002,fp1102)
> barplot(freq3,names.arg=c(1,2,3,4,5,6,7,8,9,10,11),
ylim=c(0,1),main="2002")
> barplot(freq2,names.arg=c(1,2,3,4,5,6,7,8,9,10,11),
ylim=c(0,1),main="2003")
> barplot(freq1,names.arg=c(1,2,3,4,5,6,7,8,9,10,11),
ylim=c(0,1),main="2004")

#alcune relazioni tra variabili
> par(mfrow=c(2,2))
> boxplot(eta~sex,main="",xlab="sesso",ylab="età",cex.axis=0.8)
> boxplot(eta~prof,main="",xlab="professione",ylab="età",cex.axis=0.8)
> boxplot(eta~sin,main="",xlab="sinistri",ylab="età",cex.axis=0.8)
> d<-scan("d:\\tesi\\database_axa_02.03.04(nolabel).txt")
> x<-matrix(data=d,ncol=7,byrow=TRUE)

> #crea matrice femmine
> xf<-matrix(data=NA,ncol=7,nrow=300)
> for(i in 1:300)
+ {
+ if(x[i,2]==1)
+ xf[i,]<-x[i,]
+ }
> xf<-na.omit(xf)
> xf<-as.matrix(xf)
> xf<-xf[1:113,]

> #crea matrice maschi

```

## 44 CAPITOLO 2. ANALISI ESPLORATIVA DEI CONTRATTI RC AUTO

```
> xm<-matrix(data=NA,ncol=7,nrow=300)
> for(i in 1:300)
+ {
+ if(x[i,2]==0)
+ xm[i,]<-x[i,]
+ }
> xm<-na.omit(xm)
> xm<-as.matrix(xm)
> xm<-xm[1:187,]
> pm1<-sum(xm[,4]==1)/187
> pm2<-sum(xm[,4]==2)/187
> pm3<-sum(xm[,4]==3)/187
> pm4<-sum(xm[,4]==4)/187
> pm5<-sum(xm[,4]==5)/187
> pm6<-sum(xm[,4]==6)/187
> pm7<-sum(xm[,4]==7)/187
> pm8<-sum(xm[,4]==8)/187
> pm9<-sum(xm[,4]==9)/187
> pm10<-sum(xm[,4]==10)/187
> pm11<-sum(xm[,4]==11)/187
> pf1<-sum(xf[,4]==1)/113
> pf2<-sum(xf[,4]==2)/113
> pf3<-sum(xf[,4]==3)/113
> pf4<-sum(xf[,4]==4)/113
> pf5<-sum(xf[,4]==5)/113
> pf6<-sum(xf[,4]==6)/113
> pf7<-sum(xf[,4]==7)/113
> pf8<-sum(xf[,4]==8)/113
> pf9<-sum(xf[,4]==9)/113
> pf10<-sum(xf[,4]==10)/113
> pf11<-sum(xf[,4]==11)/113
```

```
> fr1<-c(pm1,pm2,pm3,pm4,pm5,pm6,pm7,pm8,pm9,pm10,pm11)
> fr2<-c(pf1,pf2,pf3,pf4,pf5,pf6,pf7,pf8,pf9,pf10,pf11)
> par(mfcol=c(1,2))
> fr2<-fr2[-c(3,8,9,11)]
> pie(fr1,main="maschi")
> pie(fr2,main="femmine")
> par(mfrow=c(1,2))
> boxplot(hp~sex,main="",xlab="sesso",ylab="hp",cex.axis=0.8)
> boxplot(hp~prof,main="",xlab="professione",ylab="hp",cex.axis=0.8)
> par(mfrow=c(1,2))
> boxplot(hp~sin,main="",xlab="sinistri",ylab="hp",cex.axis=0.8)
> boxplot(cl~sin,main="",xlab="sinistri",ylab="classe",cex.axis=0.8)
> scatterplot.matrix(dati,groups=anno,col=rep("black",7),cex=1.2)
> cor(dati)
> library(ellipse)
> plotcorr(cor(dati),main="Matrice di Correlazione")
```



# Capitolo 3

## Analisi Multivariata

### 3.1 Ricodifica delle variabili del database

Allo scopo di effettuare un'analisi esplorativa grafica - anche ai fini della costruzione di modelli - che coinvolge i dati a disposizione in questa tesi, è necessario categorizzare alcune delle variabili. Ciò allo scopo di facilitare la maneggevolezza della base dei dati e di utilizzare opportune procedure di analisi multivariata. Per la costruzione di variabili categoriche si utilizza il comando di R `factor()` con la definizione delle etichette (`labels`).

La variabile età, opportunamente ricodificata, viene raggruppata in tre classi: valore 0 (età compresa tra i 18 e i 30 anni), valore 1 (età compresa tra i 31 e i 50 anni) e valore 2 (per le età maggiori di 50 anni). La variabile professione, invece, viene ad assumere valori 1 (dipendente), 2 (casalinga), 3 (pensionato/a) e 0 (altro). Quest'ultima codifica è stata definita selezionando le categorie professionali più frequenti (appunto dipendente, casalinga e pensionato/a) e definendo quelle meno frequenti come una categoria a parte (altro).

## 3.2 Analisi delle Corrispondenze

Un profilo generale dei clienti della compagnia assicurativa lo si ottiene incrociando tutte le variabili presenti nella base di dati. Per fare ciò è necessario introdurre una tecnica statistica più raffinata: l'*Analisi delle Corrispondenze Multiple* (ACOMU, Greenacre 1993; Analisi esplorativa di dati multidimensionali, Fabbris L., tra i molti riferimenti recenti che si potrebbero dare).

L'analisi delle corrispondenze multiple è una tecnica statistica molto potente che può essere utilizzata quando si devono analizzare matrici di dati in cui vi siano in prevalenza variabili categoriali (cioè che non si manifestino come serie progressive di numeri, ma come qualità 'discrete', nettamente separate le une dalle altre).

Interessante è il prodotto grafico dell'analisi delle corrispondenze, che permette di rappresentare - su un numero ridotto di assi - le relazioni intercorrenti tra le modalità delle variabili analizzate. Su questa base, in seguito, si potranno formulare ipotesi interpretative, che potranno essere verificate con ulteriori analisi statistiche, utilizzando tecniche più potenti sotto l'aspetto inferenziale (cioè della capacità di prestarsi a generalizzazioni), ma inadatte a gestire grandi matrici di dati.

Dal punto di vista grafico, il fine è quello di creare dei raggruppamenti di categorie delle variabili utilizzate che abbiano, rispetto ad alcune variabili selezionate, il massimo di somiglianza tra loro ed il massimo di diversità tra i gruppi di categorie delle variabili così individuate.

L'ACOMU è un'analisi di tipo fattoriale che ha come scopo quello di individuare dimensioni soggiacenti alla struttura dei dati, dimensioni intese a riassumere l'intreccio di relazioni di 'interdipendenza' tra le variabili originarie. Il punto di partenza è una tabella di contingenza 'particolare', nota come tavola di Burt, in cui le modalità riga coincidono con le modalità colonna. Effettuando l'analisi delle corrispondenze multiple, si vedranno le relazioni tra le variabili categoriche rappresentate su un unico grafico ad alta risoluzione di facile lettura.

Nell'ACOMU possiamo distinguere due tipologie di variabili:

- *attive*, cioè variabili che entrano direttamente nell'analisi concorrendo alla formazione degli assi fattoriali;
- *supplementari* o *illustrative*, cioè variabili di tipo 'passivo' che sono escluse dalla fase di estrazione dei fattori, ma si utilizzano successivamente considerando la loro posizione sugli assi fattoriali come ausilio per la loro interpretazione.

Tipicamente nell'ACOMU le quote di inerzia (cioè il massimo della variabilità della matrice dei dati) spiegate dai primi fattori non sono molto elevate, specie se confrontate con quelle dell'ACS (Analisi delle Corrispondenze Semplici), a causa del grande numero di modalità, e di conseguenza di variabilità, presente nella matrice dei dati.

Talora può risultare interessante ripetere l'ACM dopo aver ripulito la matrice dei dati compattando le categorie che nella prima analisi non presentano contributi assoluti abbastanza elevati sui fattori considerati; in questo modo infatti si riduce l'inerzia globale.

Al termine i fattori da considerare si determinano in base alla loro rilevanza, ovvero alla quota di inerzia totale che spiegano.

L'ACOMU è implementata nella funzione `mca()` della libreria MASS (Venables e Ripley, 1999). L'obiettivo dell'ACOMU è simile a quello dell'ACP (Analisi in Componenti Principali) poiché si prefigge di trovare un sottospazio di rappresentazione per variabili categoriche invece che continue. Lo spazio originario ha dimensione uguale alla somma delle modalità delle variabili categoriche ed il sottospazio di rappresentazione è definito, anche in questa analisi, da assi fattoriali (o fattori). I fattori sono costruiti considerando l'ipertavola di contingenza - tavola di Burt - tra le variabili (attive). L'interpretazione degli assi fattoriali si effettua in maniera analoga all'ACP, con la differenza teorica dovuta al fatto che non è possibile in questo caso ricorrere alle correlazioni tra assi delle variabili (categoriche) e fattori, ma si considerano le coordinate dei punti modalità negli assi fattoriali. La proiezione dei punti individuo sugli assi fattoriali è utile per la successiva analisi dei gruppi.

### 3.3 Applicazione ai dati RC Auto

Si presenta di seguito l'applicazione dell'ACOMU ai dati presentati nel capitolo 2. Si sono utilizzati, a tale scopo, i comandi R presentati nel paragrafo 3.4.

Il grafico che segue contiene la rappresentazione grafica delle modalità attive sul piano definito dai primi due assi delle soluzioni ottenute con l'ACOMU.

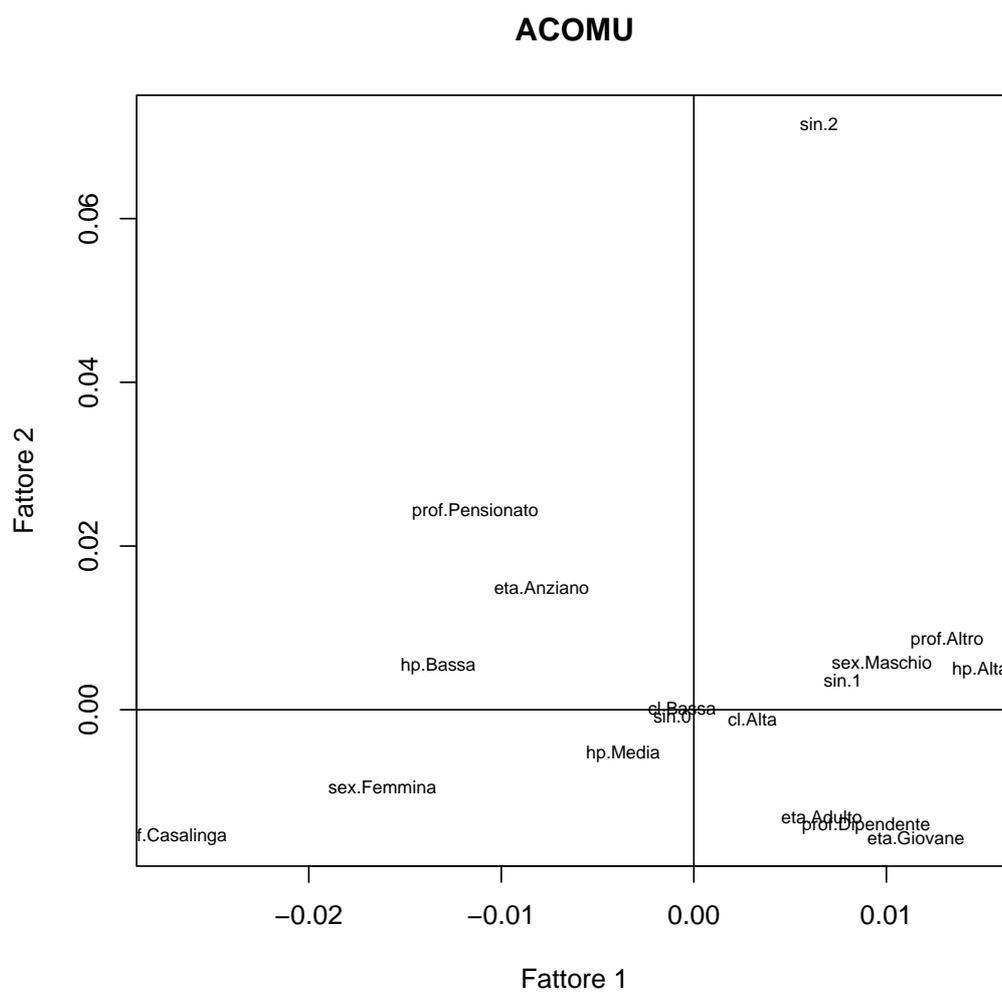


Figura 3.1: ACOMU

Si intuisce che il primo fattore rappresenta l'associazione tra il sesso, i cavalli fiscali ed il numero di sinistri. Infatti a destra si trovano principalmente i clienti maschi (*sex.Maschio*), con automobili più potenti (*hp.alta*) e meno 'virtuosi', ossia con sinistri pari a uno; mentre a sinistra quelli di sesso femminile (*sex.femmina*), con automobili meno potenti (*hp.bassa* o *hp.media*) e più virtuosi (*sin.0*). Pertanto, il sesso, il numero di sinistri ed i cavalli fiscali si dispongono prevalentemente lungo questo asse nei versi opposti: le femmine sono più virtuose ed hanno automobili meno potenti, e viceversa.

Il secondo fattore prende invece significato dalla professione associata all'età. Ad esempio, in alto a sinistra si trovano i pensionati, con età elevata (*eta.Anziano*), che rispetto al primo asse tendono ad avere auto di bassa cilindrata (*hp.bassa*); invece a destra le professioni 'altro' (che includono, ad esempio, le categorie libero professionista ed agente rappresentante), tipiche dei maschi con auto più potenti. Nella parte inferiore del grafico, invece, si dispongono a sinistra le casalinghe, caratterizzate da età variabili, mentre a destra i dipendenti. La variabile classe di merito invece è concentrata attorno all'origine. Ciò può significare che essa non è associata a nessuna delle altre variabili considerate.

Infine si osserva che per la posizione anomala del numero di sinistri pari a 2, si intuisce come questo valore potrebbe essere eliminato dall'analisi.

E' interessante notare come tali conclusioni siano molto vicine a quelle delineate al termine delle analisi esplorative, il che ci porta a poter definire i ritratti dei clienti della compagnia assicurativa.

Nel grafico sottostante si riporta la rappresentazione grafica delle modalità illustrative ottenute con l'ACOMU.

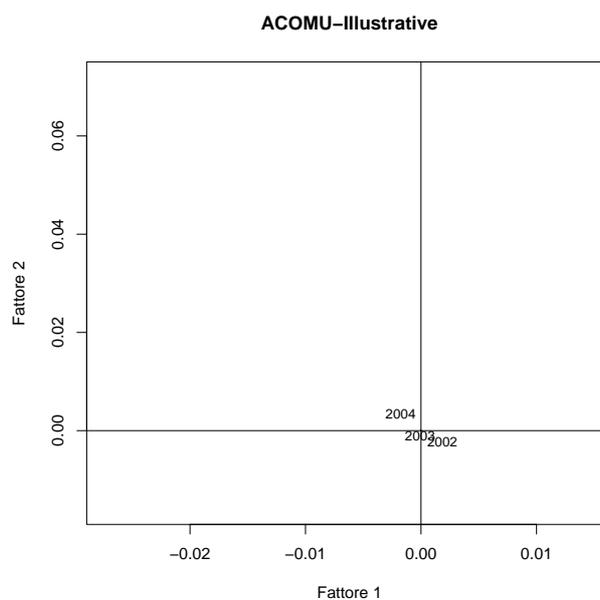


Figura 3.2: ACOMU - Illustrative

Dalla Figura 3.2 si può definire il fatto che non vi è una differenza considerevole negli anni presi in considerazione. Ovvero, la variabile anno, non incide in maniera significativa su quelle che sono le relazioni tra le variabili.

### 3.4 Routines - R

```
> dati<-read.table(...,header=T)
> attach(dati)
> estremi.classi=round(quantile(eta,prob=c(0,0.33,0.66,1)),2)
> etaclassi=findInterval(eta,estremi.classi,all.inside=TRUE)
> eta=factor(etaclassi,labels=c("Giovane","Adulto","Anziano"))
> sex=factor(sex,labels=c("Maschio","Femmina"))
> prof=factor(prof,labels=c("Altro","Dipendente","Casalinga","Pensionato"))
> cl=factor(cl,labels=c("Bassa","Alta"))
> hp=factor(hp,labels=c("Bassa","Media","Alta"))
> sin=factor(sin,labels=c("0","1","2"))
> dat=data.frame(eta,sex,prof,cl,hp,sin)
> library(MASS)
> acomu=mca(dat, nf = 2)
> plot(acomu$cs,type="n",main="ACOMU",xlab="Fattore 1",ylab="Fattore 2")
> text(acomu$cs,rownames(acomu$cs),cex=0.7)
> abline(v=0,h=0)
> plot(acomu$cs,type="n",main="ACOMU-Illustrative",xlab="Fattore 1",
ylab="Fattore 2")
> illus=predict(acomu, newdata=as.data.frame(factor(anno)),type="factor")
> text(illus[1,1],illus[1,2],label="2004",cex=0.8)
> text(illus[2,1],illus[2,2],label="2003",cex=0.8)
> text(illus[3,1],illus[3,2],label="2002",cex=0.8)
> abline(v=0,h=0)
```

## Capitolo 4

### Conclusioni

Nella tesi sono stati analizzati i dati relativi ai contratti assicurativi RC Auto. Nell'analisi preliminare si erano notate alcune relazioni tra le variabili, come ad esempio l'importanza del sesso, della professione e dei cavalli fiscali (cfr. ad esempio il boxplot in Figura 2.15). Inoltre la variabile anno non aveva mostrato nessuna influenza, in quanto i risultati erano costanti nei tre anni considerati. Le relazioni studiate nella prima parte sono state 'confermate' dall'analisi delle corrispondenze multiple. Con questo strumento si è notato come, relazionando tutte le variabili tra di loro, si sia arrivati a poter delineare dei profili dei clienti dell'agenzia, grazie a variabili e relazioni tra esse più discriminanti di altre, come per esempio le variabili sesso, professione ed hp. Inoltre si può affermare che la variabile classe di merito non distingue i clienti dei contratti e come, da un punto di vista temporale, queste non si modificano in modo significativo nei tre anni considerati.



# Bibliografia

## 4.1 Bibliografia

- Fabbris, L. (1983). *Analisi dei Dati Multidimensionali*. Cluep.
  - Pace, L., Salvan A. (1996). *Introduzione alla statistica [vol.1]: inferenza, verosimiglianza, modelli*. CEDAM.
  - Pace, L., Salvan A. (2001). *Introduzione alla statistica [vol.2]: statistica descrittiva*. CEDAM.
  - Bortot, P., Ventura, L., Salvan, A. (2000). *Inferenza Statistica: Applicazioni con S-PLUS e R*. CEDAM.
- 6 Greenacre, M. J. (1993). *Correspondence analysis in practice*. Academic Pres.

## 4.2 Siti Internet

- <http://cran.r-project.org>
- <http://www.axa-italia.it>
- <http://www.consumatori.it>
- <http://www.movimentoconsumatori.it>
- <http://www.quattroruote.it>

- <http://www.tuttoconsumatori.it>
- <http://www.unica.it>
- <http://www.sici.psy.unipd.it>